OMA05401

Self-Concept Intervention Research in School Settings:
A Multivariate, Multilevel Model Meta-Analysis

Alison J. O’Mara, Herbert W. Marsh, and Rhonda G. Craven
SELF Research Centre, University of Western Sydney, Australia

A multivariate, multilevel model meta-analysis was conducted to study the effectiveness of interventions for students in enhancing various self-concept domains. This was contrasted with a more traditional fixed effects model meta-analysis. In support of a construct validation approach to self-concept research, studies measuring self-concept outcomes directly related to the aims of the intervention (e.g. math self-concept in a math self-concept intervention) yielded higher effect sizes than those that measured outcomes of secondary or incidental relevance to the goals of the intervention. This finding concurrently supports multidimensional perspectives in self-concept research. Beyond this conceptually significant finding, findings are reported that have methodological repercussions in terms of both research methodology and intervention design. The ramifications for research methodology include the finding that higher effect sizes were observed for studies using random assignment to treatment/control conditions in the fixed effects model, but not in the multilevel model, and that direct self-concept interventions were not significant in either model. In terms of significance for self-concept intervention design, an interesting finding emerged suggesting that praise/feedback treatments were more effective in enhancing self-concept than other techniques. Throughout, the current paper contrasts the results of the commonly practiced meta-analytic method of fixed effects models with the revolutionary multilevel modeling approach. The result is a comprehensive paper that has implications for self-concept theorists, intervention designers, intervention evaluators, and meta-analysts.

The Importance of Self-concept and the Multidimensional Approach

Throughout the 20th century, self-concept has received attention from many eminent researchers due to the belief that it is fundamental to one’s personal and social development. In particular, educational researchers have established an association between self-concept and academic achievement (Delugach, Bracken, Bracken & Shicke, 1992; Marsh, 1990a; Marsh & Craven, in press; Valentine, DuBois & Cooper, 2004). Other educational outcomes linked to self-concept include coursework selection (Marsh & Yeung, 1997), educational and occupational aspirations (Marsh, 1991), academic motivation (McInerney, Roche, McInerney, & Marsh, 1997), and bullying (Marsh, Parada, Craven, & Finger, 2004; Marsh, Parada, Yeung, & Healey, 2001). Thus, enhancing self-concept is an important educational goal.

Despite the overwhelming consequence of the construct, the term ‘self-concept’ has become so commonplace that, for decades, theory development and accurate measurement were treated as superfluous to the goal of enhancing the construct (Marsh & Craven, 1997). There existed a general assumption that self-concept was unidimensional in structure, and therefore quite consistent across different contexts (see Bracken, 1996; Hattie, 1992 for reviews). As such, most self-concept interventions and their evaluation measures focused on a global, all-encompassing self-concept. It was not until 1976, when Shavelson, Hubner and Stanton developed a well-considered multidimensional model of self-concept, that the structure and measurement of self-concept became a central issue for researchers. After conducting a comprehensive review of the self-concept literature, Shavelson et al. found crucial shortfalls in the unidimensional research. In particular, they noted that there were no instruments providing clear support for the separation of self-concept into distinct domains. As a result, Shavelson et al. argued for a construct validity approach to the measurement of self-concept. They asserted:

It appears that self-concept research has addressed itself to substantive problems before problems of definition, measurement, and interpretation have been resolved. Until these problems have been dealt with in a manner made possible by advances in construct validation methodology, the generalizability of self-concept findings will be severely limited, and data on students’ self-concepts will continue to be ambiguous (p. 410).

To address these issues, Shavelson et al. (1976) developed a detailed hierarchical model of self-concept. Their rendition of self-concept suggested that the construct is multifaceted; hierarchically arranged; increasingly context-specific; and more differentiated with age. Subsequent research has vindicated their pioneering work by supporting a multidimensional structure of self-concept (Byrne, 1996b; Hattie, 1992; Marsh, 1993, 1990),
although some refinements to their theory have been made (Marsh & Shavelson, 1985). The proposed multidimensional structure provided a blueprint for a new generation of multidimensional self-concept instruments, and ensuing developments in theory, measurement and methodology have allowed considerable advances in the quality of self-concept research (see Marsh & Hattie, 1996; Byrne, 1996a, 1984). As stronger multidimensional instruments were developed, the case for a multidimensional structure of self-concept became more defined, emphasising the interdependent relation between theory development, measurement and practice (Marsh, 1990a).

The Evaluation of Interventions Using a Construct Validation Approach

As a hypothetical construct, self-concept is best understood through investigations of construct validity. Marsh (1993; Marsh & Craven, 1997) argued for a construct validity approach to intervention research in which the domains of self-concept that are most relevant to the intervention should be most affected, while less relevant domains should show less pronounced effects. In this way, the less relevant domains effectively serve as a control for response biases. However, only relatively recently have well-validated multidimensional instruments become available, and as such, construct validation approaches to understanding the multidimensionality of self-concept are only now becoming more common (see Hattie, 1992; Byrne, 1984, for reviews).

