

Interactive comment on “Demonstrating change from a drop-in engagement activity through pre- and post- graffiti walls: Quantitative linguistics and thematic analysis applied to a space soundscape exhibit” by Martin O. Archer et al.

John Coxon (Referee)

work@johncoxon.co.uk

Received and published: 19 October 2020

In general I enjoyed this manuscript and feel it would be suitable for publication in Geoscience Communication with some changes which I outline below.

95–99: I was unfamiliar with Zipf’s law, and I think applying this to evaluate public engagement in this manner is highly novel and very impressive. The less negative the exponent, the wider the vocabulary. In Figure 3, this clearly demonstrates that for the ~10 highest ranks, there was a significantly wider range of vocabulary. However, I do

C1

have some concerns about things that are unclear. Firstly, I think it is necessary to express and discuss the percentage of the words captured in the ~10 highest ranks; from an inspection of the graph it looks to me like ~90% for the “before” set and ~80% for the “after” set, so the vast majority of the language used on the walls is presumably contained in those 10 upper ranks (if I’m misinterpreting this graph, quoting these numbers explicitly is even more important). Secondly, the authors say “While the exponents in the lowest ranked segments are consistent with one another,” but the two exponents are not consistent with one another; $-1.00 \pm 0.04 \neq -0.85 \pm 0.05$. This part needs to be rewritten to acknowledge that a) although the two are similar, in fact the variety of language at rank > 10 decreased after the study compared to before; and b) what implications this has.

105: The authors say that “instead of using pre-determined qualitative codes, the analysis drew on grounded theory”. I am unfamiliar with the word “code” used in this context. Appendix B particularly, but also the citations, helped me grasp what the meaning is (i.e., for each word, identifying its theme and then within that theme whether it means one of two extremes, i.e. “quiet” or “loud”, “empty” or “full”, and “slow or busy”). However, I am concerned that readers might not be as tenacious in reading around the terminology as I was as a reviewer, and so I think the paper would be much improved with a fuller explanation here.

108–112: Tying into my criticisms above, a better way to express this would be a table which explicitly outlines which of these are themes and which are codes, I think.

113: I had not heard of Cronbach’s alpha prior to this manuscript. Most of the examples I have found of its use during my reading for this review refer to applying it to Likert-scaled quantities to compare the extent to which, say, a questionnaire of questions about a given theme are telling the author about that theme. This indicates two things; firstly, to apply it the themes must have been assigned numerical values, but it is not clear to me what those values were from the explanation herein. If the authors use Cronbach’s alpha, they need to explain in more detail how the themes were quantised

C2

in order to apply the measure (are these simply the 1/2 numbers expressed in Figure 4? If so, say that, and if not, then what are they?). Secondly, as far as I can tell from my exploration of the literature, Cronbach's alpha is applied to measure reliability, not dimensionality (as in this study); in fact, some literature I read specifically cautioned against the latter and the Cho paper cited does not seem to provide a rationale for using the measure in this way. I have no real issue with the central analysis tenet in Section 4.2, namely that the coder can differentiate to which theme a word belongs. However, if I did have an issue with it, the quoted statistic would not convince me, and should either be removed or the reasons for its validity should be expanded on significantly.

Interactive comment on Geosci. Commun. Discuss., <https://doi.org/10.5194/gc-2020-41>, 2020.