Available online at www.sciencedirect.com
ScienceDirect
Personality and Individual Differences 42 (2007) 851–858
PERSONALITY AND INDIVIDUAL DIFFERENCES
www.elsevier.com/locate/paid

The golden rule is that there are no golden rules ☆: A commentary on Paul Barrett's recommendations for reporting model fit in structural equation modelling

David Markland *

School of Sport, Health and Exercise Sciences University of Wales, Bangor George Building, Holyhead Road Bangor, Gwynedd, L57 2PZ, UK

Available online 7 November 2006

Abstract

Paul Barrett offers a challenging and timely call for a re-examination of fit assessment strategies in structural equation modelling (SEM). He points out that widely accepted cutoff values for approximate fit indices have come to be treated as if they were test statistics. Paul cites four recent studies of the behaviour of fit indices under varying data conditions which demonstrate that universal indicative cutoff values cannot be trusted. Based upon these studies, Paul advocates the abandonment of approximate fit indices and greater reliance on the chi square test and a broader assessment strategy that includes predictive accuracy. I share Paul's concerns about the lax standards often adopted in model testing and I agree with most of his arguments. However, the authors he cites in support of his recommendation to abandon approximate fit indices do not reach the same conclusion as Paul. In my response to Paul's article, I discuss some conditions under which it could be legitimate to accept a model which has failed the chi square test and I contend that approximate fit indices can play a useful part in a multi-faceted strategy for determining model adequacy, provided they are not elevated to the status of golden rules.

Keywords: Structural equation modelling; Confirmatory factor analysis; Exact fit; Approximate fit

☆ George Bernard Shaw (1903) Man and Superman.

* Tel.: +44 1248 382756.
E-mail address: d.a.markland@bangor.ac.uk

0191-8869/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.paid.2006.09.023

1. Introduction

I welcome this opportunity to comment on Paul Barrett's discussion of model fit assessment in structural equation modelling (SEM) and confirmatory factor analysis (CFA) and his recommendations authors submitting papers to *Personality and Individual Differences* and for the *Journal's* editors and reviewers. I have for some time shared Paul's view that the *Journal's* Associate Editors should seek at least a broad consensus amongst themselves on the assessment of fit issue so that submissions to the *Journal* will be met with a consistent position. Also, like Paul I have become frustrated with the slippery use of language, the shifting of goal-posts, and the selective adoption of often dubiously supported criteria that many authors adopt in order to present a rosy picture of the fit of their models. All too often it is apparent that authors adopt fit criteria that ensure that they can describe their models as fitting rather than criteria that assess whether or not their models fit. In writing this response I should say from the outset that I do not consider myself to be a methodologist or a statistical expert. Rather, my comments come from the position of an occasional user and a teacher of structural equation modelling who is relatively well-informed about the exact versus close fit debate, an Associate Editor with a responsibility for making recommendations about the fate of submissions to the *Journal*, and as a reviewer for other journals that frequently send me SEM based manuscripts.

In order to gauge what is actually currently happening within the *Journal*, before starting this commentary I conducted a search for papers describing studies using SEM and/or CFA procedures published or accepted for publication in PAID since the beginning of 2006. I found 28 such papers. In only two papers did the authors declare that the fit of their models was unacceptable. All 28 reported the chi square values and degrees of freedom for the models, but only 10 reported the chi square probabilities. Of these, only three models had non-significant chi squares, although there appeared to be an error in one paper with a table showing a non-significant chi square for one of seven alternative models tested but the text reporting that chi square was significant in all cases. In this instance the apparently exact-fitting model was highly parameterised and the authors rejected it in favour of one with a significant chi square on the grounds of parsimony. In the other two cases of non-significant chi squares, the models were also highly parameterised, yielding very few degrees of freedom (in one case only two). When reviewing or editing papers I frequently find that the authors seem to be oblivious to the fact that the fewer the degrees of freedom a model has, the less the degree of disconfirmability and that finding a good fit for a model with few opportunities for disconfirmation is scientifically of little use (MacCallum, 1995). In one paper a chi square of 332.19 with 48 degrees of freedom was described as 'small', with the conclusion that the model was acceptable. One paper even declared a 'good fit' for a model with a comparative fit index of only. 619. This brief and admittedly unsystematic examination of the state of play within the *Journal* clearly shows a considerable degree of both leniency in accepting ill-fitting models and inconsistency in what is reported and considered acceptable. If such inconsistency was evident in, say, sentencing practices in our criminal courts then the tabloid newspaper editors would be screaming that 'something must be done'. Thus Paul Barrett's call for action is, I believe, timely.

In commenting on Paul's article I would first like to make some general observations and then I will deal with each of his specific recommendations in turn. As I read it, Paul's position centres on two main issues. First that the chi square test is the only statistical test currently available for SEM and second that it is not possible, nor indeed logical, to draw universal indicative thresholds