The within-construct aspects of construct validity examine the discreteness of – and relation between – self-concept domains, while between-construct studies attempt to establish the relation between the multiple dimensions of self-concept and a host of other constructs (Shavelson et al., 1976). Intervention studies that incorporate construct validation approaches usually focus on between-construct validation. For example, Marsh and Peart (1988) demonstrated that results of a physical fitness intervention and physical fitness indicators were substantially related to physical self-concept but nearly uncorrelated with non-physical self-concepts.

Perhaps one of the best examples of the construct validation approach to intervention studies is the series of studies based on the Outward Bound program by Marsh, Barnes and Richards (1996a, 1996b). From a construct validation approach, the juxtaposition of the two studies gives them added significance, because they targeted different self-concept domains and differentially affected those domains. To elaborate, the Outward Bound standard course is a 26-day residential program based on outdoor activities, and has a predominantly non-academic focus. Gains at posttest were significantly larger for the self-concept domains predicted a priori to be most relevant to the goals of the program (i.e., non-academic self-concept domains) compared to less relevant self-concept scales (i.e., academic self-concept domains). In contrast, the Outward Bound bridging course is a 6-week residential program designed to produce significant gains in the academic domain. The results showed that academic self-concept effects were substantial and significantly larger than non-academic self-concepts effects after completion of the bridging course. The authors argued that global measures of self-esteem would have led to the conclusion that the interventions were much weaker, and the broader understanding that the match between intended goals and outcomes provides would have been lost. Thus, the two studies combined, emphasise the need to target specific self-concept domains.

The importance of the match between the self-concept domain goals of an intervention and the measurement of relevant self-concept outcomes was central to a comprehensive meta-analysis by O’Mara, Marsh and Craven (2004). They conducted a meta-analysis of 200 self-concept interventions, in which they contrasted self-concept outcomes in terms of their relevance to the intervention; i.e., they employed a construct validity approach. They reported that many of the interventions they synthesised did not use an appropriate self-concept instrument for the goals of the study, for example, conducting a reading self-concept intervention and measuring the gains using a global self-esteem scale. They showed that those interventions which precisely measured the self-concept domain that was targeted in the intervention yielded higher effect sizes than studies that used mismatched scales. By employing this innovative construct validity approach, O’Mara et al. showed that multidimensional considerations are essential in determining intervention success. However, many researchers still utilise global (unidimensional) self-concept scales, even when the intervention is clearly multidimensional in nature. This has led to an underestimation of the efficacy of self-concept interventions, and has no doubt impeded self-concept intervention development (O’Mara et al.).
Direct Versus Indirect Self-concept Interventions

Many interventions, particularly in education, attempt to enhance self-concept indirectly by enhancing the individual’s abilities (O’Mara, et al., 2004). For example, an intervention might aim to increase a student’s mathematical ability in that hope that their self-concept will increase through the satisfaction of mastery of mathematical skill. It is intuitively expected that by improving one’s performance, then self-concept levels should also increase. Given this assumption, are direct self-concept interventions (i.e., those that explicitly aim to increase one’s self-beliefs) more effective in enhancing self-concept than indirect interventions that target skill building.

Research suggests that there is a reciprocal relation between self-concept and skill building, such that direct self-concept interventions can enhance both self-concept and related performance outcomes. A substantial amount of the support for this claim comes from educational research, which specifically suggests that prior levels of academic self-concept lead to higher levels of subsequent academic achievement, beyond what can be explained by prior levels of academic achievement (e.g., Marsh & Craven, in press; Valentine, Dubois, & Cooper, 2004; Marsh, Yeung & Byrne, 1999). Marsh and Craven (in press) therefore argue that researchers and practitioners should aim to enhance academic self-concept and academic abilities concurrently. They suggest that if practitioners enhance self-concepts without improving related skills, then the gains in self-concept are likely to be short-lived because the student’s performance would not live up to their heightened self-expectations. Likewise, if practitioners improve performance without also nurturing students’ self-beliefs in their capabilities, then gains in performance are also unlikely to be sustained, and in fact may even undermine self-concept in such a way that the short-term gains in performance will be lost. For instance, Marsh and Peart (1988) found that both competitive and cooperative interventions led to positive short-term gains in physical fitness (performance), but the cooperative intervention also led to enhanced self-concept, whereas the competitive intervention impaired the self-concept needed to sustain ongoing involvement in physical activity. As such, there is evidence that self-concept is both a valuable outcome measure in its own right and an important means to enhance other desirable outcomes. This line of research supports the use of direct self-concept interventions whereby a key goal of the intervention is to enhance self-concept.

