

Constructed Measures and Causal Inference

Towards a New Model of Measurement for Psychosocial Constructs

Tyler J. VanderWeele^a

Abstract: Psychosocial constructs can only be assessed indirectly, and measures are typically formed by a combination of indicators that are thought to relate to the construct. Reflective and formative measurement models offer different conceptualizations of the relation between the indicators and what is sometimes conceived of as a univariate latent variable supposed to correspond in some way to the construct. I argue that the empirical implications of reflective and formative models will often be violated by data since the causally relevant constituents will generally be multivariate, not univariate. These empirical implications can be formally tested but factor analysis is not adequate to do so. I argue that formative models misconstrue the relation between the constructed measures and the underlying reality by which causal processes operate, but that reflective models misconstrue the nature of the underlying reality itself by typically presuming that the constituents of it that are causally efficacious are unidimensional. I discuss the ensuing problems arising from these misconstruals. I propose a causal interpretation of associations between constructed measures and various outcomes that is applicable to both reflective and formative models and is applicable even if the usual assumptions of these models are violated. I put forward an outline for a new model of the process of measure construction and discuss the practical implications of these observations and proposals for the provision of definitions, the selection of items, item-by-item analyses, the construction of measures, and the interpretation of the associations of these measures with subsequent outcomes.

Keywords: Causal inference; Constructs; Factor analysis; Formative model; Measurement; Psychosocial epidemiology; Reflective model

(*Epidemiology* 2022;33: 141–151)

Submitted July 1, 2020; accepted October 5, 2021

From the ^aDepartments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

Supported by NIH R01 CA222147.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Tyler J. VanderWeele, Harvard T.H. Chan School of Public Health, Department of Epidemiology, 677 Huntington Avenue, Boston MA 02115. E-mail: tvanderw@hsph.harvard.edu.

Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

ISSN: 1044-3983/22/331-141

DOI: 10.1097/EDE.0000000000001434

The model that dominates classical approaches to measurement and scale development, sometimes referred to as a reflective model, typically presupposes an underlying univariate latent variable that gives rise to measured indicators.^{1–4} The latent variable is then thought to correspond to some psychosocial construct of interest. It is often assumed that the underlying latent variable has causal efficacy,^{4,5} even though all we observe are its indicators.

In this article, I will present some empirical data and theoretical considerations that challenge whether either reflective or alternative so-called formative models are adequate. This will be facilitated by reviewing and deploying recently developed theory for causal inference for multiple version of treatment^{6,7} to develop alternative interpretations of exposure-outcome associations when the exposure used is a scale or index. I will show that the proposed interpretation under multiple version of treatment theory holds for both reflective and formative models, and holds more generally still.^{8–13} I will present an alternative model concerning the relationships between constructs, indicators, measures, and the true underlying constituents of reality, and I will discuss the practical implications of this for the process of measure construction.

CAUSAL INFERENCE UNDER MULTIPLE VERSIONS OF TREATMENT

The theory for causal inference under multiple versions of treatment^{6,7} was originally developed to aid interpretation in settings wherein there was no unambiguous intervention to manipulate an exposure. In such settings, because different manipulations to shift the exposure might result in different effects on an outcome, the counterfactuals or potential outcomes^{14–17} are not well-defined, and thus there is no single quantitative effect of the exposure.^{6,7,18,19} It will be argued below that these issues are relevant for most psychosocial phenomena.

Consider first settings with an unambiguous exposure-intervention. Let A denote the exposure, Y an outcome, and C a set of pre-exposure covariates. Let Y_a denote the potential outcome for Y if exposure A had been set to value a . The causal effect of a binary exposure A on outcome Y is defined, for an individual, by $Y_1 - Y_0$, and for the population by $E[Y_1 - Y_0]$. If the exposure is categorical or continuous, the values 1 and 0 can be replaced by arbitrary values, a and a^* , respectively. We say that the effect of A on Y is unconfounded given C if Y_a

is independent of A conditional on C i.e. if, conditional on C , those with and without the exposure are comparable in their potential outcomes. If this is so and the technical consistency assumption holds that when $A = a$ then $Y = Y_a$, then we have¹⁴⁻¹⁷:

$$E[Y_1 - Y_0] = \sum_c \{E[Y | A = 1, c] - E[Y | A = 0, c]\}P(c).$$

The causal effect can thus be obtained by standardizing conditional observed outcome differences across exposure groups by the proportion in each stratum of C . In practice, this is often obtained by regressing Y on (A, C) :

$$E[Y | a, c] = \beta_0 + \beta_1 a + \beta_2' c.$$

Provided the regression model is correctly specified, the causal effect is then given by: $E[Y_1 - Y_0] = \sum_c \{E[Y | A = 1, c] - E[Y | A = 0, c]\}P(c) = \beta_1$.

Now consider the setting wherein there is not a well-defined intervention on exposure A . Suppose there is some underlying “version of treatment” variable K that takes values among some set \mathcal{K} , and that for each version of treatment $k \in \mathcal{K}$, the version is sufficiently well-defined to correspond to a unique potential outcome Y_k .^{6,18} Suppose the investigator has access only to a coarsened variable A , where each value of A corresponds to one or more values of K . We might then refer to the variable A as a “composite exposure” or “compound treatment” since each value of A can come about through numerous more specific versions of treatment K .^{7,18,19} See Figure 1; the red arrows in this figure, and all subsequent figures, are those emanating from variables, related to the exposure, that are causally efficacious for the outcome.

We say there is no confounding for the effect of K on Y given C if Y_k is independent of K conditional on C . We say

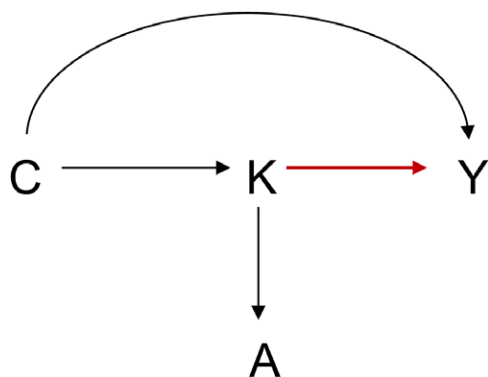


FIGURE 1. A model for multiple versions of treatment wherein the version-of-treatment variable K affects the outcome Y but is confounded by measured covariates C , with the measured exposure variable A representing a coarsening of K (Red arrows in this figure and in all subsequent figures are those emanating from variables related to the exposure that are causally efficacious for the outcome.).

