

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

Análise de um sistema de avaliação de aprendizagem para internato em pediatria baseado em exame clínico objetivo estruturado, observação de prática clínica e exame escrito

Gloria E. Sandoval¹, Patricia M. Valenzuela¹, Marcela M. Monge¹, Paulina A. Toso², Ximena C. Triviño³, Ana C. Wright⁴, Enrique Paris⁵, Ignacio Sánchez⁶, Gonzalo S. Valdivia⁷

Resumo

Objetivo: Descrever e analisar três ferramentas usadas no sistema de avaliação aplicado ao internato em pediatria por um período de 7 anos na Faculdade de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile.

Métodos: Pesquisa observacional retrospectiva projetada para as modalidades de avaliação usadas de 2001 a 2007 que foram implementadas no internato em pediatria. As ferramentas foram as seguintes: exame clínico objetivo estruturado (*objective structured clinical examination*, OSCE), exame escrito e diretrizes de observação de prática clínica diária (DOPCD). Os métodos de avaliação foram aplicados ao internato em pediatria do sexto ano, com um total de 697 alunos. A análise estatística incluiu uma avaliação descritiva, com correlação e regressões linear simples e múltipla (ANOVA), teste de Bonferroni e coeficiente alfa de Cronbach. O nível de significância foi de 5%.

Resultados: Escores de sucesso com a OSCE foram atingidos em 75,7±8%, com melhor média entre o sexo feminino ($p < 0,001$). Os escores de OSCE melhoraram após o terceiro ano de implementação. O coeficiente alfa de Cronbach foi de 0,11-0,78. O exame escrito teve escore médio de 79,8±10%, e não houve diferenças quanto a gênero. O escore médio de DOPCD 97,1±3%, com melhores resultados entre as mulheres ($p < 0,005$). A correlação entre os três métodos de avaliação mostrou uma relação positiva moderada, exceto no ano de 2007, quando a correlação foi maior ($p < 0,001$).

Conclusões: Análise do sistema de avaliação da aprendizagem foi realizada usando OSCE, exame escrito e DOPCD, que são complementares entre si, gerando bons resultados.

J Pediatr (Rio J). 2010;86(2):131-136: Internato e residência, faculdade de medicina, competência clínica, mensuração educacional, competência profissional, avaliação de programa.

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, Pontificia 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.

J Pediatr (Rio J). 2010;86(2):131-136: Internship and residency, medical school, clinical competence, educational measurement, professional competence, program evaluation.

1. MD. Assistant Professor, Department of Pediatrics, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.
2. MD. Instructor, Department of Pediatrics, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.
3. MD. Assistant Professor, Center for Medical Education, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.
4. Teacher, Assistant Professor, Center for Medical Education, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.
5. MD. Associate Professor, Department of Pediatrics, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.
6. MD. Professor, Dean Faculty of Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.
7. MD. Associate Professor, Department of Public Health, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.

Este estudo foi realizado na Faculdade de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile.

Não foram declarados conflitos de interesse associados à publicação deste artigo.

Como citar este artigo: Sandoval GE, Valenzuela PM, Monge MM, Toso PA, Triviño XC, Wright AC, et al. Analysis of a learning assessment system for pediatric internship based upon objective structured clinical examination, clinical practice observation and written examination. *J Pediatr (Rio J)*. 2010;86(2):131-136.

Artigo submetido em 02.09.09, aceito em 13.01.10

doi:10.2223/JPED.1986

Introdução

Tradicionalmente, os exames orais são usados para avaliar o domínio cognitivo e as habilidades intelectuais relacionadas a um problema clínico na conclusão de internatos. Entretanto, essa modalidade de avaliação tem limitações em razão de seu baixo grau de confiabilidade e difícil padronização^{1,2}. Por outro lado, o exame escrito (EE), uma ferramenta complementar comumente usada, avalia um espectro maior de informações cognitivas, embora dificilmente possa atingir níveis mais complexos de avaliação, como análise crítica e racionalização³. No entanto, nenhuma das ferramentas permite que o examinado demonstre sua compreensão das competências clínicas, uma vez que não reproduzem as tarefas reais que um médico deve realizar durante um encontro com um paciente⁴.

Por algumas décadas, a introdução de métodos de avaliação que simulam situações clínicas práticas tem fornecido uma solução para a limitação descrita acima. Nessa modalidade de avaliação, uma das ferramentas mais utilizadas é o exame clínico objetivo estruturado (*objective structured clinical examination*, OSCE)⁵, que, desde 1975⁶⁻⁸, tem se mostrado adequado para avaliar competências clínicas^{6,9} e demonstrou ter validade suficiente na interpretação de resultados na aplicação entre estudantes de medicina na graduação e pós-graduação^{4,10-12}. Simulações são aproximações da realidade na tentativa de reproduzir situações clínicas em condições padronizadas, permitindo, através de observação, a avaliação do alcance de objetivos específicos¹³. Elas são amplamente utilizadas para avaliar habilidades de raciocínio clínico, anamnese, exame físico, abordagem diagnóstica, posicionamento dos pacientes e realização de procedimentos, entre outros¹⁴⁻¹⁶.

Outra fonte de informações que permite a avaliação do processo de aprendizagem aplicada aos alunos durante um internato é a diretriz de observação de prática clínica diária (DOPCD)¹⁷. Com uma diretriz pré-definida e estruturada, os professores avaliam, por meio de observação direta, o desempenho dos alunos em diversas atividades clínicas. Esse método tem a vantagem de facilitar a observação do desempenho individual em situações reais, fornecendo mais informações sobre habilidades comunicativas e interpessoais e sobre profissionalismo. Entretanto, ele tem limitações referentes aos examinadores, pelo fato de que precisam ser treinados e, portanto, padronizados¹⁸.

Nenhum dos métodos descritos oferece, por si só, uma avaliação abrangente de todas as competências. Alguns estudos propõem o uso de combinações dos diferentes métodos para harmonizar a avaliação de habilidades cognitivas com a avaliação das complexas habilidades exigidas para uma adequada competência profissional. Kreiter et al.¹⁸ propuseram uma combinação dos escores obtidos na DOPCD e na OSCE para obter uma medida única, o que permitiria uma melhor representação da avaliação das habilidades clínicas do interno¹⁹.

Até 2001, o sistema de avaliação usado no internato em pediatria na Faculdade de Medicina da Pontifícia Universidad Católica de Chile (PUC), Santiago, Chile, era feito através da aplicação de duas ferramentas: DOPCD de cada etapa de

rodízio clínico e um exame oral na conclusão do programa. Naquele ano, o sistema de avaliação sofreu uma reestruturação que envolveu a implementação de um exame escrito para substituir o exame oral, aplicação de uma OSCE na conclusão do internato e manutenção da DOPCD.

O objetivo deste artigo é descrever os resultados da aplicação desse sistema de avaliação e analisar cada ferramenta e as relações entre elas por um período de 7 anos, de 2001 a 2007.

Métodos

Pesquisa observacional retrospectiva projetada para as modalidades de avaliação usadas de 2001 a 2007 e implementadas no internato em pediatria, que ocorre por 12 semanas durante o sexto ano do currículo médico. Cada grupo de internos foi identificado com um número de 1 a 4 de acordo com a sequência temporal correspondente ao longo de cada ano de estudo. A coleta de dados foi aprovada pelo comitê de ética da universidade (09-174).

Descrição das ferramentas de avaliação:

- 1) OSCE: desde 2001, foi aplicada 27 vezes: três vezes no primeiro ano e quatro vezes no ano seguinte. A OSCE foi usada na conclusão do internato em pediatria, em grupos de aproximadamente 25 internos.

Nos 3 primeiros anos, cada OSCE incluiu de 20 a 26 estações. De 2004 em diante, cada OSCE foi implementada em metade dos internos, incluindo 12 a 15 estações, e foi desenvolvida em circuitos sequenciais idênticos, nos quais não havia contato entre os internos. As estações tinham duração de 5 minutos. Cinco das estações envolviam a participação de atores profissionais especializados (APE) treinados, e foram aplicadas duas diretrizes observacionais dicotômicas de comparação sobre o desempenho do interno; uma foi aplicada por professores, e a outra, por um APE, com peso de 80 e 20%, respectivamente.

O total de OSCE consiste de 29 estações, sete das quais tendo uma versão equivalente (versões A, B e C da mesma estação). Os professores encarregados do internato criaram e selecionaram as estações, e também analisaram seus resultados com auxílio técnico em educação.

