Journal of Educational Psychology
2004, Vol. 96, No. 3, 518–522

0022-0663/04/$12.00 DOI: 10.1037/0022-0663.96.3.518

Why Multicollinearity Matters: A Reexamination of Relations Between Self-Efficacy, Self-Concept, and Achievement

Herbert W. Marsh and Martin Dowson
University of Western Sydney

James Pietsch and Richard Walker
The University of Sydney

Multicollinearity is a well-known general problem, but it also seriously threatens valid interpretations in structural equation models. Illustrating this problem, J. Pietsch, R. Walker, and E. Chapman (2003) found paths leading to achievement were apparently much larger for self-efficacy (.55) than self-concept (−.05), suggesting—erroneously, as the authors' reanalysis shows—that self-efficacy was a better predictor of achievement. However, because standard errors for these two paths were so huge (.25) thanks to the extremely high correlation between self-concept and self-efficacy ($r = .93$), interpretations were problematic. In a model comparison approach to this multicollinearity problem, constraining these two paths to be equal provided a better, more parsimonious fit to the data and also substantially reduced the standard errors (from .25 to .03).

Multicollinearity is a ubiquitous phenomenon that can produce strange, misleading, or uninterpretable results when a set of highly related independent variables is used to predict a dependent variable (Diamantopoulos & Siguaw, 2000; Pedhazur & Schmelkin, 1991). At least the detection and consequences—if not the resolution—of multicollinearity problems are well understood in traditional analyses of manifest (nonlatent) variables. However, the use of sophisticated statistical tools such as structural equation modeling (SEM) can mislead researchers into thinking that such well-known problems are no longer relevant. The purpose of this brief article is to illustrate how multicollinearity problems can contribute to inappropriate interpretations of SEM.

In a study of relations between self-concept, self-efficacy, and mathematics performance, Pietsch, Walker, and Chapman (2003) reported that self-concept and self-efficacy are highly related to each other. However, they then went on to compare the predictive utility of self-concept and self-efficacy in terms of achievement. Although they noted potential limitations due to these substantial correlations, they interpreted the results to mean that mathematics achievement was better predicted by mathematics self-efficacy than by mathematics self-concept. Here, we demonstrate that a more appropriate resolution of the problem of multicollinearity provides a very different interpretation of the results.

Herbert W. Marsh and Martin Dowson, Self-Concept Enhancement and Learning Facilitation (SELF) Research Centre, University of Western Sydney, Sydney, New South Wales, Australia; James Pietsch and Richard Walker, School of Development and Learning, Faculty of Education and Social Work, The University of Sydney, Sydney, New South Wales, Australia.

We thank Elaine Chapman and Ray Debus for helpful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Herbert W. Marsh, SELF Research Centre, University of Western Sydney, Bankstown Campus, Locked Bag 1797 Penrith South, Sydney, New South Wales 1797, Australia. E-mail: h.marsh@uws.edu.au

Mathematics (Domain-Specific) Results

This reanalysis focuses specifically on Pietsch et al.'s (2003) Model 7B (summarized in Table 4 and Figure 4 of their article; also see Model 1 in Figure 1 of the present investigation). The reader is referred to the original article for details concerning the sample, measures, and methods and for the covariance matrices used in the reanalysis presented here. On the basis of preliminary analyses, Pietsch et al. focused on self-concept items that emphasized self-perceptions of competency, and these are the focus of the present reanalysis as well. Our reanalysis replicated the standardized parameter estimates in Model 7B from the original article and also included unstandardized parameter estimates and standard errors that were not included in the published article (see Figure 1). A superficial evaluation of the path coefficients (see Figure 1) apparently supports the conclusion that self-efficacy is highly predictive of performance (standardized path coefficient = .55, $p < .05$), whereas math self-concept is not (standardized path coefficient = −.05, *ns*).

Is the path leading from self-concept to achievement really different from the path leading from self-efficacy to achievement? The observation that one path is statistically significant and the other is not, is not a sufficient basis for saying that the two paths are significantly different from each other. The appropriate test of this interpretation is to compare one model in which the two paths are allowed to differ with a second model in which the two paths are constrained to be equal. Although the apparently large difference between the two paths makes it plausible that they would be significantly different, a critical evaluation reveals some worrisome features. First, the extremely large correlation between the latent self-concept and self-efficacy constructs ($r = .93$) signals that multicollinearity is a potentially serious problem in the model. Second, the standard error for the self-efficacy-to-achievement path coefficient (.25) is extremely large. Thus, the 95% confidence interval for this coefficient is extremely wide (1.02 to 0.02). Similarly, the standard error for the self-concept-to-achievement

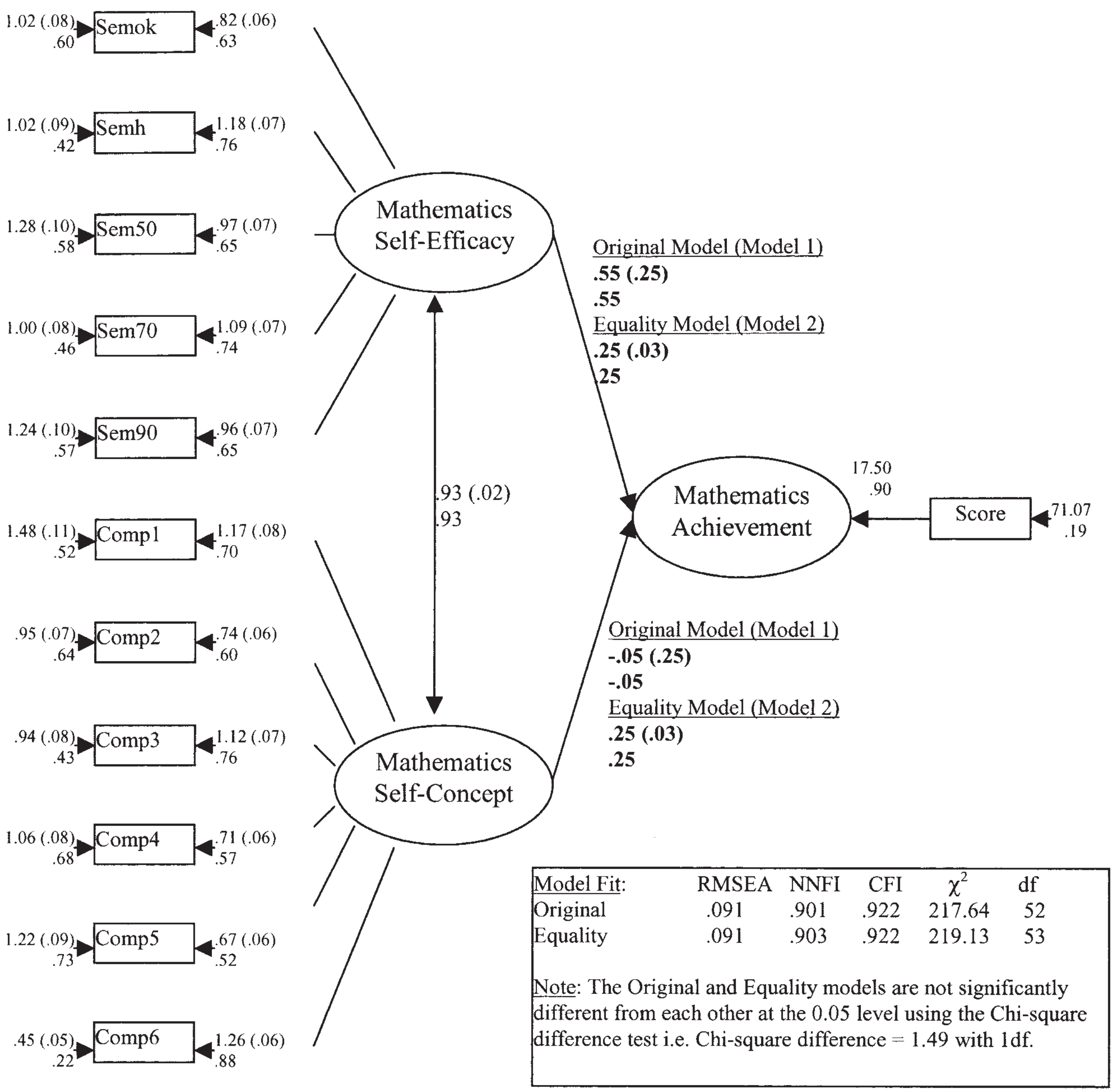

Figure 1. Models 1 and 2: Mathematics self-efficacy, self-concept, and achievement. For all parameters in the models, there are two estimates: the unstandardized estimates (with standard errors in parentheses) and the standardized estimates (with no standard errors). With the exception of the boldface figures, these estimates are essentially identical between the original and equality constrained models (also see Figure 4 in Pietsch, Walker, & Chapman, 2003, p. 600). So, for the sake of clarity in the figure, only one set of estimates (those for the original model) is presented. For definitions of items in boxes, see Appendix A of Pietsch et al. (2003, p. 602). RMSEA = root-mean-square error of approximation; NNFI = nonnormed fit index; CFI = comparative fit index; df = degrees of freedom. Adapted from "The Relationship Among Self-Concept, Self-Efficacy, and Performance in Mathematics During Secondary School," by J. Pietsch, R. Walker, and E. Chapman, 2003, *Journal of Educational Psychology, 95,* p. 600, Table 4. Copyright 2003 by the American Psychological Association.

path coefficient is very large (.25), and the corresponding confidence interval is also very wide (−.55 to .44). Because these confidence intervals are so wide, there is no sound basis for claiming from the present data that either self-concept or self-efficacy is a better predictor of performance. (Pietsch et al., 2003, did not provide standard errors for the paths, so this problem was not readily apparent.) More generally, excessively large standard errors typically indicate poor models and parameter estimates that cannot be appropriately estimated from the data (Jöreskog & Sörbom, 1993). Marsh (1989) demonstrated that excessively large standard errors might be indicative of a poor model because of an unstable solution and problems in the interpretation of the results, even when the model converges to a proper solution, provides an otherwise reasonable fit to the data, and results in apparently interpretable parameter estimates.

To test formally the statistical significance of the difference between the two path coefficients and perhaps resolve the issue of multicollinearity, we posited an alternative Model 2. In this model, the self-efficacy-to-achievement path was constrained to be equal to the self-concept-to-achievement path. This model comparison approach is appropriate when self-efficacy and self-concept factors are scaled to have a variance of 1 so that their effects are in relation to a common (standardized) metric. If the two paths really do differ significantly, then Model 1 should fit the data significantly better than Model 2. Because Model 2 is nested under Model 1, the chi-square test statistic for Model 2 cannot be any better than that of Model 1. However, if the fit of Model 2 approaches that of Model 1, then self-efficacy and self-concept do not differ in their contribution to the prediction of math achievement. The results (see Figure 1) show that the difference in chi-squares for the two models ($\Delta\chi^2 = 1.49$) is not statistically significant ($p > .05$) in relation to the difference in degrees of freedom ($\Delta df = 1$). Also, the fit statistics that take into account model parsimony (root-mean-square error of approximation [RMSEA] and nonnormed fit index [NNFI]) indicate that the fit of Model 2 is as good as, or even better than, that of Model 1. The correlation between the self-concept and self-efficacy factors is the same (.93) in the two models. However, it is very important to note that the standard errors for the path coefficients leading to mathematics achievement in Model 2 are much smaller (by a factor of almost 10) in comparison to Model 1, that is, .03 versus .25 for self-efficacy and .03 versus .25 for self-concept. Hence, at least in relation to providing reasonable standard errors, Model 2 has resolved the multicollinearity problem associated with Model 1.

In summary, on the basis of model parsimony, tests of statistical significance, fit indexes, the sizes of the standard errors, and appropriateness of interpretations, we argue that Model 2 is clearly better than Model 1. This, in turn, implies that Model 1, with its attendant problems of multicollinearity, should not be accepted as the best interpretation of the data. It is important to note that, contrary to what might be concluded on the basis of a cursory inspection of the original model, the results provide no basis for claiming that self-efficacy is a better predictor of mathematics achievement than is self-concept.

Percentages (Topic-Specific) Results

In addition to responses based on mathematics performance in general (domain-specific responses), Pietsch et al. (2003) also considered a largely parallel data set based on performance in the specific area of percentages (topic-specific responses) in which the term *mathematics* was replaced with the term *percentages*. To evaluate the generalizability of conclusions based on mathematics performance, we tested two largely parallel models (Models 3 and 4 in Figure 2) for the second set of data based on topic-specific responses to percentages. (Note: Pietsch et al., 2003, inferred percentages-competence self-concept on the basis of only five items, rather than six items, as in their mathematics-competence self-concept scale, and also considered other constructs not considered here.)

Interestingly, for the percentages data, a superficial interpretation of the results of Model 3 (see Figure 2) might suggest that topic-specific self-concept (path coefficient $= .58$, $p < .01$) is better able to predict achievement than topic-specific self-efficacy (path coefficient $= .10$, $p > .05$). Again, however, the standard errors are substantial for both self-efficacy (.17) and self-concept (.17), and the correlation between self-concept and self-efficacy is again very high (.92). For this reason, we argue that the parameters in Model 3 should be interpreted with great caution. In contrast, Model 4 (in which self-concept-to-achievement and self-efficacy-to-achievement path coefficients are constrained to be equal) is able to fit the data as well as, and even better than, Model 3, according to the NNFI. As above, we also found a remarkable reduction in the size of the standard errors associated with the paths to achievement in Models 3 and 4 (see Figure 2). Finally, there was only a very small (nonsignificant) chi-square difference between the models ($\Delta\chi^2 = 1.93$ with $\Delta df = 1$). Thus, as above, we argue that Model 4 provides a better summary of the topic-specific (percentage) responses than does Model 3. The important point again, however, is that the results provide no basis for claiming that either topic-specific self-efficacy or topic-specific self-concept is a better predictor of percentages achievement.

Conclusion

Let users of confirmatory factor analysis (CFA) and SEM beware. Whereas CFA and SEM provide extremely powerful tools to aid educational psychology researchers, many problems, such as multicollinearity, that are well known in traditional analyses of manifest (nonlatent) variables do not disappear when using more advanced latent-variable modeling techniques. These problems can render the outcomes of such analyses uninterpretable and lead to erroneous conclusions. For this reason, it is essential for users to pursue additional models to resolve these issues, ensuring that superficial examinations of structurally problematic models do not support inappropriate interpretations and propagate spurious substantive conclusions. More generally, within a construct-validity perspective, the role of researcher should be that of a skilled data detective who follows many alternative leads (Marsh, Byrne, & Yeung, 1999; Marsh & Yeung, 1997). Researchers, like detectives, should develop appropriate tests of plausible counterinterpretations of their conclusions, pursue these tests as part of an ongoing research program, and make a case for the most defensible interpretations. In defending their final models, researchers should provide a sufficiently clear audit trail, starting with a priori models, to allow the reader to evaluate the appropriateness of their alternative models and conclusions. Although discouraging to any remaining pure logical positivists, this approach to SEM more fully recognizes its role as a tool to aid substantive research, as an art form that cannot be completely codified, and as a construct-validity approach to the evaluation of interpretations of data.

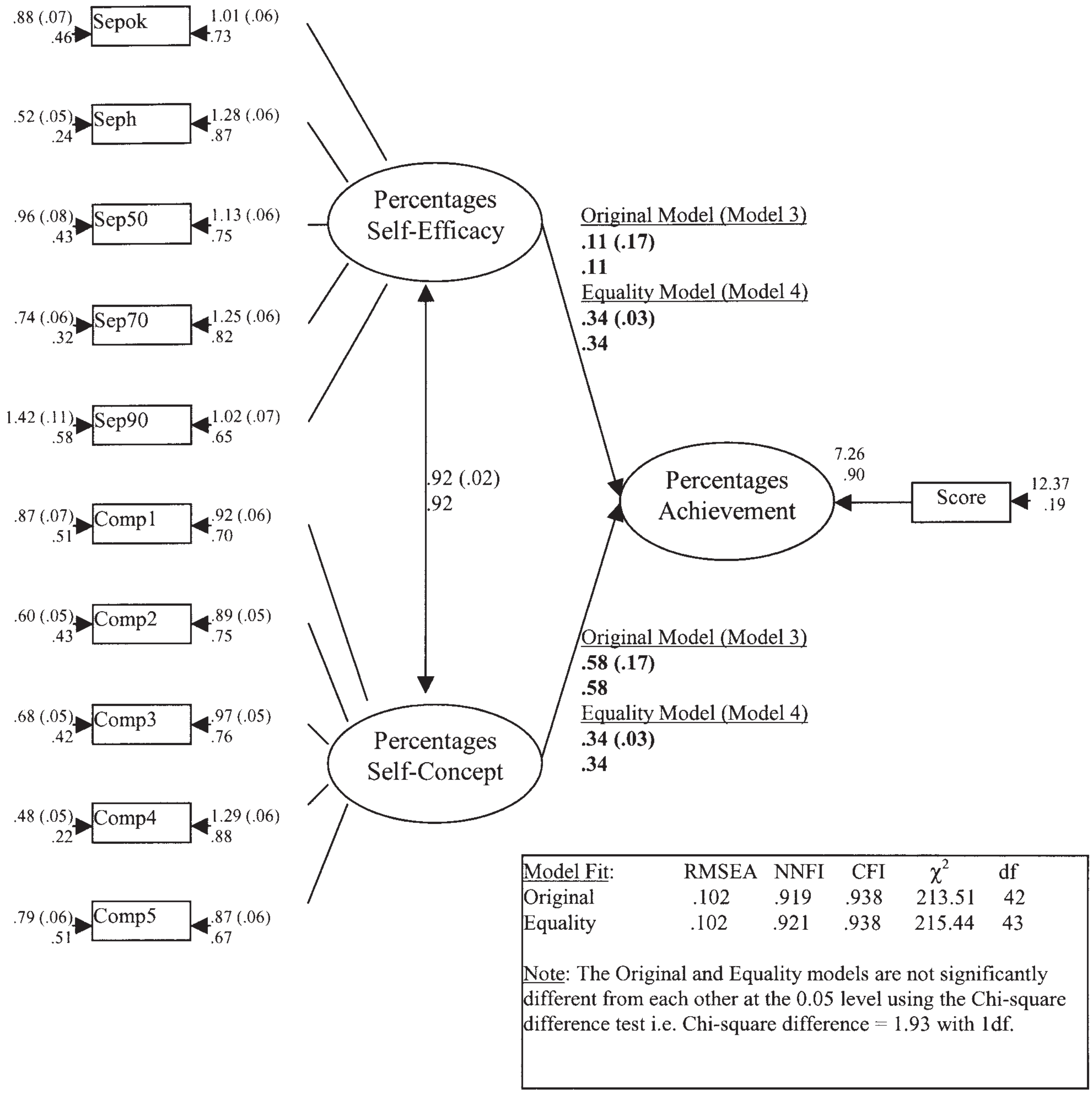

Figure 2. Models 3 and 4: Percentages self-efficacy, self-concept, and achievement. For all parameters in the models, there are two estimates: the unstandardized estimates (with standard errors in parentheses) and the standardized estimates (with no standard errors). With the exception of the boldface figures, these estimates are essentially identical between the original and equality constrained models (also see Figure 4 in Pietsch, Walker, & Chapman, 2003, p. 600). So, for the sake of clarity in the figure, only one set of estimates (those for the original model) is presented. For definitions of items in boxes, see Appendix A of Pietsch et al. (2003, p. 602). RMSEA = root-mean-square error of approximation; NNFI = nonnormed fit index; CFI = comparative fit index; df = degrees of freedom. Adapted from "The Relationship Among Self-Concept, Self-Efficacy, and Performance in Mathematics During Secondary School," by J. Pietsch, R. Walker, and E. Chapman, 2003, *Journal of Educational Psychology, 95,* p. 600, Table 4. Copyright 2003 by the American Psychological Association.

References

Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL*. London: SAGE.

Jöreskog , K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait–multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13,* 335–361.

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., & Yeung, A. S. (1997). Organization of children's academic self-perceptions: Reanalysis and counter-interpretations of confirmatory factor analysis results. *Journal of Educational Psychology, 89,* 752–759.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Erlbaum.

Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology, 95,* 589–603.

Received December 9, 2003
Revision received March 2, 2004
Accepted March 4, 2004 ■