

Formative Cows and Reflective Chicken: A Note on Mixing Measurement Modes in Structural Equation Models

Josip Mikulić¹

¹ Associate Professor, Faculty of Economics and Business - Zagreb, Croatia
(Email: jmikulic@efzg.hr)

Abstract: *This article intends to contribute to the debate on formative/reflective measurement of marketing and management constructs, and in particular, to assist researchers who are not sure which measurement mode to apply to their focal constructs. In making its points, this article uses illustrative examples, however, not in terms of abstract theoretical constructs, but rather in terms of Rossiter's analogy involving horses, yaks and other animals (Rossiter, 2005). Unlike most (or all) theoretical constructs from the field of marketing, animals are material constructs everyone knows, so one can easier imagine them in both a reflective and formative way. Consequently, the chosen type of narration should significantly increase understanding*

Keywords: *measurement mode, formative, reflective, modeling*

1. Introduction

The formative versus reflective measurement debate in the marketing and management literature has yielded valuable insight that aids researchers in choosing the proper measurement mode. The issue is of particular importance in structural equations modeling (SEM), since the reliability and, in fact, validity of model results depends on whether the proper measurement mode has been chosen. While there are several studies dealing with the general issue of reflective and formative measurement in SEM, none of them has so far focused on mixing measurement modes and its implications on SEM results. The present brief article aims to address this particular issue. Before providing an empirical example to portray the implications, this article briefly reviews both formative and reflective measurement first.

2. Formative and reflective measurement mode

2.1. Formative measurement mode

To start with, let us say we wanted to model a cow using its body-parts—i.e. its formative indicators. As the measurement unit we will define weight, however, we could have used any other measurement unit as well (e.g. height, volume, value, etc). Accordingly, would we have done well if we modeled the cow-construct with one indicator—i.e. a leg? Certainly not, because there are many other animals which have one leg and typically animals have more of them. Obviously, we did not cover the whole conceptual domain of the cow-construct (Diamantopoulos and Winklhofer, 2001). So would we have done well if we modeled the cow with two indicators—i.e. a leg and a tail? Well, we came closer, but not close enough. Moreover, the tail did not really contribute to the discriminant validity of the cow-construct, since horses have tails, as well. Nevertheless, we should acknowledge that the external validity of the cow-construct increased, since what we have measured came closer to a cow (i.e. the weight of a cow), but still we are far from perfect. We could continue this way, add a head, an eye, another eye, an eyelash, and so on, and we would probably succeed, sometime, to model the cow-

construct formatively in a way that it is different from everything else we know, with perfect external validity. However, the measurement-operationalization for our cow construct (e.g. with Likert-statements) would exhibit a questionnaire that takes hours, and, in the end, when analyzing the data, many of the cow-indicators might exhibit multicollinearity (not to speak of the possibility that some of farmers might not have surely know how heavy the liver of their cow is). Since this research is exploratory in nature, in case of multicollinearity we could factor-analyze the cow-data, and we would, probably, realize that we do not have to measure each leg or eyelash separately. It could even happen that all the cow-indicators are highly correlated, so we would receive a one-factor solution—i.e. a factor we call cow! But still, I cannot be sure that I did not leave out something important. And since, according to Diamantopoulos and Winklhofer (2005), leaving out one of the cow's formative indicators could significantly change the construct's meaning, I might in fact be measuring a horse, or even a donkey. It becomes quite clear; it is really hard to model a cow in a formative way.

2.2. Reflective measurement mode

However, everyone knows what a cow is! So why don't we simply ask farmers about the weight using a single measurement-item, e.g. on a direct rating-scale from 1 to 5, with 1 being very low weight and 5 very high weight? Though some researchers think that a single-item measure like this one would *probably* do (Drolet and Morrison, 2001), it might happen that others start attacking the results of my study, because it might not be sure that I have really measured a cow—i.e. I did not assure measurement reliability at the construct-level (e.g. Grapentine, 2001). Alright, this is not a meaningless point. Accordingly, to be sure that I have really measured what I intended to, I could specify my cow as a reflective construct, and ask the farmers e.g. about the height, length and width (and also the weight) of their animals. Three or four questions should however do, since any further one would not be likely to significantly contribute to the reliability that I have really measured a cow (Drolet and Morrison, 2001). However, as some colleagues do, I could as well ask the same question several times, but differently formulated. Though I do not really get additional information with the questions, and some respondents might think that I am making fun of them, in a reflective measurement mode the answers will produce perfect Cronbach alphas again. And this is all that (often) really matters to publish my article on cows in a top marketing journal.