that the consistency assumption holds if when $K = k$ then $Y = Y_k$. Suppose the investigator has only information on (A, Y, C) . Analogous to the formula above, it may seem natural to compute:

$$\sum_c \{E[Y | A = a, c] - E[Y | A = a^*, c]\}P(c).$$

However, it is not clear how to causally interpret this quantity when there are not well-defined interventions on A . Theory for multiple versions of treatment provides an interpretation.^{6,7} It can be shown⁶ that if the effect of K on outcome Y is unconfounded given C and if the consistency assumption holds, then

$$\begin{aligned} & \sum_c \{E[Y | A = a, c] - E[Y | A = a^*, c]\}P(c) \\ &= \sum_{k,c} E[Y_k | c]P(k | a, c)P(c) - \sum_{k,c} E[Y_k | c]P(k | a^*, c)P(c). \end{aligned} \tag{1}$$

The first expression in equation (1) is the empirical quantity we would ordinarily use to estimate effects of A on Y , if multiple versions were not an issue. The second expression provides a causal interpretation. It can be interpreted as a comparison in a hypothetical randomized trial in which, within strata of covariates C , each individual in one arm is randomly assigned a version of treatment K from the underlying distribution of K in the subpopulation with $(A = a, C = c)$, and each in the other arm is randomly assigned a version of treatment K from the underlying distribution of K in the subpopulation with $(A = a^*, C = c)$. An illustration with BMI and mortality is given in the eAppendix; <http://links.lww.com/EDE/B869>. While the original theory⁶ assumed A constituted a coarsened version of K so that the relationship between K and A was a deterministic many-to-one mapping, this assumption can be weakened. As shown in the Appendix and discussed in the eAppendix; <http://links.lww.com/EDE/B869>, beyond unconfoundedness and consistency, the only assumption needed to derive the relation in (1) is that Y is independent of A conditional on (K, C) i.e., conditional on C , A gives no information about Y once K is known. The relationship between K and A then need not be many-to-one and can, moreover, also be stochastic. This will be important in the development that follows.

The multiple versions of treatment theory allow us to make formal progress in interpreting causal effects of composite exposures. There are, however, limitations to this approach. First, when the set of versions of treatment, \mathcal{K} , is unknown, this hinders precise understanding of the interpretation. Second, with the set of underlying versions unknown, it would then effectively be impossible to implement the hypothetical randomized trials embedded within the interpretation. Third, the interpretation will vary depending on what is included in C since, once C is fixed, this may limit the range of potential versions of treatment that are possible. Fourth, with the versions of treatment unknown, it becomes difficult to substantively assess the unconfoundedness assumption and thus to know whether

the proposed interpretation is reasonable. Although the multiple versions of treatment interpretation has limitations, it may be the best we can do concerning a formal potential-outcomes-based interpretation of the quantitative effect estimate of a composite exposure,^{18,19} and may also provide insights into where to focus intervention attempts (see eAppendix; <http://links.lww.com/EDE/B869>). In the next section, we will consider how this interpretation can be applied to measures arising from reflective or formative measurement models.^{8–13}

CAUSAL INFERENCE AND LATENT VARIABLES: REFLECTIVE AND FORMATIVE MODELS

The classical model used in much measurement theory and scale development presupposes an underlying latent continuous variable η that gives rise to measured indicators (X_1, \dots, X_n) as in Figure 2A. After standardization, it is often assumed that each indicator X_i is given by a linear function of η plus random error ϵ_i :

$$X_i = \lambda_i \eta + \epsilon_i. \tag{2}$$

The random errors ϵ_i are often, but not always, assumed independent. This model forms the basis of much psychometric measure evaluation.²⁰ However, after this evaluation is complete, the measures that are used are generally just some function of the indicators (X_1, \dots, X_n) . When the indicators are on the same scale, their mean is often used. Let $A = f(X_1, \dots, X_n)$ denote the measure employed. This is typically considered an imprecise measure of the underlying latent η that corresponds to the psychosocial construct of interest. Interest often then lies in assessing the relationship of this with various outcomes.

To estimate effects, often a regression is fit of Y on (A, C) , assuming relationships depicted in Figure 2B:

$$E[Y | a, c] = \beta_0 + \beta_1 a + \beta_2 'c.$$

Provided the covariates C control for confounding, β_1 is then sometimes interpreted as the causal effect of the exposure on the outcome. Sometimes, especially when structural equation models are used, this estimation is done with

correction for measurement error, using reliabilities λ_i from the measurement model⁴ and the estimate is then interpreted as the causal effect of the latent η corresponding to the underlying construct. When the reliabilities λ_i vary across indicators and this is neglected, as would often be the case if Y were simply regressed on (A, C) , then this interpretation is problematic. However, even in this setting, a causal interpretation of β_1 is possible using multiple versions of treatment theory. Specifically, if we replace K in the previous section with η , and compare measure level $A = a + I$ with $A = a$, then if the effect of η on Y is unconfounded given C , as in Figure 2B, equation (1) above becomes:

$$\beta_1 = \sum_{\eta,c} E[Y_\eta | c] P(\eta | A = a + 1, c) P(c) - \sum_{\eta,c} E[Y_\eta | c] P(\eta | A = a, c) P(c).$$

In other words, even if reliabilities λ_i vary and this is ignored, β_1 can still be interpreted as a comparison in a hypothetical randomized trial in which, within strata of covariates C , individuals in one arm are randomized to a value of η from the actual distribution of η in the subpopulation with $(A = a + I, C = c)$, and individuals in the other arm are randomized to a value of η from its actual distribution in the subpopulation with $(A = a, C = c)$. We can apply the result from equation (1) with K replaced by η because, in Figure 2B, $A = f(X_1, \dots, X_n)$ will be independent of Y conditional on (η, C) .¹⁵ Similar remarks concerning independence pertain to the other results below.

The model in Figure 2A is sometimes referred to as a reflective model because the indicators reflect the underlying latent variable. An alternative model for measurement is sometimes called the formative model^{9–11} and is illustrated in Figure 3A. In this model, the indicators effectively together form the underlying variable of interest, which is a function of the indicators plus error:

$$\eta = \sum_i \lambda_i X_i + \epsilon. \tag{3}$$

In practice, measures are again formed as some function of the indicators $A = f(X_1, \dots, X_n)$. Sometimes it is assumed that there is no error, and the function of the indicators is itself the underlying variable of interest with $\eta = A = f(X_1, \dots, X_n)$.

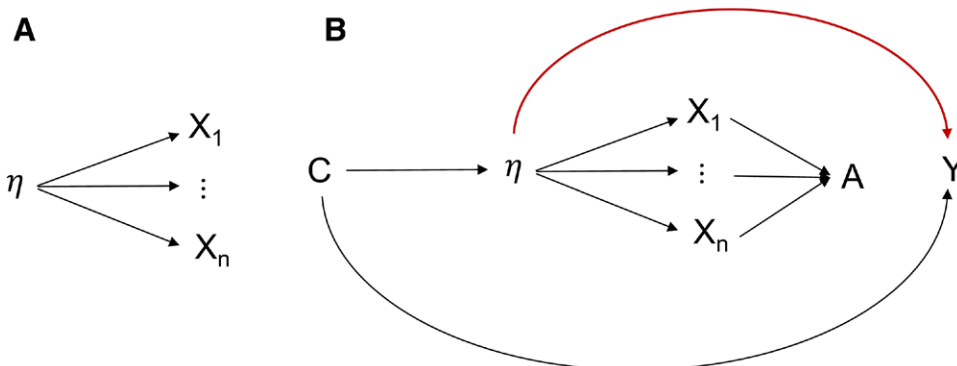


FIGURE 2. A, Basic reflective model with univariate latent variable η giving rise to indicators (X_1, \dots, X_n) . B, Structural reflective model with measure A as a function of the indicators $A = f(X_1, \dots, X_n)$ and with all causal relations concerning the indicators from prior variables C , or outcomes variables Y operating through latent variable η .

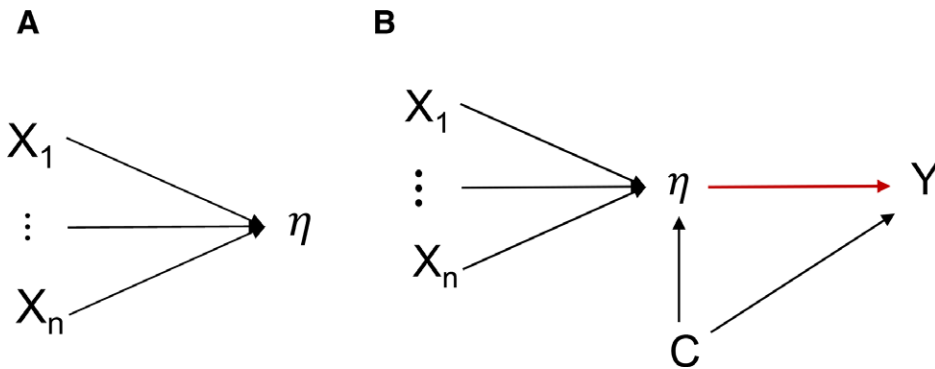


FIGURE 3. A, Basic formative model with the indicators (X_1, \dots, X_n) giving rise to a univariate latent variable η . B, Structural formative model with all causal relations with subsequent outcomes variables Y operating through latent variable η .

Considerations as to whether and when reflective or formative models are more appropriate are described elsewhere, though this continues to be debated.^{9–13}

However, in this case also the causal interpretation under multiple versions of treatment theory is applicable. Provided the effect of η on Y is unconfounded given C , as in Figure 3B, then under the regression model $E[Y|a, c] = \beta_0 + \beta_1 a + \beta_2' c$ we again have:

$$\beta_1 = \sum_{n,c} E[Y_\eta | c] P(\eta | A = a + 1, c) P(c) - \sum_{n,c} E[Y_\eta | c] P(\eta | A = a, c) P(c)$$

and β_1 can be interpreted as under the reflective model.

However, this analysis of reflective and formative models assumed that the latent η was causally efficacious. This may not be the case. Neither in Figures 2A and 3A, nor in equations (2) and (3) require that it is η , rather than (X_1, \dots, X_n) , that is causally efficacious. Consider instead the causal diagrams in Figure 4A and B. These correspond to reflective and formative models but with indicators (X_1, \dots, X_n) , rather than the latent η , having causal effects on outcome Y . Importantly, the causal diagram in Figure 4A is compatible with the reflective model in Figure 2A and with equation (2). The causal diagram in Figure 4B is compatible with the formative model in Figure 3A and equation (3). We might thus distinguish between *basic* reflective and formative models represented in Figures 2A and 3A respectively [and by equations (2) and (3)], versus what we might call *structural*^{15,21} reflective and formative models, represented by Figures 2B and 3B respectively, which additionally assume that all causal relations with (X_1, \dots, X_n) are through η (reflective)^{15,21} or that all effects of (X_1, \dots, X_n) are through η (formative). Both the structural models in Figures 2B and 3B, and also the models in Figure 4A and B, are compatible with the basic formative and reflective in Figures 2A and 3A.

In the causal models in Figure 4A and B, we might consider the effects of each indicator one by one. However, we might also instead consider measures formed as functions of the indicators, $A = f(X_1, \dots, X_n)$. If we regress Y on (A, C) using $E[Y | a, c] = \beta_0 + \beta_1 a + \beta_2' c$, we can again interpret the coefficient β_1 using MVT theory, this time taking K as the set of indicators (X_1, \dots, X_n) . In both Figure 4A and B, the effects of (X_1, \dots, X_n) on Y are unconfounded conditional on C . For

any two values $A = a + 1$ and $A = a$, we can thus interpret β_1 by equation (1) with $K = (X_1, \dots, X_n)$ and $A = f(X_1, \dots, X_n)$. The coefficient β_1 can thus be interpreted as a comparison in a hypothetical randomized trial wherein, within strata of covariates C , individuals in one arm are randomly assigned to values of (X_1, \dots, X_n) from the actual distribution of these indicators in the subpopulation with $(A = a + 1, C = c)$, and individuals in the other arm are randomly assigned to values of (X_1, \dots, X_n) from the actual distribution of these indicators in the subpopulation with $(A = a, C = c)$. The MVT interpretation is again applicable. However, now the interpretation extends to hypothetical interventions on the indicator set (X_1, \dots, X_n) , rather than the underlying latent η .

We are left with the question of which of these causal models is more reasonable. Compatibility of the data with Figure 2A and equation (2), or with Figure 3A and equation (3), tells us nothing as to whether the indicators themselves, or some underlying latent variable, is causally efficacious. The next section presents analyses concerning associations between social integration and health suggesting that, in this case, a model with a causally efficacious univariate latent might not be plausible. Critically, structural formative and reflective models are incompatible with one of the indicators being causally related to the outcome and another not, because, under the structural models, that could only be the case if one of the λ_i were 0, in which case it would not be an indicator of the underlying latent η at all.

SOCIAL INTEGRATION EXAMPLE: QUESTIONING THE MEASUREMENT MODELS

Numerous studies have examined associations between measures of social integration and subsequent health. Social integration has been conceptualized and measured in a variety of ways.^{22,23} However, evidence has been consistent across operationalizations that social participation tends to be associated with better health.²²

Chang et al²⁴ used data from the Nurses Health Study ($n = 76,362$) to examine associations between social integration and incident coronary heart disease (CHD). They used a simplified Berkman-Syme Social Integration Index in 1992 as their exposure (summing indicators, each scored 0–3, of

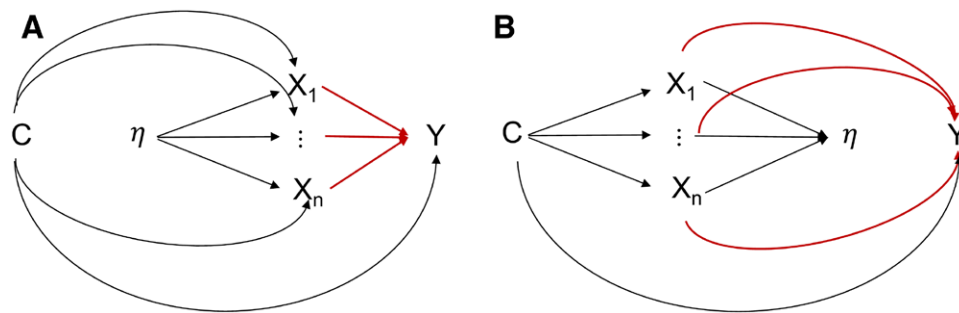


FIGURE 4. A, Basic reflective model but with the causal relations from prior variables C or outcomes Y operating directing through the indicators (X_1, \dots, X_n) rather than the latent η . B, Basic formative model but with causal effects of the indicators (X_1, \dots, X_n) on outcome Y not through the latent η .

religious service attendance, community group participation, number of close friends, and marital status), and followed incident CHD through 2014, employing proportional hazards models. After adjusting for age, education, husband's education, census-tract income, hypertension, diabetes, cholesterol, family MI history, and depressive symptoms, comparing highest versus lowest quartiles of social integration, they estimate a hazard ratio of $HR = 0.79$ (95% CI: 0.70, 0.88) for incident CHD. Under the assumption of unconfoundedness, this association could potentially be interpreted under the multiple versions of treatment theory above. However, Chang et al²⁴ also consider associations with each social integration indicator. They report evidence for an association between attending religious service more than once per week and lower CHD ($HR = 0.82$; 95% CI: 0.72, 0.93), but no notable evidence for an association with other indicators, which all have point estimates close to $HR = 1$. They report that associations are similar after adjusting for all indicators simultaneously. Similar conclusions were reached by Li et al²⁵ and VanderWeele et al²⁶ examining associations of social integration with all-cause mortality and suicide, respectively, with religious service attendance manifesting the strongest, or only, associations among the components (with marriage also protectively, but more weakly, associated with mortality).²⁵ Of course, these associations may still be confounded; moreover, the longitudinal associations of these indicators may differ with other outcomes such as happiness, income, prejudice, autonomy, etc. However, the present analyses suggest that we should be wary of assuming that "social integration" has a well-defined effect on a given outcome. From these prior analyses, it seems the indicators may be differentially associated with outcomes as in Figure 4. The structural formative model with a univariate latent is unlikely to hold.

FURTHER CRITIQUE OF THE ASSUMPTION OF AN UNDERLYING UNIVARIATE LATENT VARIABLE IN REFLECTIVE MODELS

The assumption of a univariate causally relevant latent variable is strong and will often not correspond to reality. For reflective models, the assumption is sometimes defended on

the grounds of factor analyses suggesting a unidimensional latent variable suffices to explain the covariance structure among indicators. But this does not entail a structural interpretation. The univariate factor model fitting a set of indicators [represented in Figure 2A and equation (2)] is consistent with a structural interpretation (Figure 2B) or with the latent being inert and the indicators being causally efficacious (Figure 4A). The goodness of fit of a unidimensional factor model in Figure 2A and equation (2) tells us nothing about which of these causal models is a better representation of reality. Factor analysis may be useful in generating hypotheses about underlying causally efficacious univariate latent variables, but does nothing to establish this.

The structural interpretation of a reflective model is in fact empirically testable.²¹ A structural interpretation would imply that randomized interventions that altered the latent η would have effects on the various indicators, X_i , that were proportionate to their reliabilities λ_i , which can be tested.²¹ The structural reflective model is also incompatible with one indicator having an association with an outcome, and another not, or out of proportion with their reliabilities, which can also be tested.²¹ The application of such tests to prominent scales such as the Satisfaction with Life Scale (SWLS)²⁷ indicates that the structural interpretation can be rejected:²¹ while there is evidence that four out of five indicators are associated with all-cause mortality, for one indicator, "If I could live my life over, I would change almost nothing" there is effectively no such evidence.²¹ We should be wary of assuming that a structural factor model always holds.

This does not imply the measure is bad, or that the basic univariate model is a bad fit for the covariance. The lack of a structural interpretation furthermore does not threaten the use of the scale as an outcome. Provided those using the scale equally value the individual indicators as outcomes, it is reasonable to take their average. Considered in this way, this average might then effectively be viewed as an index. Similar comments likewise pertain to other measures taken as outcomes. The interpretation of constructed measures as outcomes is arguably easier than as exposures. However, even as exposures, and even when the univariate structural

interpretation fails, the interpretation given by multiple versions of treatment theory is still applicable. But when used as an exposure, we must be careful, as even the multiple versions of treatment interpretation obscures the differential associations across indicators and obscures our capacity to discern the most relevant underlying constituents of reality. In this case, that the indicators themselves are differentially associated with all-cause mortality suggests there is no underlying univariate causally efficacious latent variable.

Similar remarks might well pertain to numerous other scales. The assumption of an underlying univariate structural factor is often just presumed, not tested. Associations between different indicators and outcomes are rarely examined. The presumption of a univariate structural latent variable may also well be unrealistic in numerous other settings.

