Reviewer's report

Title: Development process and initial validation of the Ethical Conflict in Nursing Questionnaire

Version: 1 Date: 26 January 2013

Reviewer: Joan Finegan

Reviewer's report:

This paper develops a scale to measure the type and degree of ethical conflict that nurses experience. The paper was well-written and thoughtful. It develops a scale that covers a wide range of ethical conflicts that nurses could face. While it can be applauded for its comprehensiveness, it wasn’t clear to me that their actual measure mapped on to their original intentions. As I understand it, the authors argue that past research has tended to look at frequency and degree of ethical conflict and believe that focusing on “the sole use of these two variables (i.e., frequency and degree of ethical conflict) to analyse ethical conflicts is insufficient”. Thus, their measure includes 19 scenarios that respondents evaluate in terms of the frequency with which they have experienced the event and the intensity of that experience, and then finally they must decide which moral state best describes how the experience made them feel. As I read the introduction, I was wondering how exactly these three measures would be combined, but on page eight, the authors present the formula for calculating the index of exposure to ethical conflict: the sum of the product of frequency x intensity across each of the 19 scenarios. Where does the measure of moral state fit in? How was it included in the assessment? The authors argue that this was one of the things that made their scale better than others and yet it isn’t made clear how it actually is used.

The measure of moral state also leads to another potential problem. The nurses’ task is to choose the option that best describes the moral state she/he experienced when faced with the situation, but is it reasonable to expect that respondents can make these distinctions? Are these categories indeed mutually exclusive? Did respondents often leave this question blank? In other words, did this question result in lots of missing data? This possibility leads to two concerns. First, if people have difficulty with a particular scenario, should it be included, particularly if it means that the data from these subjects was not included in the overall analysis? Second, what impact does this have on the reliability and validity analyses? Recall that these measures were only calculated on respondents who completed the entire question. Thus, rather than a sample size of 205 as indicated in the abstract, reliability and validity estimates were calculated on a subsample of 164 nurses. While there are different schools of thoughts in handling missing data, generally it is a worry if you have to eliminate subjects because you don’t know whether those with missing data varied in some systematic way from those who appeared in the final sample. In short, rather than just looking at the sample that completed the whole questionnaire, I wanted...
to know more about the responses of those who didn't complete the questionnaire. Was there one scenario that people had trouble with? If so perhaps it should be eliminated? If only one response was missing, then it seems pretty severe to eliminate the entire case. Could you do a mean substitution?

In the exploratory factor analysis, one factor emerged. It is possible that this illustrates that the construct is unitary as the authors believe, but it is also possible that it reflects some sort of response bias. Rating ethical situations is loaded with problems of social desirability. One possibility might be that people who were high in social desirability rated the scenarios and reactions in a socially desirable way. Alternatively, some people might simply rate things more extremely than other people (i.e., some use the end points of the scale and others do not). These response biases could account for the pattern emerging. Podsakoff, P.M.; MacKenzie, S.B.; Lee, J.-Y.; Podsakoff, N.P. ((October 2003). "Common method biases in behavioral research: A critical review of the literature and recommended remedies" (PDF). Journal of Applied Psychology 88 (5): 879–903) provides a thoughtful review on method biases. They discuss one statistical analysis that has been routinely done: In Harmen’s single-factor test, an exploratory factor analysis is done and then the unrotated factor structure is examined. If one factor emerges, then some take this as evidence of common method bias at work. Podsakoff et al discusses the pros and cons of Harmen’s single-factor test on page 889. More generally, I am concerned about the role that response biases may play and wonder how the authors dealt with this problem.

While I recognize that this paper represents the first step in scale construction, it would have been helpful if other scales were included so that responses could be compared. Given other scales exist to measure conflict, does this scale correlate with them? Is the new scale better at predicting outcomes than existing scales? I realize that this sort of analysis is often done on a second sample, but it would provide much stronger evidence of the utility of this scale. Nevertheless, the authors are to be commending on developing a comprehensive list of scenarios representing the ethical situations that nurses find themselves in. Learning more about this has the potential to help reduce the high stress many nurses feel at work and may even go so far as to help retain nurses in their profession.

Minor points
1. On page 10 it says that the alpha was .882 and for individual items ranged from .871 to .881 as shown in Table 2. Table 2, though doesn’t give a range of alphas for each item. It simply gives the alpha if the item was excluded from the scale. The way it is written currently leads me to think that particular items group together, but I don’t think that was the real intent.

2. Keep in mind with reliability the more items you have the more reliable the scale and 19 items is a large scale. So a high reliability is not always impressive.

3. The scale was delivered to participants in Spanish. Was anything lost in translation? That is, if a translator was given the items translated into English, would they re-translate in Spanish back to the same item?
Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests