REVIEW TITLE: Exploring students' use and the potential impact of commercial-off-the-shelf products on students' NBME and NBOME exam performance: A focused BEME systematic review

REVIEW GROUP:

Lead reviewer: Atsusi Hirumi, PhD
Professor, Instructional Design and Technology
College of Medicine, University of Central Florida (UCF)
Atsusi.Hirumi@ucf.edu, 407-417-1154

Feroza Daroowalla, MD, MPH
Associate Professor of Medicine
College of Medicine, University of Central Florida (UCF)
Feroza.Daroowalla@ucf.edu, 407-266-1100

Nyla Dil, PhD, DVM, MSc
Assistant Professor of Immunology and Medical Microbiology
College of Medicine, University of Central Florida (UCF)
Nyla.Dil@ucf.edu, 207-266-1133

David Harris, PhD
Associate Professor of Physiology
College of Medicine, University of Central Florida (UCF)
David.Harris2@ucf.edu, 207-266-1111

Andrea Berry, M.P.A.
Director of Faculty Development
College of Medicine, University of Central Florida (UCF)
Andrea.Berry@ucf.edu, 407-266-1115

Luke Horger
Instructional Designer
College of Medicine, University of Central Florida (UCF)
Luke.Horger@ucf.edu, 772-361-3166

Shalu Gillum, JD, MLS, AHIP
Medical Librarian, Head of Public Services
College of Medicine, University of Central Florida (UCF)
Shalu@ucf.edu, 407-266-1424

Juan Cendán, MD
Chairman of the Department of Medical Education, Professor of Surgery
College of Medicine, University of Central Florida (UCF)
Juan.Cendan@ucf.edu, 407-266-1141

Abstract

Commercial-off-the-shelf software developed for medical education (MedEd-COTS) has emerged as a resource used by a vast majority of medical students to prepare for national board exams. The pervasive use of MedEd-COTS outside of the formal medical school curriculum, however, has raised concerns. As more companies offer MedEd-COTS with increasing functions and features, there is a need for an up-to-date review to inform medical educators on (a) students' use of MedEd-COTS outside the formal medical school curriculum, (b) the integration of MedEd-COTS into the formal curriculum, and (c) the potential effects of MedEd-COTS usage on students' NBME or NBOME exam scores. Due to the emerging, yet limited number of studies published on either the use or integration of MedEd-COTS, a focused review of literature is proposed to guide future research and practice. Data extraction and quality appraisal will be conducted by three reviewers independently with disagreement resolved by reaching consensus with a fourth reviewer. We will complete a narrative synthesis to identify trends and issues in the findings reported by the studies included in the review.

Background to the Topic

For medical students to become Doctor of Medicine (MD) or Doctor of Osteopathic Medicine (DO) physicians, they must be selected to attend a residency program through a national matching program. The National Resident Matching Program[1] has cited performance on the high-stakes National Board of Medical Examiners (NBME) Step 1/ National Board of Osteopathic Medical Examiners (NBOME) COMLEX-USA Level 1 and NBME Step 2/ NBOME COMLEX-USA Level 2 exams as the first and fourth most considered factors when selecting medical students. The National Resident Matching Program[1] has also cited clerkship grades as the sixth most important factor for medical student selection. One component of clerkship grades is the NBME shelf exams or NBOME Comprehensive Osteopathic Medical Achievement Test (COMAT). To prepare for these NBME and NBOME high-stakes exams, it could be expected that students would heavily rely on curricular material provided by their institution. Dixon and colleagues[2] has found that preclinical course performance is positively correlated with NBOME COMLEX-USA Level 1 and 2 scores. Preclinical performance has also been positively correlated with NBME Step 1 performance[3-5]. However, within the last decade, research has shown that students are more often using commercial-off-the-shelf software developed for medical education (MedEd-COTS) developed by third-party companies instead of lecture material to prepare for national standardized medical exams[6].

