

The State of Mastery Learning in Pediatric Graduate Medical Education: A Scoping Review

Michaela Mills-Rudy ¹, Megan Thorvilson¹, Cynthia Chelf ², Stephanie Mavis³

¹Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, USA; ²Mayo Clinic College of Medicine and Science, Rochester, MN, USA; ³Department of Pediatric and Adolescent Medicine, Division of Neonatal Medicine, Mayo Clinic, Rochester, MN, USA

Correspondence: Michaela Mills-Rudy, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905, USA, Tel +1 507-266-9397, Fax +1 507-255-0602, Email mills-rudy.michaela@mayo.edu

Objective: The aim of this study was to characterize the state of mastery learning interventions, identify gaps in current approaches, and highlight opportunities to improve the rigor of ML in pediatric graduate medical education (GME) training programs.

Methods: In October 2022, we searched Ovid, PubMed, Scopus, and ERIC. Two reviewers independently screened 165 citations and reviewed the full manuscripts of 20 studies. We developed a modified data abstraction tool based on the Recommendations for Reporting Mastery Education Research in Medicine (ReMERM) guidelines and extracted variables related to mastery learning curricular implementation and design and learner assessment.

Results: Eleven studies of ML approaches within pediatric GME were included in the full review, with over half published after 2020. ML interventions were used to teach both simple and complex tasks, often in heterogeneous learner groups. While deliberate practice and feedback were consistently reported features of ML in pediatrics, opportunities for improvement include clearly defining mastery, conducting learning over multiple sessions, presenting sufficient validity evidence for assessment tools, adhering to rigorous standard setting processes, and reporting psychometric data appropriate for ML.

Conclusion: In pediatric GME, ML approaches are in their infancy. By addressing common shortcomings in the existing literature, future efforts can improve the rigor of ML in pediatric training programs and its impact on learners and patients.

Plain Language Summary: While mastery learning is a well-described, effective educational intervention utilized in multiple medical specialties, we perceived a relative lack of published studies on mastery learning in pediatric graduate medical education. Mills-Rudy's team searched the current literature to identify gaps in mastery learning approaches in pediatrics training and to highlight ways to improve the rigor of mastery learning in pediatric training programs. Their search yielded 11 studies of mastery learning approaches in pediatric graduate medical education. They identified major gaps in curriculum development and implementation as well as learner assessment. Opportunities to improve mastery learning in pediatrics include clearly defining mastery, conducting learning over several sessions, presenting sufficient validity evidence for assessment tools, adhering to rigorous standard setting processes, and reporting psychometric data appropriate for mastery learning. Future mastery learning interventions in pediatrics can address these gaps to improve the rigor of mastery learning in pediatric training programs.

Keywords: competency-based medical education, graduate medical education

Background

Educational approaches that translate to sustained effects on learners, and ultimately patients, are essential for the provision of safe health care.¹ Traditional models of graduate medical education rely on graduated clinical experience and frequently rely on approaches such as the “see one, do one, teach one” method.² In the era of rapid expansion of medical knowledge however, these approaches are insufficient to guarantee the acquisition and maintenance of clinical competence for all learners, and they create uneven educational and clinical outcomes.^{3,4}

Mastery learning (ML) is a unique educational paradigm that focuses on achievement of competency rather than time spent learning a task, such that all learners achieve the expected standard for any educational unit.^{3,5} ML utilizes baseline testing, repetitive deliberate practice, formative feedback, and rigorous assessment to a mastery standard. Importantly, mastery does not mean that a learner is training to be an expert in a task, rather, that the learner is well-prepared for the next stage of learning. In an ML environment, learners advance to the next educational unit when they are in an optimal learning zone in which they have challenged, but not exceeded, their abilities.⁶ McGaghie previously defined seven core components of ML interventions, which include baseline testing, clear learning objectives, educational activities to reach learning objectives, a set minimum passing standard (MPS), formative testing at a pre-set MPS, advancement to the next unit once reaching the MPS, and continued study on a unit until MPS is reached.⁷ In health professions education, rigorous meta-analyses suggest that in comparison with no intervention, ML interventions have large effects on skills and a moderate effect on patient outcomes; additionally, in comparison with non-mastery approaches, ML interventions have been shown to have a large effect on skills.^{8,9}

This gap is echoed in a 2021 survey of general pediatricians, in which most pediatricians still reported learning through a “see one, do one, teach one” approach, most were never formally assessed on procedural competence during residency, and all general pediatricians surveyed reported referring out at least one procedure.¹⁰ And while a paradigm shift to ML interventions is evident in surgical and medical specialties,^{9,11–15} we identified a relative paucity of studies arising from pediatric training programs. In a recent meta-analysis of ML interventions, none of the 82 studies included in included learners in pediatric GME.⁸ The unique considerations related to infant and children’s difference in anatomy, physiology, and pathophysiology may necessitate different approaches to diagnosis and management than seen in adults. ML, like all competency-based models of medical education, is founded on rigorous assessment,⁹ so given such fundamental differences in children that dictate different considerations and approaches to diagnosis and management, the use of assessment instruments developed for use in adult populations may not be appropriate. As such, assessment tools developed for use specifically in the pediatric population and standard setting based on the unique aspects of pediatric medicine are essential to translate such evidence for ML to pediatrics training.

