Performance Measurement System Implementation: 
Meta-Evaluation of Literature on Success Factors

Heather Keathley, Eileen Van Aken
Grado Department of Industrial and Systems Engineering, Virginia Tech
Blacksburg, VA, USA

Abstract

Performance measurement systems have become widely known for both the potential to make a significant impact on organizational outcomes and the challenge of actually attaining those results. While some organizations are able to effectively design, implement and use these systems, many others face challenges that lead to reduced system benefits and, in some cases, system failure. Researchers have identified factors that affect the success of performance measurement systems with particular emphasis on the implementation phase as a possible source of this problem. A systematic literature review was conducted in previous research that identified a range of factors presented in the literature. This paper continues the previous investigation by conducting an exploratory study to evaluate the use of a meta-evaluation methodology to synthesize the final paper set. This method consists of a meta-analysis of the quantitative studies and a qualitative meta-synthesis of the entire paper set. The results show that there is a dearth of quantitative studies that meet the requirements of meta-analysis which increases the importance of including the qualitative synthesis for this research area.

Keywords
Performance Measurement System, Implementation, Success Factors, Meta-Analysis, Meta-Synthesis

1. Introduction

Performance measurement (PM) systems have the potential to provide significant benefits in various organizational settings but a review of the literature reveals that many organizations are not able to completely or effectively adopt these systems [1-4]. Researchers have identified a range of factors that affect the success of PM systems throughout the lifecycle with emphasis on the implementation phase as the potential source of this failure [5, 6]. A review of the literature reveals that, while researchers generally agree that treatment of these factors should improve the outcomes achieved, there is little consensus concerning the factor definitions, the subsets of factors studied, and the strength of their effects.

A systematic literature review (SLR) was conducted in previous research to identify the breadth of factors affecting the implementation of PM systems in the literature [7]. The results showed that this research area is in a relatively early stage of development with little consistency among studies. In fact, the results showed that the number of factors studied in each paper varies significantly with no papers investigating the full set of factors identified by this study. This suggests that a detailed synthesis of the literature would provide important insights as well as support further development and consensus among researchers. Research synthesis has become increasingly prevalent with the SLR as the most commonly used method [8-10]. However, this is generally considered the first phase of a complete synthesis and additional contributions could be provided by further synthesis of the existing research.

This paper extends the previous study by investigating the applicability of two basic forms of research synthesis to this research: meta-analysis (MA) and meta-synthesis (MS) [8, 10]. MA is a method to aggregate the quantitative data provided in multiple studies using statistical hypothesis testing and effect sizes [9-11]. Conversely, MS is a qualitative method of synthesis which is more interpretive in nature [11, 12]. This mixed method approach reflects the results of the SLR which showed that more than half of the papers featured only qualitative data. Due to the scope of this paper, the basic forms of these methods are applied with more advanced methods reserved for future research. However, this work provides valuable insights and an initial step toward a complete systematic synthesis of this literature. The following sections provide a concise review of the relevant literature, description of the methodology, and the results of the exploratory evaluation. The paper concludes with an overall synthesis of the results including a discussion of the applicability of these methods to the PM system implementation success factor literature.
2. Background
It has been proposed that PM systems have a lifecycle of four phases; design, implementation, use and review [13]. While significant focus has been placed on the design and use phases, research concerning the implementation and review of these systems has only recently gained momentum in response to the relatively low success rates being reported [4, 13]. In order to better understand the success and failure of these systems, researchers have attempted to identify the enablers of and barriers to success; more generally referred to as factors. Some of the factors are identified in more general content areas such as organizational change management and organizational improvement including factors such as leadership commitment and organizational culture. In addition to these, factors specific to PM systems have also been identified including concepts such as fear of the consequences of measurement.

While the general theory in the literature is that treatment of these factors will improve the system results, there is very little empirical evidence presented to support this claim [4]. Most of the studies identified by the SLR were focused on identifying factors or testing the relationship of a small subset of factors. Of the studies investigating the results of mitigating these factors, most have a narrow sample set or are focused on a specific type of PM system which is not easily generalizable. In addition, the outcome variable in these studies is not consistent and ranges from concrete concepts, such as actual improvement in outcomes measured, to more abstract concepts, such as perceptions of implementation success. Finally, many of the factors identified are not easily observable leading to more qualitative studies or quantitative studies of subjective data. These issues increase the complexity of synthesizing the literature and further support the need for a comprehensive research synthesis.

