Meta-analysis in Information Systems Research:  
Review and Recommendations

Anand Jeyaraj  
Professor of Information Systems  
Raj Soin College of Business  
Wright State University  
3640 Colonel Glenn Highway  
Dayton, OH 45435, USA  
Phone: +1.937.775.2189  
Email: anand.jeyaraj@wright.edu  
Web: https://people.wright.edu/anand.jeyaraj

Yogesh K. Dwivedi  
Professor of Digital Marketing and Innovation  
Director of Emerging Markets Research Centre (EMaRC)  
Co-Director of Research  
School of Management, Room #323  
Swansea University, Bay Campus  
Fabian Bay, Swansea, SA1 8EN, Wales, UK  
Phone: +44 (0) 1792 602340  
Email: y.k.dwivedi@swansea.ac.uk  
Web: http://www.swansea.ac.uk/staff/som/academic-staff/y.k.dwivedi/

Abstract

Meta-analysis has gained considerable momentum in information systems (IS) research over the last two decades. As the IS discipline has matured and grappled with various applications of information technology (IT) tools, meta-analysis has served as a powerful mechanism to enable the synthesis of prior findings, reconciliation of inconsistent findings, and resolution of relationships. Prior meta-analysis studies in IS have employed derived metrics and reported effect sizes in both exploratory and confirmatory approaches, and also conducted descriptive analysis, comparison of subgroup means, regression, and structural equation modeling methods. This paper conducts a review of prior meta-analysis studies published since 2000 using a 2x2 framework and identifies the state of meta-analysis research in IS. The challenges in conducting meta-analysis research and the opportunities for meta-analysis research in IS are also identified. Based on the challenges, several recommendations to handle publication bias, inclusion and exclusion of studies, effect sizes, coding, meta-analysis modeling, and sensitivity analysis are provided. Opportunities for meta-analysis such as clarifying constructs and relationships, identifying contingencies, and testing theories to advance IS research are also identified.

Keywords: Meta-analysis, Research methods, Information Systems, Challenges, Opportunities, Recommendations.
Meta-analysis in Information Systems Research: Review and Recommendations

1. Introduction

The use of meta-analysis (MA) in information systems (IS) research has gained considerable traction over the last couple of decades. As the IS discipline continues to mature and grapple with various applications of information technology (IT) for individuals, organizations, and societies, there has been an increasing need to synthesize extant research, reconcile inconsistent empirical findings, identify gaps in knowledge, and chart paths for future research (e.g., King and He 2005). MA serves as a powerful alternative or supplement to traditional literature reviews for research synthesis in the IS discipline, possibly in the context of systematic reviews1 as well. MA studies in IS have tackled a variety of research questions dealing with behaviors of individuals, organizations, and teams (e.g., Dennis et al. 2001; Sabherwal et al. 2006; Dwivedi et al. 2019).

The advantages of MA over individual studies are well recognized despite the inherent criticisms of MA such as the comparison of “apples and oranges” and the notion of “garbage in, garbage out” (Borenstein et al. 2009). MA enables: the reconciliation of inconsistent findings across studies for the same relationship; the examination of relationships that could not be conducted within the context of a single study; and the extraction of new insights into the underlying relationships within the research area (e.g., Sabherwal and Jeyaraj 2015). Several methods for MA research are available (e.g., Glass et al. 1981; Hedges and Olkin 1985; Hunter and Schmidt 1990; Rosenthal 1991; Lipsey and Wilson 2001; Borenstein et al. 2009).

Although MA methods are typically employed on effect sizes such as correlations and group mean differences (e.g., Glass et al. 1981; Hedges and Olkin 1985; Hunter and Schmidt 1990) and inferences about the relationships and their moderators are drawn based on the aggregated effect sizes and credibility intervals, MA studies in IS demonstrate two deviations. First, studies have applied additional methods such as mean comparison tests, regression, structural equation modeling (SEM) on the corrected effect sizes produced by the MA for other research objectives (e.g.,

---

1 MA may also be employed in the context of systematic reviews, which are viewed as a method to collect and summarize all empirical evidence on a topic of interest (e.g., The Cochrane Collaboration, reported in Higgins and Green 2008, or the Campbell Collaboration). Systematic reviews appear to be gaining traction in IS research (e.g., Al-Emran et al. 2018; Asadi et al. 2017; Chu et al. 2019; Cognini et al. 2018) as does MA in the context of systematic reviews (e.g., Kapoor et al. 2014a; Tamilmani et al. 2019; Tao et al. 2020).
Sharma and Yetton 2007; Sabherwal et al. 2006; Joseph et al. 2007; Wu and Lederer 2009; Gerow et al. 2014; Dwivedi et al. 2019). Second, studies employed metrics other than effect sizes reported in prior studies to conduct MA using descriptive methods (e.g., Jeyaraj et al. 2006; Rana et al. 2015) or regression methods (e.g., Kohli and Devaraj 2003; Jeyaraj 2019). MA studies in IS thus strive to go beyond the traditional MA goals of reconciling inconsistent findings, resolving the magnitudes and directions of the relationships, and identifying moderators that may alter the effects between variables (e.g., Sabherwal et al. 2006).

MA studies typically encounter several challenges from inception to completion that need to be handled, which assume considerable importance given the scope and diversity in the application of MA methods in IS. Several studies within the IS discipline have highlighted various aspects of meta-analysis methods and offered additional guidance (e.g., Chau 1999; King and He 2005; Sharma et al. 2009; Hwang and Schmidt 2011; Hess et al. 2014; Larsen and Bong 2016; Kepes and Thomas 2018; Wu et al. 2018) in addition to the general guidelines for MA found in prior literature (e.g., Glass et al. 1981; Hunter and Schmidt 1990). Assuming the rationale for MA has been established based on opportunities or gaps in prior literature, challenges span various aspects such as the overall research approach taken for MA, the type of statistic employed for MA, the research model, and the research methods, including locating primary studies, coding the necessary statistics from studies, identifying the appropriate MA models, and analyzing the data gathered from studies.

This paper reviews the MA studies in IS published in journals since 2000. Using a 2x2 framework based on the statistics and the approaches used for MA, this paper identifies the state of MA research in IS, challenges in conducting MA research, the opportunities for MA research in IS, and several recommendations2 for MA research. The remainder of the paper is also organized along the same lines as highlighted.

2. State of Meta-analysis Research in IS

2.1. Prior Literature

MA studies in IS published in journals since 2000 were considered for this review to understand the state of research. Multiple electronic databases such as Business Source Complete, Science Direct,

---

2 While this paper is based on MA research in the IS discipline, the recommendations could be applicable to MA research in general within any discipline.
IEEE Xplore, and AIS e-library (AISeL) were used as the primary sources to identify prior studies. Keywords related to MA including “meta-analysis,” “meta-analytic structural equation modeling,” and “meta-regression” along with and “information systems” were used to search the electronic databases. Further, bibliographies of MA studies obtained through the search were screened to identify other MA studies. This meta-review is based on 100+ MA studies on various phenomena. Figure 1 shows a summary of the number of MA studies in each year.

Figure 1. MA Studies over time

---

3 MA has garnered considerable attention over the years. Hwang (1996) found six studies prior to 1996; King and He (2005) reported 34 studies between 1992 and 2004; and Kepes and Thomas (2018) showed 31 studies between 2003 and 2012.