Because of the imperative for high quality learning aligned with competency-based medical education,³ the increased application of ML within other aspects of graduate medical education (GME), and the unique aspects of pediatric medicine training, we aimed to both characterize the state of ML in pediatric training programs and identify the gaps in ML approaches in pediatric GME by highlighting opportunities to improve the rigor of conducting and reporting of mastery learning curricula in this particular field.

Methods

A scoping review was ideal to map the breadth of the literature, summarize the current literature, and identify gaps in the current mastery learning approaches in pediatrics. Using the Arksey and O’Malley framework for scoping review methodology,¹⁶ we intentionally designed a broad search strategy to screen all relevant articles related to mastery learning in pediatric graduate medical education. We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) guidelines to report our study Methods and findings.¹⁷

Research Question

We aimed to characterize the current state of mastery learning in pediatric GME by answering the following questions:

1. How is mastery learning implemented in pediatric GME?
2. What are the gaps related to curriculum development and implementation, learner assessment, and program evaluation within mastery learning settings in pediatric GME?

Search Strategy

In line with scoping review methodology, we designed an initial broad search strategy and iteratively narrowed the list of selected articles using exclusion criteria guided by the research questions. A medical librarian (CJC) conducted a literature search for publications on mastery learning and graduate medical education. Search strategies utilized

a combination of keywords and standardized index terms (see [Appendix 1](#)), which were applied in October 2022 in Ovid, PubMed (MEDLINE), Scopus (Elsevier), and ERIC. Results were limited to the English language spanning January 1974 to October 2022. We exported all results using the reference management software EndNote (Clarivate Analytics, Chandler, AZ) and removed duplicate results, leaving 165 citations. Reference lists of 18 review articles identified by the initial search strategy were hand searched for additional references, with one additional study identified. See [Figure 1](#) for the scoping review study selection overview.

Identifying Relevant Studies

The primary investigators (MAM and SCM) met at the beginning, middle, and end of the review process to develop and ensure application of a consistent shared mental model. Initial inclusion criteria comprised title and abstract review for studies that included pediatric GME learners (including pediatric residents and subspecialty fellows) and a mastery learning intervention. To ensure completeness of the initial search, articles mentioning pediatrics were kept for full text review even if the trainee type was not specified. Exclusion criteria included studies that involved only non-pediatric GME residents or fellows (ie: internal medicine residents only, medical students, non-GME health professions education learners, and health professionals) or those that had unclear study populations based on abstract only as well as those that did not mention the word “mastery”. After an independent review of the 165 included articles, the coinvestigators identified and agreed upon 20 articles for full text review.

Both MAM and SCM then conducted independent full-text reviews of the 20 articles. We did not eliminate articles based on the quality of evidence, and we included studies with negative results. We excluded abstracts for which full-text

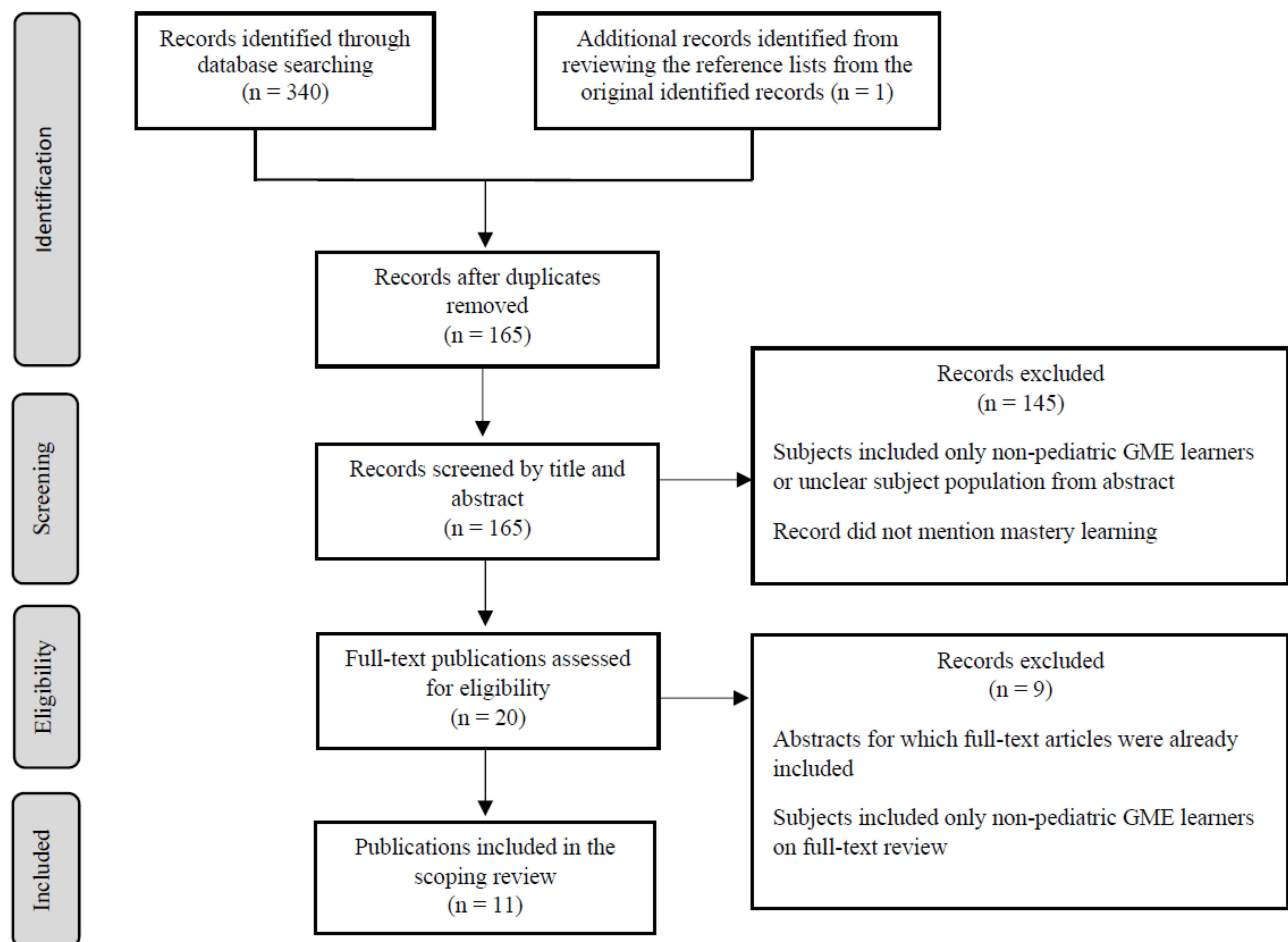


Figure 1 PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram for a October 2022 scoping review exploring the current state of mastery learning education in pediatric GME.

articles were already included, as well as articles that did not include pediatric GME learners on full-text review. This led to the selection of 11 articles for data extraction (Figure 1).

Data Extraction and Organization

We selected the Recommendations for Reporting Mastery Education Research in Medicine (ReMERM) developed by Cohen et al¹⁸ as the basis of our data extraction tool. The ReMERM guidelines were rigorously developed by experts in mastery learning and delineate the essential curricular and assessment components recommended to be reported in mastery learning research, including 38 items deemed “imperative for reporting a ML research study”. While many of these items are standard in education reporting (such as robust description and pilot testing of assessment tools, detailed reporting of rater training processes, and baseline assessment), additional elements specific to ML are emphasized and include a clear definition of mastery, the nature of practice, feedback and debriefing approaches, descriptions of standard setting methodology, a clear description of how the MPS was set and the post-training assessment using the MPS, and the number/percentage of trainees who met the mastery standard within the standard curriculum.

We reviewed the guideline author descriptions of ReMERM items and added additional items of interest to the scoping review; these included the particular pediatric GME setting, greatest validity threats to assessment tool (analyzed according to Messick’s validity framework),¹⁹ and mastery learning outcome classifications specified by Cook et al⁸ such as outcomes related to time required (such as time to tie five knots), to process (such as economy of movement during suturing), and to final product (such as knot integrity). We developed a code book (see [Appendix 2](#)) to improve inter-investigator consistency. Lead investigators (MAM and SCM) then independently extracted all data from each publication, then worked together to combine data into a single consensus response for each checklist item (see [Appendix 3](#) for core components of mastery learning in each of the included studies). The investigators reached agreement for all articles. A third author (MJT) reviewed all articles and confirmed the accuracy of final data extraction.

Collating, Summarizing, and Reporting Results

After data extraction, all authors met to synthesize the data and ensure all members fully agreed upon the interpretation of results.

Results

Included Studies

After full-text review, 11 publications met the inclusion criteria and were included in the final review. Publication dates ranged from 2011 to 2022, with six out of eleven studies being published in 2020 or later. Eight of the studies were published in general pediatrics journals, whereas three were published in pediatric subspecialty journals.

Descriptive Analysis

To characterize the state of ML within pediatric GME, we organized extracted Results from the review process into two major domains: 1) curricular design and implementation ([Tables 1 and 2](#)) learner assessment and program evaluation ([Table 2](#)).

Curricular Design

Educational Goals and Learner Groups

While ML²⁰ approaches are frequently used in simple procedural skills training (4/11, 36%), they are also increasingly described to train learners in complex and integrative tasks (7/11, 64%). Such tasks requiring various cognitive, interpersonal, and technical skills included training in status epilepticus management,²¹ medical adverse event disclosure,²² neonatal teleresuscitation,²³ and pediatric resuscitation.²⁴ All studies clearly stated their study aims and/or objectives. Most studies included pediatric residents only (8/11); amongst these studies, programs frequently included multiple postgraduate training year (PGY) levels (6/8 studies). Studies that included fellows (3 studies)^{20,23,25} were also more likely to include mixed-learner groups (ie: residents, fellows from other training programs, advanced practice

