

How Internal Medicine Clerkship Directors Are Using Entrustable Professional Activities: A National Survey Study



Katherine A. Gielissen, MD, MHS-MedEd¹ , Anna Foust, MD², Amy R. Weinstein, MD, MPH³, Nicholas Duca, MD⁴, Melissa O. Jenkins, MD⁵, Michael Kisielewski, MA⁶, Savita Misra, MA⁶, and Dana Dunne, MD, MHS-MedEd⁷

¹Emory University School of Medicine, Atlanta, GA, USA; ²University of South Alabama College of Medicine, Mobile, AL, USA; ³Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, MA, USA; ⁴Penn State College of Medicine, Hershey, PA, USA; ⁵Case Western Reserve University, Cleveland, OH, USA; ⁶Alliance for Academic Internal Medicine, Alexandria, VA, USA; ⁷Yale University School of Medicine, New Haven, CT, USA

ABSTRACT

BACKGROUND: In the present assessment environment in undergraduate medical education at U.S. medical schools, the prevalence and implementation of Entrustable Professional Activities (EPAs) in internal medicine (IM) clerkships are not well understood.

OBJECTIVE: To describe the prevalence and approach to EPA use in U.S. IM clerkships.

DESIGN: Cross-sectional, nationally representative survey of core IM clerkship directors.

PARTICIPANTS: One-hundred forty IM clerkship directors at Liaison Committee on Medical Education-accredited U.S./U.S. territory-based allopathic medical schools with membership in the Clerkship Directors in Internal Medicine (CDIM) as of December 2022.

MAIN MEASURES: Use of EPAs in IM core clerkships, including use for grading, types of EPAs, use of supportive measures for assessment, and current validity frameworks.

KEY RESULTS: The survey response was 80% (112/140); two additional respondents completed the section on EPA use ($n=114$). Approximately half of respondents (47%) reported their IM clerkship used EPAs. Among schools accredited after 1977, a higher percentage was associated with having incorporated EPAs ($p=0.03$). The Association of American Medical Colleges Core EPAs for Entering Residency (CEPAER) was the most common framework used by Clerkship Directors (CDs) for developing EPAs (55%). Most CDs (56%) used EPAs for both formative and summative assessments, and approximately half of CDs (48%) used EPAs for a portion of the final grade determination. CDs who used EPAs were no more likely to report efforts to ensure the validity of assessment, the use of faculty development, or that written assessments were a valid measure of students' performance compared to those who did not use EPAs.

CONCLUSIONS: Although EPAs have experienced substantial uptake in the IM clerkship and contribute to formative and summative assessment of learners, their use does not appear to be associated with enhanced efforts to obtain validity information.

KEY WORDS: Clerkship; Internal medicine; Entrustable professional activities; Assessment; Validity; Undergraduate medical education

J Gen Intern Med

DOI: 10.1007/s11606-024-08991-5

© The Author(s), under exclusive licence to Society of General Internal Medicine 2024

BACKGROUND

In 2014, the Association of American Medical Colleges (AAMC) introduced 13 Core Entrustable Professional Activities (EPAs) for Entering Residency (CEPAER) to improve graduates' preparedness for residency.¹ Over seven years, a 10-school consortium (the AAMC Core EPA Pilot) developed curricula and assessment strategies and rendered entrustment decisions across diverse institutions.² In kind, schools outside the pilot have implemented EPAs as part of the assessment of students, even making entrustment decisions on readiness for indirect supervision on specific tasks.³ Despite these shifts towards workplace-based assessments, validity evidence is conflicting.^{4,5} The CEPAER pilot found EPA assessments showed promise, but none of the participating schools were ready to make high-stakes decisions by the end of the 7-year study, suggesting that further development was needed to ensure medical school graduates were prepared for residency.⁶

Several studies have investigated the use of CEPAER in undergraduate medical education (UME), showing variable uptake. A 2020 survey of osteopathic schools found only 31% of respondents implemented EPAs at least "moderately," with approximately two-thirds of those schools reporting EPA outcomes.⁷ Similarly, a 2020 study of pediatric clerkship directors showed that 24% of respondents were using an EPA assessment framework in the core pediatric clerkship.⁸ Both studies called for amplified support for educators using systems-based strategies to develop valid workplace-based assessments. In parallel, there has been increasing interest in competency-based outcomes in UME. In 2021, the Undergraduate Medical Education to Graduate Medical Education Review Committee called for "[t]he UME community, working in conjunction with partners across the continuum" to "commit to using robust assessment tools and strategies, improving upon existing tools, developing new tools where needed, and gathering and reviewing additional evidence of validity."⁹

Received February 6, 2024

Accepted August 2, 2024

Published online: 15 August 2024

Prior to this study, the most recent Clerkship Directors in Internal Medicine (CDIM) Annual Survey exploring EPA use in the core IM clerkship occurred in 2015; it assessed anticipated barriers to EPAs and identified six “key” EPAs for assessment, but did not explore the extent of EPA use, and no study to date has reported EPA use in internal medicine (IM) clerkships.^{10,11} Our study’s aim was to understand the prevalence of EPA use in IM clerkships and the degree to which IM clerkship directors (CDs) evaluate the validity of their assessments. We also explored CD attitudes towards assessment and the role of validity in evaluating candidates for residency training.

