Analysis of a learning assessment system for pediatric internship based upon objective structured clinical examination, clinical practice observation and written examination

Gloria E. Sandoval,[1] Patricia M. Valenzuela,[1] Marcela M. Monge,[1] Paulina A. Toso,[2]
Ximena C. Triviño,[3] Ana C. Wright,[4] Enrique Paris,[5] Ignacio Sánchez,[6] Gonzalo S. Valdivia[7]

Abstract

Objective: To describe and analyze three tools used in the assessment system applied to the pediatric internship over a 7-year period at the School of Medicine, Pontifícia Universidad Católica de Chile, Santiago, Chile.

Methods: Retrospective observational research design for the assessment modalities implemented in the pediatric internship from 2001 through 2007. The tools were as follows: objective structured clinical examination (OSCE), written examination and daily clinical practice observation guidelines (DCPOG). The assessment methods were applied to the sixth-year pediatric internship with a total of 697 students. Statistical analysis included a descriptive assessment, with correlation and simple linear and multiple regressions (ANOVA), Bonferroni test and Cronbach's alpha coefficient. Significance level was set at $p < 0.05$.

Results: OSCE success scores were reached in 75.7±8%, with a better mean among females ($p < 0.001$). OSCE scores improved after the third year of implementation. Cronbach's alpha coefficient was 0.11-0.78. Written examination had a mean score of 79.8±10% and there were no sex differences. Mean DCPOG score was 97.1±3% and the results were better among females ($p < 0.005$). Correlation between the three assessment methods showed a moderate positive relationship except in the year of 2007, where the correlation was higher ($p < 0.001$).

Conclusions: Analysis of the learning assessment system was performed using OSCE, written examination and DCPOG, which are complementary to each other, and yielded good results.

Introduction

Traditionally, oral examinations have been used to assess the cognitive mastery and intellectual skills related to a clinical problem, upon completion of internships. However, such assessment modality has limitations due to its low degree of reliability and difficult standardization.[1,2] On the other hand, written examination (WE), a commonly used complementary tool, assesses a greater cognitive information spectrum, although it can hardly reach more complex levels of assessment, such as critical analysis and rationalization.[3] However, both tools do not enable the examinee to demonstrate his/her grasp of clinical competencies, since they do not reproduce the real tasks

132 Jornal de Pediatria - Vol. 86, No. 2, 2010

Learning assessment system for pediatric internship - Sandoval GE et al.

that a physician must undertake during an encounter with a patient.[4]

For a few decades now, introduction of assessment methods that simulate practical clinical situations has provided a solution to the limitation described above. In such assessment modality, one of the most used tools is the objective structured clinical examination (OSCE),[5] which since 1975[6-8] has proven to be adequate to assess clinical competencies[6,9] and has shown to provide enough validity in the interpretation of results upon its application among pre- and postgraduate medical students.[4,10-12] Simulations are approximations to reality attempting to reproduce clinical situations in standardized conditions, thus enabling, through observation, assessment of the attainment of specific objectives.[13] They are widely used to assess skills for clinical reasoning, anamnesis, physical examination, diagnostic approach, order placing for patients and performance of procedures, among others.[14-16]

Another source of information that enables assessment of the learning process that students undergo during an internship is the daily clinical practice observation guideline (DCPOG).[17] With a predefined and structured guideline, teachers assess, through direct observation, the performance of examinees in various clinical activities. Such method has the advantage of facilitating the observation of individual performance in real situations, thus yielding more information on communicational and interpersonal skills and professionalism. It has nonetheless limitations regarding the examiners, in view of the fact they should be trained, and therefore standardized.[18]

None of the methods described provides, on its own, a comprehensive assessment of all competencies. Some studies propose use of combinations of the different methods to harmonize the assessment of cognitive skills with the evaluation of the required complex skills for an adequate professional competence. Kreiter et al.[18] proposed a combination of the scores obtained at the DCPOG and the OSCE to obtain a unique measure enabling a better representation of the assessment of the examinee's clinical skills.[19]

Until 2001, the assessment system used in the pediatric internship at the Pontificia Universidad Católica de Chile (PUC) School of Medicine, Santiago, Chile, used to be carried out through the application of two tools, DCPOG of each of the clinical rotations and an oral exam upon completion of the program. In that year, the assessment system underwent a restructuring that involved implementation of a written exam in substitution of the oral exam, application of an OSCE upon completion of the internship, and maintenance of DCPOG.

The purpose of the present report is to describe the results of the application of such assessment system and to analyze each tool and the relations among them over a 7-year period, from 2001 through 2007.

Methods

Retrospective observational research design for the assessment modalities used from 2001 through 2007 that were implemented in the pediatric internship, which occurs over 12 weeks during the sixth year of medical career. Each intern group was identified with a number from 1 to 4 according to the corresponding time sequence across each of the study years. Data collection was approved by the ethics committee of the university (09-174).

Description of the assessment tools:

1) OSCE: Since 2001 it has been applied 27 times: three times in the first year and four times per year on the following year. OSCE was used upon completion of the pediatric internship, in groups of approximately 25 interns.

   During the first 3 years, each OSCE included 20 to 26 stations. From 2004 on, each OSCE was implemented in half of interns, including 12 to 15 stations, and was developed in identical sequential circuits, where there was no contact between interns. The stations had 5-minute duration. Five of the stations involved the participation of trained professional specialized actors (PSA), and applied two observational dichotomist comparison guidelines on the examinee's performance; one was applied by teachers and the other was applied by a PSA, weighted 80 and 20%, respectively.

   The OSCE pool consists of 29 stations, seven of which have an equivalent version (versions A, B, and C of the same station). The teachers in charge of the internship created and selected the stations, as well as analyzed their results with technical assistance in education.

2) WE: it consisted of 60 multiple-choice questions and integrated inpatient (50%), outpatient (25%) and neonatal (25%) clinical subjects in percentages proportional to the duration of clinical rotations. The pool contains 600 questions with information on the psychometric characteristics for each individual question and with regard to the complete test, each of them with its validity sources.[20]

3) DCPOG: Such guideline, common to all internships of the medical career, consisted of 10 indicators based on clinical competencies, rated according to five achievement levels: anamnesis, physical examination, diagnostic hypotheses, therapeutic plan, team work, responsibility, clinical skills, medical problem management, study habits, and rational use of resources. Each examinee was assessed with such guideline in each of the clinical rotations.

   OSCE, WE and DCPOG were used as methods complementary to each other, assessing all objectives stated in the internship program. The final grade was calculated based on the grades obtained: the DCPOG score was weighted two-thirds and the average between the WE

Learning assessment system for pediatric internship - Sandoval GE et al.

Jornal de Pediatria - Vol. 86, No. 2, 2010 **133**

the OSCE scores was weighted one-third. The event of failure to pass the examination after a rotation resulted in a re-enrollment in the same rotation for repetition. A performance in OSCE or WE under 60% was a cause of internship failure.

In the description and analysis, the grades reached by the interns were expressed as success percentage and the variables considered were addressed as continuous categorical variables.

Statistical analysis included a descriptive assessment, with correlation and simple linear and multiple regressions adjusted for selected variables. An analysis of variance (ANOVA) and a Bonferroni test were used to compare categorical variables. Correlation between results from the assessment tools was calculated with the Pearson's correlation coefficient. Pearson's score between 0.6-0.8 suggests strong correlations and values between 0.4-0.6, that is, fair correlations. However, these cut-offs are related with sample size. Cronbach's alpha coefficient was calculated to assess reliability.[13] A cut-off value of 0.4 is the lower limit stated by Barman[21] for undergraduate course OSCE in medicine. On the other hand, a value of 0.6 and over is stated by Linn & Gronlund[22] as a satisfactory performance of reliability. The statistical programs SPSS 11.0 and Stata 8.0 were used. Differences were considered statistically significant when $p < 0.05$.

Relevant academic history included the following:

a) The average of grades obtained upon taking pediatrics during the fifth year of the career, expressed with a 1 to 7 scale. The passing mark for pediatrics is 4.

b) The cumulative weighted average, which includes the grades in each of the subjects attempted by a student from the first to the fifth year of the career.

Results

Data from 697 students that took the pediatric internship from 2001 through 2007 were analyzed.

General features and prior academic history

Average grade of the pediatrics course attendants was 6.5 (Table 1). Females had a better average performance than males ($p = 0.0119$). There were no significant differences in the marks obtained upon completing the pediatrics course between the students that would later constitute the internship groups ($p = 0.29$). There were no significant differences in PPA neither between females and males nor within different internship groups ($p = 0.25$ and $p = 0.2$, respectively). However, there was a significant difference in PPA values between years, with the year of 2004 accounting for the lowest PPA ($p = 0.001$).

Objective structured clinical examination

The average success percentage for the total number of interns was 75.7% [standard deviation (SD) = 8.0; range = 35.8-91.8]. Maximum and minimum scores for each year ranged between 76 and 91.8% and between 35.8 and 64.0%, respectively. Overall, females obtained a better average than males, and such difference was statistically significant (Table 2).

Annual average success percentage evidenced a significant increase from the third year of assessment method implementation, to reach a steady state during the last 3 years (chi-square test for trend: $p < 0.001$; Table 3).

Analysis of OSCE performance for every year and for each internship group showed no significant differences between groups in any of the years under study. Cronbach's

Table 1 - General description of interns and previous academic history, between 2001 and 2007; final and weighted average mark and 95% confidence interval

| Year | Number of interns | | Final average mark in the fifth year pediatrics course (95%CI) | | Cumulative weighted average before the sixth year of the career (95%CI) | |
	Male, n (%)	Female, n (%)	Male	Female	Male	Female
2001	40 (62.5)	24 (37.5)	5.9 (5.8-6.0)	6.0 (5.7-6.2)	5.8 (5.6-5.8)	5.8 (5.7-5.9)
2002	57 (62.6)	34 (37.4)	6.6 (6.5-6.6)	6.5 (6.3-6.5)	5.8 (5.7-5.9)	5.9 (5.6-5.8)
2003	64 (56.6)	49 (43.4)	6.5 (6.4-6.5)	6.6 (6.5-6.6)	5.8 (5.6-5.8)	5.8 (5.7-5.8)
2004	64 (58.2)	46 (41.8)	6.5 (6.3-6.5)	6.6 (6.4-6.7)	5.8 (5.4-5.7)	5.6 (5.5-6.0)
2005	56 (51.4))	53 (48.6)	6.5 (6.4-6.5)	6.6 (6.5-6.7)	5.9 (5.7-5.9)	5.8 (5.7-5.9)
2006	63 (60.6)	41 (39.4)	6.6 (6.5-6.6)	6.6 (6.5-6.6)	5.8 (5.7-5.9)	5.8 (5.7-5.8)
2007	51 (48.1)	55 (51.9)	6.6 (6.5-6.6)	6.7 (6.5-6.7)	5.9 (5.8-5.9)	5.9 (5.7-5.9)
Total	395 (56.7)	302 (43.3)	6.5 (6.4-6.5)*	6.5 (6.5-6.6)*	5.8 (5.7-5.8)*	5.8 (5.8-5.9)*

95%CI = 95% confidence interval.
* Average mark (95%CI).

134 Jornal de Pediatria - Vol. 86, No. 2, 2010

Lear ning assessment system for pediatric internship - Sandoval GE et al.

Table 2 - Objective structured clinical examination, written examination and daily clinical practice observation guideline results per sex (average, standard deviation, median, range), from 2001 through 2007

	OSCE *		WE†		DCPOG‡	
	Male	Female	Male	Female	Male	Female
Average (SD)	75.2 (8.2)	77.5 (7.6)	78.5 (10.3)	79.4 (9.6)	96.6 (3.3)	97.4 (2.6)
Median (range)	75.4 (40.0-91.5)	78.2 (35.8-91.8)	80.0 (34.5-98.3)	81.8 (44.2-100)	97.1 (78.6-100)	98.6 (80.8-100)

DCPOG = daily clinical practice observation guideline; OSCE = objective structured clinical examination; SD = standard deviation; WE = written examination.
* p = 0.0008.
† p = 0.084.
‡ p = 0.0024.

Table 3 - Examinee's average success percentage in objective structured clinical examination, written examination and daily clinical practice observation guideline per year

Year	OSCE, % (SD)	WE, % (SD)	DCPOG, % (SD)
2001	68.7 (4.7)	68.3 (10.0)	97.7 (1.7)
2002	68.3 (10.3)	74.6 (8.6)	97.0 (3.2)
2003	74.7 (5.9)	79.5 (8.3)	95.2 (3.5)
2004	77.9 (6.6)	79.9 (10.1)	95.9 (3.2)
2005	79.6 (5.7)	78.5 (7.6)	97.1 (2.7)
2006	79.7 (5.5)	88.3 (7.2)	98.0 (2.0)
2007	80.3 (6.3)	78.6 (8.6)	98.2 (2.7)
Chi-square test for trend	p < 0.001	p < 0.001	p < 0.001

DCPOG = daily clinical practice observation guideline; OSCE = objective structured clinical examinationn; SD = standard deviation; WE = written examinatio.

alpha coefficient values for each OSCE application ranged from 0.11 to 0.78 and seven stations had an alpha value below 0.4. On the other hand, seven OSCE had alpha values above 0.6. Upon analysis of the impact of each station over total OSCE, seven stations had a negative effect on 21 of the 27 OSCE. When such stations were removed, an alpha value greater than 0.4 was obtained. There was no evident decrease in OSCE reliability when comparing the periods from 2001 to 2003 and from 2004 to 2007, during which time OSCE had either 20 to 26 stations or 12 to 15 stations, respectively.

Written examination

Average success rate for the interns in WE was 79.8% (SD = 10; range = 34.5-100). Maximum and minimum percentages ranged from 88.3 to 100% and from 34.5 to 68.3%, respectively. As for the performance according to sex, no statistically significant differences were found (Table 2).

The highest performance in WE was documented in 2006 and the lowest values were obtained in the first year of assessment method implementation. The difference between such years and the other years under study were statistically significant (p < 0.001) (Table 3).

Upon analyzing the performance per group each year, group 4 had a lower performance than the preceding groups (F = 9.3; p < 0.00001).

Daily clinical practice observation guideline

Average success rate among the interns following the guidelines was 97.1% (SD = 3; range = 78.6-100). Maximum percentage was 100% every year, while minimum percentage ranged between 78.6 and 92.9%. With this assessment tool, females had a better performance than males, and the difference was statistically significant (Table 2).

Results of the chi-square test for trend suggested a gradual increase between 2001 and 2007, being statistically

significant (p < 0.001). The lowest average performance was documented during 2003 and 2004, and the differences between the average performance in such years and that of the other years was statistically significant (p < 0.001).

Upon comparing the performance in DCPOG per internship group each year, there were no significant differences between groups in any of the years under study.

Comparisons between assessment tools

The highest success rates were achieved in DCPOG and the lowest were obtained for the OSCE (Table 3).

Correlations were all positive and statistically significant (Table 4). In 5 out of 7 years there was a positive correlation between WE and DCPOG; in 4 years there was a positive correlation between WE and OSCE, and in 5 years there was a positive correlation between DCPOG and OSCE. It is worth noting that in 3 years – 2003, 2005 and 2007 – all three correlations were positive and significant.

Discussion

Setting up assessment systems that are able to integrate different tools to evaluate all objectives proposed in a program is critical. The system that was implemented in 2001 for the pediatric internship meets such requirements. More than a combination of tools, such system represents tools complementary to each other to assess the required undergraduate skills for the practice of pediatrics. Thus, in such conditions, it is possible to assert the compliance with one of the most relevant sources of evidence for validity, for what the assessment of contents is comprehensive.[9,21,22]

The next step is the interpretation of results after application of various tools. This facilitates decision making on whether to continue or to modify either the tools or the whole system. Moreover, such step enables analysis of the quality of data entry on a registry.

Although the interns had the lowest average scores in OSCE as compared to the three tools that were used, the average and maximum scores increased after the first 3 years of implementation of the assessing methods. Even so, no student has attained a 100% performance. As a result of the reasonably adequate reliability indexes for a pre-graduate OSCE obtained, the study supports the use of 12 to 14 stations.[21,23]

It is important to consider four major areas of any evaluation: students, in view of their growing exposure to the method, provided by prior courses; the tool, due to improvement in the stations; teachers, in view of their training and also the experience gained over the 7 years of the internship assessment; and finally, a new approach in teaching.

The WE method also showed a gradual increase in the average success rate until 2006. Subsequently, the examination was completely modified, and as a consequence the results decreased and returned to prior levels; therefore, a continuous renewal of questions and surveillance of the question pool is considered critical. One of the weaknesses evidenced upon the result analysis was the impossibility to estimate reliability. Reliability should be calculated using the Kuder-Richardson 21 formula; however, this could not be achieved since the databases did not supply all the information required. Nevertheless, such situation was corrected as a result of the present analysis.

The high performance of students at DCPOG represents a common fact to which we were not an exception. The tool and the examiners were considered as major issues. Creation of good clinical practice assessment guidelines is a complex task, entailing difficulties to define the dimensions, success levels, format (signature, comparison)

Table 4 - Pearson's correlation coefficient between objective structured clinical examination, written examination and daily clinical practice observation guideline per year

Year	WE – DCPOG correlation	WE – OSCE correlation	DCPOG – OSCE correlation
2001	0.17	0.17	0.43*
2002	0.23*	0.13	0.25*
2003	0.24*	0.43*	0.46*
2004	0.29*	0.05	0.25
2005	0.31*	0.24*	0.32*
2006	0.11	0.39*	0.16
2007	0.48*	0.48*	0.39*
Global	0.23*	0.38*	0.28*

DCPOG = daily clinical practice observation guideline; OSCE = objective structured clinical examination; WE = written examination.
* p < 0.001.

136 Jornal de Pediatria - Vol. 86, No. 2, 2010

Learning assessment system for pediatric internship - Sandoval GE et al.

and score assignment. On the other hand, examiners might experience halo effect upon scoring, show a central tendency, benevolence or strictness, not be the same who provide tuition, or be poorly trained. In the pediatric internship, there has been an attempt to safeguard these latter aspects, as well as to ensure that only the faculty involved in the internship can be examiners. Moreover, such efforts have also targeted the provision of training in education to clinical tutors. Nevertheless, in the light of the results, we consider that this latter activity should focus on use of the guidelines since such a behavior would enable teachers to prevent appreciation errors. However, modification of the guidelines to include description of competencies and performances that should be attained would be more relevant. It was not possible to estimate reliability with Cronbach's alpha coefficient. Once again, we are facing difficulties in data entry and with databases. The fact such difficulties still remain an issue witnesses their importance.

Although measuring different domains, the positive and significant correlation between results of the assessment tools means that the best students reach the best independent results, irrespective of the assessment tool used. Finally, a nonetheless minor issue to address is the feasibility of the implementation of a system such as that described above. Obviously, such aspect should be taken into account from the beginning of the project since the resources available for our medicine schools are always limited. Such resources – human, material and financial – should be guaranteed to maintain system vitality. The pediatric internship has made arrangements to maintain, on a permanent basis, a motivated and proactive faculty for technical support in teaching; in addition, the internship has arranged use of supplies in a judicious manner to enable a continuity to achieve proposed goals.

In conclusion, analysis of the 7-year experience with the application of this assessment system for the pediatric internship at PUC allows us to state that the three methods used together are able to efficiently assess the clinical pediatric competencies required from our students.

References

1. Guerin RO. Disadvantages to Using the Oral Examination. In: Marcall EL, Bashook PG, editors. Assessing Clinical Reasoning: The Oral Examination and Alternative Methods. Evanston, IL: American Board of Medical Specialties; 1995. p. 41-8.

2. Reinhart MA. Advantages to Using the Oral Examination. In: Marcall EL, Bashook PG, editors. Assessing Clinical Reasoning: The Oral Examination and Alternative Methods. Evanston, IL: American Board of Medical Specialties; 1995. p. 31-3.

3. Martínez JM. Los métodos de evaluación de la competencia profesional: la evaluación clínica objetiva estructurada (ECOE). Educ Méd. 2005;8:S18-S22.

4. Reteguiz J, Cornel-Avendaño B. Mastering the OSCE/CSA. New York: McGraw-Hill; 1999.

5. Association of American Medical Colleges. Division of Medical Education. Medical School Graduation Questionnaire. All Schools Summary Report Final; 2007.

6. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J. 1975;1:447-51.

7. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination. Med Educ. 1979;13:41-54.

8. Barrows S. An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. Acad Med. 1993;68:451-3.

9. Carraccio C, Englader R. The objective structured clinical examination: a step in the direction of competency-based evaluation. Arch Pediatr Adolesc Med. 2000;154:736-41.

10. Rogers PL, Jacob H, Rashwan AS, Pinsky MR. Quantifying learning in medical students during a critical care medicine elective: a comparison of three evaluation instruments. Crit Care Med. 2001;29:1268-73.

11. Joorabchi B. Objective structured clinical examination in a pediatric residency training program. Am J Dis Child. 1991;145:757-62.

12. Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an objective structured clinical examination for assessing clinical performance of residents. Arch Intern Med. 1990;150:573-7.

13. Van der Vleuten CP, Swanson DB. Assessment of clinical skills with standardized patients: state of art. Teach Learn Med. 1990;2:58-76.

14. Shumway JM, Harden RM; Association for Medical Education in Europe. AMEE Guide Nº 25: The assessment of learning outcomes for the competent and reflective physician. Med Teach. 2003;25:569-84.

15. Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. Med Educ. 2004;38:199-203.

16. Petrusa ER. Clinical performance assessments. In: Norman GR, Van der Vleuten CP, Newble DI, editors. International handbook of research in medical education. Great Britain: Kluwer Academic Publishers; 2002. p. 673-709.

17. Kumar A, Gera R, Shah G, Godambe S, Kallen DJ. Student evaluation practices in pediatric clerkships: a survey of the medical schools in the United States and Canada. Clin Pediatr (Phila). 2004;43:729-35.

18. Kreiter C, Bergus G. A study of Two Clinical Performance Scores: Assessing the Psychometric Characteristics of a Combined Score Derived from Clinical Evaluation Forms and OSCEs. Med Educ. 2007;12:10.

19. Swanson DB, Clauser BE, Case SM. Clinical skills assessment with standardized patients in high-stakes tests: a framework for thinking about score precision, equating and security. Adv Health Sci Educ Theory Pract. 1999;4:67-106.

20. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ. 2003;37:830-7.

21. Barman A. Critiques on the Objective Structured Clinical Examination. Ann Acad Med Singapore. 2005;34:478-82.

22. Linn RL, Gronlund NE. Measurement an Assessment in Teaching. 8th ed. New York: Prentice-Hall. 2000.

23. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ. 2004;38:1006-12.

Correspondence:
Gloria E. Sandoval
Department of Pediatrics, Pontificia Universidad Católica de Chile
Lira 85, Santiago Centro, Santiago - Chile
Tel.: +56 (2) 354.3887
Fax: +56 (2) 638.4307
E-mail: gsandova@med.puc.cl