TOWARDS MORE COMPLEX MODELS

From the analyses above, Figure 4 seems a better representation of reality than the structural reflective and formative models in Figures 2B and 3B. However, even Figure 4 is a gross simplification. Concerning social integration, each indicator for marital status, community group participation, number of close friends, and religious service attendance corresponds to a more complex reality. Quality of marriages varies; religious services can differ dramatically in content; community groups vary from arts to sports to card games. Thus, each indicator captures only aspects of a more complex reality, as in Figure 5, wherein the indicators X_i each arise from potentially multidimensional underlying latents η_i . Models like this have been considered previously assuming univariate η_i .¹² But is it reasonable to assume each η_i is univariate? Religious services vary in length of time, in discursive content, in style of worship, in demands made by participants, etc. Even the assumption that the “latent” behind a single indicator is univariate may be wrong. Nevertheless, the multiple versions of treatment interpretation of associations between measures A and outcome Y in terms of hypothetical

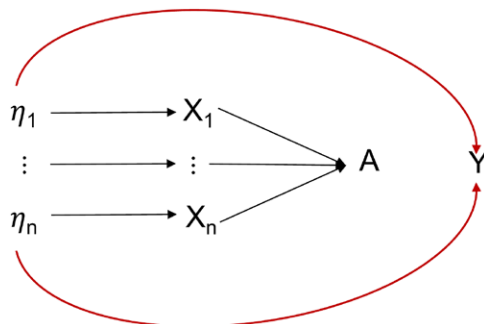


FIGURE 5. Multidimensional latent model with each indicator X_i used in forming measure A arising from a potentially multidimensional latent variable η_i , which is causally efficacious for outcome Y (Measured covariates C have been omitted for diagrammatic simplicity).

randomized trials on $\eta = (\eta_1, \dots, \eta_n)$ would still be applicable under Figure 5, and even so if each multivariate η_i affected the entire set (X_1, \dots, X_n) .

Additionally, the indicators themselves may vary over time; there will likely also be causal relations between the different aspects of the underlying reality (e.g., of social integration). Models representing this, either with only the indicators themselves or with underlying latents also, are given in Figures 6 and 7, respectively. In these figures, several things become apparent. First, if control is not made for all indicators simultaneously, then the associations of one indicator may confound that of another. For example, in Figure 6, suppose there were no effect of X_1^t on Y ; if we do not control for X_n^t in a regression of Y on X_1^t , we might observe an association between the two simply because X_n^{t-1} affects X_1^t , and X_n^{t-1} also affects X_n^t which affects Y . Second, the use of an indicator at a single time-point may be capturing, however crudely, the associations of an entire history of social participation. If we use a single composite measure that is a function of the indicators at a single time-point $A = f(X_1^t, \dots, X_n^t)$, then if temporally prior levels of these indicators $(X_1^{t-1}, \dots, X_n^{t-1})$ are causally related to the outcome Y independent of (X_1^t, \dots, X_n^t) , either directly (Figure 6) or through the latents (Figure 7), then the associations between A and Y may also partially reflect associations of the outcome with past indicators $(X_1^{t-1}, \dots, X_n^{t-1})$. Third, considerations of confounding control must take into account the time-varying nature of the indicators (and/or latents).²⁸ The multiple versions of treatment interpretation requires control of confounding for the underlying versions-of-treatment variable K . If K corresponds to a historical trajectory, its time-varying nature must be accounted for in confounding control. Unfortunately, confounding considerations become more complex with time-varying exposures^{28,29} and if confounders can themselves be affected by prior exposure levels, traditional regression-based adjustment for confounding fails; more sophisticated models are needed.^{28,29} A possibly attractive alternative is using the indicators at time t in the analysis, while simultaneously controlling for past values of the indicators (along with confounders C) at time $t - 1$.^{29,30} This proposal is discussed further in the eAppendix; <http://links.lww.com/EDE/B869>.

TOWARDS A NEW MODEL OF MEASURE CONSTRUCTION: CONSTRUCTS, INDICATORS, MEASURES, AND THE UNDERLYING CONSTITUENTS OF REALITY

Abstracting yet further from the diagrams above (and setting aside the conditioning variables), there is arguably some complex underlying reality (R in Figure 8). Certain aspects of this constitute exposure states η related to the construct of interest. The multidimensional variable η takes values in some set \mathcal{K} , each member of which defines a potential outcome for outcome Y . The variable η corresponds to

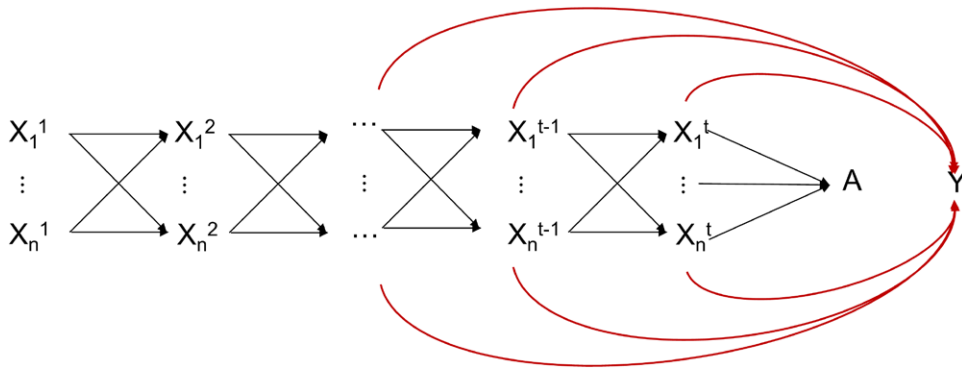


FIGURE 6. A model depicting the indicators, X_i^t used to form measures A , themselves changing over time and causally affecting one another and the outcome Y (Measured covariates C have been omitted for diagrammatic simplicity.).

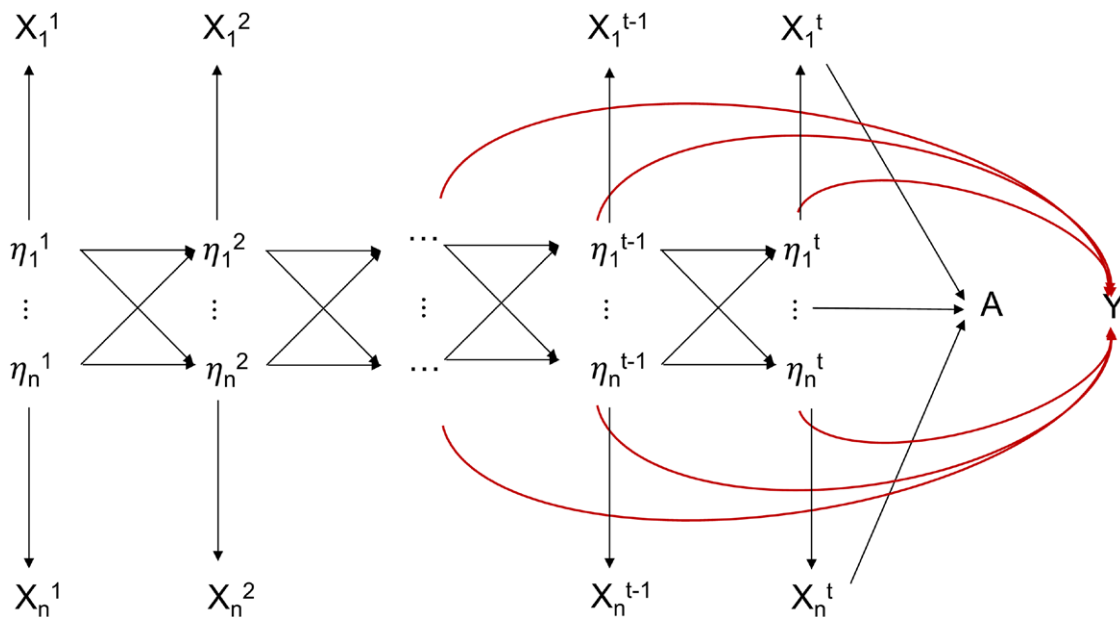


FIGURE 7. A model depicting potentially multivariate latents η_i^t giving rise to indicators X_i^t from which the measure A is formed, with the latents themselves changing over time and affecting one another as well as the outcome Y (Measured covariates C have been omitted for diagrammatic simplicity.).

the “version-of-treatment” variable K in the MVT theory; η is multivariate, not univariate. This multidimensional η gives rise to a set of observed indicators (X_1, \dots, X_n) , from which we form measures $A = f(X_1, \dots, X_n)$, either as a mean, or some other function arising from measure development processes and psychometric evaluation. We use either the indicators (X_1, \dots, X_n) or the summary measure A in analyses and examine associations with outcomes of interest, Y , controlling for other covariates, and possibly past values of (X_1, \dots, X_n) , or A , as appropriate.

Concurrent with these processes giving rise to indicators and measures is the process by which we form our concepts and constructs. The underlying constituents of reality, and our living as persons within communities, give rise to our language and the concepts embedded within it. In order to try

to systematize and study various aspects of the underlying reality, we propose constructs. Such constructs characteristically involve the systemization and reduction of our ideas, language, and concepts so as to operationalize them for use in specific modes of reasoning. However, language itself, and the concepts and derivative constructs embedded in it, of course go on to shape human behavior, the items and measures we propose, and study participants’ responses to them. These two processes are represented diagrammatically in Figure 8. In constructing measures, we hope that our measures correspond to our constructs.

The dominant measurement models – the reflective models in Figure 2 and formative models in Figure 3 – each capture aspects of these processes, but each arguably fails to acknowledge important features. Formative approaches get

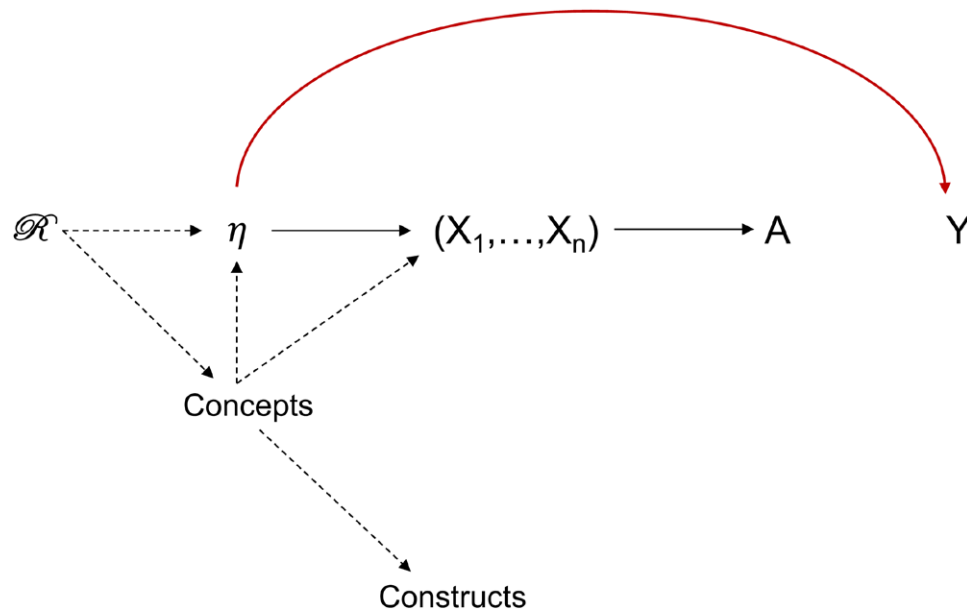


FIGURE 8. A proposed new model of measure construction wherein complex underlying reality, R contains certain aspects of this reality (represented by the multidimensional variable η) relevant to the construct. These relevant aspects of reality give rise to a set of observed indicators (X_1, \dots, X_n) , from which we form a measure A (The dotted arrows, while in some sense causal, correspond to those relations that are not explicitly between variables.).

right that our measures are always functions of our indicators. Our measure of social integration is formed by our indicators; it is not that there is a true univariate “social integration” that itself causes the indicators. However, formative models misconstrue the relation between our measures and the underlying reality to be studied. It is not that our measures, formed by the indicators, constitute (possibly subject to error) the underlying reality to be studied (as in Figure 3). It is the underlying reality that gives rise to our indicators by which we form measures.

Reflective models, in contrast, get right the fact that our measured indicators do not cause the relevant constituents of reality under study, but rather are caused by, or reflective of, these features. However, reflective models are wrong in equating the relevant aspects of reality with a univariate latent variable that corresponds to our construct.^{31–33} There is no underlying univariate latent variable that corresponds to our construct, say, of intelligence, such that “true intelligence” gives rise to the measured indicators. The underlying reality corresponding to our constructs is far more complex than a univariate variable. Models that use multiple latent variables¹² (Figure 5) more closely correspond to the underlying processes but still wrongly equate reality to a few univariate latents.

Thus, even in paradigmatic cases of the formative model, such as social integration, concerning which there is no true underlying social integration variable that “causes” the indicators, the underlying reality is nevertheless more complex than the social integration measure. Likewise, even in paradigmatic cases of the reflective model such as intelligence, concerning

which the indicators of test responses do not “cause” intelligence, it is still the case that the underlying reality is again more complex than a univariate general intelligence latent variable.^{33,34} In both cases, a complex underlying reality gives rise to our indicators from which we form measures. This is true even if we develop those measures based on psychometric approaches, such as factor analysis, arising from reflective models. Even then, the measures are still ultimately functions of the indicators, as in Figure 8. If we lose sight of that fact, we may forget that certain indicators, corresponding to particular aspects of the underlying reality, may in fact be differentially related to our outcomes of interest (as in Figures 4–7).

These issues likewise pertain to distinctions drawn between “scales” with closely related items (supposedly corresponding to reflective models) and “indices” with items that are conceptually distinct but somehow together form the construct of interest (often thought to correspond to formative models). The model for measure construction given in Figure 8 is arguably applicable to both scales and indices. In both cases, a complex underlying reality gives rise to item responses from which we form measures. The relations between the underlying processes and the formation of measures may thus be more similar for scales and indices than typically thought. Whether a measure is considered a scale or index may have more to do with the items used to construct the measure, and the use of that measure, than with the definition of the construct itself. While life satisfaction is often assessed as a scale with several related subjective indicators,²⁷ if the construct is instead assessed by life domain (work, family, health, finances, etc.),^{35,36} the measure will resemble an index. Conversely, while measures of

social integration are often assessed by domain (marital status, time with friends, religious community, etc.),^{22,23} social integration could alternatively be assessed with a series of related subjective indicators as in Duke's Subjective Social Support subscale.³⁷ In all these cases, a complex underlying reality gives rise to indicators by which we form measures. It is the conceptual relations between the items and the construct that differs, not the model of measurement per se.

All measures are formative in that they are formed from observed indicators; all measures are reflective in that they are reflective of a more complex underlying reality. The fallacy of the formative model is that the relevant underlying reality is made up of a function of our indicators; the fallacy of the reflective model is the supposition that we have imperfectly measured an underlying univariate latent variable.

IMPLICATIONS AND CONCLUSIONS

The dominance of the reflective model, and the fallacious presumption that the basic univariate factor model fitting the indicators well implies a structural interpretation, gives rise to the illusion that, in most settings, there truly is an underlying univariate latent variable adequately representing reality.^{21,38} Potential causal relations between different phenomena can reinforce this illusion further.³⁸ As argued in this paper, there are not in general adequate grounds to justify the presumption of an underlying univariate structural latent variable. But this presumption and illusion has arguably led to a related series of other subtle subsequent missteps in measure construction, conceptualization, and evaluation.

It has been argued elsewhere that current factor analysis and measure construction practices have led to the conflation of the terms “construct” and “latent variable.”^{32,39} Indeed the very term “latent construct” effectively entails the equating of a conceptual specification with a quantitative variable, generally presuming a univariate structure. We need a clear distinction between concepts and constructs, their underlying referents (η), and our attempts to measure these underlying referents (X_1, \dots, X_n).^{32,39} The conflation of “construct” and “variable,” and the presumption of a univariate underlying reality has also led to a notion that the nature of the concept is to be discovered empirically from analyses of correlations.³² Items are proposed, factor analyses implemented, and it is assumed we somehow thereby come to understand the meaning of the construct itself. This view has in turn has often led to a lack of formal definitions given for the construct under consideration,³² since, so it is thought, this is to be “discovered” empirically. While plenty of theory is often provided, formal definitions are more rare.

The lack of definitions in turn obscures the relation between items and constructs. Items that are necessary or sufficient or merely illustrative of the construct are treated interchangeably; none are related to definitions themselves. Without definitions, it becomes difficult to assess whether two different measures of allegedly the same construct are intended to assess the same thing, or whether authors have different

understandings of the construct, or whether they view the nature of the construct as something to be discovered empirically and are beginning exploration from different places. The lack of definitions also tends to lead to overly broad inclusion of items within measures. Not infrequently, conceptually distant but desirable outcomes are placed among the items (e.g., “I've been pretty successful in life” in Snyder's *hope* scale⁴⁰; or taking “Pride in your achievements” in the Connor–Davidson resilience scale⁴¹). Without definitions, criticism is more difficult; often these items are simply accepted, provided a univariate factor model accounts for covariances among indicators. Moreover, since the underlying factor is presumed univariate, item-by-item analyses, which might uncover differential relations with outcomes, are rare, thereby further obscuring the important conceptual and empirical distinctions that may be present among the items.

Much of this would benefit from change, beginning with clear definitions of the construct.⁴² Proposed items should then be derived from the definitions, with an understanding of their relationship including whether items make use of the word corresponding to the construct; whether the items are necessary, sufficient, necessary and sufficient, or merely illustrative for the construct; or whether items are intended to capture different facets of the construct. The work of analytic philosophy may be useful both in this task, and in clarifying different uses of our language and thereby facilitating particular definitions of the construct in view.^{43–45} Various measures can be proposed from item indicators on conceptual grounds. Appropriate cognitive testing and measure evaluation strategies could be developed. Factor analyses^{1–4} may be useful to assess approximate covariance dimensionality, but indication of unidimensional factor structure is neither necessary nor sufficient for using a univariate measure in analysis. It is not necessary because associations between constructed measures and outcomes can still admit a causal interpretation under the multiple versions of treatment theory⁶ above. It is not sufficient because even if the basic univariate factor model fits, the causal interpretation of the latent may not be structural.²¹ Regardless of the fit of the unidimensional model, it will still be useful to carry out item-by-item analyses, or using composites of conceptually related items, either to potentially provide some evidence for a structural univariate latent interpretation, or alternatively to uncover important distinctions between items that may be relevant in refining measure construction, understanding facets of the construct, or thinking about interventions.

A preliminary outline of a more adequate approach to the construction and use of psychosocial measures might thus be summarized by the following propositions, that I have argued for in this article: (1) Traditional univariate reflective and formative models do not adequately capture the relations between the underlying causally relevant phenomena and our indicators and measures. (2) The causally relevant constituents of reality related to our constructs are

almost always multidimensional, giving rise both to our indicators from which we construct measures, and also to our language and concepts, from which we can more precisely define constructs. (3) In measure construction, we ought to always specify a *definition* of the underlying construct, from which items are derived, and by which analytic relations of the items to the definition are made clear. (4) The presumption of a structural univariate reflective model impairs measure construction, evaluation, and use. (5) If a structural interpretation of a univariate reflective factor model is being proposed this should be formally tested, not presumed; factor analysis is not sufficient for assessing the relevant evidence. (6) Even when the causally relevant constituents of reality are multidimensional, and a univariate measure is used, we can still interpret associations with outcomes using theory for multiple versions of treatment, though the interpretation is obscured when we do not have a clear sense of what the causally relevant constituents are. (7) When data permit, examining associations item-by-item, or with conceptually related item sets, may give insight into the various facets of the construct.

A new integrated theory of measurement for psychosocial constructs is needed in light of these points – one that better respects the relations between our constructs, items, indicators, measures, and the underlying causally relevant phenomena.