2.1 Research Synthesis
A detailed review of the literature is commonly accepted as the first step in any rigorous research and is important to develop relevant research that advances a field. The concept of utilizing the existing literature as a foundation for current work has been further developed to create highly systematic methods of review which aim to utilize the existing research to its fullest extent. These developments come primarily from fields such as biology and medicine where studies are commonly quantitative and based on a more standardized format. Methods such as these are increasingly being adapted for use in other fields including the social sciences but face several challenges due to variability in these studies [11, 14]. One example of this is the SLR which originated in the medical field and is supported by the Cochrane Collaboration; an organization established in the 1970s that seeks to standardize this process in medical research [16, 17]. In response to the adoption of this method to the social sciences, the Campbell Collaboration, established in 2000, aims to standardize systematic reviews concerning ‘social interventions’ and supports a somewhat relaxed version of research synthesis including qualitative methods [18].

Research synthesis is a general term that includes a broad range of quantitative and qualitative methods. In general, research synthesis can be described as occurring in two phases; a SLR to identify and characterize the relevant studies followed by some further methodological synthesis of the content [10, 11]. The first phase focuses on defining the literature area and investigating aspects such as the primary contributions in the literature [10, 14]. In order to complete the synthesis, associated methods such as ‘meta’ methods are used to investigate more deeply into the literature to draw new insights by either aggregating or integrating existing data [10, 11, 16]. While the advantages of these methods are widely accepted, many theoretical and methodological issues with synthesizing research are actively debated and are typically mitigated through careful design of the methodology. For example, challengers suggest that the aggregation of studies results in a loss of information and the combination of studies with methodological or theoretical differences may result in invalid conclusions [10, 18, 19]. There is also a significant debate concerning biases introduced by the synthesist [18, 19]. The implications of these issues to this research are further discussed in later sections.

2.2 Previous Research
The SLR method used in the first phase of this research was adapted from Tranfield et al (2003) and featured six steps; creation of a scoping study, definition of search protocol, application of exclusion criteria, data extraction, analysis, and dissemination [7, 14]. Due to the complexity of this literature area, strict exclusion criteria were applied that required papers to have the identification of success factors or barriers of success as a primary focus of the research. Although this restriction prevented many studies from being included, it also provided for a more targeted evaluation of the literature. The final paper set from this study included 37 papers resulting in the identification of 43 individual success factors. As mentioned previously, this study also showed that the PM system implementation success factor literature is in a relatively early stage of development and would benefit from an extensive synthesis.
2.3 Meta-Evaluation
A review of the final paper set showed that approximately half of studies featured purely qualitative data. Therefore, a parallel mixed methods approach was chosen for this study which includes a meta-analysis of the appropriate quantitative data and a qualitative meta-synthesis of the entire paper set. While many methods exist for each of these analyses, the basic forms were chosen as an initial investigation of their applicability to this literature. In particular, meta-synthesis features many different approaches and is generally a less well-defined approach than meta-analysis.

Similarly to the SLR, MA originated in the medical field in response to the need for a statistically valid method for combining and generalizing research findings [9, 10, 15]. The development and use of this method in this and other scientific fields appears to have been facilitated by the convention of creating rigorous experimental designs and detailed reporting including statistical data. However, the interdisciplinary and abstract nature of PM system effects leads to a wider variety of research designs as well as significant variations in both the operationalization of variables and presentation of data. This complicates the MA procedure and could ultimately reduce the statistical validity of the results. Quantitative meta-analysis depends directly on the computation of the statistical effect size to determine the strength of a relationship. Many effect sizes exist and are classified in accordance with the experimental design characteristics and provided data [9, 10, 15]. The most well-established classifications of effect size are Pearson’s correlation, r; mean differences, d; and odds ratio, OR. These methods are considered foundational effect size classifications and are the basis for the majority of MA [10, 15]. Unfortunately, most variations of these effects sizes rely on specific experimental designs. For example, one primary requirement for class d effect sizes is the inclusion of a comparison group or control group which does not appear to be as prevalent in this research area. Another example is the required inclusion of bivariate correlation coefficients. Evaluation of this paper set suggests that most studies provide related data but omit the actual correlation matrix which challenges the application of this method.

One popular method for quantitative research in this literature is regression. This is due to the fact that most studies are using survey data to investigating a set of factors against one or more measures of implementation success. The data from these studies is typically the regression coefficient, t statistic, R² value, and F-test results, and the effect sizes typically used for these analysis come in two forms: comparative studies, i.e. ANOVA, and overall study effect sizes based on full-model assessment parameters which are most commonly R² and F-test data. Since this research is concerned with the effects of individual factors, the previously-mentioned effect sizes are not ideal. Unfortunately, a widely-accepted method for calculating the effect size for individual regression coefficients does not currently exist [10, 15, 20, 21]. A more focused review of this specific research area reveals that regularly-cited texts as late as 2009 do not discuss this issue [15] and the papers that actively research this paper are more recent ranging from 2006 to 2012 [20-23]. Two viable methods were identified in the literature [20, 21] and the proposed effect size ‘Semi-Partial Correlation Coefficient’ (SPCC) was chosen as it provides the best fit for the papers included in this study. This method utilizes the t-statistic, R² value, the number of parameters in the model, and sample size to calculate r_{sp}, in addition to the calculations and proofs for the variance and confidence intervals, which are proposed to belong to the r effect size class. Equation (1) presents the effect size calculation [20].

\[ r_{sp} = \frac{t_{p} \sqrt{1-R_{p}^{2}}}{\sqrt{n-p-1}} \]

Finally, applications of MA to synthesizing success factors in other research areas were investigated. The review found that MA is regularly applied to evaluate the effects of factors [24, 25]. In addition, papers applying meta-synthesis to this subject area were identified [26]. In these studies, the appropriate data is available and one of the common effect sizes is utilized. Although the studies are able to apply this method, they also report that many papers were eliminated due to the strict inclusion criteria.

Meta-synthesis is focused on interpretation of the literature rather than aggregating the findings [18, 19, 27]. There are many accepted methods for this which range from highly systematic, integrative techniques to more interpretive techniques [18, 19, 27]. Examples of these methods include thematic synthesis and narrative synthesis which are highly interpretive to more rigorous techniques such as grounded theory and meta-ethnography [19, 27]. Unlike MA which has a highly standardized procedure, MS requires the synthesist to make many judgments and some methods require significant expertise to fully implement. Due to the scope of this paper and the researchers’ familiarity with MS, a more general method appropriate for a relatively novice synthesist was chosen - the meta-study [27, 28].

Meta-study was originally proposed by Paterson and Canam in 2001 and features a three-phase approach termed meta-data, meta-method, and meta-theory [28]. The meta-study method asserts that each of these aspects must be
investigated individually before an overall synthesis of the literature can be obtained and uses the text of published reports as data for the qualitative synthesis. Meta-method and meta-theory investigate the appropriateness of methods and underlying paradigms from the set of studies, considers the assumptions required, and relates the findings to the larger context. Meta-data refers to the analysis of findings and allows for the selection of a qualitative synthesis method based on the goals of the synthesis. The focus of this paper is to evaluate the applicability of this method and, therefore, this method is applied at a higher level and reserves more in-depth analyses for future work.

3. Methodology
Due to the scope and exploratory nature of this paper, relatively fundamental approaches to MA and MS were chosen and some assumptions were included to allow for the evaluation of these methods which are discussed in detail in the following sub-sections. The purpose of the mixed method approach is triangulation and, therefore, the two methods are applied in parallel and then the results are integrated in the analysis. The initial phases of both MA and MS include the problem definition and collection of studies through systematic review which are considered to have been completed in the previous research. As a result, the final paper set has already been collected and this paper begins with the application of inclusion criteria or the conditions upon which studies are evaluated for inclusion in each analysis.

3.1 Meta-Analysis
The MA methodology applied in this study follows the basic principles detailed by two previously-mentioned organizations, the Cochrane Collaboration and the Campbell Collaboration, and is adapted from the Handbook of Research Synthesis and Meta-Analysis [15-17]. The basic procedure for this analysis is provided in Table 1. The method is highly systematic to reduce subjective bias [9, 10], and the random-effects model is chosen because the studies in this paper set investigate different subsets of factors as well as various operationalization of implementation success. This diversity among the publications in addition to the desire to generalize the study findings beyond the sample populations suggests that the random-effects model is most appropriate [10, 15, 17, 29].

