



EGUsphere, referee comment RC1
<https://doi.org/10.5194/egusphere-2022-494-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-494

Anonymous Referee #1

Referee comment on "Exploring TikTok as a promising platform for geoscience communication" by Emily E. Zawacki et al., EGU sphere,
<https://doi.org/10.5194/egusphere-2022-494-RC1>, 2022

Note: My familiarity with Tik Tok is minimal and my experience with it is limited to viewing videos that are occasionally forwarded to me. I am, however, familiar with other media platforms and their use for geoscience communication. Furthermore, I am a geoscientist interested in communication and using emerging platforms like TikTok.

General comments:

The manuscript presents the TikTok analytics data gathered for 48 educational geoscience videos produced and published between Oktober 2021 and February 2022 on the media platform TikTok. Given TikTok's increase in popularity as a media platform, exploring its potential as a means for geoscience communication is a very merited undertaking. There are no language issues and it's easy to follow the text. Unfortunately, the study lacks direction and it's unclear throughout the manuscript what exactly the authors intend to use the TikTok analytics data for. In other words, there's no apparent experiment design. This is reflected throughout the manuscript in several ways. These include, but are not restricted to, the following:

- 1) It's unclear why some of the x-y relationships are looked at (e.g., average view time and total views) and what the implications of a relationship would be. Many figures are redundant or show no patterns, making them seem like simple "data dumps". (For example, why show gender ratio over time when it does not change?). In other figures, the authors observe patterns I cannot see (e.g., Fig. 3).

- 2) The last criticism highlights another major flaw of the study, i.e. the complete lack of

statistical analyses that would introduce rigor and objectivity in the detection of relationships and patterns in the data. It is difficult to make specific recommendations for methods at this point since I am unsure about the overall goal and experiment design for the study. Given on what I can infer from the presentation and discussion of results, however, I imagine simple correlation analyses and ANOVAs (with "significance" test) would be suitable for the largest part.

3) The study includes US-centric assumptions that compromise important take-away messages of the study. E.g. the authors use the MST time zone (a US time zone) not just to communicate their results, but also to define the time of the day. This is done without knowledge of the location of the viewers, who may live in a completely different time zone. Since a large portion of topics are of general interest and TikTok is a global media platform, it's not unreasonable to assume the reach goes beyond the US. If it does, nothing can be said about the preferred viewing time of the audience or, by extension, the optimal time for upload. This could be avoided by good (/more thoughtful) experiment design.

4) The final conclusions and recommendations are mostly not supported by the study due to several of its flaws, the lack of clear (testable) hypotheses and lack of rigorous testing.

As such, I cannot recommend the manuscript for publication. That said, I appreciate the efforts that go into making educational videos and maintaining a presence on a media platform. With the channel and videos, the authors are on good track to test the effectiveness of TikTok (or aspects of it) and I strongly encourage the authors to do so. However, this will require a clearly thought through study/experiment design. This should include a clear idea of what questions should be addressed with the experiment, what hypotheses are tested, what the metrics are used for testing, etc.. This also includes a clear idea of how data (e.g. TikTok analytics data) can be used in it and what its limits are.

Specific comments:

The title is somewhat misleading. The study does not test the effectiveness of TikTok as a platform for geoscience communication. An important part of effective communication is what happens on the receiving side (e.g., information retention). This study looks solely at reach and basic form of engagement.

Line 16 - The term "retention" is a tad misleading. Until it was explained, I took it as a measure for how much of the video was remembered.

Line 20 – What is the “For You” page? If it’s a page with personal video suggestions, maybe use a more general description of it, such as “suggested viewing page (“For You” page on TikTok)” for those unfamiliar with TikTok. It’s explained in section 2, but I’d use a more explanatory and general term for it in the abstract so it doesn’t require knowledge of TikTok.

Line 22 – “Went the most viral” – I’m aware that such terminology has established itself fairly well by now (among specific demographics at least), but I nevertheless recommend rephrasing that for an audience that is unfamiliar with it yet have an interest in exploring such media platforms as a communication tool.

Lines 53-55 – It’s unclear what the purpose and nature of the analysis is. If it is to determine factors of a video that help maximise reach (line 55), why were engagement and view duration analysed at all? Usually, an introduction outlines the broad questions, goals and hypotheses tackled by a study. This helps guide the reader through the manuscript. At this point, I still don’t know what these are or what analyses were conducted on the 3 metrics and for what specific purpose. (Unfortunately, I also miss clarification of this down the line in the following sections).

Lines 58-77 – I appreciate the introduction to TikTok (section 2) and think it’s important. I would add some comments on TikTok’s reach. Who are the viewers, how are they distributed globally? This may be very important information for communicators who want to target specific communities.

Line 67-68 – This is a tad vague. What are these recommendation algorithms based on exactly? The authors list “user’s profile settings”, “location” and “activity”. Can users choose topics of interests in the profile settings? Since the study’s results are, to a degree, sensitive to this, I think a closer look at it and a more elaborate explanation is warranted. This will also help communicators plan their activities on TikTok better.

Line 77 – It’s unclear what is meant by searching sounds. It reads as though sounds are used as hashtags are. Is that accurate?

Lines 85-93: What is the longevity of such hashtags on TikTok? On twitter, for example, they may rapidly change and quickly lose popularity and reach. Is #LearnOnTikTok ongoing and intended as a “stable” hashtag? Was the associated campaign was a one time event or something ongoing and sustainable? If it’s ongoing and the hashtag serves almost like an education “channel” on TikTok now, this should be communicated. As a communicator, this would be very useful for me then. If it’s not expected to survive long as a hashtag, on the other hand, it doesn’t warrant as much explanation.

Lines 104-105: I disagree with the general statement that geoscience does not lend itself

to experiment-based content as much. On the contrary, since geoscience is effectively physics and chemistry applied to the Earth, many classic physics and chemistry experiments can be used to demonstrate how something works on Earth. In atmospheric sciences alone, there are plenty of pressure and moisture experiments that can be conducted (at home) to demonstrate important concepts. On the other hand, for Earthquake science, experiments involving concepts of friction and liquefaction come to mind.

Lines 123-124: Was this by design and are there studies of how the metrics used here (reach, engagement and viewing time) vary depending on the gender of the presenter? Given prevailing differences in perception of scientists of different genders, this may be important information for geoscientists who consider TikTok as a communication platform.

Section 4 (general): It's unclear what determined the choices in video characteristics and how these choices were made with a specific experiment/goal in mind. This is essentially the "experiment setup" section, but I cannot see how the choices made here serve the testing of a broader hypothesis or help answer the study's overarching question(s). For example, if one were to determine the optimal video length to maximise the engagement rate, one would vary the factor of interest and reduce the number of other free parameters (such as upload time). Reversely, If I wanted to determine the best time to upload a video, I would vary this parameter and keep others constant. I don't see this type of experiment-based and hypothesis-guided thinking in this section.

Line 175: Please express time using the primary standard UTC (or list in in brackets). As someone who's not from North America, I was completely unaware of MST.

Fig.2: I think the text-based summary of views is sufficient. The figure does not show any pattern or provides any valuable additional information and thus seems a tad redundant.

Fig 3: Unless most viewers are known to be from North America and ~ the MST time zone, I would change this to UTC. In fact, if the viewer location does not coincide with the MST zone, discussing times of the day may be misleading. Furthermore, I question the merit of presenting this data at all, since (a) there is no clear pattern except maybe an increase in views towards the evening, and (b) global viewers will likely be spread across different time zones.

Fig. 6: The ratio remains ~constant over time, making the figure rather redundant. It's sufficient to simply state the ratio, as already done in the text.

Lines 215: Please change the term "retention" to something more intuitive like "view duration". You use the latter term already. I recommend simply sticking to that, also to avoid confusion with viewer information retention (memory).

Fig. 7: It is unclear why video views is plotted against video "view duration". Is a relationship expected? Why? What would be the meaning of it?

Fig. 8: See comment on Fig. 7. It's unclear why views are plotted against engagement rate. What would the discovery of a relationship between these metrics help with? What hypothesis or idea would that address?

Fig. 9: See comment on previous 2 figures. What's would be the significance of the x-y relationship? Furthermore, the caption does not guide the reader through the figure in any way. It does not explain anything and I am left wondering why it is shown and what the authors would like me to see in it.

Section 6.4: I understand why duration and time posted may be valuable information. These may be useful for communicators in their planning to maximise views. However, I do not see the significance of a relationship between views and average % watched.

Section 7: Any thorough, literature-informed discussion of the study's limitations (and the limits to recommendations made) is missing. For example, I think it would be important to:

- Comment on importance of different metrics for assessing communication effectiveness.

- Comment on how sensitive the study's results are to the specific algorithms used by TikTok, how transparent they are and how (often) these may change. How does this affect the take-away messages of this study?

- Comment on technology exclusivity and how it limits reach.

- Comment on the effect of video content on audience demographics/nationality/location. This may be important in context of the interpretation of the viewing times also.

Lines 405-412: I question several of these recommendations (and related conclusions) based on some of the study's flaws:

(a) The recommendation to post earlier in the day seems based on the discussed relationship between time posted and video views. More precisely, it seems based on the observation that videos that received the most views were uploaded in the morning or early afternoon. By simply looking at the data plot, I do not make the same observation. In fact, it is difficult to see any significant pattern at all except maybe a general increase in video views later in the (MST) day. This contradicts the authors' findings and highlights the need for rigorous statistics to make sure a signal/observation is real. Furthermore, this recommendation is very US centric and/or vague. To elaborate: The authors use a US time zone to define the time of the day. Does the recommendation pertain to MST morning or morning in whatever time zone the communicator lives? To answer this, the authors need to establish where the views are from. If they could establish (1) that their viewers are US (or MST) based, and (2) that there is an actual relationship between views and time of day, a recommendation for upload time could be merited (albeit restricted to the US). As it is, however, no such conclusions can be drawn from the TikTok analytics data.

(b) As a communicator, I want to communicate a specific set of problems and topics. Recommendations to choose a topic based on what is newsworthy (or tied to a specific location) is therefore questionable. This recommendation seems focused on increasing views (regardless of intended subject of communication) rather than on effective communication of a specific topic. The recommendation to include gender-related topics is similarly problematic. I would like to know "What do I do to increase the number of female viewers for videos about this specific topic?" and not "What topic should I choose to increase the number of female viewers." The recommendation would be merited if the strategy is to include more gender-related topics to gain more female followers, who then view videos of other topics. However, this strategy is also questionable given that relatively few views are gathered from "following" (Fig. 5).

(c) Given how few views the videos got from hashtags (I cannot even see the percentage of it for most of Fig. 5) I do not see how the authors can list this as a recommendation here.