Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

First-Cohort Micro-Credentialing in Thoracentesis and Paracentesis: Feasibility, Uptake, Retention, and Early Implementation Barriers in Procedural Training

Matías Aguilera
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Felipe Riquelme
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Mateo Ferrer
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Felipe Araya
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Vlada Vasilyeva
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
María Inés Gaete
Pontifical Catholic University of Chile: Pontificia Universidad Catolica de Chile
Gabriel Escalona
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Francisco Garrido
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Elga Zamorano
Pontificia Universidad Católica de Chile: Pontificia Universidad Catolica de Chile
Julian Varas
jevaras@uc.cl

Pontificia Universidad Catolica de Chile Escuela de Medicina https://orcid.org/0000-0002-5828-9623

Research Article

Keywords: Micro-credentials, Surgical education, Procedural training, Simulation, Asynchronous feedback, Competency-based medical education

Posted Date: July 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10036292/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Purpose

Procedural training rarely preserves auditable evidence of observed performance. We report a first-cohort implementation of micro-credentialing as a validity-oriented pathway, examining uptake when decoupled from grades, retention as a credentialable construct, and assessor scalability.

Methods

We conducted a prospective program evaluation of a four-level pathway—simulation competency, six-month retention, peer-assessor certification, and supervised clinical performance—for thoracentesis and paracentesis. Learners completed remote preparation and uploaded simulator-based attempts for asynchronous, criterion-based assessment using a 5-point Likert-type OSATS rating scale and checklists. Credentials were summarized across an initial lock and cumulative extract. Uptake was interpreted in mature 2023–2024 cohorts; 2025 Level-1 credentials were reported descriptively because learners had insufficient follow-up for Level-2 retention and downstream progression. This was a feasibility evaluation, not an effectiveness trial.

Results

The latest extract recorded 553 Level-1, 203 Level-2, 25 Level-3, and two Level-4 credentials. Overall uptake was 51.4% (199/387), 48.2% of accepted credentials were publicly shared, and mature-cohort uptake was 53.3%. In the first retention cohort, 105/117 paracentesis learners and 98/121 thoracentesis learners met Level-2 criteria. Among voluntary returners, mean OSATS changed from 21.68 to 22.12 for thoracentesis ($p = 0.226$) and from 21.39 to 22.31 for paracentesis ($p = 0.0166$).

Conclusions

Micro-credentialing functioned as an evidence-linked assessment pathway rather than badge distribution. Moderate uptake highlights educational alignment. Retention credentialing and peer-assessor certification were feasible, but retention estimates require protocolized follow-up, and Level-3 impact on workload, turnaround, reliability, and feedback quality requires further evaluation.

Introduction

Competency-based medical education has shifted attention from time-based exposure toward demonstrated ability. Miller's framework distinguishes between knowing, knowing how, showing how, and doing, emphasizing that knowledge-based assessment is not equivalent to observed performance [1], and progression decisions should accordingly rest on observable outcomes rather than time alone [2]. In procedural training, however, the mechanisms used to document competence often remain indirect: course grades, attendance, written examinations, and checklist completion may confirm participation or knowledge but do not always preserve observable evidence that a learner can perform a

procedure to a defined standard. This matters most for psychomotor skills, where a learner may know a procedure's indications and steps yet lack evidence of safe execution. In our setting, competency frameworks for medical graduates include core procedural skills expected of general physicians; however, the ways in which these skills are taught, observed, and documented remain variable across training programs, clinical environments, faculty, and rotations [3]. Similar limitations have been reported internationally, where written examinations, grade-based systems, or single-point assessments may not adequately capture procedural readiness or the durability of performance over time [4–7]. Accordingly, course completion should not be assumed to represent evidence-linked procedural competence.

Acquiring technical skills requires deliberate practice, repeated performance with specific goals, feedback, correction, and further practice. Simulation-based education with deliberate practice has been associated with better performance than traditional clinical education alone, with reviews underscoring feedback, repetition, curriculum integration, and clearly defined outcomes [8–11]. The operational constraint is that high-quality feedback is resource-intensive: if every attempt requires real-time expert supervision, training is difficult to scale in large cohorts. Remote asynchronous feedback and video-based assessment may relieve this bottleneck, as learners record and upload performances and review structured feedback later. Prior work suggests that video without feedback is insufficient, whereas remote asynchronous feedback can improve procedural performance and support scalable training [12–15], separating the moment of practice from the moment of assessment while preserving an auditable record.

Peer or near-peer teaching and Train-the-Trainers models offer a further scalability strategy. A systematic review and meta-analysis found that peer teaching in health professions education can improve outcomes and may be comparable to conventional teaching when appropriately structured, and near-peer and peer-assessor models have expanded surgical training capacity, though their credibility depends on training, calibration, and supervision [16–18]. Peer assessment is thus best understood not as an uncontrolled substitute for expert assessment, but as a monitored mechanism to expand assessor capacity under expert-defined standards. In this context, a Train-the-Trainers course may also serve as a practical mechanism for developing and maintaining a pool of calibrated assessors with shared assessment criteria, thereby increasing the availability of trained educators to support scalable and defensible implementation [17, 18].

Digital micro-credentials have emerged as a way to make specific competencies visible and portable. In health professions education they have been described as focused, verifiable records of learning that may include criteria, evidence, issuer information, and metadata; and international frameworks emphasize clearly defined learning outcomes, transparent assessment, quality assurance, portability, and learner ownership [19–22]. Their value lies not in the badge but in the evidence pathway behind it: designed rigorously, a micro-credential indicates that a learner met a predefined performance standard rather than merely completed an activity.

In procedural training, micro-credentials may be especially useful when they connect simulation, video-based assessment, structured feedback, retention, peer teaching, and supervised clinical performance, but implementation raises unresolved questions. Can credentials be issued at scale without overwhelming faculty? Will learners claim them when they are not tied to grades? Can they document retention rather than only "test-day" performance? And how can assessment quality be preserved when video review is distributed across assessors? To examine these questions, we implemented a four-level pathway or skills tree for thoracentesis and paracentesis: simulation competence, six-month retention, peer-assessor certification, and supervised clinical performance. Rather than asking whether badges can be issued, we use this first implementation cycle to examine procedural micro-credentialing as a validity-oriented assessment pathway. Specifically, we report what the model revealed about uptake when credentialing was decoupled from grades, six-month retention, assessor scalability, early implementation barriers, and the validity safeguards needed to support this approach.

Methods

Study design

We conducted a prospective program evaluation of a micro-credentialing pathway implemented during undergraduate procedural training, focusing on feasibility, learner uptake, six-month retention, and early barriers to implementation. The initial procedural focus was thoracentesis and paracentesis, with subsequent expansion toward additional high-frequency procedural skills. Consistent with implementation-science frameworks that distinguish feasibility, adoption, acceptability, implementation quality, and effectiveness as related but distinct outcomes, the study was designed as a feasibility and program evaluation rather than a causal effectiveness trial [23, 24]. This program evaluation is reported in accordance with the Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0) [31].

Setting and participants

The intervention was implemented within a simulation-based procedural training program for medical students. Learners participated in curricular procedural training and were eligible to receive digital micro-credentials upon achieving pre-established performance thresholds. Mature Level-1 cohorts included learners from the 2023 and 2024 thoracentesis and paracentesis implementations. Level-1 credentials issued in 2025 were reported descriptively but were excluded from the mature-cohort longitudinal analysis. These credentials had not yet completed the minimum follow-up window required for Level-2 retention assessment and, consequently, could not reasonably contribute to downstream Level-3 peer-assessor certification or Level-4 supervised clinical documentation within the study period. Therefore, longitudinal uptake and progression analyses were restricted to mature 2023–2024 cohorts.

Educational intervention and workflow

The pathway combined remote preparation, deliberate practice, video-based assessment, asynchronous feedback, and criterion-based credential issuance, grounded in evidence that simulation-based education is most effective when it includes deliberate practice, feedback, and clearly defined performance criteria [8–11].

Learners first completed a remote module that included procedure-specific tutorials and common errors. They then performed and recorded a simulator-based attempt. Videos were uploaded to a secure educational platform and reviewed asynchronously, with assessors applying a 5-point Likert-type OSATS global rating scale [28] and procedure-specific checklists. Feedback was delivered asynchronously through written comments and, when available, audio or video-linked annotations.

Learners who met the predefined Level-1 competency threshold received a Level-1 micro-credential. For the credentials issued in this cohort, Level-1 required achieving at least 80% of the maximum score on both the OSATS-style global rating scale [28] and the procedure-specific checklist. For Level 2, learners were required to return at approximately six months and maintain at least 70% of their initial performance score, without critical errors, to receive the retention micro-credential. Recovery attempts could support credential acquisition but did not modify the learner's course grade. This separation of credentialing from grading was a deliberate design feature, informed by motivational frameworks in which autonomy, perceived competence, and social recognition shape engagement [25, 26].

Assessment instruments

Procedural performance was assessed with a 5-point Likert-type OSATS global rating scale [27, 28] and procedure-specific checklists. The OSATS global rating scale, originally described for surgical skills, anchors each domain on a five-point scale to capture the overall quality of technical performance [27]. In this program we used a modified five-item version covering flow, safety, instrument handling, efficiency, and procedural execution, with each item scored from 1 to 5, yielding a total score ranging from 5 to 25, consistent with prior simulation-transfer work in which this rubric structure was applied [28]. Procedure-specific checklists documented essential procedural steps, and critical-error criteria identified unsafe actions that could invalidate credential issuance even when the total score was otherwise acceptable. This combination of global rating scale, checklist, and critical-error criteria is consistent with procedure-specific validation work in simulation-based assessment [29]. For Level-2 retention, the same structure was used to compare performance over time.

Four-level micro-credentialing framework

The pathway included four progressive levels.

Level 1: Simulation competency. Learners demonstrated minimum procedural competence in a simulated environment through video-based performance assessment. The credential was issued when the learner reached the predefined operational competence threshold, defined as at least 80% of the

maximum score on both the OSATS-style global rating scale [28] and the procedure-specific checklist, with no critical errors.

Level 2: Six-month retention. Learners were reassessed at least six months after initial training. The retention credential required retaining at least 70% of the learner's own initial OSATS score and committing no critical errors.

Level 3: Peer-assessor certification. Learners who had demonstrated procedural competence could enter a Train-the-Trainers pathway to become peer assessors. This pathway included remote preparation, rubric training, review of standardized videos, and calibration against expert ratings. Peer-assessor certification required completion of calibration activities before independent review, in order to expand assessor capacity while maintaining rating quality.

Level 4: Supervised clinical performance. Learners could receive a clinical credential after completing a supervised procedure in a real clinical setting, with tutor attestation and applicable consent workflows. Level 4 required prior Level-2 retention certification but not Level-3 peer-assessor certification.

Assessor scalability and Train-the-Trainers model

A central implementation challenge was the volume of videos requiring review; relying exclusively on expert faculty would limit scalability, delay feedback, and threaten sustainability. The program therefore incorporated a peer-assessor pipeline through a Train-the-Trainers model, informed by evidence that peer teaching can be effective in health professions education when tutors are trained and supported, and by prior surgical education work showing that structured Train-the-Trainers programs can prepare learners and trainees to teach and provide technical feedback [16–18].

Peer assessors were trained to apply the OSATS global rating scale [27, 28], checklists, critical-error criteria, and feedback principles. Under expert-defined standards and using anchor videos, training included standardized performance videos, comparison with expert ratings, feedback on rating discrepancies, and calibration before independent review. Beyond expanding review capacity, this pathway created an additional competency route in which learners could progress from performing a skill to teaching and assessing it. Inter-rater reliability between peer assessors and expert raters was not formally quantified in this first cohort; assessor agreement was addressed only through this initial calibration.

Validity safeguards (as implemented)

Because this was a prospective implementation rather than a randomized trial, we treated validity as an explicit assessment concern rather than an assumption: a credential represented a criterion-based decision supported by observable performance, structured instruments, predefined thresholds, and an auditable evidence record, rather than attendance, participation, or course completion, consistent with competency-based assessment principles in which progression rests on multiple sources of evidence

rather than isolated observations [2]. Four safeguards described above operationalized this premise: separation of credentialing from grading; complementary instruments (OSATS global rating scale [27, 28], procedure-specific checklist, and critical-error criteria); calibration of peer assessors against expert ratings before independent review; and separate analysis of mature versus newly released credentials.