4 Studies related to the intellectual core, reference disciplines, and evolution of the IS discipline (e.g., Claver et al. 2000, Vessey et al. 2002, Glass et al. 2004, Larsen and Levine 2005, Larsen et al. 2008, Sidorova et al. 2008, Taylor et al. 2010, Love and Hirschheim 2016, Evangelopoulos 2016, Jeyaraj and Zadeh 2020), IS research impact (e.g., Raghuram et al. 2010; Lowry et al. 2013; Gallivan and Ahuja 2015; Huang et al. 2015) and IS research methods (e.g., Balijepally et al. 2011, Palvia et al. 2015) are excluded from this meta-review although they possess characteristics of MA.

5 More than 10 studies each were published in both Information & Management and Computers in Human Behavior. At least five studies each were published in MIS Quarterly (MISQ), International Journal of Information Management, Journal of Management Information Systems (JMIS), Journal of Information Technology (JIT), and Information Systems Frontiers. Three or more studies were published in European Journal of Information Systems (EJIS), Journal of Strategic Information Systems (JSIS), Journal of Computer Information Systems, Decision Support Systems, Communications of the Association for Information Systems, and Journal of Organizational and End User Computing. Two studies each were published in Information Systems Research (ISR), Journal of the Association for Information Systems (JAIS), Journal of Information Technology Theory and Application, Information Resources Management Journal, International Journal of Human-Computer Studies, and Pacific Asia Journal of the Association for Information Systems. One article each was published in 25+ other journals.
2.2. Organizing Framework

King and He (2005) used a qualitative-quantitative continuum to categorize the commonly used techniques for summarizing knowledge gained from prior studies as narrative reviews, descriptive reviews, vote counting, and meta-analysis, where narrative reviews are largely qualitative and meta-analysis is largely quantitative. Paré et al. (2015) developed a typology of research synthesis methods, including narrative reviews, descriptive reviews, scoping reviews, qualitative systematic reviews, meta-analysis, umbrella reviews, theoretical reviews, realist reviews, and critical reviews, of which several methods rely on qualitative analysis. Different methods for quantitative MA involving published effect sizes have been proposed (Glass et al. 1981; Hedges and Olkin 1985; Hunter and Schmidt 1990; Rosenthal 1991; Lipsey and Wilson 2001; Borenstein et al. 2009) but it is possible to conduct MA without using the effect sizes.

Extant MA research in IS can be mapped using a 2x2 matrix on two dimensions: a) the overall approach taken for meta-analysis, and b) the statistics used for meta-analysis.

- The overall approach taken for meta-analysis refers to whether the study employed exploratory or confirmatory research. Exploratory studies are largely data driven and strive to discover new knowledge about the relationships of interest, i.e., findings reported in prior studies may be treated as “data” to be explored to discover new knowledge. Confirmatory studies have a priori hypotheses about relationships and strive to validate such hypotheses. (Exploratory approach is advantageous for examining inconsistencies in prior findings, correcting for sampling and measurements errors, and assessing the impact of moderators, to generate new knowledge on the relationships being examined.)

- The statistics used for meta-analysis represents whether the study made use of the effect sizes reported in prior studies or metrics derived specifically for the analysis. Depending on the

---

6 The techniques for qualitative MA are not standardized (King and He 2005) and may be influenced by research goals. Examples of narrative reviews include DeLone and McLean (2003), Cane and McCarthy (2009), Petter et al. (2013), and Berente et al. (2019). This meta-review primarily focuses on the plethora of quantitative MA methods in IS research while recognizing the usefulness of narrative reviews in generating meta-interpretations and meta-synthesis of prior findings (Berente et al. 2019).

7 The 2x2 matrix represents one way of mapping prior MA research, assumes use of quantitative MA methods to any extent, ranging from a complete MA of effect sizes (e.g., correlations) to a rudimentary computation of metrics (e.g., frequencies), and enables the classification of the vast majority of prior studies using MA methods.
research goals, different effect sizes such as correlations, mean differences, and p-values may be reported in studies and usable in MA. Derived metrics represent statistics computed for MA using specially-coded variables or details gathered from prior studies. (A significant challenge in using reported effect sizes for MA is that the chosen effect size has to be available in the primary studies; otherwise, the study cannot be included for analysis. Derived metrics have the potential to overcome the challenge since they may not rely on the availability of effect sizes; thus, any study has the potential to be included for analysis.)

**Figure 2** shows the 2x2 matrix with the overall description for each quadrant, which represents a combination of the overall approach taken and the statistic chosen for MA. Quadrants Q3 and Q4 accommodate MA studies that employed effect sizes such as correlations (e.g., Sabherwal et al. 2006) while quadrants Q1 and Q2 subsume MA studies that used derived metrics (e.g., Rana et al. 2015). Quadrants Q1 and Q3 contain MA studies that used exploratory research (e.g., Jeyaraj et al. 2006) while quadrants Q2 and Q4 feature MA studies that undertook confirmatory research (e.g., Roberts et al. 2017).

![Figure 2. Classification Matrix for prior Meta-Analysis studies](image-url)
**Exploratory MA with derived metrics (Q1)**

These MA studies typically provide summaries of prior studies by using derived metrics that represent descriptive statistics. Legris et al. (2003) mapped the number of studies that examined specific relationships (e.g., perceived usefulness system usage, specified in the Technology Acceptance Model). Petter et al. (2013) determined the number of studies in which relationships (e.g., management support system usage) was supported and unsupported. Jeyaraj et al. (2006), Rana et al. (2015), and Baptista and Oliveira (2016) computed a “weight” for each relationship, i.e., the ratio of the number of times it was found significant to the number of times it was examined across studies. Lacity et al. (2010) and Lacity et al. (2011) defined a metric to identify the most robust findings, i.e., a positively significant result was found for a relationship in at least 80% of the times the relationship was examined in prior studies. Jeyaraj (2020) examined prior findings of the relationships found in the IS success models (DeLone and McLean 1992; 2003), and computed a “model success” metric that was based on the completeness of model specification in prior studies and the effectiveness of model paths.

**Confirmatory MA with derived metrics (Q2)**

Similar to exploratory studies with derived metrics, these MA studies use computed metrics as descriptive statistics. However, the computed metrics are then used for additional tests including hypotheses testing of proposed relationships. Kohli and Devaraj (2003) coded the number of positively significant results ($\Sigma p$), the number of negatively significant results ($\Sigma n$), and the total number of outcome variables in a study ($\Sigma v$), and derived a ratio: $\frac{\Sigma p - \Sigma n}{\Sigma v} \times 100$, which was then employed as dependent variable in a meta-regression. Sabherwal and Jeyaraj (2015) and Jeyaraj (2019) also used similar approaches to construct dependent variables for meta-regressions. The independent variables in the meta-regression models were also coded—these were largely binary variables representing relevant aspects such as data source (Kohli and Devaraj 2003), IT progress (Sabherwal and Jeyaraj 2015), and type of respondents (Jeyaraj 2019) that were considered instrumental in explaining the variation in results across studies.

**Exploratory MA with reported effect sizes (Q3)**
These MA studies are generally aimed at uncovering hitherto unknown patterns by aggregating reported effect sizes across studies and conducting subgroup or moderator analysis. While several types of reported effect sizes may be used, the zero-order Pearson correlation has been the most dominant statistic employed in these studies (e.g., Bokhari 2005; Mangalaraj et al. 2020; Tamilmani et al. 2020). Studies in this meta-review have also used regression beta coefficients (e.g., Kapoor et al. 2014a; Baptista and Oliveira 2019). Two approaches are evident among these studies. The first is restricted to the main relationships and their aggregated effect sizes—e.g., Kapoor et al. (2014b) examined the antecedents and descendants of several constructs related to innovation adoption. The second also deals with moderators such that the aggregated effect sizes for the main relationships are then examined by splitting the studies into subgroups—e.g., Gerow et al. (2014) computed the overall effect size for the relationship between alignment and performance but also analyzed the differences due to the type of respondents by splitting the studies into two subgroups to distinguish between studies with single respondents and matched pairs.

