早稲田大学 IT 戦略研究所

Research Institute of IT & Management, Waseda University

April, 2006

Apples and Oranges: Meta-analysis as a Research Method within the Realm of IT-related Organizational Innovation

Ryoji Ito
Doctoral Candidate at Graduate School of Commerce
Waseda University

RIIM Working Paper No.19

Working Paper

Apples and Oranges: Meta-analysis as a Research Method within the Realm of IT-related Organizational Innovation

by

Ryoji Ito*

Abstract. The purpose of this paper is to examine the applicability of meta-analysis to IT-related organizational innovation studies. The use of meta-analysis has been limited in IT-related organizational innovation studies (e.g., Damanpour, 1991). The reasons for the lack of use of meta-analysis in this field lie in the difficulty in comparing empirical results across the past studies because various empirical measures have been used for the same theoretical construct. The organization of this paper is as follows: First, to introduce the concepts of the apples and oranges problem in question; second, to explain the existing methods in meta-analysis to solve this problem; and finally, to critically examine Fichman (2001)'s approach that goes for resolving the apples and oranges problem in IT-related organizational innovation studies.

Keywords. apples and oranges problem; effect size; IT-related organizational innovation; meta-analysis; multiple operations; test of homogeneity

* Ryoji Ito is a doctoral candidate at Graduate School of Commerce, Waseda University.

Contents

1. Introduction and the purpose of this paper

The purpose of this paper is to consider the applicability of meta-analysis to IT-related organizational innovation studies. Meta-analysis, a quantitative or systematic form of literature review on a certain substantive question of interest is of relatively recent vintage in psychology (Glass, 1976), and is applied to various disciplines (with regard to statues of meta-analysis in several disciplines, see Guzzo, Jackson, and Katzell, 1987; Schulze, 2004; Dalton and Dalton, 2005). Thus, as Stuhlmacher and Gillespie (2005) described, it is convenient to think that this method is widely recognized as a powerful research method. Meta-analysis, however, is not without disadvantages and is the subject of harsh criticism (see Mullen, 1989, for a review). The most persistent criticism of meta-analysis has to do with apples and oranges problem, that is, the results of meta-analyses are not meaningful if they are aggregated over incommensurable study findings (Lipsey and Wilson, 2001). Consequently, to surmount the apples and oranges problem leads to the applicability of meta-analysis to IT-related organizational innovation studies.

2. Apples and Oranges problem

It is useful to commence with the introduction of general meta-analytic process. The reason comes from the fact that apples and oranges problem emerges around every corner. As noted earlier, meta-analysis is a systematic form of literature review on a certain substantive question of interest. Put differently, meta-analysis is only one of many ways to summarize, integrate, and interpret selected sets of scholarly works in the various disciplines. In consequence, there is no single correct way to perform a meta-analysis (Hall and Rosenthal, 1995). In strategic management studies, for example, a high majority of current meta-analyses utilize either the Hunter and Schmidt (1990) approach or the Hedges and Olkin (1985) approach, or some close derivation thereof. At the same time, overall, one widely accepted specification of the stages or conduct of a meta-analysis is presented by Cooper (1982). These approaches are underpinned by a similar process: formulating the problem, collecting and evaluating data, analyzing data, interpreting the result, and public presentation (for details, see Durlak and Lipsey, 1991; Cooper, 1998; Cooper and Lindsay, 1998). Among these, the main concern with this article, apples and oranges problem, mainly relates to the problem formulation stage in which the purpose of the review is elucidated and the question being asked is defined clearly enough so that only studies that address this question will be included in meta-analysis, data search stage where all relevant indexing and abstracting databases are searched and complemented by other techniques (e.g., browsing bibliographies of the articles retrieved) to ensure that all relevant published studies are retrieved, and analyzing stage where an analyst performs the statistical synthesis of study outcomes of included studies, draw appropriate inferences and conclusion, and examine threats to the validity of conclusion. In meta-analysis, the unit of analysis is the impact of variable X on variable *Y* (Rosenthal, 1984). Based on this, meta-analyst must select and define the independent and dependent variable of interest in problem formulation stage. In data search stage, if there are a limited number of study outcomes, it might make sense to broaden the definition of the independent or dependent variables. Or, it might be feasible to paint a broader picture of the construct by including multiple dependent variables. Also, when the data related to the variables are analyzed, analysts need to examine whether the data are part of same population or not. These efforts hold apples and oranges problem. It follows that the apples and oranges argument is the problem relating to “judgment calls”. Judgment calls are those decisions big and small that must be made in the absence of guidance from an objective rule or standard and introduce considerable subjectivity into the meta-analytic review process (Guzzo, Jackson, and Katzell, 1987). Accordingly, it is important to whittle away judgment calls to surmount the apples and oranges problem. Therefore, I intend to describe three “verdict apparatuses” in the remaining part of this article.

2-1. Multiple operations Approach

The first verdict apparatus is multiple operations approach, and the leading propounder is Cooper (1998). So, in this section, I describe this apparatus relying on Cooper (1998).

To some degree any synthesis of information from multiple study outcomes involves an aggregation of outcomes that are dissimilar. At the level of repeated observations of the same object, the same is true. Inevitably, mixing apples and oranges occurs in some degree. Even when studies are intended to be direct replications, exact replication probably cannot occur (Cooper, 1998).

Meanwhile, meta-analysts must be sensitive to the problem of attempting aggregation of too diverse a sampling of operations and study outcomes. Combining apples and oranges to understand something about fruit may make more sense than combining fruits and humans to understand something about organic matter. That is, it can be argued that it is a good thing to mix apples and oranges, particularly if one wants to generalize about fruit, and that studies that are exactly the same in all respects are actually limited in generalizability. Of course, the final criterion for the extensiveness of the sampling of operations is whether the level of generalization is appropriate to the question being asked and scientifically useful or not. The meta-analysts ask, “Does this level of generalization add up to our explanation and understanding of a phenomenon?”. Too diverse a sampling of study outcomes could obscure useful relationships within subgroupings of the outsomes and not provide information at the level of the more abstract categorization.

There are two potential incongruities that meta-analysts must be aware of. Those incongruities may arise because of the variety of operations in the literature. First, meta-analysts expecting to find many operations may begin a literature search with broad construct definitions. They may discover, however, that the operations used in previous relevant research are quite narrow. When such a circumstance arises, the meta-analyst narrows the conceptual underpinnings of the effort to be more congruent with existing operations. Otherwise, the conclusions will appear more general than warranted by the data. The opposite problem, using narrow concepts defined by multiple broad measures, can also confront a meta-analyst. The meta-analysts would then have faced the choice of either broadening the concept or excluding many studies. As the literature search proceeds, it is extremely important that meta-analysts take care to reevaluate the correspondence between the breadth or abstractness of their constructs and the variation in primary. In primary research, this redefinition of a problem as a study proceeds is frowned on. To the contrary, in research synthesis, it appears that some flexibility may be necessary and may indeed be beneficial.

There is a string assertion about the value of the multiple operationism (Webb, Campbell, Schwartz, Sechrest, and Grove, 1972). They define the multiple operationism as the use of many measures that share a construct definition but have different patterns of irrelevant components (Webb, Campbell, Schwartz, Sechrest, and Grove, 1972). The multiple operationism has positive consequences. The reason is that "once a proposition has been confirmed by two or more independent measurement process, the uncertainty of its interpretation is greatly reduced. The most persuasive evidence comes through a triangulation of measurement processes. If a proposition can survive the onslaught of a series of imperfect measures, with all their irrelevant error, confidence should be placed in it. Of course, this confidence is increased by minimizing error in each instrument and by a reasonable belief in the different and divergent effects of the sources of error (p. 3)".

If all or most of the measures encompassed in meta-analysis are at least minimally valid, the multiple operations can enhance construct-to-operation correspondence. According to Cooper (1998), this rationale is akin to the reasoning applied in classical measurement theory. That is to say, if a sufficient number of minimally valid items are present, small correlations between individual items on a test or questionnaire and a true score can amount to a reliable indicator. The test, or, conclusion of the meta-analysis is invalid, however, if the majority of items, or, operations bear no correspondence to the underlying construct or the items, that is, operations share a different construct to a greater degree than they share the intended one. This is true regardless of how many items, or, operations are involved. The meta-analyst needs to examine research designs for threats to the correspondence of operations and constructs. If the research designs uncovered by a literature search contain the same invalidating procedures, then the correspondence between operations and constructs is threatened.

Multiple operations go beyond introduce the potential for clearer inferences about construct variables. They are also the most important source of variance in the conclusions of different meta-analyses meant to address the same topic. A variety of operations can affect meta-analysis outcomes in the following way, variance in operational definition. The operational definitions used in two meta-analyses on the same topic can be different from one another. Two meta-analysts using an identical label for a construct can employ very different operational definitions. Each definition may contain some operations excluded by the other, or one definition may completely contain the other.

In summary, the existence of a variety of operations in research literatures presents the potential benefit of stronger inference if it allows the meta-analysts to rule out irrelevant sources of influence. However, if all or most of the operations lack minimal correspondence to the concept, or if all research designs share a similar confounding of unintended influences with intended ones, multiple operations do not ensure construct-to-operation correspondence. Again, apples and oranges criticism assumes that the meta-analyst aggregates findings of different phenomena. Sometimes that seems to be done, because the forms of operations are themselves different. The meta-analysts, however, judges these various forms as occurring in the context of a constant goal. Although studies may differ in form, it is appropriate to aggregate them if they measure the same phenomenon. A convergence, or “triangulation” of findings from methodologically varying studies lends credence to the validity of an effect (Campbell and Fiske, 1959). When a relationship remains constant though tested under a variety of circumstances, it is clearly robust.

