Reviewer's report

Title: Can the disability assessment behaviour of insurance physicians be explained? Applying the ASE model

Version: 1 Date: 13 September 2010

Reviewer: Uwe Matterne

Reviewer's report:

General comments
The authors embark on an ambitious project namely to test whether the ASE model can be applied to the prediction and explanation of insurance physicians occupational assessment behaviour. The general design is very interesting and studies like these are needed to shed light on the question as to why there is a substantial amount of variation in individual insurance physician assessment behaviour. However, the paper is afflicted with some problems. For instance, the analyses are restricted to prediction rather than explanation of assessment behaviour. Also, the method section needs major revision so that it reads more clearly. Altogether, this is a very interesting study but the article is generally difficult to read as it lacks clarity.

Specific comments

Major compulsory revision

1. The research question is not well defined. It is difficult to understand what was done before by the authors and how this and other research led to the formulation of this study’s objective. Also, the authors say that they want to illuminate the intermediary (i.e. mediating) role of knowledge and barriers. However, no meditational analysis is conducted. Tests of indirect effects could show whether such mediation takes place. It is, however, not clear whether they wanted to attempt mediation in the first place, as Figure 1 displays the two variables knowledge and barriers as moderators of the relationship between intention and behaviour. Far more effort needs to be made to provide a clear presentation of the research question in this respect.

2. Some people may not be familiar with the SEM lingo used in the manuscript; The Lisrel lingo needs to be translated into ordinary SEM vocabulary: these days there are also AMOS or Mplus users: e.g. gamma matrix = effects of exogenous on endogenous variables. There are many more examples. The reader is overwhelmed with technical language which is not necessary; just state what was done in conventional SEM language e.g. (off-diagonal parameters in PSI matrix = ?). Also, latent constructs are represented by ovals, manifest variables by rectangles; it would make an understanding of the SEM so much easier if the authors adhered to these conventions.

I also missed values for each latent variable’s disturbance (residual error), which
helps to see how much variance is explained in a dependent variable.

3. As the number of measures is substantial the authors need to find a way to describe them in a more succinct way while making sure the reader understands what they are and how they were measured. So far it is often not clear how they were actually measured, no proper examples are given. There is also too much emphasis on the Steenbeck et al. study; a study which has not been accepted for publication yet and thus cannot at present be accessed (cited about 5 times). What it reports is frequently referred to and essential for a sufficient understanding of the submitted manuscript. I also think that the measures section would definitely benefit from a generous shortening while it should more clearly describe the scales contents and their relevance to the research question. Having to read through all the details of the ASE constructs’ measurement is exhausting.

4. I have some major concerns about the measures used in this investigation. In social-cognitive research it is of paramount importance that beliefs underlying behaviour are salient, i.e. that attitudes, social norms, self-efficacy etc. are belief structures relevant to the behaviour in question and also relevant in the population investigated. Ajzen uses the TACT (correspondence in terms of target, action, context, time) principle to give some guidance as to how such salience can be achieved. Relevant beliefs are elicited by interviews and then tested in independent samples. I wonder how valid the measures are particularly as the authors mention the problems they encountered with intention. Intention is such an important concept (in the TPB and presumably in ASE too) it is seen as the most proximal determinant of behaviour. Self-efficacy was measured by a scale by Scholz et al. and probably refers to general self-efficacy as opposed to specific self-efficacy. How do you know it pertains to the assessment behaviour of insurance physicians? Why was self-efficacy not developed using the aforementioned approach (elicitation and testing)?

5. What was the objective behind the use of Homals dimensions? I know that Homals are used in homogeneity analyses. But why and how have they been used as manifest variables in the SEM?

6. Please clarify the term additive scales. Are they scales formed of a number of items? E.g. Attitude was measured by 5 additive scales (consisting of items?) and 2 Homals dimensions.

7. Another issue pertains to whether the individual scales (indicators) can be meaningfully used to represent the latent constructs. Most of the time higher values indicate more positive but sometimes more negative. E.g. in behaviour or knowledge (sometimes higher score indicate more sufficient information, sometimes higher scores less sufficient information); were some scales reversed so that higher scores always meant e.g. more sufficient information?

8. What does intention in the investigation’s context actually mean? In most social research it goes like: intention to do what. Does intention once tap: to make client return to work quickly, another time: intention to pay attention to
individual’s sickness characteristics; then: more attention paid to client’s domestic activities: What does this intention construct actually measure and how does it relate to the two behaviour constructs. Do you think the three intention items can meaningfully be combined? I think they are quite distinctly different in what they measure.

If behaviour is something like regular exercise then intention should ask something like: I intend to exercise, or how likely are you to regularly exercise.

9. Behaviour was considered a continuous measure, i.e. higher scores mean better or worse. What and how large number of items meaningfully combined, measurement model not clear; quantitative assessment of both behaviour constructs, but has to be clear that higher scores on latent construct mean better behaviour, not clear as whether can be combined; e.g. process behaviour: conflict handling seems to measure opposing qualities: once seek compromise, once more confrontational seem almost reconciliatory. That also applies to the other constructs. Were they reversed such that higher scores always represent more positive/more negative the latent underlying construct? But Table 1 negative and positive lambdas reported for most latent constructs apart from attitude.

10. Page 7: It is not clear what WIA/WAO/Wajong refer to.

11. Page 11: The abbreviation FAL is not explained.

12. Although the analysis section is very detailed it is sometimes not quite clear what was actually done; a lot of detail on the other hand can be skipped.

It is said that there were too few subjects to use a proper measurement model. Instead for each latent construct a single measurement model was estimated and the derived factor scores were then used as observed variables (indicators). I do not understand how factor scores can be used as manifest variables on an individual basis as latent factor scores are estimated over the entire sample and are a sample moment/coefficient.

As it reads it sounds like you had another measurement model of ASE constructs underlying individual factor scores? In Lisrel, input can be the raw data, but Lisrel will also accept the covariance or the correlation matrix. So why do you say you used saved factor loadings for the estimation of the structural model?

An alternative approach could be to use composite scores for the ASE constructs by adding all the individual subscales and subject them to internal consistency analyses, particularly as the sample was too small for a proper measurement model? You could have chosen the most internally consistent items for creating the composite scores. These composite scores could then be used as observed variables in the actual SEM without use of a measurement model (= path model).

13. There are conventions as to how to report analyses based on SEMs. A simple SEM (measurement and structural model as it is common standard in the literature) and from which it is clear how the latent constructs were measured and how these constructs relate to each other should contain the following: ovals (for
latent variables), rectangles (manifest variables), paths, correlations among exogenous variables; error, disturbance, explained variance etc. Figure 3 is insufficient in this respect.

14. What about missing values; did you listwise delete them; how did you handle them; imputation; FIML etc?, would be quite unusual to have a data set with no missing values.

15. Table 1 shows a standardised lambda in excess of 1. I thought standardised lambdas are in the range of 0-1.

16. page 18, about line 15. Since when do associations among disturbance terms mean associations between the actual variables? I think you mean correlations between exogenous variables, not between their disturbance terms (that term is actually reserved for endogenous variables, for exogenous variables it is just an error term).

17. page 19, discussion: please remove indirectly as you did not report any indirect effects.

18. I am not sure whether the path from behaviour assessment to self-efficacy qualifies as a feedback loop in the sense of a non-recursive model (Figure 3). However, with non-recursive models a lot of caution needs to be applied, anyway.

19. page 24, conclusion: the results from this study do not confirm the relevance of the ASE model in this setting; if at all they provide some evidence of its relevance in this setting. Important predictions from the ASE model were not supported by the data collected in this investigation.

21. I missed a comprehensive discussion of the studies’ findings in respect of its implications. Do the findings have repercussions for public health policies?

20. Title and abstract

In Title: explained refers to explained variance in outcome (dependent variable): here behaviour and intention. However, no such information is given in the article.

The abstract partly contains misleading information.

Background

The sentence ‘Very little is known about the attitudes and views that might underlie and explain the variation in the outcomes of occupational disability assessments between insurance physicians.’ is not relevant. The authors pursued a different objective and stated later in the abstract that ‘Further research is needed to determine whether the ASE variables measured for insurance physicians are related to the outcomes of occupational disability assessments.’

The term ‘behavioural traits’ is inappropriate. It refers to learned behaviour that is executed automatically. Please change.
Please remove the word 'intermediary' as their mediating role was not assessed in the study. Methods

Please change the word 'validated' to tested or something similar to avoid confusion with the more specific use of (psychometric) validation.

Conclusion

Intention was not related to content of occupational disability assessment according to Figure 3. Please change!

Minor essential revisions

The legends for the figures are insufficient. Explain exactly what coefficients are displayed.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests