Information Management Tools for Clinician Education and Practice in the Data Era

by

Emily Ann Balczewski

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in the University of Michigan
2025

Doctoral Committee:

Professor Jason S. Knight, Co-Chair
Adjunct Professor Karandeep Singh, Co-Chair
Assistant Professor Michael R. Mathis
Associate Professor Arvind U.K. Rao
Associate Professor Xu Shi
Assistant Professor Alexandra H. Vinson

Emily A. Balczewski

ebalczew@umich.edu

ORCID iD: 0000-0001-5495-9637

© Emily A. Balczewski 2025

Dedication

To my husband, Ben,

my peerless peers, Philip Barrison and Erkin Ötleş,

and the best MSTP cohort a physician scientist trainee could ask for.

Acknowledgements

None of this would have been possible without my team of research mentors. Dr. Karandeep Singh, you helped me find my feet after a rocky start and kept me constantly buoyed throughout my PhD with your enthusiasm and expertise for all things AI. Dr. Jason Knight, thank you for opening the world of rheumatology to me. Your reliability and email-based humor were important safeguards when I was uncomfortably afloat. Dr. Alexandra Vinson, much gratitude for helping Phil and I spin unbounded enthusiasm into actual expertise. The electronic flashcard project allowed me to develop skills I didn't anticipate having the opportunity to learn during my PhD but have bolstered my personal and professional toolbox significantly. Dr. Michael Mathis, I know anesthesiologists are known for balancing a lot of variables in high-pressure situations, but I nonetheless appreciate you making abundant space in your busy schedule to help me stick my PhD landing. Finally, thank you to Drs. Arvind Rao and Xu Shi for your open-minded support of the variety of methods and results that made their way into this thesis.

Two of the most important mentors during my PhD have, in fact, been my fellow learners. Phil Barrison, this thesis would not have been possible without our fireside chat and the subsequent 2+ years of hard work on which we embarked together. Dr. Erkin Ötleş, you gave me a leg up when I really needed it—thank you for helping me see my own value when I doubted it.

To the MSTP admin team – Kathy, Ron, Debra, Justine, Liz, Claire, and Gretchen – for guidance, support, and, most importantly, free and nutritious meals. To Margit Burmeister and

the DCMB team, thank you for creating a multidisciplinary department that allowed me to meet my own unique needs as a learner.

To my medical school friends—Tom, Annelie, Hannah, Erin, and many others—you made learning to be a doctor during a pandemic much less awful than it sounds. Even though you've all embarked on the next phase of your journey, you've never made me feel left behind. To my other pandemic pod members – Fred and Mindy Herrmann – thank you for giving me back my home. To my medical school mentor, Dr. David Stewart, thank you for being a true educator and providing a path for me to work through the most vulnerable sides of myself. To the best MSTP cohort™ – Hannah, Crys, Clare, Karena, Curtis, Caleb, Ben, Aaron, Jeff, and spiritual affiliates Kallen and Cyrus – wow what a ride it has been. May Journal-less Journal Club never die, even as we part ways on our individual physician scientist journeys.

To the Balczewskis, Kleises, and McMillans, thank you for forming the ever-present backdrop of family. To my husband, Ben: you complete the other side of my coin. And, as we both know, sometimes I very much need the reminder that I have other, non-academic sides. The loving, joyful, and nourishing partnership you provide makes us capable of so many incredible things, home remodels and this thesis included.

For their valuable contributions to Chapters II and III of this thesis, I thank Philip Barrison, Alexandra Vinson, Zach Landis-Lewis, Asavari Rajpurkar, Kirsten Fiestan, Pratik Vadlamudi, Matthew Sparling, and Idris Nagar. For their valuable contributions to Chapter IV of this thesis, I thank Graciela Mentz, Michael Mathis, and Karandeep Singh.

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables ix

List of Figures x

Abstract xi

Chapter 1 Introduction 1

1.1 The Data Era 1

1.2 Medical Education in the Data Era 2

1.3 Clinical Practice in the Data Era 5

1.4 Dissertation Roadmap 6

1.5 Figures 8

Chapter 2 Unboxing Electronic Flashcards: An Introduction to The Design Elements of Anki for Medical Education (DEAME) Framework 11

2.1 Abstract 11

2.2 Introduction 13

2.3 Background 14

2.3.1 What are electronic flashcards? 14

2.3.2 What is The New Taxonomy of Educational Objectives? 16

2.4 Method 17

2.4.1 Phase 1 - Initial Framework 18

2.4.2 Phase 2 - Code Team Expansion 18

2.4.3 Phase 3 - Refined Framework 19

2.4.4 Phase 4 - Final Framework 20
2.5 Results 20
2.5.1 Framework Description 21
2.5.2 Framework Challenges 27
2.6 Discussion 29
2.7 Figures and Tables 33
2.8 Appendices 47
2.8.1 Appendix A. Methods for Card Organization. 47
2.8.2 Appendix B. Codebook. 47
2.8.3 Appendix C. Codebook Examples. 48
2.8.4 Appendix D. Preliminary definitions for unseen codes. 48
Chapter 3 What is in the Black Box of Learner-Created Electronic Flashcards? 50
3.1 Abstract 50
3.2 Introduction 52
3.3 Method 54
3.3.1 Sampling 54
3.3.2 Framework Development 54
3.3.3 Framework Quantification 55
3.4 Results 55
3.4.1 Content Area 56
3.4.2 Clinical Reasoning Domain 57
3.4.3 Knowledge Domain 58
3.4.4 Cognitive Process 58
3.5 Discussion 59
3.6 Tables 62

Chapter 4 Feasibility of Estimating Cardiac Indices Using Cardiac Surgery Anesthesia Records in a Multicenter Cohort 67

4.1 Abstract 67

4.2 Introduction 69

4.3 Method 70

4.3.1 Study Design 70

4.3.2 Data Source 70

4.3.3 Study Population 71

4.3.4 Outcome Variable - Cardiac Index 71

4.3.5 Exposure Variable - CI Estimators 72

4.3.6 Risk Adjustment Variables - Patient, Case, and Intraoperative Factors 72

4.3.7 Statistical Analyses 73

4.4 Results 74

4.5 Discussion 75

4.6 Conclusion 77

4.7 Figures 78

4.8 Tables 81

4.9 Appendices 88

4.9.1 Appendix E. Full cardiac index estimation model results. 88

Chapter 5 Conclusion 89

5.1 Overview 89

5.2 Summary of Key Findings 89

5.3 EFs test a variety of knowledge in a variety of ways, including ways which may support knowledge acquisition at scale 91

5.4 EFs can test clinical and complex knowledge, but may need expert guidance and data management tools for EF curation to adequately encourage skill development. 93

5.5 Noninvasive cardiac index estimation shows initial promise, but needs future work to provide value at the patient level. 95

5.6 Final Thoughts 96

Bibliography 98

List of Tables

Table 2.1 Study deck information. 37

Table 2.2 Overview of the complete Design Elements of Anki for Medical Education (DEAME) Framework. 39

Table 2.3 *Note Components* definitions and examples. 40

Table 2.4 *Question Format* question types, encodings, and examples. 41

Table 2.5 *Answer Format* definitions and examples. 42

Table 2.6 *Knowledge Domain* definitions and examples. 43

Table 2.7 *Cognitive Process* definitions and examples. 45

Table 2.8 *Card Organization* definitions and examples. 46

Table 3.1 Study deck information. 62

Table 3.2. Quantification of content areas in electronic flashcards (EFs). 65

Table 3.3. Quantification of characteristics of electronic flashcards (EFs). 66

Table 4.1. Evaluated estimators of cardiac index. 81

Table 4.2. Number of cases and primary anesthesia attendings by study hospital. 82

Table 4.3. Pre- and post-CPB cohort characteristics. 83

Table 4.4. Correlations between estimators. 84

Table 4.5. Bland-Altman and error analysis. 85

Table 4.6. Agreement of true CI and estimated CI for low CI <2.2 L/min/m2. 86

Table 4.7. Confusion matrices for low cardiac index (CI) classification. 87

List of Figures

Figure 1.1. Examples of data sources driving discovery in the data era. 8

Figure 1.2. Example electronic flashcards (EFs). 9

Figure 1.3. Diagram of cardiac output (CO) physiology. 10

Figure 2.1 Study overview. 33

Figure 2.2 Simplified schematic of an Anki flashcard. 34

Figure 2.3 Example framework application. 35

Figure 2.4 Cueing schematic. 36

Figure 4.1. Bland-Altman plots for the Liljestrand and Zander estimator pre- and post-CPB. 78

Figure 4.2. Bland-Altman plots for remaining estimators pre- and post-CPB. 80

Abstract

Recent advancements in computing technology have transformed the field of biomedicine. Vast repositories of research and healthcare data have proliferated alongside tools to analyze them. Consequently, the role of healthcare providers has evolved to include not only direct patient care, but also the management and effective integration of data from diverse sources. To support this evolving role of physicians, significant changes in medical education to enable acquisition of knowledge at scale and clinical care to enable management of data streams have become imperative. This thesis comprises two projects which explore cutting-edge information management strategies for medical education and clinical practice.

In the first project of this thesis, we explore the design features of electronic flashcards (EFs), a technology-enhanced learning tool which shows initial promise for managing high-volume knowledge acquisition. We leverage a flexible approach to thematic analysis to identify and define the design features in EFs for undergraduate medical education (UME). The result of this qualitative study, called the Design Elements of Anki for Medical Education (DEAME) Framework, provides the first comprehensive terminology to unite the heterogeneous design features that EFs may display. This framework therefore shows that individual EFs for UME can encourage active learning of a variety of knowledge in a variety of ways, including in ways which may support acquisition of knowledge at scale such as rich metadata to organize and connect related flashcards. Furthermore, we apply this framework to a sample of EFs for UME to interrogate whether learner-created EFs might support development of core curricular

competencies such as clinical reasoning and higher-order thinking skills which are necessary for clinical practice in the data era. We find that all learner-created EF decks contain subsets of flashcards which relate to and might improve these competencies but may not do so in sufficient quantity to align with learners' and educators' goals. Our framework provides a strong foundation for future efforts to align EFs with curriculum and ensure learner engagement with such competencies.

In the second project of this thesis, we investigate the feasibility of using streams of readily available data to estimate a measure of cardiac function–cardiac index (CI)–with significant importance to clinical decision making for cardiac anesthesiologists. Using generalized linear mixed-effects models, we show that several estimators which collate noninvasive physiological measurements are independently associated with CI when adjusted for patient and procedure factors. Therefore, these estimators show future promise in predicting CI without requiring invasive measurements, as is the current gold-standard. However, we also show that our current models are poorly calibrated for individuals, and not yet sufficiently accurate to guide clinical decision making. Nonetheless, our work demonstrates advancement in the ability to help clinicians manage critical patient data with minimal added risk and with potential value for clinical practice.

These two projects use a mix of quantitative and qualitative methods to investigate state-of-the-art strategies for physicians to learn and manage vast amounts of information. While EFs and CI estimators show significant promise in supporting physicians in their role as information managers, future work is necessary to allow these strategies to achieve their full value to medical education and clinical practice.

Chapter 1 Introduction

1.1 The Data Era

In the past few decades, advances in computing technology have allowed the collection and storage of vast amounts of data.[1] In the field of biomedicine, large repositories of research and health operations data have surfaced an unprecedented amount of information about the human body (Figure 1).[2,3] New and advanced analytic methods, such as machine learning and generative artificial intelligence, have arisen to facilitate the analysis of these large, heterogeneous, and complex collections of biomedical data.[2–4] These two transformative shifts– the accumulation of data and tools to analyze it–have yielded compensatory changes in the role of healthcare providers. In addition to providing one-on-one, humanistic care to their patients, physicians must now manage large amounts of data from heterogeneous sources.[5] This management entails not only the production of new data, such as through coding and documentation of healthcare encounters, but also the "effective integration and utilization of information" to facilitate data-driven patient care.[6]

To support physicians in their new and developing role as "information managers," current approaches to education and clinical care must evolve.[5,6] Education must leverage technology to provide students with learning strategies that allow mastery of ever-increasing amounts of new biomedical discoveries.[5] Additionally, education must sharpen critical thinking and decision making skills which allow physicians to gather, appraise, and act upon data from heterogeneous sources.[5,6] In clinical practice, tools must be developed to surface timely,

actionable, and valuable information to physicians to encourage its effective integration in patient care.[7] The following sections elaborate on recent trends in medical education and clinical care which have arisen in the data era and might support the transforming role of physicians today.

1.2 Medical Education in the Data Era

Annually, over 1.5 million new scientific papers are published in biomedicine and the life sciences, representing a vast amount of novel information for physicians to assimilate.[8] This knowledge expansion has accompanied two recent shifts in medical education: the rise of eLearning and "parallel curriculum" resources among medical learners.[9–11] eLearning resources are technology-enhanced learning tools such as electronic question banks, instructional videos, podcasts, and social media.[9] The parallel curriculum includes third-party resources that exist outside of traditional curricula—and are often eLearning resources themselves.[10,11] Students report taking up such resources to facilitate learning and retention of the ever increasing amount of knowledge required by medical practice and certification exams.[9]

An archetypal example of both an eLearning and parallel curriculum resource, electronic flashcards (EFs), may be uniquely poised to facilitate knowledge acquisition at scale (Figure 1.2). For one, EFs may encourage learners to engage with established, evidence-based strategies for knowledge acquisition and retention, such as active learning, interleaving, and spaced repetition.[12–14] However, unlike other resources which may encourage similar behaviors, many collections of EFs (called "decks") are themselves very large in scale. For example, a commonly used EF deck, AnKing Step 1, has 31,475 flashcards–an order of magnitude greater than the UWorld Step 1 question bank with 3,643 questions as of April 2024.[15,16] In fact, a subset of EFs in the AnKing Step 1 deck are extracted from UWorld questions themselves.[17] However,

whereas a UWorld board-style, practice question may require users to understand and reason with many pieces of information at one time, EFs may atomize knowledge into smaller, component pieces of information. Schukow et al suggest that this breaking down of "big concepts into smaller ones" may support EF's ability to address "volume-of-information constraints" among medical learners.[18]

Indeed, preliminary evidence around EFs suggests that they generally improve measures of knowledge acquisition, including on large, summative exams such as Step 1.[19] Furthermore, several studies show a positive correlation between specific measures of EF use, like number of cards reviewed, and academic performance, suggesting a dose-dependent relationship.[20–27] Nonetheless, given that EF decks are sizable, learners may need to review large numbers of cards in order to improve their academic performance. For example, one study by Deng et al reports that reviewing 1,700 new flashcards correlated to a 1-point increase on the Step 1 exam.[20] Indeed, some learners report that the size of EF decks is a barrier to their use—a barrier that is ameliorated when learners perceive EF decks to contain high-quality questions and align strongly with curricula.[19,28,29] Therefore, for EFs to best support learners in facilitating knowledge acquisition at scale, EFs must prioritize efficiency by testing knowledge in the "right" way (e.g., with high-quality questions) and prioritizing the "right" knowledge for testing (e.g., curriculum-aligned knowledge).

However, few studies have explored how EFs actually present knowledge, including how EFs might atomize large concepts into smaller ones, test knowledge effectively, or align with curricula.[19] Indeed, many studies have no description whatsoever regarding the content or design of their EF decks.[19] When studies do describe the contents of their EF decks, there is evident heterogeneity in EF design between studies, suggesting that EFs may not employ a "one-size-

fits-all" approach to testing knowledge. For example, studies report highly different mixes of question types in EF decks, such as multiple choice, open-ended, and fill-in-the-blank.[19] No studies, to our knowledge, explicitly or empirically compare the effect of different design choices on learner behaviors and outcomes, such as time required to review individual flashcards. In this way, we have little insight into testing knowledge the "right" way to efficiently support acquisition of knowledge at scale. A key hurdle to such empirical evaluations is the lack of terminology to describe the variety of choices available for EF design. In Chapter II of this dissertation, we propose a framework which provides a terminology for EF design and may support future exploration of high-quality EFs.

In addition to creating cards that test knowledge the "right" way, EFs must prioritize the "right" knowledge for learning. Given that EFs are often reported in studies to be largely created by-learners-for-learners, they may not adequately incorporate expert perspectives in what knowledge is important to learn.[13,30,31] In fact, a key concern surrounding EFs, and other eLearning and parallel curriculum study tools, is that they primarily facilitate knowledge acquisition without actually improving the core competencies required for medical practice.[32] As students spend increasingly more time using these resources compared to expert-created curricular resources designed to facilitate competencies, this concern becomes particularly immediate.[29,33] One such competency relevant to the data era includes sharpening critical thinking and decision making skills which allow physicians to integrate information from heterogeneous sources.[5,6] If EFs could be shown to facilitate such competencies, they could have enormous promise as a learning tool in the data era, not just for knowledge acquisition, but also competency development. In Chapter III of this thesis, we assess the degree to which popular, learner-created EF decks engage with core competencies like clinical reasoning skills.

EFs have shown initial promise to support learner education in the data era. EFs are a technology-based active learning tool that have been shown to improve academic performance in the current landscape of high demands for knowledge acquisition. However, to realize their full potential in supporting students, EFs must present information to learners in an efficient manner and target learning needs beyond knowledge acquisition alone.

1.3 Clinical Practice in the Data Era

In addition to education facilitating the requisite knowledge and reasoning skills to engage in data-driven practice, clinicians may benefit from tools integrated in the clinical space. Many such recent examples exist, such as clinical decision support (CDS) tools which typically use EHR data and advanced modeling techniques to suggest actions to clinicians.[34] The recent rise in generative AI has opened another wide suite of tools, such as documentation support, which can summarize information from clinical encounters.[35] However, while there is an increasing amount and variety of such tools, a key concern around data-driven tools in the clinical space is their actual value to patient care and outcomes.[34]

The field of anesthesiology offers an interesting case study for how the data era has changed clinical practice. Anesthesiologists must monitor many streams of information about their patients during surgery in order to support their vital functioning.[36] In the data era, anesthesiologists have adopted dashboards to collate and visualize many streams of intraoperative data which may also include embedded CDS tools.[37,38] However, these dashboards may have uncertain value. For example, the ACTFAST-3 Randomized Clinical Trial demonstrated that a dashboard intervention did not actually prevent postoperative complications like hypoglycemia and hypothermia in a general surgery cohort.[39] One limitation of these dashboards may be the exclusion of relevant information.[38] For example, in cardiac surgery,

morbidity and mortality commonly stems from poor cardiac function during and after surgery, which causes low oxygen delivery to organs and organ damage.[40] Therefore, it can be presumed that monitoring cardiac function could allow clinicians early intervention in preventing and managing this low oxygen state.[41,42] However, an important measure of cardiac function is cardiac index (CI) or cardiac output (CO) which are measured through placement of an invasive pulmonary artery catheter (PAC) (Figure 1.3). Therefore, the potential complications that may result from placing a PAC must be weighed against the value that monitoring of CI/CO can provide.[43] Alternatively, the value proposition of CI/CO monitoring could be changed by allowing monitoring of CI/CO through noninvasive proxies of CI instead of direct measurement with invasive PACs. In Chapter IV, we explore the ability of noninvasive physiological data to estimate CI, while controlling for other patient and procedural factors which may impact CI measurement and are readily available in the anesthesia record.

In the data era, information management strategies like CDS tools and dashboards have the potential to support clinicians to integrate into clinical practice vast streams of information. However, these strategies may need to balance the value of surfacing information to providers with the risk of gathering that information from patients in order to demonstrate value in improving patient outcomes.

1.4 Dissertation Roadmap

This dissertation comprises two projects which present novel insights into "information management" in the contexts of medical education and cardiac anesthesiology practice. In Chapters II and III, we present mixed methods research into the design of electronic flashcards for medical education. Chapter II describes a novel approach to thematic analysis applied to >1,300 electronic flashcards for undergraduate medical education. The result of this qualitative

study is the Design Elements of Anki for Medical Education (DEAME) Framework that details seven thematic categories which group 48 different design features found in the EFs for medical education. Two of the seven thematic categories are directly derived from the *New Taxonomy of Educational Objectives* by Marzano and Kendall and provide a newly robust connection between electronic flashcards and theories of learning to describe how EFs might test knowledge.[44] Chapter III extends the qualitative work in Chapter II and provides a quantification of the types of knowledge and ways of testing it in a sample of 1,147 EFs. This quantification provides novel insight into the content and construction of highly used EF decks for medical education, including the relevance of EFs to the development of clinical reasoning and higher-order thinking skills. Chapter IV details the second project in this thesis in the context of cardiac anesthesiology clinical care. Using a large, multicenter database of cardiac surgeries, we explore the feasibility of using noninvasive data in the anesthesia record to estimate invasive measurements of cardiac index, an important prognostic measure of cardiac function. Chapter V discusses implications of the key findings from Chapters II-IV and motivates future applications and directions of these projects to support clinical education and practice in the data era.

1.5 Figures

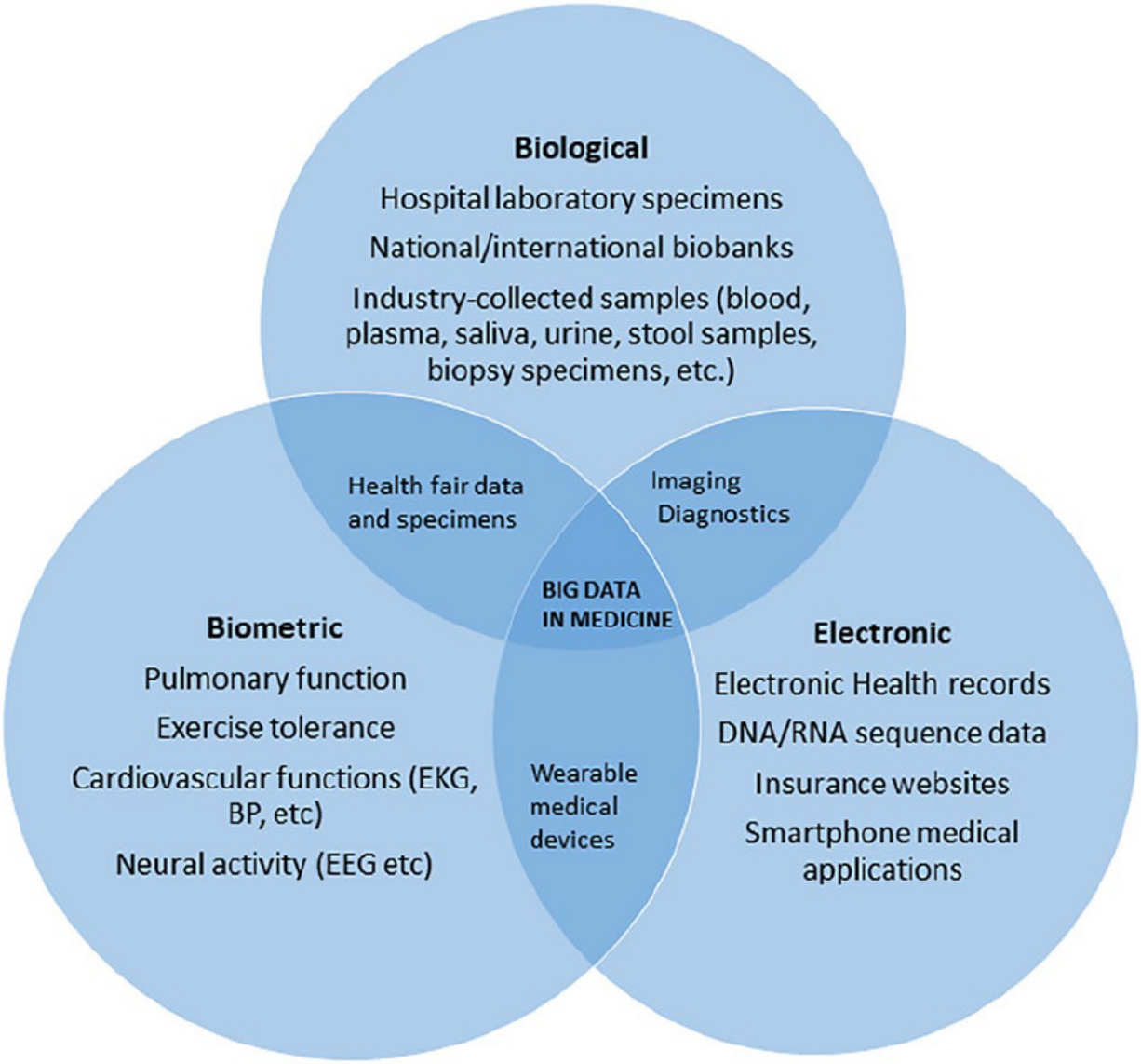


Figure 1.1. Examples of data sources driving discovery in the data era. Adapted from Mallappallil et al.[3]

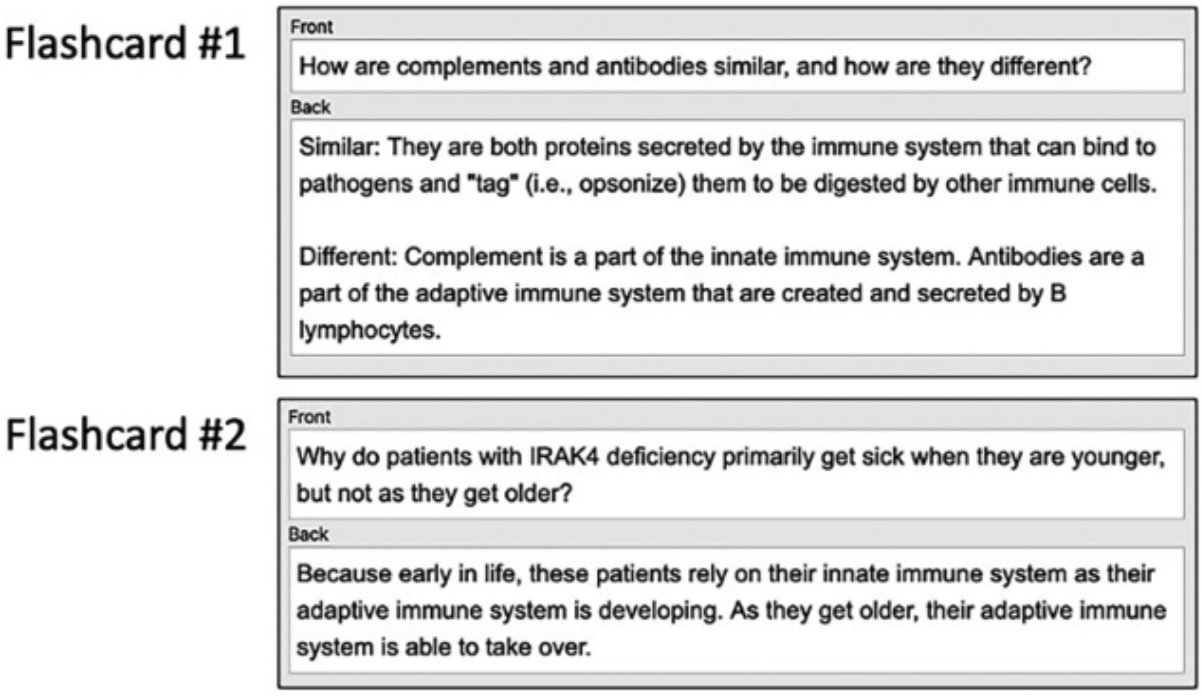


Figure 1.2. Example electronic flashcards (EFs). EFs contain questions on the front and answers on the back, adapted from Spicer et al.[13]

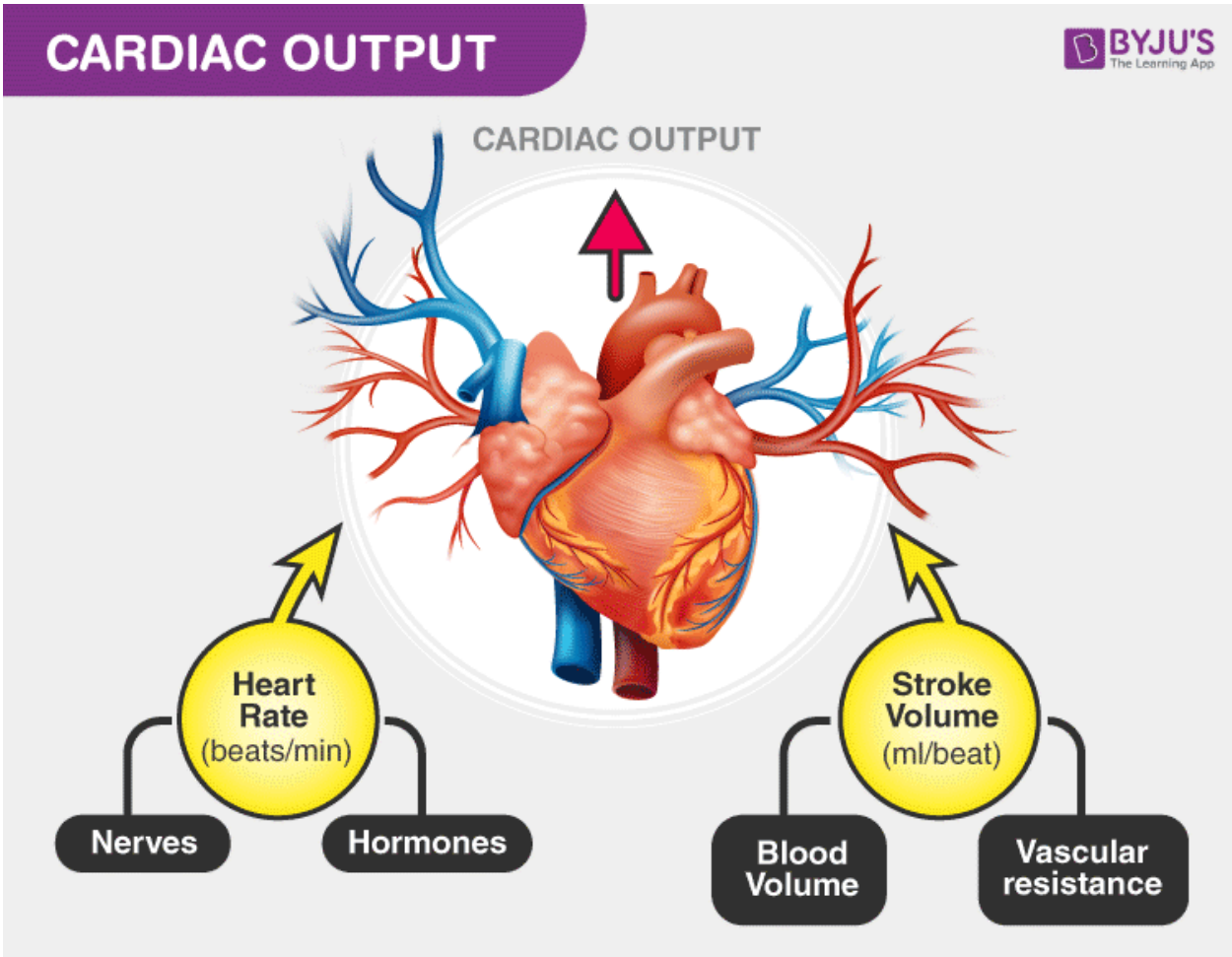


Figure 1.3. Diagram of cardiac output (CO) physiology. CO is the product of heart rate and stroke volume, which are each affected by different physiological factors. Cardiac index (CI), a related metric for cardiac function, is CO normalized by body surface area. Adapted from BYJU's The Learning App.[45]

Chapter 2 Unboxing Electronic Flashcards: An Introduction to The Design Elements of Anki for Medical Education (DEAME) Framework

2.1 Abstract

Purpose: Electronic flashcards (EFs) have rapidly ascended as an active learning resource among medical learners. Despite their significant popularity, no guidelines exist to support educators in creating high-quality, curriculum-aligned EFs for their learners. The development of such guidelines is significantly hampered by the lack of a terminology to unite the significant design heterogeneity that EFs may display. In this study, we propose a framework that offers a robust terminology for describing EF design. The framework incorporates an extension of *The New Taxonomy of Educational Objectives* (*NT*) to describe how EFs present and test knowledge– and, therefore, might align with curricular objectives.

Method: We sampled more than 1300 flashcards across six popular EF decks for undergraduate medical education (UME). Our multi-phase study used a flexible approach to thematic analysis, which involved iterative cycles of induction, deduction, and reconciliation to identify general design dimensions ("categories") and specific design elements ("codes") of EFs. These categories and codes are organized into a unifying framework, and we present definitions, examples, and possible implications for student-centered design for each.

Results: The Design Elements of Anki for Medical Education (DEAME) Framework contains seven categories and 48 codes that provide a robust foundation for conceptualizing EF design.

Two categories describe the multimedia and metadata elements in EFs. Three categories describe how EFs formulate questions and answers. Two categories extend the *NT* to describe the types of knowledge and ways of thinking likely elicited by EFs.

Conclusions: The DEAME Framework provides a much-needed terminology to describe the significant design variety in EFs for medical education. This terminology, which includes a novel application of the *NT* to categorize the knowledge representation in EFs, enables future development of best-practices for creating high-quality EFs that align with medical curricula.

2.2 Introduction

Electronic flashcards (EFs) have become a widely popular study tool among medical learners in recent years, with some studies showing adoption rates in undergraduate medical education (UME) cohorts as high as 85%,[46] rivaling the use of traditional curricular resources like lectures.[47] As EFs continue gaining traction in medical education, educational stakeholders like faculty, researchers, and administrators–who may be unfamiliar with EFs–must increasingly engage with them to support learners.

To support this critical engagement of educational stakeholders in shaping the future of EFs as a learning tool, the fundamentals of EF design need to be established. However, unlike long-standing assessment methods like multiple-choice questions,[48] EFs have no consensus terminology to describe how they organize and present knowledge.[19] Given that the nascent EF literature suggests EFs may display a wide variety of design elements–including heterogeneous question types,[21,22,28,49–56] question elements,[50,52,54–60] explanatory information,[52,61] and metadata[59]–a unifying terminology is a necessary first step to provide learners and educational stakeholders a common conceptualization of EFs. This mutual understanding may facilitate exploration into important questions around EFs to define their future role in medical education.

A question central to the concerns of both learners and educational stakeholders is: How can EFs support specific populations of learners to achieve curricular or other learning objectives? This question is particularly salient given the context in which EFs rose to popularity. EFs are a characteristic part of the "parallel curriculum," which includes third-party resources created and used outside of school-sanctioned curricula.[10,11] However, unlike many other "parallel curriculum" resources such as the UWorld question banks curated by content experts, EFs are often created by-learners-for-learners and then digitally distributed across the

globe.[62–64] As a result, the content and design of EFs may be subject to a wide degree of variability that goes unchecked by experts.[30] A handful of prior studies have used experts to align EFs with specific curricular or learning objectives,[57,65,66] but none provide explicit guidance on performing this alignment. Connections between EFs and taxonomies of learning could support the development of formal guidelines to align EFs with learning objectives. Several previous studies hint that such a connection is possible, suggesting that EFs can elicit second- or higher-order learning processes[28,52–54,67,68] and be described by taxonomies of learning like *Bloom's*[28].

In this study, we used a flexible, inductive-deductive approach to thematic analysis to create the Design Elements of Anki for Medical Education (DEAME) Framework. DEAME's main goal is to provide a terminology for the key design elements of EFs. Additionally, the framework incorporates a novel extension of *The New Taxonomy of Educational Objectives* (*NT*)[69]–an update to *Bloom's Taxonomy*–to capture the types of knowledge in EFs and how they are tested. Together, these contributions may provide learners and educational stakeholders with a shared, robust understanding of EF design fundamentals and motivate future efforts to explore student-centered and curriculum-aligned EF design.

2.3 Background

2.3.1 What are electronic flashcards?

Electronic flashcards—also called e-flashcards, digital flashcards, virtual flashcards, and interactive flashcards—facilitate active engagement in learning. While there is no consensus definition of EFs, they typically have a question ("front") and answer ("back") component like traditional paper flashcards. The electronic nature of EFs allows for additional capabilities, such as multimedia, analytics, gamification, and spaced repetition. These electronic features are facilitated by EF software platforms, such as the free-and-open-source Anki application.[70]

Additionally, many online communities, such as Reddit's /r/medschoolanki/[71] and the paid service AnkiHub,[16] provide spaces to share EF decks and circulate advice about EF use.

Research has yielded preliminary insights into learner behaviors, attitudes, and outcomes around EFs for the health professions, including medical education.[19] Learners usually demonstrate high uptake (>50% of the study population) of EFs when offered.[21,22,49,51,55–57,62,67,72–77] However, they may use EFs in differing ways, such as for massed[21,22,61,74] vs. spaced/distributed practice[28,51,78,79] and learning new knowledge[54,67,72,80] vs. rehearsing old knowledge.[25,28,78] The attitudes towards EFs among learners are largely positive, with a handful of studies suggesting EFs are perceived as helpful,[24,25,30,51,52,54,56,57,62,67,77,78,80–83] easy-to-use,[24,52,59,62,67,80] and confidence-boosting.[56,62,76,80,82,84,85] When prospectively assessing the impact of EF use on academic outcomes, most studies[11,21–23,25,51,53,60,75,81,83,85]–including all three with a randomized, controlled design[59,66,84]–show improvements in exam scores or clinical performance among EF learners. A dose-dependent effect of EF use is suggested by recent studies, which generally found positive and significant correlations between academic outcomes and specific usage metrics–such as time spent using EFs or the number of new cards reviewed.[20–27]

Nevertheless, the interpretation, extensibility, and replicability of the growing EF literature remains challenging because many studies do not report the details of the design, content, and scope of EFs used by learners. Just as an educational intervention involving ten multiple-choice questions may not have similar impacts to 30 vignette-based short answer questions, different collections of EFs may yield different impacts on learners. Clear reporting of EF-based educational interventions using shared terminologies, such as the framework proposed

in this manuscript, could clarify design best-practices for different populations of learners or curricular objectives.

2.3.2 What is The New Taxonomy of Educational Objectives?

Bloom's Taxonomy (*BT*) is often the de facto choice to characterize the types of thinking elicited by learning objectives and assessment items. However, a known limitation of *BT* is that its six levels contain a mix of types of knowledge and ways of thinking with that knowledge.[86] To address this limitation and others, several updates to *Bloom's* have been proposed, including *The New Taxonomy of Educational Objectives* (*NT*)[69] which contains two different domains that separately describe knowledge types and cognitive processes. EFs themselves may elicit different types of knowledge[87] and thinking skills[28,52–54,67,68] and, therefore, benefit from a description by the *NT*'s two-dimensional structure.

Marzano and Kendall suggest that "assessments are the logical consequence of well-articulated objectives" and furnish abundant examples converting learning objectives to example questions.[44,69] In one example of this, Dubas and Toledo created assessments from *NT*-based learning objectives for undergraduate chemistry and economics courses; in the process of assessment creation, they successfully "mapped each question/exercise/problem … to the corresponding [type of knowledge] and cognitive level" in the *NT*.[88,89] This suggests that, with some understanding of the context in which an assessment is developed and deployed, assessment items like EFs could be categorized by the *NT*. While we were unable to identify any studies that applied the *NT* fully *post hoc* to an assessment to draw conclusions about the types of learning objectives addressed by that assessment, we propose such a reversal is possible. In this manuscript, we explore the capacity of the *NT* to be adapted for this purpose, in the hope that it could support future work into the alignment of EFs with medical education curricula.

2.4 Method

To develop a framework that comprehensively and clearly formalizes the key design features in individual EFs, we assembled a team of experts in qualitative research, education, and medical education. This team incorporated inductive category generation, inductive and deductive code generation, and codebook reconciliation in a four-phase study design (Figure 2.1). Our approach was guided by the principles of thematic analysis, which Braun & Clarke define as "searching across a data set to find repeated patterns of meaning."[90] This effort resulted in a framework built upon a codebook that defines (i) categories that thematically group sets of (ii) codes that detail specific design features of individual EFs. During the development of the framework, we prioritized comprehensive category development over individual codes so that categories could be inclusive of code extensions dependent on the wide degree of variability in EF decks and platforms.

To develop the framework, we chose six Anki decks based on their routine use in UME; five were recommended by learners on Reddit, one is commonly used at our institution, and several have appeared in studies of EF use in UME.[91,92] The decks vary in structure, size, origin, and targeted learner (Table 2.1). We randomly sampled *notes* from each deck. A *note* is the Anki terminology for the base unit of card creation, and one note may give rise to multiple flashcards (Figure 2.2). We analyzed code applications using the Dedoose qualitative coding software and the R programming language.[93,94] Because one of our categories, *Card Organization*, only differed substantially between decks and did not benefit from continuous cycles of evaluating individual cards, we used only a single round of induction and deduction to develop it as described in Appendix A.

Our study was determined exempt from ongoing review by the University of Michigan IRBMED (HUM00222432).

2.4.1 Phase 1 - Initial Framework

Two researchers (EB and PB) engaged in several rounds of open coding on proportional amounts of notes from each of the six study decks, totaling 290 notes. After a completed round of coding, we reconciled code applications and memos, exploring overlap and incongruity in the meaning of categories and codes. We explored several theoretical frameworks for knowledge representation, including *Bloom's* and the Alavi and Leidner[95] conceptualization of knowledge, and settled on the *NT* as the best fit for our study aims and observations from open coding. After creating a draft codebook with inductively generated categories and inductively and deductively generated codes, we completed a fully deductive round of coding on 100 notes from a single deck (UMMS Preclinical). Discussion after this deductive round yielded 1) no new categories and 2) no consensus on changes to category and code definitions, suggesting a need to expand the coding team to encompass more perspectives. Therefore, we trained new coders in Phase 2.

2.4.2 Phase 2 - Code Team Expansion

To challenge the codes and categories in the developing framework, we assembled a team of additional coders. Four medical students (MS, AR, KF, PV) with pre-clinical and clinical training and experience with the Anki platform were selected to enrich codebook development. The initial training process consisted of two 2-hour live sessions that provided definitions, examples, active learning questions, and open discussion of codebook categories and codes. Training sessions were recorded, and training materials–including training session slide decks and primary sources for deductive codes–were deposited in a shared drive for reference. After initial training, coders participated in five training sets with 5-50 notes to code, totaling 145

notes; office hours to discuss the coding process or specific codes were offered by EB and PB once weekly. Additionally, each training set was followed by verbal and/or written feedback from EB and PB. Category and code definitions were adjusted throughout the training by incorporating coder feedback from open discussion, deliberation, and memoing. Ongoing code application fidelity was monitored by calculating ICR coefficients (Fleiss' Kappa for more than two coders and Cohen's Kappa for two coders).[96] After thorough training and team discussions in Phase 2, we established that discrepancies in code application were due to weaknesses in the code definitions, not due to coders misapplying codes per se. Therefore, we proceeded to Phase 3 to explore these weaknesses and produce a refined framework.

2.4.3 Phase 3 - Refined Framework

Following the successful onboarding of new coders, we split our full coding team into two small groups (group 1: EB, PB, AR, KF; group 2: EB, PB, MS, PV) to encourage active discussion around weak areas of the codebook. Each group had regular reconciliation meetings on a set of 30-50 notes, for a total of 290 notes in Phase 3. Discussions were guided by coder feedback on coding challenges, ICR calculations suggesting disagreement and manual inspection of note-level coding disagreements.

We continued coding rounds and discussions until we agreed that we had reached theoretical sufficiency for categories and meaning saturation for codes in the framework [97,98]. We assessed the theoretical sufficiency of framework categories through small and large group discussions with the study team. We assessed meaning saturation with small and large group discussions, logs of codebook changes, and memoing. When coders agreed that a) no new categories should be added, and b) the existing codes were as reasonable as possible given 1) the challenges of sampling such large decks for sometimes rare phenomena, and 2) the limitations of

studying, e.g., thought processes with this methodology, we completed one final round of coding to assess agreement and proceeded to phase 4.

2.4.4 Phase 4 - Final Framework

All coders read and commented on the final codebook and code examples document. When all comments were addressed, coders certified that they agreed with the contents of the codebook and felt as if it reflected the research team's work. Training materials, memos, category/code change logs, and meeting notes were reviewed to assess that their content was reflected in the final code dictionary. Though the coding team agreed that saturation had been reached, due to the large size of these decks, two coders (EB and AR) applied the final codebook to a further sample of 520 notes and reached a consensus on every card in the sample.

2.5 Results

Due to the recency of EF's rise among medical learners, many medical educators may lack a robust, confident understanding of EFs. Therefore, the primary goal of this manuscript is to present the DEAME Framework to medical educators to empower their involvement in the study and design of EFs. The framework was developed from a detailed, theory-informed, inductive-deductive process that draws on state-of-the-art learning theory by incorporating the *NT*. As such, the framework is substantial and highly granular, containing seven categories and 48 codes, many of which will apply to every EF (see Table 2.2 for a complete overview of the framework and Figure 2.3 for an example application of the complete framework to a single example flashcard). In the following section, we present each of the seven categories in the framework with a visual aid containing definitions and examples of codes in that category, as well as commentary on possible student-centered design implications. Further description of the framework may be found in our final codebook (Appendix B) and examples document

(Appendix C). At the end of this section, we discuss challenges with the framework, including limitations in using the *NT* to classify EFs.

2.5.1 Framework Description

2.5.1.1 Note Components

The *Note Components* category describes the multimedia elements present on an EF and has two subcategories: *Question Components* and *Extra Components* (Table 2.3). In practice, *Question Components* are multimedia elements on the front of an EF that are necessary or supportive to the learner in answering the question. For example, image-based *Question Formats* described in the next section would require an *Image* in the question to be answerable by a learner. In contrast, *Text Emphasis* is not required by any question types but could help the learner quickly identify the key elements of the question. After the learner reveals the answer on the back of the card, *Extra Components* can elaborate on why an answer is correct or connect learners to helpful resources.

The variety of *Note Components* offers significant potential to support learners. For example, to arrive at a medical diagnosis, clinicians need to gather and integrate many types of information, such as patient history, physical exam findings, and laboratory tests. EFs that ask learners to diagnose an illness could replicate this real-life scenario by containing any or all of the following: a text summary of the patient history, a video of key physical exam findings, and tables with laboratory findings. Additionally, many students use pre-made decks and therefore may encounter EFs with new material. Abundant explanation via *Extra Components* may allow them to successfully integrate novel material without seeking additional sources. However, when including *Note Components* on an EF, accessibility is a key consideration. Screen readers may be unable to parse content within images and, therefore, exclude visually impaired learners from

fully engaging with image-containing EFs. Likewise, audio recordings could increase accessibility for this same group.

2.5.1.2 Question Format

The *Question Format* category jointly describes the kind of question on an EF ("question type") and the syntax through which the Anki platform encodes that question ("question encoding") (Table 2.4). A key observation is that every type of question can be encoded with at least two different syntaxes. The choice of encoding impacts what the learner sees on the front and back of the card, as well as the process of card creation. For example, the *Basic* format is similar to a traditional paper flashcard in which only the answer, not the question and answer, appears on the back–and must additionally be created by inputting the question and answer into separate text boxes called fields. Conversely, the *Basic Cloze* format uses a single field for card creation and the special syntax "{{c[cloze #]::answer shown on card back::optional answer hint shown on card front}}" such that the question and answer are both visible on the card back (see Figure 2.2 for an example flashcard which uses the cloze syntax). In this way, the *Basic Cloze* format may streamline card review–such that learners do not need to switch back and forth between the question/front and answer/back to review the card fully–and card creation–which may be done in a single field.

The construction of this category as a cross-section between question type and encoding hints at a key observation about EF design: form influences function and vice versa. For example, an EF deck that aims to support learner's identification of anatomical structures could heavily incorporate *Image Occlusion* and *Image Selection* cards. Therefore, this category can complement the *NT*-derived categories below in guiding the alignment of EF design with curricular goals. Finally, the *Information Only* code in *Question Type* suggests that EFs may be

used for information storage and passive learning instead of active learning alone; other design features like *Card Organization* described below may be explored to support learners in this particular use of EFs.

2.5.1.3 Answer Format

The *Answer Format* category describes an answer's overall length and grammatical complexity (Table 2.5). Of these formats, *Forced-Choice* is unique in that it does not require the learner to form the answer in their own words but merely choose between already provided answers. The other *Answer Formats*, however, require the learner to craft their answers with varying levels of length and complexity (ranging from short, simple *Partial Complex Nominals* to longer, more complex *Short Answers* and *Lists*). In this way, different *Answer Formats* can likely be used to moderate the average difficulty and time required to answer a question. Indeed, this observation is supported by the fact that *Forced-Choice* answers, by definition, map to the lowest level *Cognitive Process* described by the *NT*, as discussed further below, while higher level *Cognitive Processes* often use *Short Answers*. Therefore, *Answer Formats*, alongside *Question Formats*, can complement the *NT*-derived categories below to align EF design with curricular goals. Together, these two categories can also represent known types of assessment questions. For example, a text-based free response question would have a *Basic* or *Basic Cloze Question Format* and a *Short Answer Answer Format*. By considering the *Question Format* and *Answer Format* in tandem, card creators can incorporate existing literature on known types of assessment questions and additionally gain insight into rendering them in the Anki platform.

2.5.1.4 Cueing

The category of *Cueing* describes how or whether the question communicates to the learner from what class of knowledge the answer should be drawn (Figure 2.4). When questions

are syntactically interrogative or imperative, they typically necessitate the inclusion of a *Cue* due to their grammatical construction. For example, "Which medication commonly causes chronic cough?" or "Name a medication that commonly causes chronic cough" both suggest that the answer should be drawn from the class of knowledge "medications." However, when the front of a card instead contains a declarative statement with a fill-in-the-blank/*Cloze* question, a *Cue* may not be present. For example, the *Cloze* question "{{c1::ACE inhibitors}} commonly cause chronic cough" does not contain a *Cue* for the answer "ACE inhibitors," which are a type of medication. When no or limited assistance is given, we say a question has no *Cueing*. When more robust assistance is given, we describe two types of *Cueing*: *Direct* and *Parallel*.

In practice, *Cueing* represents the degree and type of assistance the question provides the learner in knowing what kind of answer the question requires. When questions lack this assistance, they can be challenging (which may be a useful tool for testing a more expert audience), frustrating (because the learner must spend time figuring out what the question is asking, instead of what the answer is), or useless (because the learner cannot figure out what the question is asking, and must reveal the answer to figure it out). Questions with stronger *Cueing*, especially *Parallel Cueing*, may help trigger the answer by encouraging "cued recall," in which encountering related terms helps learners remember missing information.[99]

2.5.1.5 Knowledge Domain

The category of *Knowledge Domain* describes the type of knowledge encapsulated in the question and answer (Table 2.6). This category is derived from the *NT* and has three subcategories: *Information*, *Mental Procedure*, and *Psychomotor Procedure*. This distinction between informational and procedural knowledge is present not only in the *NT*, but also in curricular design and assessment for medical education.[100] While procedural knowledge might

seem challenging to encode in an EF, the *NT* recognizes an overlap between how brains store and engage with *Information* and *Psychomotor* or *Mental Procedure Knowledge Domains*; when a learner first encounters *Procedural* knowledge, they will often encode it in words, similar to *Information*. As they become more expert, they may engage with this knowledge automatically. Therefore, to distinguish *Procedure* from *Information*, the coding team was instructed to use their medical expertise to infer from textual descriptions when a "doer" was present in the question or answer. For example, we coded "What is the definitive diagnostic test for pulmonary embolism?" as a *Mental Procedural* question because it implies that an unnamed clinician must perform a cognitive action to choose the correct diagnostic test, "Pulmonary angiogram."

Medical learners must engage with many different types of knowledge during their studies. By recognizing the presence of *Information*, *Mental Procedures*, and *Psychomotor Procedures* in EFs, educators may better be able to target specific learning objectives and select complementary design elements to encourage their mastery.

2.5.1.6 Cognitive Process

The *Cognitive Process* category, also adapted from the *NT*, describes the type of thinking a learner might engage to generate the answer (Table 2.7). In other words, how does a question test knowledge? *Retrieval* is the least complex subcategory requiring the learner to generate an answer but not demonstrate an "in-depth" understanding of the "basic structure…or critical and noncritical elements" of the knowledge.[69] The next most complex *Cognitive Process*, *Comprehension*, requires the learner to demonstrate this "in-depth" understanding of the concepts underlying the answer, such as how concepts of *Information* are related or why a *Procedure* works. The final two *Cognitive Process* subcategories require using this "in-depth" knowledge to reason and generate new knowledge. The subcategory *Analysis* requires the learner

to reason to "generate new conclusions," and the subcategory *Knowledge Utilization* requires the learner to reason to "apply…knowledge in specific situations." Of the 13 codes shared between these four subcategories, we include nine in the DEAME Framework. Of the four excluded codes, we constructed draft definitions in our codebook for three that we felt could reasonably be rendered in the EF format but did not encounter in our sample (Appendix D).

Several important observations emerged while applying the *NT Cognitive Processes* to EFs. First, different students may employ different *Cognitive Processes* on the same card depending on their prior experience with the material. For example, if a student already knows that Myasthenia Gravis and Lambert-Eaton Syndrome belong to the category Neuromuscular Junction Diseases, then they may simply *Recall* that relationship instead of newly creating it in their minds with the *Classifying Cognitive Process*. When multiple *Cognitive Processes* were possible for a given card–typically a *Retrieval* and a higher-level *Process*–we selected the most complex one. A related observation is that while these *Cognitive Processes* are ordered by complexity, that does not mean they are ordered by difficulty. For example, using a *Retrieval* process in a multiple-choice question about advanced clinical material a learner is unfamiliar with may be easier than using a more complex *Cognitive Process* on more foundational information she learned recently.[69] Nonetheless, the two *NT*-derived categories of the framework provide exciting avenues for student-centered and curriculum-aligned EF design. For one, the presence of cards, which may elicit higher-order thinking like *Decision Making,* suggests that EFs may be able to support more advanced medical learners in graduate medical education (GME) or continuing medical education (CME).

2.5.1.7 Card Organization

The *Card Organization* category describes the two hierarchical labeling systems that can classify, arrange, or order EFs (Table 2.8). In Anki, each EF is assigned on creation to a location in an EF deck or subdeck (i.e., *Deck/Subdeck Structure*) and may optionally be assigned *Metadata Tags*. Neither organizational system appears by default to the learner on either the front or back of individual flashcards but may be used to select specific cards for review. Indeed, this function may be very valuable to support learners given the expansive size of some popular EF decks (in our study, ranging from approximately 15,000-31,000 cards as shown in Table 2.1). Different populations of learners may benefit from using *Card Organization* labels to review decks in specific and measured ways, including selecting cards relevant to their current curriculum (e.g., through *Content Area* and *Curricular Exam or Division* labels) or prioritizing important cards for emergent review (e.g., through *Importance* labels). Additionally, *Card Organization* labels that serve more clerical functions (e.g., *Status* and *Provenance*) could help support the ongoing improvement of EFs deck, particularly through crowd-sourcing models.

2.5.2 Framework Challenges

All frameworks represent an artificial way to apply organized representations to complex real-world phenomena. Frameworks are, by definition, never complete and may need consistent refining to continue to achieve their stated goals. In this way, the framework presented in this manuscript was designed in such a way as to allow for its future modification and extension. This section outlines the challenges we faced in developing the framework.

For elements of the framework that were derived from the *NT*, some of the subcategories and codes in the *NT* did not apply to the EF formats observed. For example, the only *Psychomotor Procedure* code that the coding team agreed could appear on a flashcard is a

Simple Combination Procedure, which describes physical motions performed together to complete a simple task; other codes in the *NT* for Psychomotor Procedure felt either too embodied (e.g., a *Complex Combination Procedure*, like playing basketball) or too basic (e.g., *Foundational Procedure*, like body posture or balance) to be reasonably assessed through the flashcard medium. Additionally, the *NT* contains two high-level *Cognitive Processes–Self-system Thinking* and *Metacognition*–describing the relationship between knowledge and a learner's motivations, emotions, and beliefs. While EF platforms themselves may support engagement with these *Cognitive Processes*, the coding team concluded that individual questions and answers on EFs were unlikely to support these ways of thinking.

Additionally, some elements of the framework appeared to apply to EFs but offered barriers to consensus building among the coding team. The *Cueing* category, for example, presented challenges in creating a clear boundary between *Cues* specific enough to be *Direct Cues* and *Cues* that offered no *Cueing*. Indeed, this boundary may change depending on the expertise of the learner. For example, a cardiologist may not have trouble selecting a limited number of likely candidate answer classes for "What do you do for heart failure?", but a medical student might. We encountered similar challenges creating boundaries between the *NT*'s types of *Information*, specifically of *Facts* (a kind of *Detail*) and *Generalizations*/*Principles* (two kinds of *Organizing Ideas*), as they applied to EFs. A key challenge here was the tension between the semantics of the question and our team's understanding of the content. For example, imagine a card that tests symptoms of upper respiratory infections (URIs). This card could be worded as "Rhinorrhea is a symptom of {{c1::URIs}} in >80% of cases," which is a *Fact*; this statement illustrates a specific characteristic of a person, place, or thing (in this case, a symptom, rhinorrhea). However, worded another way, "Most (>80%) {{c1::URIs}} cause rhinorrhea"

reads as a *Principle*, which is a specific type of *Generalization* that describes cause and effect relationships. This tension between the semantics of how content is represented, and the content itself is difficult to resolve in flashcards, which often contain short questions, and between coders who may have a different subjective understanding of the content. Despite this lack of overall agreement, subsets of coders could consistently apply these codes with moderate or high agreement.

In both examples, *Direct* vs. no *Cueing* and *Details* vs. *Organizing Ideas*, the coding team felt as if there were some palpable differences in how each could impact an EF learner, though we could not reach an absolute agreement. Therefore, we have retained the codes in the framework and encourage future work to improve their definitions using methods suited to this task.

2.6 Discussion

In recent years, EF use has exploded among medical learners. This popularity merits substantial involvement of educational stakeholders, so that learners may feel adequately supported in EF use. A key area of need is alignment of EFs with curricular resources, but hitherto EFs have been minimally connected to theories of learning which might support that alignment. Therefore, we propose the Design Elements for Anki in Medical Education (DEAME) Framework which offers a comprehensive terminology, including an extension of the *NT*, for learners and educational stakeholders to explore student-centered EF design principles. DEAME contains seven categories that describe overarching themes of flashcard design:

- One category–Note Components–describes the audiovisual elements of questions, answers, and explanatory information

- Three categories–Question Format, Answer Format, and Cueing–describe the formulation of questions and answers
- Two categories–Knowledge Domain and Cognitive Process–are derived from the *NT* and describe knowledge representation in EFs
- One category–Card Organization–describes how EFs are organized into decks and subdecks and assigned metadata labels

These seven categories demonstrate that EFs may display significant variety in their design. This simultaneously supports the value of a unifying framework like DEAME and highlights the significant challenge of optimizing this medium for student use. Nonetheless, each section of the DEAME Framework suggests possible avenues for exploration to develop student-centered, high-quality EFs. For example, which *Question Type* is most effective in teaching recall of anatomical structures? The framework could support exploration of research questions by providing a missing terminology to communicate EF design choices and thereby enhance the interpretability and replicability of EF studies. Because EF platforms like Anki generally contain highly empirical and granular usage metrics,[92] such exploration can likely be performed quickly and in parallel without relying on protected educational data. Notably, DEAME Framework is flexible and could be further extended alongside new understandings about EF design. Additionally, the framework borrows minimally from Anki- or medicine-specific terms and may apply to other EF platforms or fields of study.

A key strength of the terminology is that it connects EFs to a rich tradition of educational thought and research by its extension of the *NT*. Notably, this extension is a novel one, in that it represents the first time, to our understanding, that assessment items have been aligned fully *post*

hoc to the *NT*'s domains. We found this application of the *NT* to EFs for UME, overall, to be reasonably simple and highly powerful. For example, we feel the granularity of the *Cognitive Processes* in the *NT*, as compared to other taxonomies like *Bloom's*, appropriately captured the breadth of questions that we encountered in our decks. Our application of the *NT* to EFs was eased by the *NT's* rich description and abundant examples,[44] particularly because these were done in cross-section between *Knowledge Domains* and *Learning Processes*. However, our team noted that 1) few examples provided by the *NT* were directly related to medical-school content and 2) the extreme brevity of some flashcards relative to other assessment items occasionally complicated this abstraction, particularly in distinguishing the sub-types of *Information*. While our team was largely able to overcome these challenges through discussion and familiarity with both EFs and the *NT*, researchers attempting a similar application of the *NT* should consider how the *NT's* example assessment items may differ from their own.

The DEAME Framework represents a new connection forged between EFs and learning theory that motivates future work into curriculum-aligned EFs. Existing EF decks could be assessed for their current alignment with the educational goals of courses, institutions, or national/international organizations (e.g., the United States Medical Licensing Examinations) and updated to improve alignment. To enable this alignment, future work may leverage all aspects of the framework into a formal process for converting learning objectives and course outlines into flashcard decks. The development of such a formal process could rest on a qualitative observation about EFs which the terminology surfaces: that elements in the terminology are highly connected. For example, if a card creator desires a card with a *Recognizing Cognitive Process*, they must necessarily create a card with a *Forced-choice Answer*

Type. Relationships between elements could be further explored with quantitative or other qualitative methodologies to inform guidelines for EF alignment with curricula.

Our study has several limitations that motivate continued development. While we aimed to sample a variety of decks which are widely used by the UME population, we may have omitted decks with important or unique design elements. Similarly, the scope of our project excluded aspects of EFs, such as the spaced repetition algorithms integrated in EF platforms, which may meaningfully impact learner outcomes, and which educators may wish to consider when designing educational interventions with EFs. Additionally, more specific methodologies may be needed to explore the relationship of the *NT*'s domains to EFs, as learners may contextualize and engage with knowledge differently between individuals and across time. Such exploration could additionally clarify which types of medical skills and knowledge are best taught and reinforced by the EF medium.

Use of EFs among medical learners is substantial and continues to rise every year. Given that medical students who use EFs typically use them for hundreds to thousands of hours across their years of training, EFs represent a promising opportunity to improve the capacity of our future physician workforce. We hope that the DEAME framework offers important and foundational work to encourage the involvement of expert educators, researchers, and clinicians into the development and evaluation of EFs for the next generation of medical learners.

2.7 Figures and Tables

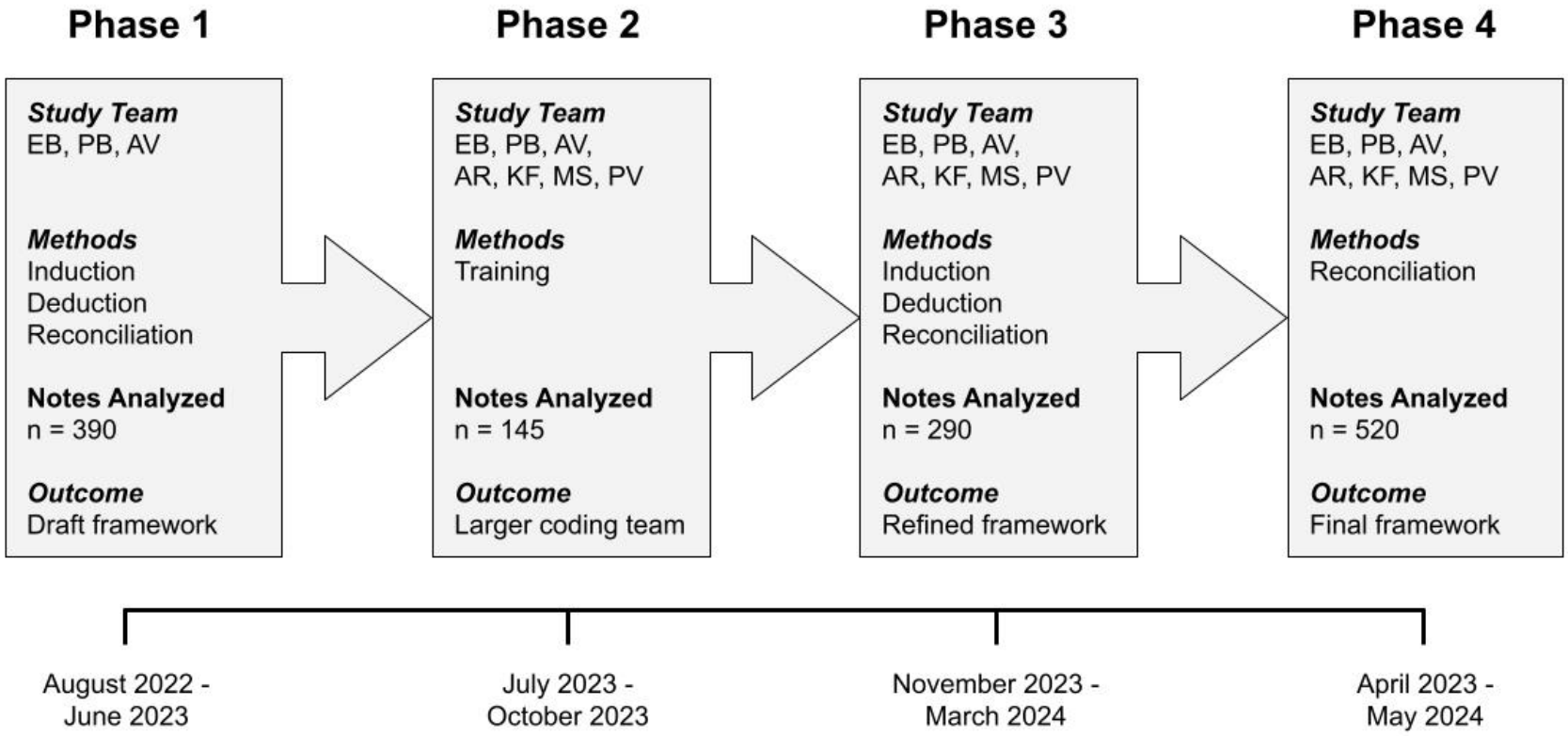


Figure 2.1 Study overview. Each phase of the four-phase study design lists the study team members involved, the methods used, the number of notes analyzed in the study sample, and the outcome of the study phase.

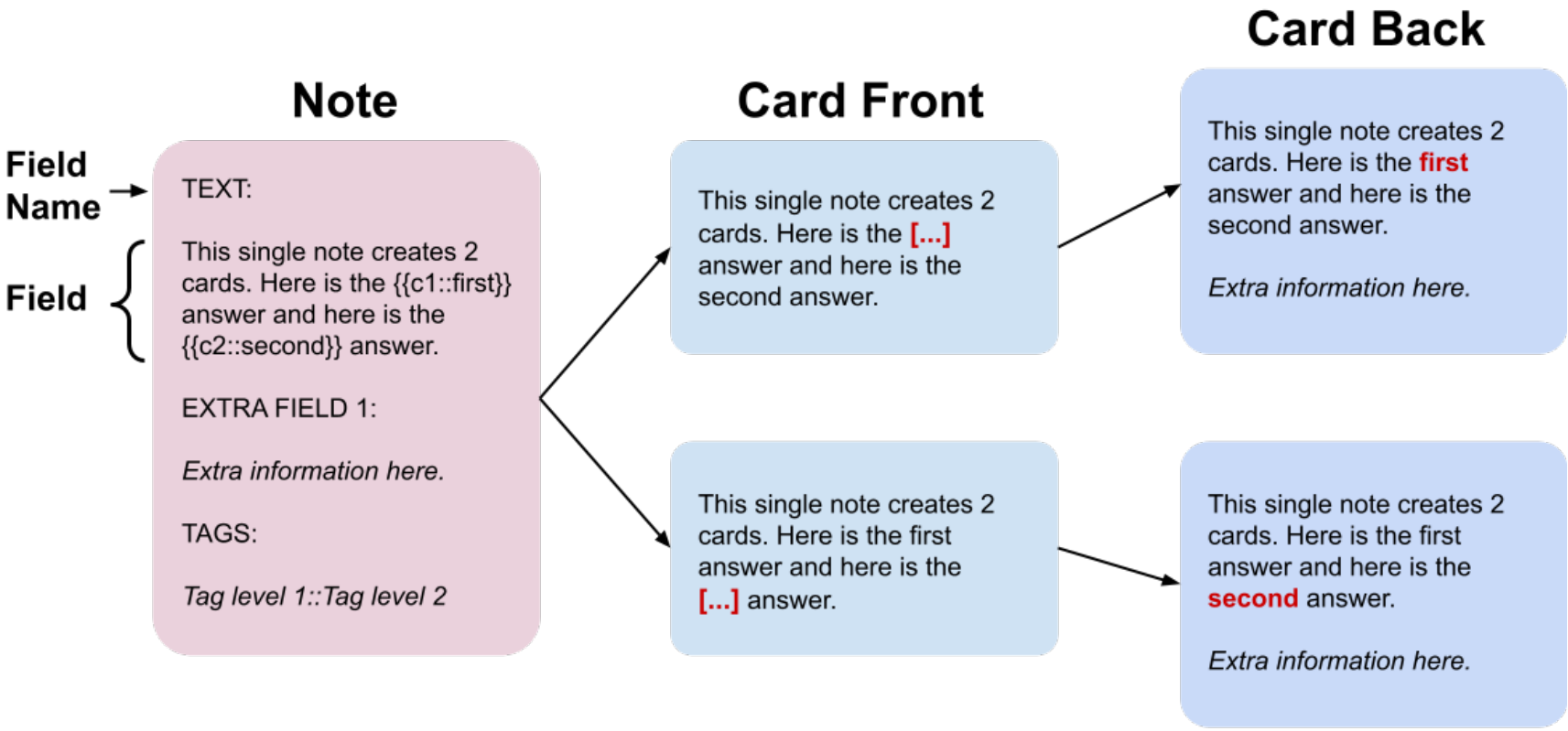


Figure 2.2 Simplified schematic of an Anki flashcard. Notes are the base unit of flashcard creation, composed of one or more fields, and each may generate more than one card. Fields are named boxes that contain text and/or multimedia, are fully customizable, and can be shown or not shown to users on the front and/or back of a flashcard. Flashcards or cards are the objects that Anki users see when they review content on the Anki platform. Cards have a frontside which shows the question and a backside that shows the answer and any extra information. This example flashcard is a fill-in-the-blank question which uses the special cloze syntax: "{{c[cloze #]::answer shown on card back::optional answer hint shown on card front}}".

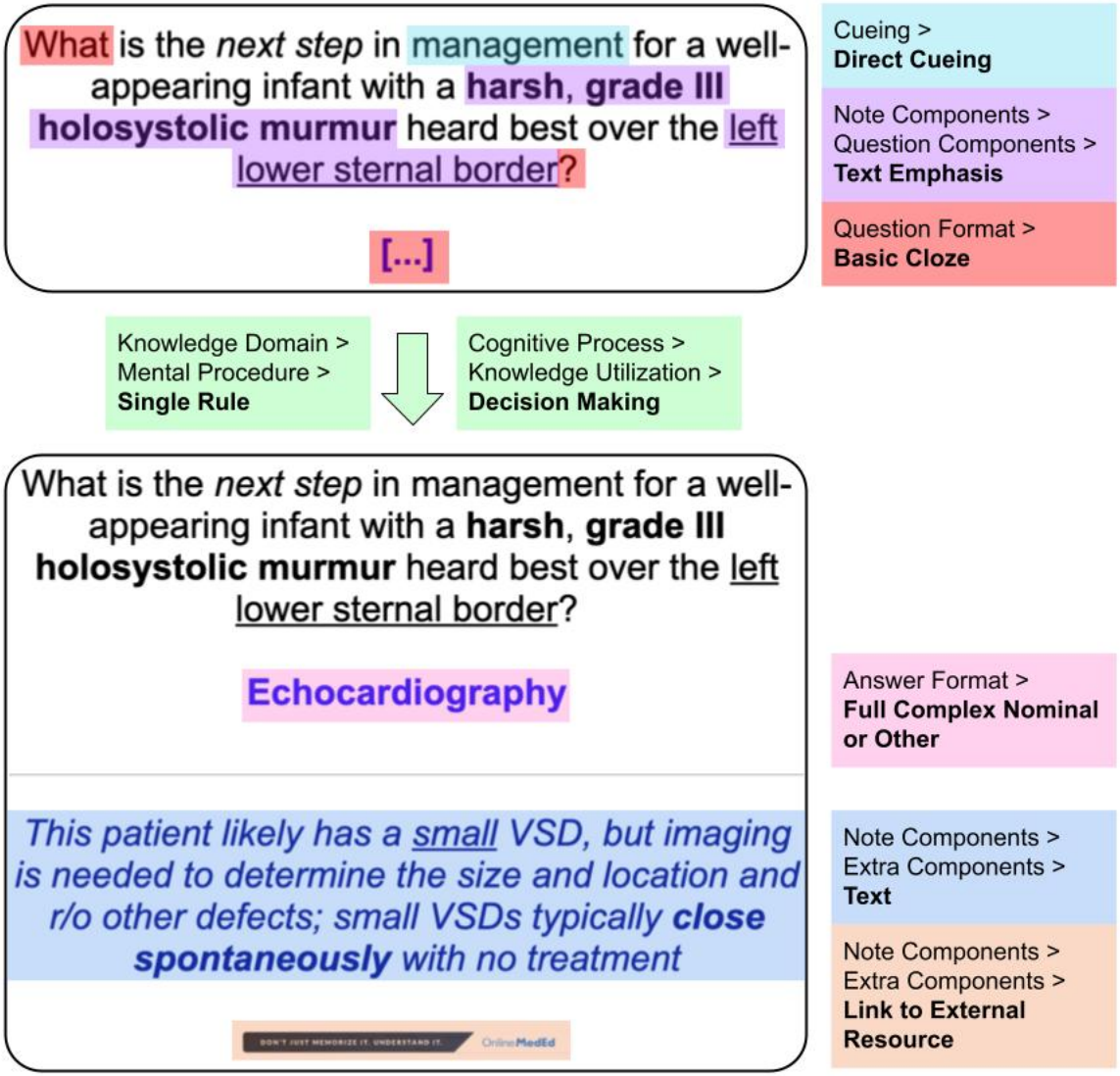


Figure 2.3 Example framework application. Example of a fully annotated Anki note, which generates a single flashcard front (rounded rectangle at top of figure) and back (rounded rectangle at bottom of figure). The colored highlights show the parts of the flashcard that were most helpful in determining applications of framework codes. The green arrow signifies the relationship between the question on the card front and answer on the card back.

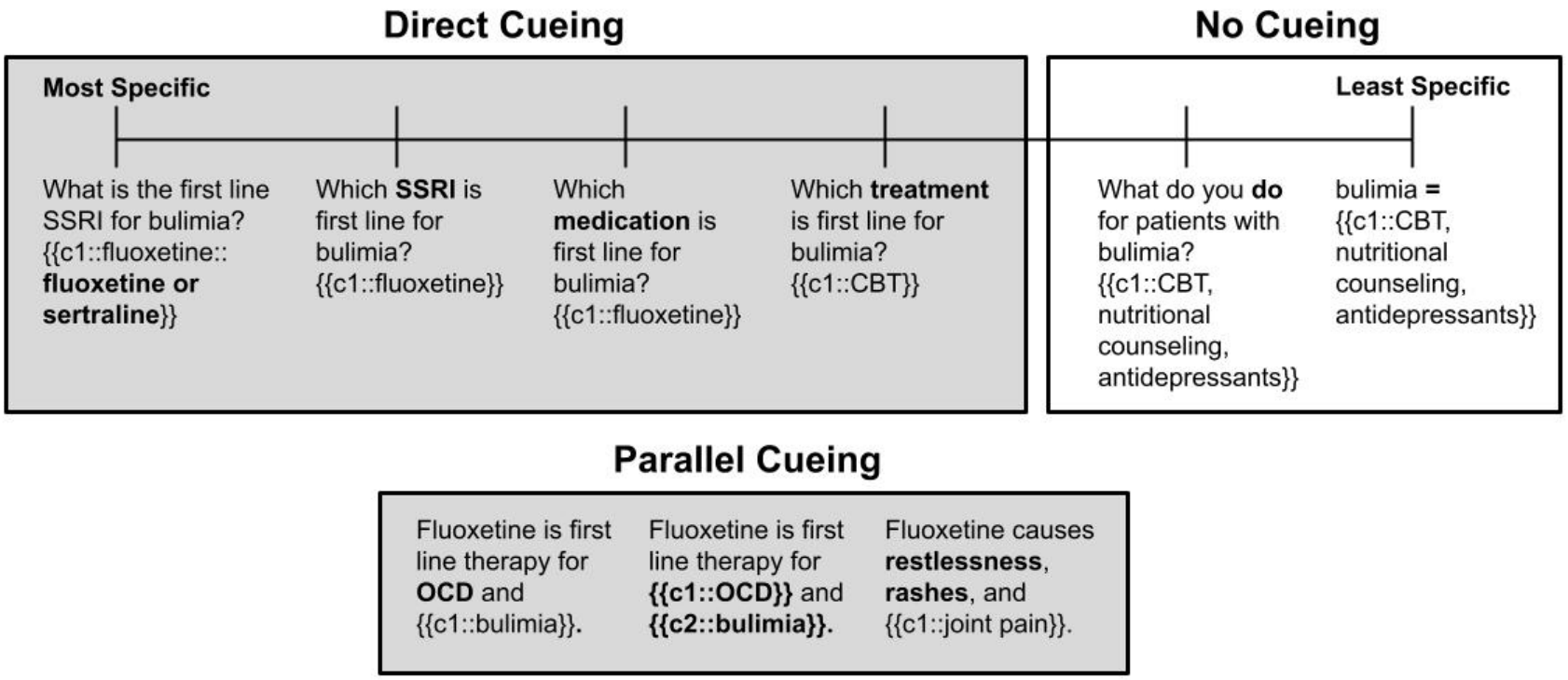


Figure 2.4 Cueing schematic. Graphic representation of the two types of *Cueing–Direct* and *Parallel*–which tell the user what class of knowledge the answer should belong to, as well as the absence of *Cueing*–no *Cueing*. *Direct Cueing* and no *Cueing* are situated at opposite ends of a continuum of specificity of *Cue*, with examples of *Cues* (bolded) provided along that continuum. *Parallel Cueing* does not exist on this same spectrum, and instead relies on words related to the answer to provide the user with a *Cue*. One may note that each example may have multiple *Cues* with similar or different types. For example, the first example of *Parallel Cueing* has a *Direct Cue* ('therapy') as well as a *Parallel Cue* ('OCD').

Deck Name	Component Subdecks	Curricular Phase	Country of Origin	Crowd-sourced	Number of Cards	Download Source	Download Date	Upload Date	Deck Version
AnKing Step 1	Lolnotacop, Zanki Step 1 & Pharmacology	Step 1	USA	X	31475	Reddit -> MEGA Drive	11 January 2023	19 Feb 2022	11
AnKing Step 2	Zanki Step 2, Cheesy Dorian, MedicalArk	Step 2	USA	X	11777	Reddit -> MEGA Drive	11 January 2023	19 Feb 2022	11
Dope: Medical Science	Bros	Step 1	Australia		21933	Reddit -> Google Drive	26 October 2022	1 May 2018	1
Dope: Clinical Medicine	Bros	Step 2	Australia		15752	Reddit -> Google Drive	26 October 2022	1 May 2018	1
Foundations		Preclinical - 2 year	USA		15154	Reddit -> Google Drive	26 October 2022	21 Jul 2019	1
UMMS Preclinical		Preclinical - 1 year	USA		28165	N/A	16 January 2023	2020-2021	1

Table 2.1 Study deck information. Details about creation, intended audience, and download of the six study decks in our sample.

Category	Subcategory	Codes	Inductive	Deductive	Deductive Source
Note Components *What multimedia elements are on the front and back of a flashcard?*	Question Components	Text Emphasis Image Table Audio	X		
	Extra Components	Text Image Table Audio Link to External Resource Extra Component in Answer			
Question Format *What type of question is on the flashcard?*		Basic Basic Cloze Cloze Image Occlusion Cloze Image Occlusion Image Selection Information Only	X	X	Basic, Cloze, and Image Occlusion from the Anki platform
Answer Format *Which form must the answer take?*		Forced-Choice Partial Complex Nominal Full Complex Nominal or Other Short Answer List	X	X	Forced-Choice from the *NT*
Cueing *How does the question inform the user which form their answer should take?*		Direct Parallel	X		
Knowledge Domain	Information	Details Organizing Ideas		X	All from the *NT*

Category	Subcategory	Codes	Inductive	Deductive	Deductive Source
What kind of knowledge is the card testing?	Mental Procedure	Single Rule Tactic Algorithm			
	Psychomotor Procedure	Simple Combination Procedure			
Cognitive Process *How is the flashcard testing a given piece of knowledge?*	Retrieval	Recognizing Recalling Executing		X	All from the *NT*
	Comprehension	Integrating			
	Analysis	Classifying Matching Generalizing Analyzing Errors			
	Knowledge Utilization	Decision Making			
Card Organization *How do flashcard decks store and categorize information?*	Deck/Subdeck Structure OR Metadata Tags	Content Area Curricular Exam or Division Third-Party Resource Difficulty Importance Status Format Provenance Deck/Subdeck Structure (for *Metadata Tags* only)	X	X	Names inductively generated from examples and descriptions in documentation of the six study decks

Table 2.2 Overview of the complete Design Elements of Anki for Medical Education (DEAME) Framework. Abbreviations: *NT = The New Taxonomy of Educational Objectives* by Marzano and Kendall (2008).

Subcategory	Codes	Definition	Example Card Front	Example Card Back
Question Components	Text Emphasis	Text in the question has formatting like bolding or italics, generally to emphasize keywords	"What **class** of drugs does bisoprolol belong to?"	
	Image Table Audio	A media object (i.e., image, table, or audio) is present in the question	"What **class** of drugs does bisoprolol belong to?" *Audio recording pronouncing bisoprolol*	
Extra Components	Text	Text is present in the answer or extra fields to provide explanation or context for the answer		Answer: "Beta-blockers " Extra Fields: "Beta-blockers have the suffix -lol"
	Image Table Audio	A media object (i.e., image, table, or audio) is present in the answer or extra fields		Answer: "Beta-blockers" Extra Fields: *Table summarizing the indications and side effects of Beta-blockers*
	Link to External Resource	A link/URL is present in the answer or extra fields		Answer: "Beta-blockers" Extra Fields: *Video link to YouTube video about Beta-blockers*
	Extra Component in Answer	Any of the above *Extra Components* is present in the answer itself, instead of extra fields		Answer: "Beta-blockers (suffix = -lol)"

Table 2.3 *Note Components* definitions and examples. Question Components are present on the front side and Extra Components are present on the back side of an EF. One or more extra fields may appear below the answer on the back side of the card. Of note, the Anki platform supports video files, though none were encountered in our sample and are therefore excluded from the framework.

Codes	Question Type	Question Encoding	Example Card Front	Example Card Back
Basic	Any except fill-in-the-blank	Question on card front and answer on card back	"What cell type is in the image below?" *Image of a red blood cell*	"Red blood cell"
Basic Cloze	Any except fill-in-the-blank	Cloze	"What cell type is in the image below? **[...]**" *Image of a red blood cell*	"What cell type is in the image below? **Red blood cell**" *Image of a red blood cell*
Cloze	Fill-in-the-blank	Cloze	"There is a **[...]** in the image below." *Image of a red blood cell*	"There is a **red blood cell** in the image below." *Image of a red blood cell*
Image Occlusion Cloze	Fill-in-the-blank	Image mask(s)	*An image contains a red blood cell and text reading "Red blood cell: average lifespan of 28 days." The first three words, "Red blood cell" are obscured by a colored rectangle.*	*An image contains a red blood cell and text reading "Red blood cell: average lifespan of 28 days."*
Image Occlusion	Identification	Image mask(s)	*An image of a red blood cell has an arrow pointing from an image label which reads "red blood cell" to the red blood cell. The image label is obscured by a colored rectangle.*	*An image of a red blood cell has an arrow pointing from an image label which reads "red blood cell" to the red blood cell.*
Image Selection	Identification	Image outline(s)	*An image of several types of blood cells has a green circle drawn around a red blood cell.*	*An image of several types of blood cells has a green circle drawn around a red blood cell.* "Red blood cell"
Information Only	Information only	Any	"**Red blood cell**: Average lifespan 28 days, lack nucleus, use hemoglobin to transport oxygen and carbon dioxide."	Same as card front

Table 2.4 *Question Format* question types, encodings, and examples. Fill-in-the-blank = question asks the learner to impute missing text; identification = question asks the learner to identify all or part of an image; information only = no question is asked; cloze = special syntax "{{c[cloze #]::answer shown on card back::optional answer hint shown on card front}}" which appears to learners as "[...]" or "[optional answer hint]" on the front of the card and "answer" on the back; image mask = a colored rectangle that covers part of an image, typically a label of a structure; image outline = a colored outline overlaid on part of an image, typically a structure.

Codes	Definition	Example Card Front	Example Card Back
Forced-Choice	Answer is chosen from fixed set of options, either directly enumerated (e.g., multiple-choice) or implied by the question (e.g., yes/no)	"Which of the following describes paralysis caused by damage to the brachial plexus? - Erb's palsy - Amyotrophic lateral sclerosis - Myasthenia gravis"	"Erb's palsy"
Partial Complex Nominal	Answer is a subset of a complex nominal	"**[...]** palsy is paralysis caused by damage to the brachial plexus"	"**Erb's** palsy is paralysis caused by damage to the brachial plexus"
Full Complex Nominal or Other	Answer is a complete complex nominal or other short phrase like a single noun, adjective, or adverb	"**[...]** is paralysis caused by damage to the brachial plexus"	"**Erb's palsy** is paralysis caused by damage to the brachial plexus"
Short Answer	Answer contains long phrases or sentences with complex grammar, typically, at minimum, a noun or noun phrase and verb	"What is Erb's palsy?"	"Paralysis caused by damage to the brachial plexus"
List	Answer contains multiple of any of the other *Answer Types* and typically has a list-like structure indicated by numbers/letters, punctuation, or spacing	"What is the differential diagnosis of Erb's palsy? **[Top 3]**"	"What is the differential diagnosis of Erb's palsy? **1) clavicular fracture** **2) osteomyelitis** **3) septic arthritis**"

Table 2.5 *Answer Format* definitions and examples. Complex nominal = a phrase with a noun and a collection of adjectives or nouns representing a single existing entity which we expect a general medical audience to recognize.[101]

Subcategory	Codes	Definition	Example Card Front	Example Card Back
Information	Details	Definitions and characteristics of specific persons, places, living and nonliving things, and events	"Gout is characterized by **[...] crystals** in the joint fluid."	"Gout is characterized by **monosodium urate crystals** in the joint fluid."
	Organizing Ideas	General statements for which examples could be provided about classes/categories/abstractions of persons, places, living and nonliving things, and events; includes associations and cause-effect relationships	"Men over 65 with **CKD** are likely to develop what rheumatic joint condition?"	"Gout"
Mental Procedure	Single Rule	A single cognitive action OR the circumstance in which that single cognitive action should be performed; an action requires a "doer" of that action, thorugh the "doer" may not be explicitly stated	"What **definitive diagnostic test** should be ordered for suspected gout?"	"Joint fluid analysis"
	Tactic	Multiple *Single Rules* which combine in no prescribed order to form a complete process for completing a cognitive task OR the circumstance in which those rules should be used	"What **panel** of diagnostic tests should be ordered for suspected gout?"	"- Joint fluid analysis - Blood/urine uric acid - ESR/CRP - Joint imaging"
	Algorithm	Multiple *Single Rules* which combine in a prescribed order to form a complete process for completing a cognitive task OR the circumstance in which those rules should be used	"What is the **treatment algorithm** for gout?"	"If no NSAID risk factors -> NSAIDS, otherwise: Monoarticular + injectible joint -> intra-articular corticosteroids Polyarticular or non-injectible joint -> oral steroids/colchicine"
Psychomotor Procedure	Simple Combination Procedure	The steps to perform a simple task comprised of physical motions OR the circumstance in which that task should be performed	"What are the **steps** to perform the Babinski maneuver?"	"- Seat/recline patient - Hold heel in one hand - Dirmly stroke bottom of the foot from heel to big toe with bottom of reflex hammer"

Table 2.6 *Knowledge Domain* definitions and examples. Subcategories and codes are derived from the *New Taxonomy of Educational Objectives* by Marzano and Kendall.

Subcategory	Codes	Definition	Example Card Front	Example Card Back
Retrieval	Recognizing	The user selects the answer from a set of provided or implied answer choices (i.e, a *Forced-Choice* answer)	"Is gout characterized by **monosodium urate crystals** in the joint fluid?"	"Yes"
	Recalling	The user produces an answer from memory, but does not necessarily understand how or why that answer is correct	"Gout is characterized by **[...] crystals** in the joint fluid."	"Gout is characterized by **monosodium urate crystals** in the joint fluid."
	Executing	The user performs a procedure without significant error but does not necessarily understand how or why the procedure works	"Compute the **corrected calcium** for the CBC results below." *Table with CBC results*	"12.2 mcg/ml"
Comprehension	Integrating	The user identifies how or why two pieces of knowledge are connected	"What is the **pathophysiology** of gout?"	"Elevated uric acid -> monosodium urate crystal deposition in joints -> joint pain/inflammation"
Analysis	Classifying	The user identifies superordinate and/or subordinate categories related to multiple pieces of knowledge	"Rheumatoid arthritis, systemic lupus erythematosis, and gout all belong to which **group** of diseases?"	"Inflammatory arthropathies"
	Matching	The user identifies important similarities and differences	"How do you **distinguish** between	"Gout: monosodium urate crystals, small joint distribution, elevated uric acid levels

Subcategory	Codes	Definition	Example Card Front	Example Card Back
		between multiple pieces of knowledge	gout and pseudogout?"	Pseudogout: calcium pyrophosphate crystals, large joint distribution, normal uric acid levels
	Generalizing	The user constructs new "rules-of-thumb" or associative/causal relationships from existing knowledge	"What can the pathophysiology of gout and rheumatoid arthritis likely tell you about how all inflammatory arthropathies develop?"	"Damage to joint epithelial cells causes release of pro-inflammatory mediators like cytokines which can create a positive-feedback cycle of further joint destruction and inflammation"
	Analyzing Errors	The user identifies whether something is true/correct or false/incorrect and explains why	"Pseudogout is characterized by elevated uric acid levels. T/F and why?"	"False. Gout is characterized by elevated uric acid levels. Pseudogout may have elevated calcium levels, but has normal uric acid levels.
Knowledge Utilization	Decision Making	The user uses existing knowledge to make a decision in a specific context	"A 65yo man with diabetes and CKD presents with acute swelling, pain, and warmth in his right DIP joint. What is the likely **diagnosis**?"	"Gout"

Table 2.7 *Cognitive Process* definitions and examples. Subcategories and codes are derived from the *New Taxonomy of Educational Objectives* by Marzano and Kendall

Subcategory	Codes	Definition	Example
Deck/Subdeck Structure OR *Metadata Tags*	Content Area	Labels that suggest content is related to general (e.g., medical specialties) or specific (e.g., classes of drugs) content areas	Rheumatology > Inflammatory Arthropathies > Gout
	Curricular Exam or Division	Labels that suggest content is related to specific exams (e.g., medical boards) or curricular divisions (e.g., preclinical sequences/courses)	Step 2 > Internal Medicine Shelf
	Third-Party Resource	Labels that suggest content is related to third-party resources, including specific pages/videos/aspects of those resources	Sketchy Pharmacology > Anti-inflammatory Drugs > Gout Drugs
	Difficulty	Labels that suggest content has a certain anticipated difficulty or rudimentariness	Foundational
	Importance	Labels that suggest content has more or less value or importance	High-Yield
	Status	Labels that suggest content should be managed in a certain way, (e.g., be deleted)	Duplicated > Delete
	Format	Labels that suggest content is rendered in a particular format (esp. a proprietary format)	Image Selection
	Provenance	Labels that suggest which user or group of users created or edited a card	Zanki
	Deck/Subdeck Structure (*Metadata Tags* Only)	*Metadata Tags* which duplicate all or part of the *Deck/Subdeck Structure*	AnKing > Zanki

Table 2.8 *Card Organization* definitions and examples. A label may appear in a hierarchical tree that describes the location of an EF (*Deck/Subdeck Structure*) or connect it to cards with similar labels across decks/subdecks (*Metadata Tags*) Any one label may serve more than one purpose, such as the label "Low-yield Surgery" suggesting a *Content Area* ("Surgery"), *Curricular Exam or Division* ("Surgery", e.g., a Shelf exam), and *Importance* ("Low-yield").

2.8 Appendices

2.8.1 Appendix A. Methods for Card Organization.

The codes for *Card Organization* were inductively generated by EB and PB. These codes were discussed with a research librarian to evaluate whether types of metadata which commonly group biomedical knowledge were omitted; no new categories or codes surfaced in this discussion, but it was determined that *Curricular Division and Exam* should be grouped into a single code, because most curricular divisions require some form of exam or assessment. EB, PB, and AR memoed independently on the alignment with these draft codes and code definitions of 1) metadata tags in a random sample of 12 notes (two from each deck) and 2) documentation from deck creators on the organizational structures in their decks. They met to discuss their memos and reconcile these definitions. During this meeting, several general observations were made, including: 1) sometimes one descriptor for a card ended up in multiple places in the tree of *Metadata Tags*, the *Deck/Subdeck Structure*, or both (i.e., *Card Organization* elements can be redundant) 2) the AnKing decks had more mechanisms for ordering and sorting cards than other decks, and 3) we as users of some of these decks previously wish we had known about some of the tags (i.e., we would have found them useful). The only new code added as a result of these discussions was *Provenance*, which was discovered by PB in the AnKing tagging documentation.

2.8.2 Appendix B. Codebook.

The full codebook for the Design Elements in Anki for Medical Education (DEAME) Framework may be found in the DeepBlue Data Repository at the following URL: https://hdl.handle.net/2027.42/196329

2.8.3 Appendix C. Codebook Examples.

Examples for every code in the Design Elements in Ankie for Medical Education (DEAME) Framework may be found in the DeepBlue Data Repository at the following URL: https://hdl.handle.net/2027.42/196329

2.8.4 Appendix D. Preliminary definitions for unseen codes.

All codes derived from the *New Taxonomy of Educational Objectives* by Marzano and Kendall (2008).

Comprehension

Symbolizing

When the learner identifies *why* or *how* two knowledge objects are connected through a symbolic explanation, such as a visual schematic of the knowledge.

Example: Draw a graphic organizer containing the screening guidelines for diabetes in adults.

Analysis

Specifying

When the learner identifies a new **application** of known generalizations. In contrast to *Generalizing*, this is primarily a deductive process, not an inductive or mixed inductive/deductive process.

Example: What can we assume about diabetes management, knowing that it is a chronic disease? {{c1::few or no definitive treatments exist, even with treatment may progress over time}}

Knowledge Utilization

Problem Solving

When the learner uses existing knowledge to overcome a barrier to typical action/behavior in a specific context.

Example: An 89 year-old man presents to the ED with shortness of breath. He has a history of hypertension and aortic valve disease. On physical exam he has notable swelling in his lower extremities and a systolic murmur. You would like to get an echocardiogram, but the machine is not working today. What other test could you use instead? {{c1::cardiac MRI}}

Chapter 3 What is in the Black Box of Learner-Created Electronic Flashcards?

3.1 Abstract

Purpose: EFs are an increasingly popular active learning tool among medical learners. Many learners use EFs created by other learners—the content and quality of which may go unchecked by experts and educators. Of central concern is whether learner-created flashcards adequately support competency and skill development in addition to facilitating knowledge acquisition. However, no prior research has empirically evaluated the knowledge and means of testing knowledge present in learner-created EF decks and their relation to skill development.

Method: Our mixed-methods study applied an inductively and deductively developed framework for EF design to a random sample of 1,179 flashcards from popular learner-created EF decks for undergraduate medical education (UME). We categorized each flashcard across four different dimensions: Content Area (e.g., flashcard is relevant to clinical vs. basic science fields), Clinical Reasoning Domain (e.g., flashcard supports reasoning like diagnosis), Knowledge Domain (e.g., flashcard tests procedural vs. informational knowledge), and Cognitive Process (e.g., flashcard involves recalling vs. applying knowledge).

Results: EFs in our sample tested a wide variety of knowledge in a variety of ways beyond simple factual recall. We report learner-created EFs can likely elicit higher-order thinking skills and test clinically relevant knowledge which may support clinical reasoning and procedural skill development. However, while Step 2 decks contained more clinically relevant and complex

flashcards than Step 1 or preclinical decks, overall clinical relevance and complexity in learner-created EFs was limited.

Conclusion: EFs have the potential to support learners in both knowledge acquisition and development of critical skills and competencies. However, learner-created EF decks may benefit from oversight by expert-educators to ensure their content and quality aligns with educational goals.

3.2 Introduction

In recent years, medical learners have adapted their learning to include a variety of third-party, "parallel curriculum" study resources.[9–11] Because these resources exist outside of school-sanctioned curricula, many educators may be "unaware of the types and amounts … [that] students are using."[102] One such parallel curriculum resource, electronic flashcards (EFs), have emerged as a highly popular active learning tool, with some studies showing adoption rates in undergraduate medical education (UME) cohorts as high as 85%,[46] rivaling the use of traditional curricular resources like lectures.[47] However, unlike curricular resources, EFs are often developed and maintained by learners themselves.[31] For example, several studies suggest that learners typically use EF decks which have been crowdsourced by other learners and distributed digitally on software platforms like Anki.[62–64,70,102] As such, popular EF decks may not have been "vetted for either quality or content by the medical school establishment."[30,31] To meet Liaison Committee on Medical Education (LCME) requirements for educational oversight and robustly support their learners, medical educators must explore what knowledge is tested and how that knowledge is tested in learner-created EFs.[31]

Clarifying the educational quality and content of learner-created EFs is critical to define their role alongside other learning modalities in medical education. In particular, EFs should be evaluated for their potential to close the classroom-clinic gap, wherein clinical aptitude—not exam scores or other measures of academic performance—is the primary benchmark of success.[32] For a resource to support clinical performance, the resource should engage the types of knowledge and ways of thinking routinely required of physicians.[103] For example, resources should prepare students for clinical reasoning tasks such as diagnosis, which requires both knowledge about diseases and their presentations as well as cognitive and procedural skills to

gather and act upon information in the clinical context.[103] EFs, including learner-created EFs, have largely been shown in literature as effective in facilitating the former: knowledge acquisition as reflected by exam scores.[19] However, can EFs contribute to development of complex cognitive and procedural skills which underlie clinical performance?

Several studies suggest that expert-created EF decks may engage with such skills. Spicer et al created curriculum-aligned EFs that "promoted higher order cognitive processes, like explaining concepts, applying content to new problems, and comparing/contrasting course material."[13] Such higher-order thinking skills (HOTS) have also been incorporated into EF creation by other groups of experts/educators.[53,54,67,68] Additionally, Schmidmeier et al found that an expert-created EF deck could facilitate acquisition and application of procedural knowledge in addition to factual or informational knowledge.[87] However, no studies, to our understanding, examine whether commonly used, learner-created decks might contain comparable or appropriate complexity. In fact, Spicer et al pursued expert creation of EFs because they noted that learner-created decks used at their US medical school too often focused on "factual recall" and "unimportant minutiae."[13]

EFs are an increasingly popular parallel curriculum resource among medical learners. Existing research on EFs has largely examined whether—not how—EFs work, making EFs feel like a "black box" to many educators. In this study, we aim to shed light on this "black box," with a particular focus on whether EFs might promote skill development in addition to knowledge acquisition. Using a theory-informed, inductively and deductively developed framework (described in Chapter II), this mixed-methods study quantifies the types of content, knowledge, and cognitive skills which may be facilitated by EFs and their relevance to clinical reasoning tasks like diagnosis.

3.3 Method

Our study was determined exempt from ongoing review by the University of Michigan IRBMED (HUM00222432).

3.3.1 Sampling

Our study sampled six Anki decks chosen for their routine use in UME. Five decks (all but the Michigan deck) are top recommendations in the popular "Med School Anki" community on the social media website Reddit,[71] and two of these (AnKing Step 1 and AnKing Step 2) have been reported in prior studies of EF use.[91,92] The final deck (Michigan) is commonly used at our institution. The study decks focus on three different phases of UME: preclinical curriculum (Michigan and Foundations), Step 1 exam (AnKing Step 1 and Dope Medical Science), and Step 2 exam (AnKing Step 2 and Dope Clinical Science). Further details about the structure, size, and origin of these decks may be found in Table 3.1.

Within each deck, we randomly sampled 150 notes, where "note" is the terminology in the Anki platform for the base unit of flashcard creation. Because a single note can give rise to multiple flashcards, the number of flashcards analyzed in each deck ranged between 164 and 264. Three flashcards in the Foundations deck have been excluded from subsequent analysis, because they simply provided information and did not ask a question, and, therefore, many aspects of our framework did not apply.

3.3.2 Framework Development

The characteristics of EFs analyzed in this study were developed through a prior qualitative study which levied a flexible approach to thematic analysis to create a framework describing the design elements in Anki flashcards for UME (see Chapter II for a description of this study). The development of two categories of EF characteristics–Knowledge Domain and

Cognitive Process–through application of the *New Taxonomy of Educational Objectives* (NT) by Marzano and Kendall[69] is comprehensively discussed in Chapter II. The remaining two categories of EF characteristic–Content Area and Clinical Reasoning Domain–were analogously developed during this study but excluded from the Design Elements of Anki for Medical Education (DEAME) framework enumerated in Chapter II because they are features of knowledge, not elements of EF design. Therefore, specifics regarding the development of Content Area and Clinical Reasoning Domain are described in Results below.