2-2. Homogeneity test Approach

As Glass, McGaw, and Smith (1981) have described, meta-analysis is geared towards generalization. Accordingly, when done well, meta-analysis makes inferences about population characteristics and relationships using sample data. By comparison, meta-analysis examines studies as the individual data points, rather than examining individuals as data points (Stuhlmacher and Gillespie, 2005). In doing so, the notion of “effect size”, the centerpiece of a meta-analytic process is useful. In meta-analysis, the finding of a study is converted into an effect size estimate. Cohen (1988) has defined an effect size as “the degree to which the phenomenon is present in the population, or the degree to which the null hypothesis is false” (pp.9-10). To put it plainly, an effect size represents the magnitude of the relationship between two variables. There are several indicators of effect size (Glass’s Δ, Cohen’s d, Hedge’s unbiased estimate of d, Pearson product moment correlation r, Z_{Fisher}, and so forth), and these metrics can be converted from one metric to the other (see Mullen, 1989; Rosenthal, 1994; Schulze, 2004, for reviews). It follows that the primary goal is to derive the best estimate of the population effect size.

In general, there are two ways to conduct meta-analysis: fixed and random effects model. Under a fixed effects model, an effect size observed in a study is assumed to estimate the corresponding population effect with random error that stems only from the chance factors associated with subject-level sampling error in that study, say, the “luck of draw” associated with sampling the subjects in each study from the population of potential subjects (Lipsey and Wilson, 2001). Put simply, usually this model is used if the analysts have reason to believe that observed effect sizes are homogeneous. On the other hand, a random effects model assumes that each observed effect size differs from the population mean by subject-level sampling error plus a value that represents other sources of variability assumed to be randomly distributed. In sum, random effects models are appropriate whenever there is reason to suspect that the observed effect sizes are truly heterogeneous, or, they are not drawn from a single population. Thus, in the case of random effects model, combining found effect sizes means assessing the average size of the real effect. Meta-analysts need to make a choice between fixed and random effects model. There are lengthy debates in the light of how to formulate questions in scientific research or whether a choice is more realistic or not (see Hedges, 1994, for a review). In the first place, however, combining effect estimates across studies is reasonable only if the studies share a common population effect size, θ. In this article, I intend to get rid of subjective judges as far as possible. One of the apparatuses to surmount judgment calls is a statistical test of homogeneity.

A statistical test for the homogeneity of effect size is formally a test of the hypothesis:

$$H_0:\ \theta_1 = \ldots = \theta_k = \theta$$

$$H_1:\ \theta_i \neq \theta, \text{ for at least one } i,$$

and the homogeneity test is based on the Q statistic (Hedges and Olkin, 1985):

$$Q = \sum_{i=1}^{k} w_i \left(T_i - \hat{\theta}\right)^2 ,$$

where T_i is the individual effect size for $i = 1$ to k (the number of effect sizes), $\hat{\theta}$ is the weighted mean effect size over the k effect sizes, and w_i is the individual weight for T_i, generally the inverse of the sampling error variance (Hedges and Olkin, 1985) or the sample size (Hunter and Schmidt, 1990). Essentially, this is the sum of squared standard normal values, which follows a χ^2-distribution with $k-1$ degree of freedom when the null hypothesis of common effect sizes for all k estimates is true. Hence, if Q exceeds the critical value for a χ^2 with $k-1$ degree of freedom, then H_0 is rejected. A statistically significant Q, therefore, indicates a heterogeneous distribution. This adds up to giving reason to use the random effects model rather than the fixed effects model. An algebraically equivalent formula for Q that is computationally convenient form is

$$Q = \sum_{i=1}^{k} w_i T_i^2 - \frac{\left(\sum_{i=1}^{k} w_i T_i \right)^2}{\sum_{i=1}^{k} w_i}.$$

Hunter and Schmidt proposed an alternative approach to homogeneity testing that does not rely on formal significance testing. Their approach separates the observed effect size variability into two components: the portion attributable to subject-level sampling error and the portion attributable to other between-study differences. Basically, if sampling error accounts for 75% or greater of the observed variability, it is unlikely —but by no means impossible— that there is a moderating variable, that is, the distribution is homogeneous. As Hunter and Schmidt (1990) described, This “75% rule” is intended as a “rule of thumb”, not a strict cut-off point and researchers with explicit a priori hypotheses about specific between study effects are encouraged to test for those relationships even if the 75% rule is exceeded. The intent of this rule, however, is to force recognition of the role of sampling error in effect size variability across studies and to discourage post hoc exploration of the relationship between study characteristics and effect size when most of the observed variability is adequately explained by sampling error.

If meta-analyst find the effect size distribution is not homogeneous, whether the test using Q or 75% rule, he or she has three options for handling the situation: 1) Analyst assumes that the variability beyond subject-level sampling error is random, that is, derives from essentially random differences among studies whose sources cannot be identified, 2) Meta-analyst continues to assume a fixed effects model, but add the assumption that the variability beyond subject-level sampling error is systematic, that is, derived from identifiable differences between studies, and 3) The analyst assumes that the variance beyond subject-level sampling error is derived partly from systematic factors that can be identified and partly from random sources. These decision branches are referred to as standard random effects model, meta-regression model, and mixed effects model, respectively (for details, see Lipsey and Wilson, 2001).

3. Aggregated measure Approach

An important construct in the IT innovation literature is the extent of organizational innovation with information technology. Many different measures are used to capture this construct, including earliness of adoption, frequency of adoption (e.g., the number of adoptions across a set of innovations), and various dimensions of extent of implementation (e.g., internal diffusion, infusion, and routinization). Some measures have a narrow focus while others aggregate innovative behaviors across a set of innovations or stages in the assimilation lifecycle. Consequently, there appear to be some significant tradeoffs involving aggregation, that is, more aggregated measures can be more robust and generalizable and can promote stronger predictive validity, while less aggregated measures allow more context-specific investigations and can preserve clearer theoretical interpretations. The goal of aggregated measure approach is to shed light on the issue of aggregation in the measurement of IT-related innovation, and in particular, to develop prescriptions for when the tradeoffs are most likely to favor aggregation. Note that aggregation can take two basic forms that aggregating innovative behaviors across innovations (such as when number of adoptions is used) and aggregating across the assimilation lifecycle within organizations (such as when behaviors that occur in both early and late stages of assimilation are reflected in the measure). I describe the third verdict apparatus, aggregated measure approach, in line with Fichman (2001) in this section.

According to Daft (1978), organizational innovation is defined as “the adoption of an idea or behavior that is new to the organization adopting it”. Viewed in this light, organizational innovation is understood in many ways. In fact, IT-related literatures show that researchers often conceptualize innovation as around the organizational initiation, adoption, and implementation of one or more emerging technologies (Prescott and Conger, 1995). If organizations take these sorts of actions earlier, more frequently, and more intensively, then they are viewed as more innovative. Table 1 exhibits that the measures of IT innovation differ along two key dimensions. In the first dimension, measures such as aggregated initiation, aggregated adoption, and aggregated implementation mix behaviors across innovations. In the second dimension, the measures mix behaviors across the assimilation lifecycle. Some measures focus on a fraction of the assimilation process in organizations, and others mix behaviors within a broader range of the process. To sum up, in fact, whether intended or not, many studies is aggregated across both innovations and stages to some degree.

Whether aggregation is permissible or not is not important. Rather, it is important to detect the circumstances in which the potential benefits of aggregation outweigh the potential costs. For this end, there are six issues to be considered: 1) the primary objective of the research, 2) the validity of generalization across assimilation stages, 3) the effects of organizational characteristics, 4) the effects of innovation characteristics, 5) the effects of innovation substitutes and complements, and 6) the effects of reporting errors and idiosyncratic adoption.

Table 1. Measures of organizational innovation

Measure	Definition
Earliness of adoption	Relative earliness of adoption within a population of potential adopters.
Internal diffusion	The extent of use of an innovation across people, projects, tasks, or organizational units.
Infusion	The extent to which an innovation's features are used in a complete and sophisticated way.
Routinization	The extent to which an innovation has become a stable and regular part of organizational procedures and behavior.
Assimilation	The extent to which an organization has progressed through the assimilation lifecycle for a particular innovation stretching from initial awareness to full institutionalization.
Aggregated initiation	The frequency or incidence of innovation initiation.
Aggregated adoption	The frequency or incidence of innovation adoption.
Aggregated Implementation	The degree of implementation of innovations that have been adopted.

Source: Fichman (2001)

1) According to Fichman (2001), underlying three objectives promote the studies of organizational innovation with information technology: (1) identifying the determinants of innovation with respect to some particular technology, (2) identifying the determinants of generally "innovative" organizations, and (3) determining the role of certain theoretical factors in innovation, but not with an overriding interest in the technology or innovative organizations par se. These styles of research are referred to as technology-focused, innovativeness-focused, and factor-focused, respectively. (1) Technology-focused studies deal with a model to explains innovation from the perspective of a particular technology or class of technologies with similar characteristics (e.g., Grover and Goslar, 1993; Howard and Rai, 1993). The challenge of these studies is to identify the outstanding explanatory factors about the nature of innovation. The goal of these studies is to maximize explanatory power for one innovation (or innovation class) that is viewed to be especially important in order to derive managerial implications for how to successfully adopt and diffuse that particular technology or technology class. However, the primary concern of these studies is not the generalization beyond the innovation at hand. In practice, these studies almost always use single measure of innovation. On the contrary, in the case of an innovation class, aggregated measures are used (Grover and Goslar, 1993). (2) The target of innovativeness-focused studies is to find the properties of organizations that

innovate over time in diverse settings. These studies typically aggregate across technologies, but they don't come in predisposed to use measures that aggregate across assimilation stages. Interestingly, there are only two studies conducted by IT researchers that utilize a more general notion of organizational innovativeness with IT as the outcome variable (Armstrong and Sambamurthy, 1999; Lind and Zmud, 1991). And, both use unconventional measures to capture this concept. (3) The main themes of factor-focused studies are to understand the role of one or more theoretical factors in determining innovation (e.g., Cooper and Zmud, 1990; Fichman and Kemerer, 1997; Grover et al., 1997; Nilakanta and Scamell, 1990; Rai, 1995; Zmud, 1982). This type of study covers a broad range of areas, from one particular factor (Bretschneider and Wittmer, 1993) to testing a more general innovation model (Grover et al., 1997). And, most of these studies aim at generalization to the level of a class of related technologies at minimum, albeit single-innovation measures is used (e.g., Cooper and Zmud, 1990). Innovation classes can be defined narrowly or broadly, and the same innovation can belong to more than one class. In regard to this point, the important pillars are to detect the level of abstraction that is consistent with the theoretical model to be tested, and to use the prominent characteristics of innovations taken at that level of abstraction to help detect which factors will be most prominent in the context of the intended study. Accordingly, the use of measures that aggregate across innovations in some class for generalizability comes about because a theoretical model sees the landscape to the degree of some innovation class. Even if that is the case, there can be compelling reasons to prefer a single innovation measure even when the intended level of generalization would permit aggregation. It would be safe to say that innovativeness-focused studies must use aggregated measures unless there is some compelling reason against their use, with other things being the same. In a similar fashion, it would be fair to say that technology-focused studies must avoid measures that aggregate across technologies unless the focus is on a technology class.

2) It would be the paramount concern for researchers who consider the role of aggregation in measuring innovation that the extent to which the underlying theoretical model can be generalized across the assimilation lifecycle within organizations. Fichman (2001) explains this point with some concrete descriptions. According to Fichman and Kemerer (1997), theories that are driven by organizational learning are generalized across stages, for significant knowledge barriers rest on all assimilation stages. Also, there is another model that holds several predictors across assimilation stages (Meyer and Goes, 1988). They are interested in several variables that are expected to have the same direction of influence, independently of assimilation stage. These two studies use a certain date when assimilation stage is reached as the outcome variable. By implication, this suggests that they aggregate innovative behaviors across all of the stages where organizations go through by that date. Summing up, even if there are other factors that have less consistency of influence, these two studies advocate that variables do exist (e.g., resources, fit,

expertise, competitive environment) that promote or hinder progress throughout the assimilation process. But, there is some counterview (e.g., Downs and Mohr, 1976; Grover and Goslar, 1993). Their notion is that aggregating across stages is problematic if important determinants of innovative behaviors have differently directioned effects. The reason comes from that the facilitating influence of variables in one stage is offset by the inhibiting effect in other stages. Eventually, this situation ends up with a loss of explanatory power and instability of results across studies. For the meantime, Damanpour (1991)'s meta-analysis of 23 studies reveals that researchers utilize the strength of effects varied depending on whether aggregated initiation, aggregated adoption, or aggregated implementation as the operation of innovation. In most cases, these effects are in the same direction. To sum up, the extent to which a theoretical model generalizes across assimilation stages depends on the context of study and the characteristics of included variables. If that generalization appears warranted, there is no harm to aggregate across stages because the tradeoffs pertaining to robustness, generalizability, and clarity of theoretical interpretation prop up to do so. To the contrary, if such generalization contradicts plausible hypotheses, aggregation across stages should be avoided.

3) Generally, several characteristics of organizations including size, structure and expertise are key determinants of innovation. It is also true that they have influences on IT-related innovation. Typically, researchers regard these characteristics as a uniform property of an organization with a single value. When he or she measures these characteristics, however, they are bound to find different values from unit to unit within an organization. Of course, different organizational units adopt different innovations. Thus, the measured values for these organizational characteristics also vary according to the innovations that a study deals with. Conventionally, the studies utilizing aggregated measures express each organizational characteristic in terms of a single overall score. The same holds for IT innovation studies. In the case where the characteristic is secondary, however, Downs and Mohr (1976) argue that this approach can average away potentially explainable variance in the observed relationship between that characteristic and measured innovation. Yet, recent evidence casts doubt on these conclusions. According to Damanpour (1991), studies using more highly aggregated measures have stronger statistical confirmation of expected theoretical relationships than that of studies using less aggregated measures. As Fichman (2001) put it, there are something that bridges the gap between Damanpour (1991) and Downs and Mohr (1976). First, in Damanpour (1991)'s meta-analysis, organizations tend to be more uniform across units in terms of secondary characteristics than Downs and Mohr (1976). There could be yet another possibility. Some of the studies included in meta-analysis only aggregated innovations adopted by a single organizational unit. And, secondary characteristics should have a constant score for all innovations in that study. Viewed in this light, most of secondary characteristics can be treated as primary characteristics. When IT innovation studies aggregate innovations, they deem it the

innovation adopted just within the IT department (Nilakanta and Scamell, 1990; Zmud, 1982; Zmud, 1984). IT innovation research has some advantage, including the rapid pace of innovation in the tools and techniques used to develop and administer IT systems, and the resulting wide variety of innovations adopted for use within the IT unit itself. To misestimate or omit the secondary characteristics holds some adverse effects, so one need to avoid them. In doing so, aggregation is of service. In consequence, the bottom line is this. The studies using aggregated measures care about the potential effects of secondary characteristics of organizations. These concerns, however, can be avoided by two ways: limiting aggregation to innovations adopted by the same organizational unit, and focusing on contexts in which secondary organizational characteristics are not likely to change significantly across innovations. In these circumstances, one can iron out the tradeoffs pertaining to robustness, generalizability, and theoretical interpretation by preferring aggregation.

4) As Downs and Mohr (1976) noted, the secondary characteristics of innovations is the innovation characteristics that vary depending on the innovation being considered. In regard to this point, Fichman (2001) cites an instance of compatibility. The reason is that the same innovation can vary highly among organizations in considering the compatibility (Meyer and Goes, 1988; Ramiller, 1994). Also, complexity, relative advantage, cost, and many other characteristics of an innovation vary greatly from organization to organization. The data volume in a single study is determined by the differences between research interests and practical limitations. Accordingly, a research model tends to omit innovation characteristics, and this situation results in a source of noise. But Fichman (2001) suggest that aggregation should moderate it. His line of thought is this. For instance, one researcher develops a model where an organization's innovative capacities predict innovation. And, there are two organizations, A and B, with different innovative capacities. Organization A has a high innovative capacity. And, this organization adopts technology Y and Z, but in the case of technology X, it defers the adoption. The reason is that the technology X is not compatible with existing needs, skills, work practices or technical infrastructure. On the other hand, organization B has a low capacity to innovate. But this organization chooses to be on the leading edge for technology X because there is a chance that it is highly compatible. When it comes to technologies Y and Z, organization B doesn't adopt them. A single-innovation design based on the technology X doesn't control for innovation characteristics. Thus, in terms of the score for innovation, it would be low for organization A compared to organization B. This does not reflect the model's prediction or the fact that organization A does tend to innovate more often than organization B. To be consistent with the conclusions of Fichman (2001), in a single innovation design, the omission of secondary innovation characteristics may introduce noise that makes it more difficult to discern the effects of included predictor variables. On the other hand, in an aggregated design, omitted secondary innovation characteristics pose less of a

problem because their effects tend to be smoothed out across innovations. Viewed in this light, a researcher who wants to aggregate across three technologies would find that innovative organization *A* has highly score compared with less-innovative organization *B*. And therefore, when secondary innovation characteristics exist and are not otherwise controlled for, the tradeoffs related to robustness and predictive validity tilt toward more aggregated measures.