Outcomes

Primary outcomes were the number of credentials issued, accepted or claimed, pending, and publicly shared, with acceptance rate defined as accepted or claimed credentials divided by issued credentials. Secondary outcomes were Level-2 retention credentials, OSATS performance at initial and retention assessment, Level-4 clinical credentials, and observed implementation barriers.

Statistical analysis

Credential counts and acceptance rates were summarized with descriptive statistics. Results were analyzed across two data extracts (an initial analysis lock and a later cumulative extract); uptake was interpreted primarily in mature 2023–2024 cohorts, separately from recently released 2025 credentials. Retention was summarized as the proportion of learners meeting Level-2 criteria by procedure. For each procedure, initial and six-month OSATS scores among learners who returned were compared with a paired t-test, after assessing the distribution of paired differences (Shapiro-Wilk); the Wilcoxon signed-rank test was used when normality was not met. These comparisons were exploratory and not adjusted for multiplicity. Statistical significance was set at two-sided $p < 0.05$. Analyses were performed in R (RStudio).

Ethics

The program and its evaluation were approved by the institutional Ethics Committee [name blinded for review] (Study ID [blinded for review]).

Results

Credential issuance and uptake

In the initial analysis dataset, 387 credentials had been issued; 199 were accepted or claimed, corresponding to an overall acceptance rate of 51.4%. Among accepted credentials, 96 of 199 were publicly shared, corresponding to 48.2%.

In the later cumulative extract, 553 Level-1 credentials had been issued, alongside 203 Level-2 retention credentials, 25 Level-3 peer-assessor credentials, and the first two Level-4 supervised clinical credentials.

For longitudinal progression analyses, mature cohorts were defined as those with sufficient follow-up time to allow Level-2 retention assessment and potential downstream progression to Level-3 peer-assessor certification or Level-4 supervised clinical documentation within the study period. Accordingly, Level-1 credentials issued in 2025 were reported descriptively but excluded from mature-cohort progression analyses because learners had not yet reached the minimum follow-up window required for Level-2 retention assessment. Among mature 2023–2024 Level-1 cohorts, acceptance was 53.3%.

Retention reassessment outcomes

A total of 203 Level-2 retention credentials were recorded in the cumulative extract. Among learners who returned for reassessment, 105 of 117 paracentesis learners and 98 of 121 thoracentesis learners met the Level-2 threshold. Among voluntary returners, mean OSATS scores changed from 21.68 to 22.12 for thoracentesis at six months ($p = 0.226$) and from 21.39 to 22.31 for paracentesis ($p = 0.0166$). These findings indicate that most reassessed learners met retention criteria, although this signal reflects voluntary return rather than protocolized follow-up of the entire original cohort.

Peer-assessor certification

The Train-the-Trainers peer-assessor pathway was implemented during the study period as part of the Level-3 micro-credentialing model. A total of 25 Level-3 micro-credentials were issued after completion of peer-assessor training and procedure-specific eligibility criteria: 13 for thoracentesis and 12 for paracentesis. These credentials support the feasibility of using Level 3 not only as a learner-recognition milestone, but also as an assessor-capacity strategy to scale video-based review and asynchronous feedback.

Early Level-4 clinical credentials

Two Level 4 supervised clinical credentials were issued, representing the first clinical attestations under the pathway. Both followed prior Level 2 certification and were issued independently of Level 3 peer-assessor certification, which is not a prerequisite for supervised clinical documentation. Each supervised clinical procedure was performed with documented patient informed consent, and standard clinical care was not altered.

Implementation barriers

Four implementation barriers were identified during the first implementation cycle: limited incentive and visibility of the credential, timing of credential release, constraints on Level 4 clinical capture, and communication barriers. Credential claiming was lower when credentials were not explicitly linked to progression decisions, portfolios, or future opportunities. Claim prompts delivered after the learning activity were associated with lower learner attention, whereas higher claiming was observed when issuance occurred close to graded milestones and was reinforced by faculty. Level 4 clinical capture was constrained by patient availability, tutor presence, consent workflow, and clinical timing. Communication barriers included missed notifications, ignored messages, and perception of platform alerts as administrative rather than educational prompts.