Table I Curricular Components of Included Studies of ML in Pediatric GME

Curricular Component	Studies Reporting (n/%)
Conceptual framework (CF)	
Reported at least one CF	10/11 (91%)
Intended educational audience	
General pediatrics training	8/11 (73%)
Subspecialty training	3/11 (27%)
Curricular focus	
Simple or singular task	7/11 (64%)
Complex or integrative task	4/11 (36%)
Journal	
Pediatrics specific	8/11 (73%)
Non-pediatrics	3/11 (27%)
Year	
Pre-2010	0/11 (0%)
2010–2014	3/11 (27%)
2015–2019	2/11 (18%)
2020–2022	6/11 (55%)
Includes mastery in title	
Yes	6/11 (55%)
No	5/11 (45%)
Setting	
Single-center	9/11 (82%)
Multi-center	2/11 (18%)
United States	9/11 (82%)
Brazil	1/11 (9%)
France	1/11 (9%)
Learners	
Residents only	8/11 (73%)
Fellows only	1/11 (9%)
Trainees and practicing physicians	2/11 (18%)
Instructional methods for desired knowledge or skill	
Videos	4/11 (37%)
Didactics	3/11 (27%)
Online modules	3/11 (27%)
Live, expert demonstration	2/11 (18%)
Quiz	1/11 (9%)
Primary deliberate practice method	
Task trainers/low fidelity mannequin	6/11 (55%)
High fidelity mannequin	1/11 (9%)
Practice on peers or real patients	2/11 (18%)
Standardized patients	1/11 (9%)
Online image and associated question library	1/11 (9%)
Simulated resuscitations	2/11 (18%)
Mastery learning deliberate practice duration	
Single session	9/11 (82%)
Longitudinal sessions	1/11 (9%)
Unclear	1/11 (9%)
Practice structure	
Individual practice	9/11 (82%)
Teams	1/11 (9%)
Unclear	1/11 (9%)

(Continued)

Table 1 (Continued).

Curricular Component	Studies Reporting (n/%)
Feedback provided during deliberate practice	
Yes	11/11 (100%)
No	0/11 (0%)
Stakes	
Voluntary	4/11 (36%)
Mandatory	4/11 (36%)
Unclear	3/11 (27%)

Table 2 Assessment Components of Included Studies of ML in Pediatric GME

Assessment Constructs	Studies Reporting (n/%)
Instrument used to assess mastery*	
Existing, published tool	1/11 (9%)
Tool modified from existing	2/11 (18%)
New tool	8/11 (73%)
Unclear	1/11 (9%)
If new assessment tool created, describes the tool development and pilot testing	
Yes	4/10 (40%)
No	6/10 (60%)
Clear definition of mastery	
Yes	3/11 (27%)
No	8/11 (73%)
Any rater training reported?	
Yes	3/11 (27%)
No	8/11 (73%)
Multiple raters assessing mastery	
Yes	8/11 (73%)
No	1/11 (9%)
Unclear	2/11 (18%)
If multiple raters, reports the reliability coefficients	
Yes	7/8 (88%)
No	1/8 (12%)
Set an MPS	
Yes	8/11 (73%)
No	3/11 (27%)
Standard setting methodology	
Mastery Angoff	4/8 (50%)
Arbitrary	4/8 (50%)
Baseline assessment using mastery checklist	
Done	7/11 (64%)
Not done	4/11 (36%)
Reported number of learners achieving MPS at baseline	
Yes	3/7 (43%)
No	4/7 (57%)
Post-curriculum assessment using mastery checklist	
Done	8/11 (73%)
Not done	3/11 (27%)

(Continued)

Table 2 (Continued).

Assessment Constructs	Studies Reporting (n%)
Reported number of learners achieving MPS after deliberate practice	
Yes	7/8 (88%)
No	1/8 (12%)
Allowed extra-time to achieve MPS	
Yes	3/7 (43%)
Unclear	4/7 (57%)
Assessment level	
Reactions	6/11 (55%)
Knowledge	3/11 (27%)
Skills in artificial settings	8/11 (73%)
Behaviors with real patients	1/11 (9%)
Effects on real patients	3/11 (27%)
For skill in artificial settings:	
Skill outcomes related to time (how long it takes to perform a task)	1/8 (12.5%)
Skill outcomes related to process (economy of movement, minor errors)	8/8 (100%)
Skill outcomes related to product (results observable after the task is done, such as knot integrity)	3/8 (38%)
For behaviors with real patients:	
Behaviors related to time	0/1 (0%)
Behaviors related to process	1/1 (100%)

Note: *One study both modified an existing tool and created a new global rating scale, leading to 12 instrument descriptions.

providers, practicing physicians) (2 studies) or include multiple PGY levels (1 study). Studies were rarely designed for a single learning level (ie: PGY-1 learners) (2/11 studies).^{21,26}

Skill Development Approaches

Foundational education was commonly provided in the form of videos (4/11 studies),^{20,26–28} didactic presentations (3/11 studies),^{22,25,29} and online modules (3/11 studies),^{23,25,28} expert demonstrations^{23,29} and quizzes²⁸ were used less often as teaching tools. Deliberate practice (DP) was utilized in all studies. Nearly all published studies describe ML occurring only over a single learning session (9/11, 82%); only one study performed multiple longitudinal learning sessions²³ and in one study, the time frame for content delivery was unclear.²⁵

Deliberate Practice, Supervision and Expert Feedback Approaches

A variety of methods and intensity of DP were reported. Most commonly, task trainers or low fidelity mannequins were utilized for DP (6/11 studies),^{20,24,26,27,29,30} however, studies also included practice on high fidelity mannequins,²¹ peers,²⁰ and standardized patients,²² as well as practice using online image libraries²⁵ and during team-based simulation training.^{23,24} When reported, the duration of DP varied substantially, between 5 and 120 minutes.

