

Translation, adaptation and validation of the Mini-Clinical Evaluation Exercise to the EU-Portuguese language

Rita SOUSA, Patrício COSTA, João CERQUEIRA, José M. PÊGO, André SANTA CRUZ, António OLIVEIRA E SILVA, Nuno SOUSA, Vítor H. PEREIRA

Introduction. The goal of this work is to validate tools to assess clinical competences of undergraduate medical students in the workplace. One of the most well-known scales is the Mini-Clinical Evaluation Exercise (Mini-CEX). This scale has been vastly studied, however, its validity is very variable amongst studies and it has never been validated to the EU-Portuguese language and context.

Subjects and methods. The translation process of the Mini-CEX was conducted by 2 bilingual individuals and overseen by four physicians specialized in medical education. We performed methods of both qualitative (translation, assessment of the translation, back translation) and quantitative nature (internal consistency, construct and content validity analysis). The scale was applied to 3rd year medical students in a simulated assessment environment with a final sample size of 818 assessments.

Results. The results show that the Portuguese version of the Mini-CEX is a valid scale and fit its purpose for the assessment of clinical competencies. The Cronbach's alpha coefficient (0.927), confirmed the internal consistency of the scale. Additionally, the validity analysis also proved to be satisfactory, with confirmatory results for all domains of the analysis.

Conclusions. This work intends to provide a scale, translated, adapted and validated to Portuguese that is focused on clinical competencies. Given the confirmatory results of the scale's validity, supporting its feasibility and applicability, we believe this tool is ready to be implemented as a complement to clinical skills assessment.

Key words. Clinical competences. Mini-CEX. OSCE. Portuguese. Workplace-based assessment.

Traducción, adaptación y validación del Mini-Clinical Evaluation Exercise (Mini-CEX) al idioma portugués europeo

Introducción. El objetivo de este trabajo es validar herramientas para evaluar las competencias clínicas de los estudiantes de medicina de pregrado en el lugar de trabajo. Una de las escalas más conocidas es el *Mini-Clinical Evaluation Exercise* (Mini-CEX). Esta escala se ha estudiado ampliamente, pero su validez es muy variable entre los estudios y nunca se ha validado para el idioma y el contexto portugués europeo.

Sujetos y métodos. El proceso de traducción del Mini-CEX fue realizado por dos personas bilingües y supervisado por cuatro médicos especializados en educación médica. Se llevaron a cabo métodos de naturaleza cualitativa (traducción, evaluación de la traducción, traducción inversa) y cuantitativa (consistencia interna, construcción y análisis de validez de contenido). La escala se aplicó a estudiantes de medicina de tercer año en un entorno de evaluación simulada, con un tamaño final de la muestra de 818 evaluaciones.

Resultados. Los resultados muestran que la versión portuguesa del Mini-CEX es una escala válida y se ajusta a su propósito para la evaluación de las competencias clínicas. El coeficiente alfa de Cronbach (0,927) confirmó la consistencia interna de la escala. Además, el análisis de validez también demostró ser satisfactorio, con resultados confirmatorios para todos los dominios del análisis.

Conclusiones. Este trabajo pretende proporcionar una escala, traducida, adaptada y validada al portugués, que se centre en las competencias clínicas. Dados los resultados confirmatorios de la validez de la escala, que respaldan su viabilidad y aplicabilidad, creemos que esta herramienta está lista para implementarse como complemento de la evaluación de habilidades clínicas.

Palabras clave. Competencias clínicas. ECOE. Evaluación basada en el lugar de trabajo. Mini-CEX. Portugués.

School of Medicine; University of Minho; Braga, Portugal (R. Sousa, P. Costa, J. Cerqueira, J.M. Pêgo, A. Santa Cruz, N. Sousa, V.H. Pereira). Life and Health Sciences Research Institute, ICVS; School of Medicine; University of Minho; Braga, Portugal (R. Sousa, P. Costa, J. Cerqueira, J.M. Pêgo, A. Santa Cruz, N. Sousa, V.H. Pereira). ICVS/3B's; PT Government Associate Laboratory; Braga/Guimarães, Portugal (P. Costa, J. Cerqueira, J.M. Pêgo, N. Sousa, V.H. Pereira). Internal Medicine Department; Hospital de Braga; Braga, Portugal (A. Santa Cruz, A. Oliveira e Silva). Cardiology Department; Hospital de Braga; Braga, Portugal (V.H. Pereira).

Corresponding author:

Dra. Rita Matos Sousa. School of Medicine. Campus de Gualtar. Universidade do Minho. 4710-057 Braga (Portugal).

E-mail:

ritasousa@med.uminho.pt

Acknowledgements:

The authors would like to thank both our translators (Dra. Isabel Matos and Dra. Edite Ferreira) for contributing to the work herein presented.

Received:

26.05.20.

Accepted:

11.06.20.



Artículo open access bajo la licencia CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

ISSN: 2014-9832

ISSN (ed. digital): 2014-9840

Declarations:

The authors have no conflicts of interest to declare. The authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work. The corresponding author grants on behalf of all authors a worldwide licence to the publishers. The authors declare that patients or the public were not involved in the design, or conduct, or reporting of our research.

© 2020 FEM

Introduction

In the past decades, research in medical education has highlighted the importance of assessment of competences with effective feedback as an important mechanism to promote the shift from 'assessment of learning' to 'assessment for learning' [1,2]. This premise, based in Miller's [3] and Mager's [4] principles changed the dogmas of the assessment methodologies, giving more emphasis to the processes that underlie learning and consolidation of knowledge. The latter is also critically promoted by a shift in the focus of assessment from acquisition of knowledge ('knows') to acquisition of competencies ('does') [1,5,6].

However, assessment of competences poses relevant challenges, especially due to the lack of appropriate tools and processes. In fact, and as an example, it has been shown that, the majority of first-year trainees in internal medicine residency program were neither observed nor received feedback more than once by a faculty member during a clinical encounter [7]. In addition, real-life scenarios may be hard to simulate in controlled environments and not fit all needs of competence training and assessment. Part of this problem is solved by the introduction of simulation scenarios, both for learning/training of skills in simulation labs and assessment processes, namely using objective structured clinical exams (OSCEs). In parallel, the focus in workplace-based assessment (WBA) grew and new tools for assessment in the workplace context appeared. WBA scales were created to foster the opportunity for the assessment of clinical skills and provide direct feedback with the final goal of improving global performance. Many tools have been developed with this premise in mind and some of the most known are i) the Mini-Clinical Evaluation Exercise (Mini-CEX) [8], ii) the Direct Observation Procedural Skills (DOPS) and iii) the case-based discussion. Each of these tools has different specific assessment domains and focuses but the goal of providing direct and clear feedback is maintained and considered essential for the good use of these methods [2,6,9].

In the Portuguese context, the case-based discussion is by far the most used method. While it may present some advantages (e.g. it is very comprehensive), it also presents several limitations: 1) it does not assess the ability to perform a clinical encounter; 2) it is time-consuming, being performed only a few times during the medical degree; and 3) it is artificial when compared with the daily clinical encounter.

The Mini-CEX scale was originally created for post-graduate assessment by the American Board of Internal Medicine, to encourage the observation of performance in short daily clinical encounters by qualified faculty. This scale focus primarily in providing useful feedback to the person being evaluated, with the main goal being the induction of change in behavior and, ultimately, the improvement of the student's clinical performance [8].

The Mini-CEX has been validated in several languages and in different populations [10-13]. Nevertheless, its validity in Portuguese language and population is still to be studied. In this work, we aim to translate, adapt and validate the Mini-CEX to the Portuguese language. For this, we applied the Mini-CEX in a Internal Medicine Department and in an OSCE.

Subjects and methods

This study took place at the School of Medicine-University of Minho (EM-UM) and at the Internal Medicine Department at the Hospital of Braga. The experimental protocol was approved by the Ethics Committee of the University of Minho (CEICVS 072/2019). The Declaration of Helsinki and the Council of Europe's Convention on Human Rights were strictly followed [14,15].

An extensive literature review regarding WBA scales, mainly the Mini-CEX, as well as their variations, design, validation process and impact evaluation, was conducted through available databases [16-20].

Mini-CEX Scale

In its original version, the Mini-CEX, developed by the American Board of Internal Medicine (Appendix 1), is an essentially formative assessment scale, composed by 7 items that assesses the competencies inherent to the clinical interview, the physical examination, the patient counseling and the clinical judgement. The last item assesses global clinical performance.

Each item is quantitatively rated in a 9-point Likert scale and qualitatively rated in three classes: 'unsatisfactory' (classification from 1 to 3), 'satisfactory' (classification from 4 to 6) and 'superior' (classification from 7 to 9). The Mini-CEX was designed to be short (10 to 20 min of observation and 5 to 10 min providing feedback) and easy to apply.

The Mini-CEX was developed in a post-graduation context, with the purpose of facilitating the formative evaluation of nuclear clinical competen-

cies and it can be used by the assessors as a routine evaluation of the students in any scenario.

Simulation scales

The OSCE consisted of six 15-minutes stations, each with 10 minutes for the clinical interview and 5 minutes for a specific task of physical examination. Four different domains were assessed using five different assessment scales. The first and second domains, Information Gathering (Hx) and Physical Examination (ExF) respectively, were assessed with checklist scales (Hx and ExF). The third domain was Communication, assessed with a Communication Assessment Scale developed at EM-UM and rated by two groups of assessors: faculty (CASF) and Standardized Patients (CASSP), all of those with assessment experience. Finally, the fourth domain was a post-encounter task. The final classification of the OSCE is the combination of the 5 different classifications of the four domains.

Scale translation

The translation of the English version of Mini-CEX to the Portuguese language was conducted by two Portuguese independent English teachers with expertise in translation. One of the translators worked together with a medical student from the EM-UM, to better understand the nuances of the topic being translated. Discrepancies between the two translators were discussed and resolved between the original translators, and with the assistance of a physician specialized in medical education. The scale was pilot-tested in a small sample of physicians who usually supervise medical students, medical students and faculty of EM-UM that provided feedback about their thoughts on each item. Suggestions from this group were assimilated into the preliminary version. This preliminary version was submitted to a committee of four physicians, experts in medical education, which resulted in the final forward-translation (Portuguese version).

After the translation process, the scale was then back-translated by the same initial team and submitted to the committee of experts, that produced the final back-translation (English version). The final back-translation was then sent to the original authors to assess if the original purpose of the scale was maintained.

After the translated scale passed through a preliminary pilot testing and subsequent revisions, we conducted a final pilot test among the intended respondents for initial validation. In this final pilot

test, the final version of the scale was administered in a workplace context, as would happen in a real WBA, and the participants were asked if the items were perceptible and it's objective clear.

Participants

To study the validity of the scale, trained faculty rated each student during a high-stakes OSCE at the final of the 3rd year of our 6-year Medical Degree. This OSCE is composed by 6-stations where students must collect the clinical history, perform a specific task of the physical examination and provide counseling about the next steps of patient management.

Each individual was assessed in 6 different stations. There was no significant time interval between assessments.

Sample size

For this work the sample size was calculated using a ratio of 10:1 (10 respondents for scale item) with a minimum of 70 respondents for a total of 7 items [21].

The final sample size of the 3rd year students assessed during the OSCE was 818 assessments with a total of 143 students. The assessments were performed by 34 faculty with experience in evaluating students' clinical performance.

Reliability-internal consistency

Internal consistency reflects the extent to which the scale items are intercorrelated, or whether they are consistent in measurement of the same construct. In this work we used the coefficient alpha (Cronbach's alpha) to estimate the internal consistency of the scale. In this analysis only items P1 to P6 were included, taken into consideration that P7 is a global performance item. Cronbach's alpha ranges from 0 to 1. Cronbach's $\alpha = 0$ indicates no internal consistency (none of the items are correlated with one another), whereas $\alpha = 1$ reflects perfect internal consistency (all the items are perfectly correlated with one another). In practice, Cronbach's alpha of at least 0.70 has been suggested to indicate adequate internal consistency. On the other hand, an alpha value that is too high ($\alpha \geq 0.90$) suggests that some questionnaire items may be redundant.

Validity

The validity of a scale assesses if that scale measures

Table I. Minimum value of CVR ($p = 0.05$).

No. of panelists	Minimum value
10	0.62
11	0.59
12	0.56
13	0.54
14	0.51
15	0.49
20	0.42

what it is intended to measure. In other words, it verifies if the inferences and conclusions made based on the results of the scale are valid. There are two major types of validity to be considered when validating a questionnaire: content validity and construct validity.

Content validity

Content validity refers to the extent to which the items in a scale are representative of the entire theoretical construct the questionnaire is designed to assess. Although the construct of interest determines which items are written and/or selected in the questionnaire development/translation phase, content validity of the scale should be evaluated after the initial form of the scale is available.

The scale was assessed by a committee of experts that judged whether the items are adequately measuring the construct intended to assess, and whether the items are sufficient to measure the domain of interest, classifying each item of the scale in three levels: not necessary, useful but not essential, essential.

The content validity ratio (CVR) was then calculated for each item by employing the Lawshe's method [22]. The CVR proposed by Lawshe is a linear transformation of a proportional level of agreement on how many experts within a panel rate an item 'essential'.

The final evaluation to retain the item based on the CVR depends on the number of total panel members. Table I shows the guideline for the valid value of CVR in order to retain the item.

Participants' satisfaction

Participant's satisfaction analysis intends to replace

to some level the previously used term 'face validity' which, in turn, refers to the degree to which the respondents judge the scale items to be valid. To assess satisfaction, the participants (both students and assessors) answered a questionnaire of four 5-point Likert scale questions and a 9-point Likert scale question (Table II), respectively, regarding their satisfaction with the scale.

Construct validity

Construct validity refers to the extent to which a measure adequately assesses the construct it purports to assess [23]. The construct validity of a questionnaire can be evaluated by estimating its association with other variables. It should be correlated positively, negatively, or not at all.

In order to assess the construct validity of the Portuguese Mini-CEX, an Exploratory Factor Analysis (EFA) was performed on a random sample of approximately 50% of the total sample ($n = 376$). EFA is a statistical analysis used to explore the underlying structure and relationship of multiple variables, that allows the reduction of the number of variables.

To test the suitability of the scale for the factor analysis, we used: 1) Kaiser-Meyer-Olkin (KMO), which measures sampling adequacy, ranging from 0 to 1, in which higher values mean higher suitability and a value of 0.6 is a suggested minimum and 2) Bartlett's test of sphericity, which tests the hypothesis that the correlation matrix is an identity matrix, which would indicate that the variables are unrelated and therefore unsuitable for structure detection. Taken together, these tests provide a minimum standard which should be passed before a factor analysis is conducted. After the suitability tests, we used the Principal Axis Factoring Analysis as the extraction method.

On a second moment, a Confirmatory Factor Analysis (CFA) was performed on the remaining variables (after the random selection for the EFA; $n = 442$). The CFA is a multivariate statistical procedure that is used to test how well the measured variables represent the number of constructs. In this case, the CFA was used to confirm the analysis performed in the previous EFA.

Criterion validity

Criterion validity refers to degree to which there is a relationship between a given test score and performance on another measure of particular relevance, typically referred to as criterion. There are two forms of criterion validity: predictive (criterion) and concurrent (criterion) validity.

The concurrent validity was assessed through the correlation between the simulation-based assessments scales usually used in the OSCE and the Mini-CEX.

The OSCE consisted of six 15-minutes stations, each with 10 minutes for the clinical interview and 5 minutes for the examination of the patient. Four different domains were assessed with five different assessment scales/tasks. The first and second domains, Information Gathering (Hx) and Physical Examination (ExF) respectively, were assessed with a checklist scale (Hx and ExF). The third domain was Communication, assessed with a Communication Assessment Scale developed in the EM-UM and rated by two types of assessors: the faculty (CASF) and the Standardized Patients (CASSP), all of those with assessment experience. And finally, the fourth domain was a post-encounter task. The final classification of the OSCE is the combination of the 5 different classifications of the four domains.

To study the correlation between the two assessment methods, we calculated the Pearson correlation coefficient (R) and the R^2 in two different analysis. For 'analysis 1' we used the following variables: HxX, ExFX, CASFX, CASSPX and P7_X (the overall clinical competence item) classification for each OSCE station. For 'analysis 2' we used the average classification of Hx_Average, ExF_Average, CASF_Average, CASSP_Average and P7_Average for each individual student and the correspondent Final OSCE Classification. R values near -1 or 1 are considered perfect correlation; strong correlation if $-0,95$ or $+0,95$; medium correlation for $-0,5$ and $+0,5$ values; weak correlation if $-0,1$ or $+0,1$ and no correlation if zero.

Normal distribution was assumed for all variables, given the high sample size ($n > 30$).

The predictive validity was not assessed in this work.

Results

Scale translation

The final forward translation and back translation of the 7 items can be seen in Table III and IV, respectively. The full documents of both forward and back translations can be found in Appendix 2.

The back translation was approved by the original authors of the scale, in the person of Dr. John J. Norcini, PhD.

The scale was piloted in 5 medical students and no major problems were detected.

Table II. Student's and assessor's questions regarding their satisfaction with Mini-CEX.

Students questions (1, low; 5, high)	1. The Mini-CEX is a practical method
	2. The answers are a fair assessment of your skills
	3. This process is useful for your personal development
	4. This process gave useful information about you to the assessor
Assessors question (1, low; 9, high)	1. Assessors satisfaction with the Mini-CEX

Table III. Final version of the translated items.

	Não satisfaz			Satisfaz			Satisfaz bastante			Não observado /aplicável
1. Relação médico-doente	1	2	3	4	5	6	7	8	9	n/o
2. Recolha de informação	1	2	3	4	5	6	7	8	9	n/o
3. Competências no exame físico	1	2	3	4	5	6	7	8	9	n/o
4. Raciocínio clínico	1	2	3	4	5	6	7	8	9	n/o
5. Aconselhamento e planeamento	1	2	3	4	5	6	7	8	9	n/o
6. Organização na abordagem ao doente	1	2	3	4	5	6	7	8	9	n/o
7. Competência clínica global	1	2	3	4	5	6	7	8	9	

Table IV. Back-translation of the items.

	Unsatisfactory			Satisfactory			Very satisfactory			Not observed / applicable
1. Doctor-patient relationship	1	2	3	4	5	6	7	8	9	n/o
2. Information gathering	1	2	3	4	5	6	7	8	9	n/o
3. Physical examination skills	1	2	3	4	5	6	7	8	9	n/o
4. Clinical judgment	1	2	3	4	5	6	7	8	9	n/o
5. Counseling and planning	1	2	3	4	5	6	7	8	9	n/o
6. Patient approach organization	1	2	3	4	5	6	7	8	9	n/o
7. Overall clinical competence	1	2	3	4	5	6	7	8	9	

Validation process

For the final validation process, we had a total of 143 subjects, 34 faculty members and 818 assessments. The missing values ($n = 101$ missing values

Table V. Descriptive statistics of the items.

	Items						
	P1	P2	P3	P4	P5	P6	P7
Mean	6.501	6.108	6.784	5.807	5.896	6.113	6.162
Median	7.00	6.00	7.00	6.00	6.00	6.00	6.00
Standard deviation	1.718	1.814	1.801	1.742	1.753	1.709	1.606
Minimum	1	1	1	1	1	1	1
Maximum	9	9	9	9	9	9	9

Table VI. Internal consistency analysis.

Item	Corrected item-total correlation	Cronbach's alpha if item deleted	Global Cronbach's alpha
P1	0.766	0.916	0.927
P2	0.822	0.909	
P3	0.601	0.938	
P4	0.854	0.905	
P5	0.828	0.908	
P6	0.870	0.903	

Table VII. Content validity. Lawshe score.

Item	Essential (n)	Useful but not essential (n)	Not necessary	Lawshe score
P1	15	1	0	0.875
P2	16	0	0	1.000
P3	15	1	0	0.875
P4	14	2	0	0.750
P5	12	4	0	0.500
P6	13	3	0	0.625
P7	13	3	0	0.625

of a total of 5,726; 1,76%) were replaced with the subject's 7 items score average. The descriptive statistics of each item can be found in Table V.

Reliability

In Table VI, we show the results of the internal consistency analysis of each item, with and without exclusion of items, as well as the corrected item-total correlation. The evaluation of the internal consistency of the scale by the Cronbach's alpha reveals values superior to 0.70, either for the global scale (0.927), as for the individual items. In 6 of the 7 items (P1, P2, P4, P5, P6 and P7), the Cronbach's alpha if item deleted was inferior to the one of the global scale, suggesting that the items in question substantially contribute to the global scale. Only P3 had a superior Cronbach's alpha if item deleted (0.938), but very close to global Cronbach's alpha. The corrected item-total correlation suggested good discriminative power, surpassing the critical value of 0.20, defined as the minimum value for a good correlation index.

Validity

Content validity was guaranteed for every item by 16 panel members, with a minimum consensus of 0.50 Lawshe score (for P5) and a maximum of 1 (for P2). The results for each item are presented in Table VII.

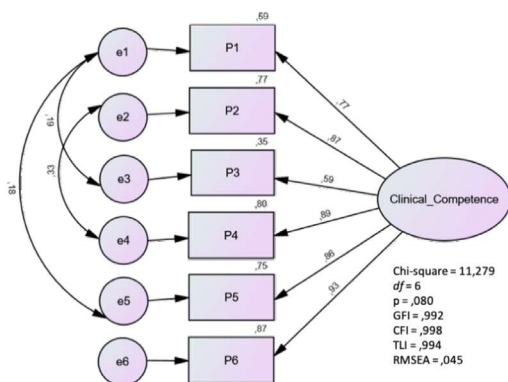
Overall satisfaction (Table VIII), assessed by both students and faculty, was considered satisfactory with an average of 4.79 out of 5 for students and an average of 7.88 out of 10 for faculty.

For the construct validity, test suitability was ensured by the KMO measure ($KMO = 0.906$) and the Bartlett's test of sphericity ($\chi^2_{(376)} = 1784.114$; $p < 0.001$).

Construct validity was firstly assessed by an EFA with a Principal Axis Factoring as the extraction method (Table IX). One factor was extracted, including items P1 to P6. From the six items analyzed, P6 is the item with the most expression, while P3 is the item with least expression.

A new variable, 'Clinical Competence' (CC), representing factor 1, was computed from the average of items P1 to P6. The correlation between CC and P7 was statistically very strong ($p < 0.001$; $R = 0.961$). In fact, this correlation is maintained for the all sample size ($p < 0.001$; $R = 0.959$), and the predictive value of the 6 items maintains its distribution when predicting P7.

To confirm the EFA, we performed a CFA on a first model compiling one latent variable with 6 observable variables, as suggested by the EFA. This model revealed inadequate goodness of fit. The post-hoc analysis of the models suggested additional covariances between e1-e3, e1-e5 and e2-e4. For that reason, a new model was developed, as seen in Figure. The second model confirms the analysis per-

Figure. Construct validity: Confirmatory Factor Analysis model.**Table VIII.** Student's and faculty's overall satisfaction.

	Average answer
Students questions (1, low; 5, high)	1. The Mini-CEX is a practical method
	2. The answers are a fair assessment of your skills
	3. This process is useful for your personal development
	4. This process gave useful information about you to the assessor
	Total
Assessors question (1, low; 9, high)	1. Assessors satisfaction with the Mini-CEX

formed with the EFA, but identifies additional covariances between items, raising awareness to their individual suitability and interdependence.

Concurrent validity

To better understand the strength of the relationship between the Mini-CEX and the OSCE results, we performed a Pearson's correlation (Table X) that demonstrated a significant correlation between both the assessment methods, not only in the final score ($R_{(143)} = 0.796$), but in all the domains assessed in the OSCE. A more detailed analysis of each OSCE station can be found in Table XI.

Discussion

This work demonstrates that the Portuguese version of the Mini-CEX has good internal consistency and reliability. This observation is in line with other validity studies, mostly in the English language [10-13,24-26]. Importantly, the use of this scale enables better assessment of clinical skills and provides relevant feedback to the students.

Regarding the validation process, and starting with the internal consistency analysis, the scale has a very high Cronbach's alpha score (0.927). However, an important reflection about the internal consistency of the Mini-CEX, is the high sample size, which might influence the analysis. Nevertheless, and to support this analysis, we analyzed the internal consistency in a much smaller sample of different students and faculty that confirmed the good internal consistency (Cronbach's alpha: 0.889; $n = 32$).

Table IX. Construct validity: exploratory factor analysis.

Factor Matrix ^a		
Item	Factor 1	Dimension
P1	0.792	Clinical competence
P2	0.842	
P3	0.625	
P4	0.889	
P5	0.859	
P6	0.924	

Extraction Method: principal axis factoring. ^a1 factors extracted; 5 iterations required.

The validity of the scale also proved very satisfying, with good content and construct validity. In fact, 16 experts were consulted to perform this qualitative analysis and no item was considered 'not necessary' by any of the panel members. Satisfaction, another subjective and qualitative analysis, also demonstrated high values, especially by the students but also by the faculty. Given that this a formative scale, this type of validity is important because it sustains the scale feasibility amongst its users, and most importantly, shows that the students perceive it as a good and fair assessment method, providing useful information from their assessors and for their personal development. Faculty satisfaction is critical to the implementation of the scale.

Table X. Overall correlations Mini-CEX and OSCE results.

		Hx average	ExF average	CASF average	CASSP average	Final score
P7 average	Pearson correlation	0.689 ^a	0.595 ^a	0.926 ^a	0.747 ^a	0.796 ^a
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000
	<i>n</i>	143	143	143	143	139

^a Correlation is significant at the 0.01 level (2-tailed).**Table XI.** Individual stations correlations Mini-CEX and OSCE results.

		Hx1	ExF1	CASF1	CASSP1
P7_1	Pearson correlation	0.543 ^a	0.025	0.798 ^a	0.469 ^a
	Sig. (2-tailed)	0.000	0.772	0.000	0.000
	<i>n</i>	141	141	141	141
		Hx2	ExF2	CASF2	CASSP2
P7_2	Pearson correlation	0.573 ^a	0.217 ^b	0.826 ^a	0.381 ^a
	Sig. (2-tailed)	0.000	0.011	0.000	0.000
	<i>n</i>	137	137	137	137
		Hx3	ExF3	CASF3	CASSP3
P7_3	Pearson correlation	0.563 ^a	0.275 ^a	0.756 ^a	0.440 ^a
	Sig. (2-tailed)	0.000	0.001	0.000	0.000
	<i>n</i>	143	143	143	143
		Hx4	ExF4	CASF4	CASSP4
P7_4	Pearson correlation	0.523 ^a	0.232 ^a	0.730 ^a	0.442 ^a
	Sig. (2-tailed)	0.000	0.006	0.000	0.000
	<i>n</i>	138	138	138	138
		Hx5	ExF5	CASF5	CASSP5
P7_5	Pearson correlation	0.456 ^a	0.225 ^b	0.822 ^a	0.613 ^a
	Sig. (2-tailed)	0.000	0.015	0.000	0.000
	<i>n</i>	117	117	117	117
		Hx6	ExF6	CASF6	CASSP6
P7_6	Pearson correlation	0.513 ^a	0.163	0.812 ^a	0.602 ^a
	Sig. (2-tailed)	0.000	0.054	0.000	0.000
	<i>n</i>	141	141	141	141

^a Correlation is significant at the 0.01 level (2-tailed); ^b Correlation is significant at the 0.05 level (2-tailed).

In fact, the application of the scale requires time from the faculty, which might present as an obstacle to its use; therefore, the high levels of satisfaction of the faculty with its use is of relevance to a successful implementation of Mini-CEX.

Another interesting finding occurred during the construct validity analysis. As previously explained, the item P7 was excluded from the EFA given that it is a global performance question. During this analysis, only one factor was extracted, which lead to a confirmatory analysis that computed a new variable 'CC'. As expected, this new variable is strictly correlated to P7, suggesting that the first 6 items are, in fact, all contributing to the global performance. In a different perspective and taking into account that P7 might be sufficient at measuring the clinical performance, one might question the relevance of the other 6 items. However, this reflection depends mainly on the purpose and context of the scale's use. If the scale is used in a formative context, in which providing quality and discriminated feedback is the main goal, we believe that maintaining all the items strengthens the scale and its purpose. On the other hand, if the scale is used exclusively for summative assessment purposes, the present analysis suggests that a single global performance 5-point Likert scale question might be sufficient.

During the EFA, the expression of each individual item was also assessed and, curiously, the item with most expression was P6, which refers to Patient Approach Organization. This finding was surprising and suggests that the student's ability of organizing the clinical encounter is the factor that influences more the faculty's opinion regarding the global performance. In contrast, P3 was the item with least expression, suggesting that the physical examination, despite being an essential part of the clinical encounter, does not influence as significantly the global performance. The CFA actually confirmed these findings and identified new covariances between variables. Again, we decided not to aggregate or exclude any of the variables keeping in mind that formative feedback is the main goal of Mini-CEX.

Finally, the decision of studying the concurrent validity was fruitful, and highly supports that the Mini-CEX is a good clinical performance scale. Looking into these results, we can understand that the Mini-CEX strongly correlates with the assessment scales used in the OSCE, particularly with the domains that are not assessed with a checklist tool (CASF and CASSP). This is also interesting because the strength of the correlation with these two domains is closer to the one seen in the final score

correlation. In the future, we believe that a global performance Likert-point question might be of value in the Information Gathering (Hx) and the Physical Examination (ExF) domains, both assessed with checklist, in the summative perspective of the Mini-CEX applied to simulation exercises.

There are several limitations in this study that should be considered when interpreting these results: i) the design of the study did not allow convergent and divergent validation mechanisms, which could be important to establish construct validity; ii) authors did not assess test-retest reliability since the scale was administered in different simulation environments, by different assessors and sequentially during an OSCE; iii) authors did not assess inter-rater reliability because each student was assessed by different assessors but in different environments and never with the same exact condition; iv) students were recruited from a single Medical School.

The strengths of the study include: multiple assessors involved; large sample size.

In summary, our study concludes that the Portuguese Mini-CEX is a valid formative scale to assess clinical performance, with good internal consistency and validity and that is ready to be implemented in the Portuguese context as a complement to other assessment methods.

References

- Swanwick T, Chana N. Workplace-based assessment. *Br J Hosp Med* 2013; 70: 5.
- Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ* 2010; 341: c5064.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990; 65 (Suppl 9): S63-7.
- Mager RF. Preparing instructional objectives: a critical tool in the development of effective instruction. 3 ed. Atlanta: CEP Press; 1997.
- Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Adv Health Sci Educ Theory Pract* 2016; 21: 455-73.
- Liu C. An introduction to workplace-based assessments. *Gastroenterol Hepatol Bed Bench* 2012; 5: 24-8.
- Day SC, Grosso LG, Norcini JJ, Blank LL, Swanson DB, Horne MH. Residents' perceptions of evaluation procedures used by their training program. *J Gen Intern Med* 1990; 5: 421-6.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. The Mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med* 1995; 123: 795-9.
- General Medical Council. Workplace based assessment: a guide for implementation. London: GMC; 2010.
- Humphrey-Murto S, Côté M, Pugh D, Wood TJ. Assessing the validity of a multidisciplinary Mini-Clinical Evaluation Exercise. *Teach Learn Med* 2018; 30: 152-61.
- Al Ansari A, Ali SK, Donnon T. The construct and criterion validity of the Mini-CEX: a meta-analysis of the published research. *Acad Med* 2013; 88: 413-20.
- Norcini JJ, Blank LL, Duffy FD, Fortna GS. The Mini-CEX: a method for assessing clinical skills. *Ann Intern Med* 2003; 138: 476-81.
- Megale L, Gontijo ED, Motta JAC. Evaluation of medical students' clinical skills using the Mini-Clinical Evaluation Exercise (Mini-CEX). *Rev Bras Educ Med* 2009; 33: 166-75.
- Convenção para a Proteção dos Direitos do Homem e da Dignidade do Ser Humano Face às Aplicações da Biologia e da Medicina: Convenção sobre os Direitos do Homem e da Biomedicina (Conselho da Europa 1997). Resolução da Assembleia da República n.º 1/2001, Diário da República – I, Série A, n.º 2, 3 de Janeiro de 2001. URL: <http://dre.pt/util/getpdf.asp?s=dip&serie=1&idddr=2001.2A&iddip=20010014>.
- Council for International Organizations of Medical Sciences. International ethical guidelines for biomedical research involving human subjects. Geneva: CIOMS; 1993.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. COSMIN checklist manual. URL: http://fac.ksu.edu.sa/sites/default/files/cosmin_checklist_manual_v9.pdf.
- Peng DX, Lai F. Using partial least squares in operations management research: a practical guideline and summary of past research. *J Oper Manag* 2012; 30: 467-80.
- Tsang S, Royse CF, Terkaw AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth* 2017; 11 (Suppl 1): S80-9.
- Taherdoost H. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *SSRN Electronic Journal* 2016; 5: 28-36.
- Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Health* 2018; 6: 149.
- Anthoine E, Moret L, Regnault A, Sbille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014; 12: 176.
- Lawsche CH. A quantitative approach to content validity. *Pers Psychol* 1995; 28: 563-75.
- Nunnally JC, Bernstein IH. Psychometric theory. 3 ed. New York: McGraw-Hill; 1994.
- Holmboe ES, Huot S, Chung J, Norcini JJ, Hawkins RE. Construct validity of the Mini Clinical Evaluation Exercise (Mini-CEX). *Acad Med* 2003; 78: 826-30.
- Durning S, Cation LJ, Market RJ, Pangaro LN. Assessing the reliability and validity of the Mini-Clinical Evaluation Exercise for Internal Medicine residency training. *Acad Med* 2002; 77: 900-4.
- Hatala R, Ainslie M, Kassen BO, Magkie I, Roberts M. Assessing the Mini Clinical Evaluation Exercise in comparison to a national specialty. *Med Educ* 2006; 40: 950-6.

Appendix 1. Original version of the Mini-CEX.

Mini-Clinical Evaluation Exercise (CEX)											
Evaluator: _____						Date: _____					
Resident: _____						<input type="radio"/> R-1 <input type="radio"/> R-2 <input type="radio"/> R-3					
Patient Problem/Dx: _____											
Setting: <input type="radio"/> Ambulatory <input type="radio"/> In-patient <input type="radio"/> ED <input type="radio"/> Other _____											
Patient: Age: _____ Sex: _____ <input type="radio"/> New <input type="radio"/> Follow-up											
Complexity: <input type="radio"/> Low <input type="radio"/> Moderate <input type="radio"/> High											
Focus: <input type="radio"/> Data Gathering <input type="radio"/> Diagnosis <input type="radio"/> Therapy <input type="radio"/> Counseling											
1. Medical Interviewing Skills (<input type="radio"/> Not observed)											
1 2 3			4 5 6			7 8 9					
UNSATISFACTORY			SATISFACTORY			SUPERIOR					
2. Physical Examination Skills (<input type="radio"/> Not observed)											
1 2 3			4 5 6			7 8 9					
UNSATISFACTORY			SATISFACTORY			SUPERIOR					
3. Humanistic Qualities/Professionalism											
1 2 3			4 5 6			7 8 9					
UNSATISFACTORY			SATISFACTORY			SUPERIOR					
4. Clinical Judgment (<input type="radio"/> Not observed)											
1 2 3			4 5 6			7 8 9					
UNSATISFACTORY			SATISFACTORY			SUPERIOR					
5. Counseling Skills (<input type="radio"/> Not observed)											
1 2 3			4 5 6			7 8 9					
UNSATISFACTORY			SATISFACTORY			SUPERIOR					
6. Organization/Efficiency (<input type="radio"/> Not observed)											
1 2 3			4 5 6			7 8 9					
UNSATISFACTORY			SATISFACTORY			SUPERIOR					
7. Overall Clinical Competence (<input type="radio"/> Not observed)											
1 2 3			4 5 6			7 8 9					
UNSATISFACTORY			SATISFACTORY			SUPERIOR					
Mini-CEX Time: Observing _____ Mins Providing Feedback _____ Mins											
Evaluator Satisfaction with Mini-CEX											
LOW	1	2	3	4	5	6	7	8	9	HIGH	
Resident Satisfaction with Mini-CEX											
LOW	1	2	3	4	5	6	7	8	9	HIGH	
Comments: _____											

Resident Signature						Evaluator Signature					

Appendix 2. Forward and Back Translations of Mini-CEX.



UNIVERSIDADE DO MINHO
Escola de Medicina

MINI EXERCÍCIO DE AVALIAÇÃO CLÍNICA: Mini-CEX

Estudante: _____

Nº mecanográfico: _____ Unidade Curricular: _____

Tutor/Formador: _____ Data: _____

Instituição (Serviço, se aplicável): _____

Contexto:

___ Internamento ___ Consulta ___ SU ___ Outro: _____

Complexidade para o nível de desenvolvimento do estudante:

___ Baixa ___ Moderada ___ Elevada

Doente:

Idade: _____ Género: _____ 1ª Obs. _____ Seguimento _____

Objetivo (assinalar todas as aplicáveis):

___ Colheita de história ___ Diagnóstico ___ Terapêutica ___ Aconselhamento

Problema médico / Diagnóstico Principal: _____

Por favor, classifique o estudante para cada um destes parâmetros, tendo em conta o seu nível de desenvolvimento.

	NÃO SATISFAZ			SATISFAZ			SATISFAZ BASTANTE			NÃO OBSERVADO / APLICÁVEL
1. Relação médico-doente	1	2	3	4	5	6	7	8	9	n/o
2. Recolha de informação	1	2	3	4	5	6	7	8	9	n/o
3. Competências no Exame Físico	1	2	3	4	5	6	7	8	9	n/o
4. Raciocínio Clínico	1	2	3	4	5	6	7	8	9	n/o
5. Aconselhamento e Planeamento	1	2	3	4	5	6	7	8	9	n/o
6. Organização na abordagem ao doente	1	2	3	4	5	6	7	8	9	n/o
7. Competência Clínica Global	1	2	3	4	5	6	7	8	9	



Pontos fortes observados

Pontos a melhorar

Tempo do Mini-CEX:

Tempo de Observação: _____ Mins

Tempo de Feedback: _____ Mins

Satisfação do Avaliador com o Mini-CEX

BAIXO 1 2 3 4 5 6 7 8 9 ALTO

Satisfação do Estudante com o Mini-CEX

BAIXO 1 2 3 4 5 6 7 8 9 ALTO

Comentários do estudante

Assinatura do Estudante

Assinatura do Tutor / Formador



UNIVERSIDADE DE MINHO
School of Medicine

Mini-Clinical Evaluation Exercise (CEX)

Student: _____

Personal Number: _____ Curricular Unit: _____

Tutor/Trainer: _____ Date: _____

Institution (Department, if applicable): _____

Setting:

___ Ward ___ In-patient ___ ED Other: _____

Complexity for the student's level of development:

___ Low ___ Moderate ___ High

Patient:

Age: ___ Sex: ___ 1st Observation: ___ Follow-up: ___

Focus:

___ Data Gathering ___ Diagnosis ___ Therapy ___ Counseling

Patient medical problem / Main diagnosis: _____

Please rate the student for each of these parameters, taking into account the student's level of development.

	UNSATISFACTORY			SATISFACTORY			VERY SATISFACTORY			NOT OBSERVED / APPLICABLE
1. Doctor-Patient Relationship	1	2	3	4	5	6	7	8	9	n/o
2. Information Gathering	1	2	3	4	5	6	7	8	9	n/o
3. Physical Examination Skills	1	2	3	4	5	6	7	8	9	n/o
4. Clinical Judgment	1	2	3	4	5	6	7	8	9	n/o
5. Counseling and Planning	1	2	3	4	5	6	7	8	9	n/o
6. Patient Approach Organization	1	2	3	4	5	6	7	8	9	n/o
7. Overall Clinical Competence	1	2	3	4	5	6	7	8	9	n/o



UNIVERSIDADE OF MINHO
School of Medicine

Strengths

Agreed actions for development

Mini-CEX Time:

Observing: _____ Mins

Providing Feedback: _____ Mins

Evaluator Satisfaction with Mini-CEX

LOW 1 2 3 4 5 6 7 8 9 HIGH

Student Satisfaction with Mini-CEX

LOW 1 2 3 4 5 6 7 8 9 HIGH

Student's Comments

Student's signature

Tutor / Trainer Signature