Discussion

The value of this first cohort lies less in the number of credentials issued than in what a real implementation reveals about the assumptions embedded in procedural micro-credentialing. Through asynchronous video-based assessment, the program could issue, claim, and retain evidence-linked credentials across successive cohorts, generate the first supervised clinical attestations, and activate an initial peer-assessor certification pathway. The contribution, however, is conceptual as much as empirical: it reframes the micro-credential as the visible output of a validity argument and exposes which components add value, which are feasible, and which require further testing.

The central point is that micro-credentialing is not badge distribution. The badge is only the visible artifact; the intervention is the evidence pathway behind it: deliberate practice, video submission, structured assessment, feedback, criterion-based issuance, and longitudinal progression. This distinction matters because digital badges without rigorous evidence may add administrative burden without improving competence documentation, whereas evidence-linked credentials can make procedural competence visible, auditable, portable, and verifiable [19–22].

Taken together, the findings provide a differentiated answer to the three questions that motivated this first implementation. First, uptake when credentialing was decoupled from grades proved to be real but moderate: learners did claim and publicly share credentials, but adoption depended less on platform availability than on educational alignment, timing, and faculty endorsement. Second, retention could be operationalized as a credentialable construct, as a substantial proportion of returning learners met predefined Level-2 criteria at six months; however, because reassessment depended on voluntary return, this should be interpreted as evidence of feasibility of retention credentialing rather than as a generalizable estimate of retained competence across the full cohort. Third, assessor scalability was not merely a planned solution: the Train-the-Trainers pathway generated an initial pool of Level-3 peer-assessor credentials among learners who had progressed through the pathway. This provides early evidence that the model can build assessor capacity from within the learner community itself. Nevertheless, the downstream effect of this peer-assessor pool on faculty workload, feedback turnaround time, inter-rater reliability, and feedback quality remains to be evaluated in subsequent cohorts.

Feasibility and scalability

The principal operational concern in video-based procedural assessment is how to review large volumes of learner videos without overwhelming faculty. In this cohort, the program began to address this challenge through a Train-the-Trainers peer-assessor pathway embedded within the Level-3 micro-credentialing model. Learners who had previously demonstrated procedural competence and retention could progress to peer-assessor certification, creating an initial pool of trained assessors from within the learner community itself. Thus, Level 3 should be interpreted not only as an additional learner-recognition milestone, but also as a capacity-building mechanism designed to support scalable video

review and asynchronous feedback. Although the full operational impact of this peer-assessor pool on faculty workload and feedback turnaround time requires further evaluation in subsequent cohorts, the presence of certified Level-3 tutors suggests that scalability beyond faculty-only assessment is feasible rather than merely theoretical.

A Train-the-Trainers pathway is nonetheless the most plausible route to that scalability: learners who demonstrate competence could be trained to assess standardized videos, deliver structured feedback, and calibrate their ratings against expert benchmarks, expanding capacity while turning teaching and assessment into credentialed competencies, consistent with near-peer education literature [16–18]. Its credibility depends entirely on quality assurance. Before peer-generated scores can support credentialing, programs must sustain initial calibration with minimum agreement thresholds, double scoring of sampled videos, drift monitoring, and feedback-quality audits; without these, scaling capacity would compromise validity.

Uptake and educational alignment

Learner acceptance was moderate, approximately 51% in the initial analysis dataset and 53.3% in mature 2023–2024 cohorts. The latter estimate excluded 2025 Level-1 credentials because learners had not yet reached the minimum follow-up window required for retention and downstream progression, rather than because of a behavioral difference between cohorts. This level of uptake should not be read as technical failure; together with the implementation barriers observed, it indicates that uptake is an educational-alignment problem rather than a platform-usability problem. Motivation theory suggests that engagement is more likely when learners perceive autonomy, competence, and a meaningful connection to a valued goal [25, 26]. Accordingly, a credential is more likely to be claimed when it is visible, endorsed by faculty, synchronized with curricular milestones, and connected to portfolios, progression, or professional identity. Programs should therefore design micro-credentials as part of a competency ecosystem rather than as isolated optional badges, with faculty endorsement at the moment of release, integration with programmatic assessment, streamlined claiming, automated reminders, and institutional recognition of credential achievement.