METHODS

CDIM is a charter organization of the Alliance for Academic Internal Medicine (AAIM), a professional association that includes academic faculty and leaders responsible for third- and fourth-year UME. CDIM has conducted annual surveys of IM CDs on topics essential to UME since 1999 (Appendix A).

Survey Development

In February 2022, CDIM opened a call for thematic survey section proposals to its members. In April, the CDIM Survey and Scholarship Committee blind-reviewed all submissions and selected three for inclusion based on relevance to the core IM training experience. The authors who proposed the EPA section (KG, AF, DD) were not involved in this selection process. From May through mid-June, the committee (AW, ND, MJ) and section authors (KG, AF, DD) revised the questions through an iterative process. From July through August, the elected CDIM Council and five CDIM experts (blinded to the committee) pilot-tested the web survey. The survey section authors then addressed and resolved comments from the pilot test, and final revisions were incorporated. To screen for problematic question content, all pilot data were analyzed for anomalies (e.g., illogical response combinations or out-of-scope values) and were resolved by the committee and section leads. All section authors, committee members, Council members, and pilot testers included subject matter experts with extensive experience in the clerkship setting.

In addition to a recurring 13-question section on “Clerkship Director and Medical School Characteristics,” the section upon which this manuscript is based contained 20 questions, including multiple-choice (single-selection and select-all-that-apply), three- and five-point Likert scale, numeric entry questions, and open-text response options, with logical skip and display patterns. Due to conditional logic or item-non-response, denominators for some questions do not sum to the total number of survey respondents. See Appendix B for the section-specific survey instrument.

Early in August 2022, MK/SM exported from the AAIM/CDIM database all individuals designated as “clerkship director” (CD) at 149 of 154 U.S. medical schools with full or provisional Liaison Committee on Medical Education (LCME) accreditation; one designated clerkship director per school was included in the survey. Generic identification numbers were appended to respondent contact files to match them back to the survey database, which included prepopulated characteristics such as medical school classification. Unique participant URLs were disseminated via email invitation through the *Qualtrics Surveys platform*. The original population of 149 was adjusted to 140 after removing nine medical schools confirmed not to have renewed their AAIM/CDIM membership at survey closure. The final population represented 91 percent (140/154) of all fully and provisionally accredited LCME schools as of the study period.

The study (22-AAIM-118) was declared exempt by Pearl IRB (U.S. DHHS OHRP #IRB00007772). Only MK and SM had access to the survey population and survey software during fielding.

The survey launched on September 7, 2022, closed on December 6, 2022, and included four email reminders to non-respondents. All email communications included voluntary opt-out links and the survey landing page displayed an informed consent statement. Contacts who voluntarily communicated they no longer were their medical school’s IM CD were replaced with the new CD of record or most appropriate next contact.

The survey dataset was stored in an encrypted drive accessible only to project personnel. Respondents’ records were merged with the complete population file to include demographic and medical school characteristics, and the study dataset was de-identified. Descriptive statistics were used to report the summary results. Fisher’s exact test or Pearson’s chi-square with one degree of freedom (or with Sidak-adjusted *p*-values for multiple comparison tests) was used to test for associations between categorical variables. Due to their non-parametric distribution, continuous variables were compared to categorical variables using the Wilcoxon rank-sum (Mann–Whitney) test, reporting median and interquartile range (IQR). Statistical significance was designated using an alpha level of 0.05. Data analysis was conducted in *Stata 16.1 SE*.

Thematic Analysis

Open-ended questions were analyzed using thematic analysis. Two study authors (KG and DD) reviewed comments to familiarize themselves with the dataset and generate initial codes. The authors discussed these codes together and grouped them into related themes, then coded independently to ensure themes/subthemes were consistently applied; discrepancies in coding were discussed among the study group until consensus was reached.

RESULTS

A summary of the 2022 CDIM Annual Survey overall results can be found in Appendix C. The survey response rate was 80.0% (112/140). Two additional individuals responded to the EPA portion of the survey and are included in this analysis ($n=114$). Of the total survey respondents, 54 (47.4%) reported their clerkship is currently using EPAs to assess medical students. Table 1 shows there were no statistically significant differences in EPA use between type of school, region, size, accreditation features, number of students who rotate on the IM clerkship at a given time, median number of years the CD has been in position, median full-time equivalent for the CD position, or use of pass/fail grading. CDs at medical schools accredited after 1977 were associated with reporting a higher likelihood of using EPAs in the IM clerkship. Among CDs who reported no EPA use (60/114), 40.0% reported no current plan to add EPAs, 8.3% reported a plan to start using EPAs, and 33.3% reported they were unsure. The remaining 18.3% reported “other” responses, and described barriers to EPA implementation congruent with the thematic analysis below.

Of CDs who reported EPA use (referred to in this manuscript as “EPA CDs”), most (55.1%) used the CEPAER guides as the sole resource for developing EPAs, followed by a combination of CEPAER and EPAs developed at their institution (36.7%) (Table 2).

Of EPA CDs, most used them for both formative and summative assessment (Table 2). Twenty-six (48.1%) reported using EPAs as part of the final clerkship grade, and a majority of EPA CDs used tiered grading (e.g., honors, letter grades). The median percentage of the grade attributable to EPAs at these schools was 65% (IQR 25), with 38.5% of these CDs using EPAs for $\leq 50\%$ of the final grade and 61.5% using EPAs for $> 50\%$ of the final grade (15.4% used EPAs for 100% of the final grade). Faculty and residents were most frequently reported as those who obtained EPA assessments of students, and were most commonly reported as individuals who received guidance on EPA assessment (Table 2). In comparative analyses, CDs who used EPAs as a component of the final grade were no more likely to report providing guidance to assessors on how to perform assessments compared to CDs who did not use EPAs for the final grade ($p=0.23$); however, there was an association with having derived EPA instruments from the literature ($p<0.01$). Pass/fail schools that used EPAs were no more likely to have specific individuals perform EPA assessments (e.g., attending physicians, residents; $p>0.05$ for all comparisons) and were no more likely to provide training to the individuals who performed EPA assessments compared to non-pass/fail schools ($p>0.05$ for all comparisons).

There was no association between EPA CDs and non-EPA CDs with respect to efforts to ensure validity of assessment, including internal review of instruments, external review of instruments, use of validated instruments from the literature, or use of faculty development compared to non-EPA CDs (Table 1). Similarly, compared to non-EPA CDs, EPA CDs

were no more likely to report monitoring assessments for validity. EPA CDs were no more likely than non-EPA CDs to report written assessments were a valid measure of students’ performance in the IM clerkship, nor were they more likely to report the Medical Student Performance Evaluation (MSPE) accurately represented a student’s level of entrustability (Table 1).

Tasks Assessed Across Respondents

All respondents were asked what CEPAER-based clinical tasks they assessed in their clerkship, regardless of whether they reported EPA use (Table 3). There were no associations between the types of tasks assessed at clerkships that formally used EPAs and clerkships that did not, with some exceptions: compared to non-EPA CDs, a higher percentage of EPA CDs reported assessment of (1) entering and discussing orders ($p=0.01$); (2) handoff communication ($p<0.01$); and (3) informed consent ($p=0.04$).

Perceived barriers to implementing EPAs in 2023 were compared to anticipated barriers described in the 2015 CDIM Annual Survey results (Table 4). These results were not tested for statistical associations because the language and methodology for each survey differed. Commonly reported barriers in the 2022 survey included lack of evaluator understanding of EPAs, insufficient faculty and resident/fellow time, and no/inadequate faculty development.

THEMATIC ANALYSIS

Reasons for Not Implementing EPAs

Reasons CDs did not implement EPAs included leadership decisions, perceived barriers, or alternative methods for assessment. Leadership decisions were due to centralized assessment systems; the school either decided against using EPAs or had not yet decided, and control over these decisions was made above the clerkship director level. Barriers included technical (no tracking system/platform), time (bandwidth), and implementation issues (faculty development, reliability questions). Alternative methods often overlapped with central control (i.e., there being a mandate for a standardized assessment form used across clerkships) but also included individual preference for other assessment systems (e.g., RIME). Some CDs reported using assessments that were essentially EPAs but they did not use that terminology or were not assessed using a supervisory scale.