**Confirmatory MA with reported effect sizes (Q4)**

Similar to exploratory studies with reported effect sizes, these MA studies are also based on reported effect sizes, but additionally test hypotheses utilizing the aggregated effect sizes obtained through a MA. Pearson correlations have been the most commonly used effect sizes in such studies (e.g., Joseph et al. 2007; Roberts et al. 2017) but significance levels have also been used (e.g., Kroenung and Eckhardt 2015). Different approaches for hypotheses testing are evident among these studies despite similarities in MA methods. Studies have hypothesized directional hypotheses for relationships—e.g., positive or negative effect between two variables—and the aggregated effect sizes resulting from the MA along with the credibility intervals have been used as evidence of support or lack of support for hypotheses (e.g., Petter and McLean 2009; Weigel et al. 2014). Studies that proposed comparative hypotheses—e.g., stronger or weaker effect for a relationship based on a moderating variable—employed additional tests such as T-tests or Q statistics on the aggregated effect sizes from the MA to show evidence of support or lack of support (e.g., Ortiz de Guinea et al. 2012; Wu and Lu 2013). In certain cases (Sharma and Yetton 2003; Sharma and Yetton 2007), the moderator variable was coded as a continuous variable, which facilitated the use of weighted least squares regression for hypotheses testing. Finally, studies also employed SEM methods to test hypotheses (e.g., Sabherwal et al. 2006; Dwivedi et al. 2019), which required the construction of a complete correlation matrix based on all the independent and dependent variables in the study.
**Key Recommendation:** Consider: a) derived metrics when all relevant studies are to be included in the meta-analysis; and b) reported effect sizes when the magnitude and direction of the relationship are to be assessed in the meta-analysis.

**Key Recommendation:** Consider: a) exploratory analysis for obtaining preliminary results; and b) confirmatory analysis for testing hypotheses regarding the relationships.

### 2.3. Research Models

MA may be used to examine different types of research models. Four types of research models can be proposed: direct effects of one or more independent variables on dependent variables (Type A); moderating effects on direct relationships between one or more independent variables and dependent variables (Type B); direct and mediating effects of independent variables on dependent variables, including interrelationships between independent variables (Type C); and moderating effects on the direct and mediating relationships between independent variables and dependent variables, including interrelationships between independent variables (Type D). Each type of research model can be examined using the exploratory or confirmatory approach and the derived metrics or reported effect sizes, as shown in the 2x2 matrix. Table 1 depicts the various types of research models seen in MA along with illustrative examples.

**Models with Direct Effects only (Type A)**

These research models examine relationships involving one or more independent variables with the same dependent variable(s). Jeyaraj et al. (2006), for instance, identified the independent variables that influenced specific dependent variables such as intention to use, adoption, diffusion, and outcomes. This was done for each relationship using a derived metric computed as the number of times it was significant across the number of times it was examined across studies. He and King (2008) examined the effects of user participation on variables such as user satisfaction, system use, team performance, project quality, and project success. The correlation effect size was used to determine the extent to which user participation impacted other variables.
Models with Moderating Effects on Direct Effects only (Type B)

These research models aim to examine the ways in which the direct effects may be altered by the presence of one or more moderators. Sabherwal and Jeyaraj (2015) conducted a meta-regression to determine the impacts of several moderators on the relationship between IT investment and the business value of IT. Moderators included IT alignment, inter-organizational IT, IT infrastructure or capability, IT adoption or use, and IT progress among others. Hameed et al. (2012) examined the relationships between several variables such as top management support and organization size on IT adoption. Subgroup analysis was conducted to assess the extent to which moderators such as the type of innovation (product vs. process), type of organization (manufacturing vs. service), and size of organization (large vs. small-and-medium) altered the main relationships. Sharma and Yetton (2003) employed regression methods to assess the moderating impact of task interdependence on the relationship between top management support and implementation success.
<table>
<thead>
<tr>
<th>Type</th>
<th>Research Model</th>
<th>Q1: Exploratory MA with Derived Metrics</th>
<th>Q2: Confirmatory MA with Derived Metrics</th>
<th>Q3: Exploratory MA with Reported Effect Sizes</th>
<th>Q4: Confirmatory MA with Reported Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| D | Moderating effects on 
models of type C | Jeyaraj (2020) | Zhang et al. (2012) | Wu and Lederer 
(2009) 
Zhao et al. (2018) | Montazami and 
Qahri-Saremi (2015) |
|---|------------------|----------------|------------------|-------------------|------------------|

**Table 1.** Research Models in Meta-Analysis
**Models with Direct and Mediating Effects (Type C)**

These models include interrelationships among the independent variables, and thus mediating effects, in addition to the direct effects on the dependent variable(s). Different analysis techniques are possible in the context of MA for these types of research models. King and He (2006), in examining the Technology Acceptance Model (TAM), analyzed the effect sizes for each bivariate relationship in the model (i.e., perceived usefulness → behavioral intention, perceived ease of use → behavioral intention, and perceived ease of use → perceived usefulness). Sabherwal et al. (2006), on the other hand, portrayed an extended IS success model that included IS success constructs refined in prior studies, user constructs (e.g., user attitude), and contextual constructs (e.g., facilitating conditions). The corrected correlation effect size for each bivariate relationship was obtained using a MA, and the collection of effect sizes were used in a meta-analytic structural equation modeling (MASEM) analysis, which yielded insights into IS success and its antecedents, and helped partially define the nomological net for IS success.

**Models with Moderating Effects on Direct/Mediating Effects (Type D)**

These research models aim to examine the ways in which the direct and mediating effects may be altered by one or more moderators. Jeyaraj (2020), in examining the IS success model (DeLone and McLean 1992; 2003), analyzed the ways in which the various bivariate relationships were altered by the moderators such as the geographic region, type of respondents, and use context. Wu and Lederer (2009) proposed that environment-based voluntariness would moderate the relationships underlying the TAM. The correlation effect sizes for all relationships in the TAM were aggregated across all studies, voluntariness was coded for each study, and weighted least squares regression techniques were used to assess the extent to which voluntariness altered the relationships found in the TAM. Montazemi and Qahri-Saremi (2015) examined the intention to use online banking using two structural models, one for the pre-adoption stage and the other for the post-adoption stage, which shared common constructs such as trust in online banking, perceived usefulness of online banking, and perceived ease of use of online banking.

**Key Recommendation:** Consider: a) direct effects or direct/mediating effects models when the basic relationships between constructs are of interest, and b) moderating effects models to assess how moderators alter the relationships between constructs.
3. Challenges in Meta-analysis Research in IS

MA undertakings are rife with challenges that need to be addressed. Before addressing the challenges, however, it is best to establish the rationale for the MA study as can be illustrated from prior studies: Kohli and Devaraj (2003) noted that studies have not demonstrated clear payoff from IT investment, Wu and Lederer (2009) argued that the role of environment-based voluntariness has not been effectively incorporated in models of technology acceptance, Wu and Du (2012) raised the relative lack of attention to system usage compared to behavioral intention, Gerow et al. (2014) noted the "alignment paradox" (i.e., IT alignment resulting in no improvement or even decline in organizational performance), and Jeyaraj (2019) observed the mixed results for the relationship between system usage and individual impact. Table 2 shows the challenges that apply to the types of MA raised by the organizing framework.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Q1: Exploratory MA with Derived Metrics</th>
<th>Q2: Confirmatory MA with Derived Metrics</th>
<th>Q3: Exploratory MA with Reported Effect Sizes</th>
<th>Q4: Confirmatory MA with Reported Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication bias</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Search methods</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Inclusion and exclusion</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Missing effect sizes</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Coding considerations</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Types of effect sizes</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Modeling considerations</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Moderator analysis</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Hypotheses testing</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Table 2. Challenges mapped by Organizing Framework

3.1. Publication Bias

Also known as the file-drawer problem, publication bias refers to the condition that studies showing significant effects for relationships of interest are more likely to be published especially in scientific journals while those with non-significant effects may remain unpublished and consigned to the file drawer (e.g., Rosenthal 1979). Funnel plots and failsafe N (e.g., Sabherwal et al. 2006; Wu and Lederer 2009; Kepes and Thomas 2018) may be used to detect issues with publication bias. To
counter publication bias, a primary recommendation is to expand the range of sources from which potential studies for the MA can be acquired (e.g., King and He 2005).

