

A Systematic Reviewing:

A Practical Example from Meta-analysis

Uni
R
853
S94
N863
2006

Norsiah Fauzan



A Systematic Reviewing:

A Practical Example from Meta-analysis

A Systematic Reviewing:

A Practical Example from Meta-analysis

Norsiah Fauzan

Universiti Malaysia Sarawak
Kota Samarahan
2006

First Published 2006
Norsiah Fauzan

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the written permission of the publisher.

Published in Malaysia by
UNIVERSITI MALAYSIA SARAWAK
94300 Kota Samarahan, Sarawak, Malaysia.

Printed in Malaysia by
THE SARAWAK PRESS SDN. BHD.
Lot 231, Jalan Nipah, Off Jalan Abell Utara,
93100 Kuching, Sarawak.

Perpustakaan Negara Malaysia

Cataloguing-in-Publication Data

Norsiah Fauzan

A systematic reviewing: a practical example from meta-analysis
/ Norsiah Fauzan.

Bibliography: p. 76-88

ISBN 983-9257-59-5

1. Systematic reviews (Medical research). 2. Meta-analysis I. Title.
610.72 uni

CONTENTS

	Preface	vi
	Tables	vii
	Figures	viii
1	Introduction	1
2	The Protocol for Systematic Review	7
3	The Practical Example of Conducting Meta-analysis: Formulation of Research Questions, the Coding, Computation of Effect Sizes, Sensitivity Analysis	13
4	Example on the Second Stage of Analysis	31
5	Analog to the Analysis of Variance	36
6	Implications of Meta-analysis for Policy and Practice	67
	Bibliography	

Preface

This book is designed to be a practical introduction to the application of systematic reviewing with meta-analysis as an example for the quantitative review. The meta-analytic technique have evolved dramatically since Gene Glass proposed the term in 1976. Since then, more flowering work on statistical methods was designed particularly for meta-analysis. The chapters will be organized according to the various Coopers' stages of research review process, (1) Problem formulation, (2) data collection and data evaluation, (3) data analysis and data interpretation and (4) presentation of report. Each of the stages were discussed with more examples placed on the stages of data analysis with the research on the area of metacognition and reading comprehension are used throughout for the illustration. The series of stages are similar with the other inquiry process that started with the problem formulation, stages of data collection and evaluation, followed by data analysis and interpretation and culminates in a final report.

The increasing number of research studies in social sciences and education led to an increase in the amount of empirical studies. As a result, it is difficult to draw an overall conclusion with regards to the effectiveness of the studies being conducted. Hence, with 'meta-analysis', the researchers and the social scientist or the policy makers would be able to synthesize the information gained from the empirical evidences to support or assist them in decision making and support their findings. This handbook will address the application of meta-analysis using the research on metacognition and reading comprehension as an illustration in answering the research questions common to social science and education. The University of Malaysia Sarawak is pleased to make this publication available to the community of postgraduate students and researchers.

Dr. Norsiah Fauzan
Universiti Malaysia Sarawak
2006

Tables

	Page
<i>Table 3.1 Summary of results from meta-analysis (dependent effect sizes)</i>	28
<i>Table 4.2 Summary of results from meta-analysis (independent effect sizes)</i>	34
<i>Table 5.1 Overall effect size estimates and outcome measures</i>	38
<i>Table 5.2 Publication Type Category: Journal articles=1, ERIC documents=2, Thesis/Dissertation=3</i>	40
<i>Table 5.3 Year of Publication Category; 1980s, 1990s and 2000</i>	42
<i>Table 5.4 Partitioning 1980s; Journals, ERIC and Dissertations</i>	43
<i>Table 5.5 Grade Level Category (Grade 2-3=1, 4-6=2, 7-9=3, 10-12=4, college=5, others=6)</i>	45
<i>Table 5.6 Ability Group Category: Mix Ability Group=1, Low Ability Group=2, Average Group=3</i>	47
<i>Table 5.7 Types of Reading Category: Types of reading materials used: Narrative=1, Expository=2</i>	49
<i>Table 5.8 Post-test Measures Category: Standardized=1, Experimenter developed=2</i>	51
<i>Table 5.9 'Instructors' Category: Classroom teachers=1, researchers=2, research assistants=3, others=4</i>	54
<i>Table 5.10 Strategy Used: Direct instruction=1, reciprocal teaching=2, generative learning and use of motivation=3, graphic/visual/imagery=4, self-instruction and verbalization=5, questioning=6</i>	56
<i>Table 5.11 Methods of Assignment Category: 'Stringent'=1, Less Strigent = 2</i>	58
<i>Table 5.12 Length of Treatment (in days) category: <15=1, 15-25=2, >25=3, 0.99=No information</i>	60
<i>Table 5.13 Number of Sessions Category: Sessions: 1-5=1, 6-10=2, >10=3, 0.99=No information</i>	63
<i>Table 5.14 Sample Size Category: 1-45=1, 46-100=2, >100=3</i>	64

Figures

<i>Figure 3.1 General Summary of Data in terms of effect sizes and 95% Confidence Intervels</i>	20
<i>Figure 3.2 Stem and Leaf Plot</i>	21
<i>Figure 3.3 Normal Quantile Plot</i>	22
<i>Figure 3.4 Heterogeneity Plot</i>	23
<i>Figure 3.5 The Inclusion and Exclusion of each of the deviated effect sizes from the studies</i>	24
<i>Figure 3.6 The gradual exclusion of deviated effect sizes from the studies</i>	25
<i>Figure 4.1 Stem and Leaf Plot</i>	32
<i>Figure 4.2 Heterogeneity Plot</i>	32
<i>Figure 5.1 Box plot and 95% CI plot of standardized mean difference, d, for the three types of publication (1=Journal. 2=ERIC documents. 3. Dissertations)</i>	41
<i>Figure 5.2 Box plot and 95% CI plot of standardized mean difference, d, for Publication Year (1=1980s. 2=1990s. 3=2000>)</i>	42
<i>Figure 5.3 Box plot and 95% CI plot of standardized mean difference, d, for different Grade Level (Grade 2-3=1, 4-6=2, 7-9=3, 10-12=4, college=5, others=6)</i>	46
<i>Figure 5.4 Box plot and 95% CI plot of standardized mean difference, d, for different Ability Groups Mix Ability Group=1, Low Ability Group=2, Average Group=3, High Ability=4</i>	48

<i>Figure 5.5 Box plot and 95% CI plot of standardized mean difference, d, for different Type of reading materials used</i>	50
<i>Figure 5.6 Box plot and 95% CI plot of standardized mean difference, d, for Type of post-test used</i>	52
<i>Figure 5.7 Box plot and 95% CI plot of standardized mean difference, d, for different instructors during the training</i>	54
<i>Figure 5.8 Box plot and 95% CI plot of standardized mean difference, d, for different strategies used during Intervention</i>	57
<i>Figure 5.9 Box plot and 95% CI plot of standardized mean difference, d, for Method of Assignment to Groups</i>	59
<i>Figure 5.10 Box plot and 95% CI plot of standardized mean difference, d, for Length of treatment (in days)</i>	61
<i>Figure 5.11 Box plot and 95% CI plot of standardized mean difference, d, for number of training sessions</i>	63
<i>Figure 5.12 Box plot and 95% CI plot of standardized mean difference, d, for sample size</i>	65

1

Introduction

The concept of meta-analysis was introduced by Glass (1979) as a means of combining results of different investigations on a related topic. It started following the raging debate in clinical psychology by Hans Eysenck in 1952, arguing that psychotherapy had no beneficial effects on patients (Eysenck, 1952). Gene V. Glass attempted to resolve the debate by statistically standardized and averaged treatment control differences for about 375 studies on psychotherapy by using “meta-analysis” as he called it, and thereby made a conclusion that psychotherapy was effective. This exercise of ‘mega-silliness’ as criticised by Eysenck is now widely accepted by many scholars as a systematic method of summarizing the results of empirical studies not only by the health sciences, but also by the social scientists in social and behavioural science. This method was further crafted by other researchers such as Hunter and Schmidt (1990) in the area of validity in the employment tests, and Rosenthal and Rubin (1978) in the area of interpersonal expectancy effects. The popular term of meta-analysis encompasses all the methods and techniques in quantitative synthesis that the author would like to share the practical experience and exploration of using this technique with the other researchers or higher degree students. These experiences might provide helpful guidance to those who are required to undertake reviews of the literature.

What is Meta-analysis?

First of all, meta-analysis is simply defined as one of the many ways to summarize, integrate and interpret the scholarly works in the variety of disciplines. Quoting Glass and the colleagues (e.g., Glass, 1976) ; meta-analysis refers to

“Analysis of analyses and the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature.”

There are two types of quantitative review procedures. The first type was introduced in the 1930s by Cochrane, 1937; Fisher, 1932; Pearson, 1933; and Tippet, 1931(cited in Lipsey and Wilson, 2001) to address the need to combine the results of a number of independent test in agricultural research. It involves the combination of probability values or Z _scores. The second type of meta-analysis is the accumulation of effect sizes, correlation coefficients or to Cohen’s d statistics.

The Applicability of Meta-analytic Technique

The use of meta-analysis is somewhat applicable to a circumscribe domain. Even though it is systematic, there are situations where the use of these techniques are not applicable. Firstly, the technique applies to quantitative research or empirical studies and cannot be used to summarize qualitative work involving theories and ethnographic studies or naturalistic inquiry. Only studies with quantitative findings using measurement variables would qualify. Secondly, this simply means that the technique would be applicable for studies producing quantitative findings. Third, it will be more appropriate to summarize different research studies with the same constructs and relationships, and configured in similar statistical forms (Lipsey and Wilson, 2001). Otherwise, this would lead to the “Apples and oranges” problem where studies with different research design and constructs were meta-analysed and compared, thus, not conceptually comparable or meaningfully compared.

In addition, practically the findings in a meta-analysis must result from comparable research design as well as be conceptually comparable. Of

course, it is inappropriate to combine study findings in different statistical forms even if they relate to the same constructs or relevant topics. The set of findings from each study, therefore need to be converted in one single index, effect size. Even then different types of study findings require different effect size statistics. For example, studies that produce **pearson** product moment correlations are meta-analysed using a different effect size statistic than studies that used proportion or reporting the pre-post mean difference between a control and treatment group.

Reasons for Meta-analysis

According to Wolf (1986), there are various potential problems with the traditional literature reviews that are addressed in a meta-analysis. These include: (1) selective inclusion of studies, often based on the reviewer's own impressionistic views of the quality of the studies; (2) differential subjective weighting of studies in the interpretation of a set of findings; (3) misleading interpretations of study findings; (4) failure to examine characteristics of the studies as potential explanations for disparate or consistent results across studies; and (5) failure to examine moderating variables in the relationship under examination. With these problems in the traditional review of scientific data, "what is needed are methods that will integrate results from existing studies to reveal patterns of relatively invariant underlying relation and causalities, the establishment of which will constitute general principles and cumulative knowledge" (Hunter et al, 1982).

Given a large amount of data to be gathered, processed, assimilated and synthesized in many different specializations, the fundamental problem as referred to by Glass et al. (1981) is the inability of the human mind to address this task reliably and validly. Very often, the conclusions in the informal and narrative techniques of reviewing are very easily influenced by prejudice and stereotyping and there is no attempt at rigorous definitions and standardization techniques (Glass, 1981). Thus, under the pressure of burgeoning research literature, the old and informal narrative techniques of research review and integration are breaking down. Glass (1981) concludes that the traditional review of scientific data has typically been done in an unscientific, impressionistic fashion and contemporary research reviewing should be more technical and statistical than it is narrative.

On the other hand, meta-analysis is a systematic way of handling information from a large study findings under review. However, it does not mean that meta-analysis cannot be used for a few as two or three

studies. The systematic coding of information in the database allows the meta-analyst to easily record the study characteristics from various research designs, retrieved, integrate and keep track of all the details. The well structured meta-analysis is a research technique in its own right (Lipsey and Wilson, 2001), hence requiring that each step such as specification of criteria, the search strategies, the formal coding of study characteristics and data analysis be documented to support the conclusions drawn from the analysis.