2.3. Is it reflective or formative?

Let us go somewhat further back in time and stick to Cronbach (1951). Do you remember the tails and legs of the previously described formative cow? As social science researchers, we know that these forming parts are very likely to be highly correlated. So could I have used the same indicators for measuring my cow, but changing the measurement mode from formative to reflective? Rossiter (2005) says that researchers often do that, but he is not really convinced that this is a good approach. It certainly is not. However, we could try with reflective mode, and we must acknowledge (and Rossiter, as well) that we would probably obtain good results, because heavier legs typically go along with a heavier tail (because the whole cow is bigger/heavier—i.e. the body-parts are highly correlated). However, we rather like to think that causality goes from the legs and the tail to the cow, rather than the other way round (they form the cow, don't they?). After all, there might be cows with shorter and larger legs out there, or with less than four legs, but with equally long tails. Accordingly, this might negatively affect Cronbach alphas in reflective measurement mode, and turn me start questioning whether I really measured a cow, though I know I did. As a matter of statistical fact, Churchill (1991, p. 495) would call into question the reliability, and subsequently the validity of my cow-construct. Conversely, if I had modeled the cow formatively using the same indicators, body-part disproportions would in fact only affect the value of something that would be a cow index. That would make more sense. So what would proper reflective measurement look like? Actually we should answer the question what we actually wanted to make reliable when opting for reflective measurement mode? That a cow is a cow? Well, everybody knows what a cow is.

3. Mixing measurement modes

What this note is particularly bothered about is mixing measurement modes in comparisons of two or more focal variables. To keep it simple, in this illustrative example we will use only two constructs. Imagine, for example, that some researcher used the previously mentioned formative cow in combination with a reflective chicken, and put both of them into the same structural equations model (SEM) to test their effects on some endogenous variable. In this case, the formative cow-conceptualization should be really, good as we will see, because, otherwise the practical implications of the study might be that a chicken can feed more farmers than a cow does. To back up this point with empirical evidence we will use an example on the nutritional value of the focal animals of our study—i.e. cows and chicken (i.e. the real data are in fact measuring the impact of airline service quality and airline image on passenger loyalty). The path coefficients were estimated using SmartPLS 2.0 software (Ringle et al., 2005). In the first example, we use a formative cow and a reflective chicken (Figure 1).

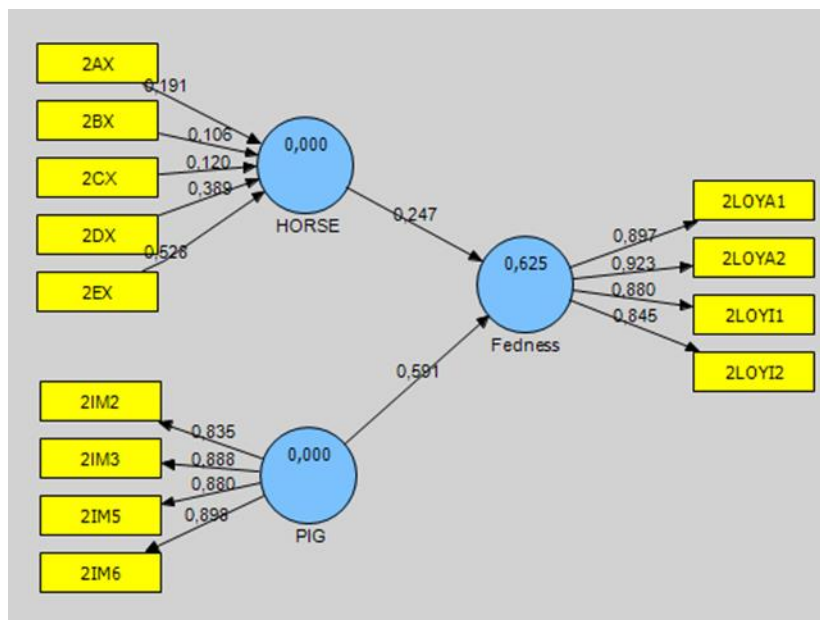


Fig 1. : The effects of cow (formative) and chicken (reflective) on the farmer's fullness

Groundbreaking results—a chicken can feed more farmers than a cow. The empirical t-values obtained with bootstrapping-procedures reveal that both animal constructs are significant, as well as all of their indicators. Moreover, the reflective chicken exhibits very good Cronbach alpha. Moreover, the animals exhibit impressive $R^2=0.625$ in the endogenous variable which is considered to describe substantial path structures (Chin, 1998).

