Published in final edited form as:

Psychol Methods. 2017 September ; 22(3): 605–608. doi:10.1037/met0000149.

Notes on measurement theory for causal-formative indicators: A reply to Hardin

Kenneth A. Bollen

Department of Psychology & Neuroscience, University of North Carolina.

Adamantios Diamantopoulos, Faculty of Business, Economics & Statistics, Department of Business Administration, University of Vienna.

Abstract

The current paper is a rejoinder to “A Call for Theory to Support the Use of Causal-Formative Indicators: A Commentary on Bollen and Diamantopoulos.” Our paper comments on the six research questions raised by Hardin (2017, this issue) in his constructive commentary on our original paper (i.e., In Defense of Causal-Formative Indicators: A Minority Report).

In the conclusion to our paper “In Defense of Causal-Formative Indicators: A Minority Report” (Bollen & Diamantopoulos, in press), we noted the calls in the literature for a moratorium on causal-formative indicators. We instead invited more discussion and constructive exchange of views and were delighted to see Andrew Hardin’s (2017, this issue) commentary that takes up the challenge. While not agreeing with all that we said, he has set a positive tone and asks valuable questions to help further advance this important research area.

We organize our responses to the issues he raises according to the questions posed in his commentary:

“Research Question 1: *How should researchers assess the reliability of indicators that do not have an associated error term?*”

The classic definition of reliability is that it is the ratio of variance of the latent variable to the total variance of an indicator (Lord & Novick, 1968; Nunnally & Bernstein, 1994). In a structural equation modeling context, this definition necessarily assumes a reflective (effect) indicator and this is evident both in the nature of the error term in reflective measurement models (i.e. error is conceived as a *residual* resulting from regressing the indicator on the latent variable) as well in the measures used to assess indicator reliability (i.e. the squared multiple correlation associated with the indicator). In other words, the conventional reliability definition is practically inseparable from the implicit idea that the indicators depend on the latent variable and it is silent on the situation when the indicator affects the latent variable.¹

Causal-formative indicators assume that the indicator itself directly affects the latent variable of interest and it has long been acknowledged that this property is *not* consistent with classical test theory (CTT) and conventional notions of reliability (e.g. see Namboodiri, Carter & Blalock, 1975, p. 70; Fornell, Rhee & Yi, 1991, p. 317). If a particular indicator is an imperfect representation of a variable that directly affects the latent variable, then a model of this relationship should acknowledge that. Specifically, in the presence of measurement error each causal-formative indicator should be modeled as a function of an underlying “perfect” (i.e. errorless) latent variable plus an error term.

To make this point concrete, suppose the “amount of social interaction” is a latent variable (L_1).² Now suppose that we have a causal-formative indicator of “time spent with friends” (x_1) that is intended to measure L_1 . If “time spent with friends” is an accurate measure, then we can allow x_1 to directly affect L_1 . Alternatively, suppose that we suspect that the time measure is inaccurate and suffers from random error such that $x_1 = L_2 + \varepsilon_1$ where L_2 is a completely accurate time measure and ε_1 is a random error with mean zero and uncorrelated with L_2 .³ In this latter case, we introduce an error term, ε_1 , for x_1 and now have L_2 directly affect L_1 . If all the causal-formative indicators (e.g., time spent with family, time spent with workmates, etc.) of amount of social interaction (L_1) contain error in a similar fashion, then the researcher should allow for that possibility. The resulting model is the higher-order factor counterpart to choosing reflective or causal-formative indicators for the first-order factors; that is, we would have a set of first-order factors and a second-order factor. The question is whether the first-order factors directly affect the second-order factor or vice versa as in a traditional second-order factor analysis.

In brief, if the causal-formative indicators directly affect the latent variable, then there is no error term for the indicators because it is the measure that has the direct effect. If, however, the indicator is a faulty representation of another variable that in turn affects the focal latent variable, then the indicator will have an error term and the latent variable that it measures will have the direct causal effect on the latent variable of interest. Thus we cannot see why “the specification of these separate latent variables changes the interpretation of the causal-formative indicators” as Hardin (2017, this issue) maintains. The interpretation of the indicators stays the same; it is only the model specification that is adjusted to take into account measurement imperfections (see Bollen & Ting, 2000; Diamantopoulos, 2006; Edwards & Bagozzi, 2000).

“Research Question 2: What are the implications of a disturbance term that can capture sources of influence related to the empirical meaning of the latent variable and the theoretical definition of the construct?”

First of all, the fact that “the missing and included causal-formative indicators are determinants of the same latent variable” does not automatically imply that the missing indicators *will* be correlated with the included ones. Second, the error term is composed of a

¹In the factor analysis literature, the error includes a specific component unique to each indicator (referred to as uniqueness or specificity) and this could be considered part of reliability. However, we do not pursue that issue here because it is not central to the current discussion.

²See Bollen & Lennox (1991) and Bainter & Bollen (2014) for further discussion of the causal indicators used in this example.

³We make the usual assumption of the error being uncorrelated with the latent variable throughout this discussion.

large number of individual omitted influences on the latent variable since it captures “all remaining causes on η (i.e. the latent variable) not represented by the indicators” (Diamantopoulos, 2011, p.336). As is the case with any multiple regression model, we assume that this error term is uncorrelated with the causal-formative indicators included in the measurement model specification. The key issue is therefore whether the error term that results when *all* these individual omitted influences are combined is correlated with the included causal-formative indicators, not whether a *specific* missing indicator is correlated with those included. Third, it could be the case that all conceptually-relevant indicators have been included but “there still may be some error left as a result of non-linear or interaction effects among the indicators” (Diamantopoulos, 2006, p. 16). Fourth, given the multiple regression specification of the causal-formative measurement model, there is always a possibility of an “omitted variables” problem and the advice here is to be careful at the indicator specification stage. In this context, and contrary to common belief, a similar problem may beleaguer reflective specifications.

Although the fact that simple (i.e. bivariate) regressions are used to link a reflective (effect) indicator to a single latent variable, it could well be the case that part of the variation in the indicator is also due to another latent variable which is correlated with the focal latent variable but has *not* been explicitly included in the model. This omission would bias the loading of the indicator as well as the interpretation of the latter. This is another good example of how a problem supposedly only applying to causal-formative indicators has an analogous counterpart for reflective (effect) indicators.

“Research Question 3: *Is the identification of a census of causal-formative indicators achievable in practice?*”

As a starting point in replying to this question, we look at the idea of a “census” of causal-formative indicators as it originated in Bollen & Lennox (1991). These authors were comparing two models. The first was a traditional one factor, four reflective (effect) indicator factor analysis model (page 306, Figure 1a). The second was a one latent variable, four causal indicator model (Figure 1b). One issue treated in the article is what happens to these two *specific* models if an indicator is removed. Bollen & Lennox (1991) stated that the remaining factor loadings in the effect indicator model should remain the same in the population. However, with the causal indicators in Figure 1b, removal of one indicator could bias the coefficients of the remaining causal indicators and their effect on the latent variable. The exception would be when the omitted causal indicator is uncorrelated with the remaining indicators. It is in the context of Figure 1b that the statement is made that a “census” of causal indicators is needed. That is, in general, all four indicators from Figure 1b form the census and all four are needed to avoid biasing the coefficients of the remaining causal indicators.

This is a far cry from requiring a comprehensive list of *all possible* causal indicators be included in a model, not least because “in practice, it can be very difficult to be *certain* that all possible causes on the latent variable are accounted for by the indicators” (Diamantopoulos, 2006, p. 12, added emphasis). Hardin’s (2017, this issue) comments following Question 3 suggest that he is interpreting the comment on “census” to mean *all possible* causal-formative indicators be included rather than interpreting the census to mean

a set of causal-formative indicators that are uncorrelated with the (composite) error of the latent variable.

In practice, it seems likely that at least some causal-formative indicators are not explicitly included in the model specification. These omitted causal-formative indicators join together with other omitted influences on the latent variable to form the error term of the latent variable. As noted in our response to Research Question 3 above, consistent estimation of the indicator coefficients requires that the *collection* of variables that form the error is uncorrelated with the included causal-formative indicators. It does not require that *each and every* omitted variable is uncorrelated with the included causal-formative indicators.

“Research Question 4: *Should contemporary definitions of predictive validity include the ability to assess the predictive capabilities of causal-formative indicators?*”

It is not exactly clear what notion of validity Hardin (2017, this issue) has in mind when he speaks of ‘predictive validity’. Predictive validity is not necessarily the same as criterion validity but a subset of the latter (‘concurrent validity’ is another form of criterion validity - see Carmines & Zeller, 1979; Zeller & Carmines, 1980). Moreover, Hardin’s (2017, this issue) statement that “establishment of predictive validity occurs when a latent variable is able to predict other relevant variables” also applies to nomological validity if ‘relevant’ is intended to mean ‘theoretically relevant’.

The above notwithstanding, predictive validity typically refers to the relationship between *two* (distinct) *constructs* (whether measured with reflective or causal-formative indicators) and not to the relationship between a construct and its indicators; as Netemeyer, Bearden & Sharma (2003, p. 76) point out, “predictive validity refers to the ability of a measure to effectively predict some subsequent and temporally ordered criterion”. We are therefore somewhat confused when – referring to causal-formative indicators - Hardin (2017, this issue) states that “the weights associated with these indicators provide important predictive information” and that they “can be individually assessed for their predictive capabilities”. Is Hardin’s (2017, this issue) reference to predictive ability intended to apply to the link between the causal-formative indicators and the focal latent variable that these indicators measure or to the link between the causal-focal indicators and *other* (i.e. downstream) latent variables that are outcomes of the focal latent variable? In the former case, the focus would be on indicator validity along the lines discussed by Bollen (1989; 2011), whereby the direct structural relation between each indicator and the focal latent variable can be interpreted as a validity coefficient. In the latter case, the focus would be on whether proportionality constraints actually hold, that is, whether the impacts of the causal-formative indicators on outcome latent variables are fully mediated by the focal latent variable. While these are two (very) different issues, neither of them poses any obstacles to the assessment of predictive validity at the latent variable level. Whether the predictor latent variable is measured with reflective (effect) indicators or with causal-formative indicators, predictive validity can readily be assessed by examining the magnitude of the structural path with the criterion latent variable and the amount of variance explained in the latter. We therefore fail to see a compelling reason why “future research should revisit classic definitions of predictive validity to ensure that they accurately account for the use of causal-formative indicators”.

“Research Question 5: *Should causal-formative indicators be evaluated using classic definitions of construct validity?*”

Again, we are somewhat confused with Hardin’s (2017, this issue) definition of construct validity as “whether a measure measures what it *purports* to measure”. This is an (over-arching) definition of validity *in general* (e.g. see Carmines & Zeller, 1979, p. 12) rather than construct validity in particular. The latter is typically defined as the extent to which a measure behaves in a *theoretically* sound manner and is further decomposed into convergent validity, discriminant validity and nomological validity (e.g. see Peter 1981)). So the notion of construct validity goes far beyond the interpretational confounding issue raised by Hardin (2017, this issue) (and to which we offered a detailed discussion in our original paper that need not be repeated here).

Having said that, we would like to make just one point in relation to the ‘old chestnut’ argument brought up by Hardin (2017, this issue) that “changes in reflective (effect) indicator loadings (that fall within accepted retention guidelines) are unlikely to alter the empirical meaning of the latent variable, even when estimated under these changing conditions”. If this is the case, then *why* is there an extensive literature on measurement equivalence or invariance that *specifically* focuses on assessing comparability of *reflective* (effect) indicators across different settings/contexts (e.g. countries or respondent types – see Vandenberg & Lance (2000) for a review)? *Why* would one be concerned with such issues as metric invariance or scalar invariance (Steenkamp & Baumgartner 1998; Myers et al. 2000) if the empirical meaning of a (reflectively-measured) latent variable is unlikely to be different across settings? The fact of the matter is that changes in indicator coefficients *can* occur under changing conditions and the resulting differences may indeed raise questions about the stability of the empirical meaning of the focal relevant variable. But this is not an exclusive problem with causal-formative indicators and applies to reflective (effect) indicators too.

“Research Question 6: *Is the use of causal-formative indicators consistent with scientific generalization?*”

In the discussion of this question, Hardin (2017, this issue) wonders whether causal-formative indicators are consistent with scientific generalizations. He suggests that they might not be for several reasons: 1) the lack of an error term influencing the causal-formative indicator, 2) the dependence of the coefficients from such variables to a latent variable on the “context” of the problem, and 3) whether disturbance terms capture “varying sources of influence across different contexts”. With regard to 1), the coefficients of causal indicators are likely biased if the indicators contain sufficient measurement error. This is true of any explanatory variable, so there is nothing special about the causal-formative indicator model. The researcher would need to take account of the error along the lines we discussed under Research Question 1 above. Points 2) and 3) leave the meaning of context ambiguous, but if it refers to their being different populations with different structural parameters, then we would expect coefficients to differ. However, we see no reason to expect that causal-formative indicators are more likely to experience different contexts than are models with reflective (effect) indicators. Although we should be concerned about different contexts

affecting our estimates, this is a general warning applying no matter what the type of indicator is involved (see also Research Question 5 above).

Concluding Thoughts

We thank Andrew Hardin (2017, this issue) for the opportunity to address the questions he raises on causal-formative indicators. Our preceding responses aimed at clarifying the issues associated with the six questions he posed and we hope that we have been successful in doing so.

We end with two points. First, causal-formative indicators indeed have some properties that differ from reflective (effect) indicators. For instance, reflective (effect) indicators typically correlate with each other while causal-formative indicators need not. Relatedly, typical reliability coefficients assessing internal consistency are inappropriate for causal-formative indicators. Thus it is very important for the researcher to be clear as to what kind of indicator is involved in a particular application. While, as extensively discussed in our original paper, the decision whether to use causal-formative or reflective (effect) indicators should be based on theoretical grounds, Bollen & Ting (2000) also offer a procedure for empirically distinguishing between causal-formative and reflective (effect) indicators.

Second, several of the problems attributed as being unique to causal-formative indicators are not. Such problems have their counterparts in reflective (effect) indicators. For instance, measurement invariance in different populations and the potential for omitted variables affecting the relation between indicator and latent variable are potential problems for both causal-formative *and* reflective (effect) indicators. But we do agree with Hardin (2017, this issue) that it is time to bridge the “great divide” and further advance the conceptual and methodological underpinnings of measurement models that include causal-formative indicators.

References

- Bainter Sierra & Bollen KA (2014). Interpretational Confounding or Confounded Interpretations of Causal Indicators. *Measurement: Interdisciplinary Research & Perspectives*, 12, 125–140. [PubMed: 25530730]
- Bollen KA (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bollen KA (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, 35, 359–372. <http://search.proquest.com/docview/869511412?accountid=14682>
- Bollen KA & Diamantopoulos A (in press). In defense of causal-formative indicators: A minority report. *Psychological Methods* (same issue as this manuscript and the commentary by Hardin).
- Bollen KA, & Lennox R (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305–314. 10.1037/0033-2909.110.2.305
- Bollen KA, & Ting KF (2000). A tetrad test for causal indicators. *Psychological Methods*, 5, 3–22. [PubMed: 10937320]
- Carmines EG, & Zeller RA (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage publications 10.4135/9781412985642
- Diamantopoulos A (2006). The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modelling in Management*, 1, 7–17. 10.1108/17465660610667775

- Diamantopoulos A (2011). Incorporating formative measures into covariance-based structural equations models. *MIS Quarterly*, 35, 335–358.
- Edwards JR, & Bagozzi RP (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174. [PubMed: 10937327]
- Fornell C, Rhee BD, & Yi Y (1991). Direct regression, reverse regression, and covariance structure analysis. *Marketing Letters*, 2, 309–320. 10.1007/BF02404080
- Hardin A (in press). A call for theory to support the use of causal-formative indicators: A commentary on Bollen and Diamantopoulos. *Psychological Methods* (same issue as the original and follow-up Bollen and Diamantopoulos papers).
- Lord FM, & Novick MR (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley
- Myers MB, Calantone RJ, Page TJ Jr, & Taylor CR (2000). An application of multiple-group causal models in assessing cross-cultural measurement equivalence. *Journal of International Marketing*, 8, 108–121. 10.1509/jimk.8A108.19790
- Namboodiri K, Carter N, & Blalock LF (1975). *Applied multivariate analysis and experimental designs*. New York, NY: McGraw-Hill.
- Netemeyer RG, Bearden WO, & Sharma S (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage Publications.
- Nunnally JC, & Bernstein IH (1994). *Psychometric theory*. New York, NY: McGraw-Hill.
- Peter JP (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research*, 18, 133–145. 10.2307/3150948
- Steenkamp JBE, & Baumgartner H (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–90. 10.1086/209528
- Vandenberg RJ, & Lance CE (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70. 10.1177/109442810031002
- Zeller RA, & Carmines EG (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge, UK: Cambridge University Press.