Available online at www.sciencedirect.com

ScienceDirect

Journal of CONSUMER PSYCHOLOGY

Journal of Consumer Psychology 20 (2010) 208 – 214

Methods Dialogue

Structural equation models are modelling *tools* with many ambiguities: Comments acknowledging the need for caution and humility in their use

Richard P. Bagozzi

University of Michigan, Ann Arbor, MI, United States

Recived 10 March 2010; accepted 10 March 2010
Available online 10 April 2010

Abstract

My goal is to provide background and perspective on the use and interpretation of structural equation models (SEMs). SEMs are complex procedures with many assumptions, intricacies, and pitfalls. I hope to give a commentary that complements the primers done by Iacobuci and deepen the users' knowledge of the procedures. But I acknowledge that this effort is at best an incomplete introduction into SEMs and cannot do justice to the many issues (and controversies) associated with it.

© 2010 Society for Consumer Psychology. Published by Elsevier Inc. All rights reserved.

> Better for us, perhaps, it might appear,Were there all harmony, all virtue here;That never air or ocean felt the wind,That never passion discompos'd the mind.But all subsists by elemental strife;And passions are the elements of life.The gen'ral order, since the whole began,Is kept in nature, and is kept in man.
> Alexander Pope,"Essay on Man" (1732)

Dawn Iacobuci provides two primers that should prove useful for researchers new to structural equation modeling (Iacobucci, 2009, 2010). My aim will be to comment on her many suggestions by giving background, clarifications, cautions, and insights not mentioned by her. To make things easy to keep track of, I will simply follow the order of topics covered by Iacobucci: first in "Everything You Always Wanted to Know..." (Iacobucci, 2009) and then in "Structural Equations Modelling..." (Iacobucci, 2010). However, nothing will be said about straightforward or less controversial issues treated by her in the interests of brevity.

Everything You Always Wanted to Know...

Motivation for SEMs

Iacobucci points out that SEMs take into account measurement error and that regression of a dependent variable on a factor measured by, say, 3 indicators overcomes problems of multicollinearity that might arise if one were to regress the dependent variable on the 3 indicators directly. Of course, if the 3 indicators were really redundant measures of one concept or construct, it might be better to regress the dependent variable on one variable formed by the average of the indicators, for the case where regression analysis is used. SEMs go further than this practice and better take into account measurement error.

It is useful to consider two kinds of multicollinearity and how SEMs might overcome common problems in multiple regression models. Imagine that we wish to represent the social identity of customers with a company and see its effects on both customer extra role behaviors (e.g., recommending the company's products to other customers) and customer purchases. Social identity has been hypothesized to consist of three distinct components: cognitive identity (sometimes termed, identification), emotional identity (often called, affective commitment), and evaluative identity(alternatively characterized as collective or group esteem). With multiple measures of each component, multiple regressions of extra role behaviors and purchases on the measures would likely affect the precision of regression parameter estimates and obscure true effects and even possibly change the signs of true effects.

To deconstruct the issue of multicollinearity in this specific case, consider Fig. 1. Here the 3 components or dimensions of customer social identity with a company are first-order social identity factors. Extra-role behaviors and purchases are then, in a

E-mail address: bagozzi@umich.edu.

1057-7408/$ - see front matter © 2010 Society for Consumer Psychology. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.jcps.2010.03.001

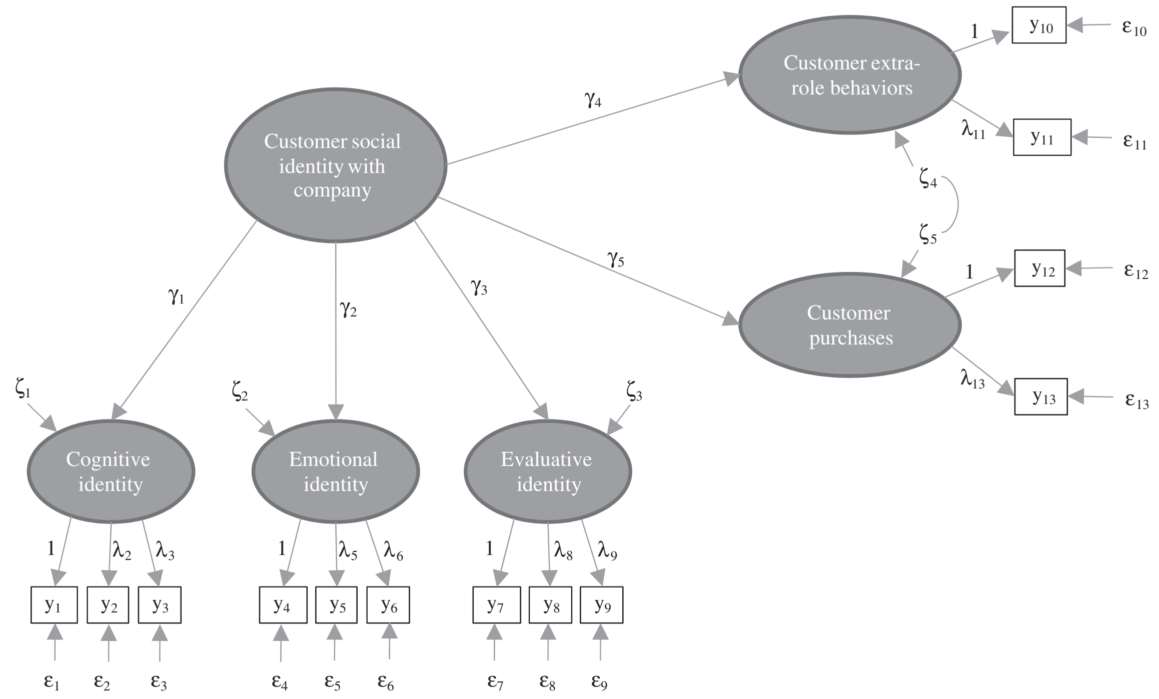

Fig. 1. Two kinds of multicollinearity and a second-order confirmatory factor analysis approach for predicting dependent variables.

sense, regressed on social identity. How are multicollinearity problems circumvented here? Well, one type of multicollinearity occurs as redundancy in multiple measures of the *same* construct (s). Cognitive, emotional, and evaluative identities each have 3 measures. To regress extra-role behaviors and purchases directly on these measures, as in multiple regression, would likely result in multicollinearity problems. The 3 first-order identity factors not only eliminate these problems but at the same time take into account the unreliability present in measures y_1–y_9. However, to the extent that the 3 first-order factors are highly correlated (in the sense discussed in the following paragraph), regression of extra-role behaviors and purchases directly on these 3 factors would also likely lead to multicollinearity problems. This second kind of multicollinearity happens because of redundancy across measures of *different* constructs, where the constructs are highly correlated due to their nature (in this case because they represent similar, but distinct, mental events or states).

As shown in Fig. 1, extra-role behaviors and purchases are regressed on a single independent or exogenous variable, social identity, and multicollinearity is not an issue. Social identity gets its meaning indirectly through the measures of its 3 components, and directly predicts or influences the dependent variables. It should be noted, speaking loosely, that multicollinearity is likely to occur whenever the measures of independent variables correlate more highly amongst themselves than they do with the measures of the dependent variables.

The second-order factor approach shown in Fig. 1 is most valid and conceptually meaningful when the first-order factors loading on the second-order factor can be interpreted as subdimensions or components of a more abstract, singular construct. Moreover, the first-order factors should be relatively highly and similarly intercorrelated. When either or both of the above mentioned conceptual and empirical criteria are not met, the second-order factor approach to multicollinearity will not be justified.

Causal Models?

The next topic Iacobucci addresses is causality, where she notes that the issues for SEMs are especially daunting because multiple dependent variables are frequently investigated and hence multiple causal claims made in any application. Indeed, because researchers using SEMs often interpret relationships between exogenous and endogenous latent variables as causal relationships, and at the same time, assert that relationships between latent and manifest variables are also causal, the causal claims are particularly acute and in need of discussion.

Consider first the point of view that relationships among exogenous and endogenous latent variables are causal. Iacobucci seems to conclude that causality is allowable to claim under experimental conditions, but in the more common case where correlations are relied upon, she believes one should term such relationships, predictive. I tend to agree that the use of SEMs in experimental contexts allows one to speak of causality, at least we might say that the relationships in such cases "approach" causality in meaning. But even in the experimental case, a number of issues deserve consideration. First, in any experiment, there are likely to be one or more threats to validity, some unknown or unaddressed, and as a consequence, a failure to reject the null hypothesis (i.e., the specified model) perhaps is at best fallible and tentative evidence for causality. Second, among the many competing and arguably incomplete interpretations of causation in the philosophy of science, one that appears to best fit the meaning of causality for the types of experiments done in consumer psychology is termed, the manipulability model. It is not feasible to discuss this perspective on causation in depth here, but I wish to point out that the assumptions and implications of this model constrain the meaning of causation and at the same time harbor advantages and disadvantages vis-a-vis other perspectives. So it seems to me that even under experimental conditions, it is best not to be overly sanguine that one is truly observing causality, and further, depending on which of the many competing models of causality one embraces, it is important to acknowledge that non-experimental approaches (e.g., so called quasi-experiments, cross-sectional and longitudinal surveys, even qualitative research methods) might satisfy some requirements for causality better than experimental methods in certain instances (see Bagozzi, 1980, Ch. 1, for an introduction to competing models of causation). All of this is simply to say that, in particular studies, it will not in general be indisputably clear that an experimental approach accords with all criteria for causation better than a non-experimental approach. It all depends on what model of causality is followed and how well the competing methods satisfy the criteria under the models. Note that SEMs can be and have been applied to experimental data, so we are not, in these comments, necessarily comparing ANOVA analyses of traditional experimental data to SEM analyses of augmented experimental data (augmented by multiple measures of manipulation checks, covariates, and dependent variables, say).

There is a further issue worthy of comment concerning inferences of causality between latent variables. Consider the following definition of causality, which is perhaps general enough so as not to run afoul of most of the competing models of causality. Causality is

> “the relation between two events that holds when, given that one occurs, it produces, or brings forth, or determines, or necessitates the second; equally we say that once the first has happened the second must happen or that the second follows on from the first... [furthermore, causation] suggests that states of affairs or objects or facts may also be causally related.” (Blackburn, 1994, p. 59).

Most philosophers of science seem to regard causality as something we infer between physical or material entities or changes therein, hence reference above to “events,” “states of affairs,” “objects,” “facts,” But latent variables are abstractions or unobservables and are nonmaterial, though we sometimes hope they “represent” or “capture” variance observed in manifest variables. The “causal” parameter one derives from estimation of SEMs (γ’s and β’s in LISREL notation) are inferred statistics from relationships amongst (material) manifest variables (i.e., observations or measures of events, states of affairs, etc.). The parameter estimates of causal relationships between latent variables might be best construed as imperfect, fallible signs of whatever causal process one is studying that occurs between the measures of causes and effects. SEMs may be specified to correct for random and systematic errors, and thus γ’s and β’s may be purged of such errors, but it is important to recognize that causality is at best estimated (inferred) from the data and predicated on the nature of the data and methodological conditions applied in any study. Researchers using SEMs, even under the control of experimental conditions, should not go too far in suspending conceptual, empirical, and methodological beliefs and assumptions, and over claiming causality.

Now to the claim that the relationship between a latent variable and its manifest or measured variables is causal. This is a longstanding point of view exposited early-on by Blalock, Bollen, Heise, and many others from the 1970s and continuing to the present by Edwards and Bagozzi (2000), Jarvis, MacKenzie and Podsakoff (2003), and nearly every treatment of the relationship in the literature. I now regard my use of “causal language” (e.g., Edwards & Bagozzi, 2000) to characterize the relationship between latent and manifest variable as being misguided. It seems to me that the relationship in question is not causal, per se, but rather one of hypothetical measurement. That is, the relationship is between an abstract, unobserved concept and a concrete, observed measurement hypothesized to measure the concept; the relationship is part logical, part empirical, and part theoretical (conceptual), with the inferred factor loading representing, in and of itself, only part of empirical meaning of the relationship (see Bagozzi, 1984, 2007, 2010, and discussion of “correspondence rules” therein).

In sum, SEMs represent different relationships that require a healthy application of interpretation. Perhaps most generally, relationships amongst latent exogenous and endogenous variables are best construed as imperfect representations of *causal relationships*, with those founded on experimental data coming closest to achieving the designation or accolade, causal, and those arising from survey research given less credence as causal. Further, cross-sectional survey data in this regard might be better interpreted as yielding evidence for *functional relationships* or alternatively relationships believed to be consistent with causal relationships as far as they go but not sufficiently strong enough to suggest causality to the degree that experiments do. Longitudinal survey data potentially support stronger interpretations than cross-sectional data but weaker interpretations than experimental data in the typical case. To avoid categorical thinking leading to either overly strong claims of causality for experimental research or premature dismissal of survey research as giving no support whatsoever for causal claims, I believe that it is best to think about a causal-like continuum, marked by relatively strong (experimental) and relatively weak (“functional relationships” in surveys) labels as endpoints, and longitudinal surveys somewhere in between strong and weak. Another method with intermediate claims of causal credence might be field or quasi-experiments, where somewhat less control than pure experiments is afforded (e.g., testing hypotheses across multiple groups in naturalistic settings or between groups formed fortuitously akin to controlled experiments). SEMs apply in all these cases and suggest differing bases and different degrees of evidence for concluding causality, with the possibility that some sub-criteria for causality might be better met in naturalistic field experiments and longitudinal designs than found in pure experiments in certain albeit relatively rare cases.

What about the designation, predictive? If a study tests a theory and exogenous and endogenous variables are linked significantly according to the theory, I think we might term the relationship, *explanatory* (e.g., ξ explains η), and then decide whether or not, or to what degree, if any, causality can be claimed. When the exogenous and endogenous variables are separated in time, the relationship might be called an "*explanatory prediction.*" But I prefer to use the term, *prediction*, by itself to suggest the case where an existing theory happens to lead to the forecast or discovery of a new phenomenon or outcome. This latter usage is consistent with some philosophy of science characterizations of what constitutes a (strong) theory. That is, a theory that explains what it is supposed to explain is given less acclaim than one that also leads to new discoveries or predictions. To keep this important distinction concerning theories, it seems best to speak of explanatory prediction and prediction, as pointing to still another continuum for interpretive purposes.

The Measurement Model: Confirmatory Factor Analysis

Iacobucci mentions that nonsignificant loadings on a factor may occur for measures that, in fact, measure other factors or alternatively are simply poor measures of the factor and could be dropped. I wish to mention a relatively common incidence falling under her general observation. Sometimes two or more loadings are high in value on a factor, whereas two or more other loadings are low (but still significant). It may be the case that the measures associated with the low loadings are simply inadequate measures of the factor and therefore might be deleted from further analysis. But it might also be the case that the measures associated with the low loadings actually measure another factor, not originally specified, that is significantly correlated with the originally hypothesized factor. Indeed, if the measures associated with the low loadings were originally proposed to measure one factor, along with the measures associated with the high loadings, then it may be the case that the proposed factor (and hence the hypothesized concept) is multidimensional, not unidimensional, and two factors could be modeled to capture this case. Of course, the multidimensionality of the hypothesized concept should make sense and be interpretable, if the multiple factors are to be retained.

The Structural Model: Path Analysis

Iacobucci notes that when a hypothesized path turns out to be nonsignificant in a path analysis this is "a diagnostic clue that the model may be mis-specified and the theory needs re-thinking." This is a plausible conclusion to be sure, yet it may be alternatively the case that the theory could have merit, but the particular data at hand fail to support it. Further testing may be needed or a deeper analysis required of the data and context of study to verify whether the test was a valid one or whether some threat to validity was at fault.

The Full SEM Model…

Modification indexes (MIs) are the last topic considered by Iacobucci in the first primer. I agree with her that caution must be used when relying on MIs and wish to add two points. First, in any run, only the single highest MI is (potentially) interpretable; interpreting the next highest MIs in the same run could be misleading, so it is best to wait to the next run, after relaxing the constraint associated with the highest MI on the first run, to examine MIs and inspect the highest MI again. Second, if one has a good fitting model (before examining any MIs), then it is possible that inspecting MIs and relaxing a constraint will produce nonsensical or misleading findings (e.g., out of range parameter estimates, illogical signs for parameters, theoretical or empirical contradictions). Generally, further improvements to an existing good-fitting model risk capitalizing on chance. Again, sound theoretical and/or methodological criteria should guide the use of MIs and their interpretation.

Structural equations modeling…

Fit indices

At this time, the generally recognized and recommended fit indexes to present are the χ^2-test, *df*, and *p*-value; RMSEA; NNFI (or TLI); CFI (or RNI); and the SRMR, where Hu and Bentler (1998, 1999) suggest that for a "good" model, the χ^2-test should be non-significant with $p \geq 0.05$, RMSEA ≤ 0.06, NNFI ≥ 0.95, CFI ≥ 0.95, and SRMR ≤ 0.08. My (highly subjective) experience suggests that the ≤ 0.08 criterion for the SRMR is too liberal and that values ≤ 0.07 may be better to rely upon, because as the average of all residuals, the SRMR risks yielding a "satisfactory" value when one or more residuals are "too high" but get overlooked when averaged with all others. An SRMR value ≤ 0.07 at least reduces this possibility somewhat in comparison to one $\leq .08$. On the other hand, some have argued that the other cut-off values are too conservative under certain conditions (e.g., Marsh, Hau & Wen, 2004), suggesting, for example, that values of the NNFI and CFI less than 0.95 may be meaningful (e.g., $\geq .90$). Last, although one might be tempted to rely on only two goodness-of-fit indexes (see Hu & Bentler's, 1999, two-index strategy: i.e., "use the SRMR plus one other index"), it seems best to examine all 4 indexes noted above (plus of course the χ^2-test; see Fan & Sivo, 2005).

A common outcome in everyday research is that the χ^2-test is significant, and one or more of the 4 indexes is unacceptable under the rules of thumbs mentioned above. What should one do in such circumstances? The meaning of a significant χ^2-test is particularly critical, for as Iacobucci notes, it is the only statistical test amongst the fit indexes provided by LISREL and other similar programs. With very large samples, because the χ^2-test is proportional to sample size, there is a danger of rejecting a valid model; with small sample sizes, there is a danger of accepting an invalid model, on the basis of the χ^2-test. But what is too large or too small a sample? (see brief discussion in the next section of the paper). Until this issue and the meaning of the χ^2-test in relation to SEMs become more resolved, I would (tentatively) recommend that one rely on the other "fit" indexes (i.e., RMSEA, NNFI, CFI, SRMR) when a significant χ^2-test results and the sample size is "large." To build-in some (minimal) safeguards, for most modeling, it would be advisable to use Hu and Bentler's (1999) more conservative criteria (except for the SRMR as noted above, where it should be ≤ 0.07) and to (ideally) hope that all 4 indexes are "acceptable." This risks deviating from scientific and statistical practice (in that the results of the χ^2-test are ignored and the meaning and validity of the 4 indexes are in need of further study), but this recommendation seems more demanding than current practice and is consistent with the spirit of use of the proposed fit indexes (which may be flawed). A problem occurs when one or more fit indexes are "unacceptable"; here one might bring other issues to bear. For example, a value for the RMSEA somewhat above the $\leq$0.06 standard might be allowed, if a model has a small number of degrees of freedom and the sample size is "small"; or for MTMM matrix models where models are specified to be complex a priori (i.e. they are non-parsimonious by definition), the RMSEA and NNFI, which penalize for model complexity, might be discounted. Also if one or so indexes is "unacceptable" one might allow this in "exploratory" research, when the indexes are near, but below, the borderline of recommended index standards. All this is fuzzy and unscientific, but the state of the art at the moment. In most cases, we do not really test models, unfortunately, but rather fit them to data, and this weakens the validity of our research.

