Reviewer's report

Title: Identity theory and education completion: A Text Analytics study of 4,432 online discussants of an ‘Understanding Dementia’ Massive Open Online Course

Version: 1 Date: 25 August 2014

Reviewer: Dirk Van Rooy

Reviewer's report:

Major Compulsory Revisions

1. My greatest concern is that the methods, as they are described currently, seem to be less than systematic. A more detailed description of the analysis (why it was chosen, what the limitations are, etc.) to minimize impressions of subjectivity would improve the paper significantly. The text currently mentions that "Elements of the methods described below have been published by us using other datasets [author citation to be included after review]“, which makes it impossible to fully evaluate the method at present.

2. The introduction fails to provide a clear definition of the question(s) being asked. Because of that, the paper appears unstructured and lacks a clear conceptual structure. It is only when we get to the section “The importance of identity to education completion and disadvantage” that a somewhat clearer narrative seems to develop. It would be much better I believe to start with this section, and then clearly specify how the proposed methodology and analysis address the issues. Which brings us to the next point -

3. The paper seems to be focused (amongst other things) on how learners that are able to integrate the course material in terms of a relevant self-narrative are more likely to complete a course. But it does not really provide a method that allows studying this issue in a systematic way, which leads to conclusions such as “Understanding Dementia MOOC’s success in achieving high completion rates for those with lower formal education appears to be linked to its diverse use of biomedical and experiential or quality of life knowledge”.

4. The discussion and conclusions go beyond the data, and are often confusing and difficult to follow. No conclusive quantitative analysis are provided to back up qualitative interpretations of the figures. For instance, linguistic analysis often applies some form of principal component analysis or factor analysis to explore the relationship between the different factors that emerge from the explorative phase of the text analysis, and how they relate to for instance demographic variables. Such an approach would also allow to study interactions and mediations. Although the conceptual mapping suggests a number of potentially interesting findings in terms of the spatial distribution of completers and non-completers, there is no real quantitative analysis to provide more conclusive and reliable interpretations. Instead, the text resorts to qualitative, verbal
interpretations of the figures. Employing more quantitative analysis would allow making more conclusive statements about the issues at hand.

4. The limitations of the current approach could be spelled out better. For instance, how does the applied analysis compare to other linguistic analysis? Why was this approach chosen?

5. The results and discussion sections are poorly structured. It might be better to map it more closely to the analysis phases introduced in the methodology. Also, a number of issues are introduced that were not properly foreshadowed in the introduction (for instance “achievement gap”), which creates a disjointed narrative, and makes it harder for the reader to follow the text. Also, the text regularly has statements such as “This raises questions about the significance of such typical storylines involving self-referential language in successful completion of MOOCs by those with lower education levels” without providing any systematic follow-up (which is partly the result of the lack of conclusive findings).

6. The study does not seem to be embedded well within an existing literature. This is related to the failure to clearly define the questions the paper is addressing. I believe the current analysis paper would best be written up as an explorative paper, applying novel methodologies. But in that case, the methodology and its limitations needs to be better defined (see points 1 and 3). At the moment, the paper is caught between an explorative analysis and an attempt to address a number of vaguely issues.

Minor Essential Revisions.
1. Perhaps it would be helpful to include examples of what constituted the “large data-set of discussion posts”.
2. It is mentioned that the “me-concept” map involved minimal manual edition, but it is not clear what that actually involves as the methods are described unclearly.
3. Figure 4 “suggest” that “system”, “food” and “eat” are among the concepts most often paired with the “me-concept”. It is not clear to me how the authors conclude from this figure that “experiential, life-matters and quality of life concepts” are most often paired with the “me-concept”.

Level of interest: An article of limited interest

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