



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Patterson, S;Shaw, L;Rank, MM;Vaughan, B

Title:

Assessments Used for Summative Purposes during Internal Medicine Specialist Training:
A Rapid Review

Date:

2023-10-01

Citation:




Patterson, S., Shaw, L., Rank, M. M. & Vaughan, B. (2023). Assessments Used for Summative Purposes during Internal Medicine Specialist Training: A Rapid Review. *Education Sciences*, 13 (10), <https://doi.org/10.3390/educsci13101057>.

Persistent Link:

<https://hdl.handle.net/11343/339493>

Review

Assessments Used for Summative Purposes during Internal Medicine Specialist Training: A Rapid Review

Scott Patterson, Louise Shaw , Michelle M Rank *  and Brett Vaughan 

Faculty of Medicine, Dentistry and Health Sciences, Centre for Digital Transformation of Health, University of Melbourne, Carlton, VIC 3010, Australia; scott@arpost.net (S.P.); louise.shaw@unimelb.edu.au (L.S.); brett.vaughan@unimelb.edu.au (B.V.)

* Correspondence: michelle.rank@unimelb.edu.au

Abstract: Assessments used for summative purposes of patient-facing clinical competency in specialist internal medicine training are high-stakes, both to doctors in training, as it is a prerequisite for qualification, as well as their community of prospective patients. A rapid review of the literature evaluated methods of assessments used for summative purposes of patient-facing clinical competency during specialist internal medicine training in Australia. Four online databases identified literature published since the year 2000 that reported on summative assessment in specialist medical training. Two reviewers screened and selected eligible studies and extracted data, with a focus on evidence of support for the criteria for good assessment as set out in the 2010 Ottawa Consensus framework for good assessment. Ten eligible studies were included. Four studied the mini-clinical evaluation exercise (mini-CEX), two the Royal Australasian College of Physicians short case exam, three a variety of Entrustable Professional Activities (EPAs) or summative entrustment and progression review processes, and one a novel clinical observation tool. The mini-CEX assessment demonstrated the most evidence in support of the Ottawa criteria. There was a paucity of published evidence regarding the best form of summative assessment of patient-facing clinical competency in specialist internal medicine training.

Keywords: rapid review; summative assessment; clinical competency; medical training



Citation: Patterson, S.; Shaw, L.; Rank, M.M.; Vaughan, B. Assessments Used for Summative Purposes during Internal Medicine Specialist Training: A Rapid Review. *Educ. Sci.* **2023**, *13*, 1057. <https://doi.org/10.3390/educsci13101057>

Academic Editor: Han Reichgelt

Received: 9 August 2023

Revised: 16 September 2023

Accepted: 22 September 2023

Published: 20 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Internal medicine training in Australia is structured and accredited by the Royal Australasian College of Physicians (RACP), under federal government regulation administered by the Australian Medical Council (AMC). There is evidence to support a correlation between success during clinical training and indices of preventative care and chronic disease management [1,2]. Assessment of medical knowledge and theory during training may be reliably and feasibly evaluated through written examination [3]. However, assessment of patient-facing clinical competencies, such as taking a history or performing a physical examination, requires observation by a skilled assessor [4]. This type of assessment may also involve a real patient in a genuine healthcare environment or a simulated participant (actor) and require assessors who are generally specialists, further increasing logistical difficulty and cost.

Summative assessment evaluates student learning against a defined benchmark or series of learning goals [5]. Assessment used for summative purposes in medical education is high stakes for learners, as the information gained is used to make decisions regarding progression through training or certification [6]. This is especially true in specialist medical training, where summative assessment is additionally high stakes to the wider community since it sets a standard for knowledge and clinical care [7]. A variety of assessment instruments for summative purposes are used in specialist Internal Medicine training programmes (for example, in Australia, New Zealand, the United Kingdom, the United

States of America, and Canada). These include a variety of assessments of patient-facing clinical competency, notably long cases, short cases, objective structured clinical examinations (OSCEs), mini-clinical evaluation exercises (mini-CEX), and other work-based assessments [4,8] (see Supplementary Material Table S1). In Australia, the assessment of patient-facing clinical competency is made by a combination of traditional long and short cases in an examination termed the Divisional Clinical Examination [9]. However, this examination has been criticised for inadequate reliability [10,11].

