Reviewer's report

Title: Undergraduate Educational Environment, Preparedness for Postgraduate Clinical Training, and Pass Rate on the National Medical Licensure Examination in Japan

Version: 1 Date: 17 December 2009

Reviewer: Edward Krupat

Reviewer's report:

The manuscript by Tokuda et al is a potentially important piece of work, as it contains the results of a survey of all PGY-1’s in an entire country. The authors are to be commended for this coordinated effort, however an effort of this magnitude requires that the paper based on it is also substantial. Having said this, although I found the paper generally well-written, there are many areas in which I believe that it needs improvement, clarification, and expansion. Let me outline these:

• Two of the key themes the authors come back to throughout the paper are “learning environment” and “engagement,” yet I think that the authors need to expand on what is meant by these, especially the first. Learning environments can be defined and can vary on so many dimensions that the authors need to tell us more about what makes a learning environment “good” or “poor,” and how or why the learning environment should be associated with the other outcomes they measured. In essence, the authors seem to have an implicit model of how all these variables fit together, and I would think that it would serve the paper well for them to explain just how and why the variables they have measured should or might fit together.

• I am not personally familiar with the DREEM, but the authors note that it has 5 domains and therefore I would assume that it generates sub-scores on these domains as well as a total score. If it does and the scores are not too highly inter-correlated (if they are we should be told so), then I would expect that the analyses would consider these domains to inform us about whether specific aspects/domains of the learning environment were associated with the other outcomes they measured. In essence, the authors seem to have an implicit model of how all these variables fit together, and I would think that it would serve the paper well for them to explain just how and why the variables they have measured should or might fit together.

• Toward the end of the paper (in the limitations) the authors note that there is an extremely important difference between preparedness for clinical training and PERCEIVED preparedness. Yet up to that point, virtually throughout the whole paper, the authors fail to insert “perceived” before preparedness, and they sometimes write as if actual preparedness is what they were dealing with. This care in language is important in avoiding the mistaken conclusion that it is the students’ actual preparedness that they have measured. By the way, is there reason to believe that actual and perceived preparedness are related--there is literature on the accuracy of self-perceptions, most notably done by the psychologist David Dunning (see in particular his paper on this comparing people

• In the limitations, the authors also warn us about drawing causal conclusions based on the association between variables. Yet I believe that they use language that implies causality in several places. For instance, in the last sentence of their Intro their language not only implies cause and effect, but also (as noted above) confuses actual with perceived preparedness in noting that a lack of association between (perceived) preparedness and NMLE scores would suggest the need to re-think the exam. It would seem as if the value of the exam could be judged according to many other criteria, and that I would be far more likely to question why students’ perceptions were not predictive of the scores before I questioned the validity of the exam, nor would I expect that changing the exam would increase the relationship (because it is not necessarily a causal relationship).

• A high survey response rate is always difficult to get, so I sympathize with the authors, but an overall rate of 36% is fairly low. However, here the authors could tell us more. I would like to know: what was the range of response rates across schools; what percentage of the PGY-1 population was male and female, and what percent of the respondents were males and female. If told this, we could at least determine if the respondents were demographically similar to the larger population (at least in gender, although they may have other information on the population of PGY-1’s which could be compared to determine if the sample of respondents was representative of the total population of PGY-1’s).

• In the discussion, the authors note that the level of self-reported preparedness was lower in Japan than in the US. I do not know the extent to which this can be accounted for by differences in the way that various cultures express modesty in their opinions of self, but if the authors are to make cross-cultural comparisons, which I believe is a good idea, they should also cite literature or at least speculate whether these are likely “real” differences in preparedness and perceptions of preparedness, or simply the result of different cultural norms about expressions of self-confidence or mastery.

• The authors note many times that the medical schools in Japan vary greatly in the extent to which they have adopted the Model Core Curriculum. If this is the case, then at the least the authors ought to find some way of characterizing the medical schools (based on the opinions of the school deans? based on a review of their curricula? based on the opinion of respected third parties?) This would allow them to determine the extent to which perceived preparedness, DREEM scores, and pass rates are associated with adoption and implementation of the new model, which could become a key finding that would pull all the results together.

• Since it appears that the national pass rate is around 90% or more, the value of this outcome is limited for analysis purposes (if there is little variability in a measure, then associations between it and other variables will be small and/or of limited use). However, I am certain that students get an actual score, which is a continuous variable that undoubtedly has a good deal of range. Therefore, I would strongly advise the authors to re-analyze using actual scores rather than Pass-Fail, which might make their results look greatly different.
In short, this project is a notable one, and its findings very much deserve to be published. However, I believe that the authors need to do a considerable job of re-writing and possibly some re-analysis, after which I believe it could be a paper with considerable impact.

Level of interest: An article of outstanding merit and interest in its field
Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.