Most studies also reported directly-supervised deliberate practice (9/11 studies). However, one study reported automated feedback but unsupervised learning (via a question bank with pre-scripted expert feedback)²⁵ and one study did not report any supervised training.²² Feedback was a consistent element of each study, although the quality of feedback and the reporting of feedback varied considerably, most commonly, in the reported level of expertise of those providing feedback.

Assessment and Evaluation in Mastery Settings

Definition of Mastery and Standard Setting Process

Only three of the included studies clearly defined mastery.^{20,23,24} While a majority of the studies report a minimum passing standard (MPS) (8/11, 73%), only half of those studies reported an appropriate standard-setting methodology for a ML intervention.^{20,21,24,29} Of the studies that report utilizing the Mastery Angoff method to define the MPS (4/11), only

two describe an accurate Mastery Angoff process^{20,24} (ie: defining cut scores for the well-prepared, not the average²⁹ or minimally competent,²¹ learner).

Assessment Domains and Levels

Studies commonly investigated skill outcomes related to process (8/11 studies) and less frequently studied skill outcomes related to final product or time required. Whereas most studies assessed skills in artificial settings (8/11, 73%), one study assessed behaviors with real patients²³ and three studies measured effects on real patients.^{26,27,29} Beyond performance assessments, additional testing and evaluations included knowledge tests (3/11, 27%)^{22,25,27} and surveys of learner reactions (6/11, 55%).^{20–22,25,27,28}

Validity Evidence Provided for Use of Assessment Tools

Most studies relied on assessment tools created or modified specifically for the ML intervention. One study utilized a previously published performance checklist without modification,²² and two modified prior studies.^{27,30} Critical threats to assessment validity were identified in 9 of 11 studies; threats most commonly related to content evidence and rater response process evidence. For example, of the ten studies with newly developed or modified assessment tools, only four reported both assessment tool development and pilot testing data as recommended in the ReMERM guidelines.^{20,24,25,27} Additionally, while most studies had multiple raters to assess learner performance and reported the interrater reliability, only three studies reported any details of the rater training process.^{21,22,24}

Assessment Timing and Frequency

Seven of the eleven studies (64%) performed a baseline assessment using a mastery checklist. While three studies reported the number of learners who achieved the minimum passing standard at baseline assessment,^{20,21,29} four studies instead reported the mean baseline score for the learner group(s).^{22,24,27,30} Most studies (8/11, 73%) reported using the mastery checklist as the post-curriculum assessment.

Discussion

Mastery learning²⁰ has been shown to be an effective teaching model in GME, and in this scoping review, we demonstrate that within pediatric GME, such curricula are infrequently reported and often nonadherent to guiding principles and recommended reporting standards for ML. The question is not whether ML is “effective” or not, but rather when, how, and why ML can be ideally applied to produce meaningful and durable changes in skill.³ Increasing the number and quality of reports of ML is essential to improving provider skill, patient safety, and patient outcomes.

The current state of ML in pediatrics is characterized by relatively few attempts to improve skills across a variety of topics, ranging from training of singular tasks (such as lumbar puncture) to training of more complex activities (such as leading resuscitations or discussing adverse events with parents). Studies frequently involved heterogeneous learner groups and were typically conducted in a single learning session. Though all studies included deliberate practice and expert feedback, the intensity, duration, and nature of feedback varied across studies.

By broadly scoping the current literature and evaluating study adherence based on standards for high quality ML education research, we were able to recognize consistent threats to robust ML in general and within subspecialty pediatric training. We noted two significant gaps related to curricular design (failing to provide a clear definition of mastery and limiting opportunities for guided practice and expert feedback to a single learning session) and three gaps related to assessment (insufficient development and/or reporting of validity evidence for use of any assessment tool, lack of rigorous standard setting methodologies, and absence of or inaccurate reporting of baseline and post-course assessment data). Adherence to the ReMERM checklist and mastery principles will allow future education researchers to improve the quality of ML research for pediatric trainees.

To advance ML within pediatric GME, these gaps, as well as potential opportunities for curriculum developers, education researchers, and education leaders, are presented in [Table 3](#). The rationale for improving each of these areas is also related to the principles of ML and its underlying theoretical basis.