There is considerable variation in the sources used for acquiring studies for MA research within the IS domain. For instance, Kohli and Devaraj (2003) and Sabherwal and Jeyaraj (2015) used journals, conference proceedings, doctoral dissertations, book chapters, and working papers; Lowry et al. (2017) and Baptista and Oliveira (2019) used journals, conference proceedings, doctoral dissertations, and book chapters; Sharma and Yetton (2003) and Gerow et al. (2014) used journals, conference proceedings, and doctoral dissertations; Dennis et al. (2001) and Kapoor et al. (2014a) used journals and conference proceedings; and Kroenung and Eckhardt (2015) and Jeyaraj (2019) identified studies from journals only to identify studies for MA. Variation can be seen in the outlets allowed for the search. For instance, Kroenung and Eckhardt (2015) preselected a set of 14 journals from which to identify articles while Jeyaraj (2019) did not restrict the journals.

A significant challenge is the non-availability of or lack of access to research that found non-significant or adverse effects for relationships since such research may remain unpublished. This challenge could be addressed in a few ways. Kohli and Devaraj (2003) contacted active researchers in the topic area of their research to solicit working papers. Institutional or other (e.g., SSRN) repositories might have access to working papers that may not have been published. Researchers may also solicit unpublished work from researchers through listserves such as AISWorld.

**Key Recommendation:** Identify studies from a broad and unrestricted range of sources, including doctoral dissertations, book chapters, and working papers, to minimize publication bias.

### 3.2. Search Methods

A common approach to identify studies has been the use of electronic databases such as EBSCO, ABI/INFORM, SCOPUS, ACM, IEEE, JSTOR, Web of Science, PsychInfo, and ScienceDirect for journals, AISeL for conference proceedings, and ProQuest Dissertation Abstracts for doctoral dissertations (e.g., Sabherwal et al. 2006; Roberts et al. 2017; Baptista and Oliveira 2019; Dwivedi et al. 2019; Mandrell et al. 2020; Tao et al. 2020). Harzing's Publish or Perish and Google Scholar databases were also used to identify studies (e.g., Roberts et al. 2017; Dwivedi et al. 2019). Sharma and Yetton (2003) examined bibliographies of shortlisted studies for MA to identify other studies while
Mandrella et al. (2020) screened studies included in prior MA in the relevant area. Roberts et al. (2017) searched for articles in press in the “AIS Basket of 8” journals. Wu and Lederer (2009) conducted a manual search of the back issues of journals that were not available on the electronic databases and also contacted researchers in the IS community through the AISWorld mailing list.

The combination of keywords used for the search on electronic databases assumes considerable importance for MA. While keywords are heavily dependent on the specific research area handled in the MA, efforts to more accurately and completely specifying the set of keywords for search may be crucial for identifying the studies for MA. It is dependent on the domain knowledge to some extent. For instance, “system use”, a construct typically found in individual-level studies of technology acceptance (e.g., Jeyaraj et al. 2006), may be known variously as "system usage" and "system utilization" and also be conceptualized in different ways such as “frequency of use”, “extent of use”, and “duration of use” (e.g., Burton-Jones and Straub 2006). The search results would likely differ depending on the combination of keywords chosen since “system use” is not the name universally used across all studies. The results are also dependent on the particular databases used since not all search engines are created equal. For instance, EBSCO allows the keywords to be applied to various fields (e.g., abstract, title, text) and the results to be restricted by other characteristics (e.g., scholarly works, publication year) but AISeL provides a more limited set of options for the search.

**Key Recommendation:** Employ a combination of both electronic and manual search for articles to maximize the number of studies that could be included in the meta-analysis.

### 3.3. Inclusion and Exclusion

MA requires researchers to judiciously juggle the inclusion and exclusion of prior studies for the analysis. The process of inclusion and exclusion could be iterative as well as subjective, and the criteria for inclusion and exclusion may evolve as the MA effort is underway. For instance, Kapoor et al. (2014a) mentioned a decision made to include studies published since Rogers (1995) in the search although the original aim was to find studies published since Rogers (2003).

Within the IS domain, MA studies have generally provided clarity on the ways in which the sample of studies for the analysis was finalized. The criteria for inclusions were explicitly laid out while the criteria for exclusion could have been implicit at times. The criteria typically related to whether the
studies: a) dealt with empirical research (i.e., conceptual studies and case studies were excluded), b) examined the constructs and relationships of interest for the MA (i.e., study excluded otherwise), and c) reported the effect sizes necessary for the MA (e.g., study included if Pearson correlation was available but excluded otherwise) (e.g., Sabherwal et al. 2006; Wu and Lederer 2009; Roberts et al. 2017; Dwivedi et al. 2019). During the coding process later, studies may be excluded if they used the same data set as another study already included in the MA sample to preserve the independence of observations in the MA (e.g., Sabherwal et al. 2006).

However, inclusion and exclusion criteria at a more macro level are seen in MA studies within the IS domain, especially in the context of systematic reviews that seek to identify the entire corpus of relevant literature. The large volume of potential studies identified using multiple electronic databases exacerbate the situation further. For instance, Kapoor et al. (2014a) mentioned that only a proportion of the retrieved studies were available for download; Tamilmani et al. (2019) described the removal of duplicate studies found from different databases; and Tao et al. (2020) highlighted the exclusion of studies based on the titles and abstracts of studies and the inclusion of studies written in English.

**Key Recommendation:** Specify the exclusion and inclusion criteria such that the sample of studies in the meta-analysis is an appropriate representation of prior research in the chosen area.

### 3.4. Missing Effect Sizes

As highlighted earlier, studies are typically excluded from the MA sample when effect sizes required for the MA are not reported. For instance, Sabherwal et al. (2006) dropped Sethi and King (1994) from the MA since the correlation between user information satisfaction and independent variables was not reported but included Santhanam et al. (2000) as the correlation between user satisfaction and top management support was available. Studies, especially those that employ regression or SEM, may not always report the correlation matrix that can provide necessary effect sizes for a MA. Consequently, a significant proportion of empirical studies may not be included in the MA sample, which could bias the results of the MA study.

Although this approach of excluding studies is common, MA studies within the IS domain have adopted different strategies to overcome exclusion of studies due to missing effect sizes. One
strategy has been to code statistics that may be converted to Pearson correlation (assuming that the primary statistic needed for the MA is the Pearson correlation). Wu and Lederer (2009), for instance, provided several conversion formulas to compute the correlation from other statistics that could be coded from the studies. Another strategy has been to directly contact the authors of the published studies and ask if the correlations may be shared (although the general assumption in MA is to employ the published statistics). For instance, Roberts et al. (2017) and Mandrella et al. (2020) contacted the authors to obtain correlations that were not reported.

**Key Recommendation:** Adopt different strategies to the extent possible to handle missing effect sizes and include studies in the meta-analysis sample.

### 3.5. Coding Considerations

Gathering data from the studies included in the MA is a significant endeavor, generally known as coding. It can be straightforward in certain respects: a) to identify characteristics of the empirical study such as the type of respondents, the geographic region, and the period of data collection, and b) to capture details of the relationship such as the sample size, the reliabilities of constructs, and the effect size.

Additional steps may be needed in other situations. For instance, Sabherwal and Jeyaraj (2015) used a derived metric to represent the business value of IT (BVIT), which represented the combined result of all relationships between IT and payoff within the same study. That derived metric required the authors to code the result of each relationship as negative (n), positive (p), or zero (z), which were used to compute the BVIT as: \( \frac{(\sum p - \sum n)}{(\sum p + \sum n + \sum z)} \times 100 \). To differentiate between developed and developing regions, Mandrella et al. (2020) employed the classification by the International Monetary Fund while Sabherwal and Jeyaraj (2015) relied on the classification by the United Nations. Sabherwal and Jeyaraj (2015) used the distinctions between the mainframe, personal, network, and mobile computing eras to represent IT progress over time. Lacity et al. (2010) and Lacity et al. (2011) coded prior findings on IT and business process outsourcing and classified the variables examined in studies into different meta-categories. To verify the accuracy of

---

8 Glass et al. (1981), Hedges and Olkin (1985), Lipsey and Wilson (2001), and Borenstein et al. (2009) may be used to obtain formulas for converting between effect sizes.
the coding effort, Lacity et al. (2010) and Lacity et al. (2011) also contacted the authors of a randomly-selected subset of studies that were included in the analysis and solicited their opinions on the extent to which they agreed with the categorizations.

Prior conceptualizations may also be used in the coding process. Roberts et al. (2017) used the classification of IT innovations by Swanson (1994) to distinguish between Type I, Type II, and Type III innovations examined in prior studies. Wu and Lederer (2009) coded voluntariness using the descriptions found in each study and the four-item scale developed by Moore and Benbasat (1991). Sharma and Yetton (2003) used the six-item scale developed by Pearce et al. (1992) to code task interdependence based on the descriptions available in each study. Sharma and Yetton (2007) used a similar approach to code technical complexity based on measures developed by Attewell (1992), Premkumar and Roberts (1999), and Cho and Kim (2001). Gerow et al. (2014) coded alignment on several dimensions using the framework by Henderson and Venkatraman (1999). These typically require multiple coders, who independently code the necessary variables based on the definitions. The inter-rater reliability can be assessed using Cohen’s kappa (e.g., Sabherwal et al. 2006) to gauge the level of agreement between the coders. Disagreements in coding could be resolved via verbal discussion (e.g., Roberts et al. 2017).

Coding rules may need to be developed and refined during the coding process, which may result in an iterative coding effort. For instance, Sabherwal and Jeyaraj (2015) developed several rules to navigate the coding process such as: code the results based on regression analysis when the study reported both correlation and regression results, code the results for full regression models when the study reported both partial and full regression models, and code results based on longitudinal analysis when the study reported results from both cross-sectional and longitudinal models. Lacity et al. (2010) and Lacity et al. (2011) describe the development of a master list of coded variables that were iteratively refined over time.

As highlighted earlier, duplicate studies should be excluded from the MA to ensure independence of observations (e.g., Sabherwal et al. 2006). Duplicate studies result from multiple studies using the same dataset (e.g., Roberts et al. 2017 included Rai et al. 2009 but not Rai et al. 2006 in the MA). Independence of observations may also be adversely affected when a single study contributes multiple correlations for the same relationship (e.g., Roberts et al. 2017 reported multiple correlations from Chatterjee et al. 2001). This requires the computation of a composite effect size.
(e.g., Sabherwal et al. 2006; Roberts et al. 2017). In situations where a single study provided the results of multiple datasets separately (e.g., Sabherwal and Chan 2001; Venkatesh et al. 2003), then each dataset should be coded separately, i.e., one study contributes multiple observations (e.g., Sabherwal and Jeyaraj 2015).

**Key Recommendation:** Develop appropriate coding strategies and rules to consistently gather the data from studies for the meta-analysis.

### 3.6. Types of Effect Sizes

MA methods have classically portrayed two types of effect sizes that can be used to aggregate prior findings. The first involves differences in means between groups such as Cohen’s d, Glass’ Δ, and Hedges’ g while the second deals with the Pearson correlation for a relationship between two constructs (Glass et al. 1981; Hedges and Olkin 1985; Hunter and Schmidt 1990). The use of correlation effect sizes has been the more dominant trend in MA within the IS discipline (e.g., Sabherwal et al. 2006; Joseph et al. 2007; Roberts et al. 2017; Dwivedi et al. 2019; Mandrella et al. 2020; Tamilmani et al. 2020). This trend may be partly due to the nature of IS inquiry which emphasizes the study of relationships between constructs, influence of factors on dependent variables, and mediating and moderating effects on relationships to a greater extent than focus on the differences between two or more groups. There are some exceptions however: Lamb et al. (2018) employed Cohen’s d while Kroenung and Eckhardt (2015) used the significance level (p-value) for the MA.

Studies also utilized the regression (beta) coefficient as the effect size for MA (e.g., Kapoor et al. 2014a; 2014b; Rana et al. 2015; Baptista and Oliveira 2016; Tamilmani et al. 2019; Baptista and Oliveira 2019). Precedents for using beta coefficients are scattered in prior literature. Wu and Lederer (2009) highlighted that the beta coefficient may be converted to correlation in a limited setting, i.e., the regression model has only a single independent variable. King and He (2006) demonstrated that there were no significant differences between correlations and beta coefficients in a specific context, i.e., the analysis of a specific set of relationships portrayed by the TAM. Peterson and Brown (2005) argued that beta coefficients can be used to impute missing correlation or difference measures. However, these situations are atypical for the IS domain, which largely involves complex research models involving mediators (and moderators at times)—e.g., the Unified
Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al. 2003). Regression slopes and intercepts are generally not comparable across studies (Hunter and Schmidt 2004). Further, the beta coefficients were not used for imputation in a few instances but as the primary effect size in several MA studies in IS. While the use of beta coefficients has the potential to increase the number of studies included in the MA (i.e., studies that reported regression coefficients but not correlations need not be excluded but could be retained in the meta-analysis), the synthesis of regression slopes in a MA is not well understood and is subject to a variety of problems (Becker and Wu 2007). Roth et al. (2018) urged a return to the practice of using correlations in MA.

**Key Recommendation:** Conduct meta-analysis using the reported correlations or group mean differences as the effect sizes consistent with practice.

### 3.7. Modeling Considerations

MA methods distinguish between fixed-effect and random-effects models based on the assumptions about the population from which the studies are drawn (Hedges and Vevea 1998; King and He 2005). Fixed-effect models assume that one true effect size underlies all studies included in the analysis whereas random-effects models allow for the true effect sizes to differ between studies (Hunter and Schmidt 2000; Borenstein et al. 2010). Both models assume heterogeneity but differ in attribution—fixed-effect models account for within-study variation whereas random-effects models allow for both within-study and between-study variations (e.g., Borenstein et al. 2010).

Random-effects models are generally more appropriate since MA studies synthesize findings from studies that differ in the types of participants, technologies, and locations. For instance, Tamilmani et al. (2019) conducted a MA of hedonic motivation based on the UTAUT2 model (Venkatesh et al. 2012), in which the primary studies included in the sample had dealt with different types of participants (e.g., students, consumers, working professionals, citizens), different types of technologies (e.g., mobile health, learning management, smart phones, wearable healthcare, banking, mobile social networking), and different locations (e.g., Malaysia, South Africa, China, Germany, Spain, Bangladesh, USA).

