Knowledge as a formative construct: a good alpha is not always better

Matthias Stadler, Michael Sailer, Frank Fischer

Angaben zur Veröffentlichung / Publication details:

Stadler, Matthias, Michael Sailer, and Frank Fischer. 2021. "Knowledge as a formative construct: a good alpha is not always better." *New Ideas in Psychology* 60: 100832. https://doi.org/10.1016/j.newideapsych.2020.100832.

Knowledge as a formative construct: A good alpha is not always better

Matthias Stadler *, Michael Sailer, Frank Fischer

Ludwig-Maximilians-Universität München, Munich, Germany

Coefficient alpha (Cronbach, 1951) is certainly one of the most important and pervasive statistics in research involving test construction and use (Edwards et al., 2019). Alpha is commonly reported for the development of scales intended to measure attitudes and other affective constructs. However, in the literature on the development of tests of student domain knowledge, some articles have cited Cronbach's alpha as an indicator of instrument quality (for a first overview see Taber, 2018). Alpha as an indicator of instrument quality assumes an underlying construct that is reflected in different observable behaviors (i.e., a reflective model; Chin, 1998). However, domain knowledge, defined as the realm of knowledge that individuals have about a particular field of study (Alexander, 1992), covers declarative (knowing that), procedural (knowing how), and conditional (knowing when and where) knowledge. As such, it often encompasses a rich array of unrelated aspects or areas. In this short article, we want to make the argument that the assumption of a reflective model is not necessarily correct for the assessment of domain knowledge. Domain knowledge may instead be formed by the manifest observations chosen for the instrument (i.e., a formative model; Chin, 1998). This has implications for both the assessment and modeling of domain knowledge.

1. Formative constructs

A person's knowledge about any specific topic cannot be observed directly. Rather, we need to observe their behavior to be able to deduct their knowledge from it. Such an unobservable variable is known as a latent variable (Borsboom et al., 2003). With the increasing popularity of structural equation modeling (SEM) techniques, researchers are able to better assess both structural and measurement models of latent variables such as domain knowledge (e.g., Park, Suh, & Seo, 2018), yet they tend to focus on the structural model rather than fully consider the relations between measures and their relevant latent constructs (Petter et al., 2007). A theory can be defined as a statement about relations among constructs within a set of assumptions about boundaries and constraints (Bacharach, 1989). According to this definition, a theory consists of two parts: On the one hand, it needs to specify relations between latent constructs, and on the other hand, it needs to describe relations between latent constructs and their manifest indicators (Bollen & Lennox, 1991). Whereas educational researchers have often placed great emphasis on defining and explaining causal relations between constructs, the direction and nature of relations *between constructs and their indicators* has received far less attention. These relations are important, though, as they constitute a secondary theory that bridges the gap between measurable empirical phenomena and abstract theoretical

* Corresponding author. Leopoldstr. 13, 80802, München, Germany.
E-mail address: Matthias.Stadler@lmu.de (M. Stadler).

constructs (Roberts & Thatcher, 2009).

As Bollen and Lennox (1991) described in their seminal article on construct measurement, manifest indicators can be the effects or the causes of latent variables (see Fig. 1). A common example of a reflective model in which the manifest indicators represent the effect of a latent construct involves the concept of general intelligence (e.g., McGrew, 2009). According to this concept, performance in virtually all cognitive tasks is affected by an underlying common factor. Changes in general intelligence should therefore result in changes in performance across all tasks, whereas a change in performance on one task (e.g., from cheating) would not result in changes in general intelligence. By contrast, formative models propose that manifest indicators form a latent construct or cause the changes that occur in it. An example is socioeconomic status (SES), where indicators such as education, income, and occupational prestige are items that cause or form the latent construct SES (Chin, 1998). If a person loses his or her job, the person's SES would be negatively affected. But a negative change in an individual's SES does not necessarily imply that the person lost his or her job. Moreover, a change in one indicator (e.g., income) does not necessarily imply a similar directional change in the other indicators (e.g., education or occupational prestige).

As the manifest indicators chosen by the researcher form a formative construct, there need to be persuasive arguments for the theoretical merit of each indicator. Sticking to the example of SES, the inclusion or exclusion of income as an indicator substantially changes the resulting construct with arguments both for and against the inclusion depending on the research question (for a more detailed discussion see Daly et al., 2002).

2. Domain knowledge as a formative construct

Theoretically, domain knowledge is often conceived as a construct formed by several aspects that do not necessarily have to be associated with each other. For instance, knowledge about biology may be defined as knowledge in five distinct biological disciplines (e.g., ecology, evolution, genetics and microbiology, morphology, and physiology; Großschedl et al., 2018). The resulting latent construct of domain knowledge about biology is formed by this definition, and choosing different disciplines would change the assumed latent construct. Likewise, the causal direction implied by the model is that the manifest observations form the latent construct. For example, teaching a student about genetics would increase his or her latent domain knowledge about biology but would not necessarily increase the student's knowledge about ecology. Thus, domain knowledge about biology would be considered a formative construct.

