Journal of Computer Information Systems

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucis20

Improving IS Practical Significance through Effect Size Measures

Nik Thompson, Xuequn Wang & Richard Baskerville

JOURNAL OF COMPUTER INFORMATION SYSTEMS
https://doi.org/10.1080/08874417.2020.1837036

Improving IS Practical Significance through Effect Size Measures

Nik Thompson[a], Xuequn Wang [b], and Richard Baskerville[c]

[a]Curtin University, Bentley, Australia; [b]Murdoch University, Murdoch, Australia; [c]Georgia State University, Atlanta, Georgia, USA

ABSTRACT

Evidence-based practice in management assigns a high value to research results as a guide to practices that have been rigorously shown to be effective. To emphasize the practical relevance and outcomes for information systems research, statistical research should generally report its effect sizes. Specifying effect sizes not only reveals the utility of our results, but it also enables evidence-based practitioners to easily compare the known effects of different interventions applied in different studies. Effect size reporting has become a standard practice in many fields, however, though information systems researchers have adopted many other elements of statistical rigor, effect sizes are often overlooked. This paper surveys the current use of effect size calculations in information systems research, explains how such effects sizes are calculated, offers recommendations on when each of the different formulae is appropriate, and provides foundational work toward an index of expected effect sizes in information systems research.

KEYWORDS

Effect size; p-value; Cohen's d; statistical significance; practical significance

Introduction

It has been a long-standing and venerated belief among Information Systems (IS) researchers that statistical significance (p-values) and explained variance (R-squared) are the most essential indicators of credibility in research outcomes in probabilistic inductive-statistical research. But this belief is growing obsolete. In behavioral science research, effect size (e.g., Cohen's *d*) is growing more important as an essential indicator of credibility in research outcomes. The increasing availability of large data sets is one important reason why the importance of statistical significance has been diminished. The importance of explained variance has been diminished by the need to support evidence-based practice. The importance of effect values is increased by the need for evidence that the claimed relationships between the constructs actually *matter*. Effect size needs to be more widely reported in IS research.

The increasing attention to effect size is at least partly a consequence of the widespread interest in evidence-based practice.[1] With roots in evidence-based medicine, the management field has been developing its parallel: evidence-based management.[2] But while medical research spans a large body of clinical research with known, valuable effect sizes,[1] management research produces work with unknown or trivial effect sizes.[3] Managers who attempt evidence-based management struggle to find the evidence in the management research literature. Comparisons of effect size in the meta-analysis of the management literature is problematic because relatively few studies publish their effect sizes.

These shortcomings in reports from the IS research community are largely similar to those of the management literature. Mohajeri, Mesgari and Lee[4] observe, "The history of quantitative research in the IS field and beyond reveals not only disputes over the adequacy of statistical significance to warrant the scientific merits of research, but also pleas for drawing attention to practical significance, as well as a lack of distinction between relevance and practical significance."

The distinction between statistical significance, practical significance, and relevance is important in evidence-based practice. Statistical significance regards whether the reported results are sufficiently "discernable" in statistical terms to actually exist; practical significance regards whether the magnitude of the reported results is sufficiently "impressive" to actually matter; relevance regards whether the reported results are sufficiently "understandable" and "useful" to have impact in the solution of certain problems.[4]

Fortunately, IS researchers do not have to invent new techniques for calculating and reporting effect size. We can learn from the experience in psychology. The research community in psychology encountered this struggle early in the history of evidence-based practice. Researchers in psychology developed techniques to improve the calculating and reporting of their effect sizes, producing research that is not only scientifically rigorous but also appealing as a basis for clinical practice.[5]

In their basic form, traditional tests of significance yielding a p value presume that the null hypothesis is true in the population, and then, based on result and sample size, calculate the probability of the findings.[6] The applicability of p values alone in determining research applicability has been subject to several decades of discussion. It has been suggested that demonstrating statistical significance alone does not directly translate to practical significance, relevance, or replicability.[7] Given a large enough sample size, statistical significance in the form of a p value will *always* be found.[8]

CONTACT Nik Thompson nik.thompson@curtin.edu.au School of Management, Curtin University , Bentley, Western Australia 6102, Australia

[1]This has not always been the case. As recently as 2012 medical journal editors were calling for greater reporting of effect size in their discipline.[49] The successes and improvements in other disciplines should serve as encouragement that information systems can follow suit.

It is possible to create sensitive experiments using several ways, including increasing the sample size, e.g.,[9] Increasing the sample size has the result of yielding estimates so precise that even minuscule differences between treatments or groups become statistically significant. This issue is further confounded by what Cohen[6] described as the *nil* hypothesis problem: an all too common scenario in which researchers set the null hypothesis to be something extremely unlikely, and thus likely to be rejected.

Today, online survey panels and market research firms bring large sample behavioral research within the grasp of most researchers. Lin et al.[10] suggested that addressing the *p* value problem may increase the credibility of IS research, citing approaches taken in other disciplines to make statistical reporting more thorough. Notable in their discussion is the role of effect size as a straightforward means to provide the reader with evidence of the magnitude of any differences observed.

In this paper, we do not critique the statistical techniques in widespread use, but instead focus on actionable ways in which the practical significance of research findings can be strengthened, often with minimal effort required from the researcher. The cornerstone of this effort to improve practical significance lies in effect sizes. Effect sizes provide a measure of the size of the observed effect, independent of the influence of sample size. Many of the developments in effect size reporting and calculation stem from the work of Cohen[6,11] who explains how effect size reporting provides insights into the observed effects. For instance, a low effect size may guide researchers to better understand the context of findings which may appear statistically significant. Perhaps more interestingly, a large effect size found alongside non statistically significant results may reveal opportunities for further research into an area that may otherwise be disregarded.[12]

To summarize, the main objective of this paper is to discuss the advantages of effect size and suggest reporting effect size and its confidence intervals whenever possible in future IS research. Our study thus emphasizes the importance of effect size and describes how it can further advance IS research.

The rest of this paper is organized as follows: We start by describing the concept of effect size, its advantages, and provide some notes on interpretation. Next, we report on our review detailing the extent to which effect sizes have been reported in IS literature over the last decade. Finally, we discuss how reporting of effect size can advance IS research along and provide recommendations for IS researchers.

Effect size: Concept and its advantages

The concept of effect size

To quote Hojat and Xu "Effect size can be considered an index of the extent to which the research hypothesis is true, or the *degree* to which the findings have practical significance in the study population".[13] Where a *p* value demonstrates statistical significance, the corresponding effect size measure demonstrates the magnitude of the phenomenon in the population (i.e., practical significance). Effect size can thus show how effectively a theory explains or predicts empirical observations in inductive-statistical research.[14]

Table 1. Effect size and value for small, medium, and large effects.[16]

Test	Measure	Effect Size Small	Medium	Large
Measures of Standardized Differences				
m_A vs. m_B for independent means	$d = \frac{m_A - m_B}{\sigma}$	.20	.50	.80
r_A vs. r_B for independent rs	$q = z_A\text{-}z_B$	.10	.30	.50
P_A vs. P_B for independent proportions	$h = \phi_A\text{-}\phi_B$	.20	.50	.80
Chi-square for goodness of fit and contingency	$w = \sqrt{\sum_{i=1}^{k} \frac{(P_{1i} - P_{0i})^2}{P_{0i}}}$	.10	.30	.50
One-way analysis of variance	$f = \frac{\sigma_m}{\sigma}$	.10	.25	.40
Measures of Association Strength				
Product moment r	r	.10	.30	.50
Multiple and multiple partial correlation	$f^2 = \frac{R^2}{1-R^2}$	.02	.15	.35
η^2	$\eta^2 = \frac{SS_{effect}}{SS_{total}}$	.01	.06	.14

Based on our review, two categories of effect size measures are popularly used in IS and will thus be focused on in our discussion: When comparing the differences between two groups, effect size estimates are often based on standardized differences between the means. Measures of standardized mean differences include Cohen's[11] *d, f, g, h, q*, and *w*[2] On the other hand, if the variables being considered are continuous or have more than two levels, effect size estimates commonly describe the proportion of variability that can be accounted for by each variable[12] or the strength of association. Measures of association strength include Cohen's[11] f^2 and *r*. There are also other measures of effect magnitude such as Cohen's[11] U_1, U_2, and U_3 which do not fall into these two major categories[15] (refer to Table 1 for examples of all categories).[3] Depending on the statistical test used, the approaches for calculating effect size may vary, and there are often multiple approaches that may be suitable in a given study. These measures can also be converted between each other (refer to Table 2 for example).

Each measure has certain limits and no measure is most suitable for all conditions.[17] Take η^2 as an example (Table 3). There are different measures of η^2, such as η^2, partial η^2 (i.e., η^2_p), and generalized η^2 (i.e., η^2_G). Here η^2 describes the proportion of the *total* variance accounted for by the effect under investigation, while η^2_p describes the proportion of variance associated with the

Table 2. Formulae for calculating effect size directly and indirectly.[12]

From this statistic	To this statistic: Cohen's d	Point biserial r	η^2 with similar group sizes
Direct formula	$d = \frac{m_A - m_B}{\sigma}$	$r = \sqrt{\frac{SS_{effect}}{SS_{total}}}$	$\eta^2 = \frac{SS_{effect}}{SS_{total}}$
Cohen's d	-	$r = \frac{d}{\sqrt{d^2+4}}$	$\eta^2 = \frac{d^2}{d^2+4}$
Point biserial r	$d = \frac{2r}{\sqrt{1-r^2}}$	-	$\eta^2 = r^2$
η^2 with similar group sizes	$d = \frac{2\sqrt{\eta^2}}{\sqrt{1-\eta^2}}$	$r = \sqrt{\eta^2}$	-