Retention as a credentialable construct

Level 2 findings matter because they move the credential beyond “test-day” performance and establish retention as something a credential can certify, to our knowledge a novel step in procedural training. The present data, however, cannot establish retention itself. Reassessment was limited to learners who returned voluntarily; the observed OSATS changes were under one point on a 25-point scale and near the competence ceiling; they were reported without confidence intervals or effect sizes, and they were unadjusted for intercurrent clinical exposure; an increase among self-selected returners is at least as consistent with selection of more competent, more motivated learners as with true retention. The contribution is therefore the construct, retention as a credentialable level, together with the design required to estimate it cleanly: protocolized six-month reassessment of all evaluable learners, with

voluntary return analyzed separately as adoption, and adjustment for baseline OSATS, skill, site, cohort, and intercurrent clinical exposure.

Implications for assessment validity

The validity of micro-credentialing does not depend on the badge but on the assessment system behind it. A credential becomes trustworthy when it links a defined skill to a clear performance standard, observable evidence, trained assessors, and transparent metadata; in this sense it is the visible output of a validity argument rather than an independent intervention. Our implementation surfaced three validity threats.

The first is self-selection: learners who claim credentials or voluntarily return for reassessment may be more motivated than those who do not, so claiming and voluntary return are adoption outcomes, not direct competence measures, and future studies should include protocolized reassessment to estimate retention independently of voluntary uptake.

The second is that assessor variability in video-based assessment scales only if assessors are trained, calibrated, and monitored. Assessors in this cohort were calibrated against agreed standards before independent review, but inter-rater reliability was not formally quantified. The Train-the-Trainers pathway produced an initial set of Level-3 peer-assessor credentials, showing that a peer-assessor pipeline can be activated; however, its impact on reliability, workload distribution, turnaround time, and feedback quality remains to be measured. Quantifying inter-rater reliability and maintaining a monitored peer-assessor pipeline through anchor videos, agreement thresholds, double scoring, drift monitoring, and feedback-quality audits are therefore priorities for the next phase [17, 18].

The third is clinical-opportunity bias, the two Level 4 credentials show that clinical attestation is feasible, but real procedures are opportunistic and vary across rotations, sites, and patient availability; Level 4 should therefore be interpreted as structured documentation of supervised opportunities, not as evidence of independent clinical transfer, and its expansion will require a simplified but rigorous workflow for minimal viable evidence, tutor attestation, and consent. Future analyses should adjust for clinical exposure and distinguish opportunity to perform from competence when performance occurs.

Taken together, these threats imply that future micro-credentialing studies should move beyond simple counts of credentials issued or claimed. Stronger evidence will require linking credential data to performance scores, retention, assessor reliability, feedback quality, learner motivation, clinical exposure, cost, and equity across training sites.

External validity and generalizability

The model may transfer to other high-frequency procedural skills that can be practiced in simulation, recorded on video, assessed with structured instruments, and later observed clinically; the program has already expanded beyond thoracentesis and paracentesis. Transfer to other institutions will depend on

local curricular alignment, assessor capacity, technology infrastructure, consent workflows, and whether micro-credentials are recognized as meaningful evidence within the institution.

Implications for the next phase

Several safeguards that a credential-validity argument ultimately requires were only partially implemented in this first cohort and define the next phase. Assessment quality should be monitored continuously rather than only at calibration, through periodic double scoring, inter-rater agreement monitoring, drift detection, recalibration of low-agreement assessors, and feedback-quality audits. Retention should be estimated through protocolized six-month reassessment of all evaluable learners, with voluntary return analyzed separately as an adoption outcome. Causal inference should be strengthened through multicenter cluster randomization, centralized allocation, stratification by university, site, and curricular year, delayed access for controls, tracking of contamination and losses, intention-to-treat analysis, and exposure-based sensitivity analyses, with statistical models reflecting the hierarchical structure of the data, learners within clusters, clusters within sites, and repeated observations across skills and time, consistent with recommendations for cluster-randomized trials and the evaluation of complex interventions [24, 30].