Sample size

Iacobucci suggests that a minimum sample size is 50, that "samples of 50 to 100 can be plenty," and that rules of thumb (e.g., $n \geq 200$) are "simplistic." There is an arbitrariness to rules of thumb concerning sample size. Pressed, I would have to say that very rarely would a sample size below 100 or so be meaningful, and that one should endeavor to achieve a sample size above 100, preferably above 200. In any case, other issues may be more important under certain circumstances. A critical issue is the distributional properties of measures (and the error terms of equations), because the frequently used maximum likelihood (ML) procedure requires multivariate normality. Kurtosis is especially a problem in this regard, but the ML estimation technique has been shown to be fairly robust to departures from normality. So even with a relatively small sample size, the ML procedure may be satisfactory, if the distributional properties of measures (and error terms?) are satisfactory or not too far out of range. Things get worse when the ratio of sample size to number of parameter estimates is too small; Bentler's recommendation of "5 to 1, preferably 10 to 1" is a conservative one in my opinion, as I have found satisfactory models with ratios near 2 to 1 on occasion. Again the distributional properties of measures are key, not sample size, per se.

Different data scenarios

It is not true that one can incorporate moderating variable hypotheses, with SEMs containing latent variables, by using procedures that are the same as found in multiple regression. To use SEMs to model interactions, one must specify complex models and reparameterize the models so to speak to take into account nonlinearities in parameter estimates (except in the case of multi-sample analyses which exhibit their own limitations, as developed below). For example, in a study of the self-regulation of dieting, my colleagues and I investigated a model wherein resistance to eating temptations, a latent variable, interacted with felt subjective norms, another latent variable, to influence intentions to diet, amongst other determinants (e.g., Bagozzi, Moore & Leone, 2004). To specify this model, we had to correlate 4 error terms amongst manifest measures and allow the measures of the latent moderator, which were products of measures of the latent variables entering the interaction, to load on 3 factors each. A number of nonlinear constraints on parameters had to be made to achieve a correct specification, including 3 factor loadings and 3 error covariances that were the products of other parameters, two error terms that were each the sum of 5 terms containing products of parameters and squares of parameters, and the variance of the latent moderator equal to the product of the latent variables entering the interaction plus the square of the covariance between the latter (see Bagozzi et al., 2004, for the specification and LISREL input program implementing it). This model fit the data well based on the 4 fit criteria mentioned above (RMSEA = 0.04, NNFI = 0.99, CFI = 0.99, SRMR = 0.04), but not on the basis of the $\chi^2(64, N=609)=129.23$, $p<0.01$; however, it seems appropriate to accept the model, given the large sample size and the "relatively" small χ^2-value. Substantively, a significant interaction effect was found (γ_3=-.16, SE = 0.08, $t=-2.11$), which implies that the impact of felt normative pressure to diet on intentions to do so, increases the greater the ability to resist temptation to eat in various unhealthy ways (the negative sign was due to reverse coding of resistance to temptation). In sum, this example points to the need for caution and care in specifying moderating variable effects in SEMs, as the ramifications are different than in classic multiple regression analysis.

There is another way to conduct tests of moderating variable effects in SEMs: namely, by use of multiple groups. For example, if one desired to test the moderating effect of gender on the attitude to behavior relationship and the sample size was too small to use asymptotic distribution free methods, then one might regress latent behavior on latent attitude in equations capturing this specification separately for men and women (and include other antecedents and consequences, if appropriate). A comparison of regression parameters across gender groups (after first demonstrating that measurement invariance occurs across gender) would demonstrate that gender moderates the effects of attitude on behavior, if the latent variable regression parameter was significantly different between men and women. This approach has the disadvantage of not taking into account all information available in the data. If no significant interaction is found, then we cannot rule-out the possibility that this was due to a failure to take into account enough information. On the other hand, when significant differences are found for parameters across groups, one might conclude that the interaction effect implied has some merit.

Iacobucci also considered longitudinal data. Fig. 2 shows a respecification of her Fig. 3 in "Structural Equations Modeling…" Although her model can be estimated and tested, the respecification avoids problems with contemporaneous inferences (i.e., interpreting the effects of cognitions at time 1 on affect at time 1; cognitions at time 2 on affect at time 2, which may be hard to justify in surveys), permits the implementation and test of a cross-lagged panel design (i.e., the effects of both cognition at time 1 on affect at time 2, and affect at time 1 on cognition at time 2, which provide sounder ways to test causality than contemporaneous designs), and allows one to test all direct and indirect effects of

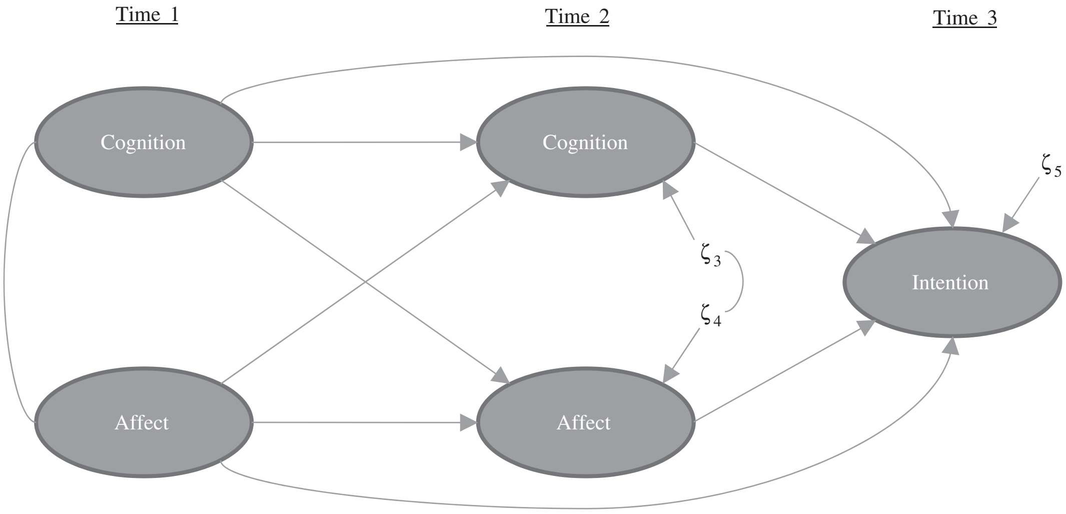

Fig. 2. Respecification of repeated measures example.

cognition and affect on intentions (which is a sort of full-information operationalization of Baron and Kenny's piece-meal test of mediation).

Other longitudinal designs are possible as well. SEMs make feasible the modeling of first- and higher-order autoregressive or Markov models (termed Simplex models by psychometricians). The fullest models and the most informative ones require 4 or more time points. Note, too, that it is possible to run circumplex models with SEMs. In all these models, SEMs offer the advantage of representing and correcting for measurement error.

Perhaps the most controversial, yet timely, topic broached by Iacobucci concerns the use of reflective versus formative indicators. Formative indicators have a long history, going back to the 1960s, but recently have come under attack, with Howell, Breivik and Wilcox (2007) taking the position that formative indicators are fatally flawed and should never be used. The controversy is over the meaning of formative measurement and a number of confounds seemingly inherent in commonly used formative applications. The issues are complex and not fully resolved; and matters are muddied because the formative model is intuitive, relatively easy to specify and program, and will run in most instances, leaving many users and readers with the impression that such models are appropriate and valid. One shortcoming with formative models that is becoming generally recognized is that classic conceptions of internal consistency reliability and construct validity (convergent and discriminant validity) do not apply. With regard to the general formative approach, I have taken a somewhat less polarized point of view than others and maintained that it may have a place in research from the point of view of the limited canonical correlation model analogue in SEMs (see Bagozzi, 2007, 2010; Bagozzi, 1980). However, the aforementioned limitations of formative measurement and the confounds recently identified should be taken seriously, with caution advised in the use of formative models (Bagozzi, 2007).

The final issue mentioned by Iacobucci that I wish to touch upon is the use of PLS. This, too, is a contentious topic and one still in a state of flux. I wish to mention that when sample sizes are small, PLS may be the only viable alternative. In some applications, too, PLS yields identical or near identical results as SEM analyses, as we found in our treatment of experimental data by both LISREL and PLS (Bagozzi, Yi & Singh, 1991). On the other hand, and in addition to the problems Iacobucci mentioned, one might warn that the weighted sums of observed variables that PLS produces and uses, in a manner analogous to incorporation of factor scores, will yield biased regression parameters. Note, however, that PLS has been the model of choice of researchers publishing in information systems journals, which is a practice that appears to be more a matter of tastes than anything else, because many of the treatments in IS could have been done with SEMs. I am inclined to recommend against blanket favoritism for or exclusion of PLS and plead for tolerance amongst reviewers and readers, as long as the assumptions undergirding the approaches are reasonably met.

