

A Comprehensive Multilevel Model Meta-Analysis of Self-Concept Interventions [R]

Alison J. O'Mara¹, Herbert W. Marsh¹, and Rhonda G. Craven²

¹*University of Western Sydney, Australia and University of Oxford, UK,*

²*University of Western Sydney, Australia*

The efficacy of self-concept interventions has previously been examined through traditional meta-analytic methods, and a host of moderators of intervention outcomes have been identified (O'Mara, Marsh, & Craven, 2004; Haney & Durlak, 1998; Hattie, 1992). However, traditional meta-analytic models have increasingly been criticized because they fail to account for the nested structure of effect sizes within studies, thereby violating statistical assumptions of independence. The multilevel model approach to meta-analysis takes into account the hierarchical structure of meta-analytic data, thus providing findings that are more statistically sound. Consequently, the present study applies the multilevel model technique to the analysis of the self-concept intervention literature. The overall mean effect size of .47 suggests a moderate impact of interventions on self-concept at post-test, and analyses show that intervention effects are maintained at follow-up. Other moderators examined include the construct validity approach to the multidimensionality of self-concept; the focus of the intervention on self-concept; the use of random assignment to treatment and control groups; the control group type; treatment type; and the treatment administrator. Intra-class correlations and the variance explained by each moderator model are presented to emphasise the importance of using a multilevel model approach to meta-analytic research. It is concluded that multilevel models provide a more accurate understanding of the self-concept intervention literature than traditional meta-analytic models. Suggestions for future self-concept intervention design and evaluation are provided.

Positive self-concept is desirable for one's development, and so self-concept enhancement is a major focus of concern for many researchers. For instance, high self-concept is associated with feelings of self-worth and high motivation to achieve. Self-concept is also an important mediating construct, which impacts upon a host of psychological and behavioural outcomes across a variety of settings. Components of self-concept have been shown to predict various performance outcomes above and beyond prior performance in the task. Some examples include the performances of elite swimmers at international events (Marsh, 2002); job performance and job satisfaction (Judge, Erez, & Bono, 2001); and academic achievement (Marsh, Byrne, & Yeung, 1999). Thus, the enhancement of self-concept is desirable, as it promotes positive outcomes (see Marsh & Craven, 2006, for review) and protects against negative outcomes (Branden, 1994).

Construct Validation and Self-concept Interventions

Despite the extensive attention paid to the outcomes associated with self-concept, the self-concept enhancement literature has suffered from adherence to outdated models of the structure of the construct. Relevant literature reviews (e.g., Marsh & Craven, 1997) have typically been unable to conclude whether interventions are effective, and this could be due to the way in which self-concept has historically been commonly measured (see O'Mara, Craven & Marsh, 2004). A fundamental problem with the self-concept enhancement literature is that, for a long time, there has existed an assumption that self-concept has a unidimensional structure, suggesting that it is consistent across different contexts (see Bracken, 1996; Hattie, 1992 for reviews). As a result, self-concept interventions have typically been evaluated using global self-concept scales. However, recent developments in self-concept theory (e.g., Marsh & Shavelson, 1985) and measurement (e.g., Marsh, 1993) have spawned research in favour of a multidimensional structure of self-concept (see Marsh, Craven & Martin, 2005, for a review; also see Byrne, 1996; Hattie, 1992). In their pivotal review of the self-concept literature, Shavelson, Hubner and Stanton (1976) 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).

Construct validation approaches have been advocated as a way of better understanding the hypothetical construct of self-concept (Marsh, 1993). Such an approach applied to intervention research would mean that the domains of self-concept that are most relevant to the intervention should be most impacted upon by the intervention, while less relevant domains should show less change (Marsh & Craven, 1997). There are two types of construct validity: convergent and discriminant validity. Convergent validity is the contrast of two variables or outcomes that one would expect to be related. When applied to self-concept interventions, we would expect self-concept domains related to the intervention to show a marked improvement after intervention. Discriminant validity is the contrast of two distinct variables or outcomes. In self-concept interventions, discriminant validity would be evidenced where unrelated self-concept domains do not improve after the intervention. 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.

However, only recently have adequate multidimensional instruments and appropriate statistical analysis procedures become available, which has heretofore meant a dearth in multidimensional self-concept research employing this approach (see Hattie, 1992; Byrne, 1984, for reviews). 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 (1986a, 1986b). 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 post-test 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 six 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 academically-oriented 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 construct validity approach 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. They reported 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, Marsh and Craven (2004) showed that multidimensional considerations are essential in determining intervention success. However, many researchers still utilise unidimensional 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., 2004).

Previous Meta-analyses of Self-concept Interventions

The current paper is a re-analysis of a previous meta-analysis by O'Mara, Marsh and Craven (2004), which used a fixed effects model method. The earlier meta-analysis was inspired by two prior self-concept intervention meta-analyses by Hattie (1992) and Haney and Durlak (1998). These three meta-analyses, and their results 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 interventions. Effect sizes in this meta-analysis were based on global self-esteem scores, implying a unidimensional approach. Hattie found a weak to moderate increase in self-concept across all interventions, as evidenced by an average effect size of 0.37. Teachers were found to be the least effective treatment administrators.

Haney and Durlak's (1998) Meta-Analysis

Haney and Durlak (1998) performed a fixed effects meta-analysis of 99 self-concept intervention studies published prior to 1992. Like Hattie (1992), Haney and Durlak used global self-esteem scales. The overall mean effect size in this analysis was 0.27, which is slightly smaller than what Hattie found. Treatments 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, $d = 0.10$). Also, studies that randomly assigned participants to treatment or control conditions yielded higher effect sizes than studies employing non-random assignment to conditions.

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

O'Mara, Marsh and Craven (2004) conducted a fixed effects meta-analysis of 145 primary studies, which reported 200 self-concept interventions for children and adolescents. They utilised a multidimensional, construct validity approach. They found an overall moderate effect size of .27, which is equal to that reported in Haney and Durlak (1998). The multidimensional structure of self-concept was supported as 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. As in the Haney and Durlak study, interventions whose focus was on self-concept enhancement had a higher mean effect size (.41) than indirect interventions (.18). Also, random group assignment strategies produced higher mean

effect sizes than non-random assignment procedures. In contrast to Hattie's (1992) findings, teachers were found to be the most effective change agents (.47).

Multilevel Modelling Meta-analysis

Historically, meta-analytic research has been dominated by a fixed effects model approach. The three previous meta-analyses of self-concept interventions described above all used this method. The fixed effects model assumes that all studies included in a meta-analysis are drawn from the same population (Hedges & Vevea, 1998). The implication is that the studies are equivalent except for sampling error variance. Realistically, though, most meta-analyses include distinct studies in their sample. For instance, the included studies might employ different study designs, intervention techniques, and/or measurement instruments. Thus, it is necessary to incorporate random error variance to account for these variations between studies (Raudenbush, 1994). This allows generalisation to studies not included in the sample (Lipsey & Wilson, 2001). A new, and increasingly popular, way of conducting meta-analysis is through applying a multilevel model approach to meta-analysis (Draper, 1995; Severiens & ten Dam, 1998; Swanborn & de Glopper, 1999). Meta-analytic databases are inherently hierarchical in nature because effect sizes are nested within studies (Goldstein, 1995; Hox, 2002; Hox & de Leeuw, 2003; Raudenbush & Bryk, 2002). The benefits of conducting a multilevel model meta-analysis have been explored by various researchers (Raudenbush & Bryk, 2002; Goldstein, Yang, Omar, Turner, & Thompson, 2000; Hox & de Leeuw, 2003; Kalaian & Raudenbush, 1996; Severiens & ten Dam, 1998; Swanborn & de Glopper, 1999; Van den Noortgate & Onghena, 2003; see also O'Mara, Marsh & Craven, 2005 for summary).

Perhaps the most critical benefit of multilevel modelling meta-analysis for the present study is that multilevel modelling circumvents the issue of independence (Kalaian & Raudenbush, 1996). Meta-analyses frequently have multiple effect sizes per study, such as when they have multiple treatment groups or multiple outcome measures. Effect sizes from the same study will be more similar to each other than to those from different studies. Thus, these effect sizes may be correlated due to their shared source (Cooper, 1998). Traditional meta-analysis assumes that the effect sizes are independent from each other, whereas multilevel modelling builds into the model the potential for multiple effect sizes per study. This technique accounts for dependencies in the data whilst also allowing each study to contribute different effect sizes (Dickerson & Kemeny, 2004), allowing a multivariate analysis of the data (Hox & de Leeuw, 2003). By modelling the nesting of levels within studies, Bateman and Jones (2003) suggest that the accuracy of the estimation of standard errors on parameter estimates, and the assessment of the significance of explanatory variables, are improved.