Table 3 Major Gaps and Opportunities Within ML Interventions in Pediatric GME

Domain	Gaps in Current Approaches	Potential Ways to Improve the Quality and Rigor of ML	Why is this Important?
Curriculum 1	Vague or incomplete reporting of mastery, where mastery indicated what the learner is well prepared to do	Answer the question, “What is the learner going to be well prepared to do after this ML curriculum?” In early course design Clearly define mastery in ML publications	Clarity of what is meant by “mastery” is critical for all local curricular stakeholders, as well as for others looking to translate such curricula
Curriculum 2	Overreliance on single session interventions to achieve durable results	Ensure course designs are longitudinal in nature, allowing for repeated practice and feedback not only within a session, but in multiple sessions, potentially with multiple sources of feedback	Expertise of any skill requires repeated and nuanced practice Emphasis on short term improvements may actually undermine goal for long term skill
Assessment 1	Limited validity evidence has been developed for assessment tools utilized in ML settings	Collect sufficient validity evidence for any assessment tool utilized in a ML setting, such that the degree of validity evidence aligns with degree of entrustment afforded by passing the mastery course	Sufficient threats to an assessment’s validity undermine the use of results obtained from that assessment Rigorous assessment tools themselves can drive learning
Assessment 2	Insufficient rigor in determining the MPS	Utilize established standard setting processes (such as the Mastery Angoff or Patient-Safety approach) to set the MPS	MPS too high may impose significant barriers to appropriate learner entrustment to advance, while MPS too low may introduce patient safety concerns
Assessment 3	Infrequent baseline testing and/or accurate reporting	Assess learners and report findings according to ML standards (baseline and post-course number/percentage who achieve the MPS)	Range restriction is expected as course goal is uniform achievement (measures of central tendency not appropriate markers for programmatic evaluation)

Improvements Related to Curriculum Development and Implementation

Opportunity #1: Improve Conceptualization and Clarity of the Term “Mastery”

Why? Fewer than one third of published ML interventions in pediatric GME provide a clear definition of mastery. While it is tempting to equate mastery with a high level of expertise, Yudkowsky et al emphasizes that in ML, mastery simply implies “readiness to proceed to the next phase of instruction”.⁵ This nuance is important as completion of a mastery learning intervention does not imply mastery over the intended material. Instead, ML interventions aim to create learners who are well prepared to succeed in the next stage of training. Aligning with the competency-based medical education framework, time commitment to learn and practice is variable in the ML model, while outcomes are uniform amongst learners. While a clear conceptualization of mastery will clearly drive course design, it is also critical for robust and appropriate feedback during deliberate practice,³¹ learner assessment (in setting the MPS, as the shared mental model for whom the “well prepared learner” is and does),⁶ and for reducing confusion about consequences of course completion (especially the degree of content expertise of the learner).

Opportunity #2: Allow for Repeated Practice Not Only Within a Single Session, but in Multiple Sessions

Why? Published studies demonstrate an overreliance on single-session interventions failing to demonstrate durable retention of skill. Ericsson has postulated that the development of expertise in a given area is related not only to the amount of practice in which the individual engages, but also the quality of these extensive experiences and of the coaching received during such training.³² Additionally, it is also clear that conditions that maximize performance in initial training stages may not always maximize long-term learning and retention of such skills.^{33,34} As such, rapidly passing an MPS may actually hinder long term mastery goals. Thus, course designers should consider use of multiple learning sessions to allow for incorporation of such effortful deliberate practice, exposure to different learning conditions (including of task and of coach), and ultimately opportunities to improve retention of skills in the long-term.

Improvements Related to Learner Assessment

Opportunity #3: Ensure Sufficient Validity Evidence for Any Assessment Tool Utilized in a ML Setting

Why? In this review, we found that validity evidence for new or modified tools are infrequently supplied in studies detailing ML interventions. Sound assessment is a central inference in the validity of ML paradigms,⁶ but threats to validity abounded across included published studies and includes major threats to all areas of Messick's unifying theory of validity. As new assessment tools were frequently developed, these most commonly reflect insufficiently developed content evidence and rater response process evidence. Validity threats can lead to inappropriate conclusions about individual learners or ML programs from assessment scores, including both unsafe entrustment for clinical duties (if incorrectly assigning the learner as passing) or undue restriction from clinical activities (if incorrectly assigning them as failing). Rigorous assessment is critical in all educational programs; however, in ML it may be even more critical as uniform competence is the goal, restriction of score ranges is expected, and learners are entrusted to perform varying levels of the task in clinical medicine.^{5,6}

The validity of any assessment tool or process is often framed as an argument, in which one gathers differing types of evidence to support the conclusion that the interpretation of any score leads to defensible decisions regarding those assessed.³⁵ Additionally, if readiness for the next stage implies a clinically important increase in clinical responsibility, stronger validity evidence is required. For example, an ML curriculum intended to prepare pediatric residents to safely and independently perform infant lumbar punctures (i.e: without direct supervision) would demand a higher level of validity evidence than an ML curriculum intending to prepare interns to perform the same procedure on a simulator. Finally, in contrast to other educational methods, learners in ML are actively encouraged to examine such assessment tools; as such, the quality of the tool may directly influence the quality of learning.

Of note, educational assessment tools are never “validated” but rather, educators can gather more or less validity evidence for their use and the interpretation of learner scores in a particular context. It is incumbent upon educators to ensure that the level of validity evidence gathered is appropriate for the ML context, including the learners and the consequences of passing an ML course. Future researchers and curriculum designers can elevate their ML curricula by ensuring that sufficient validity evidence for an assessment instrument's intended use exists before that tool is used to determine a learner's readiness for the next phase of training.