Another advantage of using meta-analysis is the use of effect size to show the effect of the treatment. Glass defined an effect size using the formula:

$$\text{Effect size (Es)} = \frac{\text{experimental mean (x)} - \text{control mean (x)}}{\text{Standard deviation of control group}}$$

The advantage of translating the difference between experimental and control groups into this metric is that the effect sizes can be compared across studies that use vastly different dependent measures. Furthermore, the advantage of an effect size metric is that it is standard deviation units and can be interpreted as a change in the percentile ranking of the “average” size in the experimental group (Marzano, 1998). For example, if the mean score of the subjects in the experimental group is .50. **standard deviation above the mean of subjects in the experimental groups.** The average student in the experimental group is at the 69th percentile of the control. In addition, it is an alternative to statistical significance testing which has been criticised by Carver (1978) as a corrupt form of scientific method. The synthesis of results from various primary studies using meta-analysis provided the researchers with an overview of findings in which statistical significance of individual results has no part (Coe, 1998). Using statistical significance to differentiate studies that find effects than those that do not can be quite misleading. Statistical significance reflects not only the magnitude of the estimated effect and the sampling error around the estimate, which depends almost entirely on the sample size. Studies with small sample size may find effects that are not statistically significant because of low statistical power (Lipsey and Wilson, 2001). In the explication of meta-analysis (1984), Fitz-Gibbon criticised significance testing as depending heavily on sample size and stated that to examine the *magnitude of the effect* is far more important than discussing the statistical significance. Significance testing, therefore left out the most important information: the size of the effect.

Finally, meta-analysis is a new method of quantitative review defined by Glass (1976) as the analysis of analyses. Due to the staggering number of research studies in social and behavioural science there is a need for a scientifically sound review method. Glass pointed out the contemporary research reviewing should be more technical and statistical than it is narrative. Very often, narrative research reviews make no attempt at rigorous definition and standardization of techniques and the conclusions are liable to prejudice and stereotyping to a degree that would be unforgivable in primary research itself (Glass, 1981). In addition, because of the changes in reviewing, Cooper (1982) commented that the researchers can no longer take the conclusions of reviews at face value and, therefore, research reviewers must pay the same attention to rigorous methodology that is required of primary researchers.

Meta-analysis, however is not without criticism. Proponents of quantitative synthesis such as Cooper (1984), Glass et al. (1981) and Jackson (1980) criticised traditional reviews for using unsystematic and poorly specified criteria for including studies and for using statistical significance as the only criterion of treatment effects. However, critics of meta-analysis such as Cook & Leviton, 1980, Eysenck 1978, Slavin, 1984) are equally justified in objecting to a mechanistic approach to literature synthesis that sacrifices most of the information contributed in the original studies and includes studies of questionable methodological quality and questionable relevance to the issue at hand. (Slavin, 1986). Slavin (1986) commented that the mis-uses of meta-analysis in education do not justify a return to traditional review procedures and proposed the “Best-Evidence Synthesis” as an alternative to Meta-analytic and Traditional reviews. The main idea behind this procedure is to add to the traditional scholarly literature review application of rational, systematic methods of selecting studies to be included and use of effect size (rather than statistical significance alone) as a common metric for treatment effects. In best-evidence synthesis, effect size data will be used as an adjunct to a full discussion of the literature being reviewed. (Slavin, 1986). In other words, the method combines the quantification of effect sizes and systematic study selection procedures of quantitative synthesis with the attention to individual studies and methodological and substantive issues typical of the best narrative reviews. (Slavin, 1986).

Organization of the Book

The purpose of this book is to introduce meta-analysis by using an example developed by the author while doing a higher degree on “The Effects of Metacognitive Strategies on Reading Comprehension”. Chapter Two will provide the framework and the procedures employed for conducting the integrative research review. These include the specification of the problem and the formulation of research questions. The second section describes the methods or steps during the meta-analysis such as sources of data, criteria for study inclusion, the variables coded from each study and the computation and analyses of effect sizes, confidence intervals and the search for publication bias. The results of exploratory data analyses and homogeneity analysis using Analog to the Analysis of variance will be exemplified in Chapter Three with summary of results from dependent effect sizes. In Chapter Four, the author will describe the second stage of the meta-analysis. Finally, the final chapter (Chapter Six) will discuss the implication of using meta-analysis in research, and for policy and practice.

2

The Protocol for Systematic Review

There are various alternative models in conducting a review. It can be categorized into two, the narrative and the systematic review, the latter depending entirely on the design of the studies.

Traditionally, one form of literature review is that carried out by an acknowledge expert in the field (Badger et al., 2000). It relies heavily on the professional judgement that reflect the preferences and values of the individual review. In other words, the outcome of the review derives from the status of the review rather than from the process whereby the review was generated. It can be a problem when the harder evidence is needed for the treatment effect related to the area being reviewed. Reflecting from the author's experience, the process is synonymous with the fishing trip in a particular part of the ocean. There's no question of how the trawl of literature was carried out and assembled. Suffice to say it reflects the way many of us actually work on the literature. Working on Chapter Two during a higher degree was difficult for the author. There was no guideline as to the search for the literature relevant to the research. It was an attempt to give an account of the author's intellectual journey in a particular field and a mixture of establishing one's academic lineage and giving due credit

to the sources of ideas and concepts. On the other hand, the reader would try to locate the author's position and how well the theory was understood.

These are very common for those undertaking higher degrees by research whose expectations are often unclear. But, the valuable metacognitive experience of the author was simply different from the first attempt in writing a qualitative review in Chapter Two of the theses, where systematic review was conducted to assess the available evidence from the literature related to the effects of metacognitive strategies in reading comprehension. In this chapter, the author will describe the protocol for systematic reviewing of the literature using meta-analysis as experienced while doing a higher degree. The experience might help the graduates, research students and other researchers in conducting a systematic review of the literature. The next section will describe the guidelines on conducting a systematic research review with reference to the examples and experience during the research.

The Protocol

Purpose

Like any other research process, it started with the purpose of the review. In this case, the author was doing the effectiveness research and, therefore it is essential that a systematic approach is adopted. In the example given, the overall domain of interest was to review the effectiveness of the *metacognitive strategies that are likely to facilitate reading comprehension*. A quantitative synthesis using meta-analysis was adopted to review the primary research studies related to the purpose of the study.

Problem formulation

At this stage, the variables involved in the inquiry were given conceptual operational definitions. In other words, the operations included in the definitions must be specified. This is important because the research reviewer would have to evaluate the concept relevance of different operations as they are encountered in the search for relevant studies. And usually, the reviewer will encounter different operations defining the same concept. As a result, two reviewers employing identical conceptual definition and reviewing the same set of studies may still reach different conclusions (Cooper, 1982). Hence, the reviewers will have to describe and report the

operational variations related to the concept and also the variations in study methods that were related to the study outcome. With reference to the research conducted by the author, the definition and the definitional vagaries in metacognition were first described in the qualitative report before formulating the research questions. Having the prior specification of operation will help the reviewer in the search for the relevant studies to be included in the meta-analysis.

Formulation of Research Questions

Being clear with the purpose and the conceptual definition of the related topic under reviewed that distinguish the relevant from the irrelevant studies. Hence, it provides a very clear structure of what evidence is to be included in the review. The formulation of research questions at this stage was supposed to be easier. However, my experience illustrates that it is not as easy as it sounds. Very often, the question and the conceptual definitions were revisited. These might be due to a large amount of literature involved and differences in the operational detail of the definition. Formulating the research questions and setting the parameters at the outset do provide a certain boundary, though it was not definite. A modification to the approach would be not to feel bound by the original research questions. Minor changes or amendments would be expected to be made to the research questions and initial criteria for the inclusion and exclusion of the studies.

The Data Collection Stage

At this stage, the reviewer would determine the procedures used to find relevant evidence. It is also crucial to determine and specify the target population which includes those elements the inquirer hopes to represent in the studies. The identifying population depends on the choice of which and how many data archives to search and the types of units sampled by primary researchers. This is complicated by the fact the reviewer wants the finding to pertain to all previous research on the problem and the hope that the retrieved studies will allow generalization to the unit of analysis that interests the topic area (Cooper, 1982). On the other hand, the search can be kept manageable if the criteria for inclusion and exclusion are carefully selected. There are a few techniques to review information on a research problem. As suggested by Cooper (1982), these includes: (1) the invisible college approach, where the scientists working on a similar problem are usually

aware of each other and attempt to systematize their contacts by exchanging reprints with one another; (2) the ancestral approach, where information is retrieved by tracking citations from one study to another; (3) the descendancy approach, or the science citation indexes, to retrieve studies that cite papers central to a topic; (4) the use of abstracting services; and (5) the online computer search. From the author's experience, the approach 2-5 were employed in addition to the traditional hand-search methods and advice from the supervisor or colleagues who are well informed on the topic being searched.