One previous review synthesized the results of research on medical students' use of MedEd-COTS to prepare for NBME exam performance. In short, McGaghie, Downing, and Kubilius[7] found that (a) commercial coaching was well-received by medical students but did not significantly improve Step exam performance, (b) research lacked rigor and control, and (c) almost no details were given about the form or conduct of the courses. It is important to note that the 2004 review focused only on the use of test preparation books and commercial coaching courses. Furthermore, the 2004 review did not distinguish medical students' use of MedEd-COTS outside of medical curriculum versus as a prescribed, integrated part of the medical school curriculum. Our review will build on the previous review by: (a) examining the impact of new features of MedEd-COTS, such as question banks, flashcards, and multimedia that have evolved over the past decade; and (b) distinguishing studies that examined students' use of MedEd-COTS on their own time versus the integration of MedEd-COTS into the formal medical school curriculum.

MedEd-COTS offer education to medical students through a combination of integrated features, such as question banks, videos, flash cards, text, note-taking, data analytics, and commercial coaching. Several of these features expose students to an instructional stimulus through flashcards or sample test questions and then provide feedback on student's responses through data analytics. The MedEd-COTS features of question banks and flashcards utilize the instructional strategies of spaced repetition by prompting students to review information over spaced periods of time. The spaced instructional stimulus-response sequence is a strategy posited by behaviorist learning theory[8] that helps explain the positive relationship between students' use of MedEd-COTS and their test performance. Furthermore, the positive impact of MedEd-COTS question banks may be explained by the testing effect because they stimulate active retrieval under medical exam conditions[9-10]. Both MedEd-COTS question bank and flash card use has been positively correlated with increased NBME Step 1 performance and testing effect[11]. Volk and colleagues[12] also identified that medical students' combined usage of a MedEd-COTS review book and online question bank usage was correlated with an increase in the NBME surgery shelf exam score. An interesting relationship has also been identified with NBOME exam performance; students' preference for the UWorld question bank over the COMBANK question bank is associated with higher scores on the COMLEX-USA Level 1 exam[13].

Interactions with faculty across Colleges of Medicine suggest that students and faculty have conflicting views on the value of MedEd-COTS. Students are concerned that faculty may not be adequately preparing them for Step and are purchasing subscriptions to MedEd-COTS to optimize their chances of passing the exam on their own. They also express concern for instructional content and activities that are not apparently related directly to Step. Faculty, in turn, are concerned that MedEd-COTS may contain errors or omissions, and that students are viewing these resources as a primary source of knowledge that offers more high yield learning experiences than the coursework and materials that are generated and prescribed by faculty. Findings from the review may yield insights on how to address such concerns. Answers to the proposed research questions may help inform faculty and students on the advantages and disadvantages of using certain MedED-COTS features to prepare for NBME or NBOME examinations.

Review Questions, Type of Review, and Key words

Our focused review will address research and literature on (a) medical student use of MedEd-COTS outside the formal medical school curriculum, and (b) the integration of MedEd-COTS as an explicit component of the medical school curriculum. Preliminary scoping of related articles indicate the following questions may yield practical information useful to both educators and researchers interested in these two emerging areas of study. The primary research question asks:

- Is there a relationship between students' use of MedEd-COTS and their performance on undergraduate NBME or NBOME exams?

The scoping search did not yield any experimental studies to determine if there is a causal relationship between students' use of MedEd-COTS and their test performance. However, there are three quasi-experimental studies and surveys that suggest there may be a relationship between students use and test performance.

Secondary research questions are organized based on the nature and purpose for reviewing identified studies. Articles examining students' use of MedEd-COTS outside the formal curriculum and articles studying the integration of MedEd-COTS into the formal curriculum will be examined by exploring answers to the following secondary questions:

- Which companies and features are most valued/used by undergraduate medical students who use MedEd-COTS outside of the formal medical school curriculum to prepare for NBME and NBOME exams?
- When do undergraduate M.D. and D.O. medical students use MedEd-COTS?
- Why do students use MedEd-COTS (other than to prepare for standardized tests)?
- What factors influence students' use of MedEd-COTS?
- What research questions on MedEd-COTS usage and outcomes have been posited by post-2004 studies?
- What methods were used to answer research questions proposed by post-2004 studies (including subjects, designs, instrumentation, procedure, limitations)?
- What gaps or opportunities remain for advancing our knowledge of the use of MedEd-COTS outside of the formal curriculum or as integrated component of formal curriculum to prepare for NBME and NBOME exams?

Articles that examine the integration of MedEd-COTS into the formal curriculum will be examined with two additional secondary questions:

- Why are faculty integrating the use of MedEd-COTS into the formal medical school curriculum (e.g., improve NBME or NBOME test performance)?
- Which MedEd-COTS companies and MedEd-COTS features are most valued/used by faculty to enhance medical education?
- What instructional approaches, theories, and strategies are being used to ground and guide the integration of MedED-COTS, if any?)

Keywords:

- MedEd-COTS: Commercial-off-the-shelf products developed for undergraduate medical education standardized test preparation, such as but not limited to Anki, UWorld, AMBOSS, Lecturio, Sketchy Medical/Sketchy Micro, Pathoma, Firecracker, Goljan, Osmosis, Kaplan QBank, and First-Aid. To be considered a MedEd-COTS, the application must be commercially developed and/or offered by commercial companies. A question bank that is solely student-authored and not offered as a commercial product would not be classified as a MedEd-COTS.
- Integration: A medical school provides medical students access to certain MedEd-COTS features and recommends that students utilize these features.
- Test Preparation: Undergraduate medical students have to prepare for and pass several licensing exams. During the undergraduate medical curriculum, M.D. students must prepare for and take NBME Step 1, Step 2, and shelf exams while D.O. students prepare for and take NBOME COMLEX-USA Level 1, Level 2, and COMAT exams. The NBME Step 3 and NBOME COMLEX-USA Level 3 exams have not been included in this definition because they are generally taken after the medical curriculum during residency.
- NBME Step 1 and Step 2 exams: Licensing exams that M.D. undergraduate students must pass to qualify for medical licensure.

- NBME Shelf exam: Licensing exams that M.D. undergraduate students take during their clerkship portion of the medical curriculum. NBME Shelf exams comprise a portion of medical students' clerkship grades at many medical schools.
- NBOME COMLEX-USA Level 1 and 2 exams. Licensing exams that D.O. undergraduate students must pass to qualify for medical licensure.
- NBOME COMAT exam: Licensing exams that D.O. undergraduate students take during their clerkship portion of the medical curriculum. NBOME COMAT exams comprise a portion of medical students' clerkship grades at many D.O. medical schools.
- Question bank: A feature of MedEd-COTS that provides students with practice questions and feedback similar to the questions they will complete in Step exams
- Flash cards: A feature of MedEd-COTS that provides students with opportunities to practice active retrieval and review of medical information
- Video lectures: A feature of MedEd-COTS that provides students with recorded lectures or animations that explain medical concepts.
- Audio lectures: A feature of MedEd-COTS that provides students with recorded lectures without a visual component
- Data analytics: A feature of MedEd-COTS that provides students data on their usage of the products features (e.g. how many videos watched, how many practice questions answered, percentage of correct answers to questions, etc.)
- Commercial coaching: A face-to-face feature of MedEd-COTS that involves students learning from a live lecture format in person or through the internet

Study Selection Criteria

Inclusion criteria:
- Quantitative studies that involve numerical data in survey, correlational, quasi-experimental, or experimental research
- Qualitative studies that involve data that are words or images gathered for study designs such as case study, ethnography, survey, narrative reviews, or extant data analysis
- Studies examining the target population of first year, second year, third year, or fourth year M.D. or D.O. medical students or librarians at M.D. or D.O. schools
- Studies assessing students' use of MedEd-COTS outside or within their medical school curriculum to prepare for the NBME Step 1, Step 2, or Shelf exams or the NBOME COMLEX-USA Level 1, Level 2, or COMAT exams
- Research published in English
- To build upon research provided by the 2004 review on MedEd-COTS, our literature review date range will be confined to 2004 to present
- Research on MedEd-COTS usage in the United States or Canada (M.D. schools are in both countries while D.O. schools are only in the United States)

Exclusion criteria:
- Studies that examine medical residents or undergraduate students before matriculation into medical school
- Studies that do not have an abstract
- Studies with no empirical data including commentaries and opinion pieces
- Studies not published in English

Search Sources and Strategies

Databases:
- Embase
- ERIC
- MEDLINE (PubMed)
- Ovid
- PsycInfo
- Web of Science
- Open Grey
- ProQuest Dissertations & Theses Global

Conference proceedings (reviewing available online abstracts from last five years):
- General Practice Education and Training Ltd
- Association for the Study of Medical Education (ASME)
- International Association of Medical Science Educators (IAMSE)
- Association of American Medical Colleges (AAMC)

Search engine:
- Google Scholar

Search strategy:
- (((students, medical[mesh] OR medical student* OR student OR education, medical, undergraduate [mesh])) AND ("board review" OR board OR "step 1" OR "step one" OR "step 2" OR "step two" OR "comlex" OR "comat" OR usmle OR "licensing exam" OR test OR exam OR shelf OR nbme OR "subject exam" OR licensure, medical[mesh] OR educational measurement[mesh] OR "United States Medical Licensing Exam*" OR "comprehensive osteopathic medical achievement test*")) AND (((commerc*) OR ("question bank") OR ("question banks") OR (flashcard) OR ("flash card") OR ("test prep*") AND (medical)))

Notes regarding search strategies:
- The Medical Subject Heading (MeSH) keyword Students, Medical was omitted from the ERIC, PsycInfo, and Google Scholar searches.
- The reference lists of included articles will be examined for additional relevant sources.
- Our search terms originally included MedEd-COTS company names (UWorld, Kaplan, First Aid, Goljan, Anki, Firecracker, Sketchy, Pathoma, Boards and Beyond, Fundamentals of Pathology, and USMLE-Rx), but that did not generate additional search results. Thus, the names were not included in the search strategies.
- Our search terms also originally included the MedEd-COTS features "video*" and "data analytics" in previous search strategies, but we found those search terms did not generate unique search results.

Data Extraction

Two reviewers will independently review the title and abstract of all generated articles in the search. If a paper's title and abstract meets all inclusion criteria, the full body of the article will be requested to review. If disagreement occurs at this stage between reviewers, a third reviewer will offer his/her opinion to reach consensus. It is important to note that any stage of disagreement in data extraction or quality appraisal, we will use the formula created by Miles and Huberman[14] to ensure interrater reliability:

- reliability = number of agreements / (number of agreements + disagreements)

Where there is disagreement on any data extraction or quality appraisal question between coders, an additional reviewer will mediate the disagreement until consensus is reached. According to research by Miles and Huberman[14], satisfactory IRR will be established if there is at least 80% agreement on 95% of the codes.

Data extraction will be completed using a form created in Qualtrics so that reviewers may quickly share and review extracted data. The data form will be organized with a list of headers for all articles, and additional headers depending on whether the paper is classified as examining MedEd-COTS usage or integration into medical curriculum.

The following headings from the standard BEME coding sheet will be utilized for all papers[15].

- Title
- Authors
- Location of study
- Study aims/objectives

If a paper is classified as a usage article (medical students' use of MedEd-COTS outside of their prescribed curriculum), headings will be included to address the following questions when applicable.

- Which MedEd-COTS companies and MedEd-COTS features are most valued/used by undergraduate medical students to prepare for NBME/NBOME exams?
- When do undergraduate M.D. and D.O. medical students use MedEd-COTS?
- Why do students use MedEd-COTS (other than to prepare for standardized tests)?
- What factors influence students' use of MedEd-COTS?
- What research questions on MedEd-COTS usage and outcomes have been posited by post-2004 studies?
- What methods were used to answer research questions proposed by post-2004 studies (including subjects, designs, instrumentation, procedure, data analysis, limitations)?
- What gaps or opportunities remain for advancing our knowledge of the use of MedEd-COTS outside of the curriculum to prepare for NBME/NBOME exams?

If a paper is classified as an integration article (a medical school provides medical students access to certain MedEd-COTS features and recommends that students utilize these features), reviewers will answer all aforementioned headings for usage article research as well as headings included to address the following questions below if applicable.

- Which MedEd-COTS companies and MedEd-COTS features are most valued/used to enhance medical education, and for what specific purpose (e.g., improve NBME or NBOME test performance)?
- What instructional approaches, theories, and strategies are being used to ground and guide the integration of MedED-COTS, if any?)
- What gaps or opportunities remain for advancing our knowledge of the integration of MedEd-COTS to facilitate undergraduate medical education?

Three reviewers will independently complete data extraction. Any disagreement will be resolved through discussion with a fourth reviewer until a full consensus is reached.

Scoping Search:

A scoping search was conducted on September 2nd, 2019, which utilized three databases (i.e., ERIC, MEDLINE (PubMed), and PsycInfo). The scoping search yielded 6,801 papers, and 21 were considered relevant for the study. Of the 21 papers, 1 paper was the previous 2004 systematic review by McGaghie et al. and 14 studies identified in the scoping review addressed student outcomes on standardized tests in various ways:

- Three (3) quasi experimental studies calculated and tested for the significance of correlations between students' use and NBME test performance, and
- Ten (10) surveys calculated correlations between survey items and NBME test performance and one (1) survey calculated correlations between survey items and NBOME test performance.

In addition, the following information pertaining to the secondary questions was addressed in the 21 papers found from the scoping review (excluding the 2004 systematic review):

- Fourteen (14) studies identified MedEd-COTS companies that students used,
- Sixteen (16) studies identified the MedEd-COTS features that students used,
- Ten (10) studies examined when students used MedEd-COTS,
- Two (2) studies examined student learning strategies associated with MedEd-COTS,
- Three (3) studies asked students to rate the usefulness of specific MedEd-COTS companies and features,
- One (1) study identified factors that influenced students' decision to use MedEd-COTS, and
- Two (2) quasi-experimental studies described the instructional strategy that was used to integrate the MedEd-COTS into the formal curriculum.

The research design of the 21 studies identified in the scoping review are as follows:

- One (1) systematic review,
- Sixteen (16) surveys and of those surveys, eleven (11) surveys conducted correlational analysis,
- Three (3) quasi experimental studies, and
- One (1) brief narrative review of qualitative data (from medical student blogs and forums, and commercial websites) regarding study resources.

Quality Assessment

In addition to the proposed data extraction, each paper's risk of bias will be assessed in the Qualtrics form. Headers in the form have been extracted from the standard BEME coding sheet (BEME Collaboration, 2012).

- Well defined problem and related purpose and/or research question(s) (Strongly disagree to strongly agree scale)
- Appropriateness of study design (Strongly disagree to strongly agree scale)
- Implementation of study design (Strongly disagree to strongly agree scale)
- Appropriateness of data analysis (Strongly disagree to strongly agree scale)
- Appropriateness and strength of findings (1 to 5 low to high scale)
- Additional comments regarding the overall strengths and weaknesses of the papers

Three independent reviewers will independently make these judgements for quality of the papers. If there is disagreement, discussion will be initiated with a fourth reviewer. A research paper will be judged as having "low quality" if its average score for quality appraisal is below 3. If a low quality article is used to support a conclusion in the narrative synthesis, the low quality of the article will be noted in the narrative synthesis. No paper will be excluded due to results of the quality assessment.

If the reviewers agree that insufficient information was provided to assess the quality of the paper, the paper's author(s) will be contacted to request additional information. We will also use appraisal data to qualify our own findings by noting when findings from low quality articles were reported or otherwise used to form conclusions on MedEd-COTS usage or its potential impact on NBME or NBOME exams.

Synthesis of Extracted Evidence

We will complete a narrative synthesis to examine the data. The scoping review identified a variety of quantitative and qualitative data of MedEd-COTS usage. However, due to limitations with the number and nature of the studies, we cannot analyze these data using either meta-analysis or meta-ethnology.

The quantitative studies examining the outcome of students' use of MedEd-COTS on NBME or NBOME exam performance did conduct inferential tests to determine the significance of correlations calculated between students' use of MedEd-COTS and their exam performance. However, we can neither pool the studies nor the data due to the wide range of interventions (e.g., variability of MedEd-COTS companies and features), and differences in research design across studies.

The qualitative studies identified from our refined scoping search reported survey and interview data on (a) why students chose to use specified MedEd-COTS and MedEd-COTS features, (b) when they used them, and (c) how these variables affected various outcomes (Step success, exam performance, etc.). After further examining the nature of these studies, we determined that none utilized a qualitative design as their primary research method. Thus, we decided to eliminate the meta-ethnography from our protocol.

We propose to conduct a narrative synthesis to identify patterns, and to discuss trends and issues in the findings reported by reviewed studies to answer our primary and secondary research questions. In other words, a narrative analysis will be conducted to identify themes in the following areas:

- The MedEd-COTS companies that students use/value or that faculty integrate/value
- The MedEd-COTS features that students use/value or that faculty integrate/value
- When students use MedEd-COTS
- Standardized test (NBME or NBOME) outcomes for students who use MedEd-COTS outside or within the formal curriculum
- Learning theory used to predict or justify the integration of MedEd-COTS into the curriculum
- Instructional strategy used to integrate the use of MedEd-COTS into formal curriculum
- Research questions and methods used by studies to assess students' use of MedEd-COTS since 2004 due to advancement in MedEd-COTS technology and medical school integration of MedEd-COTS
- Gaps or opportunities that remain for research that advances the knowledge of undergraduate usage of MedEd-COTS and its impact on NBME and NBOME exam performance

To structure our review, we will address four elements associated with the narrative synthesis process[17,18] by:

1. Developing a theory of how the intervention works to help determine (a) why MedEd-COTS usage may or may not improve students' performance on NBME and NBOME exam, and (b) how widely applicable the findings may be;
2. Developing a preliminary synthesis of findings to group studies and tabulate results in order to identify patterns in reported findings and methods;
3. Exploring relationships in the data using visualization tools and subgroup analysis to help (a) organize and present results, (b) identify reasons for differences in the direction and size of effects within and between studies, and (c) explain differences outcomes and methods within and between studies; and
4. Assessing the robustness of the synthesis to determine the overall strength of evidence on which conclusions are based. If a low quality article is used to support a statement or conclusion in the narrative synthesis, the article's quality rating will be noted.

Transfer to Practice

This review will seek to inform medical school faculty about: (a) which MedEd-COTS are being used, how are they being used and why outside of the formal medical school curriculum; (b) what particular features of MedEd-COTS (if any) may impact students' preference and test performance; (c) how MedEd-COTS are being integrated into the curriculum and with what result; and (d) what methods are being employed to study the use and integration of MedEd-COTS.

Knowledge of students' use of MedEd-COTS, and what particular features may impact students' preference and exam scores may help medical school educators (a) formulate effective strategies for integrating the use of MedEd-COTS into the curriculum, and (b) guide and advise students on the use of MedEd-COTS for Step exam preparation both in and out of class.

Knowledge of how MedEd-COTS are being integrated along with related results will also help educators formulate strategies for formally incorporating the use of MedEd-COTS into the curriculum, including what to do, and what not to do before, during and after students use to enhance student engagement and learning.

Project Timetable

The review is proposed to be completed within 10 months. Due to the focused scope of this systematic review and the investment of human resources, the review objectives are achievable within the timeframe depicted in Table 1.

Table 1. Timeframe for completing major tasks associate with the focused systematic review

	7/19	8/19	9/19	10/19	11/19	12/19	1/20	2/20	3/20	4/20
Protocol	X	X	X	X	X					
Data searching and selection					X	X				
Data extraction Quality appraisal							X	X		
Data Synthesis								X	X	
Write up								X	X	X

Conflict of interest statement

There are no conflicts of interest, including financial interest, to disclose.

Plans for updating the review

1.5 years after publication, we will repeat the literature search to identify and assess new research that has been published on the use and integration of MedEd-COTS. In 2 years, we will add a supplement to the original review highlighting new findings in this emerging field. At this point, we will also determine if there are sufficient quantitative studies with homogeneous variables to warrant a meta-analysis that may provide more precise estimates of effect.

The supplement (and meta-analysis if completed) will include references to research that was not published at the time of the original review.

Changes to the Protocol

Minor changes to the protocol will be recorded with accompanying rationale for the modifications. BEME will be consulted for approval before any significant changes are made to the protocol.

References

1   National Resident Matching Program, Data Release and Research Committee: Results of the 2018 NRMP Program Director Survey. National Resident Matching Program, Washington, DC. 2018.
2   Dixon D. Relation between variables of preadmission, medical school performance, and COMLEX-USA levels 1 and 2 performance. The Journal of the American Osteopathic Association. 2004;104(8):332-336.
3   Coumarbatch J, Robinson L, Thomas R, Bridge PD. Strategies for identifying students at risk for USMLE step 1 failure. Family medicine. 2010;42(2):105-110.
4   Gohara S, Shapiro JI, Jacob AN, et al. Predictors of success on the United States Medical Licensing Examinations (USMLE). Learn Assist Rev (TLAR). 2011;16(1):11–20.

5 Haight SJ, Chibnall JT, Schindler DL, Slavin SJ. Associations of medical student personality and health/wellness characteristics with their medical school performance across the curriculum. Academic medicine : journal of the Association of American Medical Colleges. 2012;87(4):476-485.

6 Burk-Rafel J, Santen SA, Purkiss J. Study Behaviors and USMLE Step 1 Performance: Implications of a Student Self-Directed Parallel Curriculum. Academic medicine : journal of the Association of American Medical Colleges. 2017;92(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 56th Annual Research in Medical Education Sessions):S67-s74.

7 McGaghie WC, Downing SM, Kubilius R. What is the impact of commercial test preparation courses on medical examination performance? Teaching and learning in medicine. 2004;16(2):202-211.

8 Winn, W. (1990). Some implications of cognitive theory for instructional design. Instructional Science, 19, 53–69.

9 Roediger HL, 3rd, Butler AC. The critical role of retrieval practice in long-term retention. Trends in cognitive sciences. 2011;15(1):20-27.

10 Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. Psychological science. 2006;17(3):249-255.

11 Deng F, Gluckstein JA, Larsen DP. Student-directed retrieval practice is a predictor of medical licensing examination performance. Perspectives on medical education. 2015;4(6):308-313.

12 Volk AS, Rhudy AK, Marturano MN, Ott L, DuCoin C. Best Study Strategy for the NBME Clinical Science Surgery Exam. Journal of surgical education. 2019.

13 Vora A, Maltezos N, Alfonzo L, Hernandez N, Calix E, Fernandez MI. Predictors of scoring at least 600 on COMLEX-USA Level 1: successful preparation strategies. The Journal of the American Osteopathic Association. 2013;113(2):164-173.

14 Miles, MB. & Huberman, AM. (1994). Qualitative Data Analysis (2nd edition). Thousand Oaks, CA: Sage Publications.

15 BEME Collaboration. BEME Coding Sheet – BEME Collaboration. www.bemecollaboration.org/downloads/749/beme4_appx1.pdf. Published 2012. Accessed June 11, 2019.

16 Baños JH, Pepin ME, Van Wagoner N. Class-Wide Access to a Commercial Step 1 Question Bank During Preclinical Organ-Based Modules: A Pilot Project. Academic medicine : journal of the Association of American Medical Colleges. 2018;93(3):486-490.

17 Ryan R; Cochrane Consumers and Communication Review Group. 'Cochrane Consumers and Communication Review Group: data synthesis and analysis'. http://cccrg.cochrane.org, June 2013 (accessed Sept. 2, 2019).

18 Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. a product from the ESRC methods programme. Version 1. ESRC, 2006.