However, though the statistics say that everything is fine, a look at my cow-construct reveals that I forgot the legs! Obviously, this is the reason for such troubling results. Well, I could return to the farmers and collect the necessary information, but I would certainly not be sure that I did not forget another important cow-indicator, again. So, why not rather try with a reflective cow (Figure 2)?

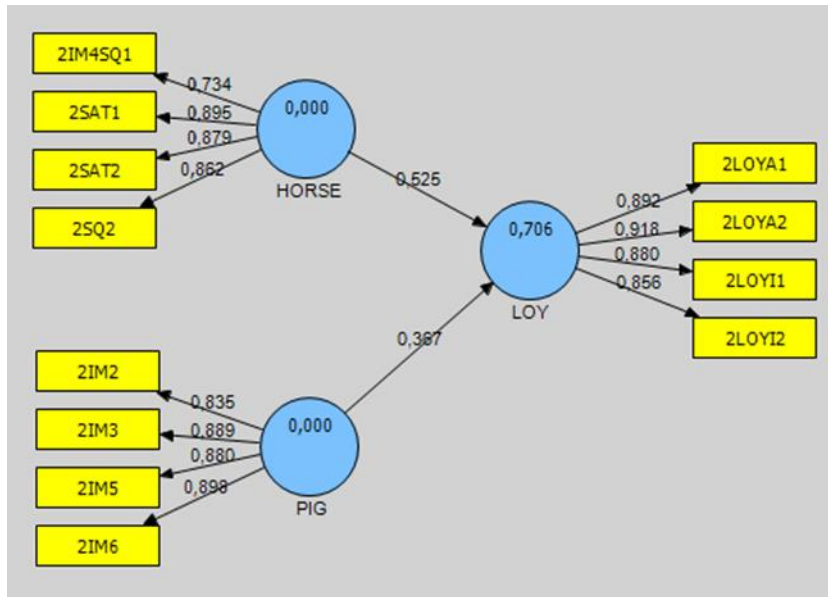


Fig 2. The effects of cow (reflective) and chicken (reflective) on the farmer's fullness

Well, these results seem to fit reality better. The reflective animals even exhibit an impressive R squared of 0.706, which is an almost too perfect model for social science researchers!

What, however, if my reflective animal conceptualizations simply were not good? Well, this is hard to tell. However, given all the exploratory and conclusive research on the cow/chicken problem, the reflective measurement operationalizations should be quite good by now. After all, usually people know that a cow is a cow. Nevertheless, we should never stop questioning the reliability and validity of their measurement operationalizations. Moreover, although we all know as researchers, we should acknowledge that there is obviously a strong need to use the same conceptualizations of cows and chicken when making comparisons across different studies (meta-studies), regardless whether these are reflective or formative. Then, and only then, we can, altogether, create reliable theory. And, of course, as soon as someone finds a significantly better conceptualization, we all who are interested in cows and chicken should start using that one, rather than to act blind and stick to dated models.

References

- [1] Diamantopoulos, A., and Winklhofer, H.M. (2001). 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>
- [2] Rossiter, J. R. (2005). Reminder: a horse is a horse. *International Journal of Research in Marketing*, 22(1), 23.
<https://doi.org/10.1016/j.ijresmar.2004.11.001>
- [3] Drolet, A. L., & Morrison, D. G. (2001). Do we really need multiple-item measures in service research?. *Journal of Service Research*, 3(3), 196-204.
<https://doi.org/10.1177/109467050133001>
- [4] Grapentine, T. H. (2001). A practitioner's comment on Aimee L. Drolet and Donald G. Morrison's "Do we really need multiple-item measures in service research?". *Journal of Service Research*, 4(2), 155-158.
<https://doi.org/10.1177/109467050142007>
- [5] Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, 295(2), 295-336.