Department of Education

UNIVERSITY OF BATH

Working Papers Series

International and Global Issues for Research

Critique of Peer-reviewed Articles on John Hattie's Use of Meta-Analysis in Education

Siebert J. Myburgh

No. 2016/3 December 2016

The working papers series aims to recognise the excellent work produced by the large community of graduate students and distinguished associates of the Department of Education at the University of Bath. The series has been designed to create opportunities to disseminate high quality research through our Departmental webpages, in a timely manner. The working papers series aims, in particular, to reflect and contribute to the global standing of the Department of Education as a leader of research in the areas of activity of its research clusters:

Internationalisation and globalisation of education

Educational leadership, management and governance

Language and educational practices

Learning, pedagogy and diversity

Abstract

John Hattie's review of the educational research of the last 15-20 years encompasses an 800 plus meta-analyses, of which the synthesis was published in his 2008 book, *Visible Learning*. His rank-ordered effect size findings resulted in his work being described and presented among educationalists as *Hattie's List*. In an attempt to curb the media hype and argue against uncritical acceptance of Hattie's views by policy-makers, several critics challenged the ability of meta-analysis to arrive at conclusive findings in education. These discussions among Snook *et al.* (2009, 2010), Hattie (2010) and Terhart (2011) reveal chronic issues with the methodology in meta-analysis and address philosophical, theoretical, and practical implications. The prevailing notion in this paper maintains that knowledge is what our best theories tell us and this position serves as a safeguard against meta-analysis' heavy reliance on empiricism. The paper fully elaborates on the strengths and weaknesses of meta-analysis and a proper role for meta-analysis is cast in light of a critical-realist perspective as opposed to one that subscribes to doctrines of empiricism. It is in this tradition that meta-analysis serves as a means to an end as opposed to becoming an end in itself.

Introduction

The purpose of this paper is to challenge school leaders, teacher practitioners, and educational policy makers to not only know what the research says, but more importantly, to also critique the assumptions that drive the methodologies by which the research is reviewed. One such review method is meta-analysis. Over the last few decades meta-analysis has surfaced as the primary vehicle by which research findings are synthesized. Meta-analysis, to put it simply, is a "statistical procedure for combining results from different studies" (Wachter 1988, p. 1407). More scientifically, it is a "research synthesis that uses formal statistical procedures to retrieve, select and combine results from previous separate studies" (Wachter 1988, p. 1407). Meta-analyses are "a form of synthesizing past literature in a systematic manner and by its nature exclude qualitative research" (Hattie 2010, p. 88). As such, meta-analysis relies heavily on empirical data.

While the term was coined by Glass in 1976, it must be noted that the technique itself was not new (Bangert-Drowns 1986). It was, however, "offered as a rigorous and objective alternative"(Haig 1996, p. 195) to other forms of research review like narrative reviews and vote-taking from box score tallies of significance test outcomes (Hedges 1982). Critics have

always considered the former too subjective while the latter "has been faulted for its bias in favor of large sample studies" (Haig 1996, p. 195). Meta-analysis is considered by some as the "most important advance in the past decade" (Slavin 1984, p. 6), but it has also "generated a great deal of controversy" (p. 6) since its Glassian inception. In keeping with the rich metaphorical language used by critics over the years, the ongoing discussion of meta-analysis in educational circles could well be described as the modern-day equivalent of Tolkien's one-ring-that-rules-them-all. The arguments for and against the method have often split the literature into Gollum-like opposition within itself. This is especially true in education.

Any serious discussion of the use of meta-analysis as a valid methodology to produce conclusive findings in the educational research community must of necessity consider John Hattie's synthesis of 15 years of empirical research, the essence of which was published in 2008 as *Visible Learning*, a 1,700 reference-list book. The scope of this project is "hitherto unique" (Terhart 2011, p. 433) and is considered a "milestone in the research and debate on the conditions for successful learning in schools" (p. 425). Hattie's mega-analysis includes over 800 meta-analyses, involves approximately 240 million students and narrows 52,637 studies down to 146,142 effect sizes which in turn are distilled into 138 ranked topics (Hattie 2010). This "veritable treasure" (Terhart 2011, p. 433) is the result of "decades of careful research" (Snook *et al.* 2009, p. 93) and demands the attention of researchers, theorists and practitioners alike.

This paper is important because it requires the educational community to consider the critique Hattie received from a few of his New Zealand colleagues from Massey University (Ivan Snook, John O'Neill, John Clark, Anne-Marie O'Neill, Roger Openshaw). These scholars paid him "the courtesy of subjecting his conclusions to critical scrutiny in a spirit of mutual truth-seeking" (Snook *et al.* 2009, p. 93). This paper also shares insights from professor Ewald Terhart from the Institute of Educational Science in Munster, Germany, who challenged the notion that Hattie had found the holy grail of research in teaching (Terhart 2011). The aforementioned articles were respectively published in the *New Zealand Journal of Educational Studies* and the *Journal of Curriculum Studies*. The intention of this paper is to join the discussion and argue critical points in order to arrive at and provide a proper perspective on the place, role and contribution of meta-analysis in the complex field of education.

This paper is especially important because it will be scrutinizing meta-analysis' heavy dependence on quantitative data in producing a synthesis of research. It is established in the end that meta-analysis' obsession with empiricism strips its statistical data of meaning because the methodology used, neglects the original theory that drives the primary studies it seeks to review (Haig 2009, 2013). Meta-analysis, therefore, is considered a threat to the ontological views and epistemological values of the critical realist who "have placed stock in the explanatory standing of theories" (Haig 2009, p. 219). The specific research question, then, seeks to determine whether meta-analysis should have a place among social scientists, especially those in the educational research community. While it has established itself in the field of medicine, business and other hard sciences to some degree (Snook *et al.* 2009), this paper will critique the professional contributions of those that have written extensively and thought-provokingly regarding meta-analysis' impact on education.

It must be noted too that this author brings his own preconceived ideas and underlying assumptions to the discussion. These will hopefully be dealt with objectively as they surface, but the reader might find that the struggle to keep the joys and pains of personal experience at bay proved more of a challenge. As a passionate, enthusiastic elementary and middle school leader who is in pursuit of high standards in an international American school setting, the lure of outcomes of "educational significance" (Fitz-Gibbon 1984, p. 136) that accompany meta-analyses serves as a strong inspiration.

