Open Peer Review on Qeios

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

Gregory E. Gilbert, Lisa A. Paganotti[1], Anubhav Agarwal[2], Jeffrey Salalac[3], Christine Gamble[4], Eric B. Bauman, Jacqueline Gallardo[2]

1 George Washington University
2 Mount Sinai South Nassau Community Hospital
3 St. John's Episcopal Hospital
4 Spartanburg Regional Medical Center

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

Background: The widespread use of social media has led to exploration of its use as an educational tool to engage learners and enhance interactive learning with its use in the healthcare literature dating back to 2008. If an intervention can impact clerkship Shelf Exam scores as a measurable short-term goal, it may have a downstream impact on USMLE or COMLEX-USA examination scores. The purpose of the investigation was to evaluate if Twitter could be used in undergraduate allopathic and osteopathic medical students to enhance learning outcomes defined as scores and passage on NBME Shelf examinations taken directly after the intervention period.

Methods: This investigation used a quasi-experimental study design. Ninety questions covering 30 different Family Medicine topics were developed through a psychometrically sound process. Following sample size calculations 32 3rd-year allopathic and osteopathic participants were recruited from a Family Medicine clerkship at a community-based hospital in the Northeast. Participants were sent tweets during two weeks of their Family Medicine clerkships. Upon completion of their clerkship, surveys were emailed to participants containing questions from ane an engagement survey. A historical comparison group of individuals completing their rotations immediately prior to the intervention was used to assess a difference in Shelf examination scores. A two-group Wilcoxon-Mann-Whitney test was used to examine differences.

Results: There was no statistical evidence of a difference between percentile scores *P*-value=.157). The CLES for COMAT/NCME Shelf Exam percentile was 62% - a meaningful effect size.

Discussion: This study adds support for the use of Twitter to improve learning outcomes during medical student clerkships. Our study built upon this finding by adding the bidirectional aspect of Twitter communication. The change in scores change may have been due to student engagement. Twitter demonstrated a social media approach to providing students with anytime-anyplace and just-in-time experience consistent with Bauman's Layered-Learning Model. The chief limitation of this study was medical student in-person clerkships were paused during the COVID pandemic and therefore our study enrollment was also paused and posed many more logistical challenges. The results of this investigation further the potential for Twitter to be used as an inexpensive educational intervention to modestly improve standardized exam scores with the potential to be used in a broader fashion (e.g., COMLEX and USMLE Step Scores). Future research can build upon this study by applying this intervention to other locations with larger samples. This study should be replicated for each core clerkship rotation for COMLEX and USMLE Step preparation.

Gregory E. Gilbert[1*], **Lisa A. Paganotti**[2], **Anubhav Agarwal**[3], **Jeffrey Salalac**[4], **Christine Gamble**[5], **Eric B. Bauman**[7], and **Jacqueline Gallardo**[7]

1. Gregory E. Gilbert, EdD, MSPH, PStat®
   Biostatistician
   SigmaStats™ Consulting, LLC.
   1865 Bairds Cove
   Charleston, SC USA 29414
   +1 843 870 2305
   sigmastatsconsulting@gmail.com
   ORCID ID: 0000-0003-0879-5496
2. Lisa A. Paganotti, PhD, PA-C
   Assistant Professor
   School of Medicine and Health Sciences
   Department of Health, Human Function and Rehabilitation Science
   The George Washington University
   1918 F Street NW
   Washington, D.C. 20052
   +1 313 570 5176
   lisaannbuckley@gwu.edu
3. Anubhav Agarwal, MD, CAQSM
   Associate Residency Program Director
   Coordinate for Primary Care Sports Medicine
   Clerkship Director Medical Student Education, Department of Family Medicine
   Clinical Assistant Professor, Icahn School of Medicine at Mount Sinai
   Mount Sinai South Nassau
   One Healthy Way
   Oceanside, NY 11572
   +1 516 255 8400
   Anubhav.agarwal@snch.org
4. Jeffrey Salalac MD
   PGY1

Department of Family Medicine

St. John's Episcopal Hospital

327 Beach 19th St

Far Rockaway, NY 11691

+1 714 328 2073

jeffreysalalac@gmail.com

5. Christine Gamble, MD

   Spartanburg Regional Health System

   Department of Family Medicine-Obstetrics Fellow

   101 E Wood St

   Spartanburg, SC 29303

   +1 702 204 2318

   gamblechr@gmail.com

6. Eric B Bauman, PhD, FSSH, PMHNP-BC, APNP

   Clinical Playground, LLC.

   1615 Maple Str

   Middleton, WI USA 53562

   +1 608 469 6200

   ebauman@clinicalplayground.com

7. Jacqueline Gallardo, MD

   Academic Chief ResidentDepartment of Family Medicine

   Mount Sinai South Nassau

   One Healthy Way

   Oceanside, NY 11572

   +1 909 576 9247

   jacquelinegallardo@gmail.com

* **Corresponding author**

Keywords: Bauman's Layered Learning Model, Twitter, Spaced Learning, Shelf Exam, COMAT, USMLE

Social media platforms such as Twitter, Facebook, and Slack have become widely popular and their use has increased exponentially over the last two decades.[1] Twitter has 330 million active monthly users with roughly half of users tweeting daily.[2] The widespread use of social media has led to exploration of its use as an educational tool to engage learners and enhance interactive learning with its use in the healthcare literature dating back to 2008.[3][4] Through an understanding of digital connectivism, faculty can use social media to compliment traditional teaching techniques and facilitate student-centered learning. Twitter has been used in medical education as a pedagogical platform; however, the use of social media as an innovative learning tool to leverage learning outcomes in undergraduate medical education is largely unexplored.[5][6][7][8][9][10] The current research has studied perceptions of Twitter use and student engagement, but has not linked it to learning outcomes. Digital natives seek routine incorporation of social media in their education. Medical students desire efficient learning and social media can be a tool to satisfy this need.[11]

In medical education, allopathic medical students take National Board of Medical Examiners (NBME) Shelf Exams and osteopathic students take the Comprehensive Osteopathic Medical Achievement Test (COMAT) after each core rotation. These NBME Shelf Exam scores correlate with, and help students stay on track for passing the United States Medical Licensing Examinations (USMLE) and Comprehensive Osteopathic Medical License Examination of the United States (COMLEX-USA).[12][13] If an intervention can impact clerkship Shelf Exam scores as a measurable short-term goal, it may have a downstream impact on USMLE or COMLEX-USA examination scores. The purpose of the investigation was to evaluate if Twitter could be used in undergraduate allopathic and osteopathic medical students to enhance learning outcomes defined as scores and passage on NBME Shelf examinations taken directly after the intervention period.

Literature review

A search of the literature using PubMed and the following composite terms “(Twitter [title/abstract]) AND ((medical[title/abstract]) AND ((education[title/abstract]) OR knowledge[title/abstract]))” in November of 2022, revealed 413 articles. Of those articles, four had to do with direct learning outcomes, and one (Reames et al.) dealt with Shelf Exam scores.

The four studies examining outcomes include Webb et al. who performed a study using 116 volunteer medical students in a clinical medicine course.[14] They compared students who used Twitter frequently to those who used Twitter little or none of the time using an in-class photo quiz. Students retained less information from Twitter than from the classroom images as demonstrated on their quiz. Hennessy et al. used Twitter to attempt to enhance learning outcomes in neuroanatomy teaching and were successful at relieving anxiety and increasing morale throughout the course, but didn’t increase exam scores.[15] Jurivich et al. conducted a “controlled, prospective, 2-year cohort observational study…to test whether weekly geriatric questions delivered through Twitter Poll could improve geriatrics knowledge during an internal medicine clerkship for third-year medical students.”[16] Pre- and post-rotation multiple-choice test results were compared and demonstrated an increase in knowledge in the intervention group compared to the control group.[16] Jurivich et al. were the first to highlight the use of Twitter to improve learning outcomes in medical clerkship students. Reames and colleagues used a pre-test/post-test prospective observational study using medical student volunteers from surgical clerkships.[17] They sent three new tweets per day with succinct, objective surgical facts and assessed aggregate test scores for participating students and historical controls.[17] They found Shelf Examination scores were not significantly different in the intervention group. This study was most closely aligned with our target population and intention to attempt to increase Shelf Examination scores for medical students. To build upon the above studies, the authors used an intervention implementing bidirectional Twitter communication to assess if this can enhance Shelf Examination scores.

Theoretical framework

Cognitive psychology, neuroscience, and learning research are the scientific basis for education. The study of these subjects is an empirical science with a solid theoretical foundation and research-based uses with application to higher education.[18] It is imperative education be more deliberate and systematic in its use of evidence-based methods. To this end, this investigation uses two theoretical frameworks, spaced learning and Bauman's Layered-learning Model (BLLM).[19][20][21] Spaced learning is used to convey and embed content in short-term memory to long-term memory, while BLLM is the paradigm under which the learning occurs.

In 2005 Fields wrote a paper describing the neurological basis for the transition of short-term memories to long-term memories.[22] Subsequently, capitalizing on Fields' work, Kelley & Whatson used three repeated stimuli separated by ten-minute intervals to test the possibility of enhancing the encoding of long-term memories in students. They were successful in demonstrating this possibility.[23] Spaced learning repeats learning content three times with two ten-minute breaks. The breaks consist of activities unrelated to the learning content. The breaks are key to spaced learning, and theories of why spacing is effective abound. The different theories of spaced learning may or may not work together yielding the memory advantage produced by spaced practice.[24] One of the more prominent theories is repeating an item may prompt retrieval of previous presentation, thus engaging a process to enhance memory.[25] In essence this practice retrieval process readies the learner for subsequent instances when relevant material will need to be recalled later in a more situated circumstance, for example during an examination or in a clinical setting. In other words, spaced learning may be viewed as "practicing forgetting". Toppino & Gerbier offer additional theories explaining the benefit of spaced practice for long-term retention.[26]

Bauman's Layered-learning Model describes a matrix for scaffolding didactic materials delivered through traditional teaching methods such as lecture, discussion and reading assignments with the presentation of additional course content by leveraging multimedia educational technology.[19][20][21] In this way the model presupposes traditional didactic learning techniques, such as knowledge transfer through scholarly reading and interaction with faculty and staff, still have relevance in the modern classroom. However, the model leverages contemporary educational technology to scaffold the transfer of knowledge to the learners in a situated and multimodal approach. The importance of the situated context of the content is not to be overlooked. Using multimedia approaches to teaching and learning including traditional educational communication such as reading assignments, small group discussions, as well as contemporary digital media approaches like instant messaging platforms prepares learners for contemporary practice environments. The contemporary practice environment is dependent on the ability of clinicians to access information from multiple sources and often synchronous to evolving clinical encounters. By providing students with anytime-anyplace and just-in-time learning experiences, BLLM provides this investigation with a dissemination and evaluation of knowledge model consistent with actual practice settings.

Bauman's Layered-learning Model does not replace traditional reading assignments with technology, but rather provides an approach to learning by increasing access to content and avoiding the pitfalls of privileging information. By making content available through digital and mobile media and through supervised clinical experiences students begin to achieve real-time, anyplace learning experiences. Learners can leverage an array of resources including, faculty-led classroom experiences, books (print or digital), social media, and multimedia games or simulations to provide as-needed or just-in-time learning to meet course objectives. Within the context of BLLM, the role of the faculty member shifts from the position of sage on the stage to a guide who determines how best to convey knowledge by leveraging digitally enhanced learning tools and techniques.[27][28] Given the plethora of information available to students, the role of faculty is to teach students how to vet, evaluate, and distill available content to support curriculum goals and objectives, ensuring student success. The lack of knowledge in learning outcomes regarding the use of Twitter or similar social media platforms, combined with their widespread availability, and cost-effectiveness provided motivation to study Twitter as a student-centered intervention to enhance learning.

Methods

Design, setting & sample

This investigation used a quasi-experimental study design. Participants were recruited from a Family Medicine clerkship at a community-based hospital in the Northeast and were in their third year of training in an allopathic or osteopathic curriculum. Prior to the intervention, a survey that contained the questions in Table 1 was emailed to each participant. Participants responded to the survey via Qualtrics survey software (Provo, UT) for the first ten months of the intervention and then responded via Survey Monkey (San Mateo, CA). Respondents were sent tweets during two weeks of their Family Medicine clerkships. Upon completion of their clerkship, a survey was emailed to participants via Qualtrics survey software for the first ten months and then changed to Survey Monkey containing questions from the MSCES™. A historical comparison group of individuals completing their rotations immediately prior to the intervention was used to assess a difference in Shelf examination scores. A two-group Wilcoxon-Mann-Whitney test with a .05 one-sided significance level will have 80% power to detect an effect size of 1.0 when the total sample size of 32.[29][30][31][32] Cohen (1988) endorses large sample sizes such as 1.0 for educational interventions.[29]

Question development

Reflecting the theoretical framework of spaced learning, the goal was to have three questions reflecting each topic. High-yield topics from the Family Practice Shelf Exam were identified.[33] Initial questions were proposed by two Family Medicine practitioners and study authors (AA, CG). Two additional parallel questions were developed by three clinician authors (EBB, LAP, & JS). All questions were reformatted by (LAP & JS) to adhere to the 280- character limit of Twitter. Questions comprised three 280-character components per question stem, multiple choice answers (a minimum of four distractors), and an explanation of the correct answer. Potential answers were presented in alphabetical order. All multiple-choice questions were reviewed and finalized by AA and CG yielding 30 topics with two parallel questions each. This process resulted in a 90-item question bank.

