Individually Weighted-Average Models: Testing a Taxonomic SEM Approach Across Different Multidimensional/Global Constructs Because the Weights "Don't Make No Nevermind"

Abstract

From the time of William James, psychologists have posited individually importance-weighted-average models (IWAMs) in which weighting specific attributes by individual measures of importance improves prediction of the global outcome measures. Because IWAMs cause much confusion, we briefly review a general taxonomic paradigm and structural equation models for testing IWAMs, and demonstrate its application for 2 simulated and 3 diverse "real" data applications (multidimensional measures of self-concept, quality of life, and job satisfaction). Consistent across the real data applications and previous research more generally, there is surprisingly little support for IWAMs when tested appropriately. In these diverse tests of IWAMs we integrate new approaches such as exploratory structural equation modeling (SEM), alternative approaches to constructing latent interactions, application of bifactor models, modeling method and item-wording effects, and the juxtaposition of model evaluation in relation to goodness of fit (typically used in SEM studies) and variance explained (typically used in multiple regression tests of IWAMs).

Keywords: exploratory structural equation modeling, individually weighted-average models, latent interaction effects, multidimensional and global constructs

Individually Weighted-Average Models: Testing a Taxonomic SEM Approach Across Different Multidimensional/Global Constructs Because the Weights "Don't Make No Nevermind"

This study is a substantive-methodological synergy in which new and evolving structural equation modeling (SEM) methodology is used to address substantively important issues (Marsh & Hau, 2007). The particular focus is on stronger SEM tests of individually importance-weighted-average models (IWAMs) in which weighting specific attributes by individual participant ratings of the importance of each attribute improves prediction of the global outcome measures. We begin by providing an overview of apparently similar methodological issues related to testing IWAMs that have independently been identified in diverse research literatures with surprisingly little cross-citation in relation to these problems. In the research reviewed here, there is little support for IWAMs, reminiscent of the classic review by Wainer (1976) in which he concluded that humans are so bad at differential weighting variables that it is best to ignore the weights because they "don't make no nevermind" (colloquially meaning "it makes no difference"; Wainer, 1976, p. 213)—a phrase that we also use in the title of this article as a tribute to Wainer. We then present a broadly applicable methodological approach to problems identified in each of these disciplines based on evolving SEM approaches to testing latent-variable interactions and demonstrate its application in four studies. The first is a simulation study using a didactic approach to illustrate this approach with clear documentation that makes it easy for applied researchers to apply the approach to their own research. We then analyze "real" data applications that are realistically complex, demonstrating variations in the application of our IWAM approach from three diverse research literatures. Although each of the real data applications is substantively meaningful in its own right, our focus in on the methodological issues raised by each of these applications that are broadly relevant to other research and on how they can be addressed within the context of our latent variable approach to IWAMs. Dating back at least to James (1963), psychologists and applied social science researchers more generally have posited weighted-average models for data analysis. In the general paradigm, a set of specific attributes (e.g., multiple domains of self-concept, quality of life, job satisfaction) are related to one or more global outcome measures (e.g., global self-esteem, life satisfaction, overall job satisfaction). According to the individually weighted-average model (IWAM), weighting each of the specific components by the importance placed on the component by each individual will provide a better prediction of the global outcome measure than if the weight assigned to each component is the same across all individuals. Despite the apparent simplicity

of this model, it has caused much confusion in how to test the model and what constitutes support for it. Our purpose is to briefly review a general paradigm and SEMs for testing IWAMs, demonstrate its application with simulated data, and then illustrate its versatility in three diverse real data applications.

INDIVIDUALLY WEIGHTED-AVERAGE MODELS IN SELF-CONCEPT RESEARCH

We begin with a brief review of the use of of IWAMs in self-concept research. This research is relevant in that the IWAM has a particularly long history and has been the basis of many studies, critical reanalyses, and highly contested debates that still have not been completely resolved in relation to the substantive literature. More generally, this research literature is indicative of the surprising complexity involved in tests of IWAMs in other areas of research. In self-concept research (Marsh, 2008) IWAMs are widely attributed to James (1963), who proposed that the best representation of a person's overall self-evaluation is an appropriately weighted average of self-evaluations in specific domains. James noted that because a person cannot be all things, each individual must select carefully "the strongest, truest, deepest self on which to stake his salvation" (p. 310), so that "I, who for the time have staked my all on being a psychologist, am mortified if others know much more psychology than I. But I am contented to wallow in the grossest ignorance of Greek" (p. 310). Objective accomplishments are evaluated in relation to internal frames of reference so that "we have the paradox of a man shamed to death because he is only the second pugilist or the second oarsman in the world. Yonder puny fellow, however, whom everyone can beat, suffers no chagrin about it, for he has long ago abandoned the attempt to 'carry that line'" (p. 310). Putting the two ideas together, James concluded that our sense of self "depends entirely on what we back ourselves to be and do" (p. 310). Despite this long history, the lack of empirical support for IWAMs, as embodied in the Jamesian perspective, was highlighted in the extended debate that largely appeared in the Journal of Personality and Social Psychology between Marsh (1986, 1993, 1994, 1995, 1996; Marsh & Sonstroem, 1995; also see Marsh, 2007) and Pelham (Pelham, 1991, 1993, 1995a, 1995b; Pelham & Swan, 1989), which included multiple original studies, reanalyses, responses, and counterresponses using diverse data sets collected by both researchers. When rigorously evaluated with appropriate statistical tools, individually importance-weighted averages tend to predict self-esteem less well— certainly no better—than averages of self-concepts in specific domains that simply ignore

importance ratings. Marsh (1995) concluded that even though there were lingering areas of disagreement, both he and Pelham agreed that support for the Jamesian perspective and the individual importance hypothesis (the basis of the IWAMs presented here) was surprisingly weak. Similarly, Pelham (1995b, p. 1165) acknowledged that if "James were around today, I suspect that he might feel that it has been embarrassingly difficult for us to uncover support for one of his simplest psychological insights" Hardy and Moriarty's (2006) review similarly concluded that support for IWAMs remained elusive.

Nevertheless, despite the apparent lack of empirical support for this Jamesian perspective as embodied in IWAMs, there is a dramatic disjuncture between the accepted psychological wisdom of many leading selfesteem researchers and actual research findings. Thus, in Kernis's (2006) monograph Self-Esteem: Issues and Answers, some of the world's leading self-esteem researchers cited some version of this Jamesian perspective as a well-established psychological principle without considering dissenting evidence (e.g., Harter, 2006; Mruk, 2006; O'Brien, Bartoletti, & Leitzel, 2006; Owens & McDavitt, 2006; Rhodewalt, 2006; Showers & ZeiglerHill, 2006; Tevendale & Dubois, 2006; Vonk, 2006). Indeed, within self-esteem research circles, the Jamesian perspective continues to be widely cited as a well-established psychological principle, one that has a solid theoretical and empirical basis and has withstood the test of time for more than a century. In dramatic contrast, there is little rigorous empirical support for this widely held assumption, and apparently none that suggests that it is either strong or robust (see Marsh & Hattie, 1996). Thus, Hattie (2003) concluded that the logic of this Jamesian perspective is so intuitively compelling that it has "been one of the more enduring claims in the psychological literature" (p. 137) even though there is little empirical support for it.

As repeatedly emphasized by Marsh and colleagues (Marsh, 1993, 1994, 1995, 2008; Scalas, Marsh, Nagengast, & Morin, 2013; Scalas, Marsh, Vispoel, Morin, & Wen, 2017; Scalas, Morin, Marsh, & Nagengast, 2014) in the self-concept literature, the apparent problem has been in the failure to specify an appropriate statistical model with which to test the theoretical predictions. Indeed, there have been several examples in publications claiming to support the individual weighted-average model that were shown to provide little if any support when the appropriate statistical model is applied (see Marsh, 1995, 1996, 2008; Scalas et al., 2013). Based on this ongoing research in relation to self-concept theory, a more appropriate statistical model has evolved that provides clearer tests of the model and clarity to potential areas of confusion. In this investigation, we more fully develop a taxonomic approach to testing IWAMs based on simulated data, briefly review how apparently similar

problems have arisen in diverse areas of research where IWAMs have been applied, and demonstrate its versatility in three diverse real data applications (multiple and global measures of self-concept, quality of life, and job satisfaction).

INDIVIDUALLY WEIGHTED-AVERAGE MODEL: A GENERAL PARADIGMATIC APPROACH

Historically the IWAM has been applied to manifest measures; either single-item ratings or scale scores representing each of the specific components, their importance, and the global outcomes they are designed to predict. More recently, stronger latent variable models are used in which some or all of the components are represented by latent variables based on multiple indicators. In the latent variable version of the paradigm model (Figure 1) there are three specific domains, each based on multiple indicators. Critical components are as follows:

- Actuals (act1, act2, and act3 in the upper left corner), which represent, for example, ratings of the multiple domains of a multidimensional construct. In self-concept research these might be academic, physical, or social self-concept, but could also represent other areas of research; dimensions of job satisfaction (e.g., pay, relations with colleagues, working conditions), or quality of life (e.g., work, family, leisure).
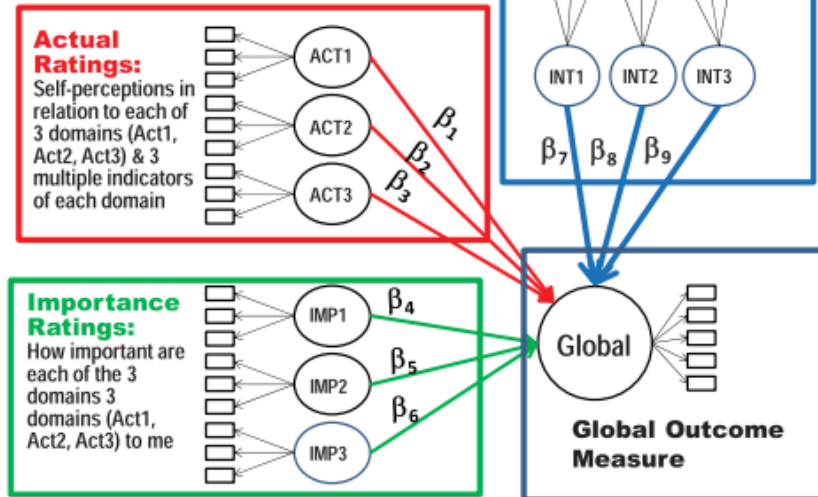
FIGURE 1 General linear structural equation model for testing individually weighted-average model. Actual ratings in different applications could refer to multiple domains of self-concept, multiple domains of job satisfaction, multiple domains of quality of life, or self-perceptions of multiple domains in other areas for which individuals also rate the importance of each of the domains. Although not presented here, correlations within each of the three sets of predictor factors (i.e., covariances among Act1, Act2 and Act3) are freely estimated.

- Importance (imp1, imp2, imp3 in the lower left corner), which represent ratings of the importance that each individual places on each domain (e.g., how important is physical competence to you).

- Actual-by-importance interactions (int1, int2, int3 in the upper right corner) representing the multiplicative combination of the actual and importance rating for each domain. These test the critical assumption of the IWAM, the effect of an act domain is moderated by its importance. For example, the IWAM predicts that if an individual perceives the physical domain of self-concept to be most important, then the physical domain should contribute more positively to the prediction of global self-esteem than do other domains—that the interaction is statistically significant and positive. In relation to a simple-slopes perspective (Aiken & West, 1991), the slope of the regression line relating physical self-concept to global self-esteem is significantly steeper for individuals who perceive the physical domain to be more important.

- Global outcome (in the lower right corner) and the path coefficients ($\beta6$–$\beta9$) relating each of the latent variables (actuals, importance, and interactions) to the global measure. The critical assumption is that at least some of the paths leading from the latent interaction terms to the global outcome ($\beta6$–$\beta9$) are positive, statistically significant, and sufficiently large to be substantively meaningful. In particular, even if statistically significant, a

6

negative path (assuming that the domain is positively oriented) provides clear support against the IWAM prediction—perhaps even more negative than if the path were nonsignificant.

Of course, this paradigmatic model is highly flexible in terms of the number of domains, the number and nature of the global outcomes, nature of the measurement model underpinning it, and the inclusion of additional variables. However, even this simple model provides a good starting point for illustrating the confusion that has resulted from IWAMs.

The starting point for IWAMs is a well-fitting measurement model. Indeed, many of the problems in the application of IWAMs (particularly when based on manifest variables in multiple regression analyses rather than latent variables in SEM analyses) stem from the failure of the underlying measurement model. Thus, for example, if the multiple indicators of the multiple specific domains do not accurately reflect the constructs they are designed to measure, or the multiple constructs are so highly correlated they cannot be adequately distinguished, then it makes little sense to apply the full IWAM. Thus the famous philosopher and storyteller, Mark Twain, is reputed to have said, "The thirteenth stroke of the clock is not only false of itself, but casts grave doubt on the credibility of the preceding twelve" (see https://en.wikipedia.org/wiki/Thirteenth_stroke_of_ the_clock). For now, let us assume that there is a well-fitting measurement model in relation to conventional indexes of fit, and that the various components can be adequately distinguished from one another.

A typical starting point is to evaluate how well the actuals are able to predict the global outcome measure ($\beta 1$–$\beta 3$ in Figure 1). A naive interpretation might be that there is support for the weighted-average model if these paths explain a significant amount of variance in the global outcome and are significantly different from each other (i.e., the model with $\beta 1 = \beta 2 = \beta 3$ can be rejected). However, even this simple starting point is fraught with interpretational difficulties that are the basis for much of the confusion in the application of this model.