Brief Summary of Research Articles

Fundamental to this work is the importance Hattie ascribes to meta-analysis as a vital method for knowing what the research says. This led him to confidently state major educational conclusions, some of which include "the enormous power of quality teaching, that variance among teachers is one of the most critical influences on student learning and achievement, and that there are strategies that seem to invoke more learning than others" (Hattie 2010, p. 86). The media especially (e.g. *Times Educational Supplement, November 2008*) have focused much on the clearly delineated factors for successful learning in schools as outlined in Hattie's original 2008 publication (Hattie 2013). This and the fact that educational policy makers are enamored with Hattie's findings compelled Snook *et al.* and Terhart to alert the educational community regarding their concerns. In critiqueing the high calibre work of a prolific and influential research reviewer like professor John Hattie, the

authors hope to inform, extend and challenge the debate regarding the adequacy of meta-analysis to speak authoritatively and conclusively in educational circles (Hattie 2010, Snook *et al*. 2009, Snook *et* al. 2010, Terhart 2011).

Every effort is made in this paper to create an engaging conversation among Hattie's perspective, the authors' viewpoints and this writer's critique. The dialogue is structured and organized into three main discussion points. The first deals with chronic issues that have always plagued meta-analysis since its inception as a research methodology. The topics include "garbage in garbage out"; publication bias; comparing "apples and oranges" and the use of effect size. Second, questions of a philosophical nature are discussed, specifically those of empirical bias and the limitations of scope. It highlights assumptions, premises and tenets that underlie Hattie's theoretical context of the empirical data base of his mega-analysis (Terhart 2011). Third, the practical aspects of meta-analysis are considered as Hattie's own concessions concerning existing caveats are brought to light to debunk the hyped-up and unsupported claims propagated by the media (Snook *et al.* 2009). While the riveting discussions between Hattie, Snook *et al.* and Terhart will not put a decisive end to the disagreement on meta-analysis in the educational community, this writer will offer a critical-realist perspective on meta-analysis as the way forward.

Article Critique

Chronic Issues

The immediate concern of this critique is to consider the critics' reasons for their bleak prognosis of meta-analysis, but the reader will do well to note from the outset that a death certificate has not been issued for meta-analysis as of yet and a post-mortem as such is therefore not available (Eysenck 1978; Slavin 1984).

Garbage In Garbage Out

The phrase "garbage in garbage out" (Snook *et al.* 2009, p. 96) originated with Hans Eysenck's now famous dispute of one of the first outcome studies of meta-analysis (Eysenck 1978). The basic idea behind the phrase questions meta-analysis' use of exhaustive eligibility criteria to include all available experimental research studies. It also objects to the indiscriminate inclusion of poorly designed studies (Snook *et al.* 2009). When Snook *et al.* (2009) emphasize the Evidence Informed Policy Network's definition of research synthesis as using "best available evidence" (p. 94), they pose an outright challenge to Hattie for

neglecting to include "much about moderators of research findings" (p. 95) in his meta-analysis. They categorically state that Hattie's meta-analysis "does not exclude poor or inadequate studies" (Snook *et al*. 2009, p. 94) and maintain that it is a "meta-analysis (however well designed) of poorly designed studies" (p. 96) that "will inevitably lead to unreliable conclusions" (p. 96). Hence the term "garbage in garbage out" (Snook *et al.* 2009, p. 96).

Hattie (2010), on the other hand, claims that meta-analysis has built-in controls to secure quality and states that "one of the very powers of meta-analysis is to deal with this issue" (p. 88). Unfortunately his casual, throw-away comment that the issue of experimental design and the quality of studies "have been dealt with elsewhere by others" (Hattie 2010, p. 88) appears to be elusive. The ensuing back-and-forth between the scholars centers the debate on what seems to be an irreconcilable difference between them. Hattie's (2010) statement that "the major advantage of meta-analysis is that extreme findings are moderated by a larger sample of studies, even though we may learn much from the extreme studies" (p. 90) is fiercely contested and opposed by Snook *et al.* (2010). They prefer instead to put their trust in "extreme" (Hattie 2010, p. 90) studies that are "large scale, well designed and carried out over seveal years" (Snook *et al.* 2010. p. 96) as opposed to conclusions from "multiple studies, many of which maybe of poor quality" (Snook *et al.* 2010, p. 96). Hattie's (2010) assertion that "it is not correct to see meta-analysis as a blancmange where all is combined and details are not explored" (p. 91) does not satisfy Snook *et al.* (2010). They specifically note a particular case where lower quality studies are dropped from a meta-analysis and the resulting achievement difference between charter and regular schools are significantly lowered as a consequence (Snook *et al.* 2010). They also deduce from Hattie's remarks on the poor research literature on learning styles that the validity of findings in a meta-analysis "cannot be maintained when there are questions about the underlying quality of some of the studies he includes" (Snook *et al.* 2010, p. 94).

Critique:

A more in-depth understanding of meta-analysis design is required to put the "garbage in garbage out" (Snook *et al.* 2009, p. 96) claim in perspective and reduce anxiety levels. The literature is indeed "well rehearsed" (Snook *et al.* 2010, p. 96). Much precaution is taken to avoid an irresponsible, mindless and haphazard inclusion of studies. For example, meta-analysis is dependent on primary, experimental studies that are well-designed, consisting of

concepts that carry adequate operationalization that yield an effect and able to find and list context variables that influence outcomes (Fitz-Gibbon 1984). In this respect meta-analysis is accountable to the hallmarks of professional standards. If dependence on such primary studies poses a challenge to the educational research community's values, then meta-analysis is indeed a "symptom" (Wachter 1988, p. 1408). It will be shown, rather, that it is a "response" (Wachter 1988, p. 1408).

Another example is the high priority of meta-analysis to first define the scope of the task and then outline inclusion criteria accordingly (Borenstein *et al.* 2011). When meta-analysis focuses on trials, the search for primary studies must include those that use random samples. Other criteria that enter the decision-making process for including primary studies in a meta-analysis are the ability to detect an interpretable effect (or operationalization of concepts), relevance, similarity and bias. The safeguard to this process is transparency, which has often led to a reduction of the large pool of initial studies (Borenstein *et al.* 2011).

Perhaps it is necessary to be clear on what poor quality is not. All studies have limitations and weaknesses, but these do not render them unfit (Bryman 2012). Neither do the varying results of different study outcomes within a meta-analysis. As long as researchers are able to ask why and determine which context variables (e.g. status of the researcher, location of the participants, etc.) have an effect, meta-analysis can facilitate the generating of hypotheses and produce new theory (Fitz-Gibbon 1984).

Meta-analysis, however, is not immune to an inclusion of studies of poor quality for the same reason poor quality studies exist. Standards are upheld by people and where they fail, meta-analysis will fail too. "The response of the research community is to locate these flaws, consider their impact for the study in question, and (hopefully) take steps to avoid similar mistakes in the future" (Borenstein *et al.* 2011, unpaginated). The other side of the coin is that bad research is more frequently exposed through meta-analysis than any other method of review. The research community ought to be confident that stellar, convincing arguments will win the day against any poorly designed, flawed and below-par studies that might be included in a meta-analysis (Wachter 1988). Snook *et al.* (2009) powerfully demonstrate this in their critique of homework and class size. It is therefore a mistake to see "proponents of meta-analysis as enemies" (Wachter 1988, p. 1408) and not allies of the broader scientific community.

The fact that Hattie did not detail these discussions in his book does not mean his meta-analysis is insufficiently or poorly done. While he responsibly pointed out the absence

of quality studies to support the literature on learning styles, he perhaps failed to live up to the high standards of transparency in not presenting a clearer picture of studies related to homework and class size. His critics deserve credit in referencing convincing arguments from primary studies to reveal what was missing in Hattie's meta-analysis.

Publication bias

Snook *et al.* (2009) point out that Hattie's work is plagued by a common malady that affects most other meta-analyses, that of publication bias. In stating that "there is a heavy reliance on published results" (Snook *et al.* 2009, p. 97), the authors imply that Hattie's meta-analysis is potentially vulnerable to the following claims:

1. It fails to include well-designed studies that have been cut from journal publications because they don't support "favored conclusions" (Snook *et al.* 2009, p. 97).
2. It gives too much credit to weak published studies, e.g. the literature on learning styles.
3. It is not current because his meta-analysis mostly includes research "from the 1980s and 1990s" (Terhart 2011, p. 428).

The fact that Hattie appears to negate the possibility of bias by referring readers to the "rich literature" (Snook *et al.* 2010, p. 95) on the topic, lacks conviction in the opinion of Snook *et al.* (2010). Whereas the criticism with "garbage in and garbage out" (Snook *et al.* 2009, p. 96) was an issue of being too inclusive, the criticism with publication bias is one of exclusion. The culprit, however, is not meta-analysis.

Critique:

The traditional approach to scientific research, one that most journal editors have subscribed to, is to rely on statistical significance testing to determine whether the null hypothesis can be safely rejected. In other words, unless a study sample boasts a probability level of at least 95 % that it is not likely to be different from the population, it would in all likelihood not be included in a journal publication (Carver 1993). The problem with statistical significance is that it is directly related to sample size and many studies that produce true effects are unfortunately overlooked because of their small sample size (Fitz-Gibbon 1984). Snook *et al.* (2009) are therefore rightly concerned that Hattie's reliance on published studies might have caused an omission (Fitz-Gibbon 1984). There may be so many null effect studies

in the researchers' file drawers that it poses a threat to the published studies as not being the real representation of findings. "The extreme view of this problem, the 'file-drawer problem,' is that the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show non-signifcant (e.g., p > 0.5) results" (Rosenthal 1979, p. 638). Snook *et al.* (2009) believe the file-drawer problem causes a bias in Hattie's meta-analysis.

"In the past there was very little one could do to assess the net effect of studies, tucked away in file drawers" (Rosenthal 1979, p. 638), but now "one can establish reasonable boundaries on the problem and estimate the degree of damage to any research conclusion that could be done by the file drawer problem" (p. 638). These parameters are calculated by fail-safe numbers, which in their most complex form use sophisticated statistical weighting of null studies, but a more friendly formula is N = 5n + 10 (Rosenthal 1979). Translated this means that the mimimum number of studies that need to be in the file drawers before they pose a "rival hypothesis" (Rosenthall 1979, p. 640) need to number at least 15. Rosenberg (2005) considers this number "fairly arbitrary" (p. 467) because a meta-analysis on 500 journal publications are hardly "unrobust" (p. 467) even though the missing studies may amount to more than 2,510. It remains that research topics with a small number of published studies are vulnerable and susceptible to the file-drawer threat (Rosenthall 1979).

John Hattie's reference to the "rich literature" (Snook *et al.* 2010, p. 95) on publication bias is a satisfactory claim that his meta-analysis is probably safe from the file-drawer threat. Probably, because it must be remembered that "a fail-safe calculation is neither a method of identifying publication bias nor a method of accounting for publication bias that does exist. It is simply a procedure by which one can estimate whether publication biases – if they exist – can be safely ignored" (Rosenberg 2005, p. 467). It can therefore be stated with confidence that Hattie's meta-analysis, although vulnerable to the charge of publication bias, has every chance to be considered robust and not biased.

Snook *et al.* (2009) are correct to be suspicious of Hattie's synthesis of meta-analysis, but it must be noted that no other method of research review fares any better. Although the problem is not caused by meta-analysis, the method is singled out because of its claim of stricter adherence to objectivity (Borenstein *et al.* 2011). The real culprit responsible for creating this large bias is that "corrupt form of the scientific method" (Carver 1978, p. 378) that relies on statistical significance testing. It is meta-analysis, however, that offers a radical shift away from statistical significance testing with an improved new metric that measures

educational significance (Fitz-Gibbon 1984). This will be discussed in more detail in meta-analysis' use of effect size.

Comparing Apples and Oranges

No methodology of science is considered respectable in the research community when experiment design is unable to control a rampant blooming of subjectivity. This always affects replicability and generalizability, two corner stones ensuring validity in the research process (Fitz-Gibbon 1984). Snook *et al.* (2010) insinuate that this loss of objectivity describes the state of educational research. Their skepticism of educational research in general and meta-analysis more specifically is based on the viewpoint of Graham Nuthall who "after forty years of close-grained research in classrooms" (Snook *et al.* 2010, p. 96) cautioned that one has to be "very selective in identifying those that have something trustworthy to say about classroom teaching and learning" (p. 96). Their concern also is a lament of the sometimes inept, clumsy attempts of meta-analysis in superimposing its sophisticated hermeneutics of statistical interpretation on educational variables that are "often poorly conceptualised and the studies often far from rigorous" (Snook *et al.* 2010, p. 96).

While it is true that meta-analysts rely on coding technicians to define concepts, variables and categories (Wachter 1988), it has to be conceded that in education these are often not as clearly defined as, for example, in the field of medicine. Snook *et al.* (2009) compare the vagueness of education's abstract qualities (small vs. large class sizes, traditional vs. open education, child-centered vs. teacher-centered education, etc.) with the more careful manipulation of medical drugs. The lack of a "clear operational definition of … variables" (Snook *et al.* 2009, p. 97) in education results in a "continuum and, therefore, subjective judgments have to be made" (p. 97). Snook *et al.* (2009) show how the effect of whole language on learning is a case-in-point of vagueness and suspect that "this category of problem might be widespread" (p. 97). The dynamic interaction between variables in the behavioral sciences furthermore confounds this issue and adds to the complexity (Snook *et al.* 2010). In this regard the authors make an interesting connection between John Hattie's "*Visible Learning*, based on international meta-analyses" (Snook *et al.* 2010, p. 96) and Graham Nuthall's "*The Hidden Lives of Learners*, based on decades of research in New Zealand classrooms" (p. 96). This, in addition to the fact that many researchers define and design their studies differently, constitute an inordinate, if not impossible comparison in the view of Snook *et al.* (2010). They consider the above as "perhaps the major problem in meta-

analysis in educational research" (p. 96) and express their surprise that Hattie doesn't comment on this.