Question testing

The candidate questions were entered into Qualtrics and distributed to third- and fourth-year medical students for pilot testing and item analysis at a separate institution. Students were encouraged to answer the questions, but not required. Twenty-two students started answering the questions and only six (27%) completed all 90 questions. An item analysis was done on the completed questions. Twenty-three of the 90 questions (26%) were deemed "easy" (i.e., a difficulty index of 85% or greater). Where possible "easy" questions were replaced with vetted, more difficult questions covering the same concept. Where not possible, questions were rewritten.

Intervention

During third-year Family Practice clerkships at a suburban hospital in the Northeast, formulated questions were sent to students. The intervention lasted two weeks out of the six-week rotation. These questions were sent via Twitter by three of the authors (AA, CG, and JG). This was repeated every day Monday through Friday for two weeks during the Family Practice clerkship. The spaced learning framework dictated our format of three spaced learning topics interrupted by 10-minute breaks, when the learner is doing something other than studying in-between the content. Content experts (AA, CG, and JG) were available to respond to learner comments and questions during these Twitter conversations. The intervention encouraged ongoing Twitter conversation to get learners to engage with the material. This was facilitated by the study authors to promote clinical thinking and decision-making by answering questions or asking additional questions. This two-way conversation was a key aspect of the study design to build upon the work of Reames et al.[17] The content expert faculty (AA, CG, and JG) viewed and responded to the Twitter feedback once daily.

Outcome Measures

For this investigation COMAT and NBME Shelf Exam scores were defined in two ways: raw scores and pass/fail. Passing scores for NBME were defined as a raw score of 59 or above out of a maximum of 95 (63%).[34] COMAT scores are reported as a raw score with a mean of 100 with a standard deviation of 10.[35][36] Raw COMAT scores were converted to percentile scores with passing defined at or above the 63rd percentile.

Covariates

To better be able to judge similarities between the comparison and intervention groups demographic variables and other pertinent variables were collected including curriculum; first portion of medical licensing exam score; number of rotations completed; previous Twitter use; how likely a participant is to comment on Twitter; and whether Family Medicine was a student's top residency selection. All information was self-reported.

Analysis

Descriptive statistics were calculated – means (standard deviations) and medians (interquartile range) for quantitative data and absolute frequencies and percentages (relative frequencies) for categorical data. Chi-squared will be used to test differences in demographic variables.

Covariates and dependent variables were tested for normality using normal probability plots and the Anderson-Darling (AD), Shapiro-Francia (SF), and Shapiro-Wilk (SW) normality tests.[37][38][39] The Anderson-Darling test is the recommended empirical distribution function test by Stephens compared to other tests of normality giving more weight to the tails of the distribution than the Cramer-von Mises test.[40] The Shapiro-Francia test was chosen because of its known performance and the Shapiro-Wilk test was chosen because it is one of the best-known tests for normality.[37]

The data for this study were comprised of independent data, therefore when judged to be normally distributed, Student's *t*-test was used for assessing an increase in means.[41] However, for data not normally distributed the Wilcoxon-Mann-Whitney test was used.[42][43] Since the only interest was in evaluating whether there was an increase in the two groups, a one-sided test was used. For normally distributed data Cohen's *d* was calculated as the effect size and interpreted according to Sawilowsky.[30] For data not normally distributed, the common language effect size (CL) was used.[44] The CL gives the percentage of observations in the post-intervention period greater than the pre-intervention period.[45] All analyses were performed using R v.4.2.1.

In April 2019, the American Statistical Association (ASA) formally strongly advocated abandoning the following terms, "statistical significance", "significantly different," "p<0.05," and "nonsignificant". Wasserstein and colleagues continue imploring researchers to remove these terms and similar terminology in favor of reporting exact *P* values and interpreting difference from a practical perspective.[46] In light of this formal stance by the ASA no interpretation of statistical significance will be made. However, this should not be interpreted as abandoning *P* values. The ASA only condones the abandonment of the dichotomization of the concept of "significance"; however, it wholeheartedly endorses the reporting of all *P* values whether less than or greater than an *a priori* alpha level.

Results

This investigation was approved by the Western Institutional Review Board and was conducted in accordance with the tenets espoused in the Declaration of Helsinki.[47]

Table 1. Characteristics of comparison group and Twitter intervention group at one community-based hospital in the Northeast (n=46)

		Treatment		
		Comparison (*n*=17)	Intervention (*n*=29)	*p*
Age [% (n)]	18-24	12 (2)	28 (8)	.282[a]
	25-34	89 (15)	72 (21)	
Gender [% (n)]	Male	59 (10)	52 (15)	.762[a]
	Female	41 (7)	48 (14)	
Race [% (n)]	Asian	41 (7)	38 (11)	1.000[a]
	African American or Black	6 (1)	10 (3)	
	White	41 (7)	38 (11)	
	Other	6 (1)	10 (3)	
	Decline to answer	6 (1)	3 (1)	
Ethnicity [% (n)]	Hispanic	6 (1)	7 (2)	.622[a]
	Middle Eastern/North African	6 (1)	19 (5)	
	Non-Hispanic	88 (15)	74 (20)	
Curriculum [% (n)]	Allopathic	65 (11)	41 (12)	.221[a]
	Osteopathic	35 (6)	59 (17)	
USMLE [M (SD)] – MD		217 (12.3)	221 (11.9)	.547[b, c, d]
COMPLEX [M (SD)] - DO		610 (65.8)	587 (95.3)	.618 [b, c, d]
Rotations [% (n)]	0	19 (3)	0 (0)	.104[a]
	1	25 (4)	17 (5)	
	2	6 (1)	21 (6)	
	3	25 (4)	10 (3)	
	4	13 (2)	17 (5)	
	5	6 (1)	24 (7)	
	6	0 (0)	3 (1)	
	7	6 (1)	0 (0)	
	8	0 (0)	3 (1)	
	9	0 (0)	3 (1)	
Twitter [% (n)]	Yes	59 (10)	38 (11)	.225[a]
	No	41 (7)	62 (18)	
Family Medicine Rank Top Choice [% (n)]	Yes	35 (6)	45 (13)	.555[a]
	No	65 (11)	55 (16)	
COMAT Score [M (SD)]		102 (.6)	110 (4.8)	<.001 [b, c, d]
NCME Shelf Exam [Mdn (IQR)]		72 (10)	72 (6)	.931[e, f]
COMAT/NCME Shelf Exam Percentile [Mdn (IQR)]		50.95 (39.20)	72.60 (46.75)	.157[e, f]

[a] *Fisher's exact test*

[b] *Judged to be normally distributed with the Shapiro-Wilk test and normal probability plot (COMPLEX: p=.479, USMLE: p=.547, COMAT: p=.145)*

[c] *Welch two-sample t-test*

[d] *Common Language Effect Size*[44] *states the probability a randomly selected score from the intervention group will exceed a randomly sampled score from the comparison population: COMAT: 90%, NCME Shelf Exam: 48%, COMAT/NCME Shelf Exam Percentile: 62%*

[e] *Judged not to be normally distributed with the Shapiro-Wilk test and normal probability plot (NCME Shelf Exam: .p=021, COMAT/NCME Shelf Exam percentile: p=.008)*

[f] *Wilcoxon-Mann-Whitney test*

Average age and demographic makeup between the historical comparison and the intervention group were similar (Table 1). Average NBME Shelf Exam scores for the historical comparison were 74 (SD=9.9) with median scores 72 (IQR=10). Average NBME Shelf Exam scores for the intervention group were 73 (SD=8.6) with median scores 72 (IQR=6). NBME Shelf Exam scores were not normally distributed (AD: *P*-value=.041; SF: *P*-value =.019; SW: *P*-value=.021). There was no statistical evidence of a difference between the intervention and comparison groups (*P*-value=.931).

Average COMAT scores for the historical comparison were 102 (SD=0.6) with median scores 102 (IQR=0.5). Average COMAT scores for the intervention group were 110 (SD=4.8) with median scores 110 (IQR=4.5). COMAT scores were judged to be normally distributed (AD: *P*-value=.395; SF: *P*-value =.100; SW: *P*-value=.145). There was statistical evidence of a difference between the intervention and comparison groups (*P*-value=<.001). The CLES was 90% which can be interpreted as 90% of randomly selected COMAT scores from the intervention group will be higher than scores from a randomly selected sample from the comparison group. This can be interpreted as a very meaningful difference in terms of COMAT scores.

Students in the comparison group scored an average of 48% (SD=28.1%) while students in the intervention group scored much higher on average (M=61%, SD=31.3%). The median percentile scores for the comparison group was 50.95% (IQR=39.20%) whereas the median percentile score for the intervention group was 72.60% (IQR=46.75%). As expected, percentile scores were judged to be not normally distributed (AD: .007; SF: .026; and SW: .009). There was no statistical evidence of a difference between percentile scores (*P*-value=.157). The CLES for COMAT/NCME Shelf Exam percentile was 62% - a meaningful effect size.

Discussion

Twitter has 330 million active monthly users with roughly half of users tweeting daily.[2] The widespread use of social media has led to exploration of its use as an educational tool to engage learners and enhance interactive learning with its use in the healthcare literature dating back to 2008.[3],[4] Prior research has studied perceptions of Twitter use and student engagement, but only two linked it to learning outcomes.

Jurivich et al. found increases in geriatric knowledge among medical students during their internal medicine clerkship due to a weekly Twitter multiple-choice question intervention.[16] This study adds support for the use of Twitter to improve learning outcomes during medical student clerkships.