Assessment literature is moving more toward programmatic assessment rather than the study of the properties of individual assessment tools [12]. Programmatic assessments support a portfolio of assessments as being more reliable and authentic than a single assessment [13–15]. The change in focus to programmatic assessment emphasises the breadth of assessment types, each with different properties and capabilities to sample the whole curriculum. Single assessments are educationally valuable; multiple assessments combined into a whole can justify summative decisions, with the highest-stakes decisions requiring the most supportive data [13,14,16]. In the programmatic assessment model, individual assessment tools do not need to be as thoroughly analysed as the entire program of assessment. Due to the significant short- and long-term consequences of decisions justified by high-stakes assessments, there is intense scrutiny from all stakeholders: learners, educators, accredited training organisations, and regulators. These are arguably the most important decisions made during medical training [7].

Given the complexity of assessing patient-facing clinical competency and the high stakes involved during specialist internal medicine training, from both the learner and community perspective, it is important that the best methods of assessment are chosen. The Ottawa Consensus identified seven criteria for good assessment: validity, reproducibility, equivalence, feasibility, educational effect, catalytic effect, and acceptability [17]. The most important of the Ottawa criteria for summative assessment is validity (supports the use of the results of an assessment for a defined purpose), reproducibility (assessment results would not change under similar circumstances), and equivalence (assessment scores will be the same if administered in different institutions or cycles of testing), in addition to feasibility (the assessment is practical, realistic and appropriate for the circumstances) and acceptability (stakeholders find the results and assessment process to be credible) [18].

This rapid review aimed to explore the published literature on methods of assessment used for summative purposes of patient-facing clinical competency during specialist internal medicine training, using the critical lens of the criteria for good assessment outlined in the 2010 Ottawa consensus statement [17].

2. Materials and Methods

A rapid review was performed to obtain a contemporary assessment of the literature in a 3-month time period. Rapid reviews are an efficient tool to gather evidence quickly [19]. This can be achieved because a rapid review uses a simplified methodology compared to a systematic review [20]. This review followed the Cochrane Rapid Review Methods Group guidance for rapid reviews [21]. Our rapid review sought to answer the question: ‘What is known about assessments used for summative purposes in specialist internal medicine patient-facing clinical competency?’

Eligibility criteria: The review was limited to studies that were written in the English language, peer-reviewed, published since the year 2000, and with full-text access available. Additional inclusion criteria are that the study was empirical research (quantitative, qualitative, or mixed methods), the context was postgraduate or specialist medical training (that is, post-initial medical qualification), the assessment was for summative purposes (final/progression determining), was a competency assessment (‘shows how’ or ‘does’) of patient-facing skills. Excluded studies were reviews and quality improvement studies, not predominantly adult or pediatric internal medicine (for example, surgery, anaesthetics, intensive care, and family medicine). We did not include programmatic assessments as this is a broad assessment type that involves longitudinal collection of multiple assessment

points and types not limited to summative assessment and thus sits outside of the scope of this rapid review. Eligibility criteria were modified iteratively, as the search strategy was trialled to ensure that the literature comprehensively covered the research question and the volume of literature was manageable within a three-month timeframe. Some assessment tools in medical education have formative uses (or, in some cases, are used for both formative and summative purposes). However, for the purpose of this review, only assessments that were used with a summative purpose or intent were included. Therefore, how the tool was used determined whether the study was included in the review. For example, the Royal Australasian College of Physicians does not use the mini-CEx as a summative assessment, but the mini-CEx is used with summative intent in the USA training system. The review, therefore, includes some studies from the USA that reference the use of the mini-CEx but none from Australia.

Literature search: An academic librarian was consulted to construct suitable search terms, with iterative modification as required in setting the eligibility criteria. A search strategy was trialled using a single database, MEDLINE. Only 17 results were generated using the search terms: “summative” AND “internal medicine” AND “performance OR competen* OR assessment OR exam” AND “postgraduate* OR postgraduate* OR graduate*”. As there was insufficient literature to form a satisfactory review, the search parameters were expanded beyond internal medicine. The search strategy was not expanded beyond specialist training (to consider training prior to completion of an initial medical qualification), nor to include assessments without a summative purpose, such as those typically included in programmatic assessment regimes. The rationale for this restriction was that summative assessment in a specialist setting is high-stakes and the final hurdle to independent practice relevant to future public health care quality. A search strategy replacing the search term “internal medicine” with “medicine OR medical OR clinical” generated 329 results. This search strategy was then generalised to four databases which comprised the final search: MEDLINE, PubMed, Web of Science, and Scopus. This search strategy was then generalised to four databases which comprised the final search and yielded a total of 1269 studies imported for screening as follows: MEDLINE (n = 351), PubMed (n = 388), Web of Science (n = 236), and Scopus (n = 294). A further 35 records were identified through handsearching via Google Scholar and web searching. Moreover, 816 duplicate studies were removed, leaving 488 studies. No studies that met all eligibility criteria were excluded based on the unavailability of full text.

Study selection: All papers identified from the final search strategy that included summative assessment in all specialist training programmes were uploaded to Covidence—a software package that streamlines evidence synthesis at each step of the systematic review process. For the remaining 488 studies, two reviewers independently performed title and abstract screening and resolved all conflicts. Moreover, 455 studies were excluded. The most common reason for exclusion was related to the population under assessment being focused on a trainee program other than specialist medicine. References lists from the studies that reached full-text review were also reviewed for any potentially relevant studies, and these were added to the full-text review set. Two reviewers performed full-text reviews and resolved conflicts.

Data extraction: A single reviewer performed data extraction using a piloted data extraction table: date, jurisdiction, specific population and inclusion/exclusion criteria, number of participants, assessment under study, outcome measure/comparator, and attributes relevant to the Ottawa criteria for good assessment. A second reviewer checked the accuracy and completeness of data extraction. All ten studies progressed to this stage.

Risk of bias assessment: Risk of bias assessment was omitted because no judgements were being made about the included study quality.

Data synthesis: For each Ottawa criterion, whether the study provided evidence of support or evidence of concern was assessed. The studies were also assessed for whether support or concern was implied or discussed without supportive evidence.

3. Results

Overall, 1269 records were identified from database searching. An additional 35 records were identified through handsearching. After duplicate, title, and abstract screening, 34 studies proceeded to full-text review, and 10 studies were included. A PRISMA flow diagram is detailed in Figure 1.

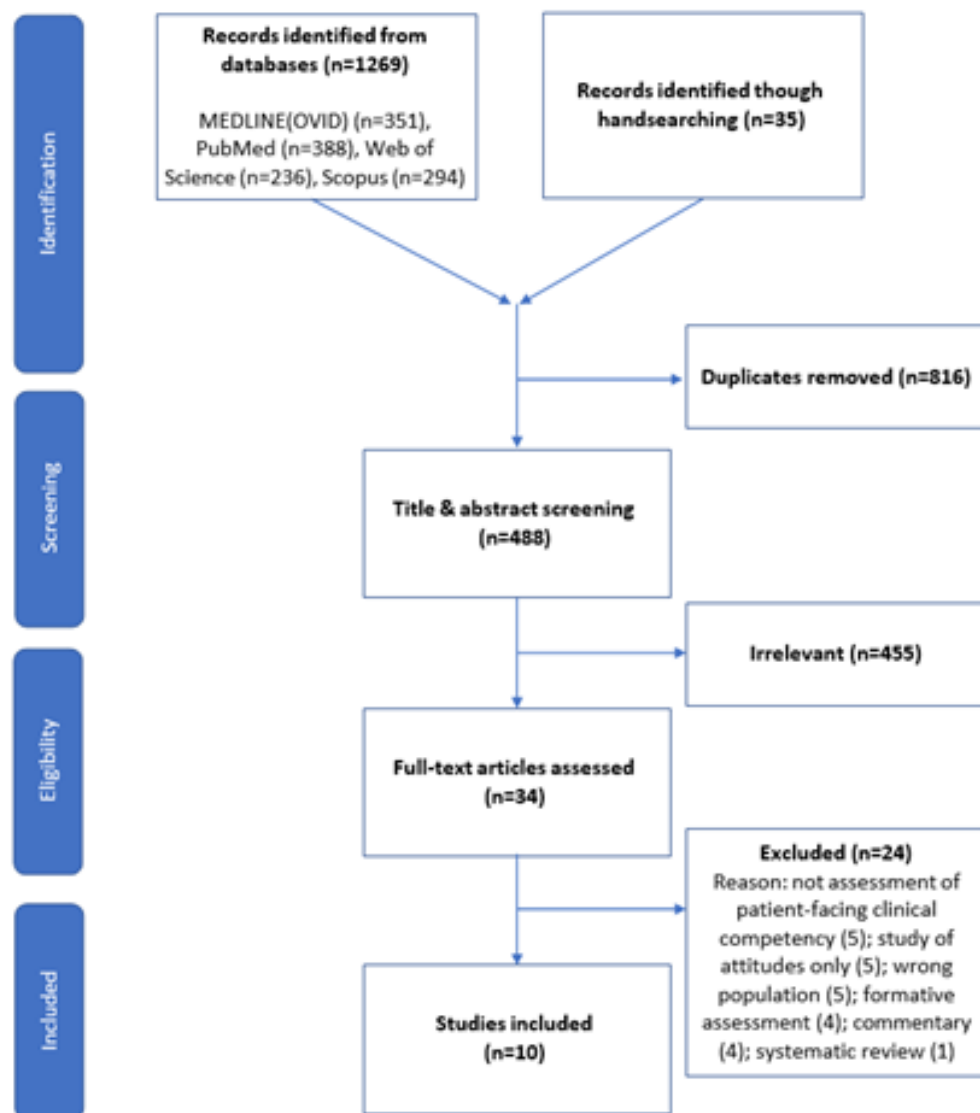


Figure 1. PRISMA diagram of study selection.

4. Study Characteristics

Half (50%) of the included articles were from the USA [22–26], with the remainder from Australia [11,27], The Netherlands [14,28], and Canada [29]. As per the selection criteria, all studies were performed predominantly in an internal medicine training context. Half (50%) of the studies related to adult internal medicine [14,22,23,26,29], three were related to paediatric internal medicine [24,25,28], and two to a combination [11,27]. Trainee cohorts under study ranged from 22 to 1190 individuals (median 105). Six studies had a cross-sectional design [14,22,25,27–29], whereas three were single cohort studies [23,24,26]. See Table 1.

Table 1. Table of study characteristics.

Study ID	Title	Country	Study Design	Population Description	Number of Participants
Durning et al., 2002 [22]	Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training	United States	Cross-sectional study	Postgraduate year doctors (“residents”) at a medical centre	23
Hatala et al., 2006 [29]	Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination	Canada	Cross-sectional study	Postgraduate year 4 resident doctors preparing for RCPSC IM examination	22
Moonen-vanLoon et al., 2013 [14]	Composite reliability of a workplace-based assessment toolbox for postgraduate medical education	Netherlands	Cross-sectional study	Dutch residents at 59 hospitals	953 total (466 first-year subset)
Norcini et al., 2003 [23]	The mini-CEX: a method for assessing clinical skills	United States	Cohort study	Non-systematic volunteer residents at 21 programs	421
Schumacher et al., 2018 [24]	Influence of Clinical Competency Committee Review Process on Summative Resident Assessment Decisions	United States	Cohort study	Paediatric residency programs in Association of Pediatric Program Directors (APPD) Longitudinal Educational Assessment Research Network (LEARN)	463 residents, 155 CCC members, 14 PDs
Schumacher et al., 2019 [25]	Justifications for Discrepancies Between Competency Committee and Program Director Recommended Resident Supervisory Roles	United States	Cross-sectional study	Paediatric residency clinical competency committee members and program directors at 14 USA residency programs	98
Smith et al., 2017 [26]	Successful Implementation of a Direct Observation Program in an Ambulatory Block Rotation	United States	Cohort study	Internal medicine residents rotating through an ambulatory care block	57 residents and 14 faculty
Smit et al., 2019 [28]	A National Process to Enhance the Validity of Entrustment Decisions for Dutch Pediatric Residents	Netherlands	Cross-sectional study	Program directors, attending staff, and residents from all Dutch paediatric programs	112 residents and 37 faculty
Wilkinson et al., 2008 [11]	Reliability of the long case	Australia	Other: Statistical analysis of results	RACP Clinical Exam sitters 2005 (Aus only) and 2006 (Aus/NZ)	1190 examinations (915 Adult, 273 Paediatric, 2 supplementary)
Wilkinson et al., 2010 [27]	The reliability of long and short cases undertaken as practice for a summative examination	Australia	Cross-sectional study	DCE candidates from 5 hospitals	59

5. Assessment and Comparator/Outcome

The most frequent assessment method reported in the reviewed literature was the Mini Clinical Examination Exercise (mini-CEX) in four studies [14,22,23,29]. Three studies investigated Entrustable Professional Activities (EPAs) and/or Clinical Competence Committee (CCC) decision-making [24,25,28]. Two studies investigated the Royal Australasian College of Physicians (RACP) Divisional Clinical Examination (DCE) short case [11,27]. One study described a novel clinical observation tool. [26]. No studies investigated the Objective Structure Clinical Examination (OSCE). Seven studies used the assessment(s) under study to determine reliability without other outcome measures or comparators. Three studies each used another assessment as a comparator (specifically: American College of Physicians In-Training Examination; Canadian Royal College of Physicians and Surgeons of Canada Internal Medicine examination; and RACP DCE). See Table 2.

Table 2. Assessment tool and outcome/comparator of included studies Mini-CEX: Mini Clinical Examination Exercise; EPA: Entrustable Professional Activities; CCC: Clinical Competence Committee decision making. Short case: Royal Australasian College of Physicians (RACP) Divisional Clinical Examination (DCE) short case; RCPSC: Royal College of Physicians and Surgeons of Canada; ITE: American College of Physicians–American Society of Internal Medicine In-Training Examination; MEF: American Board of Internal Medicine’s (ABIM’s) monthly evaluation form (MEF); PD: Program Directors Supervisory role.

Study ID	Assessment	Outcome/Comparator
Durning et al., 2002 [22]	Mini-CEX	Itself MEF ITE
Hatala et al., 2006 [29]	Mini-CEX	RCPSC IM examination
Moonen-vanLoon et al., 2013 [14]	Mini-CEX Direct Observation Procedural Skills Multi-Source Feedback	Internal
Norcini et al., 2003 [23]	Mini-CEX	Itself
Schumacher et al., 2018 [24]	Supervisory role entrustment (summative milestone profile)	Between individual CCC members and CCC decision
Schumacher et al., 2019 [25]	CCC entrustment decision	PD supervisory role
Smith et al., 2017 [26]	Direct Observation procedural Skills	Self-reported clinical skills and faculty preference
Smit et al., 2019 [28]	EPA	Acceptability
Wilkinson et al., 2008 [11]	Short case (& long case)	Itself
Wilkinson et al., 2010 [27]	Practice short case (and practise long case)	examination short case (and exam long case)

6. Ottawa Criteria for Good Assessment

No study gave evidence for all criteria within the Ottawa framework. The most frequently referenced criterion was validity, where six studies explicitly described supportive evidence, and two others implied support for validity. Five studies explicitly reported evidence for reproducibility. Evidence for equivalence was explicitly stated in one study, implied or discussed in two, and concerns regarding equivalence were raised in another two. The Ottawa criteria for which evidence was least frequently given were educational effect, catalytic effect, and acceptability. The assessment method with the most evidence of support for the Ottawa criteria was the mini-CEX. These results are summarised in Table 3.

Table 3. Reporting of Ottawa framework criteria in included studies. Legend: evidence of support (with data): +; implied/discussed evidence of support: [+]; implied/discussed evidence of concern: [-]; no evidence or discussion: blank.

Study and Assessment Tool	Validity	Reproducibility	Ottawa Framework Criteria				Acceptability
			Equivalence	Feasibility	Educational Effect	Catalytic Effect	
Mini-CEX							
Durning 2002 [22]	+	+		[+]			
Hatala 2006 [29]	+	+			[+]		
Moonen-vanLoon 2013 [14]	[+]	+	[+]	+			
Norcini 2003 [23]	+	+	[+]	[+]		[+]	+
Clinical Competency Committee							
Schumacher 2018 [24]			[-]				
Schumacher 2019 [25]	+						
Novel							
Smith 2017 [26]	[+]			+	+	+	+
Entrustable Professional Activities							
Smit 2019			[-]	+		+	+
RACP short case							
Wilkinson 2008 [11]	+		+	+			
Wilkinson 2010 [27]	+	+		[+]	[-]	[+]	

7. Discussion

This rapid review of assessments used for summative purposes of patient-facing clinical skills during specialist internal medicine training in the US, Australia, Canada, and the Netherlands found limited literature to support the efficacy of any single assessment method. Only ten eligible English language publications were found that had been published since the year 2000. From the perspective of the Ottawa criteria of good assessment, the reviewed studies focused on the validity or reliability of the assessment method, with less emphasis on the criteria of educational effect, catalytic effect, and acceptability.

Although all elements within the Ottawa framework are relevant, certain elements have varying importance depending on the purpose of the assessment under consideration. Formative assessment is particularly valued for the educational and catalytic effect on driving current and future learning [18]. This is less relevant for assessments used for summative purposes, and therefore, it is not surprising that these two elements had the least supportive evidence in the studies reviewed. Validity is essential to all assessments, and reproducibility and equivalence are the two other most important elements in effective summative assessment methods. It is notable that only one of the ten reviewed studies explicitly provided evidence for equivalence, with two others providing implied support or discussion. Two further studies implied or discussed evidence of concerns regarding Equivalence. Future work could move past unpacking the Ottawa criteria towards a more integrative review that considers the extent to which the evidence on a suite of clinical assessments provides guidance and narrative on how best to assess patient-facing competencies.

The most reported tool to summatively assess patient-facing clinical skills was the mini-CEX, followed by the RACP DCE short case. Interestingly, literature concerning the OSCE did not meet the specified criteria despite it being in use for many years and employed in the Membership of the Royal Colleges of Physicians (MRCP) Practical Applications of Clinical Examination Skills (PACES) examination in the United Kingdom [30,31]. The review also did not reveal any literature on other work-based assessments that may be used for summative purposes, such as direct observation of procedural skills (DOPS) or Case-based discussion. There is also no mention of 'eportfolio' as a summative tool in itself. Some of the most recently published studies explored the combined use of multiple workplace-based assessments, often individually with a primary formative purpose, combined to make summative assessment decisions, often by a Clinical Competency Committee [24,32].

No assessment of the quality of the published studies was made, and therefore, it is impossible to determine the assessment tool with the strongest evidence of support. However, the mini-CEX had the most (numerically) evidence of support, which is in keeping with the considerable literature on the mini-CEX in other disciplines. A systematic review of tools for direct observation and assessment of medical trainees found the mini-CEX to have the strongest validity evidence [33]. A recent review of the mini-CEX found generally good evidence for reliability, validity, feasibility, and acceptability, with less evidence for educational impact [34]. Notably, none of the 58 studies in this published review reported use primarily for summative decision-making [34].

The scope of this rapid review was deliberately narrow, which may have impacted the availability of published studies on summative assessment of patient-facing clinical competency assessment during Internal Medicine specialist training. However, in Australia, Internal Medicine specialists are the largest cohort of specialist doctors after Family Medicine specialists, and therefore, the size of the trainee cohort is not well correlated with the relative lack of investigative evidence [35]. In comparison to the assessment of knowledge or theoretical applications of knowledge, the assessment of performative competencies is difficult, presents considerable logistic difficulties, and is inherently time-consuming and costly [8]. Research on the assessment of performative competencies likely inherits many of these difficulties, and the high-stakes nature of summative assessment further exacerbates these concerns.

Narrowing the search to only assessments used for summative purposes may have excluded a large body of literature on the use of workplace assessments that are used more

formatively but also used as part of programmatic assessment to make final decisions on specialists' progression. The limitations imposed on the search criteria may also mean that opportunities were missed to explore how to better undertake clinical-facing assessments within programmatic assessment. The high-stakes nature of the summative assessments at the late stage in specialist medical training programs should be a powerful driver for a body of evidence to justify training and assessment programs. However, there may be concerns that such research could undermine existing long-standing protocols and expose a training program to challenge, and overall, such evidence is lacking [36]. In jurisdictions where such concerns hinder research, there may be a need for a regulatory imperative to require an evidence-based assessment system.

8. Strengths and Limitations of This Review

A rapid review format was chosen given the desire for a contemporary assessment of the literature. However, the narrow search criteria adopted may have excluded studies that explored educational impact and acceptability. The narrow search terms may also have contributed to the exclusion of studies from 'non-western' settings. To provide more meaningful information on how a suite of assessments provides guidance and a narrative on how best to assess patient-facing competencies, a narrative or integrative review may have been a better choice for the literature review. Internal medicine training in resource-rich, developed countries is highly regulated, and there are a limited number of accredited training providers, often only one per country. As such, there may be commercial-in-confidence concerns that have prevented the publication of data that might support (or otherwise) systems of assessment in active use. This context also makes it unlikely that systems of assessment could ever be directly compared within one jurisdiction, as this would result in trainees within a single training program being subjected to different assessments to measure acceptable clinical performance.

9. Conclusions

Summative assessment in specialist medical training is high stakes to learners, educators, and the wider community of prospective future patients. There is a paucity of literature regarding assessments for summative purposes of patient-facing clinical competencies during internal medicine training. The existing, limited literature supports the use of mini-CEX as a valid, reliable, and feasible tool. The lack of supportive literature suggests a systematic impediment to research in this area. Systems of summative assessment require an evidence base to justify their use. The absence of a satisfactory evidence base may be an opportunity for a regulatory imperative.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/educsci13101057/s1>. Table S1: Methods of summative assessment of clinical training during postgraduate internal medicine specialization.

Author Contributions: Conceptualization, S.P., L.S., B.V. and M.M.R.; methodology, S.P., L.S., B.V. and M.M.R.; formal analysis, S.P. and B.V.; investigation, S.P. and B.V.; resources, S.P., L.S., B.V. and M.M.R.; validation S.P., L.S., B.V. and M.M.R.; data curation, S.P. and B.V.; writing—original draft preparation, L.S.; writing—review and editing, S.P., L.S. and M.M.R.; supervision, L.S., B.V. and M.M.R.; project administration, S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank academic librarians for their assistance in the literature search.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kawasumi, Y.; Ernst, P.; Abrahamowicz, M.; Tamblyn, R. Association between physician competence at licensure and the quality of asthma management among patients with out-of-control asthma. *Arch. Intern. Med.* **2011**, *171*, 1292–1294. [CrossRef] [PubMed]
2. Tamblyn, R.; Abrahamowicz, M.; Dauphinee, W.D.; Hanley, J.A.; Norcini, J.; Girard, N.; Grand'Maison, P.; Brailovsky, C. Association between licensure examination scores and practice in primary care. *JAMA* **2002**, *288*, 3019–3026. [CrossRef] [PubMed]
3. Miller, G.E. The assessment of clinical skills/competence/performance. *Acad. Med.* **1990**, *65*, S63–S67. [CrossRef] [PubMed]
4. Boursicot, K.; Kemp, S.; Wilkinson, T.; Findyartini, A.; Canning, C.; Cilliers, F.; Fuller, R. Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference. *Med. Teach.* **2021**, *43*, 58–67. [CrossRef] [PubMed]
5. Boud, D. Assessment and learning: Contradictory or complementary. In *Assessment for Learning in Higher Education*; Routledge: London, UK, 1995; pp. 35–48.
6. Wass, V.; Bowden, R.; Jackson, N.; Jameson, A.; Khan, A. The principles of assessment design. In *Assessment in Medical Education and Training Oxford*; Radcliffe: Oxford, UK, 2007; pp. 11–26.
7. Touchie, C.; Kinnear, B.; Schumacher, D.; Caretta-Weyer, H.; Hamstra, S.J.; Hart, D.; Gruppen, L.; Ross, S.; Warm, E.; Ten Cate, O. On the validity of summative entrustment decisions. *Med. Teach.* **2021**, *43*, 780–787. [CrossRef] [PubMed]
8. Boursicot, K.; Etheridge, L.; Setna, Z.; Sturrock, A.; Ker, J.; Smee, S.; Sambandam, E. Performance in assessment: Consensus statement and recommendations from the Ottawa conference. *Med. Teach.* **2011**, *33*, 370–383. [CrossRef] [PubMed]
9. The Royal Australasian College of Physicians. Adult Medicine DCE. n.d. Available online: <https://www.racp.edu.au/trainees/examinations/divisional-clinical-examination/adult-medicine-dce> (accessed on 21 July 2022).
10. Gorman, D.; Scott, J. *Time for a Medical Educational Change in Time*; Wiley Online Library: Hoboken, NJ, USA, 2006; pp. 687–689.
11. Wilkinson, T.J.; Campbell, P.J.; Judd, S.J. Reliability of the long case. *Med. Educ.* **2008**, *42*, 887–893. [CrossRef] [PubMed]
12. Van Der Vleuten, C.P.; Schuwirth, L.W. Assessing professional competence: From methods to programmes. *Med. Educ.* **2005**, *39*, 309–317. [CrossRef]
13. Schuwirth, L.W.; Van Der Vleuten, C.P. Current assessment in medical education: Programmatic assessment. *J. Appl. Test. Technol.* **2019**, *20*, 2–10.
14. Moonen-van Loon, J.; Overeem, K.; Donkers, H.; Van der Vleuten, C.; Driessen, E. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Adv. Health Sci. Educ.* **2013**, *18*, 1087–1102. [CrossRef]
15. van der Vleuten, C.P. Revisiting ‘Assessing professional competence: From methods to programmes’. *Med. Educ.* **2016**, *50*, 885–888. [CrossRef] [PubMed]
16. van der Vleuten, C.P.; Schuwirth, L.; Driessen, E.; Dijkstra, J.; Tigelaar, D.; Baartman, L.; Van Tartwijk, J. A model for programmatic assessment fit for purpose. *Med. Teach.* **2012**, *34*, 205–214. [CrossRef] [PubMed]
17. Norcini, J.; Anderson, B.; Bollela, V.; Burch, V.; Costa, M.J.; Duvivier, R.; Galbraith, R.; Hays, R.; Kent, A.; Perrott, V. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Med. Teach.* **2011**, *33*, 206–214. [CrossRef] [PubMed]
18. Norcini, J.; Anderson, M.B.; Bollela, V.; Burch, V.; Costa, M.J.; Duvivier, R.; Hays, R.; Palacios Mackay, M.F.; Roberts, T.; Swanson, D. 2018 Consensus framework for good assessment. *Med. Teach.* **2018**, *40*, 1102–1109. [CrossRef] [PubMed]
19. Ganann, R.; Ciliska, D.; Thomas, H. Expediting systematic reviews: Methods and implications of rapid reviews. *Implement. Sci.* **2010**, *5*, 56. [CrossRef] [PubMed]
20. Khangura, S.; Konnyu, K.; Cushman, R.; Grimshaw, J.; Moher, D. Evidence summaries: The evolution of a rapid review approach. *Syst. Rev.* **2012**, *1*, 10. [CrossRef] [PubMed]
21. Garritty, C.; Gartlehner, G.; Nussbaumer-Streit, B.; King, V.J.; Hamel, C.; Kamel, C.; Affengruber, L.; Stevens, A. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *J. Clin. Epidemiol.* **2021**, *130*, 13–22. [CrossRef] [PubMed]
22. Durning, S.J.; Cation, L.J.; Markert, R.J.; Pangaro, L.N. Assessing the reliability and validity of the mini—Clinical evaluation exercise for internal medicine residency training. *Acad. Med.* **2002**, *77*, 900–904. [CrossRef]
23. Norcini, J.J.; Blank, L.L.; Duffy, F.D.; Fortna, G.S. The mini-CEX: A method for assessing clinical skills. *Ann. Intern. Med.* **2003**, *138*, 476–481. [CrossRef]
24. Schumacher, D.J.; King, B.; Barnes, M.M.; Elliott, S.P.; Gibbs, K.; McGreevy, J.F.; Del Rey, J.G.; Sharma, T.; Michelson, C.; Schwartz, A. Influence of clinical competency committee review process on summative resident assessment decisions. *J. Grad. Med. Educ.* **2018**, *10*, 429–437. [CrossRef]
25. Schumacher, D.J.; Poynter, S.; Burman, N.; Elliott, S.P.; Barnes, M.; Gellin, C.; Del Rey, J.G.; Sklansky, D.; Thoreson, L.; King, B. Justifications for discrepancies between competency committee and program director recommended resident supervisory roles. *Acad. Pediatr.* **2019**, *19*, 561–565. [CrossRef] [PubMed]
26. Smith, J.; Jacobs, E.; Li, Z.; Vogelmann, B.; Zhao, Y.; Feldstein, D. Successful implementation of a direct observation program in an ambulatory block rotation. *J. Grad. Med. Educ.* **2017**, *9*, 113–117. [CrossRef] [PubMed]
27. Wilkinson, T.; D’Orsogna, L.; Nair, B.; Judd, S.; Frampton, C. The reliability of long and short cases undertaken as practice for a summative examination. *Intern. Med. J.* **2010**, *40*, 581–586. [CrossRef] [PubMed]
28. Smit, M.P.; de Hoog, M.; Brackel, H.J.; Ten Cate, O.; Gemke, R.J. A national process to enhance the validity of entrustment decisions for Dutch pediatric residents. *J. Grad. Med. Educ.* **2019**, *11*, 158–164. [CrossRef] [PubMed]

29. Hatala, R.; Ainslie, M.; Kassen, B.O.; Mackie, I.; Roberts, J.M. Assessing the mini-clinical evaluation exercise in comparison to a national specialty examination. *Med. Educ.* **2006**, *40*, 950–956. [[CrossRef](#)]
30. Harden, R.M.; Stevenson, M.; Downie, W.W.; Wilson, G. Assessment of clinical competence using objective structured examination. *Br. Med. J.* **1975**, *1*, 447–451. [[CrossRef](#)] [[PubMed](#)]
31. Membership of the Royal Colleges of Physicians of the United Kingdom. MRCP(UK) Examinations. Available online: <https://www.mrcpuk.org/mrcpuk-examinations> (accessed on 8 April 2021).
32. Goldhamer, M.E.J.; Martinez-Lage, M.; Black-Schaffer, W.S.; Huang, J.T.; Co, J.P.T.; Weinstein, D.F.; Pusic, M.V. Reimagining the Clinical Competency Committee to Enhance Education and Prepare for Competency-Based Time-Variable Advancement. *J. Gen. Intern. Med.* **2022**, *37*, 2280–2290. [[CrossRef](#)] [[PubMed](#)]
33. Kogan, J.R.; Conforti, L.; Bernabeo, E.; Iobst, W.; Holmboe, E. Opening the black box of clinical skills assessment via observation: A conceptual model. *Med. Educ.* **2011**, *45*, 1048–1060. [[CrossRef](#)]
34. Mortaz Hejri, S.; Jalili, M.; Masoomi, R.; Shirazi, M.; Nedjat, S.; Norcini, J. The utility of mini-Clinical Evaluation Exercise in undergraduate and postgraduate medical education: A BEME review: BEME Guide No. 59. *Med. Teach.* **2020**, *42*, 125–142. [[CrossRef](#)]
35. Australian Institute of Health and Welfare. Medical Practitioners Workforce 2015, What Types of Medical Practitioners Are There? Available online: <https://www.aihw.gov.au/reports/workforce/medical-practitioners-workforce2015/contents/what-types-of-medical-practitioners-are-there> (accessed on 18 October 2022).
36. Hutchinson, L.; Aitken, P.; Hayes, T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. *Med. Educ.* **2002**, *36*, 73–91. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.