Most importantly, the $\beta 1$ through $\beta 3$ paths in Figure 1 are largely irrelevant to tests of the IWAM. In particular, they represent normative differences in the relation between each of the specific domains and the global outcomes that generalize across individuals in the group being tested. A finding that $\beta 1 > \beta 2 > \beta 3$ does support weighting the components differently at the normative group level, but not at the individual level. To clarify this critical issue, assume that the first domain (Act 1) is seen as most important by the group as a whole in terms of predicting self-esteem ($\beta 1 > \beta 2 > \beta 3$). However, all individuals are assigned the

same higher value for β1 and the same lower value for β2 and β3— independent of their individual importance ratings. For example, even if a particular individual rated Act 3 as most important and Act 1 as least important, this individual would still be assigned the same values (β1 > β2 > β3) as everyone else. In contrast, the IWAM would require that the importance for this individual should be higher for Act 3 than Act 1, even though this was different for the group as a whole. To clarify this issue, Marsh and colleagues (Marsh et al., 2008; Scalas et al., 2013; Scalas, Morin et al., 2014) referred to this as the normatively weighted-average model (WAM) to distinguish it from IWAMs. Thus, in the normative WAM the weights can differ for the different domains, but not according to different individuals. The different weights for each domain can be determined a priori on the basis of theory or design, or empirically based on information from the domain ratings using techniques such as factor analysis or a multiple regression that determines an empirically optimal set of weights in relation to global outcomes. Support for any of these normative WAMs would argue for the usefulness of a normative WAM in which every individual had the same (total-group or normative) weight for any particular domain, but not an IWAM.

Support for IWAMs requires that for individuals who place more importance on the first domain than do other individuals (i.e., higher values of imp1 in Figure 1), the contribution of the first domain (act1 in Figure 1) should be greater—positive if act1 is high and negative if act1 is low. In terms of the paradigmatic model, support for the IWAM requires that the actual-by-importance interactions are significantly positive and make a substantively meaningful contribution to variance explained in the global outcome. The IWAM is not supported if the set of interaction effects does not contribute to variance explained in the outcomes, or if the direction of the interaction effects is not positive (assuming that all the constructs are positively oriented). Operationalizing Latent Interaction Effects A complication in Figure 1 is the estimation of latent interactions that are critical in testing IWAMs. The estimation and interpretation of interaction effects is an important concern in psychology and the social sciences more generally, but particularly in relation to latent variables this remains an area in which best practice is still evolving (see overview by Marsh, Hau, Wen, Nagengast, & Morin, 2013; Marsh, Wen, & Hau, 2006; Marsh, Wen, Nagengast, & Hau, 2012). Although alternative approaches are available, there are two broad approaches that have been used to test IWAMs and latent interactions more generally, the product-indictor approaches (Jöreskog, 1998; Kenny & Judd, 1984; Marsh, Wen, & Hau, 2004), and the latent moderated structural equation modeling (LMS) approach (Klein & Moosbrugger, 2000; Klein & Muthén, 2007). In this investigation we used the unconstrained

product-indicator approach (Figure 1; see also Supplemental Appendix 1D, for present parallel analyses based on LMS). Nevertheless, there are complications involved with both approaches. One important limitation of the LMS approach is that it is only readily available with the Mplus package and specialized software. In contrast, the product-indicator approach is easily implemented with any standard SEM software. Also, the LMS approach as operationalized in Mplus is very numerically intensive and might not be realistic when there are more than three or four latent interactions in the same model (depending on the number of domains, the IWAM might involve as many as eight or more latent interactions—one for each domain). Furthermore, the LMS approach does not have a properly defined null model and thus does not provide general fit statistics (although information criteria such as Akaike's information criterion can be used to compare models). Nevertheless, the LMS approach overcomes many of the limitations in how indicators are formed in the product-indicator approach. In the product-indicator approach the latent interaction factors are constructed and identified with product indicators. Each product indicator is a cross-product of actual and importance indicators. In Figure 1, for example, there are three indicators of act1 and three indicators of imp1. These are used to define three interaction indicators such that int1i = act1i × imp1i where i varies from one to three when there are three indicators. This is reasonable when the act indicators are matched to the imp indicators as in this example. However, when the number of imp indicators differs from the number of act indicators, there is no completely unambiguous way to match imp and act to form latent variable indicators. Although evolving best practice has provided a number of heuristic strategies on how to form product indicators (Marsh et al., 2012; Wu, Wen, Marsh, & Hau, 2013), this remains a complication that depends on the nature of the data available (see Supplemental Appendix 1D and 1G for further discussion).

OPERATIONALIZING TESTS OF THE INDIVIDUALLY WEIGHTED-AVERAGE MODEL

In operationalizing tests of IWAMs we emphasize both goodness of fit and variance explained in outcomes. Initially, goodness of fit is important in terms of establishing a baseline measurement model in which all factors are merely correlated. Indeed, constructing a good-fitting, theoretically defensible measurement model is a critical—but often neglected—step in testing IWAMs. Although not a problem for these simulated data sets for which the measurement model is well-defined, this is likely to be a critical issue in substantive applications based on real data (see subsequent discussion). Given a well-defined, baseline measurement model (Model 0 in Table 1), we then test a general taxonomy in which

we evaluate models (Models 1–7 in Table 1) based on different combinations of the three sets of path coefficients leading to the global outcome (β1–β9 in Figure 1). Although there are many ways in which these tests could be operationalized, we offer several strategies that we suggest would be useful for applied researchers.

Appropriate Standardized Solution in Product-Indicator Models

In multiple regression analyses, standardized effects based on standardized variables (with M = 0, SD = 1) facilitate the comparison of effects, particularly when the original scales are based on different metrics. Similarly in SEM, it is customary to report completely standardized solutions in which both latent factors and observed variables are standardized. However, the appropriate standardized solution for an interaction model is not typically provided by commercial SEM packages. Indeed, the so-called standardized solutions that are provided are typically wrong and should not be used. Although there are a number of strategies to overcome this issue (see Marsh et al., 2013; Marsh et al., 2012; Wen, Marsh, & Hau, 2010), the approach used here is to begin with standardized variables for all indicators, and then follow these steps:

1. Define latent interaction indicators as the cross-product of the actual and importance indicators such that int1i = Zact1i × Zimp1i, where the i indicators of each construct (the boxes in Figure 1) are standardized, but the interaction factor is not restandardized.

2. Fit the measurement model (Model 0 in Table 1) in which the metric is established by fixing latent variances to 1.0 (rather than fixing one of the factor loadings to an arbitrary value such as 1.0). This will transform the unstandardized solution into a standardized solution.

3. In subsequent SEMs (Models 1–7 in Table 1), use the factor loadings from this Model 0 as the starting values (these are conveniently saved by Mplus, but values from the output can be used as well). For each latent factor (e.g., the 10 latent factors in Figure 1 representing actual, importance, interaction, and global factors), fix one of the factor loadings (conveniently the largest) to values from the measurement model (not to 1.0 as is typical), and freely the estimate of the latent variance term. The results for the unstandardized solution using this approach will result in appropriately standardized parameter estimates. Wen et al. (2010) showed that traditional standard errors based on this approach closely approximated those obtained from a bootstrap approach, suggesting that tests of significance of parameter estimates using this approach were trustworthy. We also

note this approach to standardization is often useful even when there are no latent interactions in the model. (This approach is illustrated with simulated and real data analyses and syntax for each study presented in the Supplemental Materials.)

TABLE 1
Taxonomy of Models Used to Test the Individually Weighted-Average Model (IWAM)

Model	Actual (Act)			Importance (Imp)			Actual-by-Importance Interaction (Int)		
	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
Model 0 measurement	0	0	0	0	0	0	0	0	0
Model 1 full model	Free	Free	Free	Free	Free	Free	Free	Free	Free
Model 2 act only	Free	Free	Free	0	0	0	0	0	0
Model 3 imp only	0	0	0	Free	Free	Free	0	0	0
Model 4 int only	0	0	0	0	0	0	Free	Free	Free
Model 5 act & imp	Free	Free	Free	Free	Free	Free	0	0	0
Model 6 act & int	Free	Free	Free	0	0	0	Free	Free	Free
Model 7 imp & int	0	0	0	Free	Free	Free	Free	Free	Free

Note. Depicted is the pattern of path coefficients relating each of the nine predictor variables to the global outcome factor (see Figure 1); each coefficient is freely estimated or constrained to be zero. Model 0 is a measurement model in which all latent factors are merely correlated (i.e., there are no path coefficients). Model 1 is the full model in which all the path coefficients are freely estimated. In each of the other models, one or two sets of the three sets of path coefficients are constrained to be zero. Comparison of the different models provides estimates of the total and unique effects that can be explained by each set of coefficients, as well as the change in fit associated with the various sets of constraints. Thus, for example, Model 2 provides estimates of the total effects associated with the set of three actual factors (including variance that is shared with the other two sets), whereas the difference between Models 1 and 7 provides an estimate of the unique effects associated with the actual factors (i.e., variance that cannot also be explained by variance associated with the other two sets). This emphasis on total and unique components of variance explained is typical in multiple regression analyses based on manifest variables, but strangely is often ignored in SEM.

Taxonomy of Models Used to Test IWAMs

In a preface to the introduction of this taxonomy of IWAMs, it is relevant to distinguish between total and unique variance explained by different sets of predictors, and to juxtapose traditional practice in multiple regression analyses of manifest variables and SEM analyses of latent variables. This issue is not only very important in relation to IWAMs, but also in SEM more generally. For multiple regression analyses of IWAMs, the results include correlations between predictor variables and outcomes as well as the path coefficients. Particularly when there are many predictors, it is useful to evaluate how much variance can be explained by a set of predictors (i.e., the total variance explained by the predictors without controlling for other predictors) and how much variance can be explained by the set of predictors over and above what can be explained by other variables in the predictor equation (i.e., the unique variance explained by a set of predictor variables after controlling for the other predictors). In traditional multiple regression packages (e.g., SPSS) this is done by introducing sets of variables in a stepwise fashion (evaluating the change in variance explained in each step) and a summary of the final regression model that includes the variance uniquely explained by each set of variables. Historically, this type of unpacking of the multiple regression prediction equation has been important in the evaluation of support for IWAMs (e.g., Marsh, 1986, 1993, 1995). However, this type of information is not so readily available in SEMs. In most SEM studies the main focus is on global goodness of fit and the path coefficients rather than the variance explained by a set of predictors. Consistent with approaches to the evaluation of IWAMs based on multiple regression, in latent IWAMs it is also relevant to evaluate the total and unique variance that can be explained by domains and the interaction effects. The overarching rationale of the taxonomy of IWAMs presented here is to provide this decomposition of effects in relation to variance explained in outcome variables as well as goodness of fit. More specifically, starting with a well-defined, baseline measurement model (Model 0 in Table 1), we then evaluate the contribution of each of the three sets of parameters to goodness of fit and variance explained in the global outcome using a taxonomy of models (Models 1–7 in Table 1).

Model 1

We begin with Model 1 in which all paths ($\beta1–\beta9$) are estimated. This full SEM is equivalent to the measurement model (Model 0) in that the number of estimated parameters, degrees of freedom, and goodness of fit are all the same. The major difference is that covariances between the three sets of predictor factors and the global outcome factor are now represented as path coefficients (i.e., $\beta1–\beta9$ in Figure 1). Hence, so long as the measurement model fits the data well, so will Model 1.

Model 2

In Model 2, only paths leading from the actual latent factors to the outcome variable (β1–β3 in Figure 1) are freely estimated, whereas the paths from the importance factors (β4–β6) and interaction factors (β7–β9) are constrained to be zero. For this model we provide path coefficients (β1–β3), the change in goodness of fit based on the Wald test (Muthén &, Muthén, 2015; also see subsequent discussion) compared to the full model (Model 1), and the variance explained by the three paths included in the model. In this way, Model 2 provides an estimate of the total variance that can be explained by the actual factors, without taking into account the variance explained by the importance and interaction factors. Also, the difference in variance estimates based on Models 1 and 2 provide an estimate of the variance explained by the paths not included in Model 2 (β4–β9). **Models 3 and 4**

Similar to Model 2, Models 3 and 4 provide estimates of the variance explained and fit of models based on only the importance factors (β4–β6; Model 3) or only the interaction factors (β7–β9; Model 4). Model 4 is potentially interesting as a "pure" weighted-average model in that it only includes the effects of actual factors weighted by the importance factors. Although useful in terms of variance explained, it is also clear that the variance explained by the interactions confounds variance attributable to main effects (which are absent in this model) and variance that is uniquely explained by the interaction effects. However, an important contribution of the paradigmatic approach is that it provides estimates of the total variance explained by a set of parameters (including variance that can be explained by other variables) and unique variance explained by a set of parameters (excluding variance that can be explained by other variables). Furthermore, in much substantive research, Model 4 is a typical operationalization of the weighted-average model (i.e., the average of each actual domain weighted by its importance). Nevertheless, a cautious interpretation of this model is warranted in that interactions potentially confound the "main" effects of the actual and importance factors with variance that can be uniquely attributed to the interactions after controlling for the main effects.