Critique:

Hattie's silence on this "major problem in meta-analysis in educational research" (Snook *et al.* 2010, p. 96) is not a case of shying away from a reluctant concession, but rather speaks of being stunned regarding the lack of historical knowledge in this area. Meta-analysis, as a method of research review, must by nature combine "results from experiments on the same phenomenon conducted under different conditions" (Wachter 1988, p. 1407). Haig in his 1995 article on statistical methods categorically states that "meta-analysis is concerned with the statistical analyses of the results of data analyses from many individual studies in a given domain for the purpose of integrating or synthesizing those research findings" (Haig 1995, p. 195). He goes on to say that it is in the consultation of a variety of studies "to ascertain the existence of robust empirical regularities, a task which can be usefully viewed as the statistical analogue of experimental replication" (Haig 1995, p. 96) that meta-analysis does its "most important work" (p. 96). The studies, therefore, must be different (Borenstein 2011). The notion that meta-analysis, at times, involves different studies should not be presented as a problem that has been overlooked. In fact, Glass contended decades ago that "combining across diverse studies is no worse than combining across subjects" (Slavin 1984, p. 7). History bears out Hattie's silence on the matter.

This issue of comparing "apples and oranges" (Snook *et al.* 2009, p. 96) is essentially an issue related to subjectivity, replicability and generalizability and is indeed an important one that affects the entire educational research community, not just meta-analysis. The solution rests in the validity of primary educational studies and only then in meta-analysis' statistical sophistication. It is noteworthy that Fitz-Gibbon (1984), an exponent of meta-analysis, does not share Nuthall's skepticism regarding the research on classroom teaching and learning. She hails meta-analysis for giving new significance to the "well-controlled small-scale experiment" (Fitz-Gibbon 1984, p. 135). Meta-analysis compares the differences in effect across a variety of primary educational studies and each time asks whether the effect still holds up and whether it is educationally significant (Fitz-Gibbon 1984). The resulting dispersion usually has central tendencies that "induce convergence of views" (Wachter 1988, p. 1408), but most often the real value lies in the consideration of the variety of "context variables" (Fitz-Gibbon 1984, p. 138) across "different subjects, treatments and occasions,

and with different methods" (Haig 1995, p. 217). The identification of context variables is conditional in "generalizing results" (Fitz-Gibbon 1984, p. 139). The replicability of meta-analysis, therefore, is not just statistically induced. It is rooted in the "internal replication" (Haig 1995, p. 217) of the stable results obtained through primary studies and perpetuated through "external replication" (p. 217) as described above. This strength of meta-analysis to formally assess consistency "from one type of study to the next" (Borenstein 2011, unpaginated) is called generalizability. If the skeptic is still stuck on the inordinate comparison of vague concepts, then perhaps meta-analysis' purpose is not fully understood.

The Use of Effect Size

Snook *et al.* (2009) identify the use of effect sizes as a "major aim" (p. 94) of Hattie's project. While "meta-analysis comes in a variety of forms and boasts an increasingly ramified methodology" (Haig 1995, p. 195), effect size in principle is "the mean of the experimental group minus that of the control group divided by the control group's standard deviation" (Slavin 1995, p. 10). An effect size of 1, for example, means that if the control group's mean is at the 50th percentile on a normal distribution of reading test scores, then the experimental group's reading scores would rise one standard deviation and distribute around the 84th percentile (Coe 2002). Students not exposed to the treatment, will most likely not see a similar increase in learning.

Snook *et al* (2009) go to great length by devoting several pages of their critique to illustrate that the use of effect size is "a hazardous exercise" (p. 94). For them, "the grouping of various factors into one 'effect size' often hides the complexity of the issues" (Snook *et al.* 2010, p. 97). They bolster their arguments by quoting from convincing case studies (homework and class size) and use commanding evidence from Hattie's own meta-analysis. For example, the 0.29 average effect size on homework in Hattie's meta-analysis fails to do justice to all the underlying nuances and the 0.2 effect size for class size is not at all neglible when longitudinal studies are considered (Snook *et al.* 2009). Also, the research on charter and regular schools show how effect size increased when studies of poor quality were omitted.

Critics further question the categorization of effect size according to magnitude. It seems arbitrary to refer to small, moderate and large effects. Compare, for example, a small effect size taken from a large randomized sample with a large effect size taken from a small fixed sample. The former is "of considerable significance" (Snook *et al.* 2009, p. 94) while

the latter is "meaningless at best" (p. 94), but "when lumped together to produce an average" (p. 94) it loses communicative value.

Despite Terhart's (2011) warnings that effect size is not causal, the natural inclination of the general public is to continue to use it in a way that encourages "silver bullet responses to complex educational problems" (Snook *et al.* 2010, p. 97). Hattie's use of effect size in his book is perceived to present itself as a panacea to the educational community, but the authors maintain that practitioners and policy makers will experience a short-lived placebo effect. The media hype feeds into the ambition of power hungry policy makers who rush teachers and practitioners into action devoid of understanding. Hattie (2010) claims this is an unintended consequence.

Critique:

Criticism of meta-analysis' use of effect size often tries to kill two (or three) birds with one stone. The critique of Snook *et al.* (2009, 2010), Hattie (2010) and Terhart (2011) in this section will not revisit previous issues, but limit itself to the use of effect size per se.

Meta-analysts use effect size to transform study outcomes to "a common metric so that they can be compared" (Bangert-Drowns 1991, p. 2) and to arrive objectively at a synthesis of the research of all primary studies. This major improvement stands in sharp contrast to the most popular and most traditional form of research review, that of vote-taking from box score tallies of significant test outcomes. Methods of statistical significance deal with the probability or likelihood that the random selection of samples was due to a sampling error. If it is found that a sampling error has indeed occurred, then the experimental group will not be considered a true representation of the population and the recorded effects will therefore not be counted as statistically significant. The study and its effect will be discarded as a chance event and will not be included in research reviews. This tendency to prefer statistical significance over educational significance in educational research is considered a "corrupt form of the scientific method" (Carver 1979, p. 378). Meta-analysis "has been introduced in part to overcome this problem" (Haig 1996, p. 196). Effect size is not interested in statistical significance, but focuses on the presence (and size) of an effect and therefore constitutes a real improvement for an authentic review of the research.