Weighting schemes differ between the MA methods (e.g., Hedges and Olkin 1985; Hunter and Schmidt 1990; Borenstein et al. 2009) but may be used to specify fixed-effect and random-effects
models. The Q and I² tests could be used to assess heterogeneity and determine if fixed-effect or random-effects models are acceptable for the MA (e.g., Huedo-Medina et al. 2006). However, the choice between fixed-effect and random-effects models should be based on an understanding of the population of studies rather than statistical tests (e.g., Borenstein et al. 2009; Hwang and Schmidt 2011).

In addition to corrections for sampling errors that can handle within-study and between-study variances, MA methods allow for corrections for measurement errors, range variations, dichotomizations, deviations from construct validity, and transcriptional or extraneous errors (Hunter and Schmidt 1990). Measurement errors may be handled by correcting each observed correlation using the reliabilities of the constructs if the reliabilities are available, or by correcting the aggregated observed correlation using the average of the reliabilities available across all observations for the relationship (e.g., Roberts et al. 2017). MA studies in the IS discipline predominantly attended to sampling errors and measurement errors in primary studies (e.g., Sabherwal et al. 2006; King and He 2006; Wu and Lederer 2009; Gerow et al. 2014; Dwivedi et al. 2019). This is perhaps not as surprising since the vast majority of constructs in IS research are measured using continuous or Likert scales, and corrections for dichotomizations or range variations may not be as crucial.

**Key Recommendation**: Employ random-effects models in meta-analysis to handle both within-study and between-study variances.

### 3.8. Sensitivity Analysis

Outliers and influential cases (e.g., Aguinis et al. 2013; Kepes and Thomas 2018; Wu et al. 2018) typically exist within data, including those used for MA. Extremely small or large effect sizes, or the inclusion of non-Pearson correlations such as tetrachoric correlations may artificially inflate the results (Hunter and Schmidt 2004). Since a typical MA with effect sizes may be influenced by the sample size and the construct reliabilities, it may be useful to examine those data for anomalies. For instance, Sabherwal et al. (2006) successively dropped one observation at a time based on the
weighted deviates to identify potential outliers\(^9\) and compared the resultant correlations for the homogeneous set of observations (i.e., observations excluding outliers) and the set of all coded observations. Techniques such as winsorizing, trimming, and standardized residuals may be used to detect and handle outliers in MA (e.g., Huffcutt and Arthur 1995; Sharma and Yetton 2011; Ada et al. 2012). However, labeling effect sizes as outliers is problematic since extreme values may be due to large sampling errors, which can be corrected using meta-analysis methods (Hunter and Schmidt 2004). Since MA methods strive to synthesize all available observations, it becomes a challenge to selectively exclude observations. Further, other factors such as common method variance (e.g., Sharma et al. 2009; Cram et al. 2019) and reliability (e.g., Hess et al. 2014) could also impact results and may be considered for sensitivity analysis.

**Key Recommendation:** Conduct sensitivity analysis with caution to ensure that outliers or other factors do not adversely impact results.

### 3.9. Moderator Analysis

The identification of moderators is a significant exercise in MA. Moderators have the potential to explain differences in the magnitudes and signs (i.e., positive or negative) of the effect sizes or derived metrics representing relationships (e.g., King and He 2005; Hwang and Schmidt 2011). To enable moderator analysis, appropriate moderator variables need to be defined and coded—e.g., Wu and Lederer (2009) created a variable to represent voluntariness; Hameed et al. (2012) distinguished between large organizations and small-and-medium enterprise (SMEs); Roberts et al. (2017) captured the stage of assimilation differentiating between Guttman and non-Guttman scales; Lamb et al. (2018) separated three-dimensional and two-dimensional games; and Zhang et al. (2018) represented national culture.

Different approaches could be employed to detect if moderator effects are supported. The most common approach is to evaluate the credibility interval surrounding the corrected effect size and the percentage of variance explained statistic (e.g., Hunter and Schmidt 1990). Moderator effects are signified when the credibility interval includes 0 in its range and the percentage of variance explained is low (e.g., Roberts et al. 2017). When moderator effects are detected, the overall set of

\(^9\) Sabherwal et al. (2006) also excluded observations in which the construct reliabilities were below 0.60, which itself is below the recommended threshold of 0.70 (Nunnally 1978).
observations is split into multiple subgroups, and the MA methods are applied on the subgroups (e.g., Wu et al. 2011; Hameed et al. 2012; Lowry et al. 2017; Lamb et al. 2018). Another approach to confirm moderators is the use of regression methods (e.g., Hwang and Schmidt 2011). In such regression models, the dependent variable could be the reported effect sizes (e.g., Sharma and Yetton 2003) or the derived metrics (e.g., Sabherwal and Jeyaraj 2015) and the significance of the moderator variable can be used to conclude if moderators are influential. Regression methods have gained momentum in MA studies in IS (e.g., Kohli and Devaraj 2003; Sharma and Yetton 2007; Wu and Lederer 2009; Zhang et al. 2018; Jeyaraj 2019).

**Key Recommendation:** Conduct moderator analysis to conclude if the variation in results of primary studies may be influenced by moderators.

### 3.10. Hypotheses Testing

The results of MA in the IS domain have been used for hypotheses testing. There is considerable diversity in the types of hypotheses examined: a) directional hypothesis for a specific relationship (e.g., Weigel et al. 2014 expected a negative relationship between complexity and IT adoption while Petter and McLean 2009 proposed a positive relationship between system quality and system use); b) moderating effect of a variable on a specific relationship (e.g., Sharma and Yetton 2007 proposed that task interdependence will moderate the relationship between training and IS implementation success while Roberts et al. 2017 hypothesized that innovation type will moderate the relationship between knowledge diversity and IT assimilation); and c) inter-relationships between constructs (e.g., Sabherwal et al. 2006 tested a model of IS success involving 10 constructs and several mediating effects while Dwivedi et al. 2019 proposed a research model which incorporated attitude into the base UTAUT model).

Methods used to evaluate the hypotheses were diverse as well. A barebones MA (e.g., Hunter and Schmidt 1990) that corrected sampling errors only or also corrected for measurement errors was sufficient to evaluate directional hypotheses. For instance, Petter and McLean (2009) found a positive corrected correlation for the relationship between system quality and system use through a MA and concluded support for the hypothesis. Hypotheses with moderators were examined using different approaches, influenced partly by the moderator type and partly by the type of statistics chosen for MA. Roberts et al. (2017) coded a binary moderator to represent the type of IT
innovation and used the correlation effect size as the statistic for the MA; thus, the hypothesis test was based on a subgroup MA. Sharma and Yetton (2007) coded a scale-based moderator for task interdependence and the correlation effect size as the statistic; hence, the hypothesis test was based on a weighted least squares regression. Sabherwal and Jeyaraj (2015) used a binary moderator to represent IT progress and employed derived metrics to represent BVIT; and the hypothesis test was based on an ordinary least squares regression. Both Wu and Lu (2013) and Managalaraj et al. (2017) had coded binary moderators and used correlation effect sizes, but Wu and Lu (2013) used T-tests and Mangalaraj et al. (2020) used Z-tests to test the difference in the means of effect sizes due to the moderators. When research models involving inter-relationships between multiple constructs were proposed, meta-analytic structural equation modeling (MASEM) was more commonly used (e.g., Sabherwal et al. 2006; Joseph et al. 2007; Montazemi and Qahri-Saremi 2015; Hamari and Keronen 2017; Dwivedi et al. 2019).