Model 5

In Model 5, paths leading from the actual (β1–β3) and importance (β4–β6) latent factors to the outcome variable are freely estimated, whereas the paths from the interaction factors (β7–β9) are constrained to be zero. Again we provide path coefficients, the change in goodness of fit based on the Wald test compared to the full model, and the variance explained by the paths included in Model 5. In this way, support for the comparison of Model 5 (with no latent interactions) and Model 1 (the full model) provides a particularly important test of

the IWAM. The difference in variance explained by the two models is an estimate of the variance uniquely attributable to the interactions. The IWAM is supported when the change in goodness of fit for Models 1 and 5 is significant, the change in variance explained by Models 1 and 5 is substantively meaningful, and the direction of the path coefficients leading from the interaction factors to the global outcome is positive.

Models 6 and 7

Similar to Model 5, Models 6 and 7, compared to Model 1, provide estimates of variance uniquely explained by the importance factors (Model 6, which excludes the importance factors) and the actual factors (Model 7, which excludes the actual factors). Whereas this distinction between total variance explained by a set of predictors and variance uniquely explained by a set of predictors is a typical concern in applied multiple regression, it has been given surprisingly little attention in SEM.

Operationalizing Models 1 Through 7 in the Taxonomy of Models

For Models 2 through 7, we conducted two complementary analyses. First, we used the same syntax as for Model 1, but changed paths to covariances (as in the measurement model). Thus, for example, in Model 2 the six paths ($\beta4$–$\beta9$) were specified as covariances rather than paths. In this way only the actual factors ($\beta1$–$\beta3$) contributed to the prediction of the global factor, and variance explained was thus based on only these factors. However, because the relations represented by the excluded paths ($\beta4$–$\beta9$) are included in the model as covariances, Models 1 and 2 are equivalent in the sense that the number of estimated parameters, degrees of freedom, and goodness of fit are all the same. Indeed, in this sense all eight models, including the measurement model, are equivalent. This strategy has the advantage of providing estimates of the covariances between the excluded latent predictor factors and the global outcome variables, after controlling for the included latent factors (in subsequent tables these covariances are shaded to distinguish them from the path coefficients). Importantly, however, this approach holds the model constant when estimating changes due to excluding or including different sets of parameters (i.e., they are merely transformed rather than excluded), whereas it is possible that completely eliminating different parameters (even in nested models) could result in offsetting changes in other parameter estimates that cascade through the model. However, because the goodness of fit for all the models is the same, this approach does not provide a test of the change in fit due to completely eliminating the parameters in question. Thus, we also reran each analysis using the Wald test to evaluate the change in fit associated with constraining the paths to be zero (and not representing these as covariances).

STUDY 1: TESTING INDIVIDUALLY WEIGHTEDAVERAGE MODELS WITH SIMULATED DATA

Study 1: Simulated Data With Meaningful Interactions

We begin by demonstrating the application of this approach to IWAMs with two simulated data sets based on Figure 1. The two population generating models are essentially the same, except that one is designed to support the IWAM (i.e., latent interactions, β7–β9 in Figure 1 are positive, and substantively meaningful) whereas the other is not (β7–β9 are zero in the population generating model). Approximating true population-level data, each of these simulated data sets is based on a very large number of cases (N = 100,000; see Supplemental Appendix 1A for the syntax used to generate the two data sets that can be used to replicate and extend analyses presented here; also see Supplemental Appendix 1B and 1C for syntax examples of the models presented below). For purposes of instruction, we present the results in detail.

Model 0

We begin with tests of the measurement model (Model 0) for the simulated data with substantively meaningful interactions in support of the IWAM. Although this is typically a critical step in tests of the IWAM, for these simulated data it is trivial to show that the measurement model provides a good fit, $\chi 2$ (419) = 451, p = .130, root mean square error of approximation (RMSEA) = .001, comparative fit index (CFI) = Tucker–Lewis Index (TLI) = 1.00 (see Supplemental Appendix 1E for a brief discussion of goodness of fit).

Model 1

In Model 1 (the full model), consistent with the population generating model, all the paths from the actual (β1–β3) and interaction (β7–β9) are statistically significant and meaningfully large. The positive direction of these paths—particularly the interaction effects—is consistent with the IWAM. The paths from the importance factors (β4–β6) are much smaller in size. The variance explained by the three sets of factors in the full Model 1 (mult R2 = .648 in Table 2) is substantial and all the paths from the interaction factors to the global outcomes are positive. Thus, Model 1 provides preliminary support for the IWAM.

Models 2 Through 4

Consistent with the interpretation of paths in Model 1, the variance components for these models show that a substantial amount of variance can be explained by the actual factors (.478, Model 2) and interaction factors (.152, Model 4), whereas the variance explained by the importance factors is smaller but still statistically significant (.081, Model

3). Similarly, the Wald tests show that each of these three models provides a significantly poorer fit to the data if any of the sets of predictors is excluded. It is also interesting to note that path coefficients in these models tend to be somewhat larger than the corresponding paths in Model 1. Although this will typically be the case, the extent of these differences will depend on the sizes of covariances among the different factors (as is the case in any regression model). The sizes of the residual covariances (those shaded in gray in Table 2) provide an indication of the effect of constraining the corresponding path to be zero to the fit of the model. In general, these residual covariances will be smaller than the corresponding covariances in the measurement model (Model 0 in Table 2), but again these differences will depend on the sizes of relations among the predictor factors as well as what paths are included in the model. **Models 5 Through 7**

This set of models provides estimates of the variance uniquely explained by each of the combinations of any two of the three sets of predictor factors. Comparison of variance components based on these models with that in the full Model 1 provides an estimate of how much variance is uniquely explained by each set of predictors. For example, in Model 5 with paths leading from the actual ($\beta1$–$\beta3$) and importance ($\beta4$– $\beta6$), latent factors, the variance component is .495. The difference between this value and the value for the full Model 1 (.648 – .495 = .153) is an estimate of the variance uniquely explained by the interaction factors that were excluded from Model 5. Similarly, the results indicate that variance uniquely explained by the importance factors (Model 6) is only .017 (.648 – .631), and that due to actual factors (Model 7) is .414 (.648 – .234). However, due in part to the artificially large sample sizes, the fits of Models 2 through 7 differ significantly (all ps < .001) from that of the full model.

TABLE 2
Tests of Individually Weighted-Average Models (IWAMs) Based on Two Simulated Data Sets

Model	Actual (Act)			Importance (Imp)			Actual-by Importance Interaction (Int)			Summary	
	Act1	Act2	Act3	Imp1	Imp2	Imp3	Int1	Int2	Int3	Multiple R²	Wald(df) p-value
Simulated Data Set 1 (with latent interactions)											
0 measurement	.519	.547	.437	.212	.177	.075	.190	.316	.130	N/A	0
1 full model	.300	.366	.237	.124	.059	−.003	.188	.316	.133	.648(.003)	0
2 act only	.338	.377	.220	.109	.059	.001	.190	.317	.134	.478(.003)	8350(6) .000
3 imp only	.454	.492	.413	.211	.176	.075	.191	.316	.128	.081(.003)	20772(6) .000
4 int only	.519	.548	.438	.215	.174	.075	.189	.314	.129	.152(.003)	26171(6) .000
5 act & imp	.302	.364	.236	.119	.063	−.004	.190	.317	.133	.495(.003)	3768(3) .000
6 act & int	.338	.378	.221	.133	.055	.001	.188	.315	.133	.631(.003)	5108(3) .000
7 imp & int	.454	.494	.413	.215	.172	.076	.190	.315	.128	.234(.003)	20772(3) .000
1X full act =	.302	.302	.302	.124	.081	−.026	.189	.316	.134	.646(.003)	427(2) .000
Simulated Data Set 2 (with no latent interactions)											
0 measurement	.565	.592	.474	.233	.188	.081	.002	.001	.004	N/A	0
1 full model	.328	.395	.257	.133	.064	−.004	.002	.003	.008	.584(.003)	0
2 act only	.369	.408	.239	.121	.060	.001	.001	.002	.008	.564(.003)	6755(6) .000
3 imp only	.494	.534	.448	.232	.186	.082	.003	.002	.002	.096(.003)	23949(6) .000
4 int only	.565	.592	.474	.233	.188	.091	.002	.001	.004	.000(.003)	32667(6) .000
5 act & imp	.328	.395	.256	.133	.064	−.004	.002	.003	.008	.584(.003)	1(3) .844
6 act & int	.365	.402	.236	.119	.059	.000	.001	.002	.008	.564(.003)	6753(3) .000
7 imp & int	.487	.526	.442	.229	.184	.080	.003	.002	.002	.096(.003)	23949(6) .000

Note. See Figure 1 and Table 1 for a depiction of the models, factors, and variables. Coefficients shown in bold are statistically significant ($p < .05$). Shaded parameters represent covariances, whereas those not shaded are path coefficients. Multiple R^2 is the variance in the global outcome explained by predictor variables (unshaded coefficients). The Wald test is based on the difference between the full model (Model 1) and each subsequent model. For Data Sets 1 and 2 respectively, $\chi^2(419) = 451$ & 448, $p > .10$; root mean square error of approximation = .001, comparative fit index = 1.000; Tucker–Lewis Index = 1.00.

Of course, there are a huge number of additional models that could be tested in addition to those included in the taxonomy of variables. For present purposes we consider Model 5X that is similar to Model 5 in that the latent interactions that are critical to the IWAM are excluded. However, whereas in Model 5 the paths from the latent interaction factors were constrained to be zero, in Model 5X latent interactions are not even included in the model. Although the χ^2 (665, 211) for this model is statistically significant due to the huge N, the fit indexes (RMSEA = .005, CFI = 1.000, TLI = .999) indicate that the fit of this model is remarkably good. Hence, an inappropriate conclusion based on the fit of Model 5X might be that the fit is so good that it is not necessary test for latent interactions. This illustrates the point noted by Mooijaart and Bentler (2010) and others that conventional fit indexes are not sensitive to multiplicative relations between latent variables. This also demonstrates the importance of evaluating total and unique variance components associated with each of the three sets of predictor variables in relation to the IWAM taxonomy of models.

Study 1B: Simulated Data With Interactions Specified to Be Zero

The second simulated data set is similar to the first with the major exception that the latent interactions in the population generating model were specified to be zero. Without evaluating each of the separate models in detail, we point to several key features. In particular, the Wald test comparing results of Models 1 and 5 is nonsignificant, indicating that constraining the latent interaction effects to be zero has no influence on goodness of fit.

18

Similarly, the variance component for Model 5 (with latent interactions constrained to be zero) is the same as for full Model 1 (with latent interactions freely estimated). Hence the variance uniquely attributable to the latent interactions is 0. Finally, the variance explained by Model 4 (latent interaction effects only) was 0, indicating that the total effect of the latent interactions was also 0. It is also of interest to fit an additional Model 1X in which the three paths leading from actual factors (β1–β3 in Figure 1) are constrained to be equal. The Wald test (552, 2, p < .0001) based on this constrained model shows that the fit of this model is significantly poorer than that of Model 1, indicating the paths are significantly different. In this special case where the latent interactions are not statistically significant, there is evidence in favor of a normative WAM but not an IWAM. Importantly, however, this support for a normative WAM should not be confused with support for an IWAM. In particular, whereas an optimally weighted combination of the three actual factors performs better than an equally weighted combination, the same higher weight for act2 (β2 = .395) is the same for every case, as are the lower weights for the other two actual factors (β1 = .328, β2 = .257). In this sense, act1 is more important across the group as a whole, but the contribution of actual1 is the same for individuals who rate this factor as relatively more or less important; the effects of actuals 1–3 are not moderated in the importance ratings.

The juxtaposition between the results based on the two simulated data sets is also important in that it clearly demonstrates that the taxonomic approach to IWAMs presented here is able to demonstrate support for the IWAM when the model is true, as well as provide evidence against it when the model is false. This is important in that particularly based on results of the IWAM in self-concept research, there is little support for IWAMs based on either latent SEM emphasized here or manifest (multiple regression) models that are the basis of much previous research. We now turn to three real data applications of the IWAM and offer suggestions about how to overcome the issues involved in the analysis of each of these data sets; suggestions that might have broad applicability to real data from other areas of research involving weighted models.

STUDY 2: TESTING THE INDIVIDUALLY WEIGHTED-AVERAGE MODEL: SELF-CONCEPT DATA

Overview of the Data Set

Marsh (2008) argued that to the extent that there is any support for the IWAM at all, it might only apply to a few narrowly focused self-concept domains (those that are unimportant

to most people but very important to a few, such as spiritual self-concept or, perhaps, physical self-concept) and for limited subgroups of individuals, and obviously does not have generalizability over different self-concept domains and different individuals as originally envisioned by James (1963). Indeed, even James focused on his self-beliefs in relation to his competence in the Greek language, a very narrowly focused domain that was likely to be highly important to a few individuals but unimportant to many— including James. Following Marsh (2008), Scalas et al. (2013) evaluated support for the IWAM for components of self-concept specifically selected as being most likely to support the model (see their article where the sample, materials, and results are presented in more detail). Here we reanalyze these data specifically chosen to be optimal in terms of being likely to support IWAMs in relation to selfconcept theory and research.

Briefly the sample consisted of UK adolescents (n = 402; 13–15 years old) who completed self-concept responses to positively worded self-concept items in three specific domains: physical (six items; e.g., I feel good about who I am physically), academic (three items; e.g., I learn quickly in most academic subjects), and spiritual (six items; e.g., I am a better person as a consequence of my spiritual/religious beliefs). Students also completed the Rosenberg Measure of Global Self-Esteem Inventory (Rosenberg, 1965) consisting of five positively and five negatively worded items (e.g., Overall, most things I do turn out well). All items were rated on a 6-point Likert-type scale ranging from 1 (false for my actual self) to 6 (true for my actual self). For the importance ratings, the original items were re-presented with instructions to rate how important each item was to them rather than how they actually saw themselves. Another 6-point Likert-type scale was used ranging from 1 (not important) to 6 (important).

