BMC Medical Education

RESEARCH

Open Access

Using the OSCE to assess medical students' communication and clinical reasoning during five years of restricted clinical practice

Armijo-Rivera Soledad[1*], Zamorano Saavedra Catalina[2], Vicencio-Clarke Scarlett[3], Behrens Pérez Claudia[4] and Pérez-Villalobos Cristhian[5]

Abstract

Background Objective and Structured Clinical Examination (OSCE) is a widely used evaluation method for health profession students. In Chile, physicians OSCEs have incorporated clinical reasoning and patient-centered communication assessment since 2015. The overall impact of the COVID-19 pandemic on undergraduate medical clinical practice remains unknown, in particular whether this context disproportionately affected lower-income regions, as was the case analysed in this study.

Aim This research compares the OSCE patient-centered communication and clinical reasoning results among five cohorts of intermediate-level medical students with restricted clinical practice in Chile.

Methods We designed an observational study analyzing five cohorts of fourth-year medical students with different clinical practice opportunities, with 3rd to 4th-year progression analysis in two cohorts. Adaptations to compensate lack of clinical practice hours include high-fidelity simulation and theoretical discussion of clinical cases, as well as formative OSCE at third-year level in two cohorts. Communication in OSCE was assessed using the Communication Assessment Tool (CAT) and Clinical Reasoning (CR) with a register form (that includes a global score, History Taking, Differential Diagnosis, and Therapeutic Plan subscores). Descriptive statistics and central tendency measurements were applied to analyze CAT and CR scores, along with Kruskal-Wallis tests and Wilcoxon's test for paired sample analysis.

Findings The 2018 cohort shows the lowest results in both variables although it was the cohort with more clinical practice opportunities. The higher CAT result was in the 2021 cohort, while the 2022 cohort exhibited a significantly higher CR score ($p < 0.05$). There is a linear tendency to grow over the years for both measures. The 2023 cohort shows significant improvement between third-year and fourth-year OSCEs in all items of CAT and a detriment in CR, while the 2022 cohort shows a significant increment in CR and four items of CAT ($p < 0.05$).

Conclusion The students with fewer hours of clinical practice showed similar CR and communication results to the prepandemic group, probably because of the adaptations implemented. Simulation was an effective alternative

*Correspondence:
Armijo-Rivera Soledad
s.armijo@udd.cl; soledad.armijo@gmail.com

Full list of author information is available at the end of the article

practice in a restricted clinical practice context. Component-by-component analysis and trends of change are a better approach to assessing progression than global scores.

Keywords Communication, Patient-centered communication, Clinical reasoning, OSCE, Medical students

Background

The Objective and Structured Clinical Examination (OSCE) described by Harden [1] constitutes an evaluative form for large groups of health profession students [2, 3]. There are practical guides for planning, implementation, and analysis of OSCE [4]. Recent expert consensus on performance assessment identifies that the validity of an OSCE relates to different elements, including the assessment of different skills, the use of rating scales more than a checklist, and the relationship with other variables such as clinical performance and the curriculum itself [5].

Patient-centered communication and clinical reasoning (CR) assessment using the Communication Assessment Tool (CAT) [6] and a CR register have been reported as part of young physicians OSCEs in Chile [7]. Patient-centered communication considered six interconnected dimensions: (i) exploring the disease and the disease experience; (ii) understanding the whole person; (iii) finding common ground between physician and patient; (iv) incorporating prevention and health promotion; (v) improving the physician-patient relationship; and (vi) "being realistic" about personal limitations or availability of time and resources [6]. In Chile this competence has been assessed using the Spanish version of the Communication Assessment Tool, because it is simple to understand, and consists of 14 descriptors on a 5-point scale (1 = poor; 5 = excellent), it is feasible to be applied for the standardized patients in a short period of time. Also, the instrument has shown to be reliable and valid and has validity evidence in the Chilean context and in OSCE [6].

CR assessment in OSCE was reported for the first time among recent graduates in Chile in 2015 [7]. The CR register form includes sections for history-taking skills, differential diagnosis, and treatment plans. The overall percentages of achievement in the exam were high, with the outcome in anamnesis having a high influence on the global outcome [7].

Clinical training in Chile is founded on traditional clinical practice, mainly in-hospital and ambulatory care clinical units, with some simulation activities. Otherwise, OSCEs are usually in person in our country. In response to the COVID-19 pandemic, and in the same manner that other countries [8], the Chilean Ministry of Health issued internal directives to universities, restricting student participation in clinical practice. These restrictions were in place from March 2020 to December 2022, limiting the possibility of implementing OSCE in person.

However, at intermediate levels with restricted clinical practice, some universities implemented remote simulation with a formative intention to address history-taking skills, with good results compared to non-pandemic training [9].

Although the pandemic's effect was global, some authors assumed that affected mainly lower-income regions due to limitations in infrastructure and access to new technologies, which contributed to the lack of patient teaching encounters [8].

During the planning of this study, our team did not find any evidence comparing intermediate level medicine OSCE outcomes in the context of restricted clinical practice due to the pandemic. Additionally, we did not find any evidence regarding patient-centered communication and CR achievement, nor the validity of the CAT at intermediate levels of a medical curriculum. This research aims to compare the OSCE patient-centered communication and CR results among five cohorts of medical students at the University Catholic of the North, during a period of restricted clinical practice in Chile. A secondary objective was to add the psychometric characteristics of the CAT instrument applied in intermediate-level OSCE.