Perceived Benefits of Implementing EPAs

The most frequently cited benefit was related to the concrete, realistic, observable behaviors outlined in the EPAs. CDs found EPAs offered granularity that improved the

Table 1 Characteristics and Validity Efforts of CDs Who Use EPAs and Do Not Use EPAs as Part of Their Core Internal Medicine Clerkship

	Use of EPAs (n = 54) n (%)	No EPAs (n = 60) n (%)	P-value*
Medical school characteristic			
Type (NCES-IPEDS)			0.45
Public	34 (63.0)	33 (55.0)	
Private	20 (37.0)	27 (45.0)	
U.S. census region			
Northeast	12 (22.2)	19 (31.7)	0.30
Midwest	13 (24.1)	14 (23.3)	0.99
South	21 (38.9)	19 (31.7)	0.44
West**	8 (14.8)	8 (13.3)	0.99
Size (number of students)			
Large (≥ 100)	20 (37.0)	29 (48.3)	0.26
Medium (65–95)	17 (31.5)	19 (31.7)	0.99
Small (< 65)	17 (31.5)	12 (20.0)	0.20
Year of initial accreditation (LCME)			
≤ 1942	26 (48.2)	32 (53.3)	0.71
1943–1966	6 (11.1)	10 (16.7)	0.43
1967–1976	8 (14.0)	12 (20.0)	0.62
1977–2022	14 (25.9)	6 (10.0)	0.03
Accreditation status as of 2022 (LCME)			0.67
Full	51 (94.4)	58 (96.7)	
Provisional	3 (5.6)	2 (3.3)	
Clerkship director gender is female	33 (61.1)	35 (58.3)	0.85
Clerkship grading structure			0.63
Pass/fail	10 (18.5)	9 (15.0)	
Not pass/fail	44 (81.5)	51 (85.0)	
Survey question			
Measures undertaken to ensure validity of assessment instruments***, ^			(◇)
Internal review of assessment instruments	26 (55.3)	35 (63.6)	0.95
External review of assessment instruments	12 (48.9)	26 (47.3)	0.99
Use of validated instruments from the literature	8 (17.0)	7 (12.7)	0.99
Faculty development on use of instruments	14 (29.8)	19 (34.6)	0.99
Other	2 (4.3)	2 (3.6)	0.99
None of the above	8 (17.0)	9 (16.4)	0.99
To what extent do you believe faculty development has improved the quality of assessments you obtain?			
To no extent	1 (2.1)	5 (9.8)	0.21
To a small extent	17 (36.2)	17 (33.3)	0.83
To a moderate extent	19 (40.4)	15 (29.4)	0.29
To a great extent	6 (12.8)	7 (13.7)	0.99
Unsure	4 (8.5)	7 (13.7)	0.53
To what extent is assessment information in your clerkship monitored for validity?			
To no extent	10 (18.5)	14 (23.3)	0.65
To a small extent	10 (18.5)	16 (26.7)	0.37
To a moderate extent	15 (27.8)	15 (25.0)	0.83
To a great extent	6 (11.1)	5 (8.3)	0.75
Unsure	13 (24.1)	10 (16.7)	0.36
To what extent are written evaluations a valid measure of students' performance?			
To no extent	0 (0)	1 (1.7)	0.99
To a small extent	6 (11.1)	7 (11.7)	0.99
To a moderate extent	36 (66.7)	40 (66.7)	0.99
To a great extent	8 (14.8)	8 (13.3)	0.99
Unsure	4 (7.4)	4 (6.7)	0.99
To what extent does the MSPE accurately represent students' level of entrustability?			
To no extent	0 (0)	2 (3.3)	0.50
To a small extent	9 (16.7)	13 (21.7)	0.64
To a moderate extent	32 (59.3)	28 (46.7)	0.19
To a great extent	6 (11.1)	4 (6.7)	0.51
Unsure	7 (13.0)	13 (21.7)	0.32

*Dichotomous test (Fisher's exact; two-sided), unless indicated by a ◇, where Pearson's chi-square with Sidak-adjusted p-values were used

**Includes one medical school from an Unincorporated Territory, to ensure data confidentiality

***Respondents could select more than one item

^Excludes 12 respondents who reported "Unsure;" n = 47 and 55, respectively

Abbreviations: *NCES-IPEDS* National Center for Education Statistics—Integrated Postsecondary Education Data System; *LCME* Liaison Committee on Medical Education; *MSPE* Medical Student Performance Evaluation

Table 2 Responses to Key Questions on EPA Development and Use by Internal Medicine Clerkship Directors