Opportunity #4: Improve the Rigor of Standard Setting Process

Why? While most studies report setting an MPS, fewer than 20% of published studies correctly utilized and reported an accepted standard setting process for ML. Similar to evidence for the assessment tool itself, a defensible approach to the standard setting is essential such that GME learners who pass an ML curriculum are appropriately entrusted to advance to the next stage of practice, while those who do not pass are appropriately restricted from certain clinical activities. To address this gap, educators should utilize an established standard-setting methodology, such as Mastery Angoff or Patient-Safety approach, to set the MPS for their intervention.^{5,36} Additionally, attention to the mental model of experts involved in standard setting exercises is critical. Given the mastery framework and the goal for learners to be “well-prepared” for the next stage, neither the average learner nor the minimally competent learner is the appropriate reference standard in these exercises.⁵

Opportunity #5: Assess Learners and Report Findings According to ML Standards

Why? While many of the interventions included in this scoping review completed some form of baseline testing, fewer than half reported their results in the recommended ML format per ReMERM standards.¹⁸ In ML, learners either pass and move on or fail and repeat the educational interventions, so scores are often restricted significantly in range and often much higher than in educational interventions in which uniform achievement of competence is not the goal.⁵ As such, in ML the number or percent of learners who achieve the MPS at baseline is significantly more interesting than a group's mean or median knowledge or skill level.¹⁸ Post-course data should be evaluated similarly, as high mean or median scores post-test are less meaningful than understanding the number and percent of learners who pass the course within the expected time frame (and those who require additional time).

Strengths and Limitations

The strengths of this study are adherence to Arksey and O'Malley scoping review strategy, assistance of medical librarian in design of the search strategy, and the use of ReMERM guidelines as the basis of the extraction tool. Limitations include publication bias, which may limit the total number of implemented but not published educational interventions occurring within pediatric GME.

Conclusion

Rigorous curriculum design and assessment is critical to improving learners' skill and patient outcomes. While the number of published studies utilizing ML frameworks are increasing, future studies can advance the field of ML in pediatrics by improving the quality of study design and consistency of reporting. We propose five areas of focus that those conducting ML curricula in the contemporary era can target to improve the rigor of pediatric GME.

Disclosure

The authors report no conflicts of interest in this work.

References

1. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Medical education featuring mastery learning with deliberate practice can lead to better health for individuals and populations. *Acad Med.* 2011;86(11):e8–9. doi:10.1097/ACM.0b013e3182308d37
2. Halstead WS. The Training of the Surgeon. *Bull.* 1904;15:267–275.
3. McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med.* 2015;90(11):1438–1441. doi:10.1097/ACM.0000000000000911
4. Smith MM, Secunda KE, Cohen ER, Wayne DB, Vermylen JH, Wood GJ. Clinical experience is not a proxy for competence: comparing fellow and medical student performance in a breaking bad news simulation-based mastery learning curriculum. *Am J Hosp Palliat Care.* 2022;40(4):423–430.
5. Yudkowsky R, Park YS, Lineberry M, Knox A, Ritter EM. Setting mastery learning standards. *Acad Med.* 2015;90(11):1495–1500. doi:10.1097/ACM.0000000000000887
6. Lineberry M, Soo Park Y, Cook DA, Yudkowsky R. Making the case for mastery learning assessments: key issues in validation and justification. *Acad Med.* 2015;90(11):1445–1450. doi:10.1097/ACM.0000000000000860
7. McGaghie WC, Siddall VJ, Mazmanian PE, Myers J. Lessons for continuing medical education from simulation research in undergraduate and graduate medical education: effectiveness of continuing medical education: American college of chest physicians evidence-based educational guidelines. *Chest.* 2009;135(3):62s–68s. doi:10.1378/chest.08-2521
8. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med.* 2013;88(8):1178–1186. doi:10.1097/ACM.0b013e31829a365d
9. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med.* 2011;86(6):706–711. doi:10.1097/ACM.0b013e318217e119
10. Iyer MS, Way DP, Schumacher DJ, Lo CB, Leslie LK. How general pediatricians learn procedures: implications for training and practice. *Med Educ Online.* 2021;26(1):1985935. doi:10.1080/10872981.2021.1985935
11. Barsness KA. Achieving expert performance through simulation-based education and application of mastery learning principles. *Semin Pediatr Surg.* 2020;29(2):150904. doi:10.1016/j.sempedsurg.2020.150904
12. Nataraja RM, Webb N, Lopez PJ. Simulation in paediatric urology and surgery. Part 1: an overview of educational theory. *J Pediatr Urol.* 2018;14(2):120–124. doi:10.1016/j.jpuro.2017.12.021
13. Barsuk JH, Cohen ER, Feinglass J, McGaghie WC, Wayne DB. Use of simulation-based education to reduce catheter-related bloodstream infections. *Arch Intern Med.* 2009;169(15):1420–1423. doi:10.1001/archinternmed.2009.215
14. Barsuk JH, Cohen ER, Wayne DB, Siddall VJ, McGaghie WC. Developing a simulation-based mastery learning curriculum: lessons from 11 years of advanced cardiac life support. *Simul Healthc.* 2016;11(1):52–59. doi:10.1097/SIH.0000000000000120
15. Cohen ER, Barsuk JH, Moazed F, et al. Making July safer: simulation-based mastery learning during intern boot camp. *Acad Med.* 2013;88(2):233–239. doi:10.1097/ACM.0b013e31827bfc0a
16. Arksey HOML, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* 2005;8(1):19–32. doi:10.1080/1364557032000119616
17. Tricco ACLE, Zarin W, O'Brien KK, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* 2018;169(7):467–473. doi:10.7326/M18-0850
18. Cohen ER, McGaghie WC, Wayne DB, Lineberry M, Yudkowsky R, Barsuk JH. Recommendations for reporting mastery education research in medicine (ReMERM). *Acad Med.* 2015;90(11):1509–1514. doi:10.1097/ACM.0000000000000933
19. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830–837. doi:10.1046/j.1365-2923.2003.01594.x
20. Ballard HA, Tsao M, Robles A, et al. Use of a simulation-based mastery learning curriculum to improve ultrasound-guided vascular access skills of pediatric anesthesiologists. *Paediatr Anaesth.* 2020;30(11):1204–1210. doi:10.1111/pan.13953
21. Malakooti MR, McBride ME, Mobley B, Goldstein JL, Adler MD, McGaghie WC. Mastery of status epilepticus management via simulation-based learning for pediatrics residents. *J Grad Med Educ.* 2015;7(2):181–186. doi:10.4300/JGME-D-14-00516.1