3.3.3 Framework Quantification

Once thematic saturation was determined to have been reached in the DEAME framework qualitative study, a quantification of framework categories was undertaken by EB and AR on a novel sample of EFs.[90] Notably, EB and AR participated critically throughout the prior qualitative study and had substantial experience with framework categories and their application to EFs. EB and AR each coded half of the study sample independently and then met to reconcile code applications. Through discussion, they were able to reach consensus on all code applications without necessitating substantial changes to framework definitions. Codes were applied in the Dedoose qualitative coding software and analyzed in the R programming language.[93,94] For each deck, we report the number and percentage of flashcards with a given characteristic. When comparing groups of decks (e.g., preclinical vs. Step 1 decks) during analysis, we report the mean of percentages for each group.

3.4 Results

A total of 1,179 electronic flashcards were analyzed across six commonly used EF decks for UME.

3.4.1 Content Area

All study decks covered a variety of content areas across different basic science and clinical fields (Table 3.2) Fields were identified inductively when flashcards with novel content were encountered and named according to standard terminology (e.g., medical fellowships) where appropriate, resulting in a total of 45 fields. Fields which are established areas of clinical practice were classified as "Clinical Fields" (n = 29), excepting Histology and Pathology, Nutrition, Radiology, and Toxicology which were included as "Basic Science Fields" because they are cross-cutting, foundational, and often incorporated early in preclinical curricula. The remaining fields were classified as "Basic Science Fields" (n = 16). The most common Basic Science Fields among all decks were Anatomy (mean = 17%), Histology and Pathology (mean = 14%), and Pharmacology (mean = 10%) and the most common Clinical Fields across all decks were Neurology (mean = 16%), Gastroenterology (mean = 13%), and Infectious Disease (mean = 16%). The fields of Internal Medicine, Family Medicine, and Geriatrics were collapsed into a single code of exclusion called "General Adult Medicine" due to their broad applicability and substantial overlap with related fields like Cardiology. As such, relevant subspecialties may be combined with General Adult Medicine to roughly correlate the relevance of our study sample to these general fields. For example, 83% of all flashcards in our study sample may pertain to Internal Medicine after combining the observed frequencies of General Adult Medicine and 11 related subspecialties.

Most flashcards in our sample (86%) contained more than relevant content area. Therefore, we additionally identified the most prominent (i.e., "primary") content area among these (Table 3.3).When making this determination, we considered the direct clinical applicability of the flashcard's content. For example, a flashcard which tested the general mechanism of

action of a diabetes drug should have a primary content area of Pharmacology (a Basic Science Field) while a flashcard which tested the clinical context in which that drug should be prescribed (i.e., diabetes) should have a primary content area of Endocrinology (a Clinical Field). Preclinical decks largely had a Basic Science Field as a primary content area (mean Primary Basic Science Field = 81%; mean Primary Clinical Field = 19%), while Step 1 decks had a roughly even representation of primary content areas (mean Primary Basic Science Field = 55%, mean Primary Clinical Field = 46%). Step 2 decks predominantly pertained to Clinical Fields (mean Primary Basic Science Field = 27%, mean Primary Clinical Field = 73%).

3.4.2 Clinical Reasoning Domain

To support clinical reasoning, physicians have "information needs" relating to the following mutually exclusive categories–Diagnosis, Etiology, Management, and Prognosis–which we call Clinical Reasoning Domains.[104–106] To clarify the boundaries between these categories, we considered the category Etiology to include risk factors in addition to known cause-effect relationships, Management to include preventative management in addition to reactive management, and Prognosis to include complications resulting from an intervention in addition to outcomes resulting from a disease process. Additionally, we added a further category–Ethics–after encountering a single flashcard in the Foundations preclinical deck which engaged with ethical principles foundational to clinical reasoning. Notably, to be assigned a Clinical Reasoning Domain, a flashcard need not require the completion of a clinical reasoning skill, but simply provide information directly relevant to its execution. For example, a flashcard which correlated a histological finding to a disease would support "Diagnosis," without necessitating that the user diagnose that disease from, e.g., a clinical vignette. Likewise, a flashcard which tested general steps for histological analysis without correlation to a specific

disease would not be assigned a Clinical Reasoning Domain. All decks contained flashcards (mean = 54%) that furnished information supportive of clinical reasoning (Table 3.3). Step 2 decks had the highest percentage of flashcards relevant to Clinical Reasoning Domains (mean = 82%), followed by Step 1 (mean = 55%) and preclinical decks (mean = 25%). The most common Clinical Reasoning Domains were Diagnosis (mean = 24%) and Etiology (mean = 14%).

3.4.3 Knowledge Domain

The Knowledge Domain category is adapted from the NT and describes three types of knowledge which a flashcard might prompt: Information, Mental Procedure, and Psychomotor Procedure (see Chapter II for full definitions and examples). Information describes "characteristics" of and "relationships" between "persons, places, living and nonliving things, and events."[69] Most flashcards across all decks (mean = 85%) tested Informational knowledge (Table 3.3). Therefore, a minority of flashcards across all decks (mean = 15%) tested Procedural knowledge which describes how or why a task or action should be performed.[69] Notably, almost 1/3 of flashcards in the Step 2 decks (mean = 30%) incorporated Procedural knowledge, in contrast to preclinical (mean = 8%) and Step 1 (mean = 7%) decks. Across all decks, Mental Procedures (mean = 14%), which include cognitive actions like diagnosis and management, were remarkably more common than Psychomotor Procedures (mean = 1%), which include physical actions like physical exam maneuvers.

3.4.4 Cognitive Process

The Cognitive Process category is also adapted from the NT and describes the type of thinking a learner might engage to generate an EF answer.[69] The four codes are roughly ordered by cognitive complexity from the least complex, Retrieval, to the most complex, Knowledge Utilization. Retrieval requires the learner to generate an answer but not demonstrate an "in-

depth" understanding of the knowledge's "basic structure…or critical and noncritical elements"[69] and can be considered a lower-order thinking skill (LOTS).[69] The remaining three levels can likewise be considered higher-order thinking skills (HOTS). Comprehension requires the learner to demonstrate an "in-depth" understanding of pieces of knowledge by understanding their whys or hows.[69] Finally, Analysis requires the generation of "new conclusions" from existing knowledge and Knowledge Utilization requires application of existing knowledge to "specific situations."[69] Of note, we coded the highest Cognitive Process a flashcard might elicit when it is first seen, though multiple may be possible over time and across learners. Retrieval was the most common Cognitive Process across all study decks (mean = 80%), but all decks contained a minority of HOTS (mean = 20%) (Table 3.3). HOTS were most prominent in the Step 2 decks (mean = 29%) compared to the Step 1 (mean = 16%) and preclinical decks (mean = 17%).

3.5 Discussion

Preliminary research suggests EFs are an effective learning strategy, but little is known about the content and quality of EF decks for medical education. Of particular concern is whether popular learner-created EFs support clinical skill development, in addition to knowledge acquisition. This mixed methods study explores the content areas, types of knowledge, and thinking skills tested in six popular learner-created Anki decks for UME. We show that EFs in our sample display significant variety across these dimensions, testing three different types of knowledge relevant to 45 clinical and basic science fields and five clinical reasoning tasks with four levels of likely elicited cognitive processes. In other words, learner-created EFs may support more than simple "factual recall" of limited content areas.[13]

Indeed, we find that the types of knowledge grow in clinical relevance and the ways of testing knowledge grow increasingly complex when learner-created EFs target later (i.e., the

Step 2 exam) vs. earlier (i.e., the Step 1 exam and preclinical curricula) phases of the UME curriculum. In this way, learner-created EFs appear to mimic the changing goals and expectations of learners as they progress through training. However, are these learner-created EFs sufficient to support learners in achieving mastery in the clinic as well as the classroom?

Indeed, while we show that learner-created EFs in our sample *can* engage with complex and clinical content, we still find that many EFs do not. For example, over half of all EFs in our sample focused primarily on testing basic science rather than clinical content. Additionally, about half of all flashcards did not have direct relevance to clinical reasoning tasks like diagnosis and management. Similarly, while we show that EFs are able to test procedural knowledge, which can facilitate these same clinical reasoning tasks, only the Step 2 decks do this with a substantial frequency. Lastly, we find that less than a quarter of EFs in our sample likely elicit HOTS like applying knowledge to make decisions in novel situations.

A variety of factors could explain these observations, including student preference to use EFs for rehearsing simple and factual knowledge and other learning tools for engaging with more clinical and complex concepts. However, as both learners and educators, we feel the potential of EFs to facilitate active learning of the full spectrum of knowledge and skills required of physicians should be explored. This exploration may benefit from learner-educator co-creation schemes which merge students' enthusiasm for and knowledge of the EF format with educators' expertise in both content and pedagogy.[13] In this way, we could better align EFs' ability to test a variety of content in a variety of ways with the specific goals and objectives of medical education.

To enable this alignment, further work is needed to understand how EF design impacts how learners experience and learn from EFs. Indeed, a limitation of our study is that it used

qualitative consensus building to map EFs to, e.g. likely elicited Cognitive Processes, but did not validate these mappings with more advanced methodologies like think-aloud studies. Were such mappings validated, templates and guidelines for flashcard design to target each of the areas in our study could be developed. Such efforts may benefit from exploration of a wider sample of EFs, including expert-created EFs or EFs on other platforms than Anki, than the limited sample our study included.

In recent years, learners have been the leaders in advancing EFs as an effective active learning tool for medical education. However, to enhance the capacity of EFs to more holistically develop our future physician workforce, we suggest that educators must partner with students in these efforts. We hope this mixed-methods study sheds light on the "black box" of learner-created EFs and motivates educators towards this necessary engagement with EFs.

3.6 Tables

Deck Name	Component Subdecks	Curricular Phase	Country of Origin	Crowd-sourced	Number of Cards	Download Source	Download Date	Upload Date	Deck Version
AnKing Step 1	Lolnotacop, Zanki Step 1 & Pharmacology	Step 1	USA	X	31475	Reddit -> MEGA Drive	11 January 2023	19 Feb 2022	11
AnKing Step 2	Zanki Step 2, Cheesy Dorian, MedicalArk	Step 2	USA	X	11777	Reddit -> MEGA Drive	11 January 2023	19 Feb 2022	11
Dope: Medical Science	Bros	Step 1	Australia		21933	Reddit -> Google Drive	26 October 2022	1 May 2018	1
Dope: Clinical Medicine	Bros	Step 2	Australia		15752	Reddit -> Google Drive	26 October 2022	1 May 2018	1
Foundations		Preclinical - 2 year	USA		15154	Reddit -> Google Drive	26 October 2022	21 Jul 2019	1
UMMS Preclinical		Preclinical - 1 year	USA		28165	N/A	16 January 2023	2020-2021	1

Table 3.1 Study deck information. Details about creation, intended audience, and download of the six study decks in our sample.

	Preclinical				Step 1				Step 2			
Content Area	**Michigan N = 211**		**Foundations N = 264***		**AnKing Step 1 N = 189**		**Dope Medical Science N = 164**		**AnKing Step 2 N = 170**		**Dope Clinical Science N = 181**	
Basic Science Field	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*
Anatomy	27 (13%)	21 (11%)	150 (57%)	129 (51%)	16 (8%)	10 (5%)	19 (12%)	12 (8%)	4 (2%)	0 (0%)	17 (9%)	15 (9%)
Biochemistry	28 (13%)	12 (7%)	4 (2%)	4 (2%)	15 (8%)	8 (4%)	14 (9%)	7 (5%)	3 (2%)	3 (2%)	3 (2%)	0 (0%)
Biostatistics	2 (1%)	2 (1%)	5 (2%)	5 (2%)	1 (1%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Cell Biology	13 (6%)	0 (0%)	10 (4%)	0 (0%)	15 (8%)	8 (4%)	8 (5%)	6 (4%)	0 (0%)	0 (0%)	1 (1%)	1 (1%)
Embryology	5 (2%)	5 (3%)	0 (0%)	0 (0%)	3 (2%)	1 (1%)	4 (2%)	4 (3%)	2 (1%)	0 (0%)	0 (0%)	0 (0%)
Genetics	7 (3%)	3 (2%)	2 (1%)	0 (0%)	8 (4%)	1 (1%)	9 (5%)	4 (3%)	0 (0%)	0 (0%)	2 (1%)	0 (0%)
Health Systems and Safety	0 (0%)	0 (0%)	13 (5%)	11 (4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Histology and Pathology	37 (18%)	27 (15%)	49 (19%)	35 (14%)	26 (14%)	12 (6%)	28 (17%)	10 (7%)	14 (8%)	1 (1%)	13 (7%)	7 (4%)
Immunology	17 (8%)	10 (5%)	4 (2%)	0 (0%)	4 (2%)	3 (2%)	8 (5%)	2 (1%)	2 (1%)	0 (0%)	2 (1%)	1 (1%)
Microbiology	14 (7%)	5 (3%)	3 (1%)	2 (1%)	21 (11%)	17 (9%)	16 (10%)	9 (6%)	4 (2%)	2 (1%)	1 (1%)	0 (0%)
Nutrition	1 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (3%)	4 (2%)	2 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)	0 (0%)
Pharmacology	23 (11%)	20 (11%)	3 (1%)	1 (0%)	26 (14%)	18 (10%)	22 (13%)	19 (12%)	21 (12%)	8 (5%)	12 (7%)	4 (3%)
Physiology	33 (16%)	19 (10%)	18 (7%)	15 (6%)	19 (10%)	11 (6%)	13 (8%)	7 (5%)	2 (1%)	1 (1%)	8 (4%)	3 (2%)
Public Health	1 (0%)	1 (1%)	9 (3%)	5 (2%)	11 (6%)	1 (1%)	6 (4%)	1 (1%)	4 (2%)	0 (0%)	5 (3%)	2 (1%)
Radiology	0 (0%)	0 (0%)	24 (9%)	17 (7%)	1 (1%)	1 (1%)	5 (3%)	1 (1%)	12 (7%)	6 (4%)	5 (3%)	1 (1%)
Toxicology	0 (0%)	0 (0%)	2 (1%)	2 (1%)	2 (1%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Clinical Field	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*	*Any*	*Primary*
Allergy and Immunology	1 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (1%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)

	Preclinical				Step 1				Step 2			
Content Area	**Michigan N = 211**		**Foundations N = 264***		**AnKing Step 1 N = 189**		**Dope Medical Science N = 164**		**AnKing Step 2 N = 170**		**Dope Clinical Science N = 181**	
Anesthesiology	1 (0%)	0 (0%)	2 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (2%)	2 (1%)	2 (1%)	0 (0%)
Cardiology	27 (13%)	4 (2%)	7 (3%)	0 (0%)	17 (9%)	5 (3%)	7 (4%)	3 (2%)	26 (15%)	11 (7%)	22 (12%)	14 (9%)
Clinical Genetics	2 (1%)	1 (1%)	0 (0%)	0 (0%)	3 (2%)	1 (1%)	1 (1%)	0 (0%)	6 (4%)	0 (0%)	1 (1%)	1 (1%)
Dentistry	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (2%)	0 (0%)	1 (1%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Dermatology	20 (9%)	1 (1%)	10 (4%)	4 (2%)	10 (5%)	4 (2%)	4 (2%)	4 (3%)	11 (6%)	4 (3%)	8 (4%)	2 (1%)
Emergency Medicine	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	1 (1%)	0 (0%)
Endocrinology	12 (6%)	5 (3%)	13 (5%)	0 (0%)	17 (9%)	10 (5%)	14 (9%)	1 (1%)	19 (11%)	6 (4%)	8 (4%)	3 (2%)
Gastroenterology	22 (10%)	1 (1%)	15 (6%)	0 (0%)	21 (11%)	9 (5%)	17 (10%)	5 (3%)	23 (14%)	8 (5%)	49 (27%)	20 (13%)
General Adult Medicine	0 (0%)	0 (0%)	2 (1%)	1 (0%)	3 (2%)	1 (1%)	0 (0%)	0 (0%)	4 (2%)	1 (1%)	7 (4%)	2 (1%)
General and Trauma Surgery	12 (6%)	4 (2%)	53 (20%)	0 (0%)	4 (2%)	2 (1%)	2 (1%)	1 (1%)	19 (11%)	13 (9%)	29 (16%)	21 (13%)
Pediatrics	2 (1%)	1 (1%)	5 (2%)	0 (0%)	10 (5%)	1 (1%)	5 (3%)	1 (1%)	16 (9%)	6 (4%)	16 (9%)	9 (6%)
Hematology	14 (7%)	6 (3%)	13 (5%)	2 (1%)	28 (15%)	5 (3%)	27 (16%)	9 (6%)	17 (10%)	4 (3%)	11 (6%)	2 (1%)
Infectious Disease	19 (9%)	13 (7%)	4 (2%)	0 (0%)	32 (17%)	10 (5%)	26 (16%)	11 (7%)	28 (16%)	12 (8%)	20 (11%)	3 (2%)
Nephrology	19 (9%)	1 (1%)	7 (3%)	0 (0%)	14 (7%)	2 (1%)	8 (5%)	1 (1%)	11 (6%)	4 (3%)	4 (2%)	1 (1%)
Neurology	27 (13%)	4 (2%)	59 (22%)	5 (2%)	31 (16%)	6 (3%)	23 (14%)	9 (6%)	18 (11%)	8 (5%)	38 (21%)	8 (5%)
Neurosurgery	0 (0%)	0 (0%)	33 (13%)	0 (0%)	1 (1%)	0 (0%)	3 (2%)	1 (1%)	3 (2%)	0 (0%)	8 (4%)	3 (2%)
Obstetrics and Gynecology	15 (7%)	5 (3%)	32 (12%)	3 (1%)	13 (7%)	5 (3%)	8 (5%)	3 (2%)	29 (17%)	14 (9%)	9 (5%)	2 (1%)

	Preclinical				Step 1				Step 2			
Content Area	**Michigan N = 211**		**Foundations N = 264***		**AnKing Step 1 N = 189**		**Dope Medical Science N = 164**		**AnKing Step 2 N = 170**		**Dope Clinical Science N = 181**	
Oncology	7 (3%)	3 (2%)	10 (4%)	7 (3%)	21 (11%)	11 (6%)	33 (20%)	15 (10%)	16 (9%)	15 (10%)	19 (10%)	8 (5%)
Ophthalmology	8 (4%)	0 (0%)	3 (1%)	0 (0%)	1 (1%)	1 (1%)	0 (0%)	0 (0%)	3 (2%)	1 (1%)	4 (2%)	3 (2%)
Orthopedic Surgery	6 (3%)	1 (1%)	43 (16%)	1 (0%)	11 (6%)	2 (1%)	5 (3%)	0 (0%)	6 (4%)	4 (3%)	6 (3%)	4 (3%)
Otolaryngology	6 (3%)	0 (0%)	37 (14%)	0 (0%)	7 (4%)	2 (1%)	2 (1%)	0 (0%)	1 (1%)	1 (1%)	9 (5%)	6 (4%)
Pediatric Surgery	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)
Physical Medicine and Rehabilitation	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Plastic Surgery	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Psychiatry	11 (5%)	0 (0%)	14 (5%)	4 (2%)	7 (4%)	2 (1%)	7 (4%)	2 (1%)	9 (5%)	3 (2%)	7 (4%)	2 (1%)
Pulmonology and Acute Care	11 (5%)	5 (3%)	13 (5%)	0 (0%)	18 (10%)	3 (2%)	10 (6%)	1 (1%)	25 (15%)	5 (3%)	25 (14%)	10 (6%)
Rheumatology	14 (7%)	4 (2%)	0 (0%)	0 (0%)	9 (5%)	7 (4%)	9 (5%)	2 (1%)	13 (8%)	5 (3%)	6 (3%)	1 (1%)
Urology	8 (4%)	0 (0%)	13 (5%)	0 (0%)	6 (3%)	0 (0%)	2 (1%)	0 (0%)	4 (2%)	0 (0%)	10 (6%)	0 (0%)

Table 3.2. Quantification of content areas in electronic flashcards (EFs). All content areas ("any") and the most prominent ("primary") content area relevant to EFs reported in n (%) cards.

	Preclinical		Step 1		Step 2	
Characteristic	**Michigan N = 211**	**Foundations N = 264***	**AnKing Step 1 N = 189**	**Dope Medical Science N = 164**	**AnKing Step 2 N = 170**	**Dope Clinical Science N = 181**
Primary Content Area						
Basic Science Field	152 (72%)	237 (90%)	99 (52%)	94 (57%)	40 (24%)	55 (30%)
Clinical Field	59 (28%)	27 (10%)	90 (48%)	70 (43%)	130 (76%)	126 (70%)
Clinical Reasoning Skill						
Diagnosis	37 (18%)	24 (9%)	35 (19%)	37 (23%)	66 (39%)	63 (35%)
Etiology	18 (9%)	6 (2%)	35 (19%)	36 (22%)	17 (10%)	34 (19%)
Management	13 (6%)	3 (1%)	19 (10%)	5 (3%)	47 (28%)	29 (16%)
Prognosis	11 (5%)	4 (1%)	17 (9%)	12 (7%)	21 (12%)	9 (5%)
Ethics	0 (0%)	1 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
None	132 (63%)	226 (86%)	83 (44%)	74 (45%)	19 (11%)	46 (25%)
Knowledge Domain						
Information	185 (88%)	255 (97%)	170 (90%)	157 (96%)	118 (69%)	131 (72%)
Mental Procedure	24 (11%)	8 (3%)	19 (10%)	7 (4%)	52 (31%)	47 (26%)
Psychomotor Procedure	2 (1%)	1 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (2%)
Cognitive Process						
Retrieval	161 (76%)	241 (91%)	165 (87%)	135 (82%)	122 (72%)	130 (72%)
Comprehension	9 (4%)	4 (2%)	0 (0%)	7 (4%)	1 (1%)	1 (1%)
Analysis	40 (19%)	16 (6%)	23 (12%)	21 (13%)	26 (15%)	36 (20%)
Knowledge Utilization	1 (0%)	3 (1%)	1 (1%)	1 (1%)	21 (12%)	14 (8%)

Table 3.3. Quantification of characteristics of electronic flashcards (EFs). Reported in n (%) flashcards. Primary Content Area = the EF was most pertinent to a field with a Basic Science vs. Clinical focus; Clinical Reasoning Skill = the EF contained information directly supportive for executing a clinical reasoning task; Knowledge Domain = the EF tested knowledge in any of three categories derived from the New Taxonomy of Educational Objectives by Marzano and Kendall[69]; Cognitive Process = the EF likely elicited thinking process in any of four categories derived from the New Taxonomy of Educational Objectives by Marzano and Kendall.[69]

Chapter 4 Feasibility of Estimating Cardiac Indices Using Cardiac Surgery Anesthesia Records in a Multicenter Cohort

4.1 Abstract

Purpose: Cardiac indices (CIs) are an important, modifiable physiologic measure correlated to end-organ perfusion and complication rates among cardiac surgical patients. However, most patients do not have invasive monitors placed as necessary for computing CIs. Cardiac surgical patients with such monitors present an opportunity to evaluate the extent to which other routinely collected physiologic, patient, and procedure data may be used to predict CIs.

Method: Anesthesia records from 5,989 patients with CI measurements during cardiac surgery across 7 US institutions were available for study. Generalized linear models were constructed to estimate CI from four previously published estimators using intraoperative blood pressure and heart rate data, adjusted for perioperative patient and procedure factors. A Bland-Altman analysis was performed comparing estimates to true CI measurements. Additionally, the concordance between CI values and estimates in classifying patients as having low CI (<2.2 L/min/m^2) was measured.

Results: All four adjusted estimators of CI were moderately correlated with CI (Spearman correlation range = 0.31-0.44) and had reasonable performance in the study population (bias range = <0.0001-<0.0001 L/min/m^2; percentage error range = 4.0-7.9%; Mean Absolute Error (MAE) range = 0.45-0.48 L/min/m^2), with the Liljestrand and Zander estimator having the

strongest overall performance. Nevertheless, patient-level estimates had limited reliability in classifying patients as having low CI (Cohen's kappa = 0.20 and 0.26 for pre- and post-cardiopulmonary bypass values, respectively).

Conclusion: This multicenter study demonstrates the significant association of physiologic CI estimators and CI when controlling for routinely collected patient and procedure factors in the anesthesia record. However, CI estimates require calibration for prediction and classification at the patient level.

4.2 Introduction

Cardiac indices (CIs) are an important, modifiable physiologic measure correlated to end-organ perfusion as well as complications resulting from organ malperfusion among patients undergoing cardiac surgery.[42] Nonetheless, benefits of CI monitoring via pulmonary artery catheters (PACs) are counterbalanced by risks associated with placement of such monitors (e.g., arrhythmia, thrombus formation, pulmonary artery rupture), precluding routine placement on cardiac surgical patients.[43] Therefore, many techniques to estimate CI and cardiac output (CO) with noninvasive measures have been explored.[107,108]

While some of these techniques show clinical promise, many of the higher performing methods rely on proprietary physiological sensors or arterial waveforms which may not be easily captured in the anesthesia record and compared between different healthcare sites.[109] Studies which instead use routinely-collected heart rate and blood pressure data to estimate CI have largely shown limited reliability;[110–117] one study in a small cohort of patients undergoing coronary artery bypass graft (CABG) surgery reported a 49% error rate for its highest-performing estimator (Liljestrand and Zander).[118] The limited performance of these simple physiologic estimators is potentially explained by the complexity of physiology during cardiac surgery, with measurement variation potentially attributable to patient demographics, concurrent cardiac conditions, cardiovascular medication treatment interventions, and surgical procedure details.[119–121] Of the aforementioned studies of CI estimators, few have attempted to control for such factors, although one found that inclusion of age and sex significantly improved the performance of their CO estimates in a single center.[110]

Nonetheless, whereas physiologic estimators of CI have largely unproven utility to inform clinical care decisions of individual patients, the usefulness of such estimators for

population-level observational study remains unresolved. Specifically, opportunities may exist to explore clinical care questions leveraging existing electronic health record (EHR) data to understand fundamental relationships between low CI and malperfusion-related outcomes. In such observational study designs, CI estimators – if found to be accurate at the population level – may aid in uncovering relationships between hemodynamic management interventions and outcomes mediated by low CI, even in the absence of direct CI measurements.