Methods

Ethics statement

This study was performed by analyzing the anonymized database of intermediate OSCEs at a Chilean medical school. The need for consent to participate was waived by the Scientific Ethical Committee of the Faculty of Medicine of the Universidad Católica del Norte extensively revised this project, and approved the protocol (Act 07/2023). This Scientific Ethical Committee is accredited and supervised by the MINSAL (Ministerio de Salud) by their Regional Secretary of Health of the Region of Coquimbo, Chile, according to resolution Res. No. 2,204,491,026/2022.

Study design

This observational study analyzes five cohorts of fourth-year medical students at a private university in the North of Chile. The study includes an additional progression analysis based on two cohorts' formative assessments in the third year.

Academic context

The analysis considered the OSCEs of the medical program, which has a seven-year curriculum. The university formally introduced the integrated OSCE for CR and communication in 2016 as a pilot for graduates, with full

implementation of the same format intermediate level OSCE beginning in 2017. However, we omitted the 2017 cohort due to potential bias stemming from the implementation team's inexperience and students' unfamiliarity with the assessment format. The 2018 cohort was chosen as the baseline because it represents the first year with a fully established OSCE process, providing a more reliable reference point for comparison. Fourth-year students who meet all theoretical and practical requirements in the fourth year of their career, participate in intermediate-level OSCEs. However, the 2019 cohort did not undergo OSCE due to a contingency in the country that caused the suspension of academic activities. In 2020, due to the suspension of clinical practice during the pandemic, the OSCE was conducted online. Following best practices for performance assessment, a formative OSCE has been included at the third-year level since 2021.

Due to the persistence of restricted clinical practice since 2021, the leading of the simulation-based training was guided by an instructor recognized as certified educator advanced (CHSE-A), with extensive experience in competency-based assessment, and a strong background in simulation-based education. This instructor leads a process of enhancing instructional design, high-quality feedback and debriefing, and rigorous analysis of OSCE results. After that analysis, the curriculum was reorganized, adding hours of practice in theoretical sessions to analyze clinical cases and practice CR and high-fidelity simulation hours, including a three-station formative OSCE in the middle of the academic period to complement the clinical experience available for the intermediate level (Table 1). All these new simulation activities were focused on the performance observed on the last OSCE results, with an emphasis on improving the teaching-learning process on the professional-patient relationship, the interpretation of basic laboratory information (haemogram, urine, and biochemistry) applied to clinical cases of Haematologic, Endocrinologic, Rheumatologic and Nephro-urologic conditions of adult patients. The course syllabus explains how to run and participate in simulations. During the briefing, the psychological safety of participants was promoted and encouraged through icebreakers (the students shared one song heard this morning) so that they could get to know each other and identify personal moods each day to consider these elements when guiding the debriefing. After the formative OSCEs in 2022 and 2023, the instructor provided structured group feedback sessions, with individual feedback sessions for students with the lowest formative OSCE performance. Due to the same restrictions the 2021 academic year finished during April 2022, with the subsequent implementation of two OSCEs at the beginning of 2022 academic period.

Table 1 Curricular characteristics of the fourth-year clinical courses between 2018–2023

Year	Formative OSCE during the third year of career	Clinical practice hours during the fourth year first semester	OSCE modality
2018	No	190 (40 CR + 30 Sim + 80 Hosp + 40 Amb)	On-Site
2020	No	50 (20 CR + 30 Sim)	On-Line
2021	No	100 (60 CR + 30 Sim + 10 Amb)	On-Site
2022	Yes	154 (90 CR + 50 Sim + 14 Amb)	On-Site
2023	Yes	165 (90 CR + 40 Sim + 15 Hosp + 20 Amb)	On-Site

[a] CR = Clinical reasoning; Sim = High Fidelity Simulation; Hosp = Hospital; Amb = Ambulatory

Participants

We analyzed an anonymized database that included a sample of medical students from five cohorts, with five fourth-year OSCE registers and two third-year OSCE registers. Student cohorts were similar each year in terms of admissions criteria and demographics (the chi-square test conducted on the proportions of women and men across cohorts yielded a test statistic of $\chi 2 = 7.24$ with a p-value of 0.124). We only considered data from students who participated for the first time in each OSCE.

Variables

The variables assessed in OSCEs are patient-centered communication and CR, which include history-taking skills (HT), differential diagnosis (DD), and treatment plan (TP).

Assessment implementation (Data sources/measurement)

a. The OSCE consists of six stations, following a blueprint established in 2017. One station corresponds to a hospital emergency setting to assess CR, teamwork and leadership. The other five assess CR and patient-centered communication in ambulatory settings. We used the stations that assessed those two competencies, excluding the emergency stations. Four ambulatory cases were based on differential diagnoses supported by basic lab tests. Two cases correspond to a first consult that requires laboratory studies and patient education in frequent endocrine pathologies (Hypo or Hyperthyroidism) or articular pain (mono or polyarticular, with or without systemic manifestations). One case was the interpretation of a Haemogram previously ordered for a patient with suspected anemia. One case corresponds to the interpretation of basic control exams in a Diabetic and Hypertense patient (to define the therapy and further controls or derivation). The last case

corresponds to a first consultation for a Urologic emergency (haematuria, renal colic, or acute urine blockage) where it was necessary to establish a syndromic diagnosis and indicate a procedure (not perform one). The cases between the five OSCE versions were similar in complexity, according to the experts that conformed the OSCE Committee.