Previous Meta-analyses of Self-concept Interventions

It is becoming increasingly evident that major scientific issues cannot be resolved by a single study, and that advances in knowledge come from the integration of many studies (Schmidt, 1992). A meta-analysis involves synthesising results from multiple studies to observe effect sizes across those studies on the phenomenon under review (Rosenthal, 1987). Moreover, meta-analyses go beyond traditional research in that they can identify patterns that are regularly veiled by traditional null-hypothesis testing, i.e., they have more power (Schmidt, 1992). Three meta-analyses have been conducted on self-concept interventions, each with different goals and methods. The results of these meta-analyses that are pertinent to the current study are discussed below.

Hattie’s (1992) Meta-Analysis

Hattie (1992) conducted a fixed effects meta-analysis of 89 self-concept intervention studies to investigate the effectiveness of self-concept treatments. The primary focus was on the intervention methods themselves, such as the types of therapy used to enhance self-concept. Self-concept was treated from a unidimensional approach, such that effect sizes in this meta-analysis were based on global or total self-concept/self-esteem scores. Hattie found that the average effect size for treatment groups was 0.37, which (by meta-analytic standards) suggests a weak to moderate overall increase in self-concept.

Haney and Durlak’s (1998) Meta-Analysis

Haney and Durlak (1998) performed a fixed effects meta-analysis on 99 self-concept intervention studies published prior to 1992. The primary focus was on the methodology of self-concept intervention research, such
as the use of randomised control group designs. Like Hattie (1992), Haney and Durlak used global and/or total self-concept/self-esteem scales, thereby espousing a unidimensional approach to self-concept. The overall mean effect size in this analysis was 0.27 – a slightly smaller mean effect size than the Hattie (1992) meta-analysis. Of particular note, treatments found to be specifically focused on self-concept enhancement (i.e., direct studies; \( d = 0.57 \)) were significantly more effective than treatments focused on other aspects, such as social skills (i.e., indirect studies, \( d = 0.10 \)). It was reported that studies that randomly assigned participants to treatment or control conditions yielded higher effect sizes. They also reported that studies using prior empirical evidence to inform their design had greater effect sizes (0.71) than those with theoretical bases (0.50) and those with no strong rationale (0.12). Given the inconsistencies in the employment of self-concept theories in intervention research, this is indeed a significant finding.

O’Mara, Marsh and Craven’s (2004) Meta-Analysis

O’Mara et al. (2004) conducted a fixed effects meta-analysis of 145 primary studies, which reported on 200 self-concept interventions for children and adolescents. The primary focus was a conceptual approach, which entailed the construct validity approach to self-concept enhancement research. As such, multiple distinct self-concept domains were included in the meta-analysis. They found an overall moderate effect size equal to that reported in Haney and Durlak (1998) of .27. Also like the former study, interventions whose focus was on self-concept enhancement had a higher mean effect size (.41) than interventions seeking to increase self-concept indirectly (.18). Further, random group assignment strategies produced higher mean effect sizes than non-random assignment procedures.

The limitations of the aforementioned self-concept intervention meta-analyses in terms of unidimensional approaches to self-concept were addressed in the O’Mara et al. (2004) study by accounting for the multidimensionality of the self-concept construct through a construct validity approach. The multidimensional structure of self-concept was supported by the finding that specific self-concept facet outcomes most relevant to the intervention had a higher mean effect size (.43) than facets with only secondary (.18) or incidental (.09) relevance to the intervention.

New Directions for Meta-analytic Techniques: Multilevel Modeling

The three meta-analyses of self-concept interventions described above all utilised a meta-analytic technique known as a fixed effects model. An essential assumption of a fixed effects model is that all studies included in a meta-analysis are drawn from the same population (Hedges & Vevea, 1998). This implies that there is no difference between the studies beyond sampling error variance. However, in reality this is highly unlikely. Most meta-analyses include quite distinct studies in their sample, and may differ in the intervention techniques used, the evaluation instruments employed, and even the basic design of the study. As such, it is increasingly seen as important to incorporate random error variance, which allows results of a meta-analysis to be generalised to studies not included in the sample of studies analysed (such as studies that have not yet been conducted). Meta-analysts are developing innovative ways of incorporating random error variance (e.g., Raudenbush, 1994); perhaps the most comprehensive and justifiable is the multilevel model meta-analysis.

Given that participants are nested within the studies under analysis, multilevel modelling can be applied to meta-analysis (Goldstein, 1995; Hox, 2002; Bryk & Raudenbush, 1992). Indeed, meta-analysis is increasingly being seen as a special form of multilevel regression analysis (Hox & de Leeuw, 2003). There are multiple benefits of conducting a meta-analysis using multilevel modelling. First, one can account for both sample variance and systematic variance in mean effect size estimation (Severiens & ten Dam, 1998), and can account for large variation between studies (Swanborn & de Glopper, 1999). It is also possible to generalize the findings to the population since it incorporates random effects. Multilevel modelling also allows the extension of traditional random effects models to include covariates (Van den Noortgate & Onghena, 2003; Goldstein, Yang, Omar, Turner, & Thompson, 2000). In addition, traditional techniques (based on Hedges and Olkin’s 1985 formula) may give less precise and more biased estimates of between-study variance (Van den Noortgate & Onghena, 2003). The multilevel approach also allows flexibility in modelling the data, because study characteristics can easily be included in the model as explanatory variables (Hox & de Leeuw, 2003). Further flexibility is afforded when one has multiple moderator variables (Bryk & Raudenbush, 1992). Finally, and
perhaps most importantly for the present study, multilevel modelling circumvents the issue of independence that plagues meta-analysis research (Kalain & Raudenbush, 1996). Violations of independence occur when studies produce multiple effect sizes because of multiple treatment groups or multiple outcome measures. Effect sizes from the same study are likely to be correlated in a way that may distort the results of the meta-analysis (Cooper, 1998). Given that the proposed meta-analysis seeks to compare and contrast multiple self-concept outcomes, this is a major issue. Multilevel modelling accounts for dependencies in the data whilst also allowing each study to contribute different effect sizes, thus maximizing the amount of information included in the analyses (Dickerson & Kemeny, 2004). As such, it is possible with multilevel modelling to specify a multivariate outcome model (Hox & de Leeuw, 2003). Bateman and Jones (2003) noted that improved modelling of the nesting of levels within studies increases the accuracy of the estimation of standard errors on parameter estimates and the assessment of the significance of explanatory variables.

Despite the growing appreciation of the ‘natural’ relation between multilevel models and meta-analysis (Draper, 1995), traditional meta-analytic practices are still commonly used (Van den Noortgate & Onghena, 2003). A literature search on meta-analyses conducted using multilevel modelling reveals a handful of studies, most of which have been conducted only recently. Furthermore, there is no published meta-analysis of self-concept studies using multilevel modelling techniques to date. In terms of implementation, this estimation procedure is relatively new and innovative, even by meta-analytic standards. Hence the present investigation was also designed to make a significant contribution to methodology by combining multilevel modelling and meta-analysis techniques to ensure the strongest available statistical tools were employed in the present investigation.

**The Present Study**

With the conceptual underpinnings of self-concept better understood, and armed with a new generation of multidimensional self-concept instruments, it would seem self-concept interventions would be better placed than ever before to achieve their aims. Surprisingly, many self-concept researchers have failed to adopt these new theoretical perspectives and instruments. Until the 1990s, in the midst of the flurry of conceptual advances, self-concept enhancement research stagnated by perpetuating the age-old practice of using unidimensional instruments (Marsh & Craven, 1997). Now, with manifold intervention reports that utilize a unidimensional perspective, a multidimensional structure, or a mixture of the two, it is difficult to disentangle the results to identify those techniques that are most effective. Using the successful construct validity approach first implemented by O’Mara, Marsh and Craven (2004), the proposed meta-analysis will contrast various self-concept domain outcomes to better explore intervention efficacy and explicate enduring misconceptions. This will be achieved using the innovative, sophisticated multilevel model meta-analysis.

The aims of the present study are:

1. To elucidate the multidimensional structure of the self-concept construct through a construct validity approach to intervention effects, by examining the relevance of the measured self-concept to the intended goals of the intervention. This is based on a body of research supporting multidimensional models over the (still common) unidimensional perspectives of self-concept (Marsh, Craven, & Martin, in press) and the results of a prior meta-analysis using a construct validity approach (O’Mara et al., 2004);

2. To examine whether interventions directly aimed at enhancing self-concept yield higher effect sizes than those that indirectly target self-concept (e.g., through skill-building techniques). Both Haney and Durlak (1998) and O’Mara et al. (2004) have found direct interventions to have more impact upon self-concept at posttest. This is consistent with a reciprocal effects model of the relation between self-concept and performance (Marsh & Craven, in press);

3. To explore the effects of random vs. nonrandom assignment to treatment and control conditions. The group assignment procedure employed in intervention research has been implicated in distorting the observed effectiveness of interventions in meta-analysis, because outcomes may be biased due to potentially differential levels of self-concept in the groups at pre-test (Lipsey & Wilson, 2001; Matt & Cook, 1994). Two previous meta-analyses of self-concept intervention literature (O’Mara et al., 2004; Haney & Durlak, 1998) have found differential results for random and non-random assignment procedures, and will therefore be included as moderators of effect size;

4. To examine whether praise/feedback methods are effective in instigating self-concept enhancement. Hattie (1992) examined psychotherapeutic approaches to self-concept enhancement, and found that
interventions based upon transactional analyses yielded highest mean effect sizes. However, little is currently known about what treatment features that are commonly used in educational settings are most effective in enhancing self-concept. Craven, Marsh and Burnett (2003) suggested that praise and feedback interventions are particularly effective. This was supported by O’Mara, Marsh, Craven and Debus (in press), in which they found praise/feedback techniques to be more successful than all other methods, including discussion groups, counseling, and skills training; and

5) To employ the innovative multilevel modeling approach to meta-analysis, and contrast the results with a fixed effects model meta-analysis. Multilevel modeling is the cutting-edge research synthesis, which has many advantages over the prevailing fixed effects models (Hox, 2002).

Method

Sample of Studies

The present meta-analysis utilised the same meta-analytic database as used in O’Mara et al. (2004). The selection criteria used by Haney and Durlak (1998) was partially adopted in order to enhance comparability with that previous meta-analysis. As such, inclusion was restricted to studies published in English, involving children and adolescents with a mean age of 18 or younger, and including a control group drawn from the same population as the intervention group. The study also had to include a measure of self-concept or another related self term (e.g., self-esteem), which could be either a global measure or one that tapped a specific domain (e.g., academic self-concept). Unlike the Haney and Durlak meta-analysis, however, the present investigation excluded studies using inferred self-concept measures, following Marsh’s (1990) argument that self-concept inferred by others (e.g., teachers, parents) represents a different construct. Thus, 96 of the 102 studies used in the Haney and Durlak meta-analysis were included in the sample.

The present study included studies published between 1958 and 2000, whereas the prior meta-analysis only sampled up to 1992. The additional studies were acquired primarily through searches of Psychological Abstracts, Expanded Academic ASAP, Social Sciences Index, PsychINFO and ERIC databases. These sources were searched using the key words self-concept, self-esteem or self-efficacy and child, adolescent. Studies in the Hattie (1992) meta-analysis and the references of identified studies were also investigated to see if any studies mentioned in these sources satisfied the criteria. Studies selected included those that incorporated specific self-concept interventions (target studies), as well as those comprising interventions with a principal focus on constructs other than self-concept that included an outcome measure of self-concept (non-target studies). The search yielded 145 appropriate studies.

Coding Procedures

A code sheet was developed in order to capture information about potential moderator variables. A code book was also produced to assist in coding and to ensure consistency in the ratings by two coders. Coding categories were also devised to reflect the multidimensional perspective under consideration, including construct validity approaches to assessing self-concept outcomes. The coding categories included: focus of the intervention on self-concept, relevance of self-concept domain outcome, match between target intervention and self-concept domain relevance, design characteristics (randomisation and control group type), treatment administrator, prevention versus treatment studies, and the rationale used to inform the intervention.

Two stages of coding reliability checks were conducted. The first stage consisted of five rounds of pilot testing with two coders, which entailed discussions over disparity in coding and subsequent amendments to the code sheet and code book. By the end of the fifth round, the two raters had a high level of concurrent rating and the code sheet and book were thus deemed suitable for use. The second stage involved both coders coding a random selection of 52 articles, which resulted in a mean agreement rate of 92.7%. The remaining 93 studies were coded solely by the first author.
Computation of Effect Sizes

Of the 145 studies located, some had multiple treatment groups, and many measured more than one self-concept domain. Consequently, coding was conducted for 200 interventions comprising 460 unique effect sizes. Effect sizes were calculated for self-concept outcomes using the Comprehensive Meta-Analysis software program (version 1.0.23; Borenstein & Rothstein, 1999). The program derives \( d \) (the standardised mean difference), which is calculated by inputting either (1) raw mean difference and common within group standard deviations, (2) group means and standard deviations for both treatment and control groups, (3) \( t \)-values, or (4) \( p \)-values.

The most common method of calculating \( d \) was based on means and standard deviations of the treatment and control groups. When post-test only scores were available, the Comprehensive Meta-Analysis program calculated the effect size \( d \) using \( d = (M_t - M_c) / SD_c \), where \( M_t \) = mean for the target intervention group and \( M_c \) = mean for the control group and \( SD_c \) = standard deviation of the control group. When pretest and post-test scores were available for both groups, the effect size was calculated using \( d = [(M_{post} - M_{pre}) - (M_{post} - M_{pre})] / SD_{pre} \), where \( (M_{post} - M_{pre}) \) = mean difference between pre- and posttest for the target intervention group, and \( (M_{post} - M_{pre})_c \) = mean difference between pre- and posttest for the control group, and \( SD_{pre} \) = pooled pretest standard deviation. Alternative methods were used when means or standard deviations were not reported (cf. Lipsey & Wilson, 2001). The effect size was set at zero if interventions reported non-significant findings and no other applicable data was supplied. Once the Comprehensive Meta-Analysis program had calculated the unadjusted effect size, the effect sizes were imported into SPSS for analysis. SPSS macros developed by Wilson (see Lipsey & Wilson, 2001) were used to run the analyses (as discussed below), and these macros automatically adjusted the effect sizes using Hedges’ small sample size bias correction. Further, each standardised mean change effect size was weighted by the inverse of its sample error variance using the formula

\[
v = \frac{n1 + n2}{(n1)(n2)} + \frac{d^2}{2(n1 + n2)^2},
\]

where \( n1 \) = treatment group sample size, \( n2 \) = control group sample size, and \( d \) = adjusted effect size.

Fixed Effects Analyses

All fixed effects analyses were run using SPSS 12.0.1. For the overall mean effect size, a macro designed by Wilson (see Lipsey & Wilson, 2001) was used, which weighted the effect sizes by the inverse of the variance of the effects. Since many of the primary studies contributed more than one effect size to the analyses, it was necessary to adopt a shifting unit of analysis approach to minimise violations of independence (Cooper, 1998). For example, a study targeting social self-concept might report both social self-concept (target self-concept outcome) and math self-concept (nontarget self-concept outcome). In this case, the effect size for the two self-concept domains would be averaged to produce a single effect size for calculations involving the overall effect size for the sample. Thus, once the effect sizes were aggregated to produce one effect size for each intervention, the overall mean effect size for the posttest analyses was based on 145 effect sizes.

If the homogeneity results of the overall mean effect size analyses are significant, it means that the effect sizes have significant variation between them and so moderator variables need to be modeled to identify systematic differences between effect sizes. The homogeneity analyses follow a chi-square distribution (Lipsey & Wilson, 2001).

For the moderator variable model, macros specifically designed for running meta-analytic, weighted multiple regression analyses in SPSS were used (see Lipsey & Wilson, 2001). All variables were entered simultaneously into an all-inclusive regression model. However, the shifting unit of analysis (Cooper, 1998) was also adopted for the predictor variable analyses. Thus, the effect sizes were aggregated to produce an effect size for each intervention for each category of the variables considered. To further minimise violations of the assumption of independence, in both the overall effect size analysis and the predictor variable analysis, the weighting of the effect sizes was divided by the number of effect sizes in each study. This helped to reduce an excessive impact of studies contributing more than one effect size to the analyses.
Multilevel Modeling Analyses

All analyses were conducted using MLwiN 2.02 using restricted maximum likelihood estimation procedures (known as RIGLS in MLwiN). In our model, we stipulated three levels: An outcome level component (level 1), an intervention-level component (level 2), a study-level component (level 3). The first step in the analyses was to estimate the intercept-only model, which if significant, is followed by the moderator analyses (described below).

Intercept-Only Model

For the intercept-only model, no predictors were included in the model. The intercept-only model is represented by the equation:

\[ d_{ijk} = \beta_{000} + v_{0k} + u_{0jk} + e_{ijk} \]

where \( d_{ijk} \) refers to the effect size for outcome \( i \) from intervention \( j \) and study \( k \), \( \beta_{000} \) refers to the intercept (average effect size for an average outcome), \( v_{0k} \) is the random error at level 3, \( u_{0jk} \) is the random error at level 2, and \( e_{ijk} \) is the random error (residual) at Level 1. The variance of \( v_{0k} \) and \( u_{0jk} \) indicates the variability in effect sizes. The variance is partitioned into within-study and between-study variance. A chi-square test is used to test between-study homogeneity. If the studies are homogeneous (i.e., there is no between-study variance), there is no significant difference between studies on the variable of interest. Any variance in effect sizes is due purely to sampling variance (Lipsey & Wilson, 2001). As such, no further analyses will be conducted. However, a significant variance component suggests that there is variance unexplained by the model: The model is heterogeneous (Bryk & Raudenbush, 1992). In other words, the participants take on a different form for a particular variable in each study included in the meta-analysis.

Moderator Analyses

If the results of the chi-square reveal heterogeneity, potential predictors are modeled to see if they explain systematic variance between the study effect sizes (Hox & de Leeuw, 2003). For example, one would test to see if design factors moderate the effect sizes. The question would be asked: Does the self-concept outcome for different studies differ according to design type (pretest-posttest or control group)? This requires that the model be expanded to include predictor variables.

The following hierarchical model was used:

\[ d_{ijk} = \beta_{000} + \beta_1 W_{1j} + \beta_2 W_{2j} + \ldots + \beta_s W_{sj} + v_{0k} + u_{0jk} + e_{ijk} \]

where \( d_{ijk} \) is the mean effect size, \( \beta_0 \ldots \beta_s \) are the regression coefficients, \( W_{1j} \ldots W_{sj} \) are the study characteristics (predictor or moderator variables), \( v_{0k} \) is the systematic variability in study \( k \) not captured by the \( s \) predictors, \( u_{0jk} \) is the systematic variability in intervention \( j \) not captured by the \( s \) predictors, and \( e_{ijk} \) is the sampling error for study \( k \) (Bryk & Raudenbush, 1992). The intercept \( \beta_{000} \) is the estimated effect size for a study with zero values for all moderator variables. The remaining regression weights \( \beta_0 \ldots \beta_s \) indicate the amount of expected variation in the effect size for a one-unit change on each variable. A likelihood ratio test was then conducted to compare the single level linear regression model with the multilevel model, where we estimated the between-study variation in the intercepts.

Results

Fixed Effects Model

Overall Mean Effect Size

The overall mean effect size in the fixed effects model was .31 (\( p < .001 \)). The results of the homogeneity analyses were significant (\( Q(144) = 1668.34, p < .001 \)) suggest that moderator variables need to be modeled.

Moderator Analyses – Weighted Multiple Regression

The results of the fixed effects model regression analyses are shown in Table 1. As can be seen, only three of the five variables were significant predictors of effect size: whether the effect size was a target self-concept outcome, whether the intervention used random assignment to conditions, and whether the intervention primarily involved praise/feedback techniques. The use of praise/feedback was the strongest predictor of effect sizes (\( B = \)
The model accounted for 17.68% of the variance in the effect sizes, suggesting that there is still a large proportion of unexplained variance ($r^2 = .177$).

Table 1. Regression Coefficients and Standard Errors of the All-Inclusive Fixed Effects Weighted Multiple Regression Model

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>$\beta$</th>
<th>SE</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-.058</td>
<td>.063</td>
<td>ns</td>
</tr>
<tr>
<td>Target self-concept outcome</td>
<td>.225</td>
<td>.068</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Target-related self-concept outcome</td>
<td>.100</td>
<td>.070</td>
<td>ns</td>
</tr>
<tr>
<td>Direct intervention</td>
<td>.138</td>
<td>.075</td>
<td>ns</td>
</tr>
<tr>
<td>Random assignment</td>
<td>.212</td>
<td>.045</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Praise/feedback</td>
<td>1.292</td>
<td>.141</td>
<td>&lt; .05</td>
</tr>
</tbody>
</table>

Multilevel Model

Overall Mean Effect Size

Recall that the intercept-only model is used to establish the overall mean effect size and whether further moderator analyses are required based on the homogeneity results, and therefore contains no predictors. The coefficient in the intercept-only model was 0.47 ($SE = .06$), which means that the overall mean effect size for the sample of studies is .47. There is significant heterogeneity in the effect sizes, as indicated by the results of the chi-square homogeneity test ($\chi^2 (144) = 511.87, p < .001$). Therefore, predictor variables were added to the model.

Moderator Analyses – Multilevel model

Table 2 shows the results of the model including the hypothesised predictors. It can be seen in the table that target self-concept outcomes, target-related self-concept outcomes, and praise/feedback strategies all yielded significant coefficients. Direct interventions and random assignment procedures were not significant predictors of effect size. A likelihood ratio test (which follows a chi-square distribution) was conducted to see if the null hypothesis (that is, that $\beta_1 = \beta_2 = \beta_n = 0$) should be rejected. The likelihood ratio test statistics was $4160.365 - 4020.879 = 139.486$ with 6 degrees of freedom, $p < .001$. Therefore, the result was significant, which means that there are real differences between studies in the population mean of effect sizes even after the predictors are included. As such, there is still unaccounted variance in the model, although the predictor-included model is a significant improvement on the intercept-only model.

Table 2. Regression Coefficients and Standard Errors of the All-Inclusive Multilevel Model

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>$\beta$</th>
<th>SE</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-.075</td>
<td>.107</td>
<td>ns</td>
</tr>
<tr>
<td>Target self-concept outcome</td>
<td>.286</td>
<td>.044</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Target-related self-concept outcome</td>
<td>.217</td>
<td>.050</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Direct intervention</td>
<td>.116</td>
<td>.164</td>
<td>ns</td>
</tr>
<tr>
<td>Random assignment</td>
<td>.219</td>
<td>.124</td>
<td>ns</td>
</tr>
<tr>
<td>Praise/feedback</td>
<td>.465</td>
<td>.237</td>
<td>&lt; .05</td>
</tr>
</tbody>
</table>

Discussion

The results of the present investigation suggest that self-concept can be improved through enhancement treatments with children and adolescents, with a mean effect size of .31 in the fixed effects model and .47 in the multilevel model. This is a promising finding, given the multitude of self-concept interventions in practice, particularly in school settings.
The results further supported the use of construct validation approaches to the study of multidimensional self-concept in meta-analysis, which was established by O’Mara et al. (2004), since in both models, targeted self-concept outcomes were the highest mean effect sizes. Interestingly, the two models differed as to whether classification as a target-related self-concept outcome was a significant predictor of effect size, with only the multilevel model finding this variable significant. This could be explained to differences between the two models. Fixed effects models known to have wider confidence intervals than models that include random effects components (Hedges & Vevea, 1998), which the multilevel model does, and also fail to account for random variance between the interventions and studies (Raudenbush, 1994). These two features of fixed effects models mean that they are less accurate in identifying significant results. It is therefore reassuring to see that the more sophisticated and rigorous multilevel models employed here support the construct validity approach.