- 2) EE: consistia de 60 questões de múltipla escolha e integrava pacientes hospitalares (50%), ambulatoriais (25%) e neonatais (25%) em porcentagens proporcionais à duração das rotações clínicas. No total, havia 600 questões com informações sobre as características psicométricas para cada questão individual e com relação ao teste completo, cada uma com suas fontes de validação²⁰.
- 3) DOPCD: essa diretriz, comum a todos os internos da carreira médica, consistia de 10 indicadores baseados em competências clínicas, classificados de acordo com cinco níveis de realização: anamnese, exame físico, hipóteses diagnósticas, plano terapêutico, trabalho em equipe, responsabilidade, habilidades clínicas, tratamento de problemas médicos, hábitos de estudo e uso racional de recursos. Cada interno era avaliado com essa diretriz em cada rotação clínica.

OSCE, EE e DOPCD foram usados como métodos complementares entre si, avaliando todos os objetivos declarados no programa de internato. A nota final foi calculada com base nas notas obtidas: o escore DOPCD recebeu peso de dois terços, e a média entre os escores de EE e OSCE recebeu peso de um terço. Caso o aluno fosse reprovado no exame após uma rotação, devia se matricular novamente na mesma rotação e repeti-la. O desempenho na OSCE ou no EE abaixo de 60% resultava em reprovação no internato.

Na descrição da análise, as notas obtidas pelos internos foram expressas como porcentagem de sucesso e as variáveis consideradas foram tratadas como variáveis categóricas contínuas.

A análise estatística incluiu uma avaliação descritiva, com correlação e regressões linear simples e múltipla (ANOVA) ajustadas para as variáveis selecionadas. Foram usados uma análise de variância (ANOVA) e um teste de Bonferroni para comparar variáveis categóricas. A correlação entre os resultados das ferramentas de avaliação foi calculada usando o coeficiente de correlação de Pearson. O escore de Pearson entre 0,6-0,8 sugere fortes correlações e valores entre 0,4-0,6, ou seja, correlações consideráveis. No entanto, esses pontos de corte estão relacionados ao tamanho amostral. O coeficiente alfa de Cronbach foi calculado para avaliar confiabilidade¹³. Um valor de ponto de corte de 0,4 é o limite inferior sugerido por Barman²¹ para OSCE em graduandos de medicina. Por outro lado, um valor de 0,6 ou mais é proposto por Linn & Gronlund²² como desempenho satisfatório de confiabilidade. Foram usados os programas estatísticos SPSS 11.0 e Stata 8.0. As diferenças foram consideradas estatisticamente significativas quando $p < 0,05$.

Histórico acadêmico relevante incluiu o seguinte:

- a) A média de notas obtidas em pediatria no quinto ano do curso, expressas com uma escala de 1 a 7. A nota mínima em pediatria é 4.

- b) A média ponderada cumulativa, que inclui as notas em cada disciplina feita pelo aluno, do primeiro ao quinto ano do curso.

Resultados

Foram analisados dados de 697 alunos que fizeram internato em pediatria entre 2001 e 2007.

Características gerais e histórico acadêmico

A nota média dos participantes do curso de pediatria foi de 6,5 (Tabela 1). As mulheres tiveram melhor desempenho médio do que os homens ($p = 0,0119$). Não houve diferença significativa nas notas obtidas na conclusão do curso de pediatria entre os alunos que posteriormente seriam incluídos nos grupos de internos ($p = 0,29$). Não houve diferença significativa em PPA entre mulheres ou homens, nem nos diferentes grupos de internos ($p = 0,25$ e $p = 0,2$, respectivamente).

Entretanto, houve diferença significativa nos valores de PPA entre os anos, sendo que 2004 obteve o menor PPA ($p = 0,001$).

Exame clínico estruturado objetivo

A porcentagem média de sucesso para o número de total de internos foi de 75,7% [desvio padrão (DP) = 8,0; variação = 35,8-91,8]. Os escores mínimo e máximo para cada ano variaram entre 76 e 91,8% e entre 35,8 e 64,0%, respectivamente. No geral, as mulheres obtiveram uma média melhor do que os homens, e essa diferença foi estatisticamente significativa (Tabela 2).

A porcentagem média anual de sucesso evidenciou um aumento significativo a partir do terceiro ano de implementação do método de avaliação, para atingir um estado constante durante os últimos 3 anos (teste do qui-quadrado para tendência: $p < 0,001$; Tabela 3).

Tabela 1 - Descrição geral de internos e histórico acadêmico entre 2001 e 2007; nota média final e ponderada e intervalo de confiança de 95%

Ano	Número de internos		Nota média final no quinto ano do curso de pediatria (IC95%)		Média ponderada cumulativa antes do sexto ano do curso (IC95%)	
	Homens, n (%)	Mulheres, n (%)	Homens	Mulheres	Homens	Mulheres
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)*

IC95% = intervalo de confiança de 95%.

* Nota média (IC95%).

Tabela 2 - Resultados de exame clínico objetivo estruturado, exame escrito e diretriz de observação de prática clínica diária de acordo com sexo (média, desvio padrão, mediana, variação), entre 2001 e 2007

	OSCE *		EE [†]		DOPCD [‡]	
	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres
Média (DP)	75,2 (8,2)	77,5 (7,6)	78,5 (10,3)	79,4 (9,6)	96,6 (3,3)	97,4 (2,6)
Mediana (variação)	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)

DOPCD = diretrizes de observação de prática clínica diária; DP = desvio padrão; EE = exame escrito; OSCE = exame clínico objetivo estruturado (*objective structured clinical examination*).

* p = 0,0008.

† p = 0,084.

‡ p = 0,0024.

Tabela 3 - Porcentagem média de sucesso dos participantes em exame clínico objetivo estruturado, exame escrito e diretriz de observação de prática clínica diária por ano

Ano	OSCE, % (DP)	EE, % (DP)	DOPCD, % (DP)
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)
Teste do qui-quadrado para tendência	p < 0,001	p < 0,001	p < 0,001

DOPCD = diretrizes de observação de prática clínica diária; DP = desvio padrão; EE = exame escrito; OSCE = exame clínico objetivo estruturado (*objective structured clinical examination*).

Análise do desempenho em OSCE para cada ano e para cada grupo de internos não mostrou diferenças significativas em nenhum dos anos estudados. Os valores do coeficiente alfa de Cronbach para cada aplicação de OSCE variaram entre 0,11 e 0,78, e sete estações tiveram valor de alfa abaixo de 0,4. Por outro lado, sete OSCE tiveram valores de alfa acima de 0,6. Na análise de impacto de cada estação sobre a OSCE total, sete estações tiveram efeito negativo sobre 21 das 27 OSCE. Quando essas estações foram removidas, foi obtido um valor de alfa maior que 0,4. Não houve diminuição evidente em confiabilidade da OSCE quando os períodos de 2001 a 2003 e de 2004 a 2007 foram comparados; durante esse tempo, a OSCE teve entre 20 e 26 estações ou 12 a 15 estações, respectivamente.

Exame escrito

A taxa média de sucesso para os internos em EE foi de 79,8% (DP = 10; variação = 34,5-100). As porcentagens máxima e mínima variaram de 88,3 a 100% e de 34,5 a 68,3%, respectivamente.

Quanto ao desempenho de acordo com gênero, não houve diferenças estatisticamente significativas (Tabela 2).

O maior desempenho em EE foi documentado em 2006, e os menores valores foram obtidos no primeiro ano de implementação do método de avaliação. A diferença entre esses anos e os outros estudados foi estatisticamente significativa (p < 0,001) (Tabela 3).

Na análise do desempenho por grupo a cada ano, o grupo 4 teve desempenho mais baixo do que os grupos precedentes (F = 9,3; p < 0,00001).

Diretriz de observação de prática clínica diária

A taxa média de sucesso para os internos que seguiram as diretrizes foi de 97,1% (DP = 3; variação = 78,6-100). A porcentagem máxima foi de 100% a cada ano, enquanto a porcentagem mínima variou entre 78,6 e 92,9%. Com essa ferramenta de avaliação, as mulheres tiveram melhor desempenho do que os homens, e a diferença foi estatisticamente significativa (Tabela 2).

Os resultados do teste do qui-quadrado para tendência sugeriram um aumento gradual entre 2001 e 2007, sendo estatisticamente significativo ($p < 0,001$). A média mais baixa de desempenho foi documentada entre 2003 e 2004, e as diferenças entre o desempenho médio nesses anos e o de outros anos foram estatisticamente significativa ($p < 0,001$).

Na comparação de desempenho em DOPCD por grupo de internos a cada ano, não houve diferenças significativas entre os grupos em nenhum dos anos estudados.

Comparações entre as ferramentas de avaliação

As maiores taxas de sucesso foram obtidas na DOPCD, e as menores ocorreram para a OSCE (Tabela 3).

As correlações foram todas positivas e estatisticamente significativas (Tabela 4). Em 5 de 7 anos, houve correlação positiva entre EE e DOPCD; em 4 anos, houve correlação positiva entre EE e OSCE; e, em 5 anos, houve correlação positiva entre DOPCD e OSCE. Cabe observar que, em 3 anos (2003, 2005 e 2007), as três correlações foram positivas e significativas.

Discussão

O desenvolvimento de sistemas de avaliação que consigam integrar diferentes ferramentas para avaliar todos os objetivos propostos em um programa é essencial. O sistema implementado em 2001 para o internato em pediatria atende a esses requisitos. Mais do que uma combinação de ferramentas, esse sistema representa ferramentas complementares entre si para avaliar as habilidades exigidas dos alunos para a prática de pediatria. Dessa forma, em tais condições, é possível avaliar a conformidade a uma das mais relevantes fontes de evidência de validade, para o que a avaliação de conteúdos é abrangente^{9,21,22}.

O próximo passo é a interpretação de resultados após a aplicação de diversas ferramentas. Isso facilita a tomada de decisão sobre continuar ou modificar algumas das ferramentas ou todo o sistema. Além disso, esse passo permite a análise da qualidade da entrada de dados em um registro.

Embora os internos tenham tido os escores mais baixos na OSCE, comparando com as outras ferramentas usadas, os escores médios mínimo e máximo aumentaram após os primeiros 3 anos da implementação dos métodos de avaliação. Mesmo assim, nenhum aluno teve desempenho de 100%. Como resultado dos índices de confiabilidade razoavelmente adequados obtidos para OSCE, o estudo sustenta o uso de 12 a 14 estações^{21,23}.

É importante considerar quatro áreas principais em qualquer avaliação: os alunos, em vista de sua crescente exposição ao método, fornecida por cursos anteriores; a ferramenta, em razão de melhoria nas estações; os professores, devido a seu treinamento e também à experiência ganha nos 7 anos da avaliação de internato; e, finalmente, uma nova abordagem ao ensino.

O método EE também mostrou um aumento gradual na taxa média de sucesso até 2006. Subsequentemente, o exame foi completamente modificado e, como consequência, os resultados diminuíram e retornaram a níveis prévios; portanto, uma contínua renovação de questões e inspeção do conjunto de questões são consideradas cruciais. Uma das fraquezas evidenciadas na análise de resultados foi a impossibilidade de estimar a confiabilidade. A confiabilidade deve ser calculada usando a fórmula 21 de Kuder-Richardson; no entanto, isso não pôde ser atingido, uma vez que as bases de dados não forneciam todas as informações necessárias. Contudo, essa situação foi corrigida como resultado da presente análise.

O alto desempenho dos alunos na DOPCD representa um fato comum, ao qual não fomos exceção. A ferramenta e os examinadores foram considerados questões importantes. A criação de boas diretrizes de avaliação de prática clínica é

Tabela 4 - Coeficiente de correlação de Pearson entre exame clínico objetivo estruturado, exame escrito e diretriz de observação de prática clínica diária por ano

Ano	Correlação EE-DOPCD	Correlação EE-OSCE	Correlação DOPCD-OSCE
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*

DOPCD = diretrizes de observação de prática clínica diária; DP = desvio padrão; EE = exame escrito; OSCE = exame clínico objetivo estruturado (*objective structured clinical examination*).

* $p < 0,001$.

uma tarefa complexa, impondo dificuldades para definir as dimensões, níveis de sucesso, formato (assinatura, comparação) e atribuição de escores. Por outro lado, os examinadores podem sofrer o efeito halo na atribuição de escores, mostrar uma tendência central, benevolência ou rigor, não ser os mesmos que fornecem o ensino ou ter um treinamento fraco. No internato em pediatria, houve uma tentativa de proteger esses últimos aspectos e garantir que somente o corpo docente envolvido no internato pudesse exercer o papel de examinador. Além disso, esses esforços também visaram à provisão de treinamento em educação para os instrutores clínicos. Porém, à luz dos resultados, consideramos que essa última atividade deveria ter como foco o uso de diretrizes, já que tal comportamento permitiria que os professores evitassem erros de avaliação. Entretanto, a modificação das diretrizes para incluir a descrição de competências e desempenhos que devem ser atingidos seria mais relevante. Não foi possível estimar a confiabilidade com o coeficiente alfa de Cronbach. Novamente, tivemos dificuldades na entrada de dados e com as bases de dados. O fato de que essas dificuldades ainda persistem é uma prova de sua importância.

Embora mensure domínios distintos, a correlação significativa entre os resultados das ferramentas de avaliação significa que os melhores alunos obtêm os melhores resultados individuais, independentemente da ferramenta de avaliação usada. Finalmente, uma questão menor é a viabilidade da implementação de um sistema como o descrito acima. Obviamente, esse aspecto deve ser levado em consideração desde o início do projeto, uma vez que os recursos disponíveis para nossas faculdades de medicina sempre são limitados. Esses recursos – humanos, materiais e financeiros – devem ser garantidos para manter a vitalidade do sistema. O internato em pediatria tomou algumas providências para permanentemente manter um corpo docente motivado e pró-ativo para dar suporte técnico ao ensino; além disso, o internato providenciou o uso criterioso de provisões para permitir uma continuidade para atingir os objetivos propostos.

Em conclusão, a análise da experiência de 7 anos com a aplicação desse sistema de avaliação para o internato em pediatria na PUC nos permite afirmar que os três métodos usados em conjunto conseguem avaliar com eficiência as competências pediátricas clínicas exigidas de nossos alunos.

Referências

- Guerin RO. Disadvantages to Using the Oral Examination. In: Marcell EL, Bashook PG, editors. *Assessing Clinical Reasoning: The Oral Examination and Alternative Methods*. Evanston: American Board of Medical Specialties; 1995. p. 41-8.
- Reinhart MA. Advantages to Using the Oral Examination. In: Marcell EL, Bashook PG, editors. *Assessing Clinical Reasoning: The Oral Examination and Alternative Methods*. Evanston: American Board of Medical Specialties; 1995. p. 31-3.
- Martínez JM. Los métodos de evaluación de la competencia profesional: la evaluación clínica objetiva estructurada (ECO). *Educ Méd*. 2005;8:S18-S22.
- Retegui J, Cornel-Avenidaño B. *Mastering the OSCE/CSA*. New York: McGraw-Hill; 1999.
- Association of American Medical Colleges. Division of Medical Education. *Medical School Graduation Questionnaire*. All Schools Summary Report Final; 2007.
- Harden RM, Stevenson M, Downie WW, Wilson GM. *Assessment of clinical competence using objective structured examination*. *Br Med J*. 1975;1:447-51.
- Harden RM, Gleeson FA. *Assessment of clinical competence using an objective structured clinical examination*. *Med Educ*. 1979;13:41-54.
- Barrows S. *An overview of the uses of standardized patients for teaching and evaluating clinical skills*. AAMC. *Acad Med*. 1993;68:451-3.
- 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.
- 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.
- Joorabchi B. *Objective structured clinical examination in a pediatric residency training program*. *Am J Dis Child*. 1991;145:757-62.
- 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.
- Van der Vleuten CP, Swanson DB. *Assessment of clinical skills with standardized patients: state of art*. *Teach Learn Med*. 1990;2:58-76.
- 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.
- Newble D. *Techniques for measuring clinical competence: objective structured clinical examinations*. *Med Educ*. 2004;38:199-203.
- 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.
- 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.
- 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.
- 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.
- Downing SM. *Validity: on meaningful interpretation of assessment data*. *Med Educ*. 2003;37:830-7.
- Barman A. *Critiques on the Objective Structured Clinical Examination*. *Ann Acad Med Singapore*. 2005;34:478-82.
- Linn RL, Gronlund NE. *Measurement an Assessment in Teaching*. 8th ed. New York: Prentice-Hall. 2000.
- Downing SM. *Reliability: on the reproducibility of assessment data*. *Med Educ*. 2004;38:1006-12.

Correspondência:

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