The analysis variables for this study include the factors as the independent variables and the concept of implementation success as the outcome variable. With the variables defined, the additional inclusion criteria are applied to the SLR results. Only 19 of the 37 papers presented any form of quantitative data, with the most common form being descriptive statistics of subjective survey data. The criterion ‘utilizes an approved methodology’ is an adaptation that was added to remove papers with a methodology that is not supported by established MA techniques. The papers removed from this study included the following methods: specialized modeling, descriptive statistics, factor analysis, simple frequencies, mean and rank, mean and standard deviation, and evaluations of data that are not related to the strength of the factors’ effects. It is important to note that some of these methods could have been included but they would have needed to be calculated using an effect size from the class d which requires comparative analysis. None of the potential papers provided sufficient data for this additional analysis. However, there do exist conversions between the three main effect size classes and, therefore, these papers could be included in the analysis with the correlation measures if the appropriate data were reported [9, 10].
Only one paper met all of the inclusion criteria and presented the Pearson’s correlation, r. This would allow for an easy investigation of the effect sizes but a synthesis cannot be performed without multiple studies. Therefore, the last criterion was relaxed to include the papers that presented regression analyses and the SPCC was used for the effect size calculation. One negative aspect of this choice is that this coefficient has not yet been proven able to be directly converted and combined with the other effect sizes [20]. Therefore, the two effect size types cannot be combined in this study and the fourth paper was omitted resulting in a final analysis set of three papers. In addition to calculating the mean effect sizes for each factor, subsequent analyses of the effect size are conducted. This includes calculations of mean effect size variance, F-test statistics, and confidence intervals. It should be noted that the eighth step in this procedure suggests the inclusion of more advanced analyses. The extended analysis typically includes items such as analysis of biases and assumptions, e.g. publication bias, and the exploration of moderator variables [29]. Due to the limitations and scope of this study, additional analyses are reserved for future research.

3.2 Meta-Study
The methodology chosen for the MS is a meta-study which is adapted from Paterson and Canam and is shown in Figure 1. This figure shows the three dimensions of MS and their objectives. The steps for this analysis are formulating the research question, primary research selection and appraisal, meta-data analysis, meta-method analysis, meta-theory analysis, meta-synthesis, and dissemination [28].

![Figure 1: Meta-Study Framework (adapted from Paterson and Canam, 2001)](image)

This method also begins with a problem definition and review of the literature to identify relevant studies. Due to the methodological design, one of the 37 papers resulting from the SLR was removed because it was a non-academic paper that provided some data based on industrial experience but did not report any structured methodology. To begin, the meta-theory is approached by identifying the disciplines of authors, the primary paradigms present, and the structured theories identified in the papers. Then, the trends in methods use and their influence on the contributions of the papers are evaluated including an analysis of the strength of methods reported. The meta-data analysis is then conducted to identify the major themes in the findings and in the factors reported as most significant. Finally, an overall synthesis of the three dimensions is conducted.

As mentioned previously, the meta-data dimension requires the synthesist to choose a qualitative synthesis method for the analysis of the data and findings. Thematic synthesis was chosen because it is flexible, provides a more general overview of the data, is well-established, and is appropriate for synthesists with limited experience [18, 27, 28]. Due to the scope of this paper, the method used is theoretical and semantic; the goal is to investigate some existing theory in the literature instead of inductively deriving theory, and the findings are evaluated at the explicit level instead of identifying latent constructs. The approach for thematic synthesis was adapted from Braun and Clarke, 2006, and consists of the following steps; familiarize yourself with the data, generate the initial codes, search for themes in the literature, review the themes, define and name the themes, produce the report [30]. The findings and factor definitions that were identified in the previous SLR are used as the initial codes for this process.