To address the limitations of previous studies and explore this line of clinical scientific inquiry, we performed this multicenter observational study of cardiac surgical patients with direct CI measurements undergoing a variety of procedures. We hypothesized that simple physiologic CI estimators will be strongly and significantly associated with CI when adjusted for important perioperative patient and procedure information which is routinely available in the anesthesia record.

4.3 Method

4.3.1 Study Design

This study was deemed exempt by the University of Michigan IRBMED (HUM00206345). The inclusion/exclusion criteria, data collection/cleaning methods, and statistical approaches of the study were approved a priori within a peer-review forum [122]. The study manuscript adheres to the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) extension of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.[123]

4.3.2 Data Source

Data were obtained from the Multicenter Perioperative Outcomes Group (MPOG) and Society of Thoracic Surgeons Adult Cardiac Surgical Database (STS-ACSD) registries. Procedures for acquiring, validating, and transferring data from participating sites to MPOG and STS-ACSD have been previously described.[124–126] Additionally, MPOG and STS-ACSD handled missing or invalid data according to registry-specific protocols.[127,128] Data from seven US hospitals who participate in both MPOG and STS-ACSD were integrated locally for this study.

4.3.3 Study Population

Adult patients undergoing cardiac surgery from January 1, 2014 to February 1, 2022 with at least one intraoperative CI measurement were included. Cardiac surgical procedures included coronary artery bypass grafts (CABG), valve repairs or replacements, thoracic aortic procedures, or combinations thereof. The index case was used when patients underwent multiple procedures in the study period. A complete case analysis was performed, due to low rates of missingness (<95%) among model covariates. Additional exclusions are detailed in unpublished manuscript by Mathis et al.[129]

4.3.4 Outcome Variable - Cardiac Index

The primary outcome was CI. Each case was randomly downsampled to two intraoperative CI measurements (one before cardiopulmonary bypass [pre-CPB] and one after [post-CPB] when available, or a single CI measurement when only pre-CPB or post-CPB measurements were available. Start and end times of CPB were extended by 5 minutes to account for possible data entry errors and exclude any measurements taken during CPB. Measurements of CI and CO were obtained from the EHR as recorded from PA catheter-derived measurements. CO measurements obtained from these sources were converted to CI measurements by normalizing by estimates of body-surface area [130,131]. CI measurements were

assessed for normality, symmetry, and potential outliers with histograms, Q-Q plots, violin plots, and descriptive statistics.

4.3.5 Exposure Variable - CI Estimators

Four estimators of CI were computed from the following routinely available intraoperative hemodynamic factors: systolic blood pressure (SBP), mean arterial pressure (MAP), diastolic blood pressure (DBP), and heart rate (HR) and normalized by body surface area as analogous to CO versus CI measurements (Table 4.1). Measurements of these hemodynamic factors were obtained in the same minute of cardiac index measurement. Where multiple measurements were found in a single minute, the median value was used. Rarely, when no hemodynamic measurements were obtained that minute, the most recent measurement was carried forward.

4.3.6 Risk Adjustment Variables - Patient, Case, and Intraoperative Factors

In addition to CI estimators, to account for possible hospital-level variation in PAC manufacturer used and anesthesiologist-level variation in CI measurement techniques, each CI measurement was adjusted by hospital of case origin and primary anesthesia attending (Table 4.2). Additional control variables included (i) patient baseline characteristics (age, sex, race, preoperative left ventricular ejection fraction, the presence of cardiogenic shock before surgery, and type of surgical procedure(s)), (ii) intraoperative hemodynamic factors (central venous pressure, O2 saturation), and (iii) intraoperative vasoactive-inotropic score (VIS).[132] Intraoperative hemodynamic factors and current doses of vasopressors and inotropes (for VIS) were obtained in the same minute of CI measurement, and processed analogously to CI estimators. Preoperative patient and surgical procedure data were collected via the STS-ACSD;

additional pre- and intraoperative data were acquired from MPOG using validated, open-source phenotypes.[127]

4.3.7 Statistical Analyses

Statistical analyses were performed using SAS version 9.4 (SAS Institute, USA) and R version 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria). Model selection followed a multi-criteria approach, using a combination of clinical expertise and empirical selection with least absolute shrinkage and selection operator (LASSO). A multilevel analysis approach was selected after evaluating the modified intraclass correlation estimates of hospital and primary anesthesia attending on the total variation of CI. Both levels were retained for analysis after >1% of the total variance were explained within each level. Generalized linear mixed models were constructed to estimate the adjusted cardiac index using the estimators in Table 4.1, adjusted for pre- and intraoperative patient and case factors. The final model specification is as follows:

Adjusted CI = CI estimator + patient- and case-level covariates + (1 | hospital) + (1 | primary anesthesia attending) + error

A Bland-Altman analysis was performed for each model-adjusted CI estimator using the 95% confidence interval of the limits of agreement ($\pm 2.093\sqrt{3}/n^{*}s$; s = is the standard deviation of the bias; n = sample size) [133]. Performance of the model-produced CI estimates were additionally assessed with an error analysis that included median absolute percentage error (MAPE) and mean absolute error (MAE).[134] The ability of CI estimates to discriminate between clinically low CI (<2.2 L/min/m^2) and normal or high CI was assessed with sensitivity,

specificity, positive predictive value (PPV), negative predictive value (NPV), and Cohen's kappa.[96]

4.4 Results

Of the 24,726 cardiac surgical cases eligible for inclusion, 7,169 (30.0%) contained at least one CI measurement, among which 5,989 (83.7%) contained complete multivariable data estimating CI values across 7 US hospitals. The cohort included 30.4% female and 85.3% white patients with a median age of 65 years (interquartile range 55-72). Non-mutually exclusive procedures included 26.0% thoracic aortic, 32.6% CABG, and 68.7% valve repair/replacement. Additionally, the analysis cohort included patients with variable preoperative cardiac function (LVEF median 58%, interquartile range 45-65%), including 4.0% of patients with cardiogenic shock.

Among cases with complete data, 2,054 (34.3%) included CI values prior to CPB and 5,618 (93.8%; non-mutually exclusive) after CPB (**Error! Reference source not found.**). CI models adjusted for perioperative factors converged for all four estimators and improved upon the null model; full model results can be found in Appendix E. Notably, the Windkessel and Herd estimators were perfectly correlated (1.00) with each other and highly correlated (0.89) with the Liljestrand and Zander estimator and, therefore, produced similar model results (full correlation matrix in Table 4.4). Hospital explained 8.00-9.62% of variation in CI measurements across all models and primary anesthesia attending explained a further <0.0001-2.26%. Three of four estimators (excluding MAP) were statistically significant ($p < 0.05$ by unpaired t-test) in all pre- and post-CPB models. Significant covariates across all pre- and post-CPB models included: age, sex, race, preoperative LVEF, and VIS. Significant covariates in only the post-CPB models included aortic procedure type. Significant covariates in only the pre-CPB models included

presence of preoperative cardiogenic shock and CABG procedure type (excluding MAP estimator model).

Bland-Altman and error analysis are detailed in Table 4.5. Similar means and biases of predictions were observed across estimators for pre-CPB (median range 2.20-2.22; bias range <0.0001-<0.0001) and post-CPB (mean range 2.50-2.50; bias range <0.0001-<0.0001) models. Spearman correlations were weak in the pre-CPB models (range 0.31-0.38) and weak to moderate in the post-CPB models (range 0.38-0.44). Models had a low percent error with a 95% confidence interval in both the pre-CPB models (range 4.0-4.3) and post-CPB models (range 7.6-7.9). Similarly, error analysis of CI estimates showed a low MAPE (pre-CPB range 16.2-16.8; post-CPB range 15.4-15.7) and MAE (pre-CPB range 0.45-0.47; post-CPB range 0.47-0.49). Overall, the Liljestrand and Zander estimator had the strongest performance across three of four comparison measures. The Bland-Altman plot for Liljestrand and Zander can be found Figure 4.1 and the remaining estimators in Figure 4.2.

Validity of CI estimates at discriminating patients with low (<2.2 L/min/m^2) vs. not low CI was assessed (Table 4.6). Confusion matrices for each model can be found in Table 4.7. Agreement between CI estimates and CI measurements was low (kappa <0.4) across all models pre- and post-CPB. Similarly to the Bland-Altman and error analysis, the Liljestrand and Zander estimator had marginally better performance than other estimators in classifying patients as having low CI.

4.5 Discussion

In this study of 5,989 cardiac surgical cases across 7 US hospitals, we investigated the feasibility of estimating CI using data routinely available in the anesthesia record. We report that four different estimators (MAP, Windkessel, Herd, and Liljestrand and Zander) are

independently associated with CI in models that adjust for clinically relevant patient and procedure characteristics. Therefore, these estimators could enable future efforts to estimate CI among patients both with and without direct CI measurements. All four CI estimators had strikingly similar performance, but the Liljestrand and Zander estimator had the strongest Spearman correlation, MAPE, and MAE. The strength of the Liljestrand and Zander estimator compared to other CI estimators has been reported previously, and attributed to the Liljestrand and Zander estimator capturing the nonlinear relationship between vascular compliance and arterial pressure.[118]

Adjustment for expert-chosen patient and procedural factors improved model performance over the null model. Future studies should consider inclusion of similar factors when generating CI estimates. Additionally, several of these factors were significantly independently associated with CI, including the presence of cardiogenic shock, a variable frequently omitted from previous studies due to their exclusion of unstable patients. These significant associations between patient/procedural factors and CI may inform further exploration of the exact relationships between variable presentations of real-world patients and cardiac physiology.

Overall, our multilevel generalized linear models provide adequate estimations of CI in our study population, with low bias, percent error, and MAE. This is further supported by the MAPE, which is lower across all models than the reported difference detection threshold (20-26%) between serial thermodilution measurements.[135] However, as shown in the Bland-Altman plots, our models typically underestimate high and overestimate low CI measurements. Therefore, as expected, our estimates have limited value per Cohen's kappa in discriminating between patients with low vs. normal CI values–a determination with possible clinical and

prognostic value for cardiac surgery patients.[136] For our models to allow for accurate prediction and classification at the patient-level, they would benefit from calibration with an external dataset through methods like those described in Mentz et al.[137]

Our study leveraged a representative, multicenter cohort of cardiac surgery patients to identify physiological estimators, patient factors, and procedure factors which may inform future modeling efforts to estimate CI among a wider variety of patients than those who typically receive PAC-based measurements. Nonetheless, the inclusion of cardiac surgical cases alone may limit generalization of our findings to other settings for which CI may be a valuable diagnostic or prognostic tool, including other surgical and critical care domains. Additionally, our cohort has limited representation of women, minoritized racial groups, and the elderly (>80 years), and may generate inaccurate or inequitable CI estimates for these groups. Finally, while our model performed well across 7 US healthcare centers, we did not control for, e.g., different vendors, manufacturers, and models for CI measurement devices among these sites, and may have unmeasured confounding as a result.

4.6 Conclusion

This study explores the feasibility of CI estimation in a large, multicenter, representative cardiac surgery cohort using routinely collected data in the anesthesia record. Models which incorporated the Liljestrand and Zander formula for estimating CI with blood pressure and heart rate performed the strongest. All models benefited from adjustment of CI estimates with patient and procedural factors. Overall, our approach provides a strong foundation for future modeling efforts to estimate CI across a variety of cardiac surgery patients for inclusion in future observational studies that explore avoidance of hypoperfusion and adverse surgical outcomes.

4.7 Figures

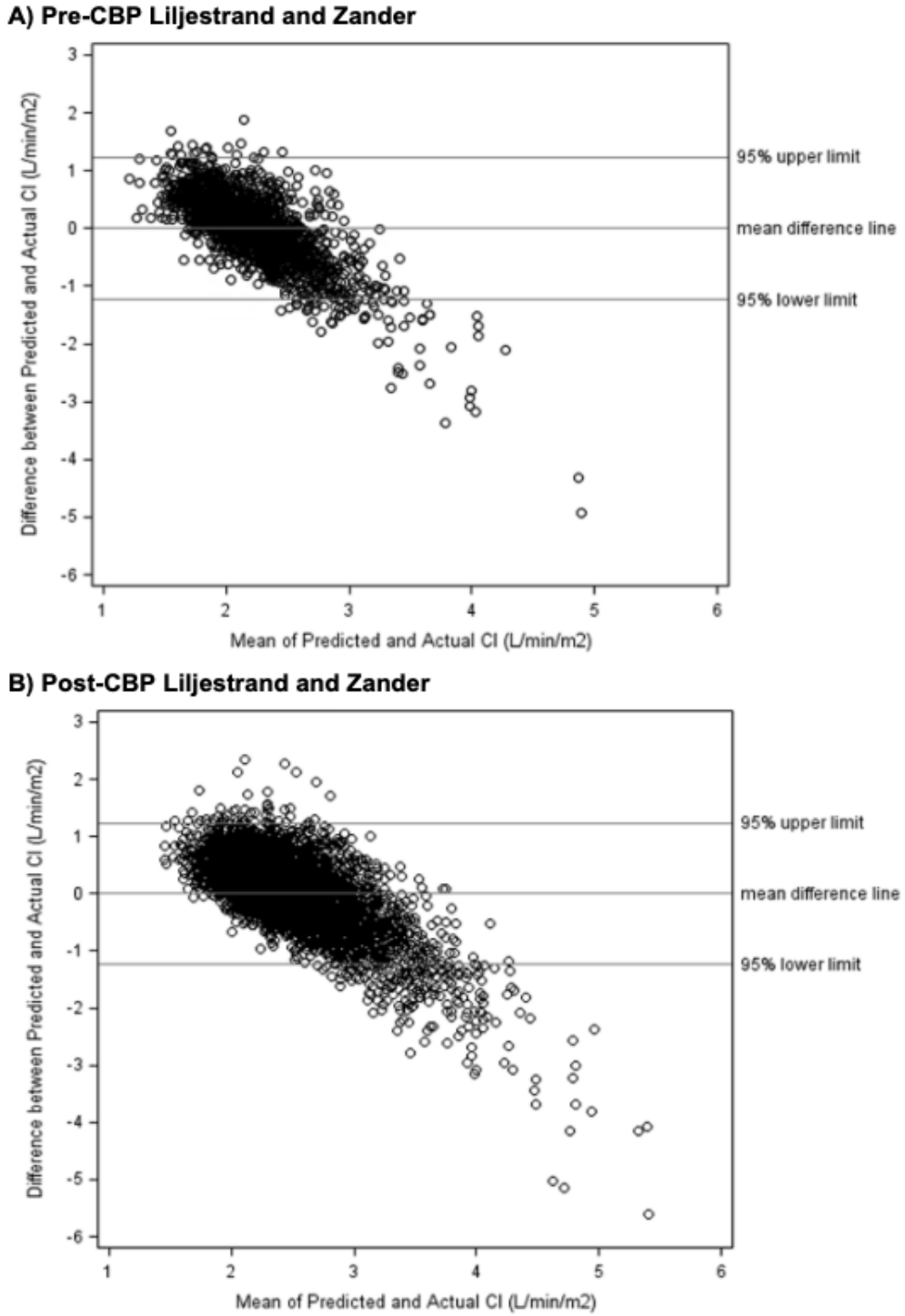
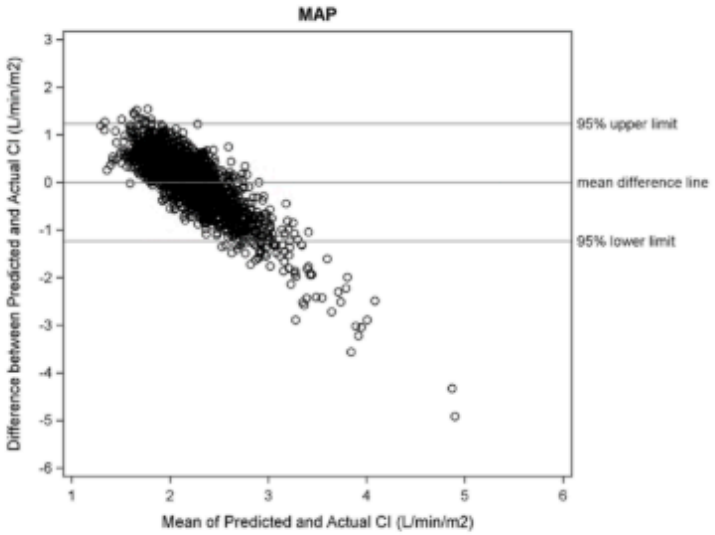
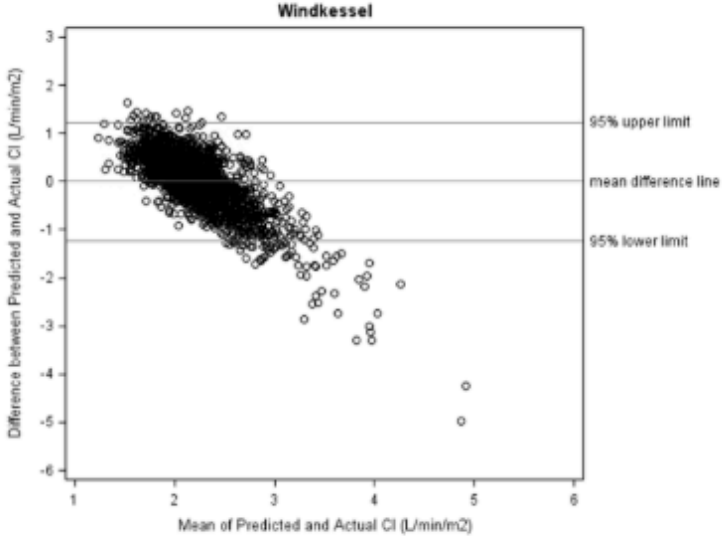


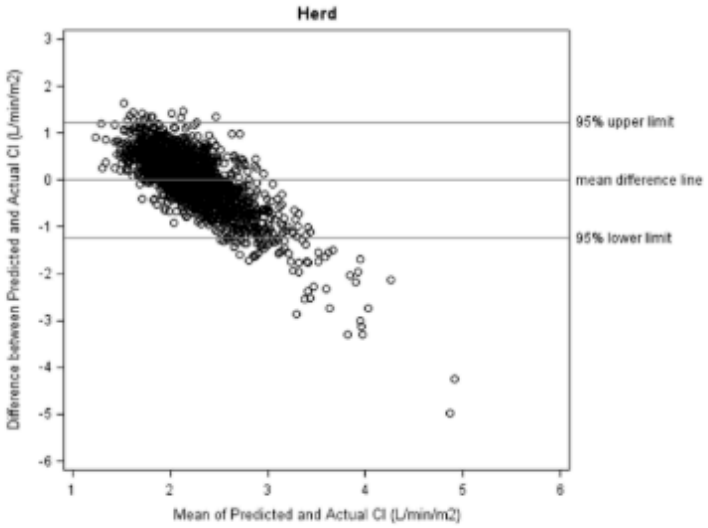
Figure 4.1. Bland-Altman plots for the Liljestrand and Zander estimator pre- and post-CPB.

CPB = cardiopulmonary bypass.

A) Pre-CBP


MAP
Difference between Predicted and Actual CI (L/min/m2)
Mean of Predicted and Actual CI (L/min/m2)
95% upper limit
mean difference line
95% lower limit


Windkessel
Difference between Predicted and Actual CI (L/min/m2)
Mean of Predicted and Actual CI (L/min/m2)
95% upper limit
mean difference line
95% lower limit


Herd
Difference between Predicted and Actual CI (L/min/m2)
Mean of Predicted and Actual CI (L/min/m2)
95% upper limit
mean difference line
95% lower limit

B) Post-CBP

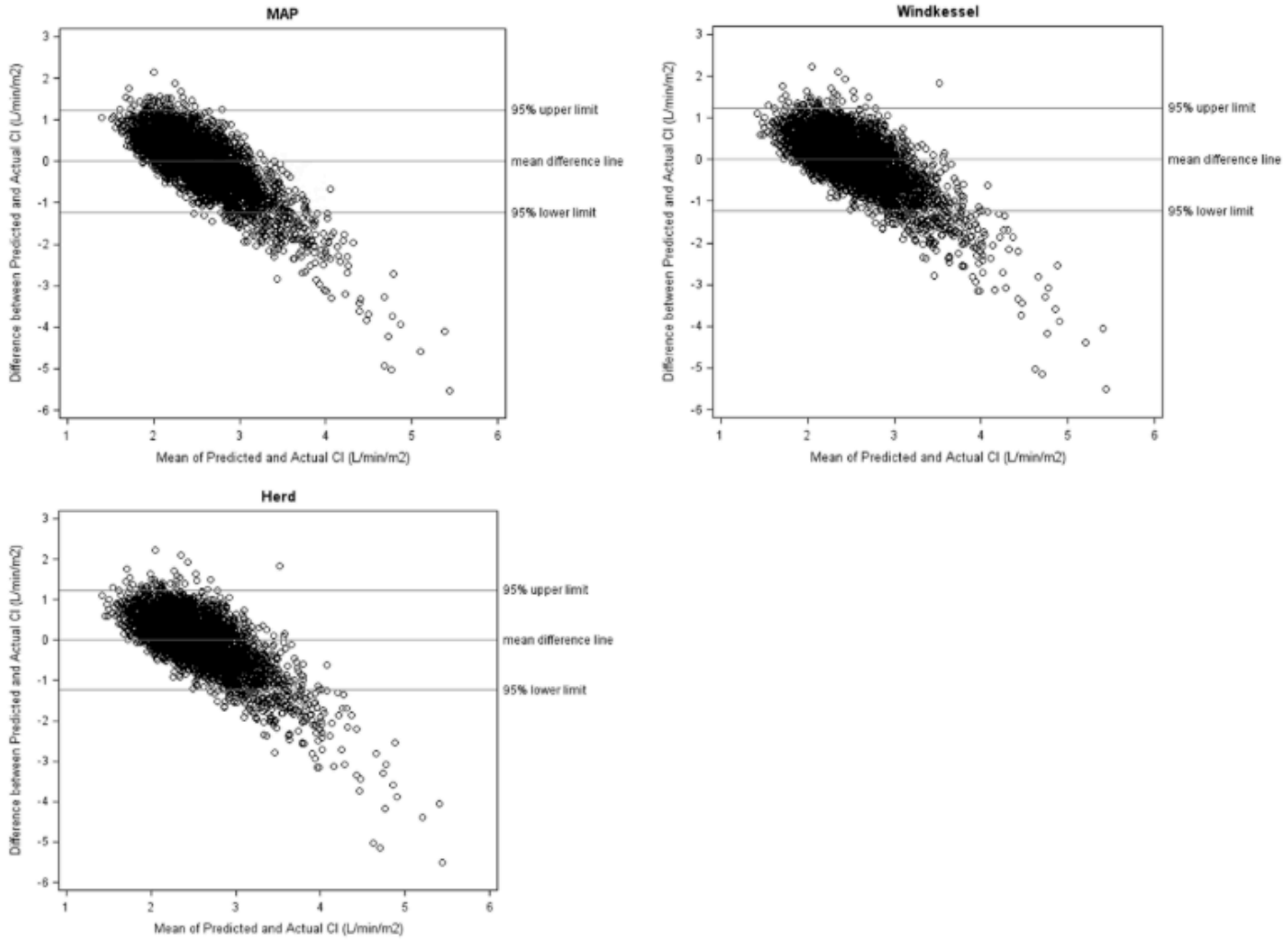


Figure 4.2. Bland-Altman plots for remaining estimators pre- and post-CPB.

CPB = cardiopulmonary bypass.

4.8 Tables

Estimator of CI	**Formula**
Mean Arterial Pressure (MAP)	(1/3 SBP + 2/3 DBP) / BSA
Windkessel [138]	(SBP - DBP) * HR / BSA
Herd [139]	(MAP - DBP) * HR / BSA
Liljestrand and Zander [113]	(SBP - DBP) / (SBP + DBP) * HR / BSA

Table 4.1. Evaluated estimators of cardiac index. CI = cardiac index; SBP = systolic blood pressure; DBP = diastolic blood pressure; MAP = mean arterial pressure; HR = heart rate; BSA = body surface area.

Hospital	Pre-CPB Patients n (%)	Post-CPB Patients n (%)	Pre-CPB Primary Anesthesia Attending n (%)	Pre-CPB Primary Anesthesia Attending n (%)
A - 84	1,219 (59)	3,919 (70)	35 (22)	36 (18)
B - 36	598 (29)	613 (11)	45 (28)	46 (23)
C - 19	73 (4)	496 (9)	29 (18)	38 (19)
D - 57	105 (5)	193 (3)	22 (14)	23 (11)
E - 23	22 (1)	271 (5)	10 (6)	22 (11)
F - 78	33 (2)	45 (1)	15 (9)	16 (8)
G - 58	4 (0)	81 (1)	3 (22)	22 (11)

Table 4.2. Number of cases and primary anesthesia attendings by study hospital. Participating health systems are Brigham and Women's Hospital, Boston, Massachusetts; Massachusetts General Hospital, Boston, Massachusetts; Oregon Health and Science University, Portland, Oregon; University of Michigan Health System, Ann Arbor, Michigan; University of Washington Medical Center, Seattle, Washington; Washington University of St. Louis School of Medicine, St. Louis, Missouri; Yale New Haven Hospital, New Haven, Connecticut.

Variable	Pre-CPB (n = 2,054)	Pre-CPB (n = 5,618)
	n (%)	n (%)
Cardiac Index (median [IQR], L/min/m2)	2.1 [1.8 to 2.6]	2.4 [2.0 to 2.9]
Cardiac Index estimator		
MAP (median [IQR], L/min/m2)	36.7 [32.0 to 42.0]	34.8 [30.6 to 39.3]
Windkessel (median [IQR], L/min/m2)	1,624.9 [1,274.0 to 2,118.8]	1,909.0 [1500.7 to 2,373.5]
Herd (median [IQR], L/min/m2)	541.6 [424.7 to 706.2]	636.3 [500.2 to 791.2]
Liljestrand and Zander (median [IQR], L/min/m2)	10.0 [8.0 to 12.5]	12.2 [9.9 to 15.0]
Age (median [IQR], years)	65 [55 to 72]	65 [55 to 72]
Sex; male	1,430 (69.6%)	3,912 (69.6%)
Race		
White	1,757 (85.5%)	4,793 (85.3%)
Black	152 (7.4%)	412 (7.3%)
Other/Multiracial	145 (7.1%)	413 (7.4%)
Type of cardiothoracic procedure		
Aortic	497 (24.2%)	1,467 (26.1%)
CABG	758 (36.9%)	1,835 (32.7%)
Valve	1,356 (66.0%)	3,869 (68.9%)
Cardiac Shock	78 (3.8%)	219 (3.9%)
Preoperative LVEF (median [IQR], %)	58 [45 to 64]	58 [44 to 65]
CVP (median [IQR], cm H2O)	11 [8 to 15]	9 [6 to 14]

Table 4.3. Pre- and post-CPB cohort characteristics.

CPB = cardiopulmonary bypass; MAP = mean arterial pressure; CABG = coronary artery bypass graft; LVEF = left ventricular ejection fraction; CVP = central venous pressure.

	MAP	Windkessel	Herd	Liljestrand and Zander
MAP	1.00	0.49	0.49	0.18
Windkessel	0.49	1.00	1.00	0.89
Herd	0.49	1.00	1.00	0.89
Liljestrand and Zander	0.18	0.89	0.89	1.00

Table 4.4. Correlations between estimators. MAP = mean arterial pressure.

	MAP		Windkessel		Herd		Liljestrand and Zander	
	Pre-CPB	Post-CPB	Pre-CPB	Post-CPB	Pre-CPB	Post-CPB	Pre-CPB	Post-CPB
Median (L/min/m2)	2.22	2.50	2.21	2.50	2.21	2.50	2.20	2.50
Bias +/- SD (L/min/m2)	<0.001 (0.63)	<0.001 (0.66)	<0.001 (0.62)	<0.001 (0.65)	<0.001 (0.63)	<0.001 (0.66)	<0.001 (0.61)	<0.001 (0.64)
Spearman Correlation	0.31	0.38	0.36	0.42	0.31	0.40	0.38	0.44
Percent Error (%)	4.0	7.6	4.0	7.8	4.0	7.6	4.3	7.9
Median Absolute Performance Error (%)	16.8	15.7	16.5	15.6	16.5	15.5	16.2	15.4
Mean Absolute Error (L/min/m2)	0.47	0.49	0.46	0.48	0.46	0.48	0.45	0.47

Table 4.5. Bland-Altman and error analysis.

Best performing values for each comparison metric highlighted in gray. MAP = mean arterial pressure; CPB = cardiopulmonary bypass; SD = standard deviation.

	MAP		Windkessel		Herd		Liljestrand and Zander	
	Pre-CPB	Post-CPB	Pre-CPB	Post-CPB	Pre-CPB	Post-CPB	Pre-CPB	Post-CPB
Sensitivity	0.57	0.24	0.61	0.28	0.61	0.28	0.62	0.31
Specificity	0.66	0.89	0.65	0.88	0.65	0.88	0.65	0.87
PPV	0.67	0.55	0.68	0.58	0.68	0.58	0.68	0.59
NPV	0.56	0.67	0.58	0.68	0.58	0.68	0.58	0.68
Cohen's Kappa	0.23	0.14	0.26	0.185	0.26	0.19	0.26	0.2

Table 4.6. Agreement of true CI and estimated CI for low CI <2.2 L/min/m2.

Best performing values for each comparison metric highlighted in gray. MAP = mean arterial pressure; CPB = cardiopulmonary bypass; PPV = positive predictive value; NPV = negative predictive value.

	Pre-bypass				Post-bypass			
MAP			Predicted				Predicted	
			Normal	*Low*			*Normal*	*Low*
	Actual	*Normal*	608	317	Actual	*Normal*	3143	398
		Low	484	645		*Low*	1583	494
Windkessel			Predicted				Predicted	
			Normal	*Low*			*Normal*	*Low*
	Actual	*Normal*	605	320	Actual	*Normal*	3119	422
		Low	445	684		*Low*	1486	591
Herd			Predicted				Predicted	
			Normal	*Low*			*Normal*	*Low*
	Actual	*Normal*	605	320	Actual	*Normal*	3119	422
		Low	445	684		*Low*	1486	591
Liljestrand and Zander			Predicted				Predicted	
			Normal	*Low*			*Normal*	*Low*
	Actual	*Normal*	599	326	Actual	*Normal*	3094	447
		Low	430	699		*Low*	1441	636

Table 4.7. Confusion matrices for low cardiac index (CI) classification. MAP = mean arterial pressure. ff

4.9 Appendices

4.9.1 Appendix E. Full cardiac index estimation model results.

Complete model results for all estimators of cardiac index (CI) before and after cardiopulmonary bypass may be found in the DeepBlue Data Repository at the following URL:

https://hdl.handle.net/2027.42/196329

Chapter 5 Conclusion

5.1 Overview

This thesis situates itself within the sweeping changes that the data era has brought to the role of physicians. We suggest that several ingredients are necessary to support physicians through their evolving role as "information managers" in addition to patient care providers. These ingredients include changes to medical education by incorporating strategies for high-volume knowledge acquisition, but not at the expense of the development of clinical reasoning and higher-order thinking skills needed to manage information. These ingredients also include changes to clinical practice by incorporating tools to help clinicians manage information, but only if those tools demonstrate sufficient value to providers and patients.

In this integrated discussion, we first summarize the key findings from Chapters II-IV which detail two projects in medical education and cardiac anesthesiology clinical practice. We then explore the implications of these projects to facilitate the changes necessary to support physicians in their roles as "information managers." Throughout this discussion, we highlight the impact of our work to the broader fields and theories which inspired it and discuss future directions for continuing our work's impact.

5.2 Summary of Key Findings

In Chapters II and III, we describe several contributions to the field of medical education and medical education research regarding electronic flashcards (EFs).

In Chapter II, we propose the Design Elements of Anki for Medical Education (DEAME) Framework to enumerate design features in a sample of EFs for UME. This framework contains seven categories that describe overarching themes of flashcard design and 48 codes that provide example design choices for each theme. This framework offers the first terminology to describe the significant heterogeneity in design that EFs can display. Additionally, the DEAME Framework represents a novel connection between theories of learning and EFs, through mapping the types of knowledge and ways of thinking likely tested by EFs to those described by the New Taxonomy of Educational Objectives (NT).[69] Finally, mapping assessment items to the NT post hoc is a novel application of the NT, and an application we largely found possible and successful through deduction and qualitative consensus building.

In Chapter III, we offer the first comprehensive and empirical evaluation of the content of learner-created EF decks for undergraduate medical education (UME). Through application of the DEAME Framework in Chapter II, we show that a sample of EF decks for UME tests a variety of knowledge in a variety of ways. While some cards contain clinically relevant and complex knowledge which may support skills like clinical reasoning with diverse sources of information, we find the overall prevalence of such cards low.

In Chapter IV, we describe contributions to the fields of medical informatics and cardiac anesthesiology regarding estimation methods for cardiac index (CI). Using generalized linear models, we show that several noninvasive, readily available physiologic estimators for CI are independently associated with CI after adjustment for relevant patient and procedure factors in a representative, multicenter cohort of cardiac surgery patients. This suggests that future CI monitoring efforts could leverage these physiologic estimators, instead of the current gold-standard invasive measures. Additionally, the statistical significance of many patient and

procedure factors in these models suggests different patient contexts explain some variability in CI measurements.

The following sections outline emergent themes from the contributions of this dissertation to physician education and practice in the data era.

5.3 EFs test a variety of knowledge in a variety of ways, including ways which may support knowledge acquisition at scale.

Previous studies have suggested that EFs are a learning strategy sought out by medical learners to facilitate learning of vast amounts of knowledge.[9] Some educators suggest that EFs fulfill this role by atomizing large concepts into smaller ones.[18] Additionally, some students perceive EFs to be most useful for knowledge acquisition when they contain high-quality questions.[19,28] However, existing literature offers limited insight to designing EFs with either of these qualities–knowledge atomization or high-quality questions. Nonetheless, several categories in the DEAME framework proposed in Chapter II suggest that EFs might be tailored to support knowledge acquisition at scale.

In regard to atomization of knowledge, the Mental Procedures category of DEAME framework shows that complex groups of procedural knowledge like Algorithms and Tactics can be broken down into their smallest possible component pieces: Single Rules. Additionally, while EFs may atomize knowledge for efficient testing, they also appear to offer mechanisms to connect atomized pieces of knowledge, should the learner need such connections to learn or integrate that knowledge. For example, flashcards may have metadata in the form of Deck/Subdeck Structure and Metadata Tags which organize related flashcards. Additionally, flashcards may contain answer explanations in the form of Extra Components which may situate

knowledge within a larger context, including through Links to External Resources which connect students to educational videos and other materials related to individual flashcards.

More broadly than knowledge atomization, the framework suggests EFs may offer other strategies to design high-quality EFs for knowledge acquisition at scale. Namely, EFs appear to demonstrate significant design flexibility, as evidenced by the many categories and codes within the DEAME framework. This suggests that different design choices could be chosen to optimize testing of certain knowledge. For example, testing of procedural knowledge could benefit from highly multimedia EFs which replicate the physical world[140]—such as videos demonstrating physical exam findings. Additionally, EFs' electronic format and significant flexibility may enable the creation of new types of questions and other design features. For example, the Image Selection question format could be reversed such that a structure name and image are provided on the card front and an outline of that structure is overlaid on the image on the card back. In that way, a user would be responsible not only for identifying a structure, but also for understanding where it begins and ends in relation to surrounding structures. This potential new Question Format might hold particular value for concepts like neuroanatomy, which may be challenging for students given the relative lack of discrete boundaries in the brain relative to other tissues.[141]

The DEAME framework supports several future directions to understand design principles of high-quality cards that might support knowledge at scale. Namely, the framework's terminology can be used as the foundation for future studies to selectively test the impact of design choices on user experiences and outcomes. For example, researchers could perform A/B testing to evaluate how quickly students learn and retain information presented by flashcards with and without certain design elements.[142] Because EF platforms like Anki generally contain highly empirical and granular usage metrics,[92] these experiments could be effectively performed

without heavy reliance on protected educational data. We hypothesize that a small number of card formats, which aggregate several intentional design choices, will be most effective in creating or editing flashcards for specific learning goals or groups of learners. These formats can then inform best-practice guidelines for EF creation by either learners or experts. Other pedagogical and assessment tools, like summative exams, are intentionally created with mixes of different content, question types, explanations, and other features to elicit certain outcomes in target populations;[48] similarly, EFs should be approached with a critical, intentional, and empirical mindset to serve the wide student population who use them.

5.4 EFs can test clinical and complex knowledge, but may need expert guidance and data management tools for EF curation to adequately encourage skill development.

As a frequently learner-driven and -created resource, EFs may not prioritize the "right" knowledge for learning, a feature of significant importance to both learners and educators.[13,28–31] In particular, prior literature argues that EFs may not adequately facilitate the development of skills necessary for clinical practice.[32] In the data era, skills of particular importance include higher-order thinking skills which require integration of different streams of information.[6] For example, borrowing from the clinical context in Chapter IV of this dissertation, cardiac anesthesiologists wishing to manage patients' cardiac function must not only consider direct measures like cardiac index (CI) but also patients' broader clinical contexts; as shown in our models, many other factors may explain some variability in cardiac function, like demographics, medication usage, and type of operative procedure. Weighing diverse and variable information requires advanced reasoning skills honed through education and clinical experience.

As the DEAME Framework shows in Chapter II, EFs might serve as an educational intervention to develop procedural knowledge and higher-order thinking skills which underlie

such advanced clinical reasoning. Nonetheless, in Chapter III of this thesis, we find that popular, learner-created EF decks may underutilize EFs' capacity to encourage such development. As learners spend increasingly more time using EFs than other resources, including curricular resources which increasingly prioritize competencies like advanced clinical reasoning, this may become a cause for concern.

In order to ensure EFs develop competencies in addition to allowing knowledge acquisition, educators may wish to formally align EFs, and particularly learner-created EFs, with curricular objectives. The connections between the NT and EFs which Chapter II discusses provides an important substrate for future work to assess and engineer such alignment. However, given the size of EF decks, educators may struggle to devote the time and resources necessary to ensure such alignment. Indeed, Spicer et al reported that their expert curation of a learner-created deck and other evidence-based learning materials was "time-intensive" and required at least 1 hour of curation for every 1 lecture hour on which materials were based.[13] Therefore, creation of high-quality, curriculum-aligned EFs—which may enhance education in the data era due to their breaking down of concepts into many small pieces—may be significantly hampered by this very characteristic.

To encourage the advancement of EFs as a learning tool through curricular alignment and prioritization of complex and clinical thinking, educators may need to incorporate novel data management tools. For example, generative AI tools have shown initial promise in categorizing and creating EFs. Pendergrast and Chalmers describe an Anki add-on called AnkiTagger that assigns metadata tags to EFs to group them by relevance to the content of didactic lectures.[143] They report that generative AI was highly effective at this task, reporting that >90% of the EFs appeared "highly relevant" to the lectures to which they were assigned. Additionally, several

add-ons like AnkiBrain have been published which allow EF creation via prompts to generative AI.[144] Therefore, educators wishing to align EFs might first use AnkiTagger to identify previously created cards which relate to their curricula and then use AnkiBrain to create new cards missing for adequate alignment. Therefore, generative AI likely presents a powerful tool for managing the size and scope of EFs. Indeed, for EFs to realize their full potential in training the physician workforce, the status quo of EF creation management must increasingly extend beyond learners themselves to encompass educators and advanced technologies.

5.5 Noninvasive cardiac index estimation shows initial promise, but needs future work to provide value at the patient level.

Like many other kinds of clinicians, anesthesiologists must manage streams of information to perform clinical care tasks.[36] Tools like CDS systems and dashboards have arisen to support clinicians in this information management, but have often been shown to offer limited value to patient outcomes.[38,39] In Chapter IV, we propose foundational work that aims to increase the value of CI monitoring to patients by estimating CI with noninvasive physiological measurements, instead of invasive direct measurements. We report that our associational model suggests that these noninvasive measurements are highly correlated with CI, even after controlling for patient and procedure factors. However, our model has limited predictive value at the patient level and would not yet offer value to clinicians seeking to understand their patients' cardiac functioning.

In order to provide more accurate, valuable estimation at the patient level, several future approaches could be explored. To increase the predictive capacity of our model at the individual level, we could calibrate our estimates with a secondary dataset.[137] Additionally, we could expand the covariate space to include other variables relevant to CI estimation. Literature

suggests that additional measures of oxygen consumption and carbon dioxide production could be likely candidates, some of which can be obtained with noninvasive monitors built into standard anesthesia systems.[145,146] Finally, we could implement other functional forms for the linear regression model—or consider other statistical or machine learning modeling approaches altogether. Among the studies which use similar noninvasive CI physiologic estimators, the random forest models by Ke et al showed among the lowest error.[110] Notably, all of these approaches aim to improve the value of CI estimation from the modeling side. Nonetheless, even when all possible modeling-focused approaches are employed to improve model performance, all estimates produced by even the best specified and most accurate models are exactly that—estimates. In other words, all output from models is imperfect, and must be interpreted with consideration of the models' biases, assumptions, etc. Indeed, a key challenged to "safe adoption of artificial intelligence/machine learning-based methods in perioperative learning health systems" includes "wide variations in the available clinician knowledge base regarding strengths and limitations" of such methods.[147]

Therefore, to maximize the value of advanced modeling approaches in the data era, educating our medical workforce to understand and robustly reason with the outputs of these models is of foremost importance. Initial efforts to teach ML/AI are underway, and will require multifaceted educational approaches—including active learning modalities like EFs—to support competencies in this space.[5,148]

5.6 Final Thoughts

This thesis embraces a diversity of research techniques applied to both medical education and clinical practice. Across both projects outlined in this thesis, we explore how the whole physician lifecycle, from education to practice, must be considered when preparing physicians

for their roles as “information managers” in the data era. Additionally, the themes emerging from these projects show how the convergence of novel technologies and evergreen human factors presents both challenges and paths forward to such preparation. We hope that this research output supports future advances in medical education and clinical practice but also underscores the concerted need for interdisciplinary research to connect these silos.

Bibliography

1. Hilbert M, López P. The world's technological capacity to store, communicate, and compute information. *Science*. 2011;332(6025):60-65. doi:10.1126/science.1200970

2. Liu Y, Chen Y, Han L. Bioinformatics: Advancing biomedical discovery and innovation in the era of big data and artificial intelligence. *The Innovation Medicine*. 2023;1(1):100012. doi:10.59717/j.xinn-med.2023.100012

3. Mallappallil M, Sabu J, Gruessner A, Salifu M. A review of big data and medical research. *SAGE Open Med*. 2020;8:2050312120934839. doi:10.1177/2050312120934839

4. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. *Nat Med*. 2020;26(1):29-38. doi:10.1038/s41591-019-0727-5

5. Alrassi J, Katsufrakis PJ, Chandran L. Technology can augment, but not replace, critical human skills needed for patient care. *Acad Med*. 2021;96(1):37-43. doi:10.1097/ACM.0000000000003733

6. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. *Acad Med*. 2018;93(8):1107-1109. doi:10.1097/ACM.0000000000002044

7. Tulk Jesso S, Kelliher A, Sanghavi H, Martin T, Henrickson Parker S. Inclusion of clinicians in the development and evaluation of clinical artificial intelligence tools: A systematic literature review. *Front Psychol*. 2022;13:830345. doi:10.3389/fpsyg.2022.830345

8. González-Márquez R, Schmidt L, Schmidt BM, Berens P, Kobak D. The landscape of biomedical research. *Patterns (N Y)*. 2024;5(6):100968. doi:10.1016/j.patter.2024.100968

9. Finn E, Ayres F, Goldberg S, Hortsch M. Brave new E-world: Medical students' preferences for and usage of electronic learning resources during two different phases of their education. *FASEB Bioadv*. 2022;4(5):298-308. doi:10.1096/fba.2021-00124

10. Coda JE. Third-Party Resources for the USMLE: Reconsidering the Role of a Parallel Curriculum. *Acad Med*. 2019;94(7):924. doi:10.1097/ACM.0000000000002722

11. Burk-Rafel J, Santen SA, Purkiss J. Study Behaviors and USMLE Step 1 Performance: Implications of a Student Self-Directed Parallel Curriculum. *Acad Med*. 2017;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

12. Yao WC, Leong KH, Chiu LT, et al. The trends in the incidence and thrombosis-related comorbidities of antiphospholipid syndrome: a 14-year nationwide population-based study. *Thromb J*. 2022;20(1):50. doi:10.1186/s12959-022-00409-8

13. Spicer JO, Ujunwa KC, Witt LS, Meka J, Gooding HC. Using instructor-developed study resources to increase evidence-based learning strategies among medical students: A mixed-methods study. *Med Teach*. 2023;45(12):1380-1386. doi:10.1080/0142159X.2023.2218537

14. French BN, Marxen TO, Akhnoukh S, et al. A call for spaced repetition in medical education. *Clin Teach*. 2024;21(1):e13669. doi:10.1111/tct.13669

15. USMLE Step 1 Practice Questions & Exams. Medical. January 7, 2020. Accessed December 29, 2024. https://medical.uworld.com/usmle/usmle-step-1/

16. AnkiHub — Collaborative Anki Decks! AnkiHub. Accessed November 20, 2024. https://www.ankihub.net/

17. UWorld2Anki. Accessed December 29, 2024. https://ankiweb.net/shared/info/271205340

18. Schukow CP, Johnson CV, Kowalski P. Is there utility for implementing digital flash card applications in pathology undergraduate and graduate medical education? *Arch Pathol Lab Med*. 2023;147(2):133-134. doi:10.5858/arpa.2022-0246-LE

19. Barrison PD, Balczewski EA, Capellari E, Landis-Lewis Z, Vinson AH. Electronic Flashcards in Health Professions Education: A Scoping Review. *Academic Medicine*. Published online January 6, 2025.

20. Deng F, Gluckstein JA, Larsen DP. Student-directed retrieval practice is a predictor of medical licensing examination performance. *Perspect Med Educ*. 2015;4(6):308-313. doi:10.1007/s40037-015-0220-x

21. Kuperstock JE, Horný M, Platt MP. Mobile app technology is associated with improved otolaryngology resident in-service performance. *Laryngoscope*. 2019;129(1):E15-E20. doi:10.1002/lary.27299

22. Lambers A, Talia AJ. Spaced Repetition Learning as a Tool for Orthopedic Surgical Education: A Prospective Cohort Study on a Training Examination. *J Surg Educ*. 2021;78(1):134-139. doi:10.1016/j.jsurg.2020.07.002

23. Kortz MW, Kongs BM, Bisesi DR, Roffler M, Sheehy RM. A retrospective and correlative analysis of academic and nonacademic predictors of COMLEX level 1 performance. *J Osteopath Med*. 2022;122(4):187-194. doi:10.1515/jom-2021-0175

24. Yüksel HG, Mercanoğlu HG, Yılmaz MB. Digital flashcards vs. wordlists for learning technical vocabulary. *Comput Assist Lang Learn*. 2022;35(8):2001-2017. doi:10.1080/09588221.2020.1854312

25. Lu M, Farhat JH, Beck Dallaghan GL. Enhanced Learning and Retention of Medical Knowledge Using the Mobile Flash card Application Anki. *Med Sci Educ*. 2021;31(6):1975-1981. doi:10.1007/s40670-021-01386-9

26. Mehta A, Brooke N, Puskar A, et al. Implementation of spaced repetition by first-year medical students: A retrospective comparison based on summative exam performance. *Med Sci Educ*. 2023;33(5):1089-1094. doi:10.1007/s40670-023-01839-3

27. Gilbert MM, Frommeyer TC, Brittain GV, et al. A Cohort Study Assessing the Impact of Anki as a Spaced Repetition Tool on Academic Performance in Medical School. *Med Sci Educ*. 2023;33(4):955-962. doi:10.1007/s40670-023-01826-8

28. Jape D, Zhou J, Bullock S. A spaced-repetition approach to enhance medical student learning and engagement in medical pharmacology. *BMC Med Educ*. 2022;22(1):337. doi:10.1186/s12909-022-03324-8

29. Makus D, Kashyap A, Labib M, Humphrey-Murto S. A curriculum ignored? The usage of unofficial commercial and peer learning resources in undergraduate medical education at a Canadian medical school. *Med Sci Educ*. 2023;33(6):1379-1388. doi:10.1007/s40670-023-01899-5

30. Loving RJ, Haeussler C, McMurray JR, Peacock B, Hsu M, Day L. Does the use of pre-made digital flashcards aid in medical students learning of anatomy. *FASEB J*. 2022;36 Suppl 1(S1). doi:10.1096/fasebj.2022.36.S1.L8037

31. Wu JH, Gruppuso PA, Adashi EY. The self-directed medical student curriculum. *JAMA*. 2021;326(20):2005-2006. doi:10.1001/jama.2021.16312

32. Davis D. The medical school without walls: Reflections on the future of medical education. *Med Teach*. 2018;40(10):1004-1009. doi:10.1080/0142159X.2018.1507263

33. Hoyt G, Adegboyega S, Constantouris G, Basu P. Study of the impact of introducing a multimedia learning tool in podiatric medical courses. *J Foot Ankle Res*. 2024;17(3):e12018. doi:10.1002/jfa2.12018

34. Ostropolets A, Zhang L, Hripcsak G. A scoping review of clinical decision support tools that generate new knowledge to support decision making in real time. *J Am Med Inform Assoc*. 2020;27(12):1968-1976. doi:10.1093/jamia/ocaa200

35. Blease C, Torous J, McMillan B, Hägglund M, Mandl KD. Generative language models and open notes: Exploring the promise and limitations. *JMIR Med Educ*. 2024;10(1):e51183. doi:10.2196/51183

36. Levin MA, Wanderer JP, Ehrenfeld JM. Data, Big Data, and metadata in anesthesiology. *Anesth Analg*. 2015;121(6):1661-1667. doi:10.1213/ANE.0000000000000716

37. Laurent G, Moussa MD, Cirenei C, Tavernier B, Marcilly R, Lamer A. Development, implementation and preliminary evaluation of clinical dashboards in a department of anesthesia. *J Clin Monit Comput*. 2021;35(3):617-626. doi:10.1007/s10877-020-00522-x

