RASCH MEASUREMENT

Transactions of the Rasch Measurement SIG
American Educational Research Association

Vol. 28 No. 1 Summer 2014 ISSN 1051-0796

Causing and Being Caused: Items in a Questionnaire May Play a Different Role, Depending on the Complexity of the Variable

I meditated on two nice, enlightening articles that appeared in RMT, 22:1 and 22:4 (2008 and 2009, respectively). The articles, written by Stenner, Stone and Burdick concerned the causal vs. correlational relationship between indicators and variables. Basically, there are variables ("index" variables) that *are caused* by their indicators (indexes) and variables (reflective) that *cause* their observable indicators (measurement indicators). In a questionnaire, "formative" items do not generate a truly "latent" variable: they are the only game in town and their result is entirely observable. They constitute a checklist more than a true measure. "Reflective" items, by contrast, do shed some light on the latent variable, and provide (an estimate of) a true measure. Corollaries of this epistemic approach are that a) formative variables are artifactual (hence, dangerous) constructs, whereas true "latent" variables do exist whichever their indicators, and b) an ideal item-response scale should be formed by "reflective" indicators as opposed to "formative" indicators.

I entirely agree with the example of the variable "socioeconomic status" given in the former article, where "education", "income", etc., really are causes of SES rather than a selection of "reflective" items. Elsewhere, I highlighted the risks of using such artifactual constructs (specifically, Quality of Life) for concealing political decisions on the rationing of healthcare resources under the guise of "objective measurement" (Tesio, 2009). By contrast, I found debatable an example given in the second Stenner article, suggesting a wrong interpretation of the FIM as a measure, whereas it should actually be considered an "index". Let's cite the text:

"The Rasch model has been shown to fit FIM data reasonably well, which indicates that the scale locations describe adequately the relative order in which these functions are lost in the aging population. The items on the top describe difficult activities, such as climbing stairs, whereas items on the bottom describe easier activities that are maintained relatively well. (Embretson, 2006, p. 52).

Contrary to a latent variable interpretation, the FIM (Functional Independence Measure) appears to be an index of motor functioning with the causal action moving from indicators to index. If the desired medical outcome is "more functional independence," then rehabilitating bladder control, walking, bathing, and so on should promote the intended outcome rather than the other way around. Alternatively, we could teach the patient to drive a motorized wheelchair but to include this as an indicator would alter the definition of "functional independence".

I think that the FIM provides evidence of the fact that being an "index" rather than a "measure" is not necessarily an all-or-nothing concept (do such phenomena really exist on this planet?). It is true that doing effective rehabilitation exercise focused on a given item (e.g. walking) does not coax the other items towards similar improvement. In a paper of mine aimed at developing a scale of balance in multiple sclerosis patients (Tesio et al., 1997) I evidenced a trouble in the final instrument, namely in the Rasch glossary, a Differential Item Functioning (DIF) between pre- and post-treatment item calibrations. Perhaps this was not a trouble in the scale, but in the treatment! My interpretation was that traditional balance training is too focused on "resistance to external perturbations", while "resistance to self-perturbations" is relatively overlooked: hence the differential changes in relative item difficulty. Going back to the walking example, "rehabilitation of walking" is an assortment of behavioral interventions entailing stimulation of balance, force, attention, motivation, communication

etc., given that it is a teaching, highly relational activity. It is reasonable to conceive that any improvement in "walking" is indeed associated with some improvement (only some, of course) in the rest of the indicators.

The articles by Stenner and colleagues made me reflect on the fact that we cannot treat "independence in daily life" *per se*, although it is the goal of our work: we can only treat "indicators", such as continence, speech, balance, etc. My feeling is that all these can be envisaged as lying on different locations along a continuum spanning between the extreme roles of "formative" vs. "reflective" variables. For instance, in the FIM scale "bladder continence" can be altered irrespective of many other cognitive and behavioral attributes (imagine a young cognitively intact paraplegic): and in fact it is prone to relevant DIF across diagnostic classes. By contrast, "lower body dressing" implies motor and sensory skills, cognition, motivation, social relationships (why dressing the lower body if not for out-of-bed mobility and social interaction?), and it is much less prone to DIF across diagnostic classes. Of course, the more we manipulate (e.g. by treatment) a "reflective" indicator, the more we can assume we are manipulating all of the other indicators, and thus we can hope that change in the target indicator will "reflect" a change of the whole variable (and will be correlated with changes in the whole item set). I suspect that the more an indicator can be assumed to belong to the person as a whole (let me call it a *high-order behavioral indicator*, see below*)*, as opposed to body parts or focal functions, the more it is *reflective* of the latent person's variable. Thus, in principle, interventions on reflective indicators are preferable. However, at least in physical and rehabilitation medicine, this raises the risk of aiming at purely "adaptive" outcomes: if the goal is "independence in daily life" after stroke, an awkward spastic gait may appear to be an outcome equivalent to a more physiologic gait, so why bother with more fine-tuned training? The latter might require work to be focused upon highly local phenomena (such as, say, passive mobility of the ankle, knee joint kinematics etc.) which would appear as roughly "formative" once added as items to the FIM. My objections are:

1. A person cannot be described by just one variable (e.g., what about "satisfaction with the outcome"? And what about "risk for fall"?). People, not statistics, must decide what variables represent the goals of treatment.
2. Latent "persons" variables are not only multiple (potentially infinite?) but can also be thought of as located along a gradient spanning from less-to-more complexity of behaviors and perceptions (Tesio, 2003). By complexity (literally, from the Latin, *cum-plexus*, "interwoven") I mean here the number and the order of interactions across "simpler" person's traits, allowing for the trait of interest. For instance, "balance" can be thought of as of lower "order", compared to "independence in daily life": the latter implies the former, not the reverse).
3. There is a complex non-linear liaison between biological (referred to "parts" of the body) and behavioral variables (referred to a unitary "person") (Granger & Linn, 2000; Tesio, 2004). As the ancient Romans said, one should distinguish between risks "quoad vitam" (threats to life) and those "quoad valetudinem" (threats to "ability"). In fact, all living beings adapt to biological troubles in order to restore behavioral competence. People, however, are unique in that they can also treat biological problems, thus aiming at "intrinsic", rather than only "adaptive" recovery.

If my objections hold, an indicator that appears to be "formative" with respect to a high-order variable, can be "reflective" with respect to a lower-order one, closer to the biological extreme. Joint pain may be "formative" (hence, a poor item) with respect to "independence in daily life", but "reflective" with respect to "perceived effectiveness of an anti-inflammatory drug".

Luigi Tesio

References

Granger C.V., & Linn R.T. (2000). Biologic patterns of disability. *Journal of Outcome Measurement, 4*(2):595-615.

Stenner J., Burdick D.S., & Stone M.H. (2008). Formative and reflective models: can a Rasch Analysis tell the difference? *Rasch Measurement Transactions, 22*(1), 1152-53

Stenner J., Stone M.H., & Burdick D.S. (2009). Indexing vs. measuring. *Rasch Measurement Transactions, 22*(4),1176-77.

Tesio L. (2009). Quality of life measurement: one size fits all. Rehabilitation makes no exception. *Journal of Medicine and the Person,7*:5-9.

Tesio L., Franchignoni F.P., Battaglia M.A., & Perucca L.(1997). A short measure of balance in Multiple sclerosis: validation through Rasch analysis. *Functional Neurology, 12*(5):255-65.

Tesio L. (2003). Measuring person's behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine,35*:1-11.

Tesio L. (2004). Bridging the gap between Biology and Clinical Medicine. Some help from Rasch measurement theory. *Journal of Applied Measurement, 5*(4):362-366.