Evaluation of Data Stage

It is at this stage that the criteria for inclusion or exclusion of studies are important. The criteria might be totally different depending on the types of literature, but there is a need to have a basis to decide on the relative merits of different pieces of research.

In fact, developing the criteria for the literature is one of the useful by-products of systematic review (Badger, 2000). Hence, critical judgements about the quality of data points is examined in light of the evidence extracted to determine if other irrelevant factors are involved. For example, the data may be been extracted from a totally different methodology as specified in the criteria, or the reports may have also omitted some discussions on the hypotheses of the research. To protect the validity of the review conclusion, the reviewer should state the evaluative criteria as objectively as possible before the literature is searched.

The Analysis Stage

Two categories of variables are involved in the analysis of data from meta-analysis: These include (1) effect sizes, the dependent variables; and (2) the descriptive variables that characterize the effect sizes and the studies that generate them, as these constitute the independent variables. The analysis proceeds by describing the distribution of selected sets of effect sizes and then examining the relationship between effect sizes and the relevant descriptive variables or characteristics of studies using the breakout tables, ANOVA comparisons and the like. However, other issues arise at this stage because each study might produce a multiple number of effect sizes from the same samples, and this might violate the assumption of independent data points which is common to most common forms of

statistical analysis. In the subsequent chapters, the author will describe the procedures of handling dependent effect sizes and creating an independent effect size from each study for analysis. In addition, the data analysis involving effect sizes is weighted analysis, where each effect size is weighted by an appropriate value, the inverse of the sampling error variance.

The Interpretation Stage

The information explosion and the lack of standardization in reviewing the literature has led to skepticism to outcome and how the reviewers arrive at a general conclusion. Selecting one or two studies from hundreds of literature on a related topic will definitely fail to portray the accumulated state of knowledge (Cooper, 1982). Quantitative reviewing techniques were suggested as a remedy to this problem (Cooper, 1982; Glass, McGaw and Smith, 1981), where findings from hundreds of data points were accumulated and synthesized into a unified statement to provide evidence on a practical significance of a related topic being reviewed. In this book, the practical example of quantitative reviewing, i.e. meta-analysis will be provided in relation to a topic on the use of metacognitive learning strategies in reading comprehension. Effect size will be used as a single index representing the effectiveness of a treatment.

Once the mean effect size has been weighted, the meta-analyst can proceed with the interpretation of effect size. In the example given in the following chapter, sensitivity analysis was conducted to avoid erroneous conclusions based on few influential observations followed by homogeneity analysis to see if the various effect sizes that are averaged into a mean value all estimate the same population effect size. This is followed by the Analog to ANOVA to explore the possible cause of variation in the distribution of effect size. Besides the Analog to ANOVA model, there are other models that could be used to examine the relationship between effect sizes and other descriptive variables such as ANOVA comparisons, multiple regression equation and the like. At this stage, we look into how the results should be interpreted and used within the context of high aggregation characteristic of meta-analysis and the study-level units of analysis as illustrated in the example throughout the book.

A frame of reference is needed to interpret effect size results from the meta-analysis. The most widely used convention for appraising the magnitude of effect sizes was established by Cohen (1980). Based on his

general observation, over a wide range of bahavioral science, standardized mean difference effect sizes fell into the following ranges:

Small	Medium	Large
ES \leq .20	ES .50	ES \leq .80

Whereas for the correlation effect size:

Small	Medium	Large
ES \leq .1 0	ES .25	ES \leq .40

In the next chapter, each of the stage will be illustrated with examples from the practical experience of conducting a meta-analysis. The synthesizing of dependent and independent effect sizes will be illustrated in two different stages or chapters.

3

The Practical Example of Conducting Meta-analysis: Formulation of Research Questions, the Coding, Computation of Effect Sizes, Sensitivity Analysis

1.0 Introduction

This chapter will describe briefly the procedures employed for conducting the integrative research review and the results of the meta-analysis. The overall domain of interest in the example illustrated in this chapter is to review the effectiveness of the metacognitive strategies that are likely to facilitate reading comprehension. A quantitative synthesis using meta-analysis was conducted to review the primary research studies related to the purpose of the study with effect size as the measure of the effectiveness of teaching metacognitive reading strategies.

The first section presents the formulation of research questions. The second section describes the methods or steps during meta-analysis such as sources of data or the literature search, criteria for study inclusion, the coding from each study, the computation and analyses of effect sizes and the sensitivity analysis. The Final section will present the results of

- (5) Characteristics of the participants in each study were coded according to grade levels, and ability groups.
- (6) Instructional strategies such as direct instruction, reciprocal teaching and generative learning in the teaching of comprehension, self-instruction, and verbalization and questioning.
- (7) Instructors.
- (8) Duration of the treatment.
- (9) Characteristics of reading materials.
- (10) Types of post-tests materials; standardized or experimenter developed tests.

- (11) Types of instructors.
- (12) Duration of treatment.
- (13) Location of the school.
- (14) Outcome measures.

3.3 Criteria for inclusion and exclusion of studies in the review

Establishing the criteria for selecting relevant studies to be included in the meta-analysis is one of the most important steps in the entire procedure. Since the focus of this meta-analysis was on the effects of metacognitive strategies on reading comprehension, the keywords used for the search of the literature were 'metacognitive' and 'reading comprehension'. Obviously, the criteria had to come from the definition of strategies that are metacognitive in nature. The identified strategies were: questioning, summarizing, paraphrasing, clarifying, predicting, identifying main ideas, text-reorganization, inferencing, hypothesizing, look back strategies, and regulating.

Firstly, the pool of relevant literature was narrowed down to include primary studies investigating the effectiveness of metacognitive strategies to improve reading comprehension. To be specific, studies included in the meta-analysis had to include an intervention particularly on the learners. This resulted in the exclusion of survey studies, case studies and comparative studies that did not present the effect of an intervention. Even the studies that focused on training the teachers on how to teach the metacognitive strategies were excluded. Suffice to say, this meta-analysis would only

include the training studies or intervention on the learners using the metacognitive strategies to improve their reading comprehension.

Secondly, the studies had to be quantitative in nature and include at least one treatment group, with either a control group or a comparison group. Studies without comparison or control group or studies with one group repeated measures were dropped out. The subjects might be randomly assigned to the treatment and control group or might be matched on reading and learning ability, or whether they were from intact classes. Studies that randomly assigned the subjects and matched on reading and learning ability were categorised as 'stringent', whereas the quasi experimental studies where assignment of students to conditions was not randomised or from intact classes were labelled as 'non stringent'.

Third, the outcome of treatment included reading measures either on *standardized tests or experimenter-developed comprehension tests*.

Fourth, the study should provide sufficient quantitative information to permit the calculation of effect sizes. These will be calculated from *the means and standard deviations* of the performances outcome for the experimental and control conditions, or from tests of the significance of the differences in performance between instructional conditions (e.g., *t tests, F statistics and proportions*).

Fifth, this meta-analysis used the correction factor to remove bias in the effect size estimates of studies with small sample sizes. Therefore, studies that failed to include the information on sample or group sizes will have to be excluded from the meta-analysis.

Finally, this meta-analysis will only include studies published in the 80s and early 2000. Popular studies in Reciprocal teaching such as *Palincsar and Brown (1984)* that frequently appeared in previous meta-analysis such as Fan (1993) and Hattie et al. (1996) were not included in the meta-analysis. However, studies that have adapted Palincsar's and Brown's strategies such as Lederer (2000), Alfassi (1998) and Hodge (1991) were selected for the meta-analysis.

Due to the criteria and the nature of meta-analysis, the results of this quantitative synthesis only represent studies that met the criteria for the inclusion. Most studies excluded from the meta-analysis were the qualitative reports or the quantitative reports that did not provide sufficient information for the computation of effect sizes.