Reames and colleagues used a pre-test/post-test prospective observational study using medical student volunteers from surgical clerkships.[17] They sent three new tweets per day with succinct, objective surgical facts and assessed aggregate test scores for participating students and historical controls.[17] They found NBME Shelf Examination scores were not significantly different in the intervention group. This study was most closely aligned with our target population and attempt to increase Shelf Examination scores for medical students. Our study built upon this finding by adding the bidirectional aspect of Twitter communication. As a result, we saw a change in unadjusted Shelf exam scores. However, we posited this change may have been due to student engagement.

The results provide further evidence of the application of BLLM where didactic materials are delivered through traditional teaching methods and are scaffolded via continual reinforcement through clinical activities, and social media. In this way, our intervention enhanced didactic learning through interaction with faculty via Twitter. The use of Twitter demonstrated a social media approach to providing students with anytime-anyplace and just-in-time experience. The anytime-anyplace and just-in-time spaced learning experience using Twitter provides an educational resource to leverage the already over-burdened medical student schedule with valuable opportunities to study. Twitter also provides undergraduate medical educators with an inexpensive resource to buttress key concepts and engage clerkship students. The use of BLLM and the application of social media in medical education underscore the need for digital immigrants and digital Neanderthals to adapt learning methodology to today's medical students (digital immigrants) avoiding digital discord.[48]

Limitations

Selection bias may be a possible explanation for the positive results seen in this study. In other words, more motivated students may have volunteered for the intervention as opposed to students who were less motivated. A positive change in Shelf Exam scores may be due to this bias.

Another limitation may be the confounder of student engagement in those students who are more engaged and thus more motivated may have better learning outcomes. However, we accounted for student engagement by assessing this factor and adjusting for it. Thus, we don't think engagement or motivation had an appreciable effect on the results. Due to the small sample size regression results must be interpreted with extreme caution.

This study was designed to be more conversational within the Twitter platform around the questions and answers, but as the study was implemented it did not evolve in this way. Rather, the Twitter posts were initially tweeted out by the study physicians and the learners responded with their answer choices but did not engage in discussion on Twitter around question content. We hypothesized that had there been more interaction around the content, learners may have taken more advantage of the intervention and further increased scores. Anecdotally, further in-person discussion occurred with clinical preceptors around the Tweeted concepts. These anecdotal conversations may have an unknown effect.

In March 2020 when the COVID-19 pandemic reached the Northeast, medical student in-person clerkships were paused and therefore our study enrollment was also paused. When medical students returned to in-person clerkships, the study enrollment began again. Two study participants started the intervention, but in-person clerkships were interrupted within the first week of their participation, therefore they were excluded from the study. We do not expect these two students would have influenced the results positively or negatively.

Anecdotally, it was observed that most students did not have a Twitter account prior to this study and had to create one for this intervention. If a platform more frequently used by students had been utilized for this study, enrollment may have increased.

One possibility for the negative findings is students were given dedicated time during their rotation day to read and respond to the Tweeted questions.

Conclusions

The results of this investigation further the potential for Twitter to be used as an inexpensive educational intervention to modestly improve standardized exam scores with the potential to be used in a broader fashion (e.g., COMLEX and USMLE Step Scores). Future research can build upon this study by applying this intervention to other locations with larger samples. Additionally, this study design could be replicated for each core clerkship rotation for COMLEX and USMLE Step preparation.

Acknowledgements

The authors are indebted to James E. Gilbert, PhD for his suggestions and valuable insights in drafting the spaced learning portion of the theoretical framework.

Contributions

GEG conceived the project idea, completed the original literature search, wrote the informed consent document, created the pre and post intervention surveys, and wrote the initial draft of the manuscript and the Methods and Results sections. LAP conceived the project idea, completed the original literature search, wrote, and edited all intervention questions, created the pre and post intervention surveys, emailed students pre and post interventions surveys, documented student participation and survey results, and wrote the initial draft of the manuscript. AA wrote all intervention questions, enrolled, and deployed the intervention with medical students, and collected NBME exam scores. JS wrote and edited all intervention questions, did the initial journal/literature review prior to the start of the project, assisted in the planning of this project, distributed intervention questions for pilot testing. CG assisted in writing intervention questions, enrolled students, and deployed the intervention with medical students. EBB conceived the idea, edited all intervention questions, edited the manuscript, and created the Twitter account. JG enrolled students and deployed the intervention with medical students. All authors suggested critical revisions and approved the final manuscript.