Conclusion

I commend Iacobucci for the service she has done to the field in preparing these primers on SEMs. She has given us comprehensive coverage of the topics and many suggestions and concerns that we all should heed. The only major topic not covered by Iacobucci and me is method bias (and the closely related issue of construct validity). This topic is becoming more central, not only in consumer psychology journals, but also in organizational behavior and other basic and applied social science journals, where some editors have begun insisting on proof that method biases do not contaminate findings before publishing research. It appears that a convergence across fields is building in this regard (for primers on the issues, see Bagozzi, 2010; Bagozzi et al., 1991; Bagozzi, Yi, & Nassen, 1999).

In conclusion, it seems best to approach SEMs with a certain amount of awe and trepidation, not so much because of the methods, per se, but more because of our own limitations, tastes, and prejudices. SEMs are important methodological tools for consumer researchers, but it is important to realize that they complement such other procedures as ANOVA, multiple regression, and qualitative methods, amongst others. The use of SEMs and interpretation of findings there from are best done dialectively with a healthy measure of skepticism and tentative confidence in their validity. Or as Alexander Pope (1732) said in his "Essay on Man,"

> All nature is but art, unknown to thee;
> All chance, direction, which thou canst not see;
> All discord, harmony not understood;
> All partial evil, universal good.
> And, spite of pride, in erring reason's spite,
> One truth is clear, 'Whatever is, is right.'[1]

References

Bagozzi, R. P. (1980). *Causal Models in marketing.* New York: Wiley.

Bagozzi, R. P. (1984). A prospectus for theory construction in marketing. *Journal of Marketing*, *48*, 11−29.

Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox. *Psychological Methods*, *12*, 229−237.

Bagozzi, R.P. (2010). Measurement and meaning in information systems and organization research. *MIS Quarterly*, forthcoming.

Bagozzi, R. P., Moore, D. J., & Leone, L. (2004). Self-control and the self-regulation of dieting decisions: The role of prefactual attitudes, subjective norms, and resistance to temptation. *Basic and Applied Social Psychology*, *26*, 199−213.

Bagozzi, R. P., Yi, Y., & Singh, S. (1991). On the use of structural equation models in experimental designs: Two extensions. *International Journal of Research in Marketing*, *8*, 125−140.

Bagozzi, R. P., Yi, Y., & Nassen, K. D. (1999). Representation of measurement error in marketing variables: review of approaches and extension to three facet designs. *Journal of Econometrics*, *89*, 393−421.

Blackburn, S. (1994). *The Oxford dictionary of philosophy.* Oxford, England: Oxford University Press.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155−174.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indices to misspecified structural or measurement model components: Rational of two-index strategy revisited. *Structural Equation Modeling*, *12*, 343−367.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, *12*, 205−218.

Hu, Li-tze, & Bentler, P. M. (1998). Fit indexes in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424−453.

Hu, LI-tze, & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1−55.

Iacobucci, D. (2009). Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask. *Journal of Consumer Psychology*, *19*, 673−680.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, *20*, 90−98.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*, 199−218.

Marsh, H. W., Hau, Kit-Tai, & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320−341.

[1] Because Pope's essay has generated a lot of controversy over the years, I wish to briefly and elliptically give my own interpretation of these verses, so as not to create more ambiguity than necessary with respect to SEMs. My intent in ending with these verses is to suggest that there is some truth behind appearances in SEMs that we may never fully know or hope to completely clarify and that it is best to begin with and hold on to some measure of humility, lest we fall into the trap of leading ourselves and others astray. My aim in this commentary has been twofold: first more to convey the complexity of issues concerning SEMs than to resolve them, per se, and second to suggest that the problems with SEMs are fundamental and only some solutions are now within our research.