Indeed, the "major aim" (Snook *et al.* 2009, p. 93) of meta-analysis is to determine effect sizes, but with an expressed purpose, that of seeking to synthesize (not summarize) research findings (Fitz-Gibbon 1984). It boasts remarkable accuracy and therefore improves

on the subjective conclusions present in narrative reviews (Wachter 1988). In this regard Borenstein (2011) quotes the work of Cooper (1997) who objected to the fact that conclusions from meta-analyses are in disagreement one third of the time with study findings from randomized trials. But this concern, however, was put in perspective at the 1996 Cochrane Collaboration in Baltimore when an expert showed that randomized trials also differ with each other about 33% of the time (Borenstein 2011). Fitz-Gibbon (1984) furthermore commented that a meta-analysis consisting of high and low validity studies hardly ever exceeded a difference of 0.1 in twelve research areas and in some research areas the difference between well-controlled and poorly designed studies was of minor significance. Not many are aware that meta-analyses are issued with a weighting scheme that allows for a transparent perusal of how effect sizes were calculated (Borenstein 2011). While it is true that meta-analysis yields reliable empirical averages, its function is not to erase differences of opinion (Wachter 1988). This brings the discussion to a secondary aim of meta-analysis, not stated as of lesser importance, but as a logical-sequential point.

It is important to note that effect size magnitude is calculated for "each of the relevant dependent variables in each study. The effect sizes are then summed and divided by the total number of effects to obtain the average size" (Haig 1995, p. 195). "Regardless of the summary number, meta-analysis should shed light on why trial results differ" (Borenstein et al. 2011, unpaginated). Effect size was never intended to be used as a magic number that would cause all studies on a topic to be turned into a conclusive summary (Borenstein 2011). If it is presented in this manner, it constitutes "dubious" (Fitz-Gibbon 1984, p. 135) practice. In "calculating effect sizes across primary studies in a common domain" (Haig 2013, unpaginated), meta-analysis maps out a data-rich pattern of the relationships between context variables, characteristics and factors. To not accept this function as a vital contribution from meta-analysis "flies in the face of widely accepted scientific practice" (Haig 2013, unpaginated). The resulting dispersion of inconsistencies, especially when substantial, facilitates the interpretation of findings (Fitz-Gibbon 1984). Snook *et al.* (2009; 2010) and Terhart (2011) should be praised for stripping the effect sizes on class size and homework in Hattie's book from its causal appearance and power. When effect sizes fail to be commensurate with the data, they land in the "unacceptable" (Fitz-Gibbon 1984, p. 135) end of the continuum.

It is appropriate to give Hattie (2010) the last word on his use of effect size. He points out that a normal distribution of the 146,142 effect sizes recorded in his 800 plus meta-

analyses shows 0.4 to be an average effect. It also roughly corresponds to the amount of learning gained in the course of a regular school year (Hattie 2013). He therefore concedes that this cut-off point is perhaps "arbitrary, but it was certainly not capricious" (Hattie 2010, p. 89). Hattie's plea that it should not be "presented as a fixed cut-score below which one should ignore and above one should cherish" (Hattie 2010, p. 89) is an ideal that unfortunately misses the mark due to the fact that other powers are at work. In publishing a book on instructional strategies that are permeated with effect size, Hattie unfortunately failed the purpose of meta-analysis to stimulate the discussion. It has resulted in "silver bullet responses" (Snook *et al.* 2009, p. 97).

Philosophical and Theoretical Issues

Hattie's meta-analysis is "a synthesis of a large number of meta-analyses of studies about education variables" (Snook *et al.* 2009, p. 93), specifically education variables that are "amenable to quantitative measurement" (p. 93) of achievement. This is an issue on two levels. First, the empirical nature of meta-analysis methodology constitutes a bias and as a result carries certain pedagogical implications. Second, Hattie himself concedes that his specific focus on achievement gains is a "limitation" (Snook *et al.* 2009, p. 95). While these two points are reflected upon separately, a combined critique is offered in order to preserve a coherent philosophical and theoretical framework.

Empirical Bias

Terhart (2011) points out the bias when he remarks that "qualitative studies are not considered" (p. 426) and early on in his critique observes Hattie's preference to summarize research results by way of a "statistical index for effect size" (p. 427) rather than the "literary way" (p. 427). He also extends an invitation to discuss "the relation between empirical research on teaching and schooling, on the one hand, and theoretical debates about the aims and qualities of the teaching-learning-process in schools, on the other" (Terhart 2011, p. 426). Terhart (2011) argues that Hattie's educational philosophy, his pedagogical views, theories of teaching and learning and notion of an ideal teacher are all firmly rooted in conclusions he derives from the enormous empirical data base obtained through meta-analysis. He uses Hattie's tables of effect size on teacher effectiveness to show the direct relationship with what Hattie considers effective teaching and learning. For example, the role of the "teacher as activator - average effect size of 0.6" (Terhart 2011, p. 433) and the "teacher as facilitator -

average effect size of 0.17" (p. 433) lead Hattie to state that "active guided instruction, which formulates clear requirements to the students and supplies them continuously with feedback about their progress and performance is more effective than unguided, facilitative, 'helping' teaching" (p. 432).

Table 3. Effect sizes for teacher as activator and teacher as facilitator.

Teacher as activator	*d*	Teacher as facilitator	*d*
Reciprocal teaching	0.74	Simulations and games	0.32
Feedback	0.72	Inquiry-based teaching	0.31
Teaching students self-verbalization	0.67	Smaller class sizes	0.21
		Individualized instruction	0.20
Meta-cognition strategies	0.67	Problem-based learning	0.15
Direct instruction	0.59	Different teaching for boys and girls	0.12
Mastery learning	0.57		
Goals—challenging	0.56	Web-based learning	0.09
Frequent effects of testing	0.46	Whole language—reading	0.06
Behavioural organizers	0.41	Inductive teaching	0.06
Average activator	0.60	Average facilitator	0.17

(Terhart 2011, p. 433)

Terhart (2011) also points out that Hattie's "ideal *born* teacher" (p. 435), the "active, responsible sentient, continuously evaluating teacher" (p. 435) is based on "the idea of scientific evidence" (p. 435). Terhart (2011) boldy states that "his interpretation of the broad empirical evidence on successful teaching and learning leads Hattie to a modernized and refined conception of authoritative, teacher-centered teaching" (p. 434). It also created a pedagogy where learners possess a mature metacognitive ability to "monitor their own learning" (Terhart 2011, p. 431). Hattie's ideal teacher works in a school dictated by evidence-driven policies from governments that demand educational institutions be faithful to research reviews produced by meta-analysis. It comes as no surprise that teaching and learning must therefore be "visible" (Terhart 2011, p. 431). The empirical influence of Hattie's meta-analysis extends beyond the sphere of the classroom, but the scope of his work doesn't.