The current meta-analysis seeks to compare and contrast multiple self-concept outcomes, thereby making violations of independence a concern. This suggests that traditional meta-analytic methods are not appropriate for this dataset. Indeed, analyses comparing components of any multidimensional construct should employ multilevel modelling meta-analytic methods, although this is not commonly the case in published psychological research. Hence, the present study contributes to meta-analytic methodology by demonstrating a construct validity approach using multilevel modelling, which is a unique innovation.

The Present Investigation

Despite the aforementioned advances in self-concept theory and instrumentation, self-concept enhancement research has continued to use outdated, superseded global self-esteem scales (Marsh & Craven, 1997). Because of the mix of unidimensional and multidimensional approaches, the reviews of the self-concept enhancement literature may be distorted (O'Mara, et al., 2004). For instance, O'Mara, et al. suggested that interventions targeting specific self-concept domains are more effective than those targeting general or global self-esteem. Thus, a literature review including both sorts of interventions might be inconclusive. Further, they noted that many interventions that were reported represented a mismatch between the goals of the intervention and the measures used to evaluate them. For example, an intervention might directly target physical self-concept, but evaluate the intervention efficacy using a global self-esteem scale. This, they argued, probably will not tap into the actual changes at post-test as a result of the intervention because the instrument is not sensitive to the particular changes.

This multilevel model meta-analysis will compare the relevance of self-concept domain outcomes in an exploration of intervention effectiveness through the construct validity approach first implemented by O'Mara, Craven and Marsh (2003; see also O'Mara, Marsh & Craven, 2004, 2005). The present study involves a reanalysis of the data of a prior self-concept intervention meta-analysis which used a construct validity approach (O'Mara, et al., 2004), this time using multilevel modelling. As in the previous study, the aims of the present study are:

- 1) To employ the innovative multilevel modelling approach to meta-analysis using specific categorical variables. This differs from an all-inclusive model by O'Mara, et al. (2005), which included dichotomous variables into one model. The present meta-analysis will analyse each variable in a separate model. As summarised above, multilevel modelling has many advantages over fixed effects models (Hox, 2002).

- 2) To explore the multidimensionality of self-concept using a construct validity approach. Previous research has found that the multidimensionality of self-concept is strongly supported (see Marsh, et al., 2005; O'Mara, et al., 2004);
- 3) To compare direct and indirect self-concept interventions. Both Haney and Durlak (1998) and O'Mara, Marsh and Craven's (2004) meta-analyses have found direct interventions to have more impact upon self-concept at post-test;
- 4) To explore the effects of random vs. non-random assignment to treatment and control conditions. O'Mara, Marsh and Craven, (2004) found random group assignment to be a more significant predictor of effect size;
- 5) To explore the control group type as a moderator of effect size. O'Mara, Marsh and Craven (2004) found that attention placebo control groups had lower mean effect sizes than waitlist or inactive controls;
- 6) To examine whether praise/feedback methods are effective. In a meta-analysis reported in Hattie (1992), psychotherapeutic approaches to self-concept enhancement were examined. However, treatment types used in educational settings have not received much attention, even though this has obvious direct applications. In a review of the literature, Craven, Marsh and Burnett (2003) lauded the success of certain praise and feedback interventions, and so we might expect such interventions to be particularly successful;
- 7) To assess whether O'Mara, et al.'s (2004) finding that teachers were the most effective treatment administrators was supported under a multilevel model; and
- 8) To test whether effects of the interventions are maintained at follow-up testing. O'Mara et al. (2004) found that the effects found at post-test did not significantly diminish by follow-up – a result we expect to emulate here.

Method

The same sample of 145 studies (200 interventions, 460 post-test effect sizes) used in the O'Mara, et al. (2004) fixed effects model meta-analysis is used here. The interventions were for children and adolescents up to a mean age of 18 years. See the previous study for description of the sample and sampling procedures, coding materials, and computation of effect sizes.

Multilevel Modelling 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), and 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_{ij} refers to the effect size for outcome i from intervention j and study k , β_{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} indicate 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 (Raudenbush & Bryk, 2002). 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 modelled to see if they explain systematic variance between the study effect sizes (Hox & de Leeuw, 2003). For example, one would test to see if random assignment to conditions moderates the effect sizes. The question would be asked: Does the self-concept outcome for different studies differ according to assignment type (random or non-random)? This requires that the model be expanded to include predictor variables. The following hierarchical model was used:

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

where d_{ijk} is the mean effect size, $\beta_0 \dots \beta_s$ are the regression coefficients, $W_{1j} \dots W_{sj}$ 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 (Raudenbush & Bryk, 2002). The intercept (β_{000}) is the estimated effect size for a study with zero values for all moderator variables. The remaining regression weights ($\beta_0 \dots \beta_s$) 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 and Discussion

Multilevel Intercept-Only Model

A baseline variance components model was used to determine how much of the total variance was partitioned into variance components associated with study, intervention and effect size. The MLwiN output produced for this model is presented in Figure 1. The mean intercept, β_0 , constituting the fixed part of the model, is .468 with a standard error (SE) of .061. This is the indicator of the overall mean effect size for the sample of studies. In other words, the overall mean effect size is approximately .47. This indicates that the fixed effects model used by O'Mara, Marsh and Craven (2004) may have actually underestimated the impact of interventions on self-concept. Given the improved accuracy in modelling the data afforded by the multilevel model over the fixed effects model, it is promising to note this larger effect size.

The intercepts for the different studies (level 3 residuals, v_{0jk}) have a variance, $\sigma^2_{v_0}$, of .186 (SE = .085). The variance is divided by the standard error, and this new figure is compared to a z-critical at the $\alpha = .05$ level to determine whether this variance is significant. This is known as the Wald test (Hox, 2002). In this particular analysis, the ratio of the parameter estimate to its SE is greater than 1.96, meaning that the intercepts of the studies considerably differ from one another at $p < .05$. The level 2 residuals, $\sigma^2_{u_0}$, have a variance of .381 with a SE of .082. This parameter estimate is large, demonstrating that most of the variation in the effect sizes could be explained by differences at the intervention level. The random parameters at level 1 are constrained to equal 1 because the variance between effect sizes is known (see Chapter 5; Hox, 2002; Raudenbush & Bryk, 2002).

The variance components shown in Figure 1 were used to compute the intraclass correlation. The total variance in the effect sizes is the sum of the level 3, level 2 and level 1 variances (.186 + .381 + 1.000 = 1.567). The between-study variance makes up a proportion of .119 of this total variance (.186 ÷ 1.567). This is known as the intraclass correlation, which suggests that 11.9% of the total variance was at the study level. Similarly, the intra-effect size correlation can be calculated by dividing the level 1 variance by the total variance (1 ÷ 1.567 = .638). Hence, 63.8% of the total variance was at the effect size level. This suggests that there could be large differences between multiple effect sizes within a given study – an unsurprising result since the effects within each study represent different self-concept domains. Finally, using the command interface in MLwiN, a chi-square was conducted (output not shown). The results of the chi-square homogeneity test suggested significant heterogeneity in the effect sizes ($\chi^2(144) = 511.87, p < 0.001$).

Maintenance of intervention effects at follow-up

The intercept-only multilevel model using the follow-up effect sizes indicated that the overall mean effect size was .02, which was non-significant ($SE = .11, p > .05$). However, the chi-square of the residuals in the multilevel model contradicted the findings of the between-studies homogeneity of the fixed effects model meta-analytic methods, with a barely significant $\chi^2(11) = 56.342, p < .05$. The intraclass correlation for level 3 was .087, suggesting that differences between studies account for 8.7% of the variance in this model. Thus, the results of the analyses of follow-up results indicate that the impact of the interventions do not significantly diminish at delayed post-test. It should, however, be noted that only 20 interventions reported follow-up results.

Multidimensionality of Self-concept

Relevance of self-concept domain outcome

Recall that target self-concept domains are those directly relevant to the intervention and hence we would expect maximum intervention change in those domains, target-related are those self-concept outcomes for which we might expect some change, whereas non-target domains would not be expected to exhibit any change after intervention. For the fixed parameters of the multilevel model with target and target-related self-concept outcomes included as dummy predictor variables, target self-concept outcomes ($\beta = .289, SE = .044, p < .05$) yield higher effect sizes than target-related self-concept outcomes ($\beta = .214, SE = .050, p < .05$). Non-target self-concept outcomes were the reference variable. Thus, given that the intercept in this model is .26, the mean effect size for target studies is .55; for target-related studies, $d = .47$; and for non-target studies, $d = .26$.

For the random parameters, the log-likelihood of the baseline model (4160.365) minus the log-likelihood of the model with explanatory variables (4028.322) gives a deviance of 132.043 ($df = 2, p < .001$). This statistically significant deviance indicates significant variance between the models, suggesting that the model with predictors fits better than the intercept-only model. The intraclass correlation for the between-studies level was .114, suggesting that differences between studies account for 11.4% of the variance in this model. The chi-square test on the variances suggests that there is significant between-study homogeneity ($\chi^2(142) = 6679.6, p < .001$).

The results further supported the use of the construct validity approach to self-concept, as targeted self-concept outcomes yielded higher mean effect sizes than target-related or non-target self-concept outcomes. This is in line with the findings of the previous meta-analysis using the fixed effects analyses. Further, it strengthens a host of research (see Marsh, et al., 2005, for review) supporting the multidimensionality of self-concept.

Match between intervention goal and self-concept outcome measure

For the fixed part of the model, only two of the categories were significant predictors of effect size. These were the group of studies focusing on a specific self-concept domain that used a measure of that domain ($\beta = 1.030$, $SE = .317$, $p < .05$, $d = 1.17$), and the group of studies in which a specific self-concept domain was targeted but measured using multiple multidimensional scales ($\beta = .749$, $SE = .316$, $p < .05$, $d = .89$). The deviance of 19.92 was significant at $p < .05$, suggesting the model with predictors included had a better fit than the intercept-only model. The intraclass correlation for level 3 was .102, suggesting that differences between studies account for 10.2% of the variance in this model. The chi-square on the residuals was significant ($\chi^2(136) = 6748.5$, $p < .001$).

The use of multidimensional approaches to self-concept intervention was supported since the only two significant predictors of effect size in the match between intervention and measurement analyses were the category of studies focusing on a specific self-concept domain that used a measure of that domain (which, incidentally, was the largest predictor of all the categories in any of the analyses) and the group of studies in which a specific self-concept domain was targeted but measured using multiple multidimensional scales. This follows the results of the previous fixed effects model meta-analysis, and also coincides with O'Mara, Craven and Marsh's (2004) argument that interventions should target and measure the same self-concept domains.

Research Methodology

Focus of intervention on self-concept

This model contained no significant categories. This is despite the model having a statistically significant improved fit over the intercept-only model, as indicated by the deviance of 91.21 ($p < .01$). In other words, the predictor-included model explains more variance than the intercept-only model, but still fails to identify significant moderator variables. The intraclass correlation for level 3 was .135, suggesting that differences between studies account for 13.5% of the variance in this model. The chi-square of the residuals was significant, $\chi^2(141) = 7053.6$, $p < .001$.

Surprisingly, the results of the present meta-analysis did not support the findings of either Haney and Durlak (1998) or O'Mara, Marsh and Craven (2004) regarding direct interventions leading to greater self-concept enhancement. The multilevel model analyses showed this variable to be non-significant for this sample, unlike the previous meta-analyses. This could be explained by differences between the two models. Fixed effects models fail to account for random variance between the interventions and studies (Raudenbush, 1994), and so are less accurate. Although this finding does not contradict the reciprocal effects model of the casual ordering between self-concept and performance (see Marsh and Craven, 2006 for review), since both self-concept and skill-based interventions seem to be equally effective in enhancing self-concept, this does highlight the importance of conducting more causal modelling research.

Group assignment procedure

Most of the random group assignment procedures yielded higher B -values than non-randomized designs, with the exception of randomly assigned groups. However, the only categories to reach statistical significance ($p < .05$) were studies that used randomly assigned individuals ($B = .838$, $SE = .409$, $d = .58$) and those that utilized intact groups and claimed to randomly assign them to conditions ($B = 1.007$, $SE = .455$, $d = .75$). The deviance value of 97.93 is significant at $p < .01$, suggesting significant improvement in model fit from the intercept-only model. The intraclass correlation for level 3 was .139, suggesting that differences between studies account for 13.9% of the variance in this model. The chi-square on the residuals was significant, ($\chi^2(137) = 6835.2$, $p < .001$).

Haney and Durlak (1998) and O'Mara, Marsh and Craven (2004) found that random assignment of participants was a significantly positive predictor of effect size using a fixed effects model. O'Mara, Marsh and Craven (2004) reported that random group assignment procedures yielded higher effect sizes than non-random procedures. This was generally found to be the case here. However, the use of multilevel modelling found only two of the categories to be statistically significant predictors. Following from O'Mara, Marsh and Craven (2004), it seems that quasi-experimental studies may lead to the underestimation of intervention effects. We therefore reiterate that researchers should employ random designs where practicable. Further, meta-analysts of intervention studies should include random assignment in their predictor models, since it appears to make a difference to the way in which self-concept outcomes are evaluated.

Control group type

Waitlist controls yielded the highest mean effect size ($d = .57$, $B = .578$, $SE = .284$, $p < .05$). The other category to reach statistical significance was the inactive control group ($d = .48$, $B = .488$, $SE = .248$, $p < .05$). The deviance statistic of 92.53

suggests a significantly better model fit over the intercept-only model ($p < .01$). The intraclass correlation for level 3 was .121, suggesting that differences between studies account for 12.1% of the variance in this model. The chi-square on the residuals was significant ($\chi^2(141) = 6916.2, p < .001$).

As in O'Mara Marsh and Craven (2004), effect sizes from intervention studies employing attention placebo controls yielded the lowest mean effect sizes. This suggests that some of the impact of interventions in waitlist and inactive control group studies could be attributable to merely being involved in an intervention. That is, receiving some sort of attention might be enough to enhance self-concept. To fully evaluate the amount of change attributable to the intervention, all future research should include an attention placebo control group.

Treatment Characteristics

The categories for this moderator were: group counselling and/or discussion, individual counselling with or without group counselling, group counselling or discussion with skills training, practice or training, praise/feedback, physically oriented, self-concept activities, social support or environmental restructure, praise with other features, social support or environmental restructure with other features, and 'other'. None of the categories were significant at the .05 level, although the deviance statistic (99.82) was significant ($p < .01$). The intraclass correlation for level two was .218, suggesting that differences between interventions account for 21.8% of the variance in this model. The chi-square of the residuals was significant ($\chi^2(133) = 6553.8, p < .001$).

Since the overall deviance statistic was significant despite a lack of significant categories, a backwards elimination method was used to determine if any categories were significant without the influence of other included variables. After backwards elimination, the only significant variable (category of studies) remaining was those studies that used praise/feedback techniques ($\beta = .479, SE = .237, p < .05, d = .92$). The deviance statistic for this model was found to be 92.23. A smaller deviance statistic suggests a better fitting model, so the model with only praise/feedback as a predictor variable is actually better than the model with more categories included. The intraclass correlation for level two of the praise/feedback only model was .237, suggesting that differences between interventions account for 23.7% of the variance in this model (also an improvement on the larger model). The chi-square of the residuals was significant ($\chi^2(133) = 6602.5, p < .001$).

Thus, praise and/or feedback techniques were found to be significant predictors of effect size. This variable was not included in the earlier meta-analysis, but was included here because of the substantive interest of the research question. This follows from the review by Craven et al. (2003), which indicated that praise/feedback techniques are effective. Such interventions seem to therefore be an easy to implement and effective way of enhancing self-concept.

Treatment Administrator

The categories for this moderator were mental health professionals, professional trainee, teacher, school counsellor, experimenter/researcher, other non-professionals, mixed administrators, and 'unspecified'. None of the β -values for the categories of this moderator were significantly different from zero, although the highest β -value was for interventions administered by the researcher themselves ($\beta = .207, SE = .295, p > .05$). However, the deviance statistic showed that the predictor-included model was a better fit than the intercept-only model (deviance = 89.77, $p < .01$). The intraclass correlation for level 3 was .159, suggesting that differences between studies account for 15.9% of the variance in this model. The chi-square of the residuals was significant ($\chi^2(137) = 7078.7, p < .001$).

The issue of the treatment administrator remains a contentious (and controversial) issue. Hattie (1992) found that teachers were the least effective administrators; O'Mara, Marsh and Craven (2004) found that they were the most effective; while the current paper suggests that there are no significant differences between treatment administrators. It is understandable that the results differ from the Hattie findings, since the sample in that study included a wider age range, which meant that administrators such as university lecturers were included in the sample. Also, the Hattie meta-analysis focused more on psychotherapeutic approaches to intervention. Thus, different samples, leading to different categorisations of the administrators, between the Hattie and O'Mara, Marsh and Craven studies might explain the distinct results. The present study could differ from the O'Mara, Marsh and Craven paper because the present paper takes into account the variance at the intervention level. Types of intervention administrators are typically associated with different treatment types. As such, accounting for error at this level might eliminate some of the statistical noise associated with either the administrator or intervention or both. This is an especially powerful example of the importance of using multilevel modelling rather than traditional methods, as the prior fixed effects results are likely to be misleading.

Conclusion

In conclusion, multilevel modelling has supported some of the findings of O'Mara, Marsh and Craven (2004). However, the findings for some variables (i.e., focus of intervention on self-concept, group assignment procedure, treatment

administrator) differed slightly. Because of its many benefits in establishing accurate results over the fixed effects model, we are confident in this latest analysis of the self-concept enhancement literature. Further, the results can be generalised to the greater population of studies as it incorporates a random error variance component.

Once again the construct validity approach and the multidimensional structure of self-concept were supported. Now more than ever it seems difficult to deny the importance of the multidimensional approach to self-concept (Marsh et al. 2005), especially in the context of interventions (O'Mara, Craven & Marsh, 2004). Most importantly, the multilevel model approach to meta-analysis has shown itself to be an essential development in meta-analytic methodology as it takes into account the nested structure of meta-analytic data. However, the relatively low variance explained in each model suggests that future research might look at additional variables, or combine a variety of variables into more complex models.

About the Authors

Alison O'Mara is a research officer at the University of Oxford and member of the SELF Research Centre, University of Western Sydney. Her substantive research focus is on self-concept, and she employs sophisticated statistical techniques such as meta-analysis, multilevel modelling, and structural equation modelling to study this and related constructs.

Professor Herb Marsh is a Professor at the University of Oxford's Department of Educational Studies. He is the founder and former Director of the SELF Research Centre, University of Western Sydney. His established research program into self-concept, motivation, identity and related constructs, as well as his work in statistical modelling, has assured his place as one of the world's leading researchers in both the broad disciplines of Education and Psychology.

Professor Rhonda Craven is Acting Director of the SELF Research Centre, University of Western Sydney, which has been ranked 7th in the world in educational psychology. As an educational psychologist, her research focuses on large-scale quantitative research studies in educational settings.

Contact Details

Ms. Alison O'Mara
Department of Educational Studies
University of Oxford
15 Norham Gardens, OX2 6PY
United Kingdom
Email: a.omara@uws.edu.au
Phone: +44 (0) 1865 274 053

References

- Bateman, I. J. & Jones, A. P. (2003). Contrasting conventional with multi-level modelling approaches to meta-analysis: Expectation consistency in UK woodland recreation values. *Land Economics*, 79, 235-258.
- Borenstein, M., & Rothstein, H. (1999). *Comprehensive meta-analysis: A computer program for research synthesis (version 1.0.23)* [Computer software]. New Jersey: Biostat, Inc.
- Bracken, B. A. (1996). Clinical applications of a context-dependent, multidimensional model of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 463–503). New York: Wiley.
- Branden, N. (1994). *Six pillars of self-esteem*. New York: Bantam.
- Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research*, 54, 427–456.
- Byrne, B. M. (1996). On the structure of social self-concept for pre-, early, and late adolescents: A test of the Shavelson, Hubner, and Stanton (1976) model. *Journal of Personality and Social Psychology*, 70, 599–613.
- Cooper, H. (1998). *Synthesizing research: A guide for literature reviews (3rd ed.)*. Beverly Hills, CA: Sage.

- Craven, R.G., Marsh, H.W., & Burnett, P. (2003). Cracking the self-concept enhancement conundrum: a call and blueprint for the next generation of self-concept enhancement research. In H.W. Marsh, R.G. Craven, & D. McInerney (Eds.) (pp. 91-126). *International Advances in Self Research. Volume 1*. USA: Information Age Press.
- Dickerson, S. S. & Kemeny, M. E. (2004). Acute stressors and cortisol response: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130, 355-391.
- Draper, D. (1995). Inference and hierarchical modelling in the social sciences. *Journal of Educational and Behavioral Statistics*, 20, 115-147.
- Goldstein, H. (1995). *Multilevel statistical models (2nd ed)*. New York: Halsted Press.
- Goldstein, H., Yang, M., Omar, R., Turner, R., Thompson, S. (2000). Meta analysis using multilevel models with an application to the study of class size effects. *Applied Statistics*, 49, 399-412.
- Haney, P., & Durlak, J. A. (1998). Changing self-esteem in children and adolescents: A meta-analytic review. *Journal of Clinical Child Psychology*, 27, 423-433.
- Hattie, J. A. (1992). *Self-concept*. Hillsdale, NJ: Lawrence Erlbaum.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486-504.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hox, J. J. & de Leeuw, E. D. (2003). Multilevel models for meta-analysis. In S. P. Reise & N. Duan (Eds.). *Multilevel modelling: Methodological advances, issues, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Judge, T. A., Erez, A., & Bono, J. E. (1998). The power of being positive: The relation between positive self-concept and job performance. *Human Performance*, 11, 167-187.
- Kalaian, H. A. & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, 1, 227-235.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Marsh, H. W. (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels. *American Educational Research Journal*, 28, 445-480.
- Marsh, H. W. (1993). Relations between global and specific domains of self: The importance of individual importance, certainty, and ideals. *Journal of Personality and Social Psychology*, 65, 975-992.
- Marsh, H. W. (2002). A multidimensional physical self-concept: A construct validity approach to theory, measurement, and research. *Psychology, The Journal of the Hellenic Psychological Society*.
- Marsh, H. W., Barnes, J., & Richards, G. (1986a). Multidimensional self-concepts: The effect of participation in an Outward Bound program. *Journal of Personality and Social Psychology*, 45, 173-187.
- Marsh, H. W., Barnes, J., & Richards, G. (1986b). Multidimensional self-concepts: A long-term follow-up of the effect of participation in an Outward Bound program. *Personality and Social Psychology Bulletin*, 12, 475-492.
- Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist*, 34, 155-167.
- Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement and adjustment*. Orlando, FL: Academic Press.
- Marsh, H.W. & Craven, R. G. (2006). What comes first?: A reciprocal effects model of the mutually reinforcing effects of academic self-concept and achievement. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.). *International Advances in Self Research (Volume 2)*. Greenwich, CT: Information Age.

- Marsh, H. W., Craven, R. G., & Martin, A. (2005). What is the nature of self-esteem: Unidimensional and multidimensional perspectives. In M. Kernis (Ed). *Self-esteem: Issues and Answers*. Psychology Press.
- [Marsh, H. W., & Peart, N. \(1988\). Competitive and cooperative physical fitness training programs for girls: Effects on physical fitness and on multidimensional self-concepts. *Journal of Sport and Exercise Psychology*, 10, 390–407.](#)
- [Marsh, H. W., & Shavelson, R. \(1985\). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 20, 107–125.](#)
- O'Mara, A. J., Craven, R. G., & Marsh, H. W. (2004, July). Unmasking the true effects of self-concept interventions and suggested guidelines for rectification. Presented at the 3rd Biennial SELF Research Centre Conference, Max Planck Institute, Berlin, Germany.
- O'Mara, A. J., Marsh, H. W., & Craven, R. G. (2004, July). Self-concept enhancement: A meta-analysis integrating a multidimensional perspective. Presented at the 3rd Biennial SELF Research Centre Conference, Max Planck Institute, Berlin, Germany.
- [O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. \(in press\). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*.](#)
- Raudenbush, S. W. (1994). Random effects model. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). New York: Russell Sage Foundation.
- [Raudenbush, S. W. & Bryk, A. S. \(2002\). *Hierarchical linear models: Applications and data analysis methods* \(2nd ed.\). Thousand Oaks, CA: Sage.](#)
- [Schmidt, F. L. \(1992\). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47, 1173-1181.](#)
- [Severiens, S. & ten Dam, G. \(1998\). A multilevel meta-analysis of gender differences in learning orientations. *British Journal of Educational Psychology*, 68, 595-608.](#)
- [Shavelson, R.J., Hubner, J.J., & Stanton, G.C. \(1976\). Validation of construct interpretations. *Review of Educational Research*, 46, 407–441.](#)
- [Swanborn, M. S. L. & de Gloppe, K. \(1999\). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, 69, 261-285.](#)
- [Van den Noortgate, W. & Onghena, P. \(2003\). Multilevel meta-analysis: A comparison with traditional meta-analytic procedures. *Educational and Psychological Measurement*, 63, 765-790.](#)

$$\text{esadj}_{ijk} \sim N(XB, \Omega)$$

$$\text{esadj}_{ijk} = \beta_{1jk} \text{cons} + e_{0ijk} \text{StdErr}_{ijk}$$

$$\beta_{1jk} = 0.468(0.061) + v_{1k} + u_{1jk}$$

$$\begin{bmatrix} v_{1k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0.186(0.085) \end{bmatrix}$$

$$\begin{bmatrix} u_{1jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.381(0.082) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ijk} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 1.000(0.000) \end{bmatrix}$$

$$-2 * \log\text{likelihood(IGLS Deviance)} = 4072.205(460 \text{ of } 460 \text{ cases in use})$$

Figure 1. MLwiN Output of the Intercept-Only Model.

Note. In the model, esadj_{ijk} is the outcome measure for the adjusted effect size i from intervention j of the k th study. XB = fixed part of the model; Ω = covariance matrix; β_{1jk} = intercept; $e_{0ijk} \text{StdErr}_{ijk}$ = standard error term, random only at the effect size level; v_{1k} = random study effect; u_{1jk} = random intervention effect; and e_{0ijk} = random effect size effect.