38. Waller RG, Wright MC, Segall N, et al. Novel displays of patient information in critical care settings: a systematic review. *J Am Med Inform Assoc*. 2019;26(5):479-489. doi:10.1093/jamia/ocy193

39. King CR, Gregory S, Fritz BA, et al. An intraoperative telemedicine program to improve perioperative quality measures: The ACTFAST-3 randomized clinical trial. *JAMA Netw Open*. 2023;6(9):e2332517. doi:10.1001/jamanetworkopen.2023.32517

40. Lobo SM, Salgado PF, Castillo VG, et al. Effects of maximizing oxygen delivery on morbidity and mortality in high-risk surgical patients. *Crit Care Med*. 2000;28(10):3396-3404. doi:10.1097/00003246-200010000-00003

41. Kobe J, Mishra N, Arya VK, Al-Moustadi W, Nates W, Kumar B. Cardiac output monitoring: Technology and choice. *Ann Card Anaesth*. 2019;22(1):6-17. doi:10.4103/aca.ACA_41_18

42. Vincent JL, Rhodes A, Perel A, et al. Clinical review: Update on hemodynamic monitoring--a consensus of 16. *Crit Care*. 2011;15(4):229. doi:10.1186/cc10291

43. Evans DC, Doraiswamy VA, Prosciak MP, et al. Complications associated with pulmonary artery catheters: a comprehensive clinical review. *Scand J Surg*. 2009;98(4):199-208. doi:10.1177/145749690909800402

44. Marzano RJ, Kendall JS. *Designing and Assessing Educational Objectives: Applying the New Taxonomy*. (Marzano RJ, Kendall JS, eds.). Corwin Press; 2008. Accessed November 11, 2024. https://books.google.com/books/about/Designing_and_Assessing_Educational_Obje.html?id=YuZ0AwAAQBAJ

45. Cardiac Output. BYJUS. July 26, 2018. Accessed January 2, 2025. https://byjus.com/biology/cardiac-output/

46. Hartman M, Debiec N, Kim E, et al. Rad Path in a flash: Creation of the first AMSER sponsored digital flashcard deck and a review of digital flashcard use among? Medical students. *Acad Radiol*. 2024;31(2):426-430. doi:10.1016/j.acra.2023.12.039

47. Halperin SJ, Zhu JR, Francis JS, Grauer JN. Are medical school curricula adapting with their students? A survey on how medical students study and how it relates to USMLE Step 1 scores. *J Med Educ Curric Dev*. 2024;11:23821205241228456. doi:10.1177/23821205241228455

48. Alliance for Clinical Education. *Handbook on Medical Student Evaluation and Assessment*. Gegensatz Press; 2015. Accessed November 20, 2024. https://books.google.com/books/about/Handbook_on_Medical_Student_Evaluation_a.html?id=bx0wCgAAQBAJ

49. Tsai S, Sun M, Asbury ML, Weber JM, Truong T, Deans E. Novel spaced repetition flashcard system for the in-training examination for obstetrics and gynecology. *Med Sci Educ*. 2021;31(4):1393-1399. doi:10.1007/s40670-021-01320-z

50. Morin CE, Hostetter JM, Jeudy J, et al. Spaced radiology: encouraging durable memory using spaced testing in pediatric radiology. *Pediatr Radiol*. 2019;49(8):990-999. doi:10.1007/s00247-019-04415-3

51. Boespflug A, Guerra J, Dalle S, Thomas L. Enhancement of customary dermoscopy education with spaced education e-learning: A prospective controlled trial. *JAMA Dermatol*. 2015;151(8):847-853. doi:10.1001/jamadermatol.2015.0214

52. Al-Rawi W, Easterling L, Edwards PC. Development of a mobile device optimized cross platform-compatible oral pathology and radiology spaced repetition system for dental education. *J Dent Educ*. 2015;79(4):439-447. doi:10.1002/j.0022-0337.2015.79.4.tb05902.x

53. Taveira-Gomes T, Prado-Costa R, Severo M, Ferreira MA. Characterization of medical students recall of factual knowledge using learning objects and repeated testing in a novel e-learning system. *BMC Med Educ*. 2015;15:4. doi:10.1186/s12909-014-0275-0

54. Pinder KE, Dick J. Creation and use of medical histology-pathology electronic flashcards. *FASEB J*. 2019;33(S1). doi:10.1096/fasebj.2019.33.1_supplement.444.21

55. Mishall PL, Burton W, Risley M. Flashcards: The preferred online game-based study tool self-selected by students to review medical histology image content. *Adv Exp Med Biol*. 2023;1406:209-224. doi:10.1007/978-3-031-26462-7_10

56. Chandler C, Parker E, Mantilla J, Rendi M. The use of anki digital flashcards for pre-clinical pathology medical education review. In: *Abstract Presented at: USCAP 109th Annual Meeting*. ; 2020.

57. DeLeon V, Ahmad N, Tran L, Thesen T. Supercharging memory retention in neurology learning: A faculty-student collaboration to apply cognitive science principles to software-based self-directed learning (S10.006). *Neurology*. 2023;100(17_supplement_2). doi:10.1212/wnl.0000000000204008

58. Ward JL, Venegrin CA. The effect of repeated review of course content on medium and long-term retention in an elective veterinary cardiology course. *J Vet Med Educ*. Published online 2023. doi:10.3138/jvme-2023-0088

59. Taveira-Gomes T, Saffarzadeh A, Severo M, Guimarães MJ, Ferreira MA. A novel collaborative e-learning platform for medical students - ALERT STUDENT. *BMC Med Educ*. 2014;14(1):143. doi:10.1186/1472-6920-14-143

60. Stephenson CL, Tadikonda P, Bui MM, Centeno B. Evaluation of innovative digital flashcards in the study of the Bethesda system for reporting thyroid cytology. *J Am Soc Cytopathol*. 2013;2(1):S13. doi:10.1016/j.jasc.2013.08.032

61. Ayoob A, Hardy P, Waits T, Brooks M. Development of a web-based curriculum to prepare diagnostic radiology residents during post-graduate year 1 by promoting learning via spaced repetition and interim testing: If you build it, will they come? *Acad Radiol*. 2019;26(8):1112-1117. doi:10.1016/j.acra.2019.02.011

62. Bow HC, Dattilo JR, Jonas AM, Lehmann CU. A crowdsourcing model for creating preclinical medical education study tools. *Acad Med*. 2013;88(6):766-770. doi:10.1097/ACM.0b013e31828f86ef

63. O'Hanlon R, Laynor G. Responding to a new generation of proprietary study resources in medical education. *J Med Libr Assoc*. 2019;107(2):251-257. doi:10.5195/jmla.2019.619

64. Underman K, Kochunilathil M, McLean L, Vinson AH. Online student culture as site for negotiating assessment in medical education. *Soc Sci Med*. 2022;310:115270. doi:10.1016/j.socscimed.2022.115270

65. Piemontesi J, Martindale A, Sagorin Z. Spaced-repetition with electronic flashcards to supplement curricular objectives: preliminary results from the UBC clerkship Anki initiative. In: *Canadian Conference for the Advancement of Surgical Education (c-Case) Virtual Conference*. CMAJ Canadian Medical Association Journal; 2020. https://www.canjsurg.ca/content/63/6_Suppl_1/S61

66. Brateanu A, Strang TM, Garber A, et al. Using an Adaptive, Self-Directed Web-Based Learning Module to Enhance Residents' Medical Knowledge Prior to a New Clinical Rotation. *Med Sci Educ*. 2019;29(3):779-786. doi:10.1007/s40670-019-00772-8

67. Sun M, Tsai S, Engle DL, Holmer S. Spaced Repetition Flashcards for Teaching Medical Students Psychiatry. *Med Sci Educ*. 2021;31(3):1125-1131. doi:10.1007/s40670-021-01286-y

68. Hasan MM, Adnan M, Waheed B. Use of an Adaptive e-Learning Platform as a Formative Assessment Tool in the Cardiovascular System Course Component of an MBBS Programme [Letter]. *Adv Med Educ Pract*. 2021;12:111-112. doi:10.2147/AMEP.S299759

69. Marzano RJ, Kendall JS. *The New Taxonomy of Educational Objectives*. Corwin Press; 2006. https://play.google.com/store/books/details?id=JT4KAgAAQBAJ

70. Anki - powerful, intelligent flashcards. Accessed February 7, 2023. https://apps.ankiweb.net/

71. Med School Anki. Accessed September 5, 2024. https://www.reddit.com/r/medschoolanki/

72. Harris DM, Chiang M. An Analysis of Anki Usage and Strategy of First-Year Medical Students in a Structure and Function Course. *Cureus*. 2022;14(3):e23530. doi:10.7759/cureus.23530

73. Hudder A, Tackett S, Moscatello K. First-year experience implementing an adaptive learning platform for first- and second-year medical students at the Lake Erie College of Osteopathic Medicine. *J Am Osteopath Assoc*. 2019;119(1):51-58. doi:10.7556/jaoa.2019.008

74. Veremis BM, Ramaswamy V, Popov V, Danciu TE. Can a spaced repetition software aid dental students? A pilot study. *J Dent Educ*. 2022;86 Suppl 1(S1):865-867. doi:10.1002/jdd.12778

75. Whitman AC, Tanzer K, Nemec EC 2nd. Gamifying the memorization of brand/generic drug names. *Curr Pharm Teach Learn*. 2019;11(3):287-291. doi:10.1016/j.cptl.2018.12.014

76. Zheng M, O'Brien K, Cuenin K, Lyon C, Bender D. Impact of test-enhanced learning as a study strategy: An exploratory study with first-year dental students. *J Dent Educ*. 2022;86(12):1611-1619. doi:10.1002/jdd.13066

77. Rana T, Laoteppitaks C, Zhang G, Troutman G, Chandra S. An investigation of anki flashcards as a study tool among first year medical students learning anatomy. *FASEB J*. 2020;34(S1):1-1. doi:10.1096/fasebj.2020.34.s1.09736

78. Landoll RR, Bennion LD, Maggio LA. Understanding excellence: A qualitative analysis of high-performing learner study strategies. *Med Sci Educ*. 2021;31(3):1101-1108. doi:10.1007/s40670-021-01279-x

79. Wothe JK, Wanberg LJ, Hohle RD, et al. Academic and wellness outcomes associated with use of Anki spaced repetition software in medical school. *J Med Educ Curric Dev*. 2023;10:23821205231173290. doi:10.1177/23821205231173289

80. Zeitlin BD, Sadhak ND. Attitudes of an international student cohort to the Quizlet study system employed in an advanced clinical health care review course. *Educ Inf Technol*. 2023;28(4):3833-3857. doi:10.1007/s10639-022-11371-3

81. Gupta S, Ojeh N, Sa B, Majumder MAA, Singh K, Adams OP. Use of an adaptive e-learning platform as a formative assessment tool in the Cardiovascular System course component of an MBBS programme. *Adv Med Educ Pract*. 2020;11:989-996. doi:10.2147/AMEP.S267834

82. Cole R, McHugh D, Netter FH. Assessing emotional stress, active recall and digital spaced-learning media in the study of thoracic gross anatomy by medical students. *FASEB J*. 2017;31(S1):580.7-580.7. doi:10.1096/fasebj.31.1_supplement.580.7

83. Mohamed I, Elsobky Y, Kreys E, Mente W. I mpact of decoding medication tradenames on students' performance in didactic & experiential rotation courses. *FASEB J*. 2021;35(S1). doi:10.1096/fasebj.2021.35.s1.01970

84. Schmidmaier R, Ebersbach R, Schiller M, Hege I, Holzer M, Fischer MR. Using electronic flashcards to promote learning in medical students: retesting versus restudying. *Med Educ*. 2011;45(11):1101-1110. doi:10.1111/j.1365-2923.2011.04043.x

85. McHugh D, Sherban TM, Rahman SH. Digital Spaced-learning Media: a Platform to Reduce Student Anxiety and Promote Proficiency in Medical Pharmacology Education. *The FASEB Journal*. 2016;30(S1):944.7-944.7. doi:10.1096/fasebj.30.1_supplement.944.7

86. Irvine J. A comparison of revised Bloom and Marzano's New Taxonomy of Learning. https://files.eric.ed.gov/fulltext/EJ1161486.pdf

87. Schmidmaier R, Eiber S, Ebersbach R, et al. Learning the facts in medical school is not enough: which factors predict successful application of procedural knowledge in a laboratory setting? *BMC Med Educ*. 2013;13:28. doi:10.1186/1472-6920-13-28

88. Dubas JM, Toledo SA. Taking higher order thinking seriously: Using Marzano's taxonomy in the economics classroom. *Int Rev Econ Educ*. 2016;21:12-20. doi:10.1016/j.iree.2015.10.005

89. Toledo S, Dubas JM. Encouraging higher-order thinking in general chemistry by scaffolding student learning using Marzano's taxonomy. *J Chem Educ*. 2016;93(1):64-69. doi:10.1021/acs.jchemed.5b00184

90. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa

91. Goldman M, Bryan J, Lucke-Wold B. Evidence-Based Algorithm "Anki" for Optimization of Medical Education: The Evolution of Knowledge Retention. Published online 2023. https://www.preprints.org/manuscript/202307.1955

92. Levy J, Ely K, Lagasca G, et al. Exploring Anki Usage Among First-Year Medical Students During an Anatomy & Physiology Course: A Pilot Study. *J Med Educ Curric Dev*. 2023;10:23821205231205388. doi:10.1177/23821205231205389

93. Salmona M, Lieber E, Kaczynski D. *Qualitative and Mixed Methods Data Analysis Using Dedoose: A Practical Approach for Research Across the Social Sciences*. SAGE Publications; 2019. https://play.google.com/store/books/details?id=inClDwAAQBAJ

94. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2023. https://www.R-project.org/

95. Alavi M, Leidner DE. Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *Miss Q*. 2001;25(1):107-136. doi:10.2307/3250961

96. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* . 2012;22(3):276-282. doi:10.1016/j.jocd.2012.03.005

97. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*. 2021;13(2):201-216. doi:10.1080/2159676X.2019.1704846

98. Hennink MM, Kaiser BN, Marconi VC. Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? *Qual Health Res*. 2017;27(4):591-608. doi:10.1177/1049732316665344

99. Adesope OO, Trevisan DA, Sundararajan N. Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Rev Educ Res*. 2017;87(3):659-701. doi:10.3102/0034654316689306

100. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63-7. doi:10.1097/00001888-199009000-00045

101. Abdullah N, Frost RA. Rethinking the Semantics of Complex Nominals. In: *Advances in Artificial Intelligence*. Springer Berlin Heidelberg; 2007:502-513. doi:10.1007/978-3-540-72665-4_43

102. Graff CJ, Kaljo K, Treat RW, Dielentheis K. ORIGINAL RESEARCH. https://wmjonline.org/wp-content/uploads/2021/120/1/8.pdf

103. Kern D, Thomas P. What do leaders need to know about curriculum planning. https://scholar.google.com/citations?user=rbyy8vkAAAAJ&hl=en&oi=sra

104. Seol YH, Kaufman DR, Cimino JJ, Johnson SB. Scenario-based Assessment of Physicians' Information Needs. In: *MEDINFO 2004*. IOS Press; 2004:306-310. doi:10.3233/978-1-60750-949-3-306

105. Young M, Thomas A, Lubarsky S, et al. Drawing boundaries: The difficulty in defining clinical reasoning. *Acad Med*. 2018;93(7):990-995. doi:10.1097/ACM.0000000000002142

106. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annu Symp Proc*. 2006;2006:359-363. https://www.ncbi.nlm.nih.gov/pubmed/17238363

107. Peyton PJ, Chong SW. Minimally invasive measurement of cardiac output during surgery and critical care: a meta-analysis of accuracy and precision: A meta-analysis of accuracy and precision. *Anesthesiology*. 2010;113(5):1220-1235. doi:10.1097/ALN.0b013e3181ee3130

108. Mathews L, Singh RKK. Cardiac output monitoring. *Ann Card Anaesth*. 2008;11(1):56-68. doi:10.4103/0971-9784.38455

109. Goodwin AJ, Eytan D, Greer RW, et al. A practical approach to storage and retrieval of high-frequency physiological signals. *Physiol Meas*. 2020;41(3):035008. doi:10.1088/1361-6579/ab7cb5

110. Ke L, Elibol A, Wei X, Cenyu L, Wei W, Chong NY. Machine learning algorithm to predict cardiac output using arterial pressure waveform analysis. In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE; 2022. doi:10.1109/bibm55620.2022.9995429

111. Sun JX, Reisner AT, Saeed M, Heldt T, Mark RG. The cardiac output from blood pressure algorithms trial. *Crit Care Med*. 2009;37(1):72-80. doi:10.1097/ccm.0b013e3181930174

112. Sun JX, Reisner AT, Saeed M, Mark RG. Estimating cardiac output from arterial blood pressurewaveforms: a critical evaluation using the MIMIC II database. In: *Computers in Cardiology, 2005*. IEEE; 2005. doi:10.1109/cic.2005.1588095

113. Koenig J, Hill LK, Williams DP, Thayer JF. Estimating cardiac output from blood pressure and heart rate: the liljestrand & zander formula. *Biomed Sci Instrum*. 2015;51:85-90. Accessed October 24, 2024. https://pmc.ncbi.nlm.nih.gov/articles/PMC5317099/

114. Kouz K, Bergholz A, Timmermann LM, et al. The relation between mean arterial pressure and cardiac index in major abdominal surgery patients: A prospective observational cohort study. *Anesth Analg*. 2022;134(2):322-329. doi:10.1213/ANE.0000000000005805

115. Caillard A, Gayat E, Tantot A, et al. Comparison of cardiac output measured by oesophageal Doppler ultrasonography or pulse pressure contour wave analysis. *Br J Anaesth*. 2015;114(6):893-900. doi:10.1093/bja/aev001

116. Zhang J, Critchley LAH, Huang L. Five algorithms that calculate cardiac output from the arterial waveform: a comparison with Doppler ultrasound. *Br J Anaesth*. 2015;115(3):392-402. doi:10.1093/bja/aev254

117. Chen T, Clifford G, Mark R. The effect of signal quality on six cardiac output estimators. *Comput Cardiol*. 2009;2009(36):197-200. https://www.ncbi.nlm.nih.gov/pubmed/20740055

118. Broch O, Bein B, Gruenewald M, et al. Accuracy of cardiac output by nine different pulse contour algorithms in cardiac surgery patients: A comparison with transpulmonary thermodilution. *Biomed Res Int*. 2016;2016(1):3468015. doi:10.1155/2016/3468015

119. Saugel B, Mair S, Götz SQ, et al. Indexation of cardiac output to biometric parameters in critically ill patients: A systematic analysis of a transpulmonary thermodilution-derived database. *J Crit Care*. 2015;30(5):957-962. doi:10.1016/j.jcrc.2015.06.011

120. Maas JJ, Pinsky MR, de Wilde RB, de Jonge E, Jansen JR. Cardiac output response to norepinephrine in postoperative cardiac surgery patients: interpretation with venous return and cardiac function curves. *Crit Care Med*. 2013;41(1):143-150. doi:10.1097/CCM.0b013e318265ea64

121. Abramov D, He KL, Wang J, Burkhoff D, Maurer MS. The impact of extra cardiac comorbidities on pressure volume relations in heart failure and preserved ejection fraction. *J Card Fail*. 2011;17(7):547-555. doi:10.1016/j.cardfail.2011.03.010

122. MPOG Current Projects. Accessed December 10, 2024. https://mpog.org/currentprojects/

123. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457. doi:10.1016/S0140-6736(07)61602-X

124. Colquhoun DA, Shanks AM, Kapeles SR, et al. Considerations for integration of perioperative electronic health records across institutions for research and quality improvement: The approach taken by the Multicenter Perioperative Outcomes Group: The approach taken by the multicenter perioperative outcomes group. *Anesth Analg*. 2020;130(5):1133-1146. doi:10.1213/ANE.0000000000004489

125. Mathis MR, Dubovoy TZ, Caldwell MD, Engoren MC. Making sense of big data to improve perioperative care: Learning health systems and the multicenter perioperative outcomes group. *J Cardiothorac Vasc Anesth*. 2020;34(3):582-585. doi:10.1053/j.jvca.2019.11.012

126. Winkley Shroyer AL, Bakaeen F, Shahian DM, et al. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: The driving force for improvement in cardiac surgery. *Semin Thorac Cardiovasc Surg*. 2015;27(2):144-151. doi:10.1053/j.semtcvs.2015.07.007

127. Phenotype Browser. Accessed October 22, 2024. https://phenotypes.mpog.org/

128. The Society of Thoracic Surgeons Adult Cardiac Surgery Database Data Collection Form. https://www.sts.org/sites/default/files/Data%20collection%20tool%20Form2_41%20final.pdf

129. Mathis MR, Mentz GB, Jie C, et al. Hospital and Clinician Practive Variation in Cardiac Surgery and Postoperative Acute Kidney Injury.

130. Burton RF. Estimating body surface area from mass and height: theory and the formula of Du Bois and Du Bois. *Ann Hum Biol*. 2008;35(2):170-184. doi:10.1080/03014460801908439

131. Ehlers KC, Mylrea KC, Waterson CK, Calkins JM. Cardiac output measurements. A review of current techniques and research. *Ann Biomed Eng*. 1986;14(3):219-239. doi:10.1007/bf02584272

132. Belletti A, Lerose CC, Zangrillo A, Landoni G. Vasoactive-inotropic score: Evolution, clinical utility, and pitfalls. *J Cardiothorac Vasc Anesth*. 2021;35(10):3067-3077. doi:10.1053/j.jvca.2020.09.117

133. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat*. 2007;17(4):571-582. doi:10.1080/10543400701329422

134. de Myttenaere A, Golden B, Le Grand B, Rossi F. Mean Absolute Percentage Error for regression models. *Neurocomputing*. 2016;192:38-48. doi:10.1016/j.neucom.2015.12.114

135. Stetz CW, Miller RG, Kelly GE, Raffin TA. Reliability of the thermodilution method in the determination of cardiac output in clinical practice. *Am Rev Respir Dis*. 1982;126(6):1001-1004. doi:10.1164/arrd.1982.126.6.1001

136. Ball L, Costantino F, Pelosi P. Postoperative complications of patients undergoing cardiac surgery. *Curr Opin Crit Care*. 2016;22(4):386-392. doi:10.1097/MCC.0000000000000319

137. Mentz G, Schulz AJ, Mukherjee B, Ragunathan TE, Perkins DW, Israel BA. Hypertension: development of a prediction model to adjust self-reported hypertension prevalence at the community level. *BMC Health Serv Res*. 2012;12(1):312. doi:10.1186/1472-6963-12-312

138. Westerhof N, Lankhaar JW, Westerhof BE. The arterial Windkessel. *Med Biol Eng Comput*. 2009;47(2):131-141. doi:10.1007/s11517-008-0359-2

139. Herd JA, Leclair NR, Simon W. Arterial pressure pulse contours during hemorrhage in anesthetized dogs. *J Appl Physiol*. 1966;21(6):1864-1868. doi:10.1152/jappl.1966.21.6.1864

140. Guze PA. Using technology to meet the challenges of medical education. *Trans Am Clin Climatol Assoc*. 2015;126:260-270. Accessed January 2, 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC4530721/

141. Sotgiu MA, Mazzarello V, Bandiera P, Madeddu R, Montella A, Moxham B. Neuroanatomy, the achille's heel of medical students. A systematic analysis of educational strategies for the teaching of neuroanatomy. *Anat Sci Educ*. 2020;13(1):107-116. doi:10.1002/ase.1866

142. King R, Churchill EF, Tan C. *Designing with Data: Improving the User Experience with A/B Testing*. O'Reilly Media; 2017. Accessed September 18, 2024. https://books.google.com/books/about/Designing_with_Data.html?id=eM-PDgAAQBAJ

143. Pendergrast T, Chalmers Z. Anki tagger: A generative AI tool for aligning third-party resources to preclinical curriculum. *JMIR Med Educ*. 2023;9(1):e48780. doi:10.2196/48780

144. AnkiBrain (AnkiChatGPT) - GPT-4, GPT 3.5 with auto-cards and document analysis. Accessed January 3, 2025. https://ankiweb.net/shared/info/1915225457

145. Arnold JH, Stenz RI, Thompson JE, Arnold LW. Noninvasive determination of cardiac output using single breath CO sub 2 analysis. *Crit Care Med*. 1996;24(10):1701. Accessed January 3, 2025. https://journals.lww.com/ccmjournal/abstract/1996/10000/noninvasive_determination_of_cardiac_output_using.17.aspx

146. Santiago Giraldo Gutiérrez D, José Velásquez Gutiérrez J, Octavio Ruiz-Villa J. Cardiac output estimation based on arterial and venous blood gas analysis: proposal of a monitoring method. *Anaesthesiol Intensive Ther*. 2021;53(2):179-183. doi:10.5114/ait.2021.105730

147. An Evolving Framework for Using Big Data Tools and Machine Learning to Enhance Perioperative Quality Improvement, Research, and Patient Safety. Anesthesia Patient Safety Foundation. June 1, 2024. Accessed January 3, 2025. https://www.apsf.org/article/an-evolving-framework-for-using-big-data-tools-and-machine-learning-to-enhance-perioperative-quality-improvement-research-and-patient-safety/

148. Tolentino R, Baradaran A, Gore G, Pluye P, Abbasgholizadeh-Rahimi S. Curriculum frameworks and educational programs in AI for medical students, residents, and practicing physicians: Scoping review. *JMIR Med Educ*. 2024;10(1):e54793. doi:10.2196/54793