References

1. ^*Micieli R, Micieli JA. Twitter as a tool for ophthalmologists. Can J Ophthalmol. 2012;47(5):410-413. doi:10.1016/j.jcjo.2012.05.005*
2. a, b*Lin Y. 10 Twitter Statistics Every Marketer Should Know in 2020. Oberlo. Published 2019. Accessed September 7, 2020. https://www.oberlo.com/blog/twitter-statistics#:~:text=Here's a summary of the,are between 35 and 65.*
3. a, b*Desselle SP. The use of Twitter to facilitate engagement and reflection in a constructionist learning environment. Curr Pharm Teach Learn. 2017;9(2):185-194. doi:10.1016/j.cptl.2016.11.016*
4. a, b*Skiba DJ. Nursing education 2.0: Twitter & tweets. Can you post a nugget of knowledge in 140 characters or less? Nurs Educ Perspect. 2008;29(2):110-112.*
5. ^*Booth R, O'Connor S. Meaningful use of Twitter in nursing education may improve student learning and should be considered as a viable educational tool to assist in the development of digital professionalism. Evid Based Nurs. 2017;20(1):28. doi:10.1136/eb-2016-102465*
6. ^*Carley S, Beardsell I, May N, et al. Social-media-enabled learning in emergency medicine: A case study of the growth, engagement and impact of a free open access medical education blog. Postgrad Med J. 2018;94(1108):92-96. doi:10.1136/postgradmedj-2017-135104*
7. ^*Kakushi LE, Évora YDM. Social networking in nursing education: Integrative literature review. Rev Lat Am Enfermagem. 2016;24:e2709. doi:10.1590/1518-8345.1055.2709*
8. ^*Logghe HJ, Selby L V., Boeck MA, Stamp NL, Chuen J, Jones C. The academic tweet: Twitter as a tool to advance academic surgery. J Surg Res. Published online 2018. doi:10.1016/j.jss.2018.03.049*
9. ^*Ross JG, Myers SM. The current use of social media in undergraduate nursing education: A review of the literature. Comput Informatics Nurs. 2017;35(7):334-344. doi:10.1097/CIN.0000000000000342*
10. ^*Sterling M, Leung P, Wright D, Bishop TF. The use of social media in graduate medical education: A systematic review. Acad Med. 2017;92(7):1043-1056. doi:10.1097/ACM.0000000000001617*
11. ^*Zucker BE, Kontovounisios C. It is time to improve the quality of medical information distributed to students across social media. Adv Med Educ Pract. 2018;Volume 9:203-205. doi:10.2147/AMEP.S155398*
12. ^*Dong T, Swygert KA, Durning SJ, et al. Is poor performance on NBME Clinical Subject Examinations associated with a failing score on the USMLE Step 3 Examination? Acad Med. 2014;89(5):762-766. doi:10.1097/ACM.0000000000000222*
13. ^*Casey PMM, Palmer BAA, Thompson GBB, et al. Predictors of medical school clerkship performance: A multispecialty longitudinal analysis of standardized examination scores and clinical assessments. BMC Med Educ. 2016;16(1):128. doi:10.1186/s12909-016-0652-y*

14. ^*Webb A, Dugan A, Burchett W, et al. Effect of a novel engagement strategy using Twitter on test performance. West J Emerg Med. 2015;16(6):961-964. doi:10.5811/westjem.2015.10.28869*
15. ^*Hennessy CM, Kirkpatrick E, Smith CF, Border S. Social media and anatomy education: Using twitter to enhance the student learning experience in anatomy. Anat Sci Educ. 2016;9(6):505-515. doi:10.1002/ase.1610*
16. a, b, c*Jurivich DA, Bande D, Theige D, et al. Integrating Geriatrics Knowledge into a Medical Student Clerkship Using Twitter Poll. J Am Geriatr Soc. 2018;66(12):2389-2393. doi:10.1111/jgs.15633*
17. a, b, c, d, e*Reames BN, Sheetz KH, Englesbe MJ, Waits SA. Evaluating the Use of Twitter to Enhance the Educational Experience of a Medical School Surgery Clerkship. J Surg Educ. 2016;73(1):73-78. doi:10.1016/j.jsurg.2015.08.005*
18. ^*Halpern DF, Hakel MD. Applying the science of learning to the university and beyond. Change. 2003;35(4):36-41.*
19. a, b*Bauman EB. Games, virtual environments, mobile applications and a futurist's crystal ball. Clin Simul Nurs. 2016;12(4):109-114. doi:10.1016/j.ecns.2016.02.002*
20. a, b*Bauman EB, Adams RA, Pederson D, et al. Building a better donkey: A game-based layered learning approach to veterinary medical education. In: GLS 10 Conference Proceedings. Carnegie Mellon University ETC Press; 2014:372-375.*
21. a, b*Bauman EB, Gilbert GE, Vaughan G. Short-term gains in histology knowledge: A veterinary gaming application. PeerJ Prepr. 2017;5:e3421v1. doi:https://doi.org/10.7287/peerj.preprints.3421v1*
22. ^*Fields RD. Making memories stick. Sci Am. 2005;(February):75-81.*
23. ^*Kelley P, Whatson T. Making long-term memories in minutes : A spaced learning pattern from memory research in education. Front Hum Neurosci. 2013;7(September):Article 589. doi:10.3389/fnhum.2013.00589*
24. ^*Kang SHK. Spaced repetition promotes efficient and effective learning: Policy implications for instruction. Policy Insights from Behav Brain Sci. 2016;3(1):12-19. doi:10.1177/2372732215624708*
25. ^*Wahlheim CN, Maddox GB, Jacoby LL. The role of reminding in the effects of spaced repetitions on cued recall: Sufficient but not necessary. J Exp Psychol Learn Mem Cogn. 2014;40(1):94-105. doi:10.1037/a0034055*
26. ^*Toppino TC, Gerbier E. About Practice. Vol 60. (Federmeier KD, ed.). Elsevier Inc.; 2013. doi:10.1016/b978-0-12-800090-8.00004-4*
27. ^*Prensky MR. Teaching Digital Natives: Partnering for Real Learning. Corwin Press, Inc.; 2010.*
28. ^*Prensky MR. Brain Gain: Technology and the Quest for Digital Wisdom. MacMillian; 2012.*
29. a, b*Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Lawrence Erlbaum Associates Publishers; 1988.*
30. a, b*Sawilowsky SS. New effect size rules of thumb. J Mod Appl Stat Methods. 2009;8(2):597-599. doi:10.22237/jmasm/1257035100*
31. ^*Dixon WJ, Massey FJ. Introduction to Statistical Analysis. 4th ed. McGraw-Hill; 1983.*
32. ^*O'Brien RG, Muller KE. Applied Analysis of Variance in Behavioral Science. Marcel Dekker; 1983.*
33. ^*National Board of Medical Examiners®. Subject Examinations: Content Outlines and Sample Items. Published 2017. Accessed January 14, 2020. http://www.nbme.org/Schools/Subject-Exams/*
34. ^*National Board of Medical Examiners. Subject Examination Program Advanced Clinical Science Examination Score Interpretation Guide NATIONAL BOARD OF MEDICAL EXAMINERS ® Subject Examination Program Advanced*

Clinical Science Examination Score Interpretation Guide.; 2013.

35. ^*National Board of Osteopathic Medical Examiners Inc. Percentile scores. COMAT Clinical Subjects. Published 2020. Accessed September 18, 2020. https://www.nbome.org/exams-assessments/comat/clinical-subjects/scores-transcripts/comat-percentile-score-conversion-table/*
36. ^*Li F, Kalinowski K, Song H, Bates B. Relationships Between the Comprehensive Osteopathic Medical Achievement Test (COMAT) Subject Examinations and the COMLEX-USA Level 2-Cognitive Evaluation. J Am Osteopath Assoc. 2014;114(9):714-721. doi:10.7556/jaoa.2014.140*
37. a, b*Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3-4):591-611. doi:10.1093/biomet/52.3-4.591*
38. ^*Shapiro SS, Francia RS. An approximate analysis of variance test for normality. J Am Stat Assoc. 1972;67(337):215-216. doi:10.1080/01621459.1972.10481232*
39. ^*Anderson TW, Darling DA. A test of goodness-of-fit. J Am Stat Assoc. 1954;49(268):765-769. doi:10.1080/01621459.1954.10501232*
40. ^*Stephens M. Tests based on EDF statistics. In: D'Agostino, Ralph B and Stephens M, ed. Goodness-of-Fit Techniques. Marcel Dekker; 1986.*
41. ^*Student. The probable error of the mean. Biometrika. 1908;6(1):1-25.*
42. ^*Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bull. 1945;1(6):80-83.*
43. ^*Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 1947;18(1):50-60. doi:10.1214/aoms/1177730491*
44. a, b*McGraw KO, Wong SP. A common language effect size statistic. Psychol Bull. 1992;111(2):361-365. doi:10.1037/0033-2909.111.2.361*
45. ^*Grissom RJ, Kim JJ. Effect Sizes for Ordinal Categorical Variables. 2nd ed. Routledge; 2012.*
46. ^*Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p<0.05." Am Stat. 2019;73(sup1):1-19. doi:10.1080/00031305.2019.1583913*
47. ^*World Medical Association. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. J Am Med Assoc. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053*
48. ^*Gilbert GE, Bauman EB. New Terms for the Educators' Digital Lexicon. Qeios. doi:10.32388/QHZ6ZW*