In scientific practice, however, domain knowledge is rarely assessed or modeled as a formative construct (Roberts & Thatcher, 2009; Taber, 2018). Rather, reflective models are assumed without any specific argument for why this choice was made (for an example see Großschedl et al., 2018). This potential mismatch between conceptualization and empirical operationalization can lead to critical misrepresentation and misinterpretation of domain knowledge as assessed with reflective models (Bollen, 2014).

3. Implications for assessment and modeling

3.1. Inadequate selection criteria

The most common implication that comes from using a reflective model to model a latent construct that is theoretically formative is that inadequate scale development procedures are thus being used (Diamantopoulos et al., 2008). As we illustrated above, formative indicators do not necessarily have to be highly correlated. In fact, correlations among formative indicators should be low enough so that they are not redundant with each other in contributing to the latent construct (Diamantopoulos & Siguaw, 2006). However, following the guidelines of scale development for reflective scales requires a maximizing of the interitem correlations (e.g., by maximizing Cronbach's alpha). As Taber (2018) summarized, a high alpha obtained from administering an instrument to a sample of students could suggest that the items measure some common factor(s) rather than unique features associated with individual test items. Alpha could then be associated as a measure of redundancy within the measure. However, measures of domain knowledge often include items that are each designed to test specific content elements. Such items are not intended to be redundant as knowledge about a domain (e.g., biology) is rarely considered a homogenous construct. Therefore, measuring a domain of interest that is considered to consist of distinct aspects, each representing related but separable units of knowledge, requires creating separate items measuring these distinct aspects. Finding that there is a high alpha coefficient across these items would actually indicate that the test is not working as intended.

As an alternative to item selection based on maximizing Cronbach's alpha, various authors have suggested that researchers should use indicator elimination based on the variance inflation factor (VIF) for formative models (e.g., Diamantopoulos & Siguaw, 2006). The VIF denotes the degree of multicollinearity (i.e., redundancy) across indicators as a value ranging from 1 (no correlation with any other indicators) to infinity (the indicator can be perfectly explained by the other indicators and is, thus, perfectly redundant). If the VIF statistic for formative measures is greater than 3.3 (i.e., more than 70% of the indicator's variance is explained by the other indicators; Diamantopoulos &

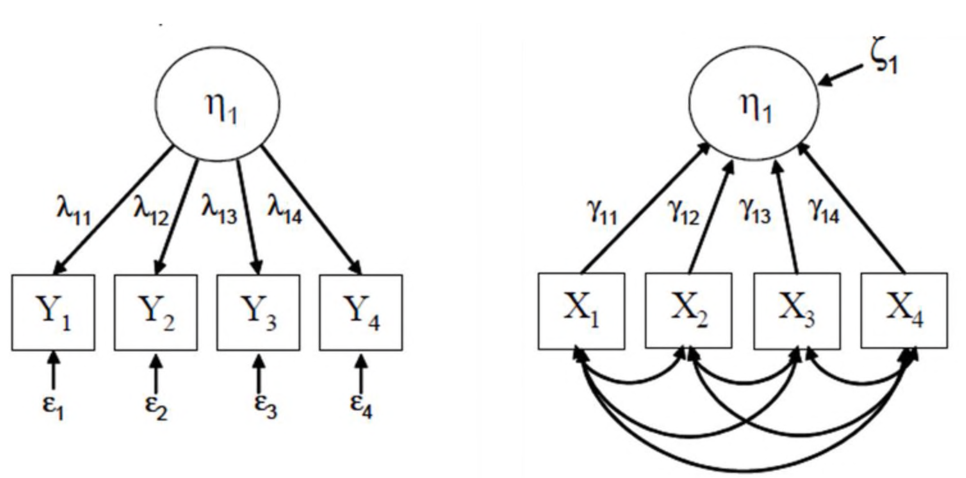

Fig. 1. Illustration of reflective (left) and formative (right) measurement models (adapted from Bollen & Lennox, 1991).

Siguaw, 2006), the researcher should address the issue by either modeling the construct as having both formative and reflective measurement items, removing correlated items if this does not affect content validity, collapsing the correlated items into a composite index, or converting the construct into a multidimensional construct (Crawford & Kelder, 2019; Petter et al., 2007).

It is of note, however, that such multicollinearity checks lead to indicator elimination on purely statistical grounds, which comes with the danger of altering the meaning of the latent construct by excluding indicators (Bollen & Lennox, 1991). However, given the pivotal importance of a clear theoretical construct definition in formative models, indicator elimination must not be divorced from conceptual considerations when a formative measurement model is involved (Diamantopoulos & Winklhofer, 2018).

3.2. Misrepresentation of constructs

A potentially even more critical implication of misspecifying formative constructs by modeling them by using reflective models is the risk of actually misrepresenting the construct. Cronbach (1951) suggested that alpha reflects "how much the test score depends upon general and group, rather than item-specific, factors" (p. 320), but on a test of domain knowledge, it may be important to include very specific items to test different knowledge components. In such a situation, a high alpha may suggest that a lot of the variance is due to general learner-related factors (e.g., intelligence, study diligence, interest) and that consequently, the instrument does not differentiate well between the different features of the conceptual material that is being tested. Rather, a latent factor from such a test could be interpreted as a measure of a general factor. In this case, a measure of domain knowledge would be misrepresented as a measure of, for instance, general intelligence.

