Reviewer's report

Title: The medical students' perspective of formal and informal mentors: a questionnaire study

Version: 3 Date: 5 August 2015

Reviewer: D. Robert Siemens

Reviewer's report:

Minor Essential Revisions

Based on response to reviews the authors modify the manuscript to redefine the faculty mentorship program to make the reading easier. However, the authors do this incompletely through the manuscript...at times using faculty in place of "formal" but in other places in the title, manuscript body and especially in the tables, the term formal is still in place. This actually makes readability more difficult. A more careful review of the manuscript and tables to make nomenclature would make readability more clear. As it is now, the manuscript is harder to read.

Major Essential Revisions.

This reviewer is not a statistician; however, I have a decent practical knowledge of statistics and in particular its use in surveys using Likert Scores. There are several excellent short online essays/publications on the best (although multiple) ways of getting the most out of this (ordinal) data. Stats of central tendency such as mean and SD and mode as the authors use are useful and most commonly used. This reviewers point in the previous review is that there is more data in their survey that the authors do not use and do not provide to an informed reader. Frequency distribution curves (how many answered 1,2,3,4,5) for example are very useful to show the overall attitudes of the students. Mean is a nice way to describe and compare groups but much data is lost. Other options (on top of using means,SD) could include an "agreement score" such as the percent that answered 5 or 4 and 5 is another way to give the reader more information. If, for example, an answer to a particular question results in a mean of 3 in two groups BUT the agreement score (those that answer "agree and strongly agree") is 45% vs 23% this may allow a bit different interpretation. It may also help demonstrate whether or not there is a tendency to "migration to the mean" in answering these questions. The presentation of mean, SD and mode and frequency is not wrong...but this reviewer feels that the authors "blanket" interpretations of differences (or lack of differences) amongst groups just based on mean is inappropriate. For example, in the results section starting at line 150 the authors use very definitive language in comparing the groups. For example, "the faculty mentor group (mean 2.5, SD 1.1) and informal mentor group (mean 2.5 SD 1.3) did not share career interests with their formal mentors (p=0.15)"

This is, without further information, an inappropriate interpretation using means. Are the authors actually saying that NONE of the students answered positively (4
or 5) to this question and only answered neutral (3) or less? I would find that unlikely although if its true then they should state that.

The same goes for the line below this with inappropriate interpretation of student career choice. The response to this point in my previous review that this would be better left to a qualitative study is nonsensical.

Another issue that is not discussed is the practical finding that the faculty members were for the most part specialists and not family doctors. More informal members were family doctors. This finding (described) in the paper is fairly intuitive. A majority of grads from medical school will be primary care physicians. A disproportionate number of clinical teachers in most Canadian med schools are specialists and not family doctors. Therefore many of these findings in the survey are a direct result of this fact. Not that this diminishes this survey results but the lack of any comment on this very obvious fact seems like a glaring inadequacy in their interpretation.

Finally, this reviewer completely disagrees with the authors justification for the use of parametric statistical tests for this ordinal data. Their justification is wrong (in my estimation). Although one of the many assumptions needed to allow the use of parametric tests such as a T test is indeed normally distributed data—it is the distribution of the answers (not the student sample or the response rate as mentioned in the authors response). Suggesting the decision of using a wrong statistical test was a priori is inappropriate. Non-parametric methods make fewer assumptions, their applicability is much wider and are more robust. To this non-expert, the parameters needed to be fulfilled to use a T test is not attained, particularly as data in a Likert scale is ordinal and not continuous. There may be instances where parametric methods can be used for ordinal, non-normally distributed data….however the responses by the authors do not convince this reviewer. The standard in all manuscripts such as these (survey studies of attitudes etc) in my quick review is to use non-parametric tests. If the authors insist on using this test I would suggest a review by an informed statistician.

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

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**

I declare I have no competing interests