[2]These measures are used when dependent variables are continuous variables. On the other hand, when dependent variables are categorical variables, measures such as the relative risk, the odds ratio, and rate ratio are used to determine whether the probability of a certain event differs across groups.[5]

[3]There are more than 70 effect size measures in the literature. For a more complete list, please refer to Table 5.1 from Kirk.[18]

Table 3. Different measures of eta squared.[12]

Measure	Formula
η^2	$\frac{SS_{effect}}{SS_{total}}$
η^2_{p}	$\frac{SS_{effect}}{S_{effect}+SS_{error}}$
η^2_{G}	$\frac{SS_B}{SS_{total}-SS_A}$

effect when the variance associated with all other effects has been removed. Therefore, when there is only one effect to investigate, η^2 will be the same as η^2_{p}. On the other hand, when there is more than one effect to examine, η^2_{p} will be larger than η^2. In such a scenario, η^2 is helpful to understand one effect relative to the total variance while $\eta^2_{\mathrm{p}}$is useful to understand one effect relative to the variance associated with the effect plus the error variance. Further, although these two measures are useful for within-study comparisons, the error variances may not be comparable in multiple studies, making between-study comparisons problematic. η^2_{G}can be helpful for between-study comparisons, which describes the proportion of variance within a study associated with the effect but without the distorting effects (i.e., variable effects in some studies but not others).

Barriers

A potential barrier for widespread reporting of effect sizes is the lack of clarity around which techniques to adopt. Kirk[18] mentions more than 70 different measures of effect magnitude, of which several may be suited for a particular study. In the absence of a publication manual, or clear direction, IS researchers may be forgiven for following the well-trodden path with their statistical reporting.

The discipline of IS draws many techniques from the psychological and behavioral sciences. For behaviorists, statistical significance is not conflated with clinical relevance, and care is taken to present results in the correct context so that the broader implications are clear to the reader. In response to Cohen[6], the American Psychological Association (APA), through their board of scientific affairs called for the inclusion of effect size estimates in scientific reporting. These guidelines have been included in the APA Publication Manual since 2001. The current edition states:

"For the reader to appreciate the magnitude or importance of a study's findings, it is almost always necessary to include some measure of effect size in the Results section. Whenever possible, provide a confidence interval for each effect size reported to indicate the precision of estimation of the effect size."[19]

Although uptake of these recommendations has been slow, much to the dismay of some statisticians, the incremental improvements are becoming apparent.

Effect size advantages

Previous quantitative studies have dominantly used classical null hypothesis significance testing to evaluate whether their hypotheses are supported. This approach has three main criticisms (summarized in Ref 16). First, null hypothesis significance testing does not tell what researchers want to test. Specifically, researchers want to test the probability that the null hypothesis is true (H_0) given a set of data (D) (i.e., p (H_0| D)). However, null hypothesis significance testing shows the probability of obtaining the data if the null hypothesis is true (i.e., p ($D|H_0$)).[4]

Second, null hypothesis significance testing is influenced by the sample size. Recent studies show that with extremely large samples, p-values quickly approach zero.[10,20] As argued by Tukey,[8] "the effects of A and B are *always* different-in some decimal place-for any A and B" (p. 100). Therefore, it is possible to find significant but not useful effects with large samples.

Third, null hypothesis significance testing involves using a fixed level of significance (e.g., .05). When hypotheses are not supported, researchers may mistakenly interpret the results as evidence for accepting null hypotheses.

Effect size is one useful approach to supplement null hypothesis significance testing and address these limitations.[15] First, the effect size directly shows the magnitude of certain effects. For example, based upon Cohen's[16] guidelines, researchers can know whether certain results are "negligible" (around .20), "moderate" (around .50), or "important" (around .80) by using Cohen's d. Therefore, effect size can convey the practical significance of the results. Second, effect size is independent of sample size and scale-free.[21] While significance can be directly influenced by the investigators' setting of N, effect size is not influenced in this way. Note that we do not propose that null hypothesis significance testing should be abandoned. Instead, we suggest supplementing null hypothesis significance testing with effect size data to show the practical significance of results.

Interpreting effect size measures

Effect size is especially useful if a discipline-specific index of prior work is available. Unfortunately, no such index exists yet for IS, despite prior work raising this issue.[22] Where such an index does not exist, it is common to use the expected magnitudes of Cohen's d as a proxy. According to Cohen,[16] a medium effect of .50 represents an effect visible to the naked eye of a careful observer, and it is confirmed that .50 approximates the average size of effects in several fields e.g.[23] Next, a small effect of .20 is selected to be noticeably smaller than medium but not so small as to be trivial. Lastly, a large effect of .80 is selected to have the same distance above medium.

Specification of expected magnitudes makes two important contributions. First, researchers have guidelines to interpret the magnitude of effect size. In other words, researchers can determine the practical or theoretical importance of certain effects and relative contributions of different factors in one study or the same factor across different studies.[12] Second, power analysis can be conducted accordingly. For example, before the study is conducted, sample size can be estimated to ensure sufficient power for detecting small, medium, and large effects.

[4]Interested readers can refer to Cohen[6] for a detailed explanation regarding why p ($H_0|D$) ≠ p ($D|H_0$).

Researchers must be mindful that small observed effects can still be important and merit continued study. In certain areas of research, small effects may well be the norm, and work may not always be consistent with the ranges described by Cohen. Rosenthal and Rubin[24] exemplified this point with a hypothetical scenario regarding disease survival rate under two medical treatments. Under Treatment A 30% of the patients live, under Treatment B 70% live. This appears to be a dramatic effect, yet for their data, $R^2 = 0.16$. It would be irresponsible to discard such findings on the basis that they fall below the 0.2 "small effect" threshold. In addition, Ellis[25] argues that small effects can also be important 1) if they "change the perceived probability that larger outcomes *might* occur" (e.g., the untimely death of schoolboy may signal an increased risk of an influenza outbreak) (p. 36), 2) if they accumulate into large effect (e.g., small energy savings per person results in large savings in total), or 3) if they can lead to technological breakthroughs (e.g., Fleming's discovery of penicillin).

Although the small, medium, and large effect sizes proposed by Cohen are helpful, they should not be perceived as invariant across contexts. Cohen's recommendation was based upon a specific area of behavioral science, and interpretation of effect size should not be based upon broad-based conventions.[26] Consistent with the discussion above, "the importance of any particular effect size depends upon the nature of the outcome studied."[27] However, researchers may often simply apply Cohen's recommendation without explicitly acknowledging that these values are based on specific literature domains. As noted by Cohen,[11] the interpretation of effect size (i.e., what would be a large, medium, or small effect size) depends on the specific research contexts, and these values may be used only when there is no better basis available.

Previous literature has used other measures to evaluate their results besides null hypothesis testing. For example, many studies use R^2 to evaluate their model. R^2 represents the total variance accounted and is useful to assess the whole model. However, R^2 has a few limitations, including sensitivity to violation of assumptions (heterogeneity of variance, balanced designs) and large standard errors.[28] Further, as an omnibus measure, it does not explain the contribution of any specific variables (e.g., independent variables or moderators).[5] Another example is Akaike's information criterion (AIC). AIC is often used for comparing non-nested models, and the lower AIC implies a better model. Although Burnham and Anderson[29] suggest that AIC change of about four to seven correspond roughly to "95% confidence" interval, there is no guideline regarding how much change in AIC represents a practically important difference. Therefore, our discussions do not focus on R^2 and AIC. Below we report on our review to assess the current practice of reporting effect size in the IS literature.

Effect size in IS: A survey of current practice

Research articles in Senior Scholars' Basket of Journals (i.e., European Journal of Information Systems [EJIS], Information Systems Journal [ISJ], Information Systems Research [ISR], Journal of the Association for Information Systems [JAIS], Journal of Information Technology [JIT], Journal of Management Information Systems [JMIS], Journal of Strategic Information Systems [JSIS], MIS Quarterly [MISQ])[6] between 2007 and 2017 inclusive were reviewed. A search was first conducted using the term "effect size." Next, further searches were carried out to ascertain if any additional papers used the terms "Cohen's d," "f squared" or "Omega squared" but had not used the words "effect size." These papers were then examined to determine if they actually had reported effect sizes, and which measures had been used. As described above, simply reporting overall R^2 does not fulfill the purpose of reporting practical significance and these are not included in the tally.

To place our findings into context, we also reviewed the contents of all published papers in our target journals to ascertain how many papers presented quantitative data reporting statistical significance where effect sizes *could* have also been reported. For this analysis, we reviewed a total of 1361 published works. The results are shown in Tables 4 and 5 (refer to the Appendix for a detailed literature summary).

Of the 1361 studies reporting statistical significance, only 61 papers reported effect size measures. While all journals have at least two papers that reported some measure of effect size, this is a relatively low proportion of the number of papers where it could have been reported. Our review found that Cohen's f^2 (42) and d (11) are the two most popular measures in the literature. As many papers reported a range of effect size, we considered the magnitude of the largest effect noted, finding that totals for small, medium, and large effects are evenly distributed at 19, 21, and 21 papers, respectively. The results of our review are presented in Appendix A which provides the foundation for an index of expected effect sizes in IS research. These different measures have been used in different ways: to 1) conduct manipulation checks, 2) assess the effect of independent variables or compare group

Table 4. Number of studies reporting effect size.

	EJIS	ISJ	ISR	JAIS	JIT	JMIS	JSIS	MISQ
Cohen's d	1	1	2	2	1	2	1	1
Cohen's f^2	2	1	6	5	1	9	8	10
η^2			1					2
Pearson's r	1			1		1		1
Odds ratio					1			1

Note: one study reports two measures.

Table 5. Number of studies reporting statistical significance.

	EJIS	ISJ	ISR	JAIS	JIT	JMIS	JSIS	MISQ
Number	174	62	300	134	46	304	70	271

[5]According to Kelley and Preacher,[50] Effect size is "a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest" (p. 140). Therefore, R^2 is less helpful because it describes the effect of *all* independent variables in the regression rather than the effect of a particular independent variable. Although the increment in R^2 can be calculated to assess the effect of certain variables, it does not represent the *rate* of change.[51] Therefore, R^2 is less helpful to understand the practical significance of certain variables.

[6]https://aisnet.org/?SeniorScholarBasket

means, 3) assess the effect of interaction effects, or 4) assess the effectiveness of alternative models.

We also look at how different measures have been used. First, Cohen's *d* has been used to check manipulations[30] or to assess the effect of independent variables (i.e., compare group means).[31] Next, Cohen's f^2 has been used to assess the effect of independent variables/moderators (e.g.,[32]) or to compare the effectiveness between different models (e.g.,[33]) η^2 is also used to assess the effect of independent variables/moderators.[34] Further, Pearson's *r* has been used to assess the effect of independent variables (often reported in meta-analysis)[35] or to assess the difference between groups of participants.[36] Lastly, odds ratio is also used to assess the effect of independent variables.

Kirk[15] found that 14 measures (not including R^2 or variance-accounted-for) are reported from 82 papers in four psychology journals in 1995. Since the review was conducted before APA's call for reporting effect size, we expect that the frequency of reporting effect size in psychology is much higher in their recent studies. On the other hand, our review in eight IS journals between 2007 and 2017 only finds 61 articles reporting five measures of effect size. This result shows that there does not appear to be a trend toward increased reporting of effect size measures. As the numbers are so small, it is not possible to draw any meaningful inference over time.

A further issue of note is that, in the papers which do include effect size, there are several limitations. Firstly, although some effect size measures are interchangeable, it is necessary to report the formula used. This disclosure is an important issue because researchers may use incorrect terminology. For example, the term Cohen's *d* may be used although the formula of other measures is employed. Secondly, the results of our literature search reveal that only one study calculated the confidence interval of effect size. This interval is useful for subsequent meta-analysis. Finally, no study reports different measures of effect size. In the one study which provided both odds ratio and Cohen's *d*, these were essentially one measure because one is calculated from the other.

In summary, our review reveals both that effect size measures are infrequently reported in the IS literature, and that where reported, there are limitations in the way it is reported. Two factors are likely to contribute to this state of affairs. Firstly, there is insufficient emphasis on this element of statistical reporting in the IS literature. Although topics such as reliability, content, and construct validity are well understood and described in influential texts (e.g.,[37]) the topic of effect size does not feature. Thus, as a discipline, IS researchers may potentially lack experience in the reporting of effect size measures. Secondly, there is no external influence on IS researchers to drive such a change in reporting techniques, unlike other disciplines. For example, there is minimal IS influence by the APA guidelines recommending the use of effect size measures. As IS lacks a single authoritative body to drive such change, it is likely that uptake will be slower.

Discussion

Advance IS research

Effect size can supplement statistical significance tests but has not been widely adopted in the IS literature. Our data reveal this level of adoption to be below 5% of the number of papers reporting significance values alone. We argue that effect size can advance the IS literature in several ways. First, effect size can show the practical significance of results. For example, Siponen and Baskerville[1] call for research focusing on intervention effect rate to demonstrate the effectiveness of theory and to achieve practical significance. Here effect sizes can be used as measures to demonstrate which theory/approach has the best intervention effect rate in a certain scenario.

Second, using effect size can help enhance theoretical precision. For example, variable X may be hypothesized to have a positive effect on variable Y greater than a lower limit. The relationship between X and Y may also be hypothesized to be larger than that between W and Y.[38] In the first scenario, researchers may hypothesize the relationship between X and Y to be larger than .20 (assume Cohen's *d* is used). If the actual effect size is medium (around .50) or large (around .80), then their hypothesis is supported. In the second scenario, researchers may hypothesize different effects of various independent variables on the dependent variable. Take Ke and Zhang[39] for example. Their study hypothesizes that the effects of various types of extrinsic motivation on task effort increase as motivation becomes more autonomous. Again, the effect size of each motivation can be calculated to assess whether the hypothesis is supported.

Third, reporting effect size measures can facilitate meta-analysis.[3] Meta-analysis aims to estimate "true effects" by integrating findings from a large number of prior studies.[40] Three pieces of information are needed to conduct meta-analysis: 1) effect size, 2) confidence interval of effect size and 3) sample size. Therefore, explicitly reporting effect size facilitates conducting meta-analysis and makes meta-analysis results more accurate.

Recommendations for future IS researchers

Reporting effect sizes can greatly advance the future practical significance of IS research.[7] Evidence-based practice is increasing practical interest in academic research in areas where the research offers real evidence. Practitioners search for treatments, action levers they can pull, that should improve their situation.[41] This search requires that they must critically appraise evidence in the research literature and combine it with evidence from their context.[42] Well-formed effect size reporting helps evidence-based practitioners compare the effects of possible action-lever variables within a single study and possibly across multiple studies. But action lever variables and their effect sizes are only one form of relevance. Practitioners can better understand the implications of effect sizes in real life because they may often possess more realistic judgments about the real-world importance of the magnitude of effect sizes compared to researchers.[4] We

[7]Please note that practical significance may not always lead to relevance, or vice versa (Mohajeri et al. Forthcoming).

suggest future research report effect sizes and their confidence intervals whenever possible. Fortunately, there are free web resources to help with related calculations e.g.[43] Also, studies have been conducted to address technical issues (e.g., non-independent data) when calculating effect size e.g.[44]

It is important to report the actual formulae used in calculating effect size. Such detail is necessary because it reveals how the reported estimates are calculated. This detail is needed even if certain measures are quite popular. For example, Fritz et al.[12] found that Cohen's d is often used as a generic term for standardized mean differences. Specifically, when researchers evaluate the effect size of mean differences, they may claim to use Cohen's d but in fact, follow the formula for Hedges' g ($g = \frac{m_A - m_B}{s}$). Such approaches make result comparisons problematic. Therefore, reporting the formula can help readers correctly interpret the results and better facilitate later meta-analysis.

Based on our review and the literature, some measures are more easily calculated than others in certain contexts. We provide a list of contexts and corresponding measures to calculate in Table 6. The list aims to provide a starting point for effect size reporting in certain contexts. Once certain measures are calculated, they can be used to derive other measures if needed. Here we do not intend to provide a complete list of contexts and measures. In other words, there may be other contexts where effect sizes can be helpful, and some other measures in the literature may be more relevant or more easily calculated.[12,15] For example, although our review shows that the IS literature has not reported Hedges' g, this measure may be more easily calculated than Cohen's d since it uses the pooled sample standard deviation instead of the standard deviation for the population as the denominator. We do not intend to limit the reporting of effect size to these measures in Table 6 only and encourage researchers to explore and report other measures wherever appropriate.

To make IS research results comparable, it can be important to report more than a single effect size measure. As discussed above, different measures have their advantages and limitations. Consider a scenario where researchers have three manipulations in their study: two manipulations from the literature and one manipulation which is newly designed. At the very least, η^2_G must be reported so the results can be compared with other contemporary or future studies. This measure can provide complementary information and help build cumulative knowledge in the literature. But if the study involves within-study comparison (i.e. to assess which manipulation has a larger effect), η^2 and/or η^2_p would also be important for making the results comparable.

Table 6. Recommended measures.

Context	Measure
Manipulation check	Cohen's d
Mean Comparison	Cohen's d
Test the effect of independent variables	Cohen's f^2, η^2, Pearson's r, odds ratio
Test the effect of moderators	Cohen's f^2, η^2
Compare alternative models	Cohen's f^2

Researchers may also describe the potential implications of effect size in their discussion section. For example, a large but nonsignificant effect may suggest opportunities for further investigation with greater power.[12] In some contexts, small effect sizes can still have a substantial practical impact.[24] On the other hand, significant results with trivial effects may be due to large sample sizes and this condition would need discussion. In other words, such results may have little value to practitioners (i.e., little practical significance). Researchers can also compare their effect sizes with those from previous studies and discuss how the results are (or are not) consistent.

Here one important question to ask is, "What is a large, medium, or small effect size for IS research?" As noted by Kirk,[15] although Cohen's definition is a good start, more systematic research is needed to extend his work in specific contexts. For example, in applied psychology, Bosco et al.[45] show that none of the existing benchmarks of effect size fit their results. Specifically, Cohen's[11] benchmarks for small, medium and large effect sizes roughly correspond to the 33rd, 73rd, and 90th percentiles of Bosco et al.[45]'s distribution, respectively (i.e., skew toward small effect sizes). Similarly, in the organizational behavior/human resources literature, Paterson et al.[46] find that the average uncorrected effect is r = .23, and Cohen's[11] benchmarks for small, medium, and large effect sizes roughly correspond to the 20th, 75th, and 95th percentiles, respectively. Therefore, we expect that the overall trend of effect size in IS research (or a sub-area) may not be consistent with the guidelines provided by Cohen either. Indeed, for IS research findings to accumulate over time, we as a discipline need to consider effect sizes to develop an understanding of the typical strengths of relationships.[47] Further, as the strength of effect size probably varies across contexts, researchers need to provide their judgment and supporting rationales to interpret the magnitude of effect sizes.[4] Reporting is the first step toward a valuable discussion of the meaning and practical significance of findings in the context of our discipline, thus future studies are needed to develop guidelines for interpreting the magnitude of effect size in the IS literature.

In their 30-year review, Aguinis et al.[26] found that the median effect size for moderation relationships (f^2) was only .002, while 72% of studies had power of .80 or higher. They question using conventional measures to interpret moderating effects. Our review also finds that most moderation effects have small or medium effect sizes. Only one study from Tiwana[48] finds a large moderation effect. This result implies that moderation effects are generally smaller and harder to detect. Therefore, a small moderation effect should not be interpreted as a trivial or non-important relationship.

