Evaluating Self-Concept Interventions from a Multidimensional Perspective:
A Meta-Analysis

Alison J. O’Mara, Rhonda G. Craven, and Herbert W. Marsh
Self-concept Enhancement and Learning Facilitation Research Centre
University of Western Sydney, Australia

Paper presented at NZARE AARE, Auckland, New Zealand November 2003
OMAO3786
Evaluating Self-Concept Interventions from a Multidimensional Perspective:  
A Meta-Analysis

Alison J. O’Mara, Rhonda G. Craven, and Herbert W. Marsh  
Self-concept Enhancement and Learning Facilitation Research Centre  
University of Western Sydney, Australia

Through a comprehensive meta-analysis, self-concept interventions that specifically target self-concept and evaluate self-concept domains relevant to the intervention were found to be more effective than interventions that do not. These findings integrate notions of domain specificity, multidimensionality, and construct validity that have not been fully incorporated into self-concept intervention and practice. Building on previous self-concept (SC) and self-esteem (SE) meta-analyses by Hattie (1986) and particularly Haney and Durlak (1998), the present meta-analysis was based on a much larger and more recent database (154 studies, 200 interventions, and 544 effect sizes), and incorporated a multidimensional perspective. Other variables found to be predictors of intervention efficacy included: treatment administrator, theoretical basis of intervention, and methodological features (type of design and control group). It was further found that benefits did not dissipate over time, although only a small number of studies conducted follow-up testing.

Self-concept is a critical goal in itself, as well as a means to facilitate other desirable outcomes in a diversity of settings. Higher levels of self-concept have been associated with many benefits, such as improved social interaction (Gurney, 1986), coping skills (Shirk, 1988) and academic achievement (Delugach, Bracken, Bracken & Shicke, 1992). Conversely, low self-concept has been linked to a wide range of maladaptive behaviours, including suicide, substance abuse and juvenile delinquency (Marsh & Craven, 1997).

Not surprisingly, then, self-concept and self-esteem (SC/SE) interventions are common, particularly in educational settings (Marsh & Craven, 1997). As a result, there exists a plethora of studies on SC/SE interventions, with mixed results (Haney & Durlak, 1998; Hattie, 1992). The conflicting results appear to be due to the theoretical and methodological weaknesses that have plagued this field of research for many years. For instance, Haney and Durlak (1998), in a meta-analysis on self-concept interventions, found that the mean effect size for self-concept (or target) studies was significantly higher than for non-target studies. The failure to distinguish between target and non-target studies (and, indeed, any core feature of the interventions) could partially explain the generally unimpressive findings in SC/SE intervention research (Marsh & Craven, 1997). Recent advances in theory (Marsh, 1990; 1993; Marsh & Hattie, 1996) and the development of multidimensional measuring instruments (Hattie, 1992) based on new theoretical models have helped to rectify these flaws.

Theoretical Considerations

Originally, self-concept was conceived to be a unidimensional construct (see Byrne, 1984). However, in their classic review of self-concept research, theory, and measurement, Shavelson, Hubner and Stanton (1976) developed a hierarchical model of self-concept that fundamentally impacted on self-concept research (Marsh & Hattie, 1996). This theory posits that general self-concept is at the pinnacle of the hierarchy, and dimensions that are more specific such as academic and social self-concept can be found lower down the structure.
Research in the 1990s (Byrne, 1996; Vispoel, 1995) showed that the proposed hierarchy was weak, although specific components of self-concept (for example, social, academic, physical, artistic and emotional) were highly differentiated (Harter, 1998; Bong, 1997; Byrne, 1996). That is, despite unconvincing support for the hierarchical structure, research has strongly supported the multidimensionality of self-concept (Hattie, 1992; Marsh and Shavelson, 1985). In educational psychology for example, many important academic outcomes (e.g., academic achievement and coursework selection) are substantially related to academic self-concept, but relatively unrelated to general and non-academic components of self-concept (Marsh, 1993; Marsh & Craven, 1997). Moreover, research has indicated that interventions are more likely to succeed if they target specific facets of self-concept (such as mathematics and physical self-concept; Marsh & Craven, 1997). The apparent corollary of a multidimensional self-concept is that the measurement of greater domains of specificity of self-concept leads to improvements in both the design and evaluation of self-concept enhancement studies.