4. Results
The results are presented in two sections; first the MA results are presented followed by the MS results. A complete synthesis of the integrated results of this study and a discussion of the applicability of these methods are included in the following sections.
4.1 Quantitative Meta-Analysis

Since only one paper fit the inclusion criteria for this study, the emergent effect size for regression coefficients was chosen for inclusion in this study. The two effect size types cannot be aggregated and, therefore, the single paper feature r is omitted from the study and reserved for future research. The three papers that were identified for this study have several beneficial characteristics for this analysis [9, 10, 15-17]. The first is that the data provided is complete and sufficient for the effect size calculation. In addition, the authors of the studies at least discussed validity and, in some cases, directly addressed it in the research design. Moderately large sample sizes are used which also strengthens the validity of the effect size calculation. Another positive aspect is that they each investigate a significant subset of factors which allows for a more useful aggregate effect size. Finally, each of these papers conducted more than one regression, i.e., evaluated more than one outcome variable, which were considered as separate entities in this evaluation.

Although these papers are useful in this analysis, there are also several challenges presented. First is the concept of implementation success. The outcome variables from these studies include perceived benefits, effectiveness of the system, use of the system, adoption, completed implementation, and organizational results. This variation is somewhat addressed by the use of the random-effects model but should be further investigated in future research. Another negative aspect is that each paper conducts a different version of multiple regression (e.g., step-wise, OLS). This may not be acceptable for a rigorous analysis but is allowed in this case so that the analysis can be conducted and the method can be evaluated for appropriateness. These two items are further discussed in a later section as a limitation to this research. Finally, the R² values presented in these studies are relatively low suggesting that there is a notable amount of unexplained variance in these models.

The effect sizes for each set of the regression coefficients were calculated according to Equation 1. Becker states that, once the effect sizes and subsequent variance and confidence intervals are calculated, the rs values can be treated as a typical r effect size [20]. Therefore, the rs values were then converted to Zr and the rest of the analysis is carried out as instructed for r class effect sizes. The interpretation of Zr is as follows; small effect ~ 0.10, medium effect ~ 0.25 and large effect ~ 0.40. Once the effect sizes were determined, the aggregation of these effects followed a typical random-effects model with inverse variance weights. To begin, the variables were grouped according to similar constructs. The effect sizes were then combined to evaluate each individual factor. The mean effect sizes were evaluated against the previously-described criteria and the strength of effects are summarized. Figure 2 includes the identified variable, as defined in the previous research [7], the mean effect size, and the ranking. It is important to note that the factor sustainability was only measured by one regression analysis and, therefore, the mean effect size is equivalent to the Zr effect size calculated previously.

To begin, the Q test for homogeneity was conducted; six of the 15 factors were proven to reject the null hypothesis of homogeneity suggesting that the variability present is more than should be expected from sampling error. These six factors are bolded in Figure 2. Next, the Z-test for significance was applied. A total of 15 individual factors were identified and tested. Of the 15, only one variable produced a p-value of less than 0.05. The factor ‘metric difficulties’, which was measured in only one study, was determined to have a significant Z-score suggesting that the
null hypothesis, i.e., that the effect is due to random variation, can be rejected. Figure 3 displays the mean effect size and confidence intervals for the 14 factors that were synthesized.

5.2 Qualitative Meta-Study

Meta-theory is evaluated first to gain a better understanding of the underlying paradigms that are framing this literature set. To begin, the academic departments of authors listed on each paper were tabulated which revealed that 30 of the 36 papers had at least one author with a management or business administration background. The next most prominent discipline is strategic manufacturing which appeared in twelve papers. This was followed by finance/economics, human resource management, and engineering management, which were identified in eight, seven, and seven papers, respectively. All other disciplines appeared in two or less papers. In addition, only 48% of the papers with more than one author had more than one discipline represented and the rest of the papers featured author groups from the same discipline. Next, the major paradigms were identified. Some of the schools of thought are obvious such as traditional PM principles, strategic/balanced PM, and change management literature. Other paradigms identified include continuous improvement, public sector reform, organizational improvement, implementation science, and adoption science. Finally, the theories explicitly identified and used in the literature are tabulated along with their relation to the context of PM system implementation which is shown in Table 3.