Measurement Model

The measurement model for these data (see syntax in Supplemental Appendix 2A) resembles Figure 1 in that there are three multi-item actual self-concept scales, three corresponding multi-item importance scales, and one multiitem global outcome (self-esteem)—although the number of items in each scale varies. However, as is typical with real data, there are several features that make the measurement model more complicated. First, responses to the Rosenberg Self-Esteem Scale typically do not result in a cleanly defined single factor. Previous research (Marsh, 1996; Marsh, Scalas, & Nagengast, 2010) argued that responses to this instrument are best represented as a global selfesteem factor and item-wording method factors reflecting the positively and negatively worded items. Following the original analyses by Scalas et al. (2013) and consistent with recommendations by Marsh et al.

(2010), here we also use a bifactor model (Reise, 2012; also see Marsh, Morin, Parker, & Kaur, 2014) in which all 10 items load on the global esteem factor, whereas positively worded items load on a positive item method factor and negatively worded items load on the negative-item method factor (see syntax in Supplemental Appendix 2A). The two method factors are constrained so as to be uncorrelated with each other and all other factors in the model.

Second, the self-concept and importance items (and thus the interaction indicators) had parallel wording. When items with the same wording are designed to measure different factors, typically there are item-wording method effects such that responses to the items are more highly correlated than can be explained by covariances among the different factors. Following Scalas et al. (2013), we included a priori correlated uniquenesses for self-concept ratings, importance rating, and the interaction indicators based on items with the same wording (see syntax in Supplemental Appendix 2A). We also note that even when self-concept and importance ratings do not have parallel wording, it is important to allow errors in the product terms to covary with the corresponding errors for the self-concept and importance indicators (see Marsh et al., 2004).

Results

Analyses presented here follow closely those presented in Study 1 with simulated data. In particular, we standardized all measured variables (M = 0, SD = 1), defined indicators of the latent interaction factors as the cross-product actual selfconcept and importance ratings, and fit a measurement model (as described previously). We then used factor loadings from the measurement model to construct appropriately standardized latent interaction models to test the IWAM corresponding to the eight models in Table 1 (see syntax in Appendices 2A and 2B). The overall measurement model provided an acceptable goodness of fit, $\chi 2$ (1330) = 2,158, RMSEA = .039, CFI = .931, TLI = .923. Next we summarize results based on the application of the taxonomy of models to these data.

Model 1

In Model 1 (the full model, Table 3), there are only three statistically significant paths, actual academic (.395), actual physical (.679), and the physical–importance interaction (–.146). Overall, the variance explained is substantial (mult R2 = .835). Although one of the latent interactions is statistically significant, the direction of the effect was negative rather than positive (i.e., the effect of physical selfconcept was smaller, not larger, for students who felt physical self-concept was more important). This is demonstrated in Figure 2, a simple-slopes graph of the interaction showing the effect of physical self-concept is greater (i.e., the

slope is steeper) for students who perceive physical self-concept as less important, opposite to the a priori prediction based on the IWAM. A literal interpretation of this effect is that the physical self-concept has a more positive effect on self-esteem for respondents who rate the domain as less important. Because the direction of the interaction is negative rather than positive, it provides clear evidence against the IWAM. In summary, Model 1 provides no support for the IWAM.

Physical SC X Imp on Global Self-Esteem
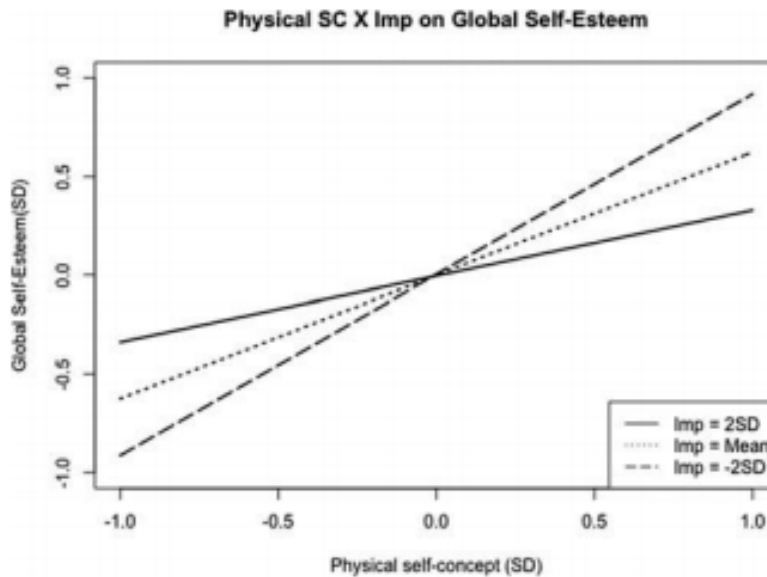
FIGURE 2   A simple-slopes graph of the regression line relating physical self-concept (SC) to global self-esteem (in standard deviation units) for individuals who rate the importance of physical self-concept +1, 0, and −1 standard deviations above or below the mean importance ratings (see Model 1, Table 3).

Models 2 Through 4

The variance components for these models show that a substantial amount of variance can be explained by the actual self-concept factors (mult R^2 = .816, Model 2); the variance components are much smaller for importance (.273, Model 3) and for interactions (.184, Model 4). However, all three variance components are statistically significant. Similarly, the Wald tests show that each of these three models provides a significantly poorer fit to the data than the full Model 1. However, although the results indicate support for the inclusion of latent interactions, this is due primarily to the physical–importance interaction in which the negative direction of the effect is opposite to that predicted by the IWAM (see Figure 2), again offering no support for the IWAM.

Models 5 Through 7

Models that include actual self-concepts explain substantial portions of data in global self-esteem: Model 5 (actual and importance; mult R2 = .820) and Model 6 (actual and interactions; mult R2 = .832).

TABLE 3
Individually Weighted Average Model for Three Self-Concept Domains and Global Self-Esteem

Model	Self-Concept			Importance			Self-Impt Interaction			Summary	
Outcome	Acad	Phys	Spirt	Acad	Phys	Spirt	Acad	Phys	Spirt	MRq	Wald(df) P-val
0 measurement	.641	.827	.216	.332	.516	.108	−.090	−.383	.173	0	0
1 full model	.395	.625	.096	−.021	−.002	−.090	.036	−.146	.050	.835(.038)	0
2 act only	.386	.679	.030	−.017	.024	−.019	−.004	−.102	.046	.816(.037)	136(6) .000
3 imp only	.485	.532	.131	.096	.472	−.019	.043	−.116	.120	.273(.056)	191(6) .000
4 int only	.584	.665	.141	.234	.295	.040	.033	−.402	.189	.184(.065)	195(6) .000
5 act & imp	.389	.646	.107	−.037	.067	−.089	−.003	−.091	.038	.820(.036)	109(3) .000
6 act & int	.396	.624	.014	−.026	−.013	−.023	.040	−.138	.057	.832(.038)	2(3) .602
7 imp & int	.482	.504	.120	.138	.351	−.035	.096	−.211	.131	.318(.061)	174(3) .000

Note. See Figure 1 and Table 1 for a depiction of the models, factors, and variables. Coefficients shown in bold are statistically significant ($p < .05$). Shaded parameters represent covariances, whereas those not shaded are path coefficients. Multiple R^2 is the variance in the global outcome explained by predictor variables (unshaded coefficients). The Wald test is based on the difference between the full model (Model 1) and each subsequent model. For Models 1 and 2, and each of the subsequent models when path coefficients constrained to be zero are replaced with covariances. Goodness of fit of the measurement model and all seven structural equation models was $\chi^2(1330) = 2,158$, root mean square error of approximation = .039, comparative fit index = .931, Tucker–Lewis Index = .923.

Indeed, the Wald test comparing Model 1 (full model) and Model 6 (actual & interaction) is not statistically significant (Wald = 2, df = 3, p = .602) and the amount of variance uniquely explained by the importance factors is close to zero (.003 = .835 − .832). However, like Model 1, the direction of the latent physical–importance interaction is negative rather than positive and none of the other latent interactions is statistically significant. It is also relevant to note that the total variance that can be explained by the importance ratings (.273 in Model 3) is substantial, but that the amount of variance uniquely explained by the importance ratings after controlling for actual and interaction effects is close to zero (.003). This illustrates that there is much multicollinearity associated with the importance ratings that is likely to complicate interpretations of the data unless appropriate statistical modeling is used.

Methodologically, analyses of these self-concept data extend the analyses based on simulated data in several important ways that are likely to have broad generalizability to other applied research. The use of the bifactor approach with the Rosenberg scale to control for item-wording method effects is likely to generalize to other outcome variables that are based on a mixture of positively and negatively worded items. Similarly, the correlated uniqueness approach used to model items based on parallel worded items is also likely to be useful when the same wording is used for items designed to assess actual and importance factors. Substantively, the results provide a clear lack of support for the IWAM consistent with a growing body of self-concept research based on manifest (multiple regression) and latent (SEM) analyses reviewed earlier.

STUDY 3: TESTING THE INDIVIDUALLY WEIGHTED-AVERAGE MODEL: QUALITY OF LIFE

Background to the Application of the IWAM to Quality of Life and Life Satisfaction

Data for this study came from Transitions from Education to Employment (TREE), a longitudinal panel study following up Swiss students who participated in the Programme for International Student Assessment (PISA 2000) and who left compulsory schools in the same year. For more details on the sample, variables, and availability of the data, see TREE (2013a, 2013b, 2013c). Data used here are based on Panel 8 collected in 2010 when most of the respondents (N = 2,751) were about 26 years of age. Quality of life was assessed in relation to four domains: employment, education, and training; partnership and children; social activities (e.g., associations, political organizations or parties, unions, political organizations, volunteer work); and leisure (hobbies, sports, recreational activities, contacts with friends). Multiple-item global outcome variables were anomie, self-esteem, depression, positive emotions, negative emotions, and positive life attitudes.

The multiple domains of life satisfaction in the TREE survey were derived from the German Socioeconomic Panel Study framework (Schimmack, 2008). However, one criticism of the original framework noted by Schimmack (2008) is that the domains are not weighted by subjective importance (Andrews & Whithey, 1976; Schimmack & Oishi, 2005; Schimmack, Diener, & Oishi, 2002), an issue addressed in TREE data by the inclusion of subjective ratings of importance. In support of WAM approaches, Schimmack et al. (2002) reported that weighted-average measures of domains added to the prediction of global life satisfaction beyond what could be explained by unweighted averages. Nevertheless, their approach did not distinguish between what here we refer to as normative weighted-average models (i.e., the regression weights are not identical for each of the domains, but are identical across individuals) and the individually weighted-average model (IWAM based on latent interactions where weights vary according to the individual importance placed in each domain) presented here. Furthermore, other researchers (e.g., Campbell, Converse, & Rodgers, 1976) have argued that there is little or no empirical support for the use of importance as a weighting factor, although Hsieh (2013) noted that the role of domain importance continues to be a topic in quality of life research. Indeed, Hsieh (2013) argued that inconclusive evidence is due substantially to the way that importance weighting is assessed. In this respect, there are many parallels between research into quality of life and self-concept research reviewed earlier—particularly potential confusion between normative WAMs and IWAMs.

Measurement Model

For each of the domains of quality of life, individuals rated satisfaction (1 = very unsatisfied to 6 = very satisfied) and importance (1 = entirely unimportant to 6 = very important) based on single item ratings. Interactions were based on the cross-product of satisfaction and importance ratings. Life satisfaction is related to six global outcomes: anomie, self-esteem, depression, positive emotions, negative emotions, and global positive life satisfaction. To facilitate interpretations, all negatively worded items and scales are reverse-scored so that higher values reflected more positive outcome. After standardizing (M = 0, SD = 1) all items, interactions were based on the cross-product of corresponding satisfaction and importance ratings.

The measurement model (see Supplemental Appendix 3A) for these data is based on Figure 1, but also incorporates a number of features that are specific to these data. First, each of the specific domains (actual and importance ratings) is represented by a single item rather than multiple indicators. Although typical in quality of life research, this is a potentially important limitation in terms of assessing the factor structure and controlling for measurement error (both unreliability in the specific domains and also potential method effects in complex measurement structures).

Second, the global outcome measures consist of a set of six global well-being measures. Although global positive life satisfaction is most closely aligned to measures of satisfaction in specific life domains, it is important to emphasize that the IWAM is easily extended to include multiple outcome measures; indeed constructs such as those considered here are frequently used in quality of life research. A typical approach might be to use scale scores to represent each of the outcomes or, perhaps, to model each as a latent factor in six separate analyses. However, there are important methodological and substantive limitations to these approaches that led us to represent all six global outcomes as latent factors in the same model. Thus, for example, this allows us to evaluate the factor structure underlying these constructs to determine whether the factors are welldefined and distinguishable. Also, although beyond the scope of this demonstration, there are many potentially interesting analyses that could not be performed if each outcome were considered in separate models (e.g., whether the pattern of paths from specific components is invariant over multiple outcomes).

For present purposes, we used exploratory structural equation modeling (ESEM) to model the factor structure underlying the six outcome measures. Although a detailed review of ESEM is beyond the scope of this investigation (see Marsh et al., 2014; Marsh et al., 2009;

Morin, Marsh, & Nagengast, 2013), Marsh and colleagues have argued that ESEM represents an optimal compromise between the flexibility of exploratory factor analysis (EFA) and the parsimony and rigor of confirmatory factor analysis (CFA). Using target rotation, the analyst can specify an a priori factor structure as in CFA, but like EFA, ESEM allows items to cross-load on different factors. Marsh et al. emphasized that the typical CFA structure is almost always too restrictive, specifying that each item loads on one and only one factor. Indeed, using the ESEM-within-CFA strategy (Marsh et al., 2014), it is possible to transform an ESEM into an equivalent CFA model. In this case, CFA is a special case of ESEM in which all cross-loadings are constrained to be zero. Simulation and a growing number of real data studies (see Marsh et al., 2014) demonstrate that ESEM almost always results in a better fit to the data than does CFA and latent factors that are more distinguishable (i.e., less correlated in that constraining nonzero factor loadings to be zero typically results in positively biased estimates of factor covariances). Although in its simplest form ESEM is the same as EFA, ESEM allows researchers to incorporate the full range of CFA and SEM models, such as those required to test IWAMs in ways that are not possible with EFAs.

Results

As in Studies 1 and 2, we standardized all measured variables (M = 0, SD = 1), defined interaction factors as the cross-product of actual and importance ratings, and fit a measurement model. The a priori ESEM measurement model (with 15 single-item factors—5 actual satisfaction, 5 importance, and 5 interaction factors—and the six global outcomes) provided an acceptable goodness of fit, $\chi 2$ (484) = 1,999, RMSEA = .034, CFI = .956, TLI = .927. (See Supplemental Appendix 3A for syntax for the measurement model and further discussion of ESEM.) We then constructed appropriately standardized latent interaction models to test the IWAM corresponding to the eights models in Table 1 summarized in Tables 4 and 5.

Model 1

In the full Model 1, the variance explained is statistically significant for each of the six outcomes (Table 4). Consistent with the design of the study, the variance component is higher for the global measure of life satisfaction (mult R2 = .447). Indeed, except for positive emotions (.301), the variance components for the other global measures of well-being are more modest (.128–.144).

The path coefficients relating all five actual satisfaction factors to the six global outcomes are nearly all positive and statistically significant (28 of 30 were significant, 2 were not; see Table 5). However, there is substantial variation depending on the outcome and the

domain. The largest paths tend to be for the global positive life satisfaction outcome (particularly paths from employment, education, and partner domains, but not from social and leisure). Nevertheless, the pattern of effects is quite differentiated. Thus, for example, employment is the best predictor of positive life satisfaction and positive emotions, but is the weakest predictor of self-esteem, whereas the social and leisure domains are the weakest predictors of positive life satisfaction but better predictors of positive and negative emotions. In contrast to the satisfaction paths, paths from the importance factors are mostly small and the majority (17 of 30) are not even statistically significant. In terms of support for the IWAM, the critical paths are from the interaction factors. However, these are even smaller (the largest was .091) and only 3 of 30 reach statistical significance.

Models 2 Through 7

It is relevant to compare variance components in Model 1 (full model) and with those in Models 2 through 7 (Table 4). This shows that much of the variance in the global outcomes that can

TABLE 4
Quality of Life Variance Components: Individually Weighted-Average Model for Five Quality of Life Facets and Six Global Outcomes

Outcome	Model 1 (SAT+IMP+Int)	Model 2 (SAT only)	Model 3 (IMP only)	Model 4 (INT only)	Model 5 (SAT+IMP)	Model 6 (SAT+INT)	Model 7 (IMP+INT)
ANOM	.128(.015)	.121(.015)	.025(.006)	.001(.002)	.126(.015)	.123(.015)	.026(.006)
SELE	.137(.016)	.105(.014)	.061(.011)	.005(.003)	.131(.016)	.109(.015)	.065(.011)
DEPR	.142(.018)	.134(.017)	.029(.008)	.005(.004)	.139(.017)	.137(.017)	.030(.008)
PANAP	.300(.020)	.266(.020)	.093(.013)	.010(.005)	.293(.020)	.272(.020)	.103(.013)
PANAN	.149(.018)	.140(.017)	.026(.007)	.006(.004)	.146(.017)	.143(.018)	.028(.008)
POSL	.438(.022)	.394(.024)	.143(.017)	.016(.008)	.432(.021)	.400(.023)	.151(.017)
Wald (df p-value)		211 (60, .000)	594 (60, .000)	540 (60, .000)	61 (30, .000)	157 (30, .000)	501 (30, .000)

Note. The five domains were employment, work; education, further education/training; partnership/children; social activities (e.g., associations, political organizations/parties, unions, political organizations, volunteer work); and leisure (hobbies, sports, recreational activities, contacts with friends). For each of the five domains, individuals rated satisfaction (SAT) and importance (IMP) based on single item ratings. Interactions (INT) were based on the cross-product of SAT and IMP ratings. The six global outcomes were ANOM = anomie; SELE = self-esteem; DEPR = depression; PANAP = positive emotions; PANAN = negative emotions; and POS = positive life attitudes. Goodness of fit of the measurement model and all seven structural equation models was χ^2 (484) = 2,299, root mean square error of approximation = .035, comparative fit index = .951, Tucker–Lewis Index = .920. The Wald test is based on the difference between the full model (Model 1) and each subsequent model. For Models 1 and 2, and each of the subsequent models when path coefficients constrained to be zero are replaced with covariances Goodness of fit of the measurement model and all seven structural equation models was χ^2(484) = 2,299, root mean square error of approximation = .035, comparative fit index = .951, Tucker–Lewis Index = .920.

TABLE 5
Quality of Life (QoL): Individually Weighted Average Model for Five QoL Facets and Six Global Outcomes

Model / Outcome	Satisfaction Ratings					Importance Ratings					Satisfaction × Importance Interactions				
	Employ	Educ	Partner	Social	Leisure	Employ	Educ	Partner	Social	Leisure	Employ	Educ	Partner	Social	Leisure
M30 ANOM	.291	.251	.151	.235	.208	.115	.116	.063	.057	.077	−.038	−.008	−.019	.005	.002
SELE	.195	.253	.179	.229	.218	.149	.138	.124	.088	.173	−.056	.001	−.054	−.015	.013
DEPR	.244	.231	.254	.258	.232	.114	.074	.097	.075	.101	−.069	−.015	−.042	−.028	−.018
PANAP	.436	.335	.234	.327	.330	.233	.176	.105	.161	.152	−.039	.057	−.033	−.050	−.053
PANAN	.264	.227	.136	.298	.278	.048	.072	−.004	.127	.107	−.028	−.048	.003	−.048	−.060
POS	.489	.430	.406	.376	.327	.235	.158	.251	.191	.188	−.094	.015	−.092	−.058	−.029
M31 ANOM	.197	.095	.036	.101	.074	.013	.064	.026	−.029	.034	−.013	.008	.018	.015	.028
SELE	.043	.137	.080	.102	.079	.077	.049	.052	−.015	.136	−.018	.024	.006	−.017	.067
DEPR	.106	.092	.177	.105	.083	.053	.003	−.006	−.006	.059	−.035	.007	.002	−.023	.028
PANAP	.325	.103	.102	.087	.175	.125	.028	.063	.057		−.007	.093	.033	−.043	.018
PANAN	.166	.070	.040	.142	.139	−.038	.026	−.070	.044	.050	−.021	−.027	.000	−.029	.011
POS	.366	.250	.297	.090	.097	.121	.017	.141	.095	.123	−.045	.055	.057	−.048	.050
M32 ANOM	.198	.122	.043	.081	.078	.041	.060	.017	−.009	.029	−.028	−.001	.008	.025	.036
SELE	.058	.168	.091	.074	.107	.106	.080	.069	.028	.122	−.055	.007	−.028	−.003	.051
DEPR	.122	.104	.170	.092	.097	.066	.030	.011	.009	.050	−.069	−.011	−.005	−.018	.018
PANAP	.346	.146	.092	.087	.179	.143	.120	.035	.087	.084	−.027	.077	.005	−.028	−.003
PANAN	.154	.079	.006	.160	.153	−.013	.028	−.042	.041	.037	−.029	−.041	.024	−.037	−.009
POS	.382	.265	.319	.093	.116	.147	.081	.132	.126	.151	−.093	.027	−.031	−.036	.014
M35 ANOM	.196	.095	.040	.099	.072	.020	.063	.016	−.031	.026	−.011	.008	.015	.024	.034
SELE	.044	.139	.082	.097	.072	.084	.048	.052	−.018	.116	−.013	.019	.006	.000	.070
DEPR	.108	.094	.177	.102	.079	.065	.004	−.003	−.007	.050	−.037	−.006	−.002	−.018	.026
PANAP	.327	.107	.108	.077	.171	.120	.075	.009	.064	.056	.019	.089	.023	−.026	.016
PANAN	.167	.070	.038	.143	.137	−.030	.031	−.066	.043	.046	−.031	−.038	−.003	−.032	.005
POS	.366	.252	.303	.081	.089	.134	.012	.112	.092	.113	−.031	.039	.040	−.030	.048

Note. The five domains were Employ = employment, work; Educ = education, further education/training; Partner = partnership/children; Social = social activities (e.g., associations, political organizations/parties, unions, political organizations, volunteer work); leisure (hobbies, sports, recreational activities, contacts with friends). For each of the five domains, individuals rated satisfaction and importance (based on single item ratings). Interactions were based on the cross-product of satisfaction and importance ratings. The six global outcomes were ANOM = anomie; SELE = self-esteem; DEPR = depression; PANAP = positive emotions; PANAN = negative emotions; POS = positive life attitudes;

be explained by full Model 1 can also be explained by the actual satisfaction ratings alone (Model 2) and, to a lesser extent by the importance factors (Model 3). Thus, for example, the variance explained in positive life satisfaction is .438 in Model 1 and .394 in Model 2 (Table 4). In contrast, variance explained in Model 3 (importance only) is .143 and that in Model 4 (interactions only) is not statistically significant. Although variance components for the other global outcomes are smaller than for positive life satisfaction, the pattern of results is similar.

In particular, none of the variance components for Model 4 (interaction only) is statistically significant for any of the six outcome measures.

Across the entire set of models, those that contain paths from the actual satisfaction factors to the global outcomes (Models 1, 2, 5, and 6 in Table 4) explain substantial amounts of variance, whereas models that do not contain these paths explain much less variance. In particular, unique variance explained by the set of interaction factors (the difference between Model 1 and Model 5) is small for all six outcomes; differences in variance components varied from .002 (anomie) to .007 (positive emotions). Also, only 2 of the 30 residual covariances relating interaction factors and outcomes are statistically significant. Nevertheless, due in part to the large sample size, the Wald test comparing Models 1 and 5 is statistically significant, providing some limited support for the IWAM.

Methodologically, analyses of these quality-of-life ratings extend the evaluations of IWAMs in several ways that could have generalizability to other research. Thus, the IWAM taxonomic approach can be applied even when actual and importance ratings are based on single-item indicators, although reliance on these manifest indicators substantially undermines the strength of the latent variable approach. Also, the use of such a diverse set of outcome variables provides a much richer framework for evaluating IWAMs. Although beyond the scope of this demonstration study, it would have been possible to evaluate the similarity in the paths across the different outcomes, to evaluate paths in relation to a higher order factor that represented all six global outcomes, or to evaluate whether specific (actual, importance, and interaction) paths were able to explain significant portions of variance in outcome factors beyond the global positive life satisfaction factor. Finally, the introduction of ESEM was important in terms of providing an acceptable-fitting measurement model compared to the highly restrictive CFA model in which all cross-loadings are constrained to be zero. Although ESEM could not be applied to the specific ratings that were based on single-item indicators in this study, ESEM offers potentially even more important advantages when the specific ratings are also based on multiple indicators (see discussion of job satisfaction ratings in Study 4). Substantively, at best the results provide very weak support for the IWAM in that across the entire set of 30 interactions (Model 1) only 3 were statistically significant and the unique variance attributable to interactions was nonsignificant and very small for all six outcomes.

STUDY 4: TESTING THE INDIVIDUALLY WEIGHTED-AVERAGE MODEL: JOB SATISFACTION

Background to the Application of the IWAM to Job Satisfaction

Job satisfaction is among the most widely studied topics on work-related attitudes and constructs (Judge, Weiss, Kammeyer-Mueller, & Hulin, 2017). The use of WAMs has a long and controversial history in the study of job satisfaction (Quinn & Mangione, 1973; Rosenberg, 1957; Vroom, 1964) that has considerable overlap with studies of IWAMs in other areas. Thus, Quinn and Mangione (1973) argued that most advocates of WAMs in job satisfaction studies used theoretical or common-sense rationales but offered limited empirical support in relation to a priori predictions. However, in their empirical study (based on alternative methods of weighting specific components by importance ratings in relation to the prediction of global job satisfaction), Quinn and Mangione (1973) found, "The data not only failed to support the hypothesis that the validity of job satisfaction ratings may be increased by weighting them by importance ratings but indicated, on the contrary, that importance-weighting actually reduced the validity of satisfaction ratings" (p. 1). They noted several statistical issues (e.g., high correlations among constructs resulting in multicollinearity, restriction of the range, appropriate statistical models, use of ipsative importance scores) similar to issues raised in the Marsh– Pelham debates (see earlier discussion) in relation to IWAMs in self-concept research. Also similar to Marsh's (2008) suggestion in relation to self-concept research (stemming from James), Quinn and Mangione (1973) argued that the strongest test of a WAM would entail a diversity of job facets, at least some of which were highly important to a few workers and highly unimportant to most, whereas most studies involved job facets selected to be at least reasonably important to all workers.