Of much importance, the results of the present meta-analysis did not support the findings of either Haney and Durlak (1998) or O’Mara et al. (2004) regarding direct interventions leading to greater self-concept enhancement, which was a significant finding in the previous meta-analyses. In the present analyses, both the fixed effects models using a shifting unit of analysis (Cooper, 1998) and the multilevel model found this variable to be non-significant. This does not, however, contradict a reciprocal effects model of self-concept and performance, which proposes that self-concept can be improved through indirect means (i.e., skill building) as well as direct effects (see Marsh & Yeung, 1997). This finding has applications for intervention designers, who should consider using both direct and indirect treatments to enhance self-concept. Given that there is a vast array of interventions that are designed to enhance self-concept implemented in educational settings, this finding could lead to a new generation of joint-focus treatment programs in schools. Future studies could look at whether non-self-concept outcomes (e.g., academic achievement) were also positively impacted on by the interventions. It would also be useful to test whether the gains in both self-concept and performance are maintained after the potential euphoric effects of the intervention wane at delayed posttest, which could help to clarify the relation as proposed by the reciprocal effects model (see Marsh & Craven, in press).

Random assignment of participants to conditions was found to be a significantly positive predictor of effect size in the fixed effects model, as was found in Haney and Durlak (1998) and O’Mara et al. (2004). However, the use of multilevel modeling found this variable to be non-significant. It has been argued elsewhere (O’Mara et al.) that quasi-experimental studies may lead to the underestimation of intervention effects, particularly if intervention groups are based on students who initially have more problems than non-randomly assigned comparison groups. It was therefore encouraged that researchers employ random designs where practicable, and that meta-analysts include random assignment in their predictor models (O’Mara et al.). Despite the lack of statistical significance in the multilevel model, we still encourage this practice. The variable “random assignment” in the present analyses encompassed a range of assignment methods, whereas O’Mara et al (2004) used more refined categories, and so the agglomeration of various types of random strategies might have led the group to be too heterogeneous to reach significance.

Praise and/or feedback techniques were found to be significant predictors of effect size in both models. This follows from reviews by Craven et al. (2003) suggesting the value of such interventions, and the results of O’Mara et al. (in press) showing praise/feedback to be the most effective treatment technique. Such interventions are generally easy to implement in the classroom, require little training and minimal cost, and so offer a promising direction for educational practitioners. It is interesting to note, however, that the multilevel model was more conservative in its estimate of the predictive ability of this variable on effect size. This emphasizes the importance of using models that account for random sources of error and violations of independence, as the multilevel model does.

In summary, multilevel modeling has proven to be a useful new direction for meta-analysis. It affords greater confidence in the accuracy of the results than a traditional fixed effects model, and also allows the results to be generalised to the greater population of studies. To the best of our knowledge, this is the first ever published three level multilevel model meta-analysis, and certainly the first multilevel model meta-analysis for self-concept interventions. Important findings in the current multilevel modeling meta-analysis support those of previous meta-analyses regarding the construct validity approach to multidimensional self-concept interventions, and the use of praise/feedback interventions. However, direct interventions were not found to have significantly higher effect sizes than indirect interventions, as has been found in prior meta-analyses. This attests to the more rigorous statistical procedures of the multilevel modeling approach, and highlights new directions for causal modeling researchers. The current study thereby offers valuable information for: self-concept theorists in terms
of the construct validity approach; intervention designers, as evidenced by the efficacy of praise/feedback interventions; intervention evaluators, from the perspective of measuring relevant self-concept domain outcomes; and meta-analysts, through the use of cutting-edge statistical procedures.

About the Authors

**Alison O’Mara** is a Doctoral candidate with the SELF Research Centre. Her research involves a series of multilevel model meta-analyses related to multidimensional self-concept. Alison has presented refereed papers at international conferences, is the past recipient of the National Leadership Scholarship, and the current recipient of a prestigious Australian Postgraduate Award.

**Professor Herb Marsh** is Director and Research Professor of the SELF Research Centre, University of Western Sydney. Largely due to his established research program into self-concept, motivation, identity and related constructs that are the focus of the SELF Research Centre, he is arguably Australia’s leading researcher in both the broad disciplines of Education and Psychology.

**Professor Rhonda Craven** is Deputy Director of the SELF Research Centre – ranked 7th in the world in educational psychology, is Associate-Professor in the School of Education and Early Childhood Studies, University of Western Sydney. As an Educational Psychologist her research focuses on large-scale quantitative research studies in educational settings.

Contact Details

Ms. Alison O’Mara  
PhD Candidate  
SELF Research Centre  
University of Western Sydney  
Bankstown Campus  
Locked Bag 1797  
Penrith South DC NSW 1797  
Australia  
Email: a.omara@uws.edu.au  
Phone: +61-414-726-535

References