Survey question	n (%)
How did you develop EPAs for your clerkship?	(n = 49)
AAMC Core EPAs only	27 (55.1)
AAMC Core EPAs and Institutional EPAs	18 (36.7)
Institutional EPAs only	4 (8.2)
Are EPAs used for formative or summative assessment?	(n = 51)
Both formative and summative	30 (58.9)
Formative only	11 (21.6)
Summative only	9 (17.6)
Neither formative nor summative	1 (2.0)
How are EPAs used for summative assessment?	(n = 54)
Not used for summative assessment	12 (22.2)
Contribute to the final clerkship grade	26 (48.2)
Used for summative assessment other than final grade	16 (29.7)
What type of grading is used on your clerkship?	(n = 54)
Pass/fail	10 (18.5)
Tiered grading*	44 (81.5)
Who submits EPA assessments of students in your clerkship?***	(n = 54)
Sub-interns/acting interns	2 (3.7)
Interns (PGY1)	30 (55.6)
Residents (PGY2 or higher, not including fellows)	44 (81.5)
Fellows	24 (44.4)
Faculty	52 (96.3)
Who receives guidance (online modules, videos, in-person instruction, formal training) on EPAs?***	(n = 54)
Sub-interns/acting interns	2 (3.7)
Interns (PGY1)	18 (33.3)
Residents (PGY2 or higher, not including fellows)	25 (46.3)
Fellows	8 (18.2)
Faculty	31 (57.4)
CD or associate CD	20 (37.0)
No guidance provided	17 (31.5)

Of a total 54 possible respondents who indicated EPA use in their IM clerkship. The number of respondents varied for each item as the table excludes respondents who reported “Unsure.”

*Included honors tiers, letter grades, and numeric scales

***Respondents could select more than one response for this item

effectiveness of direct observation, enhanced feedback specificity, and increased frequency of formative feedback (compared to prior to implementing). Process-wise, CDs appreciated the ability to gauge the progress of entrustment over time. Additionally, CDs perceived students found EPAs provided clarity of expectations. CDs also expressed that the standardization offered by EPAs helped increase the inter-observer reliability of assessors.

Disadvantages of Using EPAs

Many respondents reported concerns over the administrative burden on already busy faculty and residents and the power dynamics that can arise if students are driving the requests. CDs also expressed concern about the number of observations needed for each EPA to be reliable, and how reliability may be affected by limited exposure where multiple observations by a single observer may be challenging. CDs reported some faculty did not fully understand EPAs or

entrustment scores, necessitating faculty development, and that students saw EPAs as a “checkbox” activity, impeding organic feedback.

DISCUSSION

To our knowledge, this is currently the only nationally representative survey of EPA use in IM clerkships. Our study found nearly half of responding CDs use EPAs in some form, indicating substantial adoption since the CEPAERs were introduced in 2014. A greater percentage of CDs reported using EPAs compared to a 2020 snapshot of 135 participating LCME-accredited medical schools, which found that 34% of respondents used the CEPAER as a framework, as part of their assessments, or both.^{12,13} Interestingly, nearly all CDs in our survey reported using at least some elements of the CEPAER, although they did not uniformly refer to them as EPAs or use the entrustment framework for assessments. Together, these findings suggest a notable uptake of EPAs in general and CEPAER in particular in IM clerkships.

We found no substantial differences in demographic characteristics between schools that used EPAs compared to those that did not with one exception: schools of more recent accreditation were associated with adopting EPAs in their IM clerkship. Although such findings may suggest that more recent accreditation is associated with school’s incorporation of “newer” forms of competency-based assessment such as EPAs, this is unclear given the limitations of available data and should be further studied. Despite this, our findings show the priorities of medical school leadership have a strong effect on assessment in general and the ability to implement EPAs in particular, with many CDs reporting the presence or absence of such support as either a facilitator or barrier to their implementation plans, respectively.

Although Table 4 should be interpreted with caution (perceived barriers are not the same construct as experienced barriers), our results demonstrate substantial differences with respect to administrative support, financial support, and coordination across clerkships and campuses. Together, these results indicate schools that successfully added EPAs either experienced fewer barriers or—more likely—were supported in their efforts by their institutions. This theory aligns with a recent study demonstrating the pace of EPA implementation is dependent on system factors such as leadership involvement, investment in data management, and school-specific resources.¹⁴

EPA CDs reported assessment of additional tasks compared to those reported by non-EPA CDs. Given most EPA CDs used the CEPAER as a guide for their implementation, it is possible the list of 13 “Core” activities assisted these CDs in comprehensively planning assessment opportunities for their learners, or—as implied by our thematic findings—that EPA assessment was part of a larger school-wide effort to enhance assessment. These three EPAs (handoffs, placing orders,

Table 3 Clinical Tasks Currently Assessed as Reported by Respondent Clerkship Directors, by Categories of Those Who Were Formally Using EPAs and Those Who Were Not Formally Using EPAs as Part of Their Assessment Schema