<table>
<thead>
<tr>
<th>Theory</th>
<th>Relation to Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signaling Theory</td>
<td>Organizational communications and management behaviors are signals to the employees</td>
</tr>
<tr>
<td>Organizational Theory</td>
<td>Formal organization concepts which frame the context of the system</td>
</tr>
<tr>
<td>Grounded Theory</td>
<td>Conclusions are drawn bottom-up from the data</td>
</tr>
<tr>
<td>Institutional Theory</td>
<td>Institutional environments impact organizational structures and culture</td>
</tr>
<tr>
<td>Expectancy Theory</td>
<td>Employee motivation is driven by expected results/benefits from the system</td>
</tr>
<tr>
<td>Agency Theory</td>
<td>Describes the relationship between leadership and system participants such as employees</td>
</tr>
<tr>
<td>Theory of Constraints</td>
<td>States that systems are always under at least one constraint and works to explain the effect of the factors</td>
</tr>
<tr>
<td>Path Dependency Theory</td>
<td>Frames the system adoption process with a specific focus on technology</td>
</tr>
<tr>
<td>Traits Theory</td>
<td>Explains the personality traits of different roles of system participants</td>
</tr>
<tr>
<td>Contingency Leadership Theory</td>
<td>The best method of implementation depends on the context in which the system is placed</td>
</tr>
<tr>
<td>Leadership Theories</td>
<td>Explains the relationship between system leaders and system participants</td>
</tr>
<tr>
<td>Competency Theory</td>
<td>Details the difference between effective and non-effective leadership styles</td>
</tr>
</tbody>
</table>

Next, the meta-method analysis is conducted to identify method use trends and their influence on the development of the field. The methods used in this set are as follows in order of frequency of use; case study, conceptual framework, survey analysis, action research, interviews (semi-structured), secondary data analysis, and Delphi method. The trend analysis shows that case studies and interviews have increased over time while the others appear to have been used relatively regularly with the exception of secondary data analysis and Delphi method which were only used once each. Finally, information regarding the sample set, research design, and validity of the studies was investigated which revealed that many of these studies are conducted with narrow samples which are specific to some demographic or limited in sample size. Also, the outcome variables for these studies is not consistent, with very few of the studies addressing the complexity of defining implementation success. Most of the studies used managers or leaders who put the PM system in place as the sample population and evaluated implementations after they were completed.

The thematic meta-data analysis was then conducted which identified themes in the findings and also tabulated the factors that were identified as most influential in implementation success. All of the authors identified some factors as having a stronger impact than others and, therefore, the factors identified with the strongest effect were tabulated as shown in Figure 4. It can be seen that leadership commitment, for example, was acknowledged as having a significant impact in 21 papers. This figure shows an excerpt of the entire list of factors and represents 57% of the total factors identified. Then, the major themes in the findings were coded and the primary themes are reported in Figure 5.
Several insights can be gained from the data collected in the previous three analyses. First, the meta-theory reveals that the primary influence in this field is management and business administration which can be expected from the review of the literature. However, the influences of finance/economics and human resource management are not as strong as expected and are noted less often than strategic manufacturing. Also, the existence of engineering management and other diverse disciplines coupled with the prevalence of papers with multiple disciplines involved suggests that the research area is more developed than previously expected. Next, the review of methods reveals that the methods used in the studies are relatively limited with strong limitations noted in most research designs. The generalizability of these studies is limited and should be further decomposed in future research to identify the specific differences in findings among the various demographics. Finally, the data analysis shows that seven of the top ten most frequently investigated factors are also indicated as being the most influential. However, there appears to be a great deal of variability in reports of strength of effect among the studies. A total of 37 individual factors were reported as having the strongest effects with nine factors being identified in only one paper each. The primary findings also provided some insights concerning dichotomous views of the themes. For example, a significant portion of the papers consider implementation to be a single event while others consider this as a phase in the PM system lifecycle which will be repeated to implement updates to the system. Another example is the identification of success factors and barriers to success. Papers are included in this set that phrases all factors as one or the other while other papers identify both. However, most of the factor can be phrased as either positive or negative; for example, lack of specialized training can be a barrier to success while putting training programs in place and educating leaders can be an enabler, suggesting that neutral definitions should be considered.