Limitations of Scope

The limitation of Hattie's meta-analysis exists in that the effect sizes fail to report on real achievement gains in "new knowledge, skills and dispositions" (Snook *et al.* 2009, p. 95) because they are "generally more difficult to measure" (p. 95). Hattie also excludes the

outcomes of moral education because "this is not strictly achievement as is typically defined, these are not included in the tables" (Snook *et al.* 2009, p. 95). The bulk of Hattie's findings is "more related to the surface and deep knowing and less to conceptual understanding" (Snook *et al.* 2009, p. 95). This is also Hattie's own criticism (Hattie 2010). "Not only is much research based on surface achievement – so are most teacher-made tests, and teacher classroom questions" (Hattie 2010, p. 90). To this Snook *et al.* 2010 respond that "there is less to be drawn from his synthesis than commentators have suggested. This is a warning to policy makers rather than a criticism of John Hattie's work." (p. 96).

It is to Hattie's credit that he acknowledges that his book does not facilitate an extensive discussion on the most critical factors that "could lead to major differences to students" (Hattie 2010, p. 89). He justifies paying minimal attention to non-school factors because educators don't have control over them. Snook *et al* (2009) quote from several authoritative studies to emphasize that student influences, home conditions and other socioeconomic factors make up for approximately 70-90 % of the variance in student achievement. This means schools control only 10-30% of the variables that have an effect on student achievement. Again, policy makers are reminded that "policy decisions cannot be drawn in isolation from the background variables of class, poverty, health in families and nutrition" (Snook *et al.* 2009, p. 95). The limitations are real (Nye and Hedges 2004).

Critique:

Bryman's (2012) notion that "the practice of social research" (p. 4) is not "hermetically sealed off from the social sciences and the various intellectual allegiances that their practitioners hold" (p. 4) is a refreshing reminder to join the critics in a careful examination of Hattie's assumptions with regards to his view of the educational world and the constructs he is measuring. The fact that "methodological problems and debate are neglected" (Terhart 2011, p. 426) leaves the critical thinker to search for hints in an attempt to engage at a deeper level with the research. A few clues present themselves.

Terhart's (2011) reminder that Hattie's 800 plus meta-analyses exclude qualitative studies means that the 52,637 primary studies generated empirical data via quantitative methodology. The hypothetico-deductive method with its emphasis on statistical significance is almost exclusively associated with Hattie's mega-analysis and aligns with the view that "scientific theories are basically instruments for ordering knowledge claims" (Haig 1995, p. 191). This reveals the assumptions behind meta-analysis as a research methodology. It

therefore maintains that knowledge of the social world is discoverable a-priori, which most likely will result in a description of social reality lacking what Bryman (2012) describes as "a constantly shifting emergent property of individuals' creation" (p. 20). This in itself is not adequate proof to place Hattie in a specific ontological/epistemological camp because "empirical generalizations" (Bryman 2012, p. 11) can still be "insightful" (p. 11) if the researcher employs "super empirical virtues" (Haig 1995, p. 192) like "plausibility, fertility, explanatory depth" (p. 192), etc. However, it is interesting to note that Hattie's synthesis of the research over the last 28 years (1980-2008) produced a solid "stability hypothesis" (Terhart 2012, p. 436). Not much is new. It appears, then, that Hattie's massive empirical data base does not teem with the vital dynamics of social interaction. Meta-analysis is an ontology devoid of constructionism and sustains an epistemology, which in its final analysis, grossly neglects theory and heavily favors the empirical (Haig 2013).

Bryman (2012) reminds us of the importance of a research question in articulating a coherent philosophical and theoretical framework. "If a research question is formulated in such a way as to suggest that organizations and cultures are objective social entities that act on individuals, the researcher is likely to emphasize the formal properties of organizations or the beliefs and values of members of the culture" (Bryman 2012, p. 19). Much can be learned from Hattie's (2010) research question which focuses on the "influences of achievement for school-aged students" (p. 86). Snook *et al* (2009) qualify those influences as "variables affecting the achievement of students" (p. 93) and Terhart (2010) describes them as "the conditions of successful teaching and learning in schools" (p. 425). The culture in Hattie's teaching and learning environment is described by Terhart (2011) as a "deep belief in the attainability of successful learning through good teaching" (p. 429) and as a "belief in the infinite promotion of learning by right, the good teaching" (p. 435). The culture is "a model of visible teaching and learning" (Terhart 2011, p. 430) and one where "the teacher is in a position of dominance – but he has to tame or to hide his dominance in a student-centred manner" (p. 434). Learning as an outcome is treated as an "objective measure with specific categories of built-in essences" (Bryman 2012, p. 18), which implies that Hattie impotently deals with achievement on the most superficial level, devoid of measuring complex interaction. The organization (school) is "a technocratic, administration-driven, centralistic, and economically motivated neoliberal strategy of enhancing the efficiency of schools and teachers" (Terhart 2011, p. 434). One begins to understand why Snook *et al.* (2009; 2010) protested policy makers' indiscriminate adoption of all that is in the book.

Hattie's research question seems to be driven by a personal lament of why "we set our standards so low" (Hattie 2010, p. 86). This is his "burning social problem" (Bryman 2012, p. 4), which produced a value-laden book that contains several evaluative statements. "The synthesis of meta-analyses showed the enormous power of quality teaching, that variance among teachers is one of the most critical influences of student learning and achievement, and that there are strategies that seem to invoke more learning than others" (Hattie 2010, p. 86). Hattie's reflective engagement throughout the dialogue shows that the "untrammeled incursion of values" (Bryman 2012, p. 22) have been kept to the minimum.

Practical Issues

Hattie assures the research community that he has arrived at sound conclusions based on his confidence that his mega-analysis of meta-analyses consists of quality studies, that the effect sizes faithfully represent a review of the original data and that he adequately explores moderators (Hattie 2010). He therefore boldly states that the practical application of his research shows "the enormous power of quality teaching, that variance among teachers is one of the most critical influences on student learning and achievement, and that there are strategies that seem to invoke more learning than others" (Hattie 2010, p. 86). Snook *et al.* (2009), however, draw from the late Roy Nash's "unequaled" (p. 104) contributions to justify their skepticism of Hattie's findings. To them the rare competence to interpret and critique statistical methods and quantitative research can easily lead to propaganda. A mere reference to *the research shows* (italics mine) will not suffice and must be displaced with an understanding of how research works and "how easily it leads to error rather than truth" (Snook *et al.* 2009, p. 105). For Terhart (2011), the usefulness of Hattie's narrow scope will remain locked away until the empirical spotlight of meta-analysis also incorporates a much broader, qualitative searchlight. The question from Snook *et al.* (2010) is whether practitioners will remember Hattie's plumbline, which is a quote from John Dewey, on how to implement his research: "Evidence does not supply us with rules for action but only with hypotheses for intelligent problem-solving, and for making inquiries about our ends in education" (p. 97). The major point, then, of Hattie's meta-analysis is that it "cannot be automatically applied to practice" (Snook *et al.* 2009 p. 104) without the hard work of integrating findings with "personal beliefs, values and experience" (Snook *et al.* 2009, p. 104).