Sources of variability should also be considered when effect sizes are interpreted or compared. In other words, researchers need to consider the design and sample characteristics when interpreting effect sizes. For example, when studies are conducted to understand social media usage behaviors, one sample with experienced social media users probably has less variability than another sample with social media users from varied backgrounds.

Conclusion

The effect sizes in IS research have not been widely reported. We argue that IS is no longer an immature discipline where the objective of theory testing is simply to learn if there is any effect at all (e.g. through a *p* value). Rather, the discussion must move toward a richer perspective considering measures of impact such as effect sizes. It is a relatively simple additional step to calculate the effects sizes in addition to statistical significance and explained variance. This important research practice expands the practical significance of our research and the comparability of our studies; it better enables future IS practitioners to ground evidence-based practice on rigorous scholarly research.

Acknowledgments

The authors gratefully acknowledge the very helpful comments of Detmar Straub and Kaveh Mohajeri.

ORCID

Xuequn Wang http://orcid.org/0000-0002-1557-8265

References

1. Siponen M, Baskerville R. Intervention effect rates as a path to research relevance: information systems security example. J Assoc Inf Syst. 2018;19(4):247–65. doi:10.17705/1jais.00491.
2. Rousseau DM. The oxford handbook of evidence-based management. Oxford: Oxford University Press; 2012.
3. Schwab A. Why all researchers should report effect sizes and their confidence intervals: paving the way for meta-analysis and evidence-based management practices. Entrepreneurship. 2015;39:719–25.
4. Mohajeri K, Mesgari M, Lee AS. When statistical significance is not enough: investigating relevance, practical significance and statistical significance. MIS Quarterly. Forthcoming. https://www.misq.org/forthcoming/.
5. Ferguson CJ. An effect size primer: A guide for clinicians and researchers. Prof Psychol Res Pr. 2009;40(5):532. doi:10.1037/a0015808.
6. Cohen J. The earth is round (p<. 05). Am Psychologist. 1994;49:997–1003.
7. Cohen J. Things i have learned (so far). Am Psychologist. 1990;45(12):1304. doi:10.1037/0003-066X.45.12.1304.
8. Tukey JW. The philosophy of multiple comparisons. Stat Sci. 1991;6:100–16. doi:10.1214/ss/1177011945.
9. Geoffrey K, Wickens Thomas D. Design and analysis: A researcher's handbook. Upper Saddle River, NJ: Prentice Hall; 2004. p. 4.
10. Lin M, Lucas HC Jr, Shmueli G. Research commentary—too big to fail: large samples and the p-value problem. Info Syst Res. 2013;24:906–17.
11. Cohen J. Statistical power analysis for the behavioral sciences. Hilsdale, NJ: Lawrence Earlbaum Associates; 1988. p. 2.
12. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen. 2012;141(1):2. doi:10.1037/a0024338.
13. Hojat M, Xu G. A visitor's guide to effect sizes–statistical significance versus practical (clinical) importance of research findings. Adv Health Sci Edu. 2004;9(3):241–49. doi:10.1023/B:AHSE.0000038173.00909.f6.
14. Webster J, Starbuck WH. Theory building in industrial and organizational psychology. In: Cooper GL, Robertson IT, editors. International review of industrial and organizational psychology. Chichester (UK): John Wiley and Sons; 1988. p. 93–138.
15. Kirk RE. Practical significance: A concept whose time has come. Educ Psychol Meas. 1996;56(5):746–59. doi:10.1177/0013164496056005002.
16. Cohen J. A power primer. Psychol Bull. 1992;112(1):155. doi:10.1037/0033-2909.112.1.155.
17. McGraw KO. Problems with the best: A comment on rosenthal's" how are we doing in soft psychology?". Am Psycho. 1991;46(10):1084–86.
18. Kirk RE. The importance of effect magnitude. In: Handbook of research methods in experimental psychology. Hoboken, NJ: Blackwell Publishing; 2003. p. 83–105.
19. American Psychological Association. Publication manual of the american psychological association. 6th ed. Washington (DC): American Psychological Association; 2010.
20. Guo W, Straub D, Zhang P. A sea change in statistics: A reconsideration of what is important in the age of big data. J Manage Anal. 2014;1(4):241–48. doi:10.1080/23270012.2014.983196.
21. Coe R. It's the effect size, stupid: What effect size is and why it is important. Exeter, UK; 2002.
22. Baroudi JJ, Orlikowski WJ. The problem of statistical power in mis research. MIS Q. 1989;13(1):87–106. doi:10.2307/248704.
23. Sedlmeier P, Gigerenzer G. Do studies of statistical power have an effect on the power of studies? Psychol Bull. 1989;105(2):309. doi:10.1037/0033-2909.105.2.309.
24. Rosenthal R, Rubin DB. A note on percent variance explained as a measure of the importance of effects. J Appl Soc Psychol. 1979;9(5):395–96. doi:10.1111/j.1559-1816.1979.tb02713.x.
25. Ellis PD. The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results. Cambridge, UK: Cambridge University Press; 2010.
26. Aguinis H, Beaty JC, Boik RJ, Pierce CA. Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. J Appl Psychol. 2005;90(1):94. doi:10.1037/0021-9010.90.1.94.
27. Eden D. From the editors: replication, meta-analysis, scientific progress, and amj's publication policy. Acad Manag J. 2002;45(5):841–46. doi:10.5465/amj.2002.7718946.
28. Olejnik S, Algina J. Measures of effect size for comparative studies: applications, interpretations, and limitations. Contemp Educ Psychol. 2000;25(3):241–86. doi:10.1006/ceps.2000.1040.
29. Burnham KP, Anderson DR. Model selection and inference. New York (NY): Springer-Verlag; 1998.
30. Lowry PB, Moody GD. Proposing the control-reactance compliance model (crcm) to explain opposing motivations to comply with organisational information security policies. Info Syst J. 2015;25(5):433–63. doi:10.1111/isj.12043.
31. Kim D, Benbasat I. Trust-assuring arguments in b2c e-commerce: impact of content, source, and price on trust. J Manage Info Syst. 2009;26(3):175–206. doi:10.2753/MIS0742-1222260306.
32. Bala H, Massey AP, Montoya MM. The effects of process orientations on collaboration technology use and outcomes in product development. J Manage Info Syst. 2017;34(2):520–59. doi:10.1080/07421222.2017.1334494.
33. Fang Y, Qureshi I, Sun H, McCole P, Ramsey E, Lim KH. Trust, satisfaction, and online repurchase intention: the moderating role of perceived effectiveness of E-Commerce institutional mechanisms. MIS Q. 2014;38(2):407–27. doi:10.25300/MISQ/2014/38.2.04.
34. Wells JD, Valacich JS, Hess TJ. What signal are you sending? How website quality influences perceptions of product quality and purchase intentions. MIS Q. 2011;35:373–96. doi:10.2307/23044048.
35. He J, King WR. The role of user participation in information systems development: implications from a meta-analysis. J Manage Info Syst. 2008;25(1):301–31. doi:10.2753/MIS0742-1222250111.
36. Khan SS, Kumar R, Zhao K, Stylianou A. Examining real options exercise decisions in information technology investments. J Assoc Inf Syst. 2017;18(5):372. doi:10.17705/1jais.00459.