Baumert et al. (2009) illustrated how such a misspecification could lead to misinterpretations based on the debate on whether large-scale student assessment studies (e.g., TIMSS, PISA) measure the results of processes of knowledge acquisition or a single cognitive ability. (Rindermann (2007)) claimed that international and national assessment studies essentially measure the same thing as intelligence tests based on the large amount of variance explained by the first factor of an exploratory factor analysis with achievement subtests from PISA. As demonstrated by (Brunner (2008)), interindividual differences result from two independent cognitive abilities: general cognitive ability and domain-specific ability. These domain-specific abilities are uncorrelated and can be combined into a formative knowledge construct. Findings on the relations between students' abilities and important criteria such as students' socioeconomic status, general school satisfaction, educational aspirations, domain-specific interests, and subject-specific grades may differ substantially depending on the structural model applied.

4. Conclusion

In this short article, we argue that domain knowledge can and, more often than not, should be considered a formative construct rather than a reflective one. More specifically, we argue that most theoretical conceptualizations of domain knowledge already consider domain knowledge to be formed from distinct aspects that define the assumed construct with convincing arguments for and against the inclusion of specific items. However, the scientific practice of assessing and modeling domain knowledge does not match this theoretical conception as researchers usually seek to maximize interitem correlations rather than to optimally cover the theoretical width of their constructs. We illustrate how this misspecification can have serious implications for the interpretations of the construct of domain knowledge as inadequate selection criteria lead to limited construct operationalizations or even misrepresentations of the constructs that are being assessed. Please note, that there are other, more general issues with the use of Cronbach's alpha as well, that we could not discuss in this short article (for a summary see Trizano-Hermosilla & Alvarado, 2016). We hope that this short article will inspire researchers to review their assessments of domain knowledge and prompt them to think about whether their theoretical conceptualization of the construct is actually represented in its empirical operationalization.

Acknowledgements

This research was supported by a grant from the Deutsche Forschungsgemeinschaft DFG (COSIMA; DFG-Forschungsgruppe 2385).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.newideapsych.2020.100832.

References

Alexander, P. A. (1992). Domain knowledge: Evolving themes and emerging concerns. *Educational Psychologist, 27*(1), 33–51. https://doi.org/10.1207/s15326985ep2701_4.

Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review, 14*(4), 496–515. https://doi.org/10.5465/amr.1989.4308374

Baumert, J., Lüdtke, O., Trautwein, U., & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review, 4*(3), 165–176. https://doi.org/10.1016/j.edurev.2009.04.002

Bollen, K. A. (2014). *Structural equations with latent variables* (Vol. 210). John Wiley & Sons.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305–314. https://doi.org/10.1037/0033-2909.110.2.305

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Brunner, M. (2008). No g in education? *Learning and Individual Differences, 18*(2), 152–165. https://doi.org/10.1016/j.lindif.2007.08.005.

Chin, W. W. (1998). Commentary: Issues and opinion on structural equation modeling. *MIS Quarterly, vii–xvi*(22).

Crawford, J. A., & Kelder, J.-A. (2019). Do we measure leadership effectively? Articulating and evaluating scale development psychometrics for best practice. *The Leadership Quarterly, 30*(1), 133–144. https://doi.org/10.1016/j.leaqua.2018.07.001

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/bf02310555

Daly, M. C., Duncan, G. J., McDonough, P., & Williams, D. R. (2002). Optimal indicators of socioeconomic status for health research. *American Journal of Public Health, 92*(7), 1151–1157. https://doi.org/10.2105/ajph.92.7.1151

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1203–1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management, 17*(4), 263–282. https://doi.org/10.1111/j.1467-8551.2006.00500.x

Diamantopoulos, A., & Winklhofer, H. M. (2018). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269–277. https://doi.org/10.1509/jmkr.38.2.269.18845

Edwards, A., Joyner, K., & Schatschneider, C. (2019). A simulation study on the performance of different reliability estimation methods. https://doi.org/10.31234/osf.io/xzc52.

Großschedl, J., Mahler, D., & Harms, U. (2018). Construction and evaluation of an instrument to measure content knowledge in biology: The CK-IBI. *Education Sciences, 8*(3), 145. https://doi.org/10.3390/educsci8030145

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Park, S, Suh, J., & Seo, K. (2018). Development and Validation of Measures of Secondary Science Teachers' PCK for Teaching Photosynthesis. *Research in Science Education, 48*(3), 549–573. https://doi.org/10.1007/s11165-016-9578-y.

Petter, S., Straub, D. W., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, (31), 657–679.

Rindermann, H (2007). Theg-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*(5), 667–706. https://doi.org/10.1002/per.634.

Roberts, N., & Thatcher, J. (2009). Conceptualizing and testing formative constructs. *ACM SIGMIS - Data Base, 40*(3), 9. https://doi.org/10.1145/1592401.1592405

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7*, 769.