An overall synthesis of this work, while limited by assumptions which are discussed in the following section, provides new insights from this field that were not captured by the initial SLR. One such insight is that not all of the papers view implementation in the same way. There are variations in both the definition and in the concept of cyclical improvement of the system. This suggests that there may be several forms of implementation present in the literature which should be characterized and may involve specific subsets of factors for each version. Another insight is that many of the factors identified are from the change management literature. It is possible that these factors are being identified because the organizations are not handling organizational change correctly in general. This suggests that there may be two types of factors present in the literature; factors that are specific to PM system implementation and factors that are indicators of larger-scale organizational problems that should be addressed before implementation of any new system is considered. One important insight gained from this synthesis is that most of the studies assert that treatment of these factors is an enabler to successful implementation. However, very few studies actually investigated this effect and most studies simply attempted to identify the factors after implementation. These studies did report that treatment of the factors led to better results but this amount of evidence does not support the level of assumption seen in the literature. Finally, the papers in this study commonly focus on a very specific application area which challenges the generalizability of their work. The results show that minor changes in demographics can have a significant impact on the factors identified. An example of this is evaluation of the balanced scorecard (BSC). This framework is used in almost half of the paper set, but the factors identified in each paper varied greatly across minor differences such as between a small and medium sized company and organizations with different management styles. It is commonly accepted that factors will differ across major changes in organizational settings and across different types of PM systems but the results of this study suggest that the success factors may be more sensitive to the organizational setting than previously suggested. This could mean that the subset of factors may not be determined by high-level classifications but could, instead, be unique to each organization based on a variety of items.
6. Discussion & Conclusions

There are several characteristics of PM system implementations and the related literature that impact the types of studies being conducted. One characteristic is that the success of these implementations is not consistently defined. This concept ranges from concrete assessments of changes in organizational outcomes to more abstract concepts such as system effectiveness which complicates aggregation [4, 31]. Another characteristic is that many of the factors are not easily measurable. These factors include items such as improved decision making, increased communication/coordination among functional units, employee buy-in, and resistance to measurement [7]. The influences of discipline and paradigms also shape the types of studies and promote variability in studies. This diversity supports the development of this area but has not reached a level yet where enough studies are provided so that the dataset can be adequately segmented and further investigated. It also results in a strong inconsistency in the reporting conventions used resulting in lack of complete data in some cases. This analysis revealed that the vast majority of quantitative studies in this literature review do not provide adequate data for effect size calculations resulting in a highly limited MA. This was compensated, however, through the application of the MS which was found to be applicable to all but one of the papers and provided insights that were not discovered through the SLR process.

The results of this study are subject to a number of assumptions related to both the analysis design for this paper as well as the basic assumptions of the methods used. Most of the restrictive MA assumptions are relaxed by the use of the random-effects model. This allows for differences among studies including treatment effects. However, due to the lack of acceptable papers, assumptions concerning the legitimacy of the effect size and the inclusion of studies with slightly different research designs were necessary. Another assumption has to do with validity of the individual studies. MA accounts for validity in terms of variations and sample size but does not address the validity of tools such as survey instruments which are used commonly used in this field. This is typically alleviated by an inclusion criterion that requires the study to report on the validity of their methodologies. In practice, however, data on validity can be incomplete or completely omitted from the publication resulting in a loss of information in the analysis. MS also aims to identify and evaluate the validity of the studies which helps to make this issue more explicit. Assumptions specific to the MS include that this paper set can be synthesized as a whole instead of being segmented according to research design similarities to increase the validity and generalizability of the synthesis. Because the final paper set is relatively small, the segments do not contain enough papers for any significant synthesis. This coupled with the fact that the purpose of this paper is to evaluate this method and obtain some high-level insights to guide future research in this area led to the decision to evaluate the papers as one group.

The application of these methods was only moderately successful and required some assumptions to be made that allowed the included papers to be aggregated and synthesized. However, the method application and review of the relevant literature suggest that this method is highly appropriate for this research and can provide valuable insights. The further development of effect sizes for partial regression data as well as for other data types common to social-science research is needed to allow for more accurate calculations and statistically valid insights. Future work in this area will be conducted to address the limitations and applied assumptions as well as revise the SLR so that it is more theoretically rigorous and allows for the inclusion of more studies. This may lead to a more comprehensive understanding of the factor effects, statistically accurate results, a wider range of factors being identified, and more papers in the final set. The MA procedure should also be extended to include many of the more advanced techniques to analyze the MA results. One example of this is the investigation of the control and moderating variables present in these papers. This is also true for the MS which could include a more advanced method for the meta-data analysis and the inclusion of more sophisticated MS techniques in addition to the meta-study.

References