5) Aggregating innovation substitutes or moderate complements has advantage over aggregating unrelated innovations, aggregating strong complements, or using single-innovation measures. When an innovation has one or more substitutes diffusing at the same time, aggregation across these substitutes may dampen a subtle source of noise in the measurement of innovation. Again, Fichman (2001) explains this issue with concrete descriptions. For example, there is a pair of emerging innovations. And, these two innovations are not perfect substitutes. In this instance, an organization may adopt one or the other, but cannot adopt both of them. In this kind of environment, the one for which a researcher decides to capture the adoption is an inappropriate alternative. The reason comes from the fact that every organization that chooses the other alternative is assigned a score of low innovation. All organizations are assigned appropriate score if a researcher aggregates across both innovations. Far from it, from a perspective of a predictive validity, the aggregation of complementary innovations only provides small benefits. The aggregated measure of perfect complements is perfectly correlated with each of the individual innovation measures included in the aggregate. Consequently, if a researcher aggregates perfect complements, he or she cannot find effect on the measurement of innovation. At the same time, in the case of imperfect complements, some technologies that hold analogous properties can be complementary in varying degrees, given organizational contingencies and variations in primary and secondary organizational characteristics. There are three concerns on this point. First, in order to obtain value from implementing several imperfect or moderate complements, superior IT innovation capability is required compared to implementing only one of these applications. Second, organizations are likely to be viewed as more consistent reactor to complements compared with unrelated innovations. This leads to moderate concerns that aggregation sometimes adds up to mixing apples and oranges. And finally, there might be some sort of organizations. That is to say, these are organizations that tend to adopt certain clusters of complementary innovations. If they are apt to benefit from the adoption of each innovation in the cluster, this situation leads to the fact that there is fixated interest in such organizations. To conclude, aggregating substitutes contributes to the greatest increase in predictive validity. Meanwhile, aggregating moderate complements hardly act on predictive validity. Aggregating moderate complements, however, leads to results that have a clearer theoretical interpretation.

6) Aggregated measures introduce noise into the study of innovation under some circumstances. To the contrary, in any situation where reporting errors and

idiosyncratic adoption have serious consequences, aggregation also functions to reduce noise. According to Fichman (2001), whenever informants are asked questions about adoption of an innovation, some will misreport the fullest level of adoption. In the IT domain, this situation is of particular concern. The innovations in this domain are more complex, abstract, or multifaceted. Thus, different respondents respond to innovations differently. In order to scratch out these errors, it is important to aggregating across innovations. This is conducive to produce a more reliable overall innovation score for each organization. In addition, he suggests that serendipity can play an important role without relation to the organizational adoptions of innovation and the decisions about the continuity of assimilation. If there are some tradeoffs pertaining to robustness of measurement and predictive validity, aggregation can settle them. Therefore, aggregation is useful, unless these sources of noise are expected to be present and cannot be feasibly eradicated by other means.

4. Final remarks

These three “verdict apparatuses” are not a three-pronged approach but mutually related approaches. Nonetheless, they can be distinguished along two aspects, that is, statistical and conceptual aspect. In the case of conceptual aspect, aggregated measure approach reinforces multiple operations approach in the sense that several hurdles in the latter approach that a meta-analyst must overcome are surmount in the former approach. On the other hand, in statistical aspect, meta-analyst can get rid of some subjective judgment calls, that is, the lack of homogeneity test, because numerical information can be available. Thus, what remains to be seen are discussions of 75% rule and the decision branches when the effect size distribution is not homogeneous. On the whole, however, these three interrelated approaches should whittle away judgment calls somewhat, especially by virtue of aggregated measure approach. And therefore, it would be safe to say that meta-analysis would be applicable to IT-related organizational innovation studies.

Acknowledgments

The author wishes to thank Tomoaki Sakano, professor at Waseda University for useful comments on an earlier draft of this paper.

References

Armstrong, C. P., and Sambamurthy, V. (1999). Information Technology Assimilation in Firms: The Influence of Senior Leadership and IT Infrastructures. *Information Systems Research, 10,* 304-327.

Bretschneider, S., and Wittmer, D. (1993). Organizational Adoption of Microcomputer Technology: The Role of Sector. *Information Systems Research, 4,* 88-109.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cooper, H. (1998). *Synthesizing Research: A Guide for Literature Reviews.* ThousandOaks, Sage.

Cooper, H. M., and Lindsay, J. L. (1998). Research synthesis and meta-analysis. In: L. Beckman., and D. J. Rog (Eds), *Handbook of applied social research methods,* (pp. 315-337). Thousand Oaks, CA: Sage.

Cooper, R. B., and Zmud, R. W. (1990). Information Technology Implementation Research: A Technological Diffusion Approach. *Management Science,* 36, 123-139.

Daft, R. L. (1978). A Dual-Core Model of Organizational Innovation. *Academy of Management Journal,* 21, 193-210.

Dalton, D. R., and Dalton, C. M. (2005). Strategic management studies are a special case for meta-analysis. In: D. J., Ketchen. Jr, and D. D. Bergh (Eds.), *Research Methodology in Strategy and Management,* (Vol. 52, pp. 31-63). SanDiego: Elsevier.

Damanpour, F. (1991). Organizational Innovation: A Meta-Analysis of Effects of Determinants and Moderators. *Academy of Management Journal, 34,* 555-590.

Downs, G. W., and Mohr. L. B. (1976). Conceptual Issues in the Study of Innovation. *Administrative Science Quarterly, 21,* 700-714.

Durlak, J. A., and Lipsey, M. W. (1991). A Practitioner's Guide to Meta-Analysis. *American Journal of Community Psychology, 19,* 291-332.

Fichman, R. G. (2001). The role of aggregation in the measurement of IT-related organizational innovation. *MIS Quarterly, 25,* 427-445.

Fichman, R. G., and Kemerer, C. F. (1993). Adoption of Software Engineering Process Innovations: The Case of Object Orientation. *Sloan Management Review, 34,* 7-22.

Glass, G. V. (1976). Primary, secondary, and meta-analysis. *Educational Researcher, 5,* 3-8.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Grover, V., Fiedler, K., and Tong, J. (1997). Empirical Evidence on Swanson's Tri-Core Model of Information Systems Innovation. *Information Systems Research, 8,* 273-287.

Grover, V., and Goslar, M. D. (1993). The Initiation, Adoption and Implementation of Telecommunications Technologies in U.S. Organizations. *Journal of Management Information Systems, 10,* 141-163.

Guzzo, R. A., Jackson, S. E., and Katzell, R. A. (1987). Meta-Analysis Analysis. In: L. L. Cummings, and B. M. Staw (Eds.), *Research in organizational Behavior,* (Vol. 9, pp. 407-442). Greenwich, CT: JAI Press.

Hall, J. A., and Rosenthal, R. (1995). Interpreting and evaluating meta-analysis. *Evaluation and the Health Professions, 18,* 393-407.

Hedges, L. V. (1994). Statistical considerations. In: H. Cooper and L. H. Hedges (Eds.), *The handbook of research synthesis,* (pp. 29-38). New York: Russell Sage Foundation.

Hedges, L. V.,and Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Howard, G. S., and Rai, A. (1993). Promise and Problems: CASE Usage in the US. *Journal of Information Technology, 8,* 65-73.

Hunter, J. E., and Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings.* Thousand Oaks, Sage.

Lind, M. R., and Zmud, R. W. (1991). The Influence of a Convergence in Understanding between Technology Providers and Users on Information Technology Innovativeness. *Organization Science, 2,* 195-217.

Lipsey, M. W., and Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Meyer, A. D., and Goes, J. B. (1988). Organizational Assimilation of Innovations: A Multilevel Contextual Analysis. *Academy of Management Journal, 31,* 897-923.

Mullen, B. (1989). *Advanced BASIC meta-analysis,* Hillsdale, NJ: Erlbaum.

Nilakanta, S., and Scamell, R. W. (1990). The Effect of Information Sources and Communication Channels on the Diffusion of an Innovation in a Data Base Environment. *Management Science, 36,* 24-40.

Prescott, M. B., and Conger, S. A. (1995). Information Technology Innovations: A Classification by IT Locus of Impact and Research Approach. *Database,* 26, 20-41.

Rai, A. (1995). External Information Source and Channel Effectiveness and the Diffusion of CASE Innovations: An Empirical Study. *European Journal of Information Systems, 4,* 93-102.

Ramiller, N. C. (1994). Perceived Compatibility of Information Technology Innovations Among Secondary Adopters: Toward a Reassessment. *Journal of Engineering Technology Management, 11,* 1-23.

Rosenthal, M. C. (1994). The fugitive literature. In: H. Cooper and L. H. Hedges (Eds.), *The handbook of research synthesis,* (pp. 232-244). New York: Russell Sage Foundation.

Shulze, R. (2004). *Meta-analysis: A comparison of approaches.* Cambridge, MA: Hofrefe and Huber Publishing.

Stuhlmacher, A. F. and Gillespie, T. L. (2005). Managing conflict in the literature: Meta-analysis as a research method. *International negotiation, 10,* 67-78.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., and Grove, J. B. (1972). *Unobtrusive Measures: Nonreactive research in the social sciences.* Rand McNally &

Company: Chicago.
Zmud, R. W. (1982). Diffusion of Modern Software Practices: Influence of Centralization and Formalization. *Management Science, 28,* 1421-1431.
Zmud, R. W. (1984). An Examination of “Push-Pull” Theory Applied to Process Innovation in Knowledge Work. *Management Science, 30,* 727-738.

●早稲田大学ＩＴ戦略研究所　ワーキングペーパー一覧●

入手ご希望の方は下記までご連絡下さい.
連絡先：RIIM-sec@list.waseda.jp
www.waseda.ac.jp/projects/riim/

RIIM IT戦略研究所
Research Institute of Information Technology and Management

事務局：早稲田大学大学院商学研究科　気付
169-8050 東京都新宿区西早稲田１－６－１
連絡先：RIIM-sec@list.waseda.jp
http://www.waseda.jp/prj-riim/

WASEDA UNIVERSITY