REFERENCES

- Comrey AL, Lee HB. *A First Course in Factor Analysis*. Psychology Press; 2013.
- Kline P. *An Easy Guide to Factor Analysis*. Routledge; 2014.
- Brown TA. *Confirmatory Factor Analysis for Applied Research*. Guilford Publications; 2015.
- Bollen KA. *Structural Equations with Latent Variables*. Wiley; 1989.
- Sánchez BN, Budtz-Jørgensen E, Ryan LM, Hu H. Structural equation models: a review with applications to environmental epidemiology. *J Am Stat Assoc*. 2005;100:1443–1455.
- VanderWeele TJ, Hernán MA. Causal inference under multiple versions of treatment. *J Causal Inference*. 2013;1:1–20.
- Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology*. 2011;22:368–377.
- Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. *Psychol Methods*. 2000;5:155–174.
- Diamantopoulos A, Siguaw JA. Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration. *Br J Manag*. 2006;17:263–282.
- Coltman T, Devinney TM, Midgley DF, Venaik S. Formative versus reflective measurement models: two applications of formative measurement. *J Bus Res*. 2008;61:1250–1262.
- Diamantopoulos A, Winklhofer HM. Index construction with formative indicators: an alternative to scale development. *J Mark Res*. 2001;38:269–277.
- Edwards JR. The fallacy of formative measurement. *Organ Res Methods*. 2011;14:370–388.
- MacKenzie SB, Podsakoff PM, Jarvis CB. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *J Appl Psychol*. 2005;90:710–730.
- Morgan SL, Winship C. *Counterfactuals and Causal Inference*. Cambridge University Press; 2015.
- Pearl J. *Causality*. Cambridge University Press; 2009.
- Imbens GW, Rubin DB. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press; 2015.
- Hernan MA, Robins JM. *Causal Inference*. Chapman Hall; forthcoming.
- VanderWeele TJ. On well-defined hypothetical interventions in the potential outcomes framework. *Epidemiology*. 2018;29:e24–e25.
- VanderWeele TJ. On causes, causal inference, and potential outcomes. *Int J Epidemiol*. 2016;45:1809–1816.
- DeVellis RF. *Scale Development: Theory and Applications*. 4th ed. Sage Publications; 2016.
- VanderWeele TJ, Vansteelandt S. A statistical test to reject the structural interpretation of a latent factor model. Available at: <http://arxiv.org/abs/2006.15899>.
- Berkman LF, Kawachi I, Glymour MM, eds. *Social Epidemiology*. Oxford University Press; 2014.
- Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol*. 1979;109:186–204.
- Chang SC, Glymour M, Cornelis M, et al. Social integration and reduced risk of coronary heart disease in women: the role of lifestyle behaviors. *Circ Res*. 2017;120:1927–1937.
- Li S, Stampfer MJ, Williams DR, VanderWeele TJ. Association of religious service attendance with mortality among women. *JAMA Intern Med*. 2016;176:777–785.
- VanderWeele TJ, Li S, Tsai AC, Kawachi I. Association between religious service attendance and lower suicide rates among US women. *JAMA Psychiatry*. 2016;73:845–851.
- Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess*. 1985;49:71–75.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
- VanderWeele TJ, Jackson JW, Li S. Causal inference and longitudinal data: a case study of religion and mental health. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51:1457–1466.
- VanderWeele TJ, Mathur MB, Chen Y. Outcome-wide longitudinal designs for causal inference: a new template for empirical studies. *Stat Sci*. 2020;35:437–466.
- Maraun MD, Halpin PF. Manifest and latent variates. *Measurement*. 2008;6:113–117.
- Maraun MD, Gabriel SM. Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas Psychol*. 2013;31:32–42.
- Maraun MD. Myths and Confusions: Psychometrics and the Latent Variable Model. 2003. Available at: <http://www.sfu.ca/~maraun/myths-and-confusions.html>. Accessed 27 May 2020.
- Michell J. Constructs, inferences, and mental measurement. *New Ideas Psychol*. 2013;31:13–21.
- Fugl-Meyer AR, Bränholm IB, Fugl-Meyer KS. Happiness and domain-specific life satisfaction in adult northern Swedes. *Clin Rehabil*. 1991;5:25–33.
- Cummins R. The domains of life satisfaction: an attempt to order chaos. *Soc Indic Res*. 1996;38:303–328.
- Koenig HG, Westlund RE, George LK, Hughes DC, Blazer DG, Hybels C. Abbreviating the Duke Social Support Index for use in chronically ill elderly individuals. *Psychosomatics*. 1993;34:61–69.
- VanderWeele TJ, Batty CJK. On the dimensional indeterminacy of one-wave factor analysis under causal effects. 2020. Available at: <https://arxiv.org/abs/2001.10352>.
- Slaney KL, Racine TP. What's in a name? Psychology's ever evasive construct. *New Ideas Psychol*. 2013;31:4–12.
- Snyder CR. Conceptualizing, measuring, and nurturing hope. *J Couns Dev*. 1995;73:355–360.
- Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson resilience scale (CD-RISC). *Depress Anxiety*. 2003;18:76–82.
- Krause MS. Measurement validity is fundamentally a matter of definition, not correlation. *Rev Gen Psychol*. 2012;16:391–400.
- Wittgenstein L. *Philosophical Investigations*. Macmillan Publishing Company; 1953.
- Hanfling O. *Philosophy and Ordinary Language: The Bent and Genius of our Tongue*. Routledge; 2013.
- Maraun MD. Measurement as a normative practice: implications of Wittgenstein's philosophy for measurement in psychology. *Theory Psychol*. 1998;8:435–461.

APPENDIX

Causal Effects Under Multiple Versions of Treatment

The derivation here follows the structure of the proof of Proposition 8 given by VanderWeele and Hernán⁶ but under weaker assumptions (see eAppendix; <http://links.lww.com/EDE/B869>), requiring only that (i) Y is independent of A conditional on (K, C) ; (ii) the effect of K on Y is unconfounded given C i.e. Y_k is independent of K given C ; and (iii) the consistency assumption that when $K = k$ then $Y_k = Y$. Under these assumptions we then have that:

$$\begin{aligned} & \sum_c \{E[Y | A = a, c] - E[Y | A = a^*, c]\} P(c) \\ &= \sum_{k,c} E[Y | k, a, c] P(k | a, c) P(c) - \sum_{k,c} E[Y | k, a^*, c] P(k | a^*, c) P(c), \\ &= \sum_{k,c} E[Y | k, c] P(k | a, c) P(c) - \sum_{k,c} E[Y | k, c] P(k | a^*, c) P(c), \\ &= \sum_{k,c} E[Y_k | k, c] P(k | a, c) P(c) - \sum_{k,c} E[Y_k | k, c] P(k | a^*, c) P(c), \\ &= \sum_{k,c} E[Y_k | c] P(k | a, c) P(c) - \sum_{k,c} E[Y_k | c] P(k | a^*, c) P(c), \end{aligned}$$

where the first equality follows from the law of iterated expectations, the second from the independence of Y and A conditional on (K, C) , the third from consistency, and the fourth from unconfoundedness.