




## Exploring students' [pre-pandemic] use and the impact of commercial-off-the-shelf learning platforms on students' national licensing exam performance: a focused review-BEME guide no. 72

Atsusi Hirumi, Luke Horger, David M. Harris, Andrea Berry, Feroza Daroowalla, Shalu Gillum, Nyla Dil & Juan C. Cendán


To cite this article: Atsusi Hirumi, Luke Horger, David M. Harris, Andrea Berry, Feroza Daroowalla, Shalu Gillum, Nyla Dil & Juan C. Cendán (2022): Exploring students' [pre-pandemic] use and the impact of commercial-off-the-shelf learning platforms on students' national licensing exam performance: a focused review-BEME guide no. 72, Medical Teacher, DOI: [10.1080/0142159X.2022.2039380](https://doi.org/10.1080/0142159X.2022.2039380)


To link to this article: <https://doi.org/10.1080/0142159X.2022.2039380>

 View supplementary material 

 Published online: 10 Mar 2022.



 Submit your article to this journal 

 Article views: 235

 View related articles 

 View Crossmark data 

## Exploring students' [pre-pandemic] use and the impact of commercial-off-the-shelf learning platforms on students' national licensing exam performance: a focused review-BEME guide no. 72

Atsusi Hirumi<sup>a</sup> , Luke Horger<sup>a</sup>, David M. Harris<sup>b</sup>, Andrea Berry<sup>b</sup> , Feroza Daroowalla<sup>b</sup>, Shalu Gillum<sup>b</sup>, Nyla Dil<sup>b</sup> and Juan C. Cendán<sup>b,c</sup>

<sup>a</sup>Learning Sciences, College of Community Innovation and Education, University of Central Florida, Orlando, FL, USA; <sup>b</sup>Medical Education, College of Medicine, University of Central Florida, Orlando, FL, USA; <sup>c</sup>Interim Dean, Herbert Wertheim College of Medicine, Florida International University, Miami, FL, USA

### ABSTRACT

**Background:** Commercial-off-the-shelf learning platforms developed for medical education (herein referred to as MedED-COTS) have emerged as a resource used by a majority of medical students to prepare for licensing examinations. As MedED-COTS proliferate and include more 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' national licensing exam scores in the USA.

**Methods:** Due to the limited number of studies published on either the use or integration of MedED-COTS, a focused review of literature was conducted to guide future research and practice. Data extraction and quality appraisal were conducted independently by three reviewers; with disagreements resolved by a fourth reviewer. A narrative synthesis was completed to answer research questions, contextualize results, and identify trends and issues in the findings reported by the studies included in the review.

**Results:** Results revealed consistent positive correlations between students' use of question banks and their licensing exam performance. The limited number of integration studies, combined with a number of methodological issues, makes it impossible to isolate specific effects or associations of integrated commercial resources on standardized test or course outcomes. However, consistent positive correlations, along with students' pervasive use and strong theoretical foundations explaining the results, provide evidence for integrating MedED-COTS into medical school curricula and highlight the need for further research.

**Conclusions:** Based on findings, we conclude that students use exam preparation materials broadly and they have a positive impact on exam results; the literature on integration of MedED-COTS into formal curriculum and the use by students of resources outside of exam preparation is scant.

### KEYWORDS

Commercial learning platforms; national board exams; focused review; BEME

## Introduction

In the USA, medical students seeking a Doctor of Medicine (MD) or Doctor of Osteopathic Medicine (DO) diploma and further graduate training must successfully complete national licensing exams. Data collected by the National Resident Matching Program from graduate residency programs cites performance on these licensing exams (National Board of Medical Examiners [NBME] USMLE Step 1; National Board of Osteopathic Medical Examiners [NBOME] COMLEX-USA Level 1; NBME USMLE Step 2; and NBOME COMLEX-USA Level 2) as the first and fourth most considered factors when selecting medical students for consideration for admission to the graduate program (National Resident Matching Program 2018). The same survey cites clinical specialty course (clerkship) grades as the sixth most important factor for consideration for medical student selection into graduate residency program (2018).

### Practice points

- MedED-COTS, particularly Qbanks that are based on spaced repetition and active recall frameworks, are associated with better board exam performance.
- Institutions should consider integrating MedED-COTS formally into the curriculum or supporting their use financially.
- MedED-COTS can serve as a tool to reduce faculty efforts on development of Qbanks and monitoring of individual student performance.
- Research is needed on how to best incorporate MedED-COTS.
- Future research should focus on the consequence of the USMLE Step 1 going to pass/fail on MedED-COTS usage and capabilities.

Often, one component of clinical clerkship grades is performance on a standardized national subject exam (the NBME Subject Exam ('shelf exams') or NBOME Comprehensive Osteopathic Medical Achievement Test (COMAT)) (National Resident Matching Program 2018).

Research on how students prepare for these national standardized exams, whether for licensing or for course grades, seeks to clarify factors associated with higher scores and students' reliance on material outside the curricular resources provided by the institution. Dixon (2004) found that pre-clinical course performance is positively correlated with NBOME COMLEX-USA Level 1 and 2 scores. Pre-matriculation entry exam performance, including Medical College Admission Test (MCAT) has also been positively correlated with NBME Step 1 performance (Coumarbatch et al. 2010; Gohara et al. 2011; Haight et al. 2012). However, within the last decade, research indicates that students are more likely to supplement school curriculum with commercial learning resources to prepare for national standardized medical exams instead of lecture material (Burk-Rafel et al. 2017). For this focused review, the term MedED-COTS is defined as commercially distributed (**Commercial-Off-The-Shelf**) learning resources that are designed to help medical students prepare for national board exams and may also be used to facilitate Medical Education. Examples of MedED-COTS include AMBOSS, Aquifer, Boards and Beyond, Firecracker, Kaplan USMLE, Lecturio, OnlineMedEd, and ScholarRX. Print-based review books that may be offered in digital format, such as but not limited to First Aid and USMLE Lecture Notes, are also included (ScholarRX 2020).

One previous review synthesized the results of research on medical students' use of MedED-COTS to prepare for NBME exam performance. In their 2004 review, McGaghie, Downing, and Kubilius found that (1) commercial coaching was well-received by medical students but did not significantly improve Step (licensing) exam performance, (2) research lacked rigor and control, and (3) almost no details were given about the form or conduct of the courses (2004). However, the 2004 review centered only on the use of test preparation books and commercial coaching courses. Current generation MedED-COTS offer integrated and interactive features, such as multi-media content which is continually updated, question banks (or 'Qbanks') with spaced repetition and timed tests, capabilities for user-generated flashcards and notes, data analytics, customized content based on user performance and self-assessment, and supplementary coaching by experts. Furthermore, the 2004 review did not distinguish medical students' use of MedED-COTS outside of medical curriculum from their use as prescribed, integrated components of the medical school curriculum. Our review builds on the previous review by: (1) examining the impact of new features of MedED-COTS, such as Qbanks, flashcards, and multimedia that have evolved over the past decade; and (2) distinguishing studies that examine students' use of MedED-COTS as a supplement to the formal curriculum versus use within and integrated into the formal medical school curriculum.

Students have several motivations for using MedED-COTS. They feel that MedED-COTS will optimize their performance on the exams, and have concerns that formal

curricular content and activities are not explicitly related to licensing exam content. Faculty has concerns about errors or omissions, and feel challenged on how to integrate these resources into curricula without excluding other content and competencies. This focused review speaks to these motivations and concerns, and can inform faculty and students on the advantages and disadvantages of using certain MedED-COTS features to prepare for licensing examinations. The review summarizes MedED-COTS research before two recent significant events: (1) the COVID-19 pandemic and its requirement for remote teaching and limitations on clinical experiences, and (2) changes to the grading scale of NBME USMLE STEP 1 exam. Both events may have implications for how curricula and students utilize and integrate MedED-COTS. Due to the emerging yet limited number of studies published on the use or integration of MedEd-COTS, we conducted a focused review of literature to set the stage for future research and practice.

### Questions addressed in this review

Our focused review addressed 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 studies led to primary and secondary questions for both educators and researchers.

The primary research question for this review was: *Is there a relationship between students' use of MedED-COTS and their performance on undergraduate licensing exams?*

Secondary research questions based on identified studies were:

1. Which platforms and product features are most valued/used by undergraduate medical students who use MedED-COTS outside of the formal medical school curriculum to prepare for licensing exams?
2. When do undergraduate MD and DO medical students use MedED-COTS?
3. Why do students use MedED-COTS, other than to prepare for standardized tests?
4. What research questions on MedED-COTS usage and outcomes have been posited by post-2004 studies?
5. What methods were used to answer research questions proposed by post-2004 studies?
6. What gaps or opportunities remain for advancing our knowledge of the use of MedED-COTS outside of the formal curriculum or as an integrated component of formal curriculum to prepare for licensing exams?

Three additional questions were posited for studies on integration of MedED-COTS into the formal curriculum:

1. Why are faculty integrating the use of MedED-COTS into the formal medical school curriculum?

1. Which MedED-COTS platforms and MedED-COTS features are most valued/used by faculty to enhance medical education?

2. Are there instructional approaches, theories, and strategies grounded in research being used to guide the integration of MedEd-COTS?

## Methods

### Search strategy

A systematic review was conducted on the efficacy of MedEd-COTS in 2004 (McGaghie et al. 2004). This current review examined studies from 1 January 2004 to 29 January 2020. Only full-text papers with empirical data were included. A total of eight databases were searched: EMBASE, ERIC, MEDLINE (PubMed), Ovid, PsycInfo, Web of Science, Open Grey, and ProQuest Dissertations & Theses Global. Conference proceedings from 2016 to 2020 were searched from the following sources: International Association of Medical Science Educators (IAMSE), Association for the Study of Medical Education (ASME), and the Association for Medical Education in Europe (AMEE). The first ten pages of Google Scholar were also examined. Additionally, the reference lists of included studies were examined for relevant sources.

The full search string utilized for this review can be found in [Supplementary Appendix 1](#), along with the adaptations used for the databases searched. For example, the Medical Subject Heading (MeSH) 'Students, Medical' was omitted from the ERIC, PsycInfo, and Google Scholar searches. A 'polyglot' tool was utilized to translate search strings between databases (Bond University 2019). 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 these did not generate additional search results. Thus, names were not included in the search strategies. Our search terms also originally included the MedEd-COTS features 'video' and 'data analytics,' but we found those search terms did not generate unique search results.

### Inclusion/exclusion criteria

The inclusion criteria for this review were as follows:

1. Quantitative studies that involve numerical data in survey, correlational, quasi-experimental, or experimental research;
2. 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;
3. Studies examining the target population of first year, second year, third year, or fourth year MD or DO medical students or librarians at MD or DO schools;
4. 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;
5. Research published in English; and
6. Research on MedEd-COTS usage in the USA or Canada.

Exclusion criteria for this review were as follows:

1. Studies that examine medical residents after graduation or undergraduate students before matriculation into medical school;
2. Studies without an abstract;
3. Studies with no empirical data including commentaries and opinion pieces; and
4. Studies not published in English.

### Screening and selection of included studies

Two authors (SG and LH) independently reviewed titles and abstracts of all search results. If a paper's title and abstract met inclusion criteria, the full text of the article was reviewed. Any disagreement about inclusion was resolved by a third author (AH). A total of 4064 records were identified through the database search, and hand-searching resulted in the identification of 9 additional records. After de-duplication in EndNote reference management software, 3428 records remained. Title and abstract screening resulted in an inter-rater reliability score of 99% between the two reviewers. Differences were remediated by discussion with the third reviewer. After remediation was completed, 36 records remained. Of those 36, full-text screening was completed and 19 studies were included for data extraction. Reasons for exclusion included lack of empirical MedEd-COTS data, reference to non-commercial technology, non-empirical study design such as commentary, and inappropriate target population. The PRISMA flow diagram for article selection is shown below ([Figure 1](#)).

### Data extraction and remediation

Papers included in the review were classified as either 'usage' or 'integration,' where usage studies examined students' use of MedEd-COTS outside of their formal medical school curriculum, and integration studies involved a medical school providing students access to certain MedEd-COTS features with recommendation that students utilize those features. Data extraction instruments were created using Qualtrics survey software. A pilot review was conducted with three of the 19 included studies to refine the research questions and train coders to ensure consistency of data extraction by practicing identifying mediating variables, reviewing the difference between research problems *versus* research purpose, and setting a precedent of extracting data from studies verbatim with no subjective interpretation. Data extraction for *all* studies addressed the following:

1. Author(s) and date of publication;
2. Methods (subjects, designs, instrumentation, procedure, and limitations);
3. Whether a relationship existed between MedEd-COTS use and performance on NBME or NBOME exams;
4. When M.D. and D.O. students used MedEd-COTS;
5. Why students used MedEd-COTS;
6. Factors that influenced students' use of MedEd-COTS; and
7. Research questions on MedEd-COTS usage and outcomes proposed by post-2004 studies.

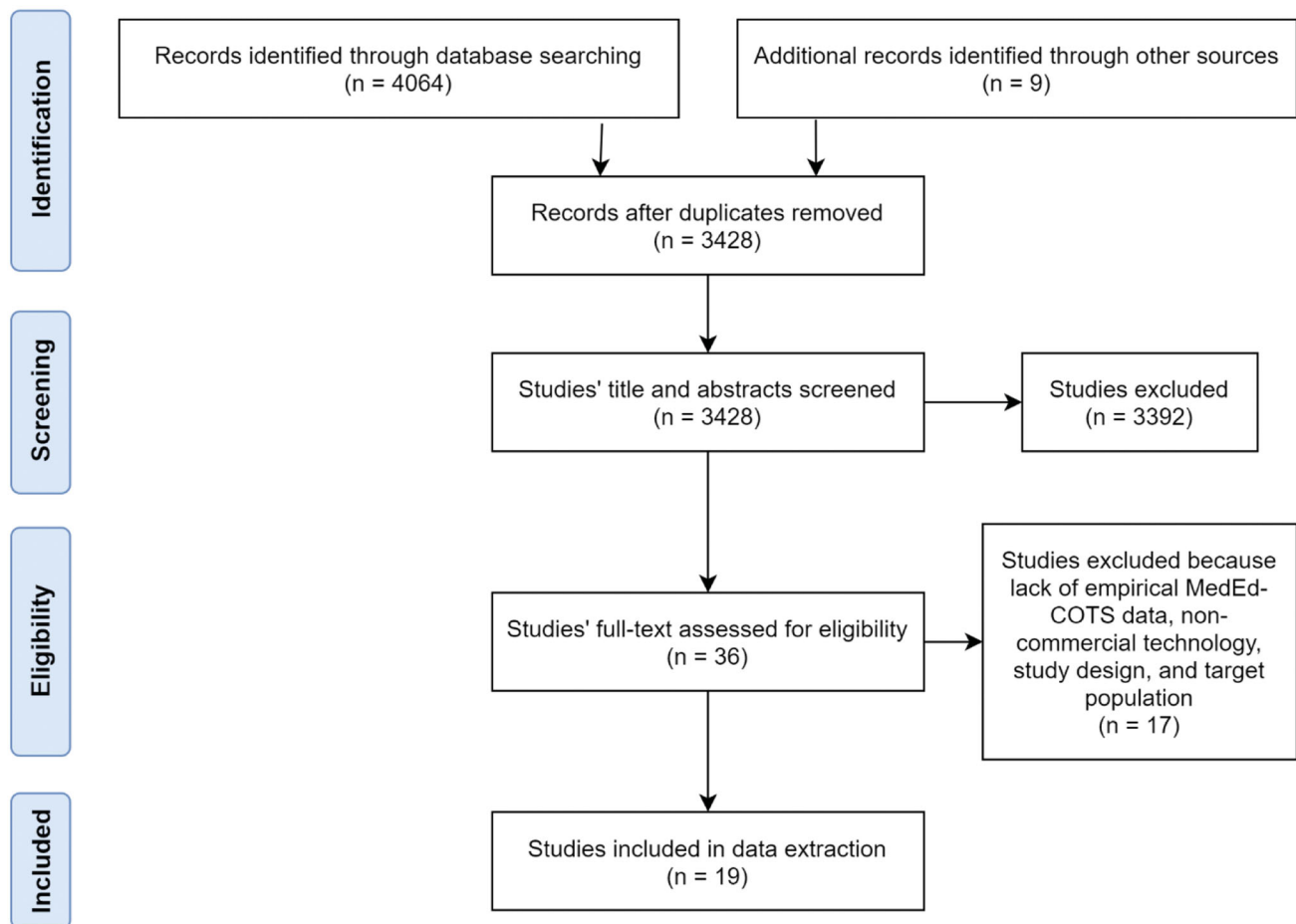


Figure 1. PRISMA scheme.

Data extraction for *usage* studies addressed the following: Which MedEd-COTS platforms and features did students prefer to use outside of the formal medical school curriculum to prepare for NBME and NBOME exams?

Data extraction for *integration* studies addressed the following:

1. Why did faculty provide student's access to MedEd-COTS or integrate the use of MedEd-COTS into the formal medical school curriculum?
2. Which MedEd-COTS platforms and features were most valued/used by faculty to enhance medical education?
3. Which instructional approaches, theories, and strategies were being used to ground and guide the integration of MedEd-COTS?

Three independent reviewers extracted data from publications into the Qualtrics surveys. Reviewers completed the appropriate forms (see [Supplementary Appendices 2 and 3](#)), depending on whether an article was designated a 'usage' or an 'integration' study. All discrepancies were remediated by discussion and resolution with a fourth reviewer.

### Data analysis and synthesis

Both quantitative and qualitative data of MedEd-COTS usage were collected. Quantitative studies examined the association of student use of MedEd-COTS with licensing exam performance. Due to the variation in research design across studies and the wide variability of MedEd-COTS

platforms and features, we could pool neither the studies nor the data.

Studies with qualitative data included survey and interview data on why students used specific MedEd-COTS and features, when they used them, and how these variables affected various outcomes (e.g. exam performance). However, since no study utilized a qualitative design as their primary research method, these data cannot be interpreted through meta-ethnography. We conducted a narrative synthesis to identify patterns, discuss trends and issues reported in the studies we reviewed, and to answer our primary and secondary research questions.

### Summary of the quality appraisal process

Three independent reviewers assessed the risk of bias of each paper according to the following dimensions from the standard BEME coding sheet (BEME Collaboration 2012):

- A. Well defined problem and related purpose and/or research question(s)
- B. Appropriateness of study design
- C. Implementation of study design
- D. Appropriateness of data analysis
- E. Appropriateness and strength of findings

Based on a protocol pre-approved by the BEME International Collaborating Centre (BICC), all studies are presented here regardless of quality appraisal, which is summarized as 'strength of conclusions' (Table 1). Each

**Table 1.** Summary of characteristics and quality appraisal of included studies<sup>a</sup>.

Author (Year)	Study type	MedED-COTS	Participants	Outcomes	Quality					Strength of conclusions
					A	B	C	D	E	
1. Banos et al. (2018)	Integration (Pre-experimental survey)	Kaplan USMLE Step 1 Qbank	185 first year medical students at a single institution	Instructor-designed exams, NBME Customized Assessments, module final grades, and USMLE Step 1 exam score	3	3	3	3	3	High
2. Bonasso et al. (2015)	Use (Non-experimental survey)	Board Review, Doctors in Training, First Aid, Kaplan, UWorld/USMLE World	164 second year medical students from two cohorts at a single institution	USMLE Step 1 exam score	3	2	3	3	3	Adequate
3. Burk-Rafel et al. (2017)	Use (Non-experimental survey)	Firecracker, First Aid, Goljan, Kaplan, Pathoma, USMLE-Rx, UWorld/ USMLE World, NBME Basic Science Exam	274 medical students from two cohorts immediately after completing pre-clinical pass-fail curriculum and Step 1 at a single institution	First attempt USMLE Step 1 exam score	3	3	3	3	3	High
4. Deng et al. (2015)	Use (Non-experimental mixed method)	Anki, Firecracker, USMLE-Rx	72 medical students who completed preclinical curriculum and had taken Step 1 in 2014 at a single institution	USMLE Step 1 exam score	3	3	3	3	2	Adequate
5. Giordano et al. (2016)	Use (Non-experimental survey)	First Aid, UWorld	81 end of second year medical students from two cohorts at a single institution	Step 1 exam score	2	3	3	3	3	Adequate
6. Zhang et al. (2005)	Use (Non-experimental mixed method)	Kaplan test preparatory course	100 third year medical students after taking Step 1 but before receiving results at a single institution	Step 1 exam score	2	3	3	3	2	Adequate
7. Jackson et al. (2019)	Use (Non-experimental survey)	Doctors in Training, Kaplan, Pathoma, Sketchy, UWorld/USMLE World, TrueLearn, NBME CBSSA, COMSAE, COMbank, COMlex, COMquest	102 third year osteopathic medical students from one class at a single institution	COMLEX-USA Level 1 and USMLE Step 1 exam scores	3	3	3	3	2	Adequate
8. Schwartz et al. (2018)	Integration (Retrospective quasi-experimental mixed method)	First Aid, Kaplan, UWorld.	995 second year medical students from 6 cohorts at a single institution	USMLE Step 1 exam score and first time pass rate	3	3	3	3	3	High
9. Vora et al. (2013)	Use (Non-experimental survey)	First Aid, Kaplan USMLE videos, Combank, COMSAE	132 third year osteopathic medical students after taking COMLEX-USA Level 1 at a single institution	COMLEX-USA Level 1 exam score	3	2	3	3	3	Adequate
10. Kumar et al. (2015)	Use (Non-experimental survey)	Practice questions (publisher not reported)	256 third year medical students from three cohorts after taking Step 1 at a single institution	USMLE Step 1 exam score	3	3	3	3	3	High
11. Kastenmeier et al. (2018)	Integration (Quasi-experimental mixed method)	WISE-MD	394 third year medical students after completing surgery clerkship and after fourth year surgery AI or	Perceived value of ILPs, self-directed learning skills and NBME surgery subject exam score	3	3	3	3	3	High

(continued)

Table 1. Continued.

Author (Year)	Study type	MedED-COTS	Participants	Outcomes	Quality					Strength of conclusions
					A	B	C	D	E	
12. Taylor et al. (2018)	Use (Non-experimental survey)	First Aid, UWorld/USMLE World, Pestana's surgery notes, NMS surgery, Wikipedia, Surgical Recall, UpToDate, Medscape, YouTube, Cancer.gov, WebMD	92/133 third year medical students at the end of the third year surgery clerkship at a single institution	NBME Surgery subject exam score	3	3	3	3	3	High
13. Volk et al. (2019)	Use (Non-experimental survey)	Online Qbanks; high yield review book, case-based exams, additional case-based books, rapid fire question and answer books (publisher not reported)	82/193 third year medical students who completed the surgery clerkship at a single institution	NBME Shelf Exam score for Surgery	3	3	3	3	3	High
14. Maholtz et al. (2015)	Use (Non-experimental mixed method)	Doctors in Training, First Aid, UWorld/USMLE World, COM bank and COMATs	399 Osteopathic medical students who had taken the COMLEX within past year in a multi-institutional study	COMLEX-USA exams (Level 1 and 2-CE) scores. Perceived usefulness of resources	2	2	1	1	1	Flawed
15. Lau and Kolli (2017)	Use (Non-experimental mixed method)	UWorld/USMLE World, Epocrates, UpToDate, and Medscape	86/112 third year medical students at a single institution	Use and perceived value of apps	3	3	3	1	3	Flawed
16. Hendrix and Hasman (2008)	Use (Non-experimental survey)	Print and Electronic (E-book, CD-ROM, Web-based) USMLE resources. Book and Flashcard NBDE resources (publishers not reported)	Librarians from 58/125 AAMC-accredited medical schools and 23/56 ADA accredited dental schools in a multi-institutional study	Provision of print and electronic exam preparation resources	3	3	3	3	3	High
17. Yavner (2016)	Integration (Sequential mixed method-Dissertation)	UWorld/ USMLE World, WISE-MD, MD on Call module	Third and fourth year medical students (n = 70 study 1, n = 52 study 2, n = 106 study 3) at a single institution	Student preference for learning materials (strategy and resource)	3	3	3	3	3	High
18. Seal et al. (2020)	Use (Non-experimental survey)	UWorld/USMLE World and NBME practice tests	399 Medical students who have taken Step 1 in a multi-institutional study	USMLE Step 1 exam score	3	2	3	2	3	Adequate
19. Andryka et al. (2014)	Use (Retrospective mixed method)	First Aid, Kaplan Lecture Notes and Qbank, Rapid Review Pathology, The Pre-Test Series, Board Review Series, High-Yield Series, USMLE Gunner, Med Master, Student Doctor Network forums, Gunner Library, NBME Practice exams, Doctors in Training, GoJian Audio Lectures, USMLE Rx	Unspecified number of medical students posting to multiple blogs in a multi-institutional study	Student reported pros and cons	1	1	1	1	2	Flawed

<sup>a</sup>A: well-defined problem; B: appropriate design; C: design implementation; D: appropriate data analysis; E: appropriate findings

dimension was assessed on a 1 to 3 low-high scale. A quality score of 3 signified that reviewers found that the paper adequately addressed a quality dimension. A score of 2 indicated that, although a paper addressed a quality dimension, there was room for improvement. A score of 1 signified a paper did not address a quality dimension. Articles with quality scores of 3 and 2 were used to identify trends and patterns in the results of this review, and to formulate conclusions. Articles with at least one quality score of 1 were not used to identify trends and patterns in the results nor used to derive conclusions, and should be interpreted with caution as denoted by a flawed strength of conclusions score in the [Table 1](#). The quality of each paper has been summarized in [Table 1](#), along with a determination of the strength of each paper's conclusions.

The papers that received a quality appraisal score of 1 and reasons for that rating are as follows:

1. *Comprehensive Osteopathic Medical Licensing Examination-USA Level 1 and Level 2-Cognitive Evaluation Preparation and Outcomes* (Maholtz et al. 2015); poor implementation of research design, lack of data analysis, and poor presentation of results.
2. *App Use in Psychiatric Education: A Medical Student Survey* (Lau and Kolli, 2017); lack of data analysis.
3. *Too Smart to Fail: Guide for the Struggling Medical Student* (Andyryka et al. 2014); a poorly defined problem, unclear research design and lack of data analysis.

Nine out of the 19 studies were found to meet all quality dimensions and thus received a quality score of 3. Eight of the 19 studies received combinations of quality scores of 2 and 3, and therefore were considered to meet quality criterion but had room for improving its study design, definition of the research problem, and/or data analysis.

Three studies received at least one quality appraisal score of 1; one study failed to meet four criteria, one study failed to meet three criteria, and one study was flagged for meeting all but one criterion. The findings of these papers are noted in the narrative synthesis, and readers are cautioned when interpreting or making recommendations on use or integration of MedEd-COTS based on any of three papers receiving a quality score of 1.

#### **Additional limitations of this focused review**

In addition to concerns for quality, the interpretation and discussion of results were also limited by the number and nature of studies included in the review. As noted in the introduction, we conducted a focused review due to the emerging nature and limited number of studies that were found that examined the use and integration of MedEd-COTS. However, the limited number of integration studies, combined with issues identified by reviewing the research questions and methods followed by each study made it difficult to answer the secondary research questions we posed in this review. Furthermore, the studies included in this review were primarily correlational in nature, based on associations found between surveys, observations and test scores. Those designated as 'integration' studies were either non-experimental or pre-experimental. This review was also limited to studies in the pre-COVID-19 era. The COVID-19 pandemic of 2020–2021 will surely give rise to a

new set of studies about how students use MedEd-COTS and how they are integrated into the curriculum due to restrictions in synchronous face to face teaching and student exposure to patients.

## **Results**

The findings are organized and presented according to the primary and secondary research questions posed by this review, beginning with key characteristics of the studies to provide context for review findings.

### **Study characteristics**

#### **Research questions addressed by reviewed studies**

[Table 1](#) summarizes the key characteristics of the 19 studies included in our review. The majority of studies (15/19) included in this review examined the association of MedEd-COTS use outside of the formal curriculum to national exam performance. Varying features of MedEd-COTS examined by the studies included: NBME practice tests, self-directed methods of testing (with or without spaced repetition), Qbanks, test preparation books and courses, and study habits.

Other research questions, not directly related to the use of MedEd-COTS, that were posed in several studies to identify predictors of licensing exam performance were: the effect of MCAT scores, length of time preparing, study format (individual or group), preclinical GPA, medical school performance, and medical school curriculum on board performance. MCAT scores remain the primary confounding variable as four papers found a positive correlation between MCAT score and USMLE Step 1 performance (Deng et al. 2015; Burk-Rafel et al. 2017; Banos et al. 2018; Seal et al. 2020) and another linked MCAT to COMLEX-USA Level 1 score (Vora et al. 2013). Another variable of interest examined by five reviewed studies was students' performance in medical school courses. In each case, performance was positively correlated to exam performance (Vora et al. 2013; Deng et al. 2015; Maholtz et al. 2015; Burk-Rafel et al. 2017; Taylor et al. 2018).

The remaining studies addressed varying research questions with little overlap among them in the outcomes studied. For example, Jackson et al. aimed to determine if retrieval practice efficacy differs amongst various MedEd-COTS (2019). Questions about the use of MedEd-COTS on preclinical academic performance and NBME clerkship exam performance (Banos et al. 2018; Taylor et al. 2018) were studied. Other studies aimed to explore how students use apps and the value of app use (Lau and Kolli 2017); factors that make a learning strategy and resource selection more or less appealing under conditions of stress and fatigue (Yavner 2016); or reviewed critical resources for helping struggling medical students (Andyryka et al. 2014). Finally, in the study by Kastenmeier et al., while MedEd-COTS were mentioned as a resource in the program being evaluated, the research question focused on the impact of individual learning plans on educational outcomes and acquisition of self-directed learning skills (2018).



### ***What methods were used in reviewed studies***

The primary outcome variable of interest was student performance on national standardized licensing exams. Ten studies correlated the use of MedED-COTS with USMLE Step 1, three with COMLEX-USA Level 1) and three with standardized clerkship subject exams (NBME Surgery Shelf Exam).

The majority of reviewed studies (15/19) looked at the use of student-selected MedED-COTS, whereas four studies examined students use of MedED-COTS prescribed by the school as part of the formal curriculum. The number of MedED-COTS considered in each study varied significantly, with six studies assessing the impact of one MedED-COTS on outcomes with the remaining 14 studies considering a range of 2-16 MedED-COTS (mean of 6).

One dissertation was reviewed that consisted of three sequential, non-experimental investigations that included both quantitative and qualitative data. Of the 18 peer-reviewed journal articles included in the review, 15 were considered non-experimental with neither comparison group nor manipulated variable, one was classified as pre-experimental with a manipulated variable but no comparison group, and two were considered quasi-experimental with manipulated variables, comparison groups but no random assignment. Otherwise, 10 of the 18 peer-reviewed articles utilized quantitative surveys to gather data, and seven studies included at least one qualitative and one quantitative measure, and were thus classified as mixed-method study type. Fifteen studies represented samples from single institutions and four were multi-institutional.

### ***Primary question: is there a relationship between students' use of MedED-COTS and their performance on licensing exams?***

Although limited in number, there is direct evidence that the use of specific MedED-COTS increases students' performance on the USMLE Step 1 exam. Six of 19 studies reported significant positive correlations between the use of Kaplan USMLE Step 1 Qbank, NBME Comprehensive Basic Science Exam, Firecracker, UWorld, Anki Flashcards, COMSAE, and USMLE-Rx and USMLE Step 1 scores. The studies attributed higher exam scores to question bank use, number of items attempted, number of unique questions completed, timing in relationship to sitting for the Step 1, and familiarity with the questions that, in turn, point to the value of using Qbanks with spaced repetition and retrieval practice (Brown et al. 2014) in achieving good USMLE Step 1 scores. Complicating this synthesis, two of the six studies also reported non-significant correlations between various MedED-COTS or MedED-COTS features and Step 1 performance. These findings were discussed in the context of using multiple MedED-COTS, and using MedED-COTS that are designed to prepare students for COMLEX examinations (Bonasso et al. 2015; Jackson et al. 2019). While most of the studies provide support for MedED-COTS use to improve Step 1 scores, it appears that the mechanisms by which each resource may improve scores vary, thus making it difficult to make a broad conclusion. Across this group of studies, effectiveness of resources can be explained by use, number of items or tests taken, and timing of use.

Beyond the MedED-COTS that are independently predictive of Step 1 success, three studies also found positive correlations for students who utilized a combination of MedED-COTS and USMLE Step 1 performance. The studies described how students had access to and used multiple MedED-COTS including Firecracker, First Aid, Goljan, Kaplan, Pathoma, USMLE-Rx, UWorld, NBME Comprehensive Basic Science exam, Doctors in training, Sketchy Micro and Pharm, TrueLearn, NBME CBSSA, COMSAE, ComBANK, COMLEX, and COMQUEST (Burk-Rafel et al. 2017; Jackson et al. 2019) thus resulting in their improved performance. While less helpful in providing practice points for faculty and students considering the use of specific MedED-COTS in the curriculum, the results reinforce the use of resources to improve test scores. As with the studies that examined the use of individual MedED-COTS, those evaluating the use of a combination of resources attributed student performance to the number of times a student read a resource and the number of unique questions completed. Across all reviewed studies, use of UWorld, Pathoma, USMLE-Rx, NBME Comprehensive Basic Science Exam, and Kaplan alone or in combination with other MedED-COTS correlate positively to USMLE Step 1 performance.

In addition to correlations with USMLE Step 1 scores, positive correlations were found between the use of MedED-COTS and COMLEX-USA Level 1 exam performance. Similar to studies whose dependent outcome was USMLE Step 1, in publications in which COMLEX-USA Level 1 is the outcome of interest, there are very few independent correlations identified. However, there were two studies that evaluated a combination of resources on COMLEX-USA Level 1 performance. NBME CBSSA, COMSAE, Sketchy Micro, UWorld and TrueLearn as well as USMLE books, Kaplan USMLE videos, COMBANK and COMSAE were positively correlated to board performance (Jackson et al. 2019; Vora et al. 2013). Overall, the studies did point to the importance of number of questions completed, practicing questions in close proximity to the exam, and completing questions in timed mode.

### ***Secondary research questions for use and integration studies***

#### ***Student perceptions of platforms and features***

There were no specific MedED-COTS preferred by students across studies in this review. However, there was a strong preference for platforms that incorporated questions. In 15 of the 19 of reviewed studies, students were given the choice of which MedED-COTS or MedED-COTS features to use. Even in the four studies where MedED-COTS resources were made available formally or integrated into the curriculum, it was still up to the students to determine which resources to use, how to use them, the frequency of use, and to what extent they utilized specific features of each program (Banos et al. 2018; Kastenmeier et al. 2018; Schwartz et al. 2018; Yavner 2016). Four out of 19 studies in the review examined student perceptions and preferences for platforms and/or features (Banos et al. 2018; Kastenmeier et al. 2018; Schwartz et al. 2018; Yavner 2016). Three of the four studies reported that students felt MedED-COTS were useful in preparing for examinations including USMLE Step 1 and COMLEX-USA Level 1 and 2,

and NBME subject exams (Lau and Kolli 2017; Kastenmeier et al. 2018; Maholtz et al. 2015), while two showed that students utilize such resources based on perceived educational value, or their own preferences for learning materials (Yavner 2016; Lau and Kolli 2017). Within the four articles, there were no commonalities in student preferences for certain MedED-COTS. However, two of the four articles emphasized student preference for resources with questions (Lau and Kolli 2017; Yavner 2016). Students also reported appreciation for their accessibility, interactivity, and ability to prepare them for exams.

Student perception of educational value or preferences for MedED-COTS centered on several separate themes. Students found MedED-COTS more than or equally effective as traditional didactic lectures (Kastenmeier et al. 2018), chose particular MedED-COTS apps useful in certain clinical rotations (Lau and Kolli 2017), and used them based on alignment with assessment, immediate learning goals, time constraints, ability to receive immediate feedback, and level of student fatigue and stress (Yavner 2016). While these findings are based on single studies, the overarching theme across the three studies point to the usefulness of MedED-COTS compared to other resources provided by the medical education curriculum. It appears that how the MedED-COTS are packaged with content, coupled with formative assessments in an easy to use format, has influenced student use.

#### ***Timing of MedED-COTS use***

Evidence points to the importance of using MedED-COTS in advance of dedicated study periods and sitting for USMLE Step 1. Six of the 19 articles included in the review evaluated the effects of timing on Step 1 performance. Four of these articles pointed to the importance of early preparation, prior to dedicated study periods, as a significant predictor of higher Step 1 scores (Giordano et al. 2016; Burk-Rafel et al. 2017; Vora et al. 2013; Kumar et al. 2015). However, one investigation suggested that studying near the exam period, versus the amount of time used for studying, was significantly correlated to board examination success (Jackson et al. 2019). While the current evidence clearly supports early preparation as an important factor in exam performance, timing of preparation is less clear, as one study (Jackson et al. 2019) supported the importance of studying near the exam period while another found that timing has no correlation with board scores (Bonasso et al. 2015).

#### ***Reasons students use MedED-COTS***

Students' primary motivation for utilizing MedED-COTS is to prepare for and excel on high stakes examinations. The primary reason students focus on excelling on such examinations relates to their ability to match into a competitive specialty or program (Zhang et al. 2005; Hendrix and Hasman 2008; Bonasso et al. 2015; Maholtz et al. 2015). Additionally, three studies pointed to lack of preparation material or experiences provided by the medical school to ensure student success on licensing or course exams (Maholtz et al. 2015; Burk-Rafel et al. 2017; Taylor et al. 2018). Related to the importance of performing well on high stakes examinations, six additional studies discussed

the role of MedED-COTS in providing high yield material and practice questions as a tool for improving knowledge retention (Andryka et al. 2014; Burk-Rafel et al. 2017; Volk et al. 2019; Giordano et al. 2016; Jackson et al. 2019; Kumar et al. 2015). While seen as somewhat different concepts, practice questions tend to focus on high yield content thus providing further evidence as to the role of Qbanks as a tool to support spaced repetition, retrieval practice, and as an important contributor to student use.

Student perceptions of which resources will best prepare them to excel on high stakes exams, that at times are informed by solid evidence and other times less informed by such systematic evidence, is an important consideration when looking at student use of MedED-COTS. With resources like 'Charting Outcomes in the Match,' students are able to determine the qualifications that program directors within their preferred specialties are seeking (National Resident Matching Program 2020). An important piece of data in 'Charting Outcomes in the Match' is the average USMLE scores for applicants who were successful at matching into the various specialties (National Resident Matching Program 2020). USMLE score data in 'Charting Outcomes in the Match' then presses medical students into avenues for achieving the highest scores by using MedED-COTS that are at their disposal (National Resident Matching Program 2020).

#### ***Secondary research questions for integration studies only***

##### ***Reasons for integrating the use of MedED-COTS into the formal medical school curriculum***

Faculty and students appear to share the same goals in integrating MedED-COTS in the formal curriculum. Although multiple variables are at play across reviewed publications, two studies discuss the role of MedED-COTS in preparing students for national board exams (Banos et al. 2018; Schwartz et al. 2018). Faculty also assume that MedED-COTS are engaging and efficacious for learning important scientific facts, concepts, and principles, thus promoting their use (Yavner 2016; Schwartz et al. 2018). The close alignment of MedED-COTS content and Qbanks with licensing exams as well as the way in which content is packaged and delivered prompts their formal and informal use during medical school. Faculty often lack the time and expertise to produce such resources thus prompting their use to fill in this gap.

##### ***MedED-COTS platforms and features***

While there was no explicit mention of faculty preferences for specific platforms or features in these studies, faculty appear to value Qbanks, clinical cases, and videos as they seek to integrate the use MedED-COTS within the curriculum to enhance exam performance. Similar to the results synthesized for answering the primary research question, Qbanks in Kaplan and UWorld were positively correlated with exam scores, thus underscoring the preference for these platforms over others discussed in this review (Banos et al. 2018; Schwartz et al. 2018). Interestingly, two studies evaluate the use of WISE-MD and particularly the clinical cases and videos as valuable study tools, thus pointing to

the value of other pedagogical approaches beyond Qbanks in meeting student and faculty goals for learning (Yavner 2016; Kastenmeier et al. 2018).

### ***Instructional approaches and theories guiding the integration on MedED-COTS***

Cognitive-based learning theories were the primary frameworks that guided the integration and study of MedED-COTS in MD curriculum. The four integration studies all report evidence-based instructional approaches or theoretical frameworks in the design of the study or as it relates to the integration of MedED-COTS into the curriculum. In two studies, such theories meaningfully contribute to the generalizability of results. While there are no themes related to the theories emphasized in these papers, both self-directed learning and cognitive load theories were posited to understand how students use questions and access content to promote retention of knowledge (Kastenmeier et al. 2018; Yavner 2016). The concepts of self-directed learning and cognitive load are supported by a less rigorous study that emphasizes deliberate practice, active retrieval and transfer-appropriate-processing, which are all cognitive-based learning theories (Banos et al. 2018).

## **Discussion**

The primary finding from our focused review was the positive correlations reported between the use of MedED-COTS and performance on licensing examinations. Seventeen of the nineteen studies included in this review were deemed adequate or high strength of evidence in support of the conclusions. While we were unable to adequately address some of our secondary research questions with the current literature, our primary research question of whether MedED-COTS improve licensing examination performance is well-supported. Students' pervasive use of MedED-COTS also provides additional evidence in support of their use and integration. Furthermore, the use of Qbanks and spaced repetition is rooted in well-established educational frameworks, including empirical research and theory on deliberate practice, spaced retrieval practice, testing, and active recall. The results of our focused review led to three areas of discussion: (1) building on existing (pre-pandemic) MedED-COTS research; (2) implications for educational practice; and (3) gaps in studies and further research needs.

### ***Building on existing (pre-pandemic) MedED-COTS research***

The prior BEME review on the use of commercial products in medical education was conducted in 2004 (McGaghie et al. 2004). Consistent with that 2004 synthesis, our review also (a) focused on the impact of commercial products on national board exams, (b) found that studies included in the review lacked rigor and control, and (c) found the lack of studies examining the integration of MedED-COTS with medical school curriculum.

Both the 2004 and our current review examined the impact of using commercial resources on students' national NBME and NBOME exam performance. A consistent feature

of studies in both reviews was the lack of control samples for the study population. Because the studies were neither prospective nor randomized, the findings were limited to associations and correlations. Another consistent finding between the 2004 BEME review and the current set of literature was the relative lack of publications on the integration of MedED-COTS into formal curricula. Rather, the literature reported on schools providing access to MedED-COTS to help students prepare for licensing exams or sought students' perceptions of MedED-COTS. Based on the reviewed literature, MedED-COTS appear to be supplementary to the formal curriculum; few pre-pandemic studies prescribed the use of MedED-COTS to facilitate the achievement of curricular goals and objectives besides student-driven use for exam preparation.

Unlike the 2004 review that found that the use of commercial educational product use was not related to test performance, we found that specific features of MedED-COTS were positively correlated to students' licensing exam performance across studies. Significant correlations were found between Qbank use, number of items attempted, number of unique questions completed, timing in relationship to sitting for the Step 1 USMLE examination, and familiarity with the questions in six studies we reviewed. The use of Qbanks is supported by the theoretical framework of deliberate practice (Ericsson et al. 1993) and spaced retrieval practice (Brown et al. 2014). Deliberate practice is focused, repetitive effort on a task to improve performance. Spaced retrieval practice, in turn, has been found to reinforce meaning, arrest forgetting, strengthen connections to prior knowledge, enhance retention, and bolster cues and retrieval (Brown et al. 2014). Deliberate practice coupled with spaced retrieval practice, testing effect and active recall (Chan 2009; Butler 2010; McDaniel et al. 2013; Nunes and Karpicke 2015) provide a strong theoretical explanation and rationale for students' use of questions banks and flashcards, and their success. These features and framework were not highlighted in the 2004 BEME review due most likely to the differences in technology and the arrival of online platforms in MedED-COTS products as discussed below.

Many of today's MedED-COTS employ technology that can facilitate established learning methods such as adaptive rehearsal and spaced repetition (Deng et al. 2015; Rana et al. 2020). Also, profound changes in the internet, multimedia authoring, and smart phone technology allow for better and more flexible access and interactions with rich, multimedia content found in modern MedED-COTS. In the 2004 BEME review, the preparation courses and resources were predominately live lecture based as opposed to online programs in which students can control the pace and timing of their work.

### ***Implications for educational practice***

Our review confirmed the use of MedED-COTS among an extraordinarily large portion of current medical students. Four studies included in this review found over 90% of students utilized MedED-COTS (Burk-Rafel et al. 2017; Vora et al. 2013; Volk et al. 2019; Seal et al. 2020). The pervasive use of MedED-COTS, coupled with evidence for their effectiveness in facilitating recall and performance on licensing

examinations, along with a strong theoretical underpinning that explains the positive results, point to several substantive implications for current educational practice.

First, we recommend that faculty consider formally integrating the use of MedED-COTS, that include features such as Qbanks and flashcards with algorithms to monitor performance and facilitate spaced repetition, early in students' program of study (before designated test preparation study periods). In addition to helping students increase test performance, such integration may also increase trust in faculty and formal curriculum by helping students meet immediate needs (e.g. for passing licensing examinations), while subsidizing the cost of accessing the commercial resources (currently borne by students), and making study practices more time-efficient.

MedED-COTS also are a tool that could save faculty time by reducing the effort put into quiz question development and monitoring of individual student performance. Many of the recent MedED-COTS have the capability to output performance reports that could help faculty identify students' deficits or misconceptions. A program of early intervention based on these assessments could address student study habits or students at risk of not being adequately prepared for future sessions. This will allow faculty to allocate time to focus on individual students as well as direct efforts to create activities that fill gaps and address higher levels of learning such as application and synthesis.

One limitation of our study in guiding educational practice is that it focused on the use of MedED-COTS for exam preparation rather than course preparation. However, there are published research studies that reflect the use of online platforms that are integrated into medical education curricula such as Aquifer. Ninety-seven percent of US allopathic schools subscribe to Aquifer based on company data in 2021 (Aquifer Inc. 2021). This platform consists of virtual patient cases that provide students exposure to clinical conditions or populations they may not have access to in patient settings. For example, Aquifer virtual cases fill a gap in geriatric training (Sehgal et al. 2019). Lecturio, an online medical education resource, has also added a series of virtual patient cases to facilitate graduate medical education as well as enhance undergraduate medical education (Lecturio GmbH 2021). These and other examples are bound to emerge in the literature as studies of outcomes, cost-effectiveness, and efficiency are undertaken and will speak to practice recommendations for educators.

Our recommendation for integration of these resources comes with the caveat that this step requires resources and infra-structure: training on use and integration, curating materials, technical support, and budgeting. Institutions need to consider how to pay for these learning platforms and how to provide time in the curriculum for their usage. The need to curate resources, by faculty, staff, or advanced students, as well as to provide the technical support necessary to ensure access and maintenance all become practical points to consider for facilitating educational practice.

### **Gaps in studies and future research needs**

The results of our focused review revealed a number of gaps in studies that point to future needs for research on the use and integration of MedED-COTS. The most

apparent gap was the limited number of empirical studies available for answering the primary and secondary research questions. In particular, the limited number of integration studies, combined with methodological issues regarding the nature of the research questions and methods, made it impossible to isolate specific effects or associations between MedED-COTS and standardized test or course outcomes. However, the consistent positive correlations found between students use of Qbanks, coupled with widespread use, and strong theoretical explanations for the reported findings provide support for integrating MedED-COTS into medical school curricula, and highlight the need for related research.

The additional secondary research questions posed by our review are still relevant, including but not limited to: Are there instructional approaches, theories, and strategies grounded in research being used to guide the integration of MedED-COTS? Substantive research on the integration of educational resources indicate that what faculty do explicitly before, during and after introducing a new resource, such as a virtual patient simulation, is just as important to student learning as the resource itself (Barry Issenberg et al. 2005). Computer assisted instruction that were simply 'add-ons' in medical education resulted in in poor integration and suboptimal learning outcomes (Haag et al. 2007). When integrating new technology into educational contexts, we know that their use and surrounding practices need to be considered (Edelbring 2010). Studies on explicit and deliberate strategies for integrating MedED-COTS are thought to be essential for understanding and advancing students' acceptance and learning.

One limitation of our review was the scant literature on the use of MedED-COTS to prepare students for the osteopathic clinical specialty COMAT exams or USMLE Step 2 examinations, which follow the clinical training of medical students. Existing research does support the use of Qbanks in general for standardized exam preparation, so it is most likely applicable to these situations as well. Current available research also does not provide insights into the financial implications of engaging with MedED-COTS; this is an area of particular interest for future research given the already substantial costs of medical education.

The need for integration research is further highlighted by increased adoption of MedED-COTS due to constraints posed by the COVID-19 pandemic as well as the change in USMLE STEP 1 exam scoring to pass/fail. Our focused review documents the results of pre-pandemic research on the use and integration of MedED-COTS. Future research is needed to determine how both students' and faculty members' perceptions and use of MedED-COTS may change during and after the pandemic, and in light of converting STEP 1 scores to pass/fail.

Finally, we suggest that investigators consider different study designs in future research. We recognize that true experiments with randomization and tight controls may not be feasible in educational settings. However, in addition to correlational studies, we recommend the use of qualitative and design-based research methods (McKenney and Reeves 2018), along with replication (Leppink and Pérez-Fuster 2017) and longitudinal studies (Balmer and Richards 2017) to generate evidence and inform the use of MedED-COTS to advance medical education.

## Conclusions

MedED-COTS are accessible and pervasive. Initially viewed as student-driven, some schools are now explicitly showing support for these platforms by making specific ones available or recommended for their students. From the student-user standpoint, a key driver appears to be access to numerous formatted questions that mimic the structure of standardized national examinations. Students report that the questions serve to put structure to the content they are processing, fill gaps in the formal curriculum, allow self-assessment, and help identify key or high-yield learning points.

Learner use of MedED-COTS does appear to influence performance on national examinations. When considering NBME USMLE Step 1 performance, studies reviewed support the positive impact of use of these platforms, number of question items or tests taken, and the timing of said interactions. Similar patterns were noted for preparatory materials aimed at the COMLEX-USA Level 1 examination with the number of questions completed, timing and format of practice playing a role in the outcome. Some MedED-COTS generally led to positive outcomes: UWorld, Pathoma, USMLE-Rx, NBME Comprehensive Basic Science Exam, and Kaplan. Timing of the interaction with MedED-COTS may influence the outcome, with positive impacts noted through early engagement as well as immediately before national exams. While there is a paucity of data related to timing of MedED-COTS use, research on retrieval practice could be examined to further ground the importance of timing.

Given the evolving features and pervasive use by students, the potential of MedED-COTS for advancing medical education appears to be significant. However, empirical research on both their use outside and integration within the formal medical school curriculum is still limited. This review summarizes MedED-COTS research before the COVID-19 pandemic and before changes to the grading of USMLE STEP 1 exam, and seeks to set the stage for guiding future research and practice by: (1) examining the relationship between students' use of MedED-COTS and their performance on undergraduate licensing exams; and (2) exploring key variables that appear to influence student-driven use outside and formal integration of MedED-COTS within medical school curriculum.

## Disclosure statement

David Harris serves as an uncompensated basic science lead on the Aquifer Sciences Initiative, part of Aquifer (non-profit organization). The authors alone are responsible for the content and writing of the article.

### Glossary

MedED-COTS: Commercially distributed (Commercial-Off-The-Shelf) learning resources that are designed to help medical students prepare for national board exams and may also be used to facilitate Medical Education.

## Funding

The author(s) reported there is no funding associated with the work featured in this article.

## Notes on contributors

**Atsusi Hirumi**, PhD, Professor of Instructional Design and Technology, Department of Learning Sciences and Educational Research, College of Community Innovation and Education, University of Central Florida, Orlando, FL.

**Luke Horger**, MA, Instructional Design and Technology, College of Community Innovation and Education, University of Central Florida, Orlando, FL.

**David Harris**, PhD, Associate Professor of Physiology, Department of Medical Education, University of Central Florida College of Medicine, Orlando, FL.

**Andrea Berry**, MPA, Executive Director, Faculty Life Office and Instructor, Department of Medical Education, University of Central Florida College of Medicine, Orlando, FL.

**Feroza Darowalla**, MD, MPH, Associate Professor of Medicine, Department of Medical Education, University of Central Florida College of Medicine, Orlando, FL.

**Shalu Gillum**, JD, MLS, AHIP, Head of Public Services for the Harriet F. Ginsburg Health Sciences Library and Assistant Medical Librarian, Department of Medical Education, University of Central Florida College of Medicine, Orlando, FL.

**Nyla Dil**, DVM, MSc, PhD, Assistant Professor of Medicine, Department of Medical Education, University of Central Florida College of Medicine, Orlando, FL.

**Juan C. Cendan**, MD, Professor of Surgery, Interim Dean, Herbert Wertheim College of Medicine, Florida International University, Miami, FL.

## ORCID

Atsusi Hirumi  <http://orcid.org/0000-0003-1125-3399>

Andrea Berry  <http://orcid.org/0000-0002-3884-4108>

## References

- Andryrka M, Wilson-Byrne MT, Fitzpatrick S, Veitia M, Orwig R, Shuler M, Franklin D. 2014. Too smart to fail: guide for the struggling medical student. *W V Med J*. 110(6):12-4, 16-9.
- Aquifer Inc. 2021. About: advancing clinical teaching & learning. [accessed 2021 Apr 13]. <https://aquifer.org/about-aquifer/>
- Balmer DF, Richards BF. 2017. Longitudinal qualitative research in medical education. *Perspect Med Educ*. 6(5):306-310.
- Banos JH, Pepin ME, Van Wagoner N. 2018. Class-wide access to a commercial step 1 question bank during preclinical organ-based modules: a pilot project. *Acad Med*. 93(3):486-490.
- Barry Issenberg S, Mcgaghie WC, Petrusa ER, Lee Gordon D, Scales RJ. 2005. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 27(1):10-28.
- BEME Collaboration. 2012. BEME coding sheet – BEME collaboration. [accessed 2019 Jun 11]. [www.bemecollaboration.org/downloads/749/beme4\\_appx1.pdf](http://www.bemecollaboration.org/downloads/749/beme4_appx1.pdf)
- Bonasso P, Lucke-Wold B, 3rd, Reed Z, Bozek J, Cottrell S. 2015. Investigating the impact of preparation strategies on USMLE step 1 performance. *MedEdPublish*. 4(1):5.
- Bond University. 2019. Systematic review accelerator. [accessed 2019 Sept 12]. <https://sr-accelerator.com/#/polyglot>
- Brown PC, Roediger HL, III, McDaniel MA. 2014. *Make it stick: the science of successful learning*. Cambridge (MA): Belknap Press of Harvard University Press.
- Burk-Rafel J, Santen SA, Purkiss J. 2017. Study behaviors and USMLE step 1 performance: implications of a student self-directed parallel curriculum. *Acad Med*. 92(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 56th Annual Research in Medical Education Sessions): S67-S74. DOI:10.1097/ACM.0000000000001916

- Butler AC. 2010. Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Psychol Learn Mem Cogn.* 36(5):1118–1133.
- Chan JC. 2009. When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language.* 61(2): 153–170.
- Coumarbatch J, Robinson L, Thomas R, Bridge PD. 2010. Strategies for identifying students at risk for USMLE step 1 failure. *Fam Med.* 42(2):105–110.
- Deng F, Gluckstein JA, Larsen DP. 2015. Student-directed retrieval practice is a predictor of medical licensing examination performance. *Perspect Med Educ.* 4(6):308–313.
- Dixon D. 2004. Relation between variables of preadmission, medical school performance, and COMLEX–USA levels 1 and 2 performance. *J Am Osteopath Assoc.* 104(8):332–336.
- Edelbring S. 2010. A threefold framework for relating to innovations and technology in education: learning from, with and about technology. *Interprofessional E-learning and collaborative work: Practices and technologies.* Hershey (PA): IGI Global; p. 23–33.
- Ericsson KA, Krampé RT, Tesch-Römer C. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev.* 100(3):363–406.
- Giordano C, Hutchinson D, Peppler R. 2016. A predictive model for USMLE step 1 scores. *Cureus.* 8(9):e769.
- Gohara S, Shapiro JI, Jacob AN, Khuder SA, Gandy RA, Metting PJ, Gold J, Kleshinski J. 2011. Joining the conversation: predictors of success on the United States Medical Licensing Examinations (USMLE). *Learn Assist Rev.* 16(1):11–20.
- Haag M, Singer R, Bauch M, Heid J, Hess F, Leven F. 2007. Challenges and perspectives of computer-assisted instruction in medical education. *Methods Inf Med.* 46(1):67–69.
- Haight SJ, Chibnall JT, Schindler DL, Slavin SJ. 2012. Associations of medical student personality and health/wellness characteristics with their medical school performance across the curriculum. *Acad Med.* 87(4):476–485.
- Hendrix D, Hasman L. 2008. A survey of collection development for United States Medical Licensing Examination (USMLE) and National Board Dental Examination (NBDE) preparation material. *J Med Libr Assoc.* 96(3):207–216.
- Jackson F, Duane E, Harmon R, Kollar RA, Rainville NM, Smith RM. 2019. Resources that improve medical board licensing examination performance. *Cureus.* 11(10):e5927.
- Kastenmeier AS, Redlich PN, Fihn C, Treat R, Chou R, Homel A, Lewis BD. 2018. Individual learning plans foster self-directed learning skills and contribute to improved educational outcomes in the surgery clerkship. *Am J Surg.* 216(1):160–166.
- Kumar AD, Shah MK, Maley JH, Evron J, Gyftopoulos A, Miller C. 2015. Preparing to take the USMLE Step 1: a survey on medical students' self-reported study habits. *Postgrad Med J.* 91(1075):257–261.
- Lau C, Kolli V. 2017. App use in psychiatric education: a medical student survey. *Acad Psychiatry.* 41(1):68–70.
- Lecturio GmbH. 2021. Lecturio medical. [accessed 2021 Aug 31] <https://www.lecturio.com/medical/>.
- Leppink J, Pérez-Fuster P. 2017. We need more replication research - A case for test-retest reliability. *Perspect Med Educ.* 6(3):158–164.
- Maholtz DE, Erickson MJ, Cymet T. 2015. Comprehensive osteopathic medical licensing examination-USA level 1 and level 2-cognitive evaluation preparation and outcomes. *J Am Osteopath Assoc.* 115(4):232–235.
- McDaniel M, Fadler CL, Pashler H. 2013. Effects of spaced versus massed training in function learning. *J Exp Psychol Learn Mem Cogn.* 39(5):1417–1432.
- McGaghie WC, Downing SM, Kubilius R. 2004. What is the impact of commercial test preparation courses on medical examination performance? *Teach Learn Med.* 16(2):202–211.
- McKenney S, Reeves TC. 2018. *Conducting educational design research.* Abingdon: Routledge.
- National Resident Matching Program. 2018. Results of the 2018 NRMP program director survey. [accessed 2021 Apr 13]. <https://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-www.pdf>
- National Resident Matching Program. 2020. Charting outcomes in the match: senior students of US medical schools. [updated Jul 2020; accessed 2021 Aug 25]. [https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2020/07/Charting-Outcomes-in-the-Match-2020\\_MD-Senior\\_final.pdf](https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2020/07/Charting-Outcomes-in-the-Match-2020_MD-Senior_final.pdf)
- Nunes LD, Karpicke JD. 2015. Retrieval-based learning: research at the interface between cognitive science and education. In: *Emerging trends in the social and behavioral sciences.* Hoboken (NJ): Wiley; p. 1–16.
- Rana T, Laoteppitaks C, Zhang G, Troutman G, Chandra S. 2020. An investigation of anki flashcards as a study tool among first year medical students learning anatomy. *FASEB J.* 34(S1):1–1.
- ScholarRX 2020. Gulf medical university college of medicine and scholarRx announce medical education partnership. [accessed 2021 Apr 15]. <https://scholarrx.com/gulf-medical-university-college-of-medicine-and-scholarrx-announce-medical-education-partnership/>
- Schwartz LF, Lineberry M, Park YS, Kamin CS, Hyderi AA. 2018. Development and evaluation of a student-initiated test preparation program for the USMLE step 1 examination. *Teach Learn Med.* 30(2):193–201.
- Seal ZA, Koek W, Sharma R. 2020. Correlation of medical college admission test scores and self-assessment materials with the United States medical licensing examination step 1 performance. *Cureus.* 12(4):e7519.
- Sehgal M, Syed Q, Callahan KE, Powers BB, Eleazer GP, Gleason LL, Ramaswamy R, Sauvigne K, Leipzig RM, Shah A. 2019. Introducing aquifer geriatrics, the American geriatrics society national online curriculum. *J Am Geriatr Soc.* 67(4):811–817.
- Taylor JA, Shaw CM, Tan SA, Falcone JL. 2018. Are the kids alright? Review books and the internet as the most common study resources for the general surgery clerkship. *Am J Surg.* 215(1):191–195.
- Volk AS, Rhudy AK, Marturano MN, Ott L, DuCoin C. 2019. Best study strategy for the NBME clinical science surgery exam. *J Surg Educ.* 76(6):1539–1545.
- Vora A, Maltezos N, Alfonso L, Hernandez N, Calix E, Fernandez MI. 2013. Predictors of scoring at least 600 on COMLEX-USA level 1: successful preparation strategies. *J Am Osteopath Assoc.* 113(2): 164–173.
- Yavner SD. 2016. *Stress, fatigue, and medical students' study resource selection: implications for the design of educational multimedia.* New York (NY): New York University.
- Zhang C, Rauchwarger A, Toth C, O'Connell M. 2005. Student USMLE step 1 preparation and performance. *Adv Health Sci Educ Theory Pract.* 9(4):291–297.