**Key Recommendation:** Examine hypotheses using regression and structural equation modeling methods depending on the type of research model.

4. **Opportunities for Meta-analysis Research in IS**

MA offers several opportunities in addition to its basic capabilities such as aggregating prior findings, resolving inconsistent findings, determining true effect sizes by correcting for sampling and measurement errors, and identifying moderators that may explain the variance in effect sizes (e.g., Hunter and Schmidt 1990).

4.1. **Unexplored Areas**

An analysis of the research areas seen in prior MA studies in IS can provide insights into potential opportunities for future research. **Figure 3** showcases the 2x2 organizing framework with illustrative studies for each quadrant, categorized by the two predominant units of analysis in IS research.
**Figure 3.** Illustrative prior MA studies by Unit of Analysis (UoA)
Prior MA studies have accorded greater attention to the individual unit of analysis, i.e., the behaviors of individuals. Few research areas garnered significant attention in MA. Several studies conducted MA of specific models such as the TAM, UTAUT and its extensions (Legris et al. 2003; King and He 2006; Schepers and Wetzels 2007; Wu et al. 2011; Dwivedi et al. 2019; Tamilmani et al. 2019; Tamilmani et al. 2020; Tao et al. 2020), Theory of Planned Behavior (TPB) (Weigel et al. 2014), and Innovation Diffusion Theory (IDT) (Kapoor et al. 2014a). Studies analyzed factors influencing IT adoption and usage (Mahmood et al. 2001; Wu and Lederer 2009; Wu and Du 2012; Zhang et al. 2012; Wu and Lu 2013; Montazemi and Qahri-Saremi 2015; Rana et al. 2015; Baptista and Oliveira 2016; Zhang et al. 2018). Studies also examined various aspects of IS success (Sharma and Yetton 2003; Bokhari 2005; Sabherwal et al. 2006; Sharma and Yetton 2007; Petter and McLean 2009; Hwang 2014; Jeyaraj 2019; Jeyaraj 2020). Other research areas include: games (Hamari and Keronen 2017; Lamb et al. 2018), piracy (Lowry et al. 2017; Eisend 2019), security compliance (Trang and Brendel 2019; Cram et al. 2019), trust (Wang et al. 2016; Yu et al. 2020; Sarkar et al. 2020), online reviews (Hong et al. 2017), user participation (He and King 2008), consumer satisfaction (Ghasemaghaei and Hassanein 2015), turnover (Joseph et al. 2007), and electronic word-of-mouth (Qahri-Saremi and Montazemi 2019; Ismagilova et al. 2019; 2020).

MA studies on the organization unit of analysis (i.e., the behaviors of organizations, typically based on decisions by key organizational decision-makers) have been few and restricted to certain research areas: IT adoption and assimilation (Jeyaraj et al. 2006; Lee and Xia 2006; Narayanan et al. 2009; Hameed et al. 2012; Roberts et al. 2017; Mangalaraj et al. 2020), IT alignment (Gerow et al. 2014), IT integration (Henningsson et al. 2018), IT and business process outsourcing (Lacity et al. 2010; Lacity et al. 2011; Lacity et al. 2016; Schermann et al. 2016), and BVIT and IT payoff (Kohli and Devaraj 2003; Lim et al. 2011; Ada et al. 2012; Sabherwal and Jeyaraj 2015; Mandrell et al. 2020). Opportunities for MA may exist in several other areas such as IT capability, IS management, IS development, IS implementation, IS governance, cybersecurity, knowledge management, enterprise systems, and e-business, several of which emerged as popular topics in the IS domain (e.g., Jeyaraj and Zadeh 2020). Few studies have dealt with other units of analysis such as groups (Dennis and Wixom 2001; Dennis et al. 2001; Ortiz de Guinea 2012) and projects (Wang and Keil 2007; Hannay et al. 2009). It may be helpful to expand research involving such units of analysis as well as consider other areas such as teams, social networks, communities of practice, non-profit organizations, SMEs, governmental agencies, and societies.
**Key Recommendation:** Consider un- or under- explored areas for potential meta-analysis studies, especially in other units of analysis, and establish rationale for the study.

### 4.2. Ambiguous Constructs

MA affords the opportunity to refine and clarify constructs, and empirically test the constructs. A few illustrations may help shed light on this aspect. Sabherwal and Jeyaraj (2015), for instance, conducted a MA on factors that explain the differences in the BVIT seen across studies. While the analysis provided useful insights, it also highlighted the diversity of measures employed to represent BVIT, which include profitability, productivity, market share, sales, revenue growth, return on assets, return on equity, performance, and several others. BVIT measures are based on primary data obtained from respondents, financial statements, and computed metrics. An analysis that can develop a set of unique constructs to represent BVIT along with a quantitative MA could lend greater clarity and guidance. Bokhari (2005) conducted a MA on the relationship between system usage and user satisfaction based on the IS success model (e.g., DeLone and McLean 1992) and found a significant positive effect for the relationship. However, the IS success model portrays user satisfaction and system usage to influence each other (i.e., both system usage → user satisfaction and user satisfaction → system usage are possible), which was not handled in Bokhari (2005). A MA could be used to isolate the different roles of both constructs over time and help resolve outstanding questions involving the two constructs. Schepers and Wetzel (2007) examined system usage and its influential factors in the context of the TAM. System usage, however, has been employed to represent a variety of behaviors such as frequency of use, time of use, extent of use (e.g., Burton-Jones and Straub 2006), not all of which may be equally appropriate. For instance, time of use may be high for individuals with less experience but low for individuals with more experience, whereas frequency of use may differ based on task requirements. MA may be used to treat the different conceptualizations separately and provide greater clarity in understanding system usage.

**Key Recommendation:** Consider constructs that may be refined and clarified through cleaner coding in the meta-analysis.
4.3. Confounding Relationships

Despite the insights gained from MA studies, there may be opportunities to further clarify the relationships. A few illustrations from prior studies may serve to highlight the potential for future studies to refine our understanding. Sharma and Yetton (2003), for instance, examined the moderating effect of task interdependence on the relationship between top management support and implementation success. While the analysis of the moderating effect was noteworthy, the measurement of implementation success was confounded by the inclusion of both system usage and user satisfaction measures, based on the argument that both measures represented the implementation stage. However, system usage and user satisfaction have been treated as distinct constructs that can influence each other (e.g., DeLone and McLean 1992). Ideally, the moderating effect of task interdependence should be examined on two relationships: top management support and system usage as well as top management support and user satisfaction. Petter and McLean (2009) conducted a MA on the relationships underlying the DeLone and McLean (2003) model of IS success. Of the 14 hypothesized relationships involving 7 constructs, all of which were expected to be positively significant, two relationships involving service quality (on system use and user satisfaction) showed confidence intervals that included 0 and hence the positive effects were not supported. But the number of observations used to compute the cumulative effect sizes for both relationships were also low. Another relationship involving service quality and intention to use was untested because only one study was available. These suggest that a closer look at the relationships involving service quality could be valuable. Tamilmani et al. (2019) analyzed the extent to which hedonic motivation, a construct defined in the UTAUT2 model, had been used in empirical studies. In so doing, the study portrayed hedonic motivation as an antecedent to behavioral intention, performance expectancy, and effort expectancy. The portrayal of hedonic motivation is different from UTAUT2 and would benefit from clarification in future studies.

**Key Recommendation:** Consider novel, untested, and reciprocal relationships between constructs for examination in a meta-analysis.