37. Boudreau M, Gefen D, Straub D. Validation in information systems research: A state-of-the-art assessment. MIS Q. 2001;25 (1):1–16. doi:10.2307/3250956.
38. Edwards JR, Berry JW. The presence of something or the absence of nothing: increasing theoretical precision in management research. Organ Res Methods. 2010;13(4):668–89. doi:10.1177/1094428110380467.
39. Ke W, Zhang P. The effects of extrinsic motivations and satisfaction in open source software development. J Assoc Inf Syst. 2010;11(12):784. doi:10.17705/1jais.00251.
40. Schmidt F, Hunter J. Are there benefits from nhst? Am Psychologist. 2002;57(1):65–66. doi:10.1037/0003-066X.57.1.65.
41. Robey D, Markus ML. Beyond rigor and relevance: producing consumable research about information systems. Inf Resou Manage J. 1998;11(1):7–15. doi:10.4018/irmj.1998010101.
42. Pfeffer J, Sutton R. Hard facts dangerous half-truths & total nonsense: Profiting from evidence-based management. Cambridge (Mass): Harvard Business School Pres; 2006.
43. Practical meta-analysis effect size calculator. 2012. Campbell Corporation; [accessed]. http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-ESTypes.php.
44. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: A practical guide for biologists. Biol Rev. 2007;82(4):591–605. doi:10.1111/j.1469-185X.2007.00027.x.
45. Bosco FA, Aguinis H, Singh K, Field JG, Pierce CA. Correlational effect size benchmarks. J Appl Psychol. 2015;100(2):431. doi:10.1037/a0038047.
46. Paterson TA, Harms P, Steel P, Credé M. An assessment of the magnitude of effect sizes: evidence from 30 years of meta-analysis in management. J Leadersh Organ Stud. 2016;23:66–81.
47. Schmidt F. Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers. Psychol Methods. 1996;1(2):115. doi:10.1037/1082-989X.1.2.115.
48. Tiwana A. Governance-knowledge fit in systems development projects. Info Syst Res. 2009;20(2):180–97. doi:10.1287/isre.1070.0164.
49. Sullivan GM, Feinn R. Using effect size—or why the p value is not enough. J Grad Med Educ. 2012;4(3):279–82. doi:10.4300/JGME-D-12-00156.1.
50. Kelley K, Preacher KJ. On effect size. Psychol Methods. 2012;17 (2):137. doi:10.1037/a0028086.
51. Champoux JE, Peters WS. Form, effect size and power in moderated regression analysis. J Occup Psychol. 1987;60(3):243–55. doi:10.1111/j.2044-8325.1987.tb00257.x.