b. ***Patient-centered communication***: The standardized patient applied the validated Spanish version of the instrument CAT [6]. This is a 14-item 5-point Likert scale, with excellent performance equal to five. We considered the main value and the number and proportion of excellent.
c. ***Clinical reasoning***: The students complete a CR register form that includes sections for HT, DD and TP, designed similarly as described for a national OSCE in Chile [7]. These registers were qualified by a trained clinical expert, using a rating score with predefined criteria for unsatisfactory, satisfactory, and excellent responses in HT, DD, and TP separately, assigning a numeric variable with values between one and three respectively. Also, a global score was calculated with a compensatory approach. An OSCE Committee, which includes an expert in Medical Education and two Internal Medicine specialists, established the qualification criteria.

Bias

To minimize bias in this study, we applied standardized assessment methods across all participants, using validated tools such as the Communication Assessment Tool (CAT) and a structured register form for CR. Statistical analyses, including Kruskal-Wallis and Wilcoxon's tests, were chosen to account for potential non-normality in the data, ensuring robust comparisons. Additionally, all students were assessed under the same conditions to reduce variability. By including only each student's first attempt, we ensure that those who did not pass are represented in the analysis, rather than selectively analyzing only successful attempts. This approach provides a more accurate reflection of overall student performance, as it prevents an artificial inflation of scores that could result from repeated attempts. However, certain biases could not be fully eliminated, particularly differences in clinical exposure due to pandemic-related restrictions. Rather than being overlooked, these factors are central to our analysis, as they provide valuable insight into how variations in learning conditions during the pandemic may have influenced student performance. These limitations are acknowledged and discussed in detail.

Study size

Based on our primary aim, and considering a medium effect size (f = 0.25), a 95% confidence interval (1- α), and an 80% power (1- β), we estimate a minimum sample size of 120 students, which applies to any secondary objective.

Statistical methods

Descriptive statistics (frequency and percentage) and central tendency measurements (Mean, Median, Standard Deviation, and Interquartile Range) were used. To analyze the normality and symmetry of the sample results Skewness, Kurtosis, and Shapiro-Wilk tests will be reported. We used the Kruskal Wallis test to compare cohorts, McDonald's Omega for the reliability of scale scores, and Wilcoxon's test for paired sample analysis. Additionally, we performed confirmatory factor analyses to assess CAT validity in this sample.

JASP 0.17.2.1 software (Amsterdam, Netherlands) was utilized, with a statistical significance threshold set at P-value < 0.05 and a 95% confidence interval (CI).

Results

Participants

The anonymized database included registers from 307 potentially eligible fourth-year medical students from five cohorts. We excluded the results of eight students from the 2022 cohort in their second attempt in the same OSCE. The final sample consisted of 299 participants, 144 women (48.16%) and 155 men (51.84%), with complete registers for five stations in each fourth-year exam (1495 OSCE stations). We analyzed an additional cluster of 605 stations applied in formative OSCE during the third year of the career in 2022 and 2023.

Main results

The average CAT score across cohorts was 4.12 ± 0.40 which represents 82.40% of the theoretical maximum. Meanwhile, the mean global score in CR was 2.2 ± 0.36 representing 73.33% of the theoretical maximum. The Shapiro-Wilk test shows that the sample differed from the normal distribution, determining that the subsequent analysis was performed using non-parametric tests.

The lower scores in CAT (mean = 3.96, 79.20%) and CR (mean = 1.99, 66.30%) were in the 2018 cohort, while the higher results in CAT in the 2021 cohort (mean = 4.32, 86.40%), and in CR in 2022 cohort (mean = 2.65, 88.3%). The graphic shows a tendency to grow over the years in both variables (Fig. 1). There are statistically significant differences between the groups in CAT ($p < 0.05$) and CR ($p < 0.05$).

Dunn's post hoc comparisons of CAT mean show a statistically significant difference between the cohorts 2018–2021, 2020–2021, 2020–2023, 2021–2022, and 2021–2023 ($p < 0.05$), where the 2021 cohort had higher scores than the others, and 2023 cohort shows a better performance than 2020 cohort. The items with lower mean value of excellents were *"Involved me in decisions*

