Review of: "Can Twitter be used to improve learning outcomes in undergraduate medical education? A pilot study"

Renslow Sherer[1]

1 University of Chicago

Potential competing interests: No potential competing interests to declare. Contributions to review by Hongmei Dong, Jonathan Lio, and Ivy Jiang.

Review of QEIOS 1.6.23

Comments for the author

Abstract:

Methods line 5, there is a typo '….questions for ane an engagement' – please revise and clarify.

What is the total N? In the abstract it says 32 were enrolled, but in table I it reports results for only 29 results, and elsewhere it says 46 people were enrolled. Please clarify.

Abstract states: "The purpose of the investigation was to study " … learning outcomes defined as scores and passage on NBME Shelf examinations…." In the Results section, scores of NBME and COMAT are reported. The paper should clarify how the two tests are related.

The acronym 'CLES' should be spelled out as Common Language Effect Size in the abstract.

It is not completely clear in the abstract what the intervention is. Are participants being tweeted questions on family medicine topics? Or is there something else in the tweets?

In the abstract, it is not clear in what ways the intervention group differed from the historical controls.

It seems useful to present the full findings in the abstract as well, if space allows, i.e. that no difference was found for NBME shelf exam scores, but meaningful differences were found for the COMAT scores. The implications of these discordant findings should be considered in the discussion.

Narrative

Pg , line 7: Define 'digital native'

Pg 4, line 14: If the goal is to find whether twitter can enhance learning outcomes, i.e. NBME Shelf scores, then the expected outcome would be superior scores in the Twitter intervention group.

The Literature Review section is valuable on this topic of great general interest to medical educators.

In the methods section, “Twitter use” does not seem to be the only difference between the two groups – the comparison group did not use Twitter AND did not learn by answering the 90 questions. It is quite possible that being required to answer the 90 questions could improve learning, whether the questions were given to students via Twitter, or text messaging, or course webpage.

Theories: ‘Spaced learning’ is poorly and inaccurately defined. Given the lack of a detailed description of the intervention, it is hard to tell if the theory was appropriately applied.

Pg 6: Since this paper is about learning outcomes using Twitter and not the study questions themselves, would a more appropriate control be a group that accessed the questions apart from social media, e.g. email?

Pg 6: What does MSCES stand for?

Pg 6: The paragraph under “Design, setting, & sample” explaining the different surveys used is confusing. Why was it changed to Survey Monkey after 10 months? How many surveys did participants complete?

Pg 6, 2nd to last line: The authors should comment on the 280 character restriction imposed by Twitter. In their judgement, did this impact the quality or utility of the questions, and more broadly the entire process? Some clinical questions may not fit into this tight corner.

Pg 7, Question testing: Comma after “where possible”

Pg 7, Intervention: As a practical matter, can the authors clarify their time dedicated to the process, including sending out the questions with spacing of 10 minutes between learning topics, during which they were available for bilateral discussion with 32 students shared by 3 faculty members, including the one daily timing of responding to Twitter feedback?

How do the investigators know that the students are doing something other than studying in-between the content?

Pg 8 Analysis: For the one-sided test, might there have been a decrease in scores in the Twitter group from being distracted by Twitter?

Pg 10, last paragraph in results: Average scores are reported without reference to whether this was for the COMAT or the NCME Shelf exam or both, please clarify.

Pg 11, para 2, line 11: The authors posit that the positive change in unadjusted Shelf same scores may have been due to student engagement. This seems likely. More comment is needed here – do the authors view that as a negative finding? Isn’t that one of the points of frequent Twitter communications, i.e. to capture or re-capture student attention to this content?

Last paragraph: The practical question above re: faculty time also applies to this notion of ‘anytime, anywhere and just in time experiences’. Is that an accurate description for this interaction? If faculty only review the twitter responses once

daily, there is an inevitable lag between question and answer, possible for hours or days.

Last paragraph – spell out BLLM, please.

Limitations: this section is well done. The authors should also consider the small sample size as a limitation, which increases the likelihood of selection bias. For the same reason, their optimism that engagement was not a factor and that their analysis accounted for it is not persuasive.

The last line in the limitations section should be repeated in the methods, as it is an important component of study implementation. Why did the authors choose the format? This seems to undermine the 'anywhere/anytime' nature of the benefit of twitter, and to introduce a less sustainable dimension to the intervention.

Descriptions of the following are unclear, incomplete, or disorganized: the study timeline (timing of intervention, the exams, the surveys), contents of the research instruments, rationales for using multiple surveys, and details of the intervention procedure.

As a minor concern, the use of too many acronyms hinders comprehension.

Summary:

This is an interesting attempt to study the use of Twitter for the concrete purpose of improving scores on NBME shelf exams and COMAT questions. However, the investigators do not appear to have studied the effects of Twitter use as a unique social media app, but rather engagement with family medicine content questions. As the authors pointed out, there was no conversation within the Twitter platform. The authors offer a useful literature review of previous modest attempts to utilize social media in medical education. In this study of 32 (or 29) students in a family medicine clerkship, they found improvements in COMAT question scores, but not in NBME scores, compared to historical controls using a structured bidirectional twitter engagement around validated test questions. The study limitations are substantial with a very small N and only a historical control group, and a likely high and un-assessed degree of unfamiliarity with Twitter among the medical students. The biggest flaw is that the actual intervention is poorly described. The actual burden on faculty for real time implementation should be more clearly described, as it is not at all convincing that this intervention is either practical or justifiable.

There are unnecessary repetitions (the paragraph that begins on the bottom of p3 and the 1st paragraph of the Discussion section). The overall quality of this paper is weak, and I would not regard its findings as meaningful or reliable as current presented.