Ongoing research about the nature of job satisfaction has not resolved controversies about the relevance of WAMs. Thus, as emphasized by Judge et al. (2002; also see Judge et al., 2017), "job satisfaction is typically characterized as a global construct that also comprises specific factors" (p. 26). In evaluating support for the facet approach to job satisfaction, Judge et al. argued that an appropriate test is to predict global job satisfaction with the specific domains (essentially Model 2 in the taxonomy), but noted that the domains are so highly correlated that an unweighted model does nearly as well (e.g., Model 2 with paths constrained to be equal). Similarly, studies that regress job satisfaction domains on global outcomes other than global job satisfaction (e.g., global self-esteem, life satisfaction, commitment) also are based on an implicit WAM in which the weights are empirically determined as regression weights. Furthermore, as noted by Judge et al. (2017), other theoretical models such as Locke's (1976) value-percept theory explicitly weight domains in relation to importance (as in the IWAM), whereas other studies evaluate individual-difference

variables (other than importance) as moderators of the relation between specific components of job satisfaction and global outcomes.

Measurement Model and Preliminary Analyses

Following from the facet approach to job satisfaction and the evaluation of their relative importance, Stamp (1997) developed the Index of Work Satisfaction (IWS) used in this investigation. The IWS is a two-part multidimensional job satisfaction instrument. Part A measures satisfaction in relation to six specific domains of job satisfaction based on responses to 44 items: pay, professional status, autonomy, organizational policies, task requirements, and interactions. Part B consists of paired-comparison ratings of all possible combinations of the six specific domains. For each of these 15 paired-comparison ratings, respondents selected the one that they felt meant the most to their feeling of job satisfaction. In this investigation, outcome measures were multi-item scales that measure global job satisfaction, global self-esteem, and global work self-concept. The participants (332 registered Australian nurses) had a mean age of 46 years old (SD = 9), were mostly female (94%), and had an average of 23 years of working experience (for more detailed descriptions of the data, measures, and participants, see Cowin, 2002; Cowin, Johnson, Craven, & Marsh, 2008).

The measurement model is based on Figure 1, but also incorporates a number of features that are specific to these data. Preliminary factor analyses (ESEMs) revealed the necessity of splitting the interactions factor into two separate factors (nurse–nurse interactions and nurse–doctor interactions), a possibility specifically noted by Stamps (1997). A distinctive aspect of these data is the use of paired-comparison ratings of the specific domains in which the most important domain of each pair is scored +1 and the other domain was scored −1. Summing the responses for each domain provides a measure of relative importance that overcomes some potential problems with Likert-scale responses. However, it also results in ipsative scores (i.e., they sum to zero for each participant) that complicate the statistical analyses. In particular, any one of the scales is completely determined by the other five scales, so that one of the scales must be excluded from the analyses to avoid a positive nondefinite matrix (Cattell, 1944; Jackson & Alwin, 1980; also see Marsh, 1993; in relation to IWAMs) needed to pursue factor analysis models.

Another complication is the construction of the interaction terms. When the number of items measuring specific domains and their importance is the same, and particularly when there is a logical matching of specific domain and importance ratings (as in the self-concept data), the construction of indicators for the interaction factors is straightforward. However,

when the numbers are not the same and there is no one-to-one matching, there has been considerable discussion of how to form the interaction indicators. In their original presentation of the product-indicator approach to latent interactions, Marsh, Wen, and Hau (2004, 2006) proposed that the strategy should use all the importance and specific-domain items, but not reuse any of the items (see Supplemental Appendix 1G for further discussion).

In this investigation where there is only one importance score per domain, it would be possible to construct 44 crossproduct terms—one for each of the specific items. However, particularly when the number of specific rating items is substantial as with these data, the resulting number of variables becomes unwieldy. An alternative used here is to construct factor scores for each of the specific domains in preliminary factor analyses, and then to use the factor scores to construct the interaction terms. Thus, the final measurement model contained a total of 75 measured variables: 44 items representing seven specific domains, five importance scores (noting that one had to be left out because they were ipsative), seven interactions (the cross-product of factor scores representing the seven factors and their corresponding importance score, noting that because there were two social relation factors but only one social relation importance score, the one importance score for social relations was used to form interactions with both the specific social relation factors), and 18 items designed to measure the three global outcomes.

We again used ESEM (see earlier discussion in relation to quality of life data) to model the factor structure underlying the three global outcome factors as well as the seven factors representing domain-specific satisfaction factors. Following the strategy used by Marsh, Nagenghast, Morin, Parada, Craven, and Hamilton (2011; also see Marsh et al., 2014), separate ESEMs were performed for the specific and global factors, but included in a single analysis (along with importance and interaction scores). In this way, specific items were allowed to cross-load on different specific factors and global items were allowed to load on different global factors, but specific items were not allowed to load on global factors and global items were not allowed to load on specific factors. Thus the specific and global factors were not contaminated by each other at the level of individual items.

Following the general approach, we constructed a measurement model to represent appropriately standardized solutions for each of the eight models in the taxonomy (Table 1). We began by standardizing all measured variables (M = 0, SD = 1), defining single-measure indicators of the seven interaction factors as the cross-product of the specific domains (a factor score for each domain) and importance ratings, and fitting a measurement model based on all 75 measured variables (as previously described). We then used factor loadings from the

measurement model to construct appropriately standardized latent interaction models to test the eight IWAM models in Table 1. The overall measurement model provided an acceptable goodness of fit, $\chi^2 (1330) = 2,158$, RMSEA = .039, CFI = .931, TLI = .923, noting that the fit for all eight models in the taxonomy is necessarily the same as the measurement (see earlier discussion). Next we summarize results based on the application of the taxonomy of models to these data.

Model 1

In the full Model 1, the variance explained is statistically significant for each of the three outcomes (Table 6). Consistent with the design of the study, the variance component is highest for the global measure of job satisfaction (mult R^2 = .809) and also very high for the job self-concept (.718), but substantially smaller for global self-esteem (.208) that was not specific to the work setting. The path coefficients relating all seven actual satisfaction factors, importance factors, and interaction factors to the three outcomes (Table 7) are quite different from those in Studies 1 through 3; only 4 of 57 path coefficients are even statistically significant. In relation to each of the three outcomes, by far the largest contribution is for the specific satisfaction rating of professional standing. Indeed, only one other path (from task to job satisfaction) is statistically significant. In terms of support for the IWAM, the critical paths are from the interaction factors. However, none of these interaction terms contributes significantly to the prediction of any of the global outcome factors.

Models 2 Through 7

The variance components for these models show that much of the variance can be explained by various subsets of the 57 paths in the full Model 1. However, given the results for Model 1, it is not surprising that all models with the 21 paths from the seven specific satisfaction domains to the three outcomes are able to explain most of the variance explained by Model 1. Indeed, the fit of Model 2 that has only 21 of these paths does not differ significantly from that of the full Model 1 with all 57 paths (Wald = 42, df = 36, p = .298).

TABLE 6
Tests of the Individually Weighted-Average Model for Job
Satisfaction Over Three Outcome Measures

Model		Summary	
Outcome	Outcome	Multiple R^2	Wald (df) p value
0 measurement	**Global esteem**	**0**	**0**
	Job self-concept	**0**	**0**
	Job satisfaction	**0**	**0**
1 full model	**Global esteem**	**.208** (.047)	**0**
	Job self-concept	**.718** (.073)	
	Job satisfaction	**.809** (.058)	
2 act only	**Global esteem**	**.160** (.047)	42.603 (36) .298
	Job self-concept	**.704** (.073)	
	Job satisfaction	**.801** (.059)	
3 imp only	**Global esteem**	**.037** (.017)	305.903 (42) .000
	Job self-concept	**.083** (.026)	
	Job satisfaction	**.045** (.020)	
4 int only	**Global esteem**	.035 (.020)	285.574 (36) .000
	Job self-concept	**.059** (.027)	
	Job satisfaction	**.103** (.035)	
5 act & imp	**Global esteem**	**.190** (.046)	20.213 (21) .508
	Job self-concept	**.711** (.071)	
	Job satisfaction	**.803** (.059)	
6 act & int	**Global esteem**	**.178** (.048)	19.949 (15) .174
	Job self-concept	**.711** (.075)	
	Job satisfaction	**.807** (.058)	
7 imp & int	**Global esteem**	**.072** (.024)	206.994 (21) .000
	Job self-concept	**.127** (.032)	
	Job satisfaction	**.134** (.037)	

Note. See Figure 1 and Table 1 for a depiction of the models. Multiple R^2 is the variance in each of the global outcomes explained by predictor variables (see Table 7). The Wald test is based in the difference between the full model (Model 1) and each subsequent model based on all three outcome measures. For Models 1 and 2, and each of the subsequent models when path coefficients constrained to be zero are replaced with covariances, $\chi^2(1986) = 2,873$, root mean square error of approximation = .029; comparative fit index = .942; Tucker–Lewis Index = .921.

Although Model 1 is able to explain marginally more variance in each of the global outcomes than Model 2, the differences in the variance components in the two models are all less than one standard error for each of the outcomes (Table 6), suggesting that the differences are due to capitalization on chance. In summary, results based on these job satisfaction data provide no support for the IWAM. Methodologically, analyses of these job satisfaction data have several distinctive features that are likely to have broad generalizability to other applications. Thus, for example, although the use of ipsative importance ratings poses statistical complications, they are easily incorporated into our IWAM taxonomy. Also, the approach used to construct interaction terms provides a practical solution to the use of the productindicator approach when there are large numbers of specific rating items. Finally, the study demonstrated how two separate ESEMs (one for specific actual rating factors and one for global rating factors) can be incorporated into a single model such that the two sets of

factors based on each set of items are not confounded. Substantively, the results provide a clear lack of support for the IWAM.

TABLE 7

Selected Individually Weighted-Average Models for the Job Satisfaction Study: Three Global Outcomes (Global Esteem, Job Self-Concept, and Job Satisfaction), Three Sets of Predictor Variables (Satisfaction, Importance, and Satisfaction-by-Importance Interactions), and Seven Domains

Model Outcome	Satisfaction Ratings							Importance Ratings					Satisfaction × Importance Interactions						
	pay	Prof	Auto	Organ	Task	Intr1	Intr2	pay	Prof	Organ	Auto	Intr	pay	Prof	Auto	Organ	Task	Intr1	Intr2
0 Esteem	−.018	**.365**	**.193**	.130	.109	**.209**	**.191**	.061	−.027	**−.083**	**−.085**	.094	.006	−.107	.017	.018	−.006	−.062	.016
Nurs SC	**.166**	**.821**	**.179**	**.251**	**.142**	**.219**	**.213**	**.276**	−.082	−.041	−.067	−.037	−.036	−.100	−.045	.041	**.084**	−.017	.012
Job sat	**.173**	**.865**	**.254**	**.305**	**.356**	**.358**	**.290**	**.208**	−.030	−.040	−.058	−.043	−.068	−.152	−.025	.081	.047	−.003	.012
1 Esteem	−.134	**.310**	.041	.013	.003	.061	.106	−.065	−.069	**−.202**	**−.187**	.026	.065	−.075	.074	.049	−.037	−.099	.057
Nurs SC	−.022	**.902**	−.156	.147	−.044	−.074	.028	.012	−.070	.011	−.050	−.038	.048	.049	.021	−.059	−.023	.023	.017
Job sat	−.048	**.865**	−.164	.070	**.182**	.058	.042	−.045	−.036	−.036	−.048	−.056	.032	−.009	.052	.031	−.058	.032	−.021
2 Esteem	−.127	**.326**	.039	.039	−.005	.051	.085	.005	−.026	**−.063**	−.071	.091	.040	−.061	.037	.012	−.042	−.052	.014
Nurs SC	−.026	**.883**	−.167	.120	−.043	−.063	.024	.036	−.067	.026	−.028	−.015	.048	.047	.007	−.025	−.019	.017	.012
Job sat	−.053	.850	**−.185**	.075	**.181**	.066	.041	−.007	−.013	.008	−.014	−.021	.021	−.008	.025	.014	−.049	.016	−.009

Note. See Figure 1 and Table 1 for a depiction of the models, factors, and variables. Coefficients shown in bold are statistically significant ($p < .05$). Shaded parameters represent covariances, whereas those not shaded are path coefficients. Also see Table 6 for summary statistics based on these and other models.

DISCUSSION: OVERVIEW, STRENGTHS, WEAKNESSES, AND DIRECTIONS FOR FURTHER RESEARCH

Dating back at least to the time of William James, psychologists and other applied researchers have posited heuristic, intuitive, theoretical models based at least implicitly on IWAMs. The IWAM is so intuitively appealing that applied researchers continue to argue for this model even when empirical support for it is largely nonexistent. Following from James's initial proposal in relation to self, there is a particularly long and at times controversial research literature on the IWAM in self-concept research. Nevertheless, despite repeated claims and counterclaims, there now exists a clear operationalization of tests of IWAMs (Figure 1), which suggests a lack of support for IWAMs in self-concept research. Furthermore, even a cursory review of the application of IWAMs in the study of quality of life and job satisfaction reveals that many of the issues identified in self-concept research, as well as the lack of support for IWAMs, generalizes to these other areas of research as well. Indeed, given the prevalence of the issue, it is perhaps surprising that there have not been more cross-citations to these related issues and similar findings across different areas of research. We suggest that one of the reasons might be that applied researchers in these fields of research have concentrated on narrowly focused issues and jargon that is idiosyncratic to that area, rather than seeking a broader, more generic methodological framework. Hence, our overarching purpose is to propose a general taxonomic paradigm for testing IWAMs, and to demonstrate its application across simulated and real data applications.

Issues Arising From the Application of the Taxonomic Paradigm for Testing IWAMs

The simulated data clearly demonstrate that the taxonomic paradigm for testing IWAMs (Figure 1 and Table 1) works for ideal data. For simulated data designed to support the IWAM the results clearly support it, and for simulated data designed not to support the IWAM the results clearly do not support it. We then applied this taxonomic approach to testing IWAMs to three diverse data sets reflecting appropriately "messy" real data. There were different methodological and statistical complications in the construction of appropriate measurement models in each of these data sets. Indeed, many of these issues and the proposed resolutions are likely to be of interest to applied researchers more generally as well as those testing IWAMs. Nevertheless, in each instance once an appropriate measurement model was constructed, application of the taxonomy and tests of the IWAMs were straightforward. From this perspective of facilitating the application of this taxonomic paradigm, it is also relevant to discuss some of the issues raised in the analyses presented here as well as strengths, limitations, and possible directions for further research.

Individually and Normatively Weighted-Average Models

In the applied and substantive research literatures into WAMs, the most persistent source of confusion is the difference between IWAMs that are our focus and what we refer to as normative WAMs. One possible interpretation of a WAM is that the paths leading from actual factors to global constructs (i.e., $\beta1$–$\beta3$ in Figure 1) are not identical. Although potentially interesting and easily tested within the framework of the taxonomic paradigm (i.e., Model 2 with equality constraints), this issue is largely irrelevant to tests of the IWAM. This is clear from a statistical perspective in that the existence or nonexistence of first-order "main" effects (that are the focus of normative WAMs) says nothing about the existence or nonexistence of interaction effects (that are the focus of IWAMs). From a more substantive perspective, the essence of the IWAM is that the relative contribution of the actual factor for a given domain differs systematically from individual to individual as a function of how important that domain is to a particular individual (i.e., the effects of actuals is moderated by importance). In contrast, in Model 2 (i.e., actuals only; $\beta1$–$\beta3$ in Figure 1) the weight for a given actual factor is exactly the same for all individuals; the weights might or might not vary for different factors but are constant across individuals for any one actual factor. Thus whether these paths are the same or different has no bearing on support for IWAMs. In this sense, a potentially important contribution of the taxonomic paradigm is to clarify this distinction and eliminate this widespread source of confusion between normative WAMs and IWAMs.

Tests of Latent Interactions

Interaction effects (i.e., β7–β9 in Figure 1) are a critical feature of the taxonomic paradigm, but also of relevance to applied and substantive researchers more generally. Although widely applied (and sometime misunderstood) in relation to manifest models, there are important advantages and additional complications in testing latent interactions. Here we used the product-indicator approach (see Figure 1) to construct latent interaction factors. Although there are alternative approaches to testing latent interactions (see overview by Marsh et al., 2013; Marsh et al., 2006; Marsh et al., 2012), the two most widely used approaches are the product-indicator approach used here and the LMS approach available in Mplus (as well some other specialized software packages). The choice of the product-indicator approach was, in part, pragmatic in that the LMS is numerically intensive and not feasible when there are more than two or three latent interactions, as will typically be the case in IWAMs (see Supplemental Appendix 1D, an LMS analysis of the simulated data).

Although somewhat tangential to this investigation, we also note that for higher order polynomials (e.g., quadratic and cubic effects) in combination with interactions or higher order interactions that involve more than two latent variables, the adverse effects due to measurement error are likely to be much larger than those with simple latent interaction because these errors aggregate multiplicatively. Research into these polynomials and higher order interactions, however, is limited. Indeed, we suspect that it would be difficult to test these with product-indicator models and might be impossible with the LMS approach as currently specified, but might be possible within the evolving Bayesian framework (Marsh et al., 2012). A complication with the product-indicator approach is how to match multiple actual indicators with multiple importance indictors. This is relatively straightforward in some applications (e.g., Studies 1 & 2), and there is a wide variety of different strategies that can be used when the matching is not straightforward. In Study 4 (job satisfaction data) we demonstrated a new approach to this problem in which the interactions were defined in relation to factor scores reflecting each specific actual factor. In this study, there was only a single importance indicator for each domain. However, if there had been multi-item importance factors, it would also be possible to construct factor scores representing both the actual and importance factors that could be used in creation of the interaction terms. This two-step approach clearly is desirable from a parsimony perspective, and also consistent with the suggestion that interaction terms should use all of the items but not reuse any of the items (Marsh, Wen, & Hau, 2004). Nevertheless, because this pragmatic approach is apparently new, there is need for further application to evaluate its appropriateness across a variety of situations and its relative effectiveness in relation to other approaches. We also note,

however, that once the appropriate measurement model has been established (which includes interaction terms), the taxonomic approach used here is appropriate no matter how the latent interactions are constructed.

Applied researchers attempting to detect latent interactions also need to be cautious in interpreting the goodness of fit of these models, which often is used as a very important (or even the sole) criterion in assessing the hypothetical model of the relationships. Using goodness of fit for models with latent interactions becomes complicated due to a number of issues. First, as in other SEM models, goodness of fit reflects the fit of the whole model and thus is not sensitive to any change related to the latent interaction paths alone. Second, the saturated and null models for the product-indicator models and LMS approaches are not well defined, thus making fit indexes difficult to interpret. Note also that conventional fit indexes as well as likelihood tests are not sensitive to nonlinear relations between latent variables (Mooijaart & Bentler, 2010). Thus, as illustrated in Study 1 with simulated data, a model with no interactions provided an exceptionally good fit to the data even though the taxonomy of models in Table 2 showed that the latent interactions were highly significant and substantively meaningful. In the approach used here, goodness of fit was an important consideration in the evaluation of the initial measurement model that was the basis of subsequent tests of IWAMs but not for a priori substantive tests of predictions based on the IWAMs. A good-fitting measurement model might or might not support IWAM predictions, and support or nonsupport of IWAM predictions does not undermine support for the fit of the measurement model. Finally, the use of highly sophisticated statistical tools such as SEMs can mislead otherwise knowledgeable applied researchers into thinking that well-known problems that exist with less complicated approaches are no longer relevant. Although we have focused on important advantages to the use of latent variable approaches to these problems, the reader should not think that the latent variable approaches are a panacea for all of the complications in evaluating interaction and nonlinear effects. Theoretical and empirical research (e.g., Aiken & West, 1991) demonstrates that there are many inherent difficulties in estimating interaction effects, even when manifest (nonlatent) variable approaches are used. Misspecified models could be a problem. Also, even when the parameter estimates are well defined, there is a range of approaches to visualize interaction effects (and their substantive importance) in manifest applications that have not been routinely implemented in latent interaction studies. These include, for example, testing the statistical significance of simple slopes (i.e., whether the effect of one predictor variable is statistically significant at a certain value of the other predictor variable; see Figure 2), regions of significance (i.e., the range of

the moderator values for which the relation between X1 and Y are statistically significant), and, when one predictor is categorical and the interaction is disordinal, the crossing point of the two regression lines (see, e.g., Preacher, Curran, & Bauer, 2006). Furthermore, the analysis of interaction effects becomes more problematic when the relation between a given independent variable and the dependent variable is nonlinear or when the correlations among the interacting variables become increasingly large. There is no reason to suggest that problems and strategies such as these are not also relevant for latent variable approaches.

ESEM Versus CFA

The IWAM can be tested with either ESEM or CFA, and the decision should be largely empirical based on a comparison of the measurement models using the two approaches. If the fit of the CFA model is not meaningfully worse than the corresponding ESEM, and the factor correlations based on the two approaches are similar, then the CFA should be preferred in terms of parsimony. Thus, for example, in Study 1 based on data simulated from an ICM-CFA model (with all cross-loadings constrained to be zero), CFA provides a more parsimonious solution that is essentially the same as the ESEM solution, and thus is preferable. However, a growing body of research with real data shows that the CFA models with no cross-loadings generally result in poor fits to the data and positively biased factor correlations that undermine the distinctiveness of the latent factors (Marsh, 2014; Morin et al., 2013). Study 4 also demonstrated an interesting variation on the traditional ESEM model in that two separate ESEM structures were incorporated into the same analysis. This was important in terms of not contaminating the outcome factors with crossloadings from the actual factors. However, this feature of ESEM would be particularly useful in studies where there are multi-item scales for both actual and importance factors. Thus, for example, it would be possible to have separate ESEMs for actual and importance factors (and even allow correlated uniquenesses for matching actual and importance items) but still constrain importance items not to load on actual factors and actual items not to load on importance factors (e.g., Marsh, Nagengast, et al., 2011). These issues are particularly important here because of the potential overlap in the different global measures of well-being and the items designed to measure them. More generally, ESEM is a useful tool to reduce the typically positively biased correlations among specific actual factors and among specific importance factors typical in IWAM studies. Again, however, it is important to emphasize that the choice between ESEM or CFA is related to which approach provides the most appropriate measurement model, whereas the application of taxonomic paradigm is essentially the same for either of these approaches.

Summary of Substantive Issues

It might seem surprising that the lack of support for IWAMs is so consistent across the three real data studies, given the diversity of research areas, the issues in each area of research, and complications specific to each of the applications considered here. However, this result is apparently consistent with research in each of these different areas. Substantively, these results call into question the usefulness of the assumptions underpinning IWAMs that have been the implicit or explicit basis of so much research in psychology as well as the social sciences more generally. Nevertheless, across many different fields of application, there seems to be ongoing confusion about how to formulate appropriate tests of IWAMs, appropriate statistical approaches to use in their tests, and appropriate interpretation of the results. In this respect, the failure to find support for IWAMs should be seen as a strength rather than a limitation of the taxonomic paradigm. Indeed, this is why it was important to demonstrate that this approach provided clear support for the IWAM based on simulated data, even when there seems to be little or no support for IWAMs based on real data. Nevertheless, our intent is not to claim that there is never support for IWAMs, but merely to provide a common framework for applied researchers to use in testing this suggestion. However, it will be interesting to see if future research is able to support IWAMs using the taxonomic paradigm approach or, indeed, if reanalyses of previous research claiming to support IWAMs actually do support it when evaluated in relation to the taxonomic paradigm approach.

Although the focus of this article is on the methodology used to test IWAMs, it is also relevant to speculate on possible substantive explanations for the failure to support IWAMs. Although there is clearly variation in the importance ratings the respondents give to different domains, this might provide support for normative WAMs (in which the rank-ordering of importance is similar across different respondents) rather than IWAMs emphasized here. It is also possible that importance ratings are not valid indicators of actual importance. Clearly the importance ratings have face validity and when there are multiple indicators of each importance factor, there is psychometric evidence that the multiple indicators form a latent factor. However, Hattie and Fletcher (2005) found that there was only weak agreement among a range of different nomographic and idiographic approaches used to assess importance (ranking, identity scenarios, magnitude scaling, personal importance ratings, paired comparisons, and the Brunswick lens modeling) and no support for IWAMs based on any of them. They concluded that importance is so elusive and difficult to articulate in

measurement terms that it might not be possible to conceive of self-concept as a weighted average of specific components of self to form self-esteem.

The problem might also be with the global outcomes being used in which participants make ratings based on irrelevant information (see Schwarz & Strack, 1999; in relation to subjective well-being and discussion of the Chameleon effect by Marsh & Yeung, 1999) rather than the rational aggregation of different components implicit in IWAMs. However, some of the outcome measures have been widely validated and not all the outcomes require participants (even implicitly) to combine multiple components (e.g., positive and negative emotions, anomie, and depression in Study 3). Although all these possibilities and many others have been considered (for further discussion see Hattie & Fletcher, 2005; Marsh, 1995, 2008; Pelham, 1995a; Scalas et al., 2013; Scalas et al., 2017; Scalas et al., 2014) even advocates of the Jamesian perspective such as Pelham (1995b) acknowledged that if "James were around today, I suspect that he might feel that is has been embarrassingly difficult for us to uncover support for one of his simplest psychological insights" (p. 1165). In his classic review of weighted-average approaches, Wainer (1976) concluded that humans are not very good at differentially weighting constructs so that simple unweighted averages consistently outperform weighted averages based on human judgment. This led him to conclude that the best way to weight constructs is to simply ignore the weights because in relation to choosing optimal weights "it don't make no nevermind" (colloquially meaning "it makes no difference"; p. 213). From Wainer's perspective, perhaps it is not so surprising that support for IWAMs is so weak. Weighted averages based on human judgment have a long history of not working, so why should this one be any different? Although there seems no completely satisfactory explanation for the failure of IWAMs, the methodology proposed here provides a more solid base to pursue these substantive concerns.

REFERENCES

Aiken, L. S., & West, S. G. (1991). Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage.

Andrews, F. M., & Withey, S. B. (1976). Social indicators of well-being. New York, NY: Plenum. Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). The quality of American life: Perceptions, evaluations, and satisfactions. New York, NY: Russell Sage Foundation.

Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, interactive. Psychological Review, 51, 292–303. doi:10.1037/h0057299

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling, 14, 464–504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 233–255. doi:10.1207/S15328007SEM0902_5

Cowin, L. S. (2002). The self-concept of nurses and its relationship to job satisfaction and retention (Unpublished PhD thesis). University of Western Sydney, Sydney, Australia.

Cowin, L. S., Johnson, M., Craven, R. G., & Marsh, H. W. (2008). Causal modeling of self-concept, job satisfaction, and retention of nurses. International Journal of Nursing Studies, 45, 1449–1459. doi: 10.1016/ j.ijnurstu.2007.10.009

Hardy, L., & Moriarty, T. (2006). Shaping self-concept: The elusive importance effect. Journal of Personality, 74, 377–402. doi:10.1111/j.1467- 6494.2005.00379.x

Harter, S. (1985). Competence as dimensions of self-evaluation: Toward a comprehensive model of self-worth. In R. Leahy (Ed.), The development of self (pp. 55–122). New York, NY: Academic Press.

Harter, S. (1986). Processes underlying the construction, maintenance and enhancement of self-concept in children. In J. Suls & A. Greenwald (Eds.), Psychological perspectives on the self (Vol. 3, pp. 136–182). Hillsdale, NJ: Erlbaum.

Harter, S. (2006). The development of self-esteem. In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 144–159). New York, NY: Psychological Press. Hattie, J. (2003). Getting self-concept back on the correct pathway for selfconcept research in the new millennium: Revisiting misinterpretations and revitalizing the contributions of James' agenda for research on self. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), International advances in self research (Vol. 1, pp. 127–150). Greenwich, CT: Information Age.

Hattie, J., & Fletcher, R. (2005). Self-esteem = success/pretensions: Assessing pretensions/importance in self-esteem. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), International advances in self research (Vol. 2, pp. 123–152). Greenwich, CT: Information Age. Hsieh,

J. Y. (2013). Impact of individual and organizational factors on job satisfaction: A comparison of multilevel models and multiple regression models using different data arrangements. Journal of Management & Organization, 19(1), 44–59. doi: 10.1017/jmo.2013.3

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1–55. doi:10.1080/10705519909540118

Jackson, D. J., & Alwin, D. F. (1980). The factor analysis of ipsative measures. Sociological Methods & Research, 9, 218–238. doi:10.1177/ 004912418000900206

James, W. (1963). The principles of psychology. New York, NY: Holt, Rinehart & Winston. (Original work published 1890)

Jöreskog, K. G. (1998). Interaction and nonlinear modeling: Issue and approaches. In R. E. Schumacker & G. A. Marcoulides (Eds.), Interaction and nonlinear effects in structural equation modeling (pp. 239–250). Mahwah, NJ: Erlbaum.

Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. Journal of Applied Psychology, 87, 530–541. doi: 10.1037/0021-9010.87.3.530 Judge, T. A., Weiss, H. M., Kammeyer-Mueller, J. D., & Hulin, C. L. (2017). Job attitudes, job satisfaction, and job affect: A century of continuity and of change. Journal of Applied Psychology, 102, 356– 374. doi: 10.1037/apl0000181

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin, 96, 201–210. doi:10.1037/0033-2909.96.1.201

Kernis, M. H. (Ed.). (2006). Self-esteem issues and answers. New York, NY: Psychological Press.

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika, 65, 457–474. doi:10.1007/BF02296338

Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. Multivariate Behavioral Research, 42, 647–673. doi:10.1080/ 00273170701710205

Locke, E. A. (1976). Organizational behavior: Affect in the workplace. Annual Review of Psychology, 53, 279–307.

Marsh, H. W. (1986). Global self-esteem: Its relation to specific facets of self-concept and their importance. Journal of Personality and Social Psychology, 51, 1224–1236. doi:10.1037/0022-3514.51.6.1224

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. doi:10.1037/0022-3514.65.5.975

Marsh, H. W. (1994). The importance of being important: Theoretical models of relations between specific and global components of physical self-concept. Journal of Sport and Exercise Psychology, 16, 306–325. doi:10.1123/jsep.16.3.306

Marsh, H. W. (1995). A Jamesian model of self-investment and self-esteem: Comment on Pelham (1995). Journal of Personality & Social Psychology, 69, 1151–1160. doi:10.1037/0022-3514.69.6.1151

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? Journal of Personality & Social Psychology, 70, 810–819. doi:10.1037/0022-3514.70.4.810

Marsh, H. W. (2007). Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology. Leicester, UK: British Psychological Society.

Marsh, H. W. (2008). The elusive importance effect: More failure for the Jamesian perspective on the importance of importance in shaping selfesteem. Journal of Personality, 76, 1081–1122. doi:10.1111/j.1467- 6494.2008.00514.x

Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), Handbook of self-concept (pp. 38–90). Oxford, UK: Wiley.

Marsh, H. W., & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. Contemporary Educational Psychology, 32, 151–171. doi:10.1016/j.cedpsych.2006.10.008

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), Contemporary psychometrics: A festschrift for Roderick P. McDonald (pp. 275–340). Mahwah, NJ: Erlbaum.

Marsh, H. W., Hau, K.-T., Wen, Z., Nagengast, B., & Morin, A. J. S. (2013). Moderation. In T. D. Little (Ed.), The Oxford handbook of quantitative methods (Vol. 2, pp. 361–386). New York, NY: Oxford University Press.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. Psychological Assessment, 22, 471-491.

Marsh, H. W., Morin, A. J. S., Parker, P., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. Annual Review of Clinical Psychology, 10, 85–110. doi:10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. Structural Equation Modeling, 16, 439–476. doi:10.1080/10705510903008220

Marsh, H. W., Nagengast, B., Morin, A. J. S., Parada, R. H., Craven, R. G., & Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. Journal of Educational Psychology, 103, 701–732. doi:10.1037/a0024122

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. Psychological Assessment, 22, 366–381. doi:10.1037/a0019225

Marsh, H. W., & Sonstroem, R. J. (1995). Importance ratings and specific components of physical self-concept: Relevance to predicting global components of self-concept and exercise. Journal of Sport & Exercise Psychology, 17, 84–104. doi:10.1123/jsep.17.1.84

Marsh, H. W., Wen, Z., & Hau, K. T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), Structural equation modeling: A second course (pp. 225–265). Greenwich, CT: Information Age.

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. Psychological Methods, 9, 275–300. doi:10.1037/1082-989X.9.3.275

Marsh, H. W., Wen, Z. L., Hau, K.-T., Little, T. D., Bovaird, J. A., & Widaman, K. F. (2007). Unconstrained structural equation models of latent interactions: Contrasting residual- and mean-centered approaches. Structural Equation Modeling, 14, 570–580. doi:10.1080/10705510701303921

Marsh, H. W., Wen, Z., Nagengast, B., & Hau, K. T. (2012). Structural equation models of latent interaction. In R. H. Hoyle (Ed.), Handbook of structural equation modeling (pp. 436–458). New York, NY: Guilford.

Marsh, H. W., & Yeung, A. S. (1999). The lability of psychological ratings: The chameleon effect in global self-esteem. Personality and Social Psychology Bulletin, 25, 49–64. doi:10.1177/0146167299025001005

Mooijaart, A., & Bentler, P. M. (2010). An alternative approach for nonlinear latent variable models. Structural Equation Modeling, 17, 357– 373. doi: 10.1080/10705511.2010.488997

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling: An introduction. In G. R. Hancock & R. O. Mueller (Eds.), Structural equation modeling: A second course (2nd ed., pp. 395–436). Greewich, CT: Information Age.

Mruk, C. J. (2006). Defining self-esteem: An often overlooked issue with crucial implications. In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 10–15). New York, NY: Psychological Press.

Muthén, L. K., & Muthén, B. (2015). Mplus user's guide. Los Angeles, CA: Muthén & Muthén.

O'Brien, E. J., Bartoletti, M., & Leitzel, J. D. (2006). Self-esteem, psychopathology, and psychotherapy. In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 26–35). New York, NY: Psychological Press.

Owens, T. J., & McDavitt, A. R. (2006). The self-esteem motive: Positive and negative consequences for self and society. In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 281–287). New York, NY: Psychological Press.

Pelham, B. W. (1991). On confidence and consequence: The certainty and importance of self-knowledge. Journal of Personality and Social Psychology, 60, 518–530. doi:10.1037/0022-3514.60.4.518

Pelham, B. W. (1993). The idiographic nature of human personality: Examples of the idiographic self-concept. Journal of Personality and Social Psychology, 64, 665–677. doi:10.1037/0022-3514.64.4.665

Pelham, B. W. (1995a). Further evidence for a Jamesian model of selfworth: Reply to Marsh (1995). Journal of Personality and Social Psychology, 69, 1161–1165. doi: 10.1037/0022-3514.69.6.1161

Pelham, B. W. (1995b). Self-investment and self-esteem: Evidence for a Jamesian model of self-worth. Journal of Personality and Social Psychology, 69, 1141–1150. doi:10.1037/0022-3514.69.6.1141

Pelham, B. W., & Swann, W. B., Jr. (1989). From self-conceptions to self-worth: On the sources and structure of global self-esteem. Journal of Personality and Social Psychology, 57, 672–680. doi:10.1037/0022-3514.57.4.672

Preacher, K. J., Curran, P., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. Journal of Educational and Behavioral Statistics, 31, 437–448. doi: 10.3102/10769986031004437

Quinn, R. P., & Mangione, T. W. (1973). Evaluating weighted models of measuring job satisfaction: A Cinderella story. Organizational Behavior & Human Performance, 10(1), 1–23. doi: 10.1016/0030-5073(73)90002-0

Reise, S. P. (2012). Invited paper: The rediscovery of bifactor measurement models. Multivariate Behavioral Research, 47, 667–696. doi:10.1080/ 00273171.2012.715555

Rhodewalt, F. (2006). Possessing and striving for high self-esteem. In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 281–287). New York, NY: Psychological Press.

Rosenberg, M. (1957). Occupations and values. New York, NY: The Free Press.

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.

Scalas, L. F., Marsh, H. W., Morin, A. J. S., & Nagengast, B. (2014). Why is support for Jamesian actual–ideal discrepancy model so elusive? A latent-variable approach. Personality and Individual Differences, 69, 62–68. doi: 10.1016/j.paid.2014.05.010

Scalas, L. F., Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Latent-variable approaches to the Jamesian model of importanceweighted averages. Personality and Social Psychology Bulletin, 39, 100–114. doi: 10.1177/0146167212465321

Scalas, L. F., Marsh, H. W., Vispoel, W., Morin, A. J. S., & Wen, Z. (2017). Music self-concept and self-esteem formation in adolescence: A comparison between individual and normative models of importance within a latent framework. Psychology of Music. Advance online publication. doi:10.1177/0305735616672317

Scalas, L. F., Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2014). Importance models of the physical self: Improved methodology supports a normative- cultural importance model but not the individual importance model. European Journal of Social Psychology, 44, 154–174. doi: 10.1002/ejsp.2001 Schimmack, U. (2008). Measuring wellbeing in the SOEP. SSRN Electronic Journal. doi:10.2139/ssrn.1306888

Schimmack, U., Diener, E., & Oishi, S. (2002). Life-satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. Journal of Personality, 70, 345–384. doi:10.1111/jopy.2002.70.issue-3

Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. Journal of Personality and Social Psychology, 89, 395–406. doi:10.1037/0022-3514.89.3.395

Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), Well-being: The foundations of hedonic psychology (pp. 61-84). New York, NY: Russell Sage.

Showers, C. J., & Zeigler-Hill, V. (2006). Pathways among self-knowledge and self-esteem: How are self-esteem and self-knowledge linked? Are these links direct or indirect? In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 216–223). New York, NY: Psychological Press.

Stamps, P. (1997). Nurses and work satisfaction: An index for measurement (2nd ed.). Chicago, IL: Health Administration Press.

Tevendale, H. D., & Dubois, D. L. (2006). Self-esteem change: Addressing the possibility of enhancing improvements in feelings of self-worth. In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 170–177). New York, NY: Psychological Press.

Transitions from Education to Employment. 2013a. Codebook TREE Survey Panel 2001 (Release 2013). Basel, Switzerland: Author.

Transitions from Education to Employment. 2013b. Concepts and scales: Survey Waves 1−8, 2001–2010. Basel, Switzerland: Author. Retrieved from

http://www.tree.unibe.ch/e206328/e305140/e305154/

files305246/TREE_2013_Concepts_Scales_English_ger.pdf(29.02. 2016)

Transitions from Education to Employment. (2013c). TREE project documentation 2000– 2012. Basel, Switzerland: Author. http://www.tree.

unibe.ch/e206328/e305140/e305154/files305251/TREE_2013_e_ Project_docu_2000- 2012_ger.pdf

Vonk, R. (2006). Improving self-esteem. In M. H. Kernis (Ed.), Self-esteem issues and answers (pp. 178–187). New York, NY: Psychological Press.

Vroom, V. H. (1964). Work and motivation. New York, NY: Wiley.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 83, 213–217. doi:10.1037/0033- 2909.83.2.213

Wen, Z., Marsh, H. W., & Hau, K.-T. (2010). Structural equation models of latent interactions: An appropriate standardized solution and its scale-free properties. Structural Equation Modeling, 17, 1–22. doi:10.1080/ 10705510903438872

Wu, Y., Wen, Z., Marsh, H. W., & Hau, K.-T. (2013). A comparison of strategies for forming product indicators for unequal numbers of items in structural equation models of latent interactions. Structural Equation Modeling, 20, 551–567. doi:10.1080/10705511.2013.824772