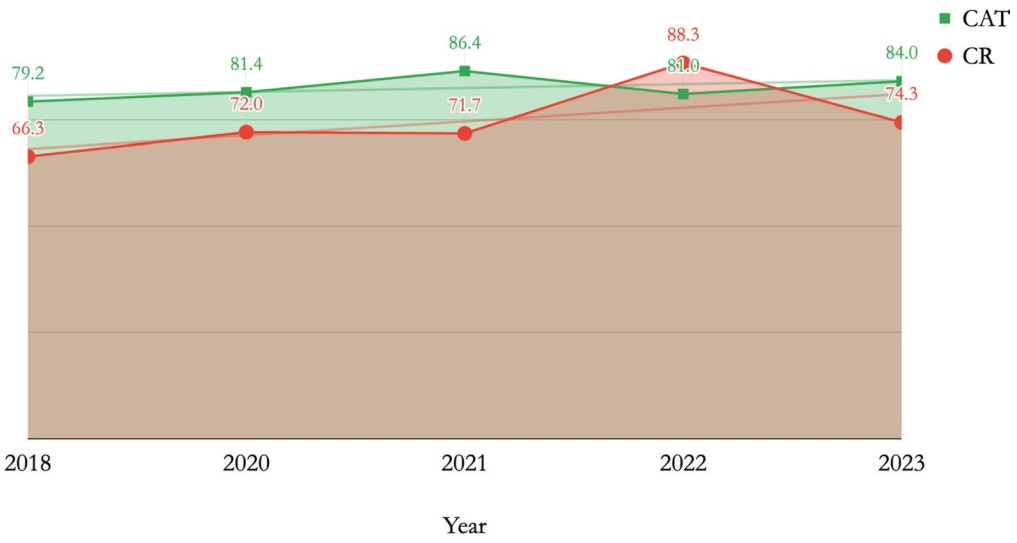

Fig. 1 Five cohorts mean scores in communication (CAT) and clinical reasoning (CR)

as much as I wanted" with 35.77%, *"Encouraged me to ask questions"* with 37.97% and *"Checked to be sure I understood everything"* with 39.25%. The items with higher mean value of excellents were *"Treated me with respect"* with 49.34, *"Let me talk without interruptions"* with 48.07% and *"Paid attention to me (looked at me, listen carefully)"* with 45.71%. The items with lower and higher numbers of excellent differ between the five cohorts (Table 2).

In the case of CR, the 2022 cohort shows a better and more significant performance than the other four ($p < 0.05$). On the other hand, the 2018 cohort exhibits statistically significant differences ($p < 0.05$) from the other four cohorts, with the lowest performance (Fig. 2).

Comparing the progression between the third and fourth year, the 2023 cohort shows significant improvement in all items of CAT, and the 2022 cohort shows statistically significant differences in items two, eight, 13, and 14. The alternative hypothesis specifies that the third-year measure for all tests is less than the fourth-year measure (Table 3).

Comparing the progression in CR score, HT and TP between the third and fourth year, the 2023 cohort shows a significant detriment while 2022 shows a significant improvement ($P < 0.05$). The tendency of the change in DD is coherent (Fig. 3).

In this sample, the CAT internal consistency expressed by McDonald's omega coefficient was excellent ($\omega = 0.97$, CI: 0.966–0.970), with a range between 0.964 and 0.970 for each item. As evidence for CAT validity in this curriculum level, confirmatory factor analyses showed an acceptable fit for a one-factor model in each of the five measurements with CFI > 0.995, TLI < 0.994, and RMSEA < 0.081.

Discussion

In this study, we found that the OSCE results in CR were lower than CAT in four cohorts, with a linear tendency to get better results from the first to the last cohort analyzed. The results of the TeleOSCE in the pandemic cohort (2020) show similar results compared with the previous generation (2018). There is a significant difference between the third and fourth-year OSCEs in the two cohorts analyzed, although the trends in these results are the opposite and the components of the CR do not change equivalently between them. Concerning the validity of the CAT, the instrument's internal consistency was high in this sample.

This study shows that the performance of medical students in patient-centered communication is higher than CR in fourth-year OSCE. These results could be attributed to the early familiarization with communication skills, which correspond to transversal competencies that have been developing since before university entrance. In patient-centered communication, there are elements common to empathic communications deployed in other contexts that are re-signified for medical practice. On the other hand, the complexity of CR as a specific process that relies on medical knowledge, may be difficult to achieve in early stages in the curriculum.

However, these patient-centered communication results are better than those of a similar OSCE applied in 2015 at the end of the medical career in Chile [6, 7]. This can result from the students' and teaching strategies' adaptation to the assessment system, following the idea that "assessment guides learning" [5]. Some of the items that show improvement reflect the model of shared decisions between the clinician and the patient, which was underscored previously in young physicians in Chile [6, 7].

Table 2 Frequency and percentage of "excellent" responses per item in the CAT scale per year

Item		Mean		2018		2020		2021		2022		2023	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
1	Greeted me in a way that make me feel comfortable	835	40.91	109	43.60	124	37.58	147	47.42	120	38.10	136	48.23
2	Treated me with respect	1007	49.34	123	49.20	159	48.18	174	56.13	163	51.75	167	59.22
3	Showed interest in my ideas about my health	851	41.70	115	46.00	123	37.39	162	52.26	117	37.14	140	49.65
4	Understood my main health concerns	857	42.00	123	49.20	126	38.30	163	52.58	111	35.24	138	48.94
5	Paid attention to me (looked at me, listen carefully)	933	45.71	112	44.80	140	42.55	167	53.87	151	47.94	160	56.74
6	Let me talk without interruptions	981	48.07	117	46.80	168	51.06	168	54.19	152	48.25	161	57.09
7	Gave as much information as I wanted	835	40.91	101	40.40	120	36.47	169	54.52	118	37.46	133	47.16
8	Talked in terms I could understand	861	42.19	107	42.80	142	43.16	151	48.71	144	45.71	151	53.55
9	Checked to be sure I understood everything	801	39.25	94	37.60	116	35.26	174	56.13	116	36.83	128	45.39
10	Encouraged me to ask questions	775	37.97	98	39.20	109	33.13	162	52.26	119	37.78	117	41.49
11	Involved me in decisions as much as I wanted	730	35.77	85	34.00	117	35.56	160	51.61	112	35.56	106	37.59
12	Discussed next steps, including any follow-up plans	844	41.35	87	34.80	136	41.34	169	54.52	126	40.00	134	47.52
13	Showed care and concern	880	43.12	105	42.00	132	40.12	172	55.48	133	42.22	146	51.77
14	Spent the right amount of time with me	899	44.05	103	41.20	141	42.86	176	56.77	133	42.22	156	55.32

n = 2041; n: Frequency; %: Percentage; CAT: Communication Assessment Tool

Similarly to other studies, we confirm that a remote OSCE is a valid instrument for assessing communication and CR skills [9–11].

In our study the student's performance at intermediate-level OSCEs was not affected negatively during and after the pandemics, as was described in internal medicine clerkship [11], and even with the reduction in hospital practice hours. On the other hand, these results contrast with other studies that show lower outcomes in pandemic cohorts [12] and with the statement that the pandemic affects more low-income countries [8].

In this study, the cohort with more high-fidelity simulation hours showed the best CR performance in OSCE in five years. Our results support the idea that simulation-based practice provides better opportunities for students to achieve the expected results at this curricular level [9]. The coherence between the teaching strategy and the assessment system may explain the result in this cohort, but another explanation could be the adoption of strategic performance during OSCEs [5]. Otherwise, instructor expertise and training plays a crucial role in the effectiveness of simulation training on students performance. Experienced simulation instructors can use ice breakers fluidly to create a better learning educational atmosphere during the briefing, provide more structured feedback, facilitate deeper reflective learning, design and implement more rigorous and high-impact simulation scenarios and OSCEs [13]. Additionally, improvements in faculty development programs and increased exposure to best practices in medical education may have strengthened the instructional quality over time.

This study also revealed that the changes between the third and fourth years compared in two cohorts were different in the analyzed groups and that the elements inside the tool used to assess patient-centered communication and CR changed in a non-equivalent manner between cohorts over the years.

In contrast with the prior approach of applying a compensatory approach [7], our results support the idea that an analysis component by component is a better approach because global scores do not capture all the exam details.

This is aligned with the recommendations to assess the OSCE as a whole system, but analyzing each station independently. One implication of these findings is that the decisions to guide the curriculum post-OSCE should be based not only on global scores and compensatory grading systems at a single OSCE but also on the analysis of the components of each competence, and in the analysis of trends of change, and the relation of this changes with the curricular opportunities. To decide just based on a compensatory approach questions the validity of the instrument in order to make high-impact decisions at this level of the curriculum.

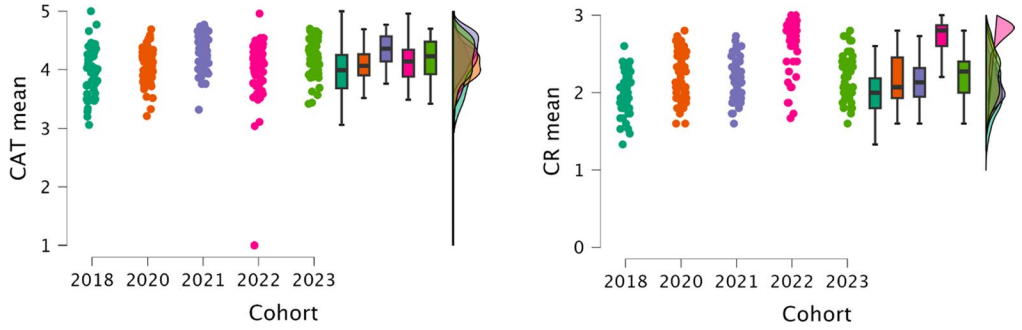

Fig. 2 Comparison of five cohorts for CAT and CR

Table 3 CAT progression in cohorts 2022 and 2023

Item	Measure 3° (2021) and 4° year (2022)					Measure 3° (2022) and 4° year (2023)				
	Mean 3°	Mean 4°	W	z	*p*	Mean 3°	Mean 4°	W	z	*p*
1	4.08	4.08	467.50	-0.56	0.29	4.07	4.35	127.50	-3.38	< 0.001*
2	4.24	4.33	342.00	-1.79	0.04*	4.05	4.46	53.00	-4.60	< 0.001*
3	4.08	4.01	581.00	-0.07	0.47	3.98	4.16	188.00	-3.14	< 0.001*
4	4.11	4.02	588.50	0.26	0.61	3.97	4.15	174.00	-2.68	0.004*
5	4.16	4.25	425.50	-1.47	0.07	4.02	4.35	90.00	-4.07	< 0.001*
6	4.27	4.26	414.00	-0.95	0.17	4.07	4.44	49.00	-4.56	< 0.001*
7	4.06	4.00	516.50	-0.50	0.31	3.80	4.10	124.50	-3.71	< 0.001*
8	4.03	4.17	269.50	-2.80	0.003*	3.86	4.29	77.00	-4.68	< 0.001*
9	3.98	3.97	423.50	-0.83	0.20	3.63	4.11	125.00	-3.96	< 0.001*
10	3.82	3.88	525.50	-1.08	0.14	3.49	4.00	145.50	-3.26	< 0.001*
11	3.60	3.39	767.00	1.25	0.90	3.22	3.53	213.00	-2.82	0.002*
12	4.03	4.02	479.00	-0.67	0.25	3.72	4.17	118.50	-3.37	< 0.001*
13	4.03	4.15	368.50	-2.07	0.02*	3.91	4.24	133.00	-3.14	< 0.001*
14	4.11	4.16	331.50	-1.91	0.03*	3.86	4.36	76.00	-4.38	< 0.001*