Limitations

This study has several limitations. It was a first-cohort program evaluation rather than a randomized effectiveness trial, so causal claims about an independent effect of micro-credentialing on competence are not warranted. The total number of learners trained and eligible was not available, so acceptance was computed on issued credentials rather than the eligible population, and a full participant-flow denominator could not be constructed. Although assessors were calibrated against agreed standards, inter-rater reliability was not formally quantified; therefore, the reliability and feedback quality of the emerging Level-3 peer-assessor pool remain to be evaluated before peer-generated scores can be used as a fully validated scalability mechanism. Retention was assessed only among learners who returned voluntarily, without protocolized reassessment of all evaluable learners; voluntary return reflects adoption and cannot be separated from technical retention in these data, and the observed OSATS changes were small, near the ceiling, and reported without confidence intervals or effect sizes. Level 4 clinical credentialing remains exploratory (n = 2), as clinical procedures are opportunistic and depend on patient availability, tutor presence, consent workflow, and rotation timing. Finally, patient outcomes and independent clinical performance were not evaluated; the findings should be read as evidence of feasibility, uptake, and a preliminary retention signal, not as proof of clinical transfer.

Conclusions

Micro-credentialing of thoracentesis and paracentesis was feasible under asynchronous video assessment with predefined performance thresholds. Moderate uptake indicates that issuing credentials is insufficient; activation depends on educational alignment, including timing, incentives, faculty endorsement, and curricular value. Six-month reassessment supports retention as a credentialable

construct, but because reassessment was limited to voluntary returners, these findings should be interpreted as an adoption-conditioned signal rather than as retention generalizable to all learners. Estimating retention will require protocolized reassessment of all eligible learners. The Train-the-Trainers pathway produced an initial pool of Level-3 peer-assessor credentials, supporting its feasibility as a capacity-building mechanism for scalable video review. Its downstream impact on faculty workload, feedback turnaround time, inter-rater reliability, and feedback quality should be evaluated in subsequent cohorts.

References

1. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65(9 Suppl). https://doi.org/10.1097/00001888-199009000-00045. S63-S67.
2. Frank JR, Snell LS, ten Cate O, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638–45. https://doi.org/10.3109/0142159X.2010.501190.
3. Asociación de Facultades de Medicina de Chile. Competencias comunes para los egresados de las Escuelas de Medicina. ASOFAMECH; versión vigente.
4. Park WB, Kang SH, Myung SJ, Lee YS. Does Objective Structured Clinical Examinations Score Reflect the Clinical Reasoning Ability of Medical Students? Am J Med Sci. 2015;350(1):64–7. https://doi.org/10.1097/MAJ.0000000000000420.
5. Sarkar A, Heidelbaugh JJ, Hallbauer G, Appelbaum N. Imprecise Clinical Assessments and Inaccurate Grades: Family Medicine Clerkship Director Perspectives. Fam Med. 2024;56(8):471–5. https://doi.org/10.22454/FamMed.2024.819598.
6. Hawkins N, Younan HC, Fyfe M, Parekh R, McKeown A. Exploring why medical students still feel underprepared for clinical practice: a qualitative analysis of an authentic on-call simulation. BMC Med Educ. 2021;21(1):165. https://doi.org/10.1186/s12909-021-02605-y.
7. Moreno MI, Segovia J, Marican Palacios VP, Villalón F. Autopercepción de competencias en procedimientos clínicos en estudiantes de medicina durante el internado. Rev Educ Cienc Salud. 2024;24(1):32–8.
8. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med. 2008;15(11):988–94. https://doi.org/10.1111/j.1553-2712.2008.00227.x.
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? Acad Med. 2011;86(6):706–11. https://doi.org/10.1097/ACM.0b013e318217e119.
10. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;27(1):10–28. https://doi.org/10.1080/01421590500046924.
11. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA.

2011;306(9):978–88. https://doi.org/10.1001/jama.2011.1234.

12. Tejos R, Crovari F, Achurra P, Avila R, Inzunza M, Jarry C, et al. Video-Based Guided Simulation without Peer or Expert Feedback is Not Enough: A Randomized Controlled Trial. World J Surg. 2021;45(1):57–65. https://doi.org/10.1007/s00268-020-05766-x.
13. Belmar F, Gaete MI, Duran V, Chelebifski S, Jarry C, Ortiz C, et al. Taking advantage of asynchronous digital feedback: development of an at-home basic suture skills training program for undergraduate medical students. Glob Surg Educ. 2023;2:32. https://doi.org/10.1007/s44186-023-00112-w.
14. Varas J, Belmar F, Fuentes J, Vela J, Contreras C, Letelier LM, et al. Improving Medical Student Performance With Unsupervised Simulation and Remote Asynchronous Feedback. J Surg Educ. 2024;81(12):103302. https://doi.org/10.1016/j.jsurg.2024.103302.
15. Figueroa Ú, Jarry C, Inzunza M, Montero I, Garrido F, Villagrán I, et al. Innovation Meets Practice: A Scalable Simulation-based Methodology for Massive Paracentesis Training. Gastroenterology. 2025;168(5):865–e8692.
16. Zhang H, Liao AWX, Goh SH, Wu XV, Yoong SQ. Effectiveness of peer teaching in health professions education: A systematic review and meta-analysis. Nurse Educ Today. 2022;118:105499. https://doi.org/10.1016/j.nedt.2022.105499.
17. Duran Espinoza V, Jaras IM, Schilling MM, Coronel BV, Riveros FB, Jarry C, et al. Train the Trainers: a comparative analysis of medical students and residents as trainers for undergraduate med-school learners in surgical skills. Glob Surg Educ. 2023;2:110.
18. Montero Jaras I, Espinoza VD, Schilling MM, Riveros FB, Fernández ÚF, Coronel BV, et al. Effectiveness of a Train the Trainers course for digital feedback in healthcare simulation. Glob Surg Educ. 2024;3:64. https://doi.org/10.1007/s44186-024-00258-1.
19. Noyes JA, Welch PM, Johnson JW, Carbonneau KJ. A systematic review of digital badges in health care education. Med Educ. 2020;54(7):600–15. https://doi.org/10.1111/medu.14060.
20. Zdunek K, Dobrowolska B, Dziurka M, Galazzi A, Chiappinotto S, Palese A et al. Challenges and opportunities of micro-credentials as a new form of certification in health science education—a discussion paper. BMC Med Educ. 2024;24:1169. https://doi.org/10.1186/s12909-024-06174-8
21. Womack-Adams K, Morbitzer KA, Ondek C, Collins H, McLaughlin JE. A review of microcredentials in health professions continuing professional development. Front Med. 2025;12:1532811. https://doi.org/10.3389/fmed.2025.1532811.
22. Council of the European Union. Council recommendation of 16 June 2022 on a European approach to micro-credentials for lifelong learning and employability. Official J Eur Union. 2022;2022/C 243/02.
23. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research. Adm Policy Ment Health. 2011;38(2):65–76. https://doi.org/10.1007/s10488-010-0319-7.
24. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350:h1258.

https://doi.org/10.1136/bmj.h1258.

25. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol. 2020;61:101860. https://doi.org/10.1016/j.cedpsych.2020.101860.
26. Chen B, Vansteenkiste M, Beyers W, Boone L, Deci EL, Van der Kaap-Deeder J, et al. Basic psychological need satisfaction, need frustration, and need strength across four cultures. Motiv Emot. 2015;39:216–36. https://doi.org/10.1007/s11031-014-9450-1.
27. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273–8. https://doi.org/10.1046/j.1365-2168.1997.02502.x.
28. Boza C, León F, Buckel E, Riquelme A, Crovari F, Martínez J, et al. Simulation-trained junior residents perform better than general surgeons on advanced laparoscopic cases. Surg Endosc. 2017;31(1):135–41. https://doi.org/10.1007/s00464-016-4942-6.
29. Gerard JM, Kessler DO, Braun C, Mehta R, Scalzo AJ, Auerbach M. Validation of global rating scale and checklist instruments for the infant lumbar puncture procedure. Simul Healthc. 2013;8(3):148–54. https://doi.org/10.1097/SIH.0b013e3182802d34.
30. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group CONSORT. CONSORT 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661. https://doi.org/10.1136/bmj.e5661.
31. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf. 2016;25(12):986–92. https://doi.org/10.1136/bmjqs-2015-004411.