Which of the following activities do you CURRENTLY formally assess in your core IM clerkship?	Core internal medicine (IM) clerkship is currently using Entrustable Professional Activities (EPAs) to assess medical students		
	Yes, n (%) n = 54	No, n (%) n = 60	P-value*
Gather a history and perform a physical examination (#1)	53 (98.2)	59 (98.3)	0.99
Prioritize a differential diagnosis following a clinical encounter (#2)	49 (90.7)	57 (95.0)	0.99
Recommend and interpret common diagnostic and screening tests (#3)	45 (83.3)	49 (81.7)	0.99
Enter and discuss orders and prescriptions (#4)	23 (42.6)	9 (15.0)	0.01
Document a clinical encounter in the patient record (#5)	48 (88.9)	51 (85.0)	0.99
Provide an oral presentation of a clinical encounter (#6)	48 (88.9)	52 (86.7)	0.99
Form clinical questions and retrieve evidence to advance patient care (#7)	41 (75.9)	37 (61.7)	0.75
Give or receive a patient handover to transition care responsibility (#8)	23 (42.6)	6 (10.0)	0.001
Collaborate as a member of an inter-professional team (#9)	44 (81.5)	43 (71.7)	0.96
Recognize a patient requiring urgent or emergent care and initiate evaluation and management (#10)	19 (35.2)	12 (20.0)	0.60
Obtain informed consent for tests and/or procedures (#11)	13 (24.1)	3 (5.0)	0.04
Perform general procedures of a physician (#12)	16 (29.6)	12 (20.0)	0.97
Identify system failures and contribute to a culture of safety and improvement (#13)	21 (38.9)	12 (20.0)	0.29

The number listed after each task corresponds to each of the 13 Core EPAs for Entering Residency (CEPAER)¹. Numbers after each description correspond to the specific CEPAER referenced

*Pearson's chi-square (1 degree of freedom): Sidak-adjusted *p*-values

Table 4 Anticipated Compared to Experienced Barriers to Implementing EPAs as Part of Internal Medicine Clerkships. Respondents to the 2015 CDIM Annual Survey Included CDs Who Had Not Implemented EPAs, Whereas the 2022 Annual Survey Respondents Included Only CDs Who Had Implemented EPAs as Part of Their IM Clerkship

2015 CDIM annual survey*		2022 CDIM annual survey**	
Anticipated barrier	%	Experienced barrier	%
Insufficient administrative support to track the student's progress within each EPA	65.1	Insufficient administrative support	18.2
Lack of financial support from the medical school	42.9	Lack of financial support	4.6
Insufficient faculty time to directly observe student progression	79.4	Insufficient faculty time	54.6
–	–	Insufficient resident/fellow time	45.5
Inadequate resources for faculty development	71.4	No/inadequate faculty development	43.2
–	–	No/inadequate resident/fellow development	43.2
Lack of clerkship director understanding regarding EPAs and milestones	7.9	–	–
–	–	Lack of evaluator understanding regarding EPAs	54.6
Need to coordinate with other clerkships to assess all of the EPAs	60.3	Lack of coordination with other clerkships to assess EPAs	29.6
Need to create equivalent systems across sites or regional campuses	34.9	Lack of equivalent systems across sites or regional campuses	11.4

*For 63 of 95 possible respondents

**For 44 of 54 possible respondents who reported "Yes" to "What barriers have you encountered since implementing an EPA system in your IM clerkship?" The 10 additional respondents reported not having encountered any barriers

informed consent) have been reported by IM program directors as having gaps between expected and reported performance in new interns¹⁵; however, there is some debate about when it is most appropriate to assess these specific CEPAER. In a recent summary report, the CEPAER Pilot suggested these three EPAs, among others, might be most appropriate for more advanced trainees such as sub-interns or early interns.⁶ It would be helpful to study how and why these CDs incorporated more advanced EPAs into their curricula.

This study explored the relationship between IM clerkship assessment and the collection of validity evidence.

Although there have been numerous studies on the validity and reliability of EPAs in UME, the existing evidence is limited,^{4,5,16–21} and a recent study investigating the reliability of assessment across multiple institutions found EPAs produced poor reliability even when considering dimensions of faculty development, assessment coaches, or entrustment scale type.²² Although it has long been theorized that "formative reports" of workplace-based assessments (such as EPAs) could be used to inform summative entrustment decisions,²³ there remains a question about whether this is consistently true for CEPAER.²² Given

these findings, it is notable the EPA CDs in our study were no more likely than non-EPA CDs to monitor the validity of their assessments despite a majority reporting EPA use as part of summative assessment and nearly half of EPA CDs using EPAs as part of the final clerkship grade.

While most of our participants used the CEPAER to develop their instruments, validity is highly context-specific and should be established and monitored in individual learning environments.²⁴ Ongoing reporting of competency-based outcomes and the associated validity evidence underlying high-stakes decisions are particularly important given shifts towards pass/fail clerkship grading in recent years.¹² It has been argued that competency-based assessment might prompt institutions to improve the quality of assessment data reported²⁵; however, our study shows EPAs were not associated with increased efforts to establish validity nor did it add additional nuance in the MSPE in regard to student entrustability. Such findings are disappointing given known issues with MSPE variability between schools, leading to concerns over grade inflation and difficulty interpreting medical student performance.^{26–28}

It should be noted that this study did not investigate the timing of when individual schools started using EPAs; therefore, it is possible some respondents did not have sufficient time to develop validity or collect outcomes data on their learners. Given the increased attention on providing high-quality information to residency training programs to best support learners in their transition to advanced training,⁹ we believe our study represents an important call to action: There is a pressing need to collect and—just as vital—share validity evidence on assessments in IM clerkships, and provide this information to stakeholders.

LIMITATIONS

Despite our 80% survey response rate, some degree of measurement error and/or bias might have been introduced by survey non-response or item non-response. Further, only IM CDs were surveyed (not all IM clerkship leadership). The final survey population represented 140 LCME fully or provisionally accredited medical schools with AAIM/CDIM membership out of 154 schools, and five of the 112 responding schools were provisionally accredited. It is possible that responses from provisionally accredited schools might have biased the results.

CONCLUSIONS

A substantial number of U.S. medical schools are using EPAs for formative and/or summative assessment in the IM clerkship and the adoption of EPAs has grown. Lack of support and infrastructure at the institutional level were barriers

to uptake. There is a lack of data demonstrating the desired outcomes of increased student readiness for residency in a reliable and valid fashion. CDs should perform these psychometric analyses to ensure that high-quality data are being generated when used for summative purposes.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11606-024-08991-5>.

Acknowledgements: The authors wish to acknowledge the individuals who developed the 2015 CDIM Annual Survey section focused on milestones and EPAs, which provided foundational information for the current study: Drs. Deborah DeWaay, Joseph Wayne, Deborah Leizman, and Carlos Palacio.

Corresponding Author: Katherine A. Gielissen, MD, MHS-MedEd, Emory University School of Medicine, Atlanta, GA, USA (e-mail: kgielis@emory.edu).

Data Availability The data used in this study were obtained from the Clerkship Directors in Internal Medicine (CDIM) database. Due to the nature of this data, which includes sensitive and potentially identifiable information, access is restricted to protect the privacy of the participants. Researchers interested in accessing the CDIM data for replication or further study should contact the Clerkship Directors in Internal Medicine (CDIM) organization directly.

Declarations:

Ethical Approval: This study and its protocol (Number: 22-AAIM-118) were submitted to Pearl IRB (registered with the U.S. Department of Health and Human Services Office for Human Research Protections as IRB00007772) for exemption determination in accordance with FDA 21 CFR 56.104 and DHHS 45 CFR 46.104 regulations, and were deemed exempt according to FDA 21 CFR 56.104 and 45CFR46.104(b)(2): (2) Tests, Surveys, Interviews on 08/15/2022.

Conflict of Interest: The authors declare that they do not have a conflict of interest.

REFERENCES

1. **Englander R, et al.** Toward defining the foundation of the MD degree: core entrustable professional activities for entering residency. *Acad Med.* 2016;91(10): p. 1352-1358.
2. **Brown DR, et al.** Finding a Path to Entrustment in Undergraduate Medical Education: A Progress Report From the AAMC Core Entrustable Professional Activities for Entering Residency Entrustment Concept Group. *Acad Med.* 2017;92(6): 774-779.
3. **Parsons AS, et al.** Summative Entrustment Decisions in UME: Outcomes of a Binding Entrustment Committee Decision. *Acad Med.* 2023;98(11S):S186-S187.
4. Dunne D, Gielissen K, Slade M, Park YS, Green M. WBAs in UME—how many are needed? A reliability analysis of 5 AAMC core EPAs implemented in the internal medicine clerkship. *J Gen Intern Med.* 2022;37(11):2684-2690.
5. **Ryan MS, et al.** A question of scale? Generalizability of the Ottawa and Chen scales to render entrustment decisions for the core EPAs in the workplace. *Acad Med.* 2022;97(4):552-561.
6. **Amiel J, et al.** Core Entrustable Professional Activities for Entering Residency: Summary of the 10-School Pilot, 2014–2021. 2022: Association of American Medical Colleges.
7. **Linsenmeyer M, et al.** Status of entrustable professional activities (EPA) implementation at colleges of Osteopathic Medicine in the United States and future considerations. *Am Osteopath Assoc.* 2020;120(11):749-760.

8. **Butani L, et al** Entrustable Professional Activity-Based Assessments in Undergraduate Medical Education: A Survey of Pediatric Educators. *Acad Pediatr.* 2021;21(5): 907-911.
9. **Lovell E, Mejciano GC.** Initial Summary Report and Preliminary Recommendations of the Undergraduate Medical Education to Graduate Medical Education Review Committee (UGRC). Coalition for Physician Accountability; July 1, 2021. Available from: <https://physicianaccountability.org/wp-content/uploads/2021/04/UGRC-Initial-Summary-Report-and-Preliminary-Recommendations-1.pdf>.
10. 2015 CDIM Annual Survey of Core Internal Medicine Clerkship Directors. Study Database. Alliance for Academic Internal Medicine: Alexandria, VA.
11. **Fazio SB, et al.** Competency-based medical education in the internal medicine clerkship: a report from the alliance for academic internal medicine undergraduate medical education task force. *Acad Med.* 2018;93(3):421-427.
12. Association of American Medical Colleges. Number of Medical Schools Using Selected Grading Systems. AAMC; March 23, 2022. Available from: <https://www.aamc.org/data-reports/curriculum-reports/interactive-data/grading-systems-use-us-medical-schools>.
13. **Ryan MS, et al.** Competency-based frameworks in medical school education programs: a thematic analysis of the academic medicine snapshots, 2020. *Acad Med.* 2022;97(11S):S63-S70.
14. **Encandela JA, et al.** Entrustable professional activities as a training and assessment framework in undergraduate medical education: A case study of a multi-institutional pilot. *Med Educ Online.* 2023;28(1):2175405.
15. **Angus SV, et al.** Internal medicine residency program directors' views of the core entrustable professional activities for entering residency: an opportunity to enhance communication of competency along the continuum. *Acad Med.* 2017;92(6):785-791.
16. **Ryan MS, et al.** Generalizability of the Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) scale to assess medical student performance on Core EPAs in the workplace: Findings from one institution. *Acad Med.* 2021;96(8):1197-1204.
17. **Rodgers V, et al.** Implementation of a workplace-based assessment system to measure performance of the core entrustable professional activities in the pediatric clerkship. *Acad Pediatr.* 2021;21(3):564-568.
18. **Cutrer WB, et al.** Assessing medical student performance of Entrustable Professional Activities: A mixed methods comparison of Co-Activity and Supervisory Scales. *Med Teach.* 2020;42(3): 325-332.
19. **Soukoulis V, et al.** The use of EPA assessments in decision-making: Do supervision ratings correlate with other measures of clinical performance? *Med Teach.* 2021. 43(11):1323-1329.
20. **Parsons AS, et al.** Concordance of Narrative Comments with Supervision Ratings Provided During Entrustable Professional Activity Assessments. *J Gen Intern Med.* 2022;37(9): 2200-2207.
21. **Keeley MG, et al.** Fidelity to Best Practices in EPA Implementation: Outcomes Supporting Use of the Core Components Framework From the University of Virginia Entrustable Professional Activity Program. *Acad Med.* 2022;97(11): 1637-1642.
22. Ryan MS, Gielissen KA, Shin D, Perera RA, Gusic M, Ferenchick G, Ownby A, Cutrer WB, Obeso V, Santen SA. How well do workplace-based assessments support summative entrustment decisions? A multi-institutional generalisability study. *Med Educ.* 2024;58(7):825-837.
23. **Ten Cate O, et al.** Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach.* 2015;37(11):983-1002.
24. **Downing SM.** Validity: on the meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830-837.
25. **Hauer KE, Lucey CR.** Core clerkship grading: the illusion of objectivity. *Acad Med.* 2019;94(4):469-472.
26. **Tisdale RL, et al.** A Retrospective Analysis of Medical Student Performance Evaluations, 2014-2020: Recommend with Reservations. *J Gen Intern Med.* 2022;37(9):2217-2223.
27. **Gielissen K, et al.** Consequence in Competency-Based Education: Individualize, but Do Not Compromise. *J Gen Int Med.* 2022;37(9):2146-2148.
28. **Wang A, et al.** Reimagining a pass/fail clinical core clerkship: a US residency program director survey and meta-analysis. *BMC Med Educ.* 2023;23(1):788.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.