Proper Role of Meta-Analysis

Premise

In order to understand the proper role of meta-analysis in the educational research community, it is crucial to identify a premise. We must ask: Who is the social scientist? (Haig 1996). This paper proposes that meta-analysis achieves its purpose if it is placed in the hands of the critical realist who believes that "science is to be understood as a value-laden, problem-oriented human endeavor which tries to construct causal explanatory theories of both phenomena and particular events" (Haig 1996, p. 196). The critical realist, then, conceptualizes problems, deliberates and formulates goals, carefully crafts research questions, engages in critical decision-making regarding methods, and skillfully adjusts goals and/or methodology to generate rich empirical data bases that will effectively facilitate causal explanation (Haig 1996). The researcher's work therefore entails more than interpretation and analysis of data. He is not only a statistician. He is first and foremost a researcher, then a statistician, or at least a researcher who is willing to call on those with the necessary statistical skills to assist him. When meta-analysis is placed in the hands of such a researcher, it will not be an end in itself, but a means to achieve the purpose.

The critical realist does not subscribe to empirical doctrines or hold to the empirical adequacy of a theory (Haig 1996). She will therefore not perpetuate the primacy of the hypothetico-deductive method with its emphasis on measures of statistical significance testing, and this belief that causation is empirical and that theories are knowledge claims (Haig 1996). "Meta-analysis itself is not an approach to theory testing" (Haig 2013, p. 18). The social scientist, however, encourages data mining through the use of reliable and sophisticated statistical techniques such as meta-analysis to explore and establish empirical regularities through which she will be able to extract a signal from the noise (Wachter 1988). In order to do this, superempirical virtues are required, e.g. the use of inference (Haig 1996). "Qualitative accounts should always accompany quantitative work" (Fitz-Gibbon 1985, p. 48). The social scientist's use of meta-analysis, therefore, will be based on the perspective of a realist who believes that knowledge is what our best theories tell us (Haig 2013).

Purpose

Haig (1996) identifies three major goals for meta-analysis:

> "(a) identifying the worth of outcome studies;
>
> (b) identifying empirical regularities and;

(c) helping in the construction of explanatory theory" (p. 195).

In order to do this meta-analysis must integrate the findings from various studies to arrive at a synthesis of the research (Wachter 1988). This has always been its purpose. It offers a more rigorous, systematic and objective way than other methods (Fitz-Gibbon 1984). Whereas narrative reviews are too subjective regarding their selection of studies, vote-taking from statistical test-taking box score tallies is too reliant on sample size. Meta-analysis' wide variety of ingenious statistical methodology allows it to turn all types of research studies into real effects, making it an effective instrument to review what the research is saying (Borenstein 2011). Essentially, then, the principal function of meta-analysis is to fashion an empirical generalization from findings of primary studies in a given domain (Haig 1996). Its main work, however, is to detect the existence of robust empirical regularities, which is the statistical analog of direct experimental replication. Once this is accomplished, it sets the stage for the social researcher to begin the most important work of abductive reasoning, which requires the use of superempirical virtues (Haig 1996).

Parameters

It is key to set clear parameters in place in order to keep statisticians from doing the work that can only be done by researchers as described above. Instead of using an exhaustive list of criteria to appear thorough and inclusive, it is best practice to exercise discretion in the selection of studies and limit the review to best studies (Slavin 1995). The professional competency of the research community is at stake if agreement cannot be reached on what constitutes a quality study. Fitz-Gibbon (1984) identifies the following crucial factors to be present: adequate design, operationalisability of variables, context variables, standard deviation of control group and effect sizes. Perhaps it is high time for the research community to establish "research councils" (Fitz-Gibbon 1985, p. 47) before findings from a meta-analysis are published. This counters Glass and Smith's (1978) criticism that the use of best studies will only lead to an "ad hoc impeaching on methodological grounds of the studies of one's enemies" (Fitz-Gibbon 1984, p. 141). An intended consequence of such an arbitration might lead to a purification of professional journals to preserve the standard on quality research and to not be swayed and influenced by political agendas (Wachter 1988). This includes consulting the file drawers to rule out publication bias (Borenstein 2011).

It is important to note Glass' original rationale of meta-analysis in order to highlight vital distinctions between explanatory functions and evaluative properties (Haig 1996). This

will further bolster the parameters for effective use. In a number of publications Glass insists that meta-analysis is an evaluative exercise rather than a scientific one. By this he meant that it is sufficient to produce an outcomes study regarding the effectiveness of a program without having to give an explanation of the causal mechanisms behind those outcomes. This perspective of the use of meta-analysis rests on a faulty understanding (Haig 1995; 2013). Only producing a list of effect sizes on treatment effects without offering insight into how they are accomplished is to fall short of the goals and aims of science, which is always to provide "causal explanatory theories" (Haig 1996, p. 196). This is not a criticism of meta-analysis as a methodology. The intention here is to establish a parameter that will help keep meta-analysis as a tool in the hands of the researcher instead of it becoming an end product in itself. Neither does this say that meta-analysis has explanatory value in and of itself. While its most important work is to detect empirical data patterns, descriptions of the relationships between these variables are not considered explanatory in nature (Haig 2013). "The stable general effects gleaned from meta-analysis are empirical phenomena, and statements about phenomena are the objects of scientific explanation; they are not the explanations themselves" (Haig 2013, p. 18). Also, meta-analysis sometimes assists with the predictive success of explanatory theory testing, but this is not its usual role. "To employ meta-analysis to assist in the predictive testing of an explanatory theory does not thereby confer an explanatory role on meta-analysis itself. One does not assign status simply on the basis of association" (Haig 2013, p. 18). Meta-analysis is a statistical method and explanation is trans-statistical (Haig 1995). This again empasizes the importance of the researcher to employ meta-analysis as opposed to meta-analysis yielding conclusive findings on its own.

Product

In stating the premise, purpose and parameters of the proper use of meta-analysis, it is obvious that meta-analysis is a methodology, or "a perspective" (Bangert-Drowns 1986, p. 388) instead of an end product. With its ability to integrate and synthesize research findings meta-analysis enables the practitioner to explore stability hypotheses, that which have been constant; and variability hypotheses, that which are changing or have not been included (Terhart 2011). It therefore affords an understanding of the field (a look back) and allows for further exploration (what still needs to be done). Boston (2002) puts it even better when she says that meta-analysis "is provisional representing the best evidence at the time it was conducted, but subject to change in the face of the evolution of the knowledge base" (p. 4).