4.4. Hidden Contingencies

One of the strengths of MA is the identification of moderators that can offer contingent explanations of established relationships in the IS domain. While moderators have been included in prior models
(e.g., Venkatesh et al. 2003), they are constrained by the research context typically restricted to specific settings (e.g., organizations, countries, systems). Since MA enables the aggregation of findings across studies, it becomes possible to examine moderators that may not be otherwise possible. For instance, Wu and Lederer (2009) and Jeyaraj (2020) incorporated the voluntariness of use as a moderator in examining the relationships underlying the TAM and the IS success model respectively. Other moderators commonly seen across MA studies include the type of respondents, type of systems or innovations, economic region, and national culture (e.g., Kohli and Devaraj 2003; Schepers and Wetzels 2007; Hameed et al. 2012; Gerow et al. 2014; Sabherwal and Jeyaraj 2015; Wang et al. 2016; Roberts et al. 2017; Zhang et al. 2018; Cram et al. 2019). Such contingencies are rather straightforward to code, analyze, and interpret but others such as task interdependence (Sharma and Yetton 2003), technical complexity (Sharma and Yetton 2007), fit model for alignment (Gerow et al. 2014), assimilation stage or measure (e.g., Roberts et al. 2017) may not be so although they can yield finer insights. Opportunities to model contingencies related to task, project, and team may be possible for MA studies.

**Key Recommendation:** Consider contingencies that can provide incisive and clear explanations of moderating effects underlying the relationships.

### 4.5. Theory Testing

MA is well positioned for theory testing\(^9\), especially those involving meta-analytic moderators. Consider the TAM (Davis 1989) as an illustration. Several studies have tested TAM in various contexts, including technologies, users, and locations (e.g., Adams et al. 1992; Szajna 1996; Venkatesh and Davis 2000). Each such study has the potential to determine if the three major relationships specified in TAM hold in the chosen context. In a MA, however, all such studies can be incorporated, moderator variables can be constructed to represent different aspects of the contexts across studies (e.g., moderator to describe the national culture), and the overall effectiveness of TAM across different contexts can be assessed. MA thus presents a significant advantage over primary studies that cannot examine across various contexts (without experiencing significant

---

\(^{10}\) It is somewhat more difficult to employ MA for theory generation (e.g., Guzzo et al. 1987) because MA relies extensively on established relationships for effect sizes that can be used for synthesis. Thus, novel direct or mediating relationships may not directly emerge from a MA. However, there are opportunities to refine and clarify relationships, including new relationships between existing constructs (Sabherwal et al. 2006) and to develop research models (Joseph et al. 2007).
costs). Hwang (2014) analyzed three different models—i.e., direct effect model, moderator effects model, and meditational model—involving the same three constructs: top management support, training, and implementation success. The study used both regression and subgroup analyses to examine the moderator effects model whereas the procedure in Viswesvaran et al. (1999) was used to examine the meditational model. Theory testing of models with interrelationships between constructs is also possible (e.g., Sabherwal et al. 2006; Joseph et al. 2007; Dwivedi et al. 2019).

**Key Recommendation:** Consider research models and research methods that enable theory testing, and possibly theory generation, through meta-analysis.

### 5. Conclusion

This paper was initiated with several goals: identify the state of MA research in IS, explicate challenges underlying MA research, extract the opportunities available for MA research, and raise usable recommendations for conducting MA research. **Table 3** presents several recommendations on various aspects related to MA.

Based on a review of prior MA studies in IS, this paper identified MA studies could use: reported effect sizes or derived metrics as the statistic for aggregation, adopt an exploratory or confirmatory approach to analysis, and examine different types of research models. Several aspects related to MA including publication bias, inclusion and exclusion of studies, effect sizes, coding, MA modeling, and sensitivity analysis are clarified in this paper. Opportunities for MA research including unexplored areas, ambiguous constructs, confounding relationships, hidden contingencies, and theory testing are also identified. This paper can serve to foster further MA research and theoretical advances in IS in the future.
<table>
<thead>
<tr>
<th>Category</th>
<th>Criterion</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>Overall approach for MA</td>
<td>Consider: a) exploratory analysis for obtaining preliminary results; and b) confirmatory analysis for testing hypotheses regarding the relationships.</td>
</tr>
<tr>
<td></td>
<td>Type of statistic for MA</td>
<td>Consider: a) derived metrics when all relevant studies are to be included in the meta-analysis; and b) reported effect sizes when the magnitude and direction of the relationship are to be assessed in the meta-analysis.</td>
</tr>
<tr>
<td></td>
<td>Research model</td>
<td>Consider: a) direct effects or direct/mediating effects models when the basic relationships between constructs are of interest, and b) moderating effects models to assess how moderators alter the relationships between constructs.</td>
</tr>
<tr>
<td>Challenges</td>
<td>Publication bias</td>
<td>Identify studies from a broad and unrestricted range of sources, including doctoral dissertations, book chapters, and working papers, to minimize publication bias.</td>
</tr>
<tr>
<td></td>
<td>Search methods</td>
<td>Employ a combination of both electronic and manual search for articles to maximize the number of studies that could be included in the meta-analysis.</td>
</tr>
<tr>
<td></td>
<td>Inclusion and exclusion</td>
<td>Specify the exclusion and inclusion criteria such that the sample of studies in the meta-analysis is an appropriate representation of prior research in the chosen area.</td>
</tr>
<tr>
<td></td>
<td>Missing effect sizes</td>
<td>Adopt different strategies to the extent possible to handle missing effect sizes and include studies in the meta-analysis sample.</td>
</tr>
<tr>
<td></td>
<td>Coding considerations</td>
<td>Develop appropriate coding strategies and rules to consistently gather the data from studies for the meta-analysis.</td>
</tr>
<tr>
<td></td>
<td>Type of effect sizes</td>
<td>Conduct meta-analysis using the reported correlations or group mean differences as the effect sizes consistent with practice.</td>
</tr>
<tr>
<td></td>
<td>Modeling considerations</td>
<td>Employ random-effects models in meta-analysis to handle both within-study and between-study variances.</td>
</tr>
<tr>
<td></td>
<td>Sensitivity analysis</td>
<td>Conduct sensitivity analysis with caution to ensure that outliers or other factors do not adversely impact results.</td>
</tr>
<tr>
<td></td>
<td>Moderator analysis</td>
<td>Conduct moderator analysis to conclude if the variation in results of primary studies may be influenced by moderators.</td>
</tr>
<tr>
<td></td>
<td>Hypotheses testing</td>
<td>Examine hypotheses using regression and structural equation modeling methods depending on the type of research model.</td>
</tr>
<tr>
<td>Opportunities</td>
<td>Unexplored areas</td>
<td>Consider un- or under-explored areas for potential meta-analysis studies, especially in other units of analysis, and establish rationale for the study.</td>
</tr>
<tr>
<td></td>
<td>Ambiguous constructs</td>
<td>Consider constructs that may be refined and clarified through cleaner coding in the meta-analysis.</td>
</tr>
<tr>
<td></td>
<td>Confounding relationships</td>
<td>Consider novel, untested, and reciprocal relationships between constructs for examination in a meta-analysis.</td>
</tr>
<tr>
<td></td>
<td>Hidden contingencies</td>
<td>Consider contingencies that can provide incisive and clear explanations of moderating effects underlying the relationships.</td>
</tr>
<tr>
<td></td>
<td>Theory testing</td>
<td>Consider research models and research methods that enable theory testing, and possibly theory generation, through meta-analysis.</td>
</tr>
</tbody>
</table>

**Table 3. Recommendations for Meta-Analysis Research**

**Acknowledgements**

The authors gratefully acknowledge the excellent comments and suggestions provided by the four reviewers, which have been immensely helpful in refining the presentation.
References