22. Aubin J, Rivolet O, Taunay AL, Ragot S, Ghazali DA, Oriot D. Benefit of simulation-based training in medical adverse events disclosure in pediatrics. *Pediatr Emerg Care.* 2022;38(2):e622–e627. doi:10.1097/PEC.0000000000002454
23. Mavis SC, Kreofsky BL, Ouk MY, Carey WA, Fang JL. Training fellows in neonatal tele-resuscitation using a simulation-based mastery learning model. *Resusc Plus.* 2021;8:100172. doi:10.1016/j.resplu.2021.100172
24. Braun L, Sawyer T, Smith K, et al. Retention of pediatric resuscitation performance after a simulation-based mastery learning session: a multicenter randomized trial. *Pediatr Crit Care Med.* 2015;16(2):131–138. doi:10.1097/PCC.0000000000000315
25. Brown KA, Riley AF, Alade KH, et al. A novel tool for teaching cardiac point-of-care ultrasound: an exploratory application of the design-based research approach. *Pediatr Crit Care Med.* 2020;21(12):e1113–e1118. doi:10.1097/PCC.0000000000002441
26. Kessler DO, Arteaga G, Ching K, et al. Interns' success with clinical procedures in infants after simulation training. *Pediatrics.* 2013;131(3):e811–820. doi:10.1542/peds.2012-0607
27. Kessler DO, Auerbach M, Pusic M, Tunik MG, Foltin JC. A randomized trial of simulation-based deliberate practice for infant lumbar puncture skills. *Simul Healthc.* 2011;6(4):197–203. doi:10.1097/SIH.0b013e318216bfc1
28. Price A, Greene HM, Stem CT, Titus MO. Sticking it straight: pediatric procedure curriculum initiative. *Pediatr Emerg Care.* 2022;38(2):79–82. doi:10.1097/PEC.0000000000002324
29. Couto TB, Reis AG, Farhat SCL, Carvalho VEL, Schwartsman C. Changing the view: impact of simulation-based mastery learning in pediatric tracheal intubation with videolaryngoscopy. *J Pediatr.* 2021;97(1):30–36. doi:10.1016/j.jpeds.2019.12.007
30. Matterson HH, Szyld D, Green BR, et al. Neonatal resuscitation experience curves: simulation based mastery learning booster sessions and skill decay patterns among pediatric residents. *J Perinat Med.* 2018;46(8):934–941. doi:10.1515/jpm-2017-0330
31. Eppich WJ, Hunt EA, Duval-Arnould JM, Siddall VJ, Cheng A. Structuring feedback and debriefing to achieve mastery learning goals. *Acad Med.* 2015;90(11):1501–1508. doi:10.1097/ACM.0000000000000934
32. Ericsson KA. The differential influence of experience, practice, and deliberate practice on the development of superior individual performance of experts. In: *Cambridge Handbook of Expertise and Expert Performance*; 2018.
33. Kapur M. Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educ Psychologist.* 2016;51(2):289–299. doi:10.1080/00461520.2016.1155457
34. Schmidt RA, Bjork RA. New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychol Sci.* 1992;3(4):207–217. doi:10.1111/j.1467-9280.1992.tb00029.x
35. Cook DABT, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 2006;119(2):166.e167–116. doi:10.1016/j.amjmed.2005.10.036
36. Barsuk JH, Cohen ER, Wayne DB, McGaghie WC, Yudkowsky R. A comparison of approaches for mastery learning standard setting. *Acad Med.* 2018;93(7):1079–1084. doi:10.1097/ACM.0000000000002182

Advances in Medical Education and Practice

Dovepress

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>