Whatever the outcome of meta-analysis, its ultimate contribution will be measured by how much it aided the social scientist to do the work of research in order to illuminate theory (Haig 2013). A strong recommendation is for all published meta-analysis' findings to be peer reviewed so that the editor, practitioner and policy-maker can benefit from expert understanding.

> Where meta-analysis is presented as an outcomes study, it should be accompanied by "thick descriptions" (Fitz-Gibbon 1985, p. 48) in order to aide the researcher in his task. Steven Goodman, in his editorial for Annals of Internal Medicine (1991) writes, Regardless of the summary number, meta-analysis should shed light on why trial results differ; raise research and editorial standards by calling attention to the strengths and weaknesses of the body of research in an area; and give the practitioner an objective view of the research literature, unaffected by the sometimes distorting lens of individual experience and personal preference that can affect a less structured review (Borenstein et al. 2011, unpaginated).

Conclusion

In critiqueing these very important and insightful peer-reviewed articles on the use of meta-analysis in education, one is tempted to return to the one-ring-that-rules-them-all metaphor. The discussion in this paper covered much ground and now that the journey is over, the question is whether Mount Doom awaits. Is there a place for meta-analysis in the social sciences and in particular the world of educational research? To be able to say *the research shows* (italics mine) reflects a natural and deep human desire for synthesis which practitioners and policy-makers desperately need. Has the much sought-after wisdom appeared along the journey? It is the position of this paper that meta-analysis supplies evidence, but it is evidence that "does not supply us with rules for action but only with hypotheses for intelligent problem-solving, and for making inquiries about our ends in education" (Snook *et al.* 2010, p. 97)

References

Bangert-Drowns, R.L., 1986. Review of Developments in Meta- Analytic Method. *Psychological Bulletin,* 99(3), pp. 388-399.

Bangert-Drowns, R.L. & Rudner, L.M., 1991. *Meta-analysis in educational research.* ERIC Clearinghouse.

Borenstein, M., Hedges, L.V., Higgins, J.P. & Rothstein, H.R., 2011. *Introduction to meta-analysis.* John Wiley & Sons.

Boston, C., ERIC Clearinghouse on Assessment and Evaluation. & United States. Office of Educational Research and Improvement., 2002. Effect size and meta-analysis. *ERIC digest.* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation,, p. 1 online resource (6 p.).

Bryman, A., 2012. Social Research Methods (4th Edition). Oxford: Oxford University Press, pp. 3-25.

Carver, R.P., 1978. The case against statistical significance testing. *Harvard Educational Review,* 48(3), pp. 378-399.

Carver, R.P., 1993. The case against statistical significance testing, revisited. *Journal of Experimental Education,* 61, pp. 287-287.

Coe, R., 2002. *It's the effect size, stupid: what effect size is and why it is important.*

Cumming, G., 2013. The new statistics why and how. *Psychological science*, p. 0956797613504966.

Denscombe, M., 2007. The Good Research Guide: for small-scale social research projects (3rd Edition). Buckingham: Open University Press, pp. 52-64.

Eysenck, H.J., 1978. An exercise in mega-silliness. *The American Psychologist,* 33(5), p. 517.

Eysenck, H.J., 1994. Meta-analysis and its problems. *BMJ: British Medical Journal,* 309(6957), p. 789.

Fitz-Gibbon, C., 1984. Meta-analysis: an explication. *British Educational Research Journal,* 10(2), pp. 135-144.

Fitz-Gibbon, C., 1985. The implications of meta-analysis for educational research. *British Educational Research Journal,* 11(1), pp. 45-49.

Gallo, P.S., 1978. Meta-analysis: A mixed meta-phor? *The American Psychologist,* 33(5), pp. 515-517.

Glaser, B., 1967. The Discovery of Grounded Theory. New York: Aldine de Gruyter, pp. 1-18.

Glass, G.V. & Smith, M.L., 1978. An exercise in mega- silliness: Reply. In: C.A. Kiesler, ed., pp. 517-519.

Haig, B.D., 1996. Statistical methods in education and psychology: A critical perspective. *Australian journal of education,* 40(2), pp. 190-209.

Haig, B.D., 2009. Inference to the best explanation: A neglected approach to theory appraisal in psychology. *Am. J. Psychol.,* 122(2), pp. 219-234.

Haig, B.D., 2013. The philosophy of quantitative methods. *The Oxford Handbook of Quantitative Methods in Psychology,* 1, p. 7.

Hattie, J., 2010. On Being a'Critic and Conscience of Society': The Role of the Education Academic in Public Debates. *New Zealand Journal of Educational Studies,* 45(1), p. 85.

Hattie, J., 2012. Know Thy Impact. *Educational Leadership,* 70(1), pp. 18-23.

Hattie, J., 2013. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge.

Hedges, L.V., 1982. Statistical Methodology in Meta-Analysis. ERIC Clearinghouse.

Mansell, W., 2008. *Research reveals teaching's Holy Grail* [Online]. United Kingdom. Available from: http://www.tes.co.uk/article.aspx?storycode=6005393 [Accessed June, 24 2014].

Nickles, T., 1981. What is a problem that we may solve it? *Synthese,* 47(1), pp. 85-118.

Nye, B., Konstantopoulos, S. & Hedges, L.V., 2004. How large are teacher effects? *Educational evaluation and policy analysis,* 26(3), pp. 237-257.

Rosenberg, M.S., 2005. The file-drawer Problem Revisited: A General Weighted Method for Calculating fail-safe Numbers in meta-analysis. *Evolution,* 59(2), pp. 464-468.

Rosenthal, R., 1979. The file drawer problem and tolerance for null results. *Psychological bulletin,* 86(3), p. 638.

Slavin, R.E., 1984. Meta- Analysis in Education: How Has It Been Used? *Educational Researcher,* 13(8), pp. 6-15.

Slavin, R.E., 1995. Best evidence synthesis: An intelligent alternative to meta- analysis. *Journal of Clinical Epidemiology,* 48(1), pp. 9-18.

Snook, I., Clark, J., Harker, R., O'Neill, A.-M. & O'Neill, J., 2010. Critic and conscience of society: A reply to John Hattie. *New Zealand Journal of Educational Studies,* 45(2), p. 93.

Snook, I., O'Neill, J., Clark, J., O'Neill, A.-M. & Openshaw, R., 2009. Invisible Learnings?: A Commentary on John Hattie's Book-'Visible Learning: A Synthesis of Over 800 Meta-analyses Relating to Achievement'. *New Zealand journal of educational studies,* 44(1), p. 93.

Terhart, E., 2011. Has John Hattie really found the holy grail of research on teaching? An extended review of Visible Learning. *Journal of Curriculum Studies,* 43(3), pp. 425-438.

Wachter, K.W., 1988. Disturbed by Meta- Analysis? *Science*, 241(4872), pp. 1407-08.