Note. For all tests, the alternative hypothesis specifies that measure 3° year is less than measure 4° year

*= statistically significant difference

W = Paired Samples Wilcoxon signed-rank test used to assess differences between paired or related samples

z = Z-score in the Wilcoxon test; the standardized value of the W statistic allows its interpretation in normal distribution

p = p value < 0.05

From our perspective, making high-impact decisions at this level of the curriculum solely based on a compensatory approach is questionable [14]. On the contrary, the educative value of a system that provides details of both communication competence and CR should be encouraged [5].

Finally, concerning the validity evidence of CAT, the results that we found are similar in reliability [6, 15], confirming that the instrument could be used in this level of the curriculum, being cautious about the passing score, because although the mean results are better than described previously in Chile [6, 7], they are still far the results described in the original article that assess Canadian experienced physicians [15].

While this study provides information on the clinical outcomes of medical students assessed in intermediate OSCEs, some limitations must be considered. Considering the restriction to clinical practices, we can´t correlate the OSCE results with reliable clinical assessment. The nature of the CR theoretical practice or the implementation of high-fidelity simulation in the fourth year or previous curriculum was not described, and there is no systematic data on the implementation and impact of these practices. It is essential to acknowledge that the absence of information about the instructors' and clinical supervisors' backgrounds and experiences limits the depth of the comparative analysis. The OSCE administration using written registers of the CR process does not collect the interaction information during the station, potentially impacting the comprehensive interpretation of the results if videos are not systematically analyzed. Finally, the findings may not be universally applicable, as the study is limited to one single medicine school in Chile. The differences between curriculum, instructors, and clinical tutors' experience and credentials as well as OSCE organization experience could affect the generalizability of the results to other institutions.

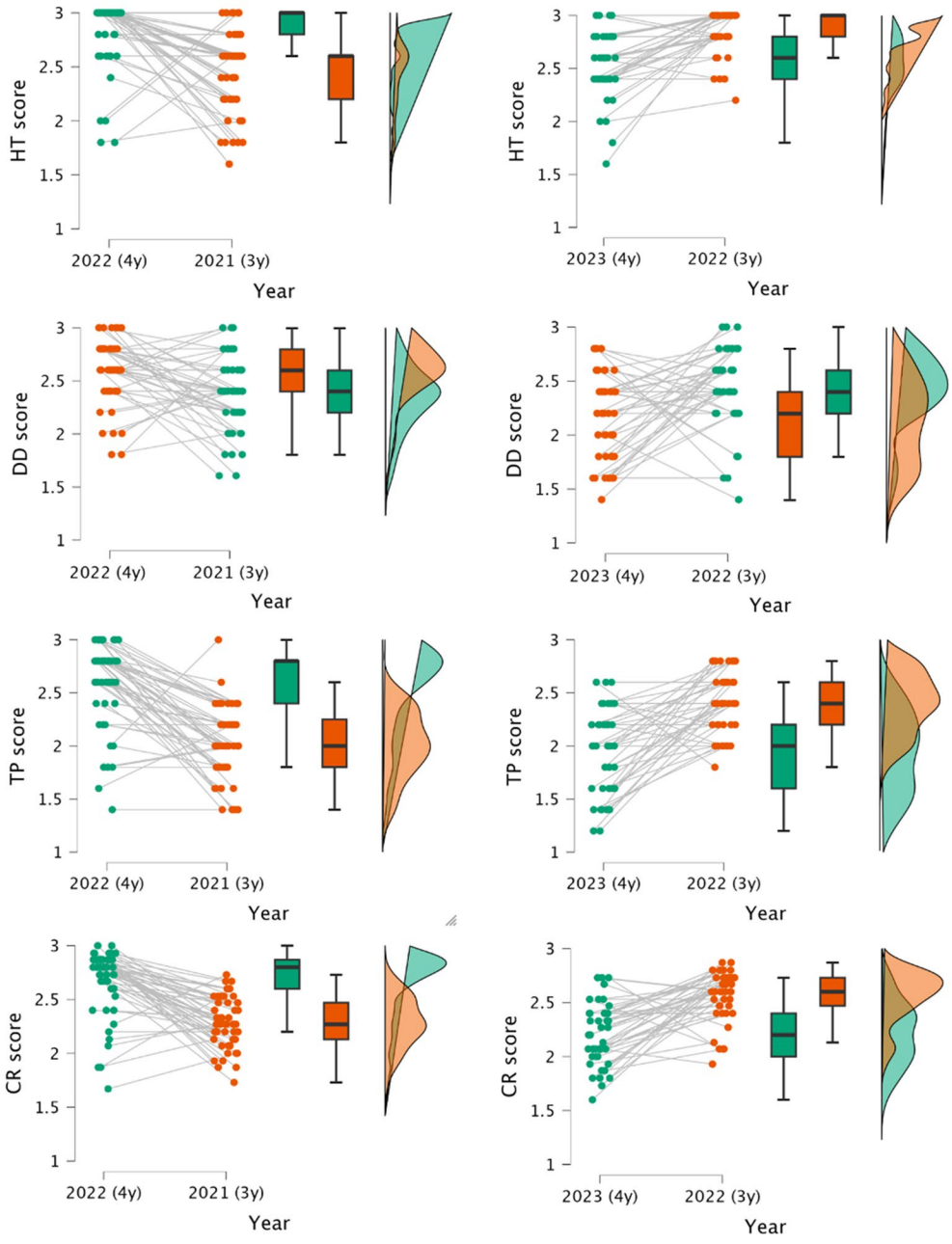

Fig. 3 Comparison of the progression of CR and its components in two cohorts (2022 and 2023)

One limitation of this study is that it was conducted at a single medical school in Chile, which may limit the generalizability of the findings to other institutions with different curricula, faculty experience, and OSCE implementation processes. Additionally, students who engaged more with simulation-based learning and adapted practices during the pandemic may have been more motivated or higher-performing, potentially leading to an overrepresentation of these students in the sample. Furthermore, the variability in clinical practice hours across cohorts due to pandemic-related restrictions may have influenced student experiences and performance beyond the impact of the curricular interventions analyzed. These factors should be considered when interpreting the results, and future research should explore their implications in broader educational contexts.

The limitations of this study reflect common challenges in medical education, where curricula rely on clinical educators who balance teaching with other professional responsibilities. Variability in clinical practice exposure, differences in instructor experience, and the lack of systematic data on simulation implementation can influence

student learning and assessment outcomes. These factors, while limiting the generalizability of our findings, are inherent to educational programs that depend on dynamic clinical environments. Addressing these challenges through standardized faculty development, clearer curricular structures, and systematic tracking of teaching methodologies could enhance the consistency and applicability of OSCE-based assessments in similar contexts.

However, one of the strengths of this study is that it compares cohorts and analyzes progression in intermediate levels of the curriculum, analyzing communication and reasoning in a complementary but independent manner, which is infrequent to be found in medical education literature. Another strength is the use of CAT, which is an instrument that has been validated in multiple languages and settings. For instance, a study conducted in Brazil provided validity evidence for using the CAT to assess residents' interpersonal communication skills under challenging simulation scenarios, underscoring its cross-cultural applicability including educative contexts [16]. Our study reinforces the idea that CAT appears suitable, valid, and reliable when used in undergraduate OSCE.

Conclusion

One of the strengths of this study is that it compares cohorts and analyzes progression in intermediate levels of the curriculum, analyzing patient-centered communication and CR elements in a complementary but independent manner, which is infrequent to be found in medical education literature. Another strength is the use of CAT, which is an instrument that has been validated in multiple languages and that appears suitable, valid, and reliable when used in undergraduate OSCE.

Simulation, especially if guided by skilled instructors, emerged as an effective alternative to compensate for the limitations imposed by restricted clinical hours in Chile, serving as an emergency solution. OSCE and TeleOSCE are assessment systems that can be used in the intermediate-level of the medical career. Combining the assessment of CR components and patient-centered communication configures a system that integrates multiple perspectives into the assessment process, capturing cognitive and non-technical skills in an integrated and independent manner. When considering the OSCE as an assessment system and introducing instruments that measure different variables, it is necessary to analyze each exam element separately and not limit it to the qualification decision or global scores.

Finally, the CAT confirms its attributes as a valid instrument to assess patient-centered communication in undergraduate students and applied by standardized patients, providing more points of view to the assessment system.

Lessons learned

- Simulated practice and CR exercises guided by skilled instructors provide opportunities to develop clinical skills at an intermediate level in restricted clinical practice contexts.
- Analyzing the internal components of OSCEs separately provides more information about progression in clinical practice than global scores.
- The communication assessment tool (CAT) is an instrument reliable and valid for promoting and assessing patient-centered communication.

Abbreviations

CAT	Communication Assessment Tool
CR	Clinical reasoning
DD	Differential diagnosis
HT	History-taking skills
OSCE	Objective and Structured Clinical Examination
TP	Treatment plan

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-025-07210-x.

Supplementary Material 1

Acknowledgements

We thank our simulated patients for taking the time and energy to assess our students and contribute to this research through their perspectives. We also thank NBME and Prof. Gregory Makoul for supporting the original research that initiated the road we have been walking since 2013.

Author contributions

The authors contribution according to CREDIT was: Conceptualization: SA Data curation: SA, CZ, SV Methodology/formal analysis/validation: SA, SV, CP Project administration: SA, CB Writing– review & editing: SA, SV, CP, CZ, CB.

Funding

No funding was available for the development of this work.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

This study was performed by analyzing the anonymized database of intermediate OSCEs at a Chilean medical school. The need for consent to participate was waived by the Scientific Ethical Committee of the Faculty of Medicine of the Universidad Católica del Norte extensively revised this project, and approved the protocol (Act 07/2023). This Scientific Ethical Committee is accredited and supervised by the MINSAL (Ministerio de Salud) by their Regional Secretary of Health of the Region of Coquimbo, Chile, according to resolution Res. No. 2204491026/2022.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details
[1]Facultad de Medicina, Universidad del Desarrollo, Avenida Plaza 680, Las Condes, Santiago 7510602, Chile
[2]Escuela de Medicina, Universidad Católica del Norte, Coquimbo, Chile
[3]Unidad de Simulación e Innovación en Salud, Universidad San Sebastián, Santiago, Chile
[4]Dirección General de Pregrado, Universidad Católica del Norte, Coquimbo, Chile
[5]Departamento de Educación Médica, Universidad de Concepción, Concepción, Chile

Received: 1 May 2024 / Accepted: 22 April 2025
Published online: 25 April 2025

References

1. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1(5955):447–51. https://doi.org/10.1136/bmj.1.5955.447. PMID: 1115966; PMCID: PMC1672423.
2. Khan KZ, Ramachandran S, Gaunt K, Pushkar P. The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: an historical and theoretical perspective. Med Teach. 2013;35(9):e1437-46. https://doi.org/10.3109/0142159X.2013.818634. PMID: 23968323.
3. Harden RM. Revisiting 'Assessment of clinical competence using an objective structured clinical examination (OSCE)'. Med Educ. 2016;50(4):376-9. https://doi.org/10.1111/medu.12801. PMID: 26995470.
4. Khan KZ, Gaunt K, Ramachandran S, Pushkar P, The Objective Structured Clinical Examination (OSCE).: AMEE Guide No. 81. Part II: organisation & administration. Med Teach. 2013;35(9):e1447-63. https://doi.org/10.3109/0142159X.2013.818635. PMID: 23968324.
5. 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(1):58–67. doi: 10.1080/0142159X.2020.1830052. Epub 2020 Oct 14. PMID: 33054524.
6. Armijo-Rivera S, Behrens C, Giaconi ME, Hurtado A, Fernandez MR, Parra P, Morales MV, Makoul G. Validación de la versión en español de un instrumento de evaluación de la comunicación centrada en el paciente en OSCE. [Validation of the Spanish version of a patient-centred communication assessment instrument in OSCEs] Educación Médica, 2021 July;22(4), 193–8. Spanish. https://doi.org/10.1016/j.edumed.2020.12.007.
7. Behrens C, Morales V, Parra P, Hurtado A, Fernández R, Giaconi E, Santelices L, Armijo S, Furman G. Diseño e implementación de OSCE para evaluar competencias de egreso en estudiantes de medicina en un consorcio de universidades chilenas [A standardized objective structured clinical examination to assess clinical competencies in medical students]. Rev Med Chil. 2018;146(10):1197–1204. Spanish. https://doi.org/10.4067/S0034-98872018001001197. PMID: 30724985.
8. Connolly N, Abdalla ME. Impact of COVID-19 on medical education in different income countries: a scoping review of the literature. Med Educ Online. 2022;27(1):2040192. PMID: 35317712; PMCID: PMC8956309.
9. Jadue M, Munitiz P, Torres L-M, Blacutt M, Rivera SA, Reyes E-P. Comparison of achievement in anamnesis skills in on-site and online formative objective structured clinical exam. Salud Ciencia Y Tecnología - Serie De Conferencias. 2023;2023(Oct 2):427. https://doi.org/10.56294/sctconf2023427.
10. Prasad L, Hockstein S, Safdieh JE, Harvey K, Christos PJ, Kang Y. An objective structured clinical exam on breaking bad news for clerkship students: In-Person versus remote standardized patient approach. MedEdPORTAL. 2023;19:11323. https://doi.org/10.15766/mep_2374-8265.11323. PMID: 37484524; PMCID: PMC10359437.
11. Yaşlı M, Koçak Algül S, Alpay-Kanıtez N. The impact of the COVID-19 pandemic on internal medicine clerkship by comparing exam results and feedback. Psychol Health Med. 2023 Jul-Dec;28(8):2375–80. Epub 2022 Sep 12. PMID: 36097435.
12. Yuan EJ, Huang SS, Hsu CA, Lirng JF, Li TH, Huang CC, Yang YY, Li CP, Chen CH. Negative effects on medical students' scores for clinical performance during the COVID-19 pandemic in Taiwan: a comparative study. J Educ Eval Health Prof. 2023;20:37. https://doi.org/10.3352/jeehp.2023.20.37. Epub 2023 Dec 26. PMID: 38148494; PMCID: PMC10810719.
13. Guerrero JG, Alqarni AS, Estadilla LT, Benjamin LS, Rani VI. Raters and examinees training for objective structured clinical examination: comparing the effectiveness of three instructional methodologies. BMC Nurs. 2024;23(1):500. https://doi.org/10.1186/s12912-024-02183-6. PMID: 39039471; PMCID: PMC11265468.
14. Malau-Aduli BS, Jones K, Saad S, Richmond C. Has the OSCE Met its final demise?? Rebalancing clinical assessment approaches in the Peri-Pandemic world. Front Med (Lausanne). 2022;9:825502. https://doi.org/10.3389/fmed.2022.825502. PMID: 35265639; PMCID: PMC8899088.
15. Makoul G, Krupat E, Chang CH. Measuring patient views of physician communication skills: development and testing of the communication assessment tool. Patient Educ Couns. 2007;67(3):333–42. https://doi.org/10.1016/j.pec.2007.05.005. Epub 2007 Jun 18. PMID: 17574367.
16. Rocha SR, Romão GS, Setúbal MSV, Lajos GJ, Luz AG, Collares CF, Amaral E. Cross-Cultural adaptation of the communication assessment tool for use in a simulated clinical setting. Teach Learn Med. 2020 Jun-Jul;32(3):308–18. Epub 2020 Feb 24. PMID: 32090632.

Publisher's note