Likewise, it has been found that the relevance of the outcome measure to the intervention is a pertinent focus in self-concept research (Hattie, 1986). For example, implementing a maths intervention, and then measuring maths self-concept will produce a higher mean effect size than if we were to measure general self-concept. Hattie (1992) suggested that global measures might give a distorted judgment of self, whereas tests with dimensional specificity are more effective in determining domain success.

Methodological Flaws
Research has found that poor methodology in SC/SE intervention research stems from small sample sizes, inadequate reporting of variables such as age and ethnicity of participants, non-randomised designs and a lack of a control group (Haney & Durlak, 1998; Hattie, 1992). Insubstantial interventions, insufficient connection between aims and the outcomes measured, and an under-use of multidimensional instruments with construct validity have also been implicated in undermining the effectiveness of SC/SE enhancement research (Hattie, 1992; Marsh & Craven, 1997).

The Next Generation in Self-Concept Enhancement Research
Multidimensional evaluative instruments have now been constructed to allow more accurate probing of the multidimensional quandary (Byrne, 1996). Consequently, the recent advances in the quality of SC/SE research are due to better theoretical models, superior measurement instruments, improved methodology, a focus on a multidimensional self-concept, and stronger interventions (Marsh & Craven, 1997). New SC/SE enhancement studies can better contribute to identifying techniques for the enhancement of self-concept and related constructs.

However, 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). To make sense of the large volume of conflicting research in the field, it is informative to statistically review developments in SC/SE enhancement research by conducting a meta-analysis. A meta-analysis is a complex statistical technique, utilising quantitative procedures to integrate and compare research results (Rosenthal, 1984).

Hattie (1986) conducted one such meta-analysis and found that cognitive treatments were more effective than affective and “other” treatment types. Furthermore, supporting Elardo and Elardo’s earlier finding (1976), Hattie reported that although many schools’ curricula
state that teachers should facilitate SC/SE enhancement, classroom teachers were actually the least effective change agents. Haney and Durlak (1998), also recognising the need for a coherent and analytical review of the SC/SE literature, performed a meta-analysis of SC/SE intervention studies. Of particular note, they observed that interventions targeting self-concept, as opposed to those that measure self-concept incidentally to another intervention, are more effective. Also, they noted that many studies fail to conduct follow-up testing.

The Present Investigation

The major improvements in SC/SE enhancement studies have occurred since Haney and Durlak’s (1998) study, which was based on pre-1992 studies. The primary purpose of the present study is to critically analyse SC/SE enhancement studies, with the necessary inclusion of the multidimensional approach that has increased in prevalence since Haney and Durlak’s (1998) study. Using meta-analytic techniques, this study aims to: a) assess the benefits of interventions targeting SC/SE versus those that aim to improve SC/SE through other means (e.g., social skills training); b) explore the construct validity approach to the study of intervention effects by assessing whether components of SC/SE most logically related to the intervention produce higher effect sizes; c) examine mechanisms for increasing the effectiveness of interventions (e.g., treatment administrator type); d) inspect the impact different design features (e.g., control group type) have on observed effect size, and e) investigate the resilience of outcomes associated with the treatment at follow-up.

Method

Sample of Studies

As a partial extension of Haney and Durlak’s (1998) study, it was necessary to adopt their selection criteria in the sampling procedure. The criteria was restricted to published studies, and consisted of: studies involving children or adolescents with a mean age of 18 or younger, studies containing at least one outcome measure of SC/SE, and studies containing a control group drawn from the same population as the intervention group. Unlike the Haney and Durlak meta-analysis, the present investigation excluded studies using inferred self-concept measures based on responses by others, as Marsh (1990) specifically argued that self-concepts inferred by others (e.g., teachers, parents) represent a different construct.

Thus, 96 of the studies used in the 1998 meta-analysis were included in the sample. Additional studies were acquired primarily through searches of Psychological Abstracts, PsychINFO and ERIC databases, and through examination of the references of identified studies. Studies included those that incorporated specific SC/SE interventions (target studies), as well as those comprising interventions with a principal focus on constructs other than SC/SE that included an outcome measure of self-concept or self-esteem (non-target studies).

Coding Procedures

Using the selection criteria above resulted in the location of 154 studies. Some studies had multiple treatment groups, and many measured more than one SC/SE domain. Consequently, coding was conducted for 200 interventions comprising 544 effect sizes. The 544 effect sizes were separated into two files: one pertaining to posttest results ($N = 460$), the other in relation to follow-up testing effect sizes ($N = 84$). These were analysed separately, and all results (below) refer to the posttest scores except where specified otherwise.
The major focus of the present meta-analysis is the inclusion of multidimensionality of self-concept. Hence, coding categories were devised to reflect these recent developments in the literature, including construct validity approaches to assessing self-concept outcomes. The coding scheme used by Haney and Durlak (1998) was consulted during the development of the code sheet in order to provide a basis of comparison with the earlier meta-analysis. The integration of features of the previous meta-analysis with recent developments resulted in a comprehensive coding system. The following subsection of coding variables were subjected to statistical analysis, as they relate to the aims of the study: focus of the intervention on self-concept; relevance of SC/SE domain outcome; match between target intervention and SC/SE domain relevance; design characteristics (randomisation and control group type); treatment administrator; prevention versus treatment studies; and the rationale used to inform the intervention.

Coding Reliability
Two stages of coding reliability checks were conducted, which involved a second coder. The first stage consisted of five rounds of pilot testing, entailing discussions over disparity in codings and subsequent amendments to the code sheet and codebook. By the end of the fifth round, the raters had a high level of concurrent rating and the code sheet and book were thus deemed suitable for use. The second stage involved the coding of evaluative variables (e.g., overall quality of the intervention) by the second coder, of a randomly selected 52 articles so that coefficient kappa could be calculated (where appropriate). Concurrently, the first author coded all 154 studies. Coding reliability was deemed satisfactory, with a mean percent agreement of 92.7%.

Computation of Effect Sizes
Effect sizes were calculated for SC/SE outcomes using the Comprehensive Meta-Analysis software program (version 1.0.23; Borenstein & Rothstein, 1999). The program derives \( d \) using the standardised mean difference, which is calculated by inputting either: means and standard deviations, \( t \)-values, or \( p \)-values. When post-test only scores are available, the effect size (\( d \)) is:
\[
d = \frac{(M_t - M_c)}{SD_{pooled}}.
\]
Where \( M_t \) = mean for the target intervention group and \( M_c \) = mean for the control group and \( SD_{pooled} \) = pooled standard deviation based on both groups.

When pretest and post-test scores are available for both groups, the effect size (\( d \)) is
\[
d = \frac{([M_{post} - M_{pre}]_t - [M_{post} - M_{pre}]_c)}{SD_{pre(pooled)}}.
\]
Alternative methods were used when means or SDs were not reported (cf. Wolf, 1986; Lipsey & Wilson, 2001). As in Haney and Durlak’s (1998) meta-analysis, the effect size was set at zero if interventions reported non-significant findings and no other applicable data was supplied. This occurred for 181 (31.6%) of the effect sizes. Of the calculable effect sizes (i.e., those not set at zero because of insufficient information), 80.6% were positive. This suggests that the procedure of setting \( d \) at zero may have underestimated the actual effect on self-concept (Durlak & Lipsey, 1991) in the present review. The sign of the difference was positive when the treatment had a larger improvement than the control, whereas a negative \( d \) value indicates a higher score for the control groups.

Categorical Fixed Effects Model Testing and Homogeneity of Effect Sizes
Categorical models are used to ascertain the association between the study features and the extent of effect sizes (Hattie, Biggs and Purdie, 1996). \( Q_B \) is used to estimate the between-groups effect, which reveals whether the average effect size differs over groups. That is, it
indicates whether the variable is a significant moderator of outcome. $Q_B$ has an approximate $\chi^2$ distribution with df = $p - 1$, where $p$ is the number of classes (Hedges & Olkin, 1985), and should be at or below the .01 significance level. The homogeneity statistic $Q_W$ was produced to determine whether each set of $d$s had an effect size consistent across the studies. $Q_W$ allows the determination of the appropriateness of the grouping of studies for between-group analysis. $Q_W$ has an approximate $\chi^2$ distribution with df = $k - 1$, where $k$ is the number of effect sizes (Hedges & Olkin, 1985). Ideally, within-group findings ($Q_W$) should be greater than or equal to the .01 probability level.

**Multiple Regression**

As recommended by Lipsey and Wilson (2001), weighted multiple regression analyses were then conducted. Each of the coded variables of interest (e.g., focus of intervention on SC/SE) were entered into an individual regression to determine which categories within a variable were most predictive of effect size. Further, to establish predictive ability across all variables of interest, an all-inclusive regression procedure was run. Because of the potential for multicollinearity when all coding variables were included in the regression analysis, a backwards elimination process was used, eliminating non-significant predictors step-by-step, until only significant predictors remained.

**Results and Discussion**

The overall mean effect size of .26 suggests that, although only a small change, interventions can lead to SC/SE improvement. There was, however, a great deal of variability in outcomes; the effectiveness of interventions was far from uniform. The statistical analyses employed helped to identify some strategies that could lead to increased benefits of SC/SE interventions.

**Focus of Intervention on Self-Concept**

Target interventions exhibited the largest mean effect size (.70). Interventions for which self-concept was an incidental variable had the lowest effect sizes (.13), followed by interventions which aimed at increasing SC/SE through non-SC/SE programs, for example social skills training (.20). This suggests that interventions are best served by targeting SC/SE through specifically-design SC/SE interventions.

**Relevance of Self-Concept Outcome**

Mean effect sizes were substantially larger for self-concept outcomes judged to be focal (most relevant) to the intervention (.40), whereas the effect sizes were smaller for self-concept outcomes judged to be secondary (.16), and smaller still for those outcomes judged to be non-relevant to the intervention (.12). This supports the construct validity approach to SC/SE research, emphasises the need to incorporate multidimensional perspectives of SC/SE. It also endorses the use of multidimensional instruments when measuring specific domain interventions to ensure that true benefits are not masked by inappropriately global scales.

**Match Between Intervention and Self-Concept Domain**

This variable was effectively a combination of the focus of the intervention on self-concept and the relevance of the outcome domain measured. Interventions that were both targeted on SC/SE and measured specific, related domain outcomes yielded the highest effect sizes of all (.89). This is substantially larger than the .26 overall mean effect size, and suggests that appropriately constructed self-concept enhancement studies make a substantial difference in intervention success. Interventions that did not adequately match the intervention to the
outcome measures in terms of multidimensionality (e.g., a physical self-concept intervention measured using a global SC/SE scale), had substantially lower effect sizes e.g., .20 for the example given). This further lends credence to the use of multidimensional measures of SC/SE outcome, and to the use of targeted interventions when seeking self-concept improvement.

_Treatment Administrators_
Teachers were found to be the most effective change agents (.46), which is promising given that they comprised the largest group of treatment administrators. This indicates that the implementation of SC/SE development programs in the schools is promising. However, there is a temptation in meta-analyses to make unqualified generalisations, and requires prudence in interpreting the results. For example, the success of teachers in the present meta-analysis could be attributable to the fact that teachers are typically more familiar to the participants than other treatment administrator types. Thus, it could be that any treatment administrator to whom participants are accustomed is more effective. Further, Strein (1988) found that treatment administrators were typically associated with certain types of treatments, which could moderate the amount of impact attributable to teachers. Further research into the interaction between these factors is warranted.

_Rationale for the Intervention_
Within this variable, the highest mean effect sizes were for interventions based on studies with both theoretical and empirical basis to the intervention design (.54). The lowest observed effect sizes were for studies with only previous research used to inform intervention design (.09). This emphasises the interdependent relationship between theory and practice (Marsh, 1990), and suggests that future research would be astute in consulting recent developments in self-concept theory, coupled with firm empirical research, when designing intervention and evaluation programs.

_Control Group Type_
Comparisons with attention placebo controls resulted in significantly lower mean effect sizes (.26) than inactive controls (.30) or waitlist controls (.34). Because of the significant difference between observed effect sizes, aggregation of scores across groups (i.e., calculating mean effect sizes based on different types of control groups) could be deleterious, as some comparisons will over- or underestimate the true effect size. Thus, features of the methodology can mask the effectiveness of interventions, and interpretation of effect sizes should be considered in light of the impact of design characteristics (Strein, 1988; Lipsey & Wilson, 2001).

_Randomisation_
Random assignment strategies consistently produced higher mean effect sizes (mean effect size range .35 to .74) than non-random assignment procedures (mean effect size range -.61 to .32). It appears that group assignment procedure impacts upon self-concept enhancement success, typically in favour of randomised designs. This suggests that quasi-experimental studies may lead to the underestimation of intervention effects.

_Prevention Versus Treatment_
Treatment interventions yielded larger effect sizes (.40) than prevention interventions (.24). However, unlike the other variables discussed above, this was not found to be a significant predictor when included in the regression model. Haney and Durlak (1998) found this
difference to be a significant predictor, so further research on this issue is necessary to clarify this quandary.

**All-Inclusive Regression**

In the all-inclusive regression, a significant $Q_R$ value (926.84; $p < .001$) indicated that significant variability in the effect sizes was explained by the model. However, the significant $Q_E$ (5548.56, $p < .001$) suggests that the effect sizes were still heterogeneous even after the variability accounted for by the predictor variables was removed. Thus, the unexplained variability was greater than would be expected from sampling error (Lipsey & Wilson, 2001). The significant categories pertained to the following variables: focus of intervention on self-concept, self-concept outcome relevance, treatment administrator, rationale for intervention, control group, group assignment procedure, and prevention versus treatment. $R^2$ for the all-inclusive model indicated that these variables accounted for 14.31% of the variance in the model.

**Follow-Up Analyses**

Separate analyses were conducted for the follow-up testing effect sizes. Analyses were only computed for studies that reported both posttest and follow-up outcome measures – a total of 12 studies comprising 20 interventions. For these follow-up studies, the weighted effect for the difference between the posttest and follow-up was .048 ($SE = .061$, $p = .43$, not significant). Thus, the results indicate that self-concept did not systematically increase or decrease during the follow-up period following the intervention. Hence, at least for the studies considered here, the self-concept intervention effects were resilient over time.

**Limitations**

A few caveats in interpreting meta-analysis results warrant discussion. Firstly, results of any meta-analysis are limited to the sorts of studies included in the sample. For instance, age restrictions in the present sample mean that these results may not be applicable to older participants. Secondly, it is impossible from this statistical review to confer causal connections upon the variables. Causal modelling analyses would allow better causal inferences to be made (Haney & Durlak, 1998), however, this requires strong theoretical understanding in order to be effective, which can informed by meta-analysis (Rosenthal, 1984).

Thirdly, the possibility of selection bias is a potential problem in any meta-analysis (see Wolf, 1986). Publication bias – when meta-analyses use only published studies – could be a problem since published studies tend to have higher effect sizes than unpublished studies (Lipsey & Wilson, 1993). Selection bias in the present meta-analysis may also stem from the exclusion of studies that did not incorporate a control group design. Although other designs have been found to be valid (e.g., time series design; Marsh, Richards & Barnes, 1986), these were excluded from the present study. A delicate balance between practicality and comprehensiveness is essential in any meta-analysis (Durlak & Lipsey, 1991), and results are interpreted in light of these restrictions.

A final admonition is that studies under review do not always report the necessary information. For instance, many studies with non-significant results did not report any data, and were thus entered as zero. As discussed above, this could drastically underestimate the true effect size.
Implications
A number of potential research areas have been posited above. The validation of the multidimensional approach emphasises the contingent association between theory, research, and practice (Marsh, 1990). It highlights the need to use multidimensional instruments for measuring specific facet outcomes, so that meaningful improvements are not lost in the translation from multidimensional interventions to unidimensional measures. Further, the greater improvements experienced by participants in treatments focusing on specific domains suggest that, despite the obvious desirability of an all-in-one, global self-concept treatment, it is likely more beneficial for the participants to receive specific self-concept domain treatments. Hay, Burn and Butler (2000) conducted one such treatment, where each of 11 treatment sessions specifically focused on one of the 11 self-domains in Marsh’s (1990) SDQII scale. Cost-benefit analysis research into this area could determine the feasibility of replacing blanket approaches to enhancing self-concept with distinct self-concept domain interventions.

References

Note. Studies used in the meta-analysis can be obtained from the authors upon request.


