梨花醫大誌 : 第 30 卷 第 1 號 2007
Ewha Med J Vol. 30, No. 1, 2007

의과대학생의 증례 구술 발표시 체크리스트 평가 및 평가자간 차이 비교

이화여자대학교 의학전문대학원 진단검사의학과학교실
허정원 · 이미애 · 정화순

= Abstract =

Evaluation of Checklist and Inter-Rater Agreement in Oral Case Presentation of Undergraduate Medical Students

Jungwon Huh · Miae Lee · Whasoon Chung
Department of Laboratory Medicine, School of Medicine, Ewha Womans University

Background : Undergraduate medical students should learn oral presentation skills, which are central to physician-physician communication. The purpose of this study was to compare checklist scores with global ratings for evaluation of oral case presentation and to investigate interrater agreement in the scoring of checklists.

Methods : The study group included twenty-one teams of undergraduate medical students who did clerkship for 2 weeks in the department of Laboratory Medicine of Mokdong Hospital, School of Medicine, Ewha Womans University from January 2005 to October 2006. Three faculty raters independently evaluated oral case presentations by checklists, composing of 5 items. A consensus scores of global ratings were determined after discusssion. Inter-rater agreement was measured using intraclass correlation coefficient (ICC). As the ICC values approaches 1.0, it means higher inter-rater agreement.

Results : The mean of consensus global ratings was significantly higher than that of checklists by three faculty raters (12.6 ± 1.7 vs 11.1 ± 2.0, $P<0.001$). Spearman's correlation coefficient between global ratings and checklist scores was $r=0.82$ ($P<0.01$). The overall scores of checklist were significantly different among three raters (12.3 ± 2.0, 10.8 ± 2.8, 10.0 ± 2.7, $P<0.05$). ICC values in the scoring of checklists were as follows ; for overall scores, 0.750 ; for individual checklist items, 0.350−0.753.

Conclusions : These results suggest that checklist scores by faculty raters could be one of the most useful tools for evaluation of oral case presentation, if checklist would be modified to make less ambiguous and more objective and faculty raters would have opportunities to be educated and trained for evaluation skills of oral case presentation.

KEY WORDS : Oral case presentation · Undergraduate medical students · Global rating · Checklist · Inter-rater agreement.

서 론

환자 증례의 구술 발표는 의사와 의사간의 의사소통에 있어 중추적인 역할을 하며, 환자의 데이터를 모으고 통합하고 해석하여 추론할 수 있는 기술을 필요로 한다[1)2)]. 따라서 의과대학생은 실습 기간 중 환자 증례의 구술 발표 능력을 습득하는 것이 중요하며, 대부분의 임상과는 실습 과정 중에 환자 증례 구술 발표 기회를 포함하고 있다.

평가는 실제로 측정하고자 하는 학습 목표를 정확히 측정할 수 있어야 하며, 평가자에 따라 평가 결과는 항상 일정해야 한다. 의과대학생의 지식 영역은 학습 목표를 바탕으로 구성된 필기시험을 통하여 객관적으로 평가할 수 있으며, 수치로 표현되는 정량적 평가이므로 평가자마다 차이도 적다. 그러나 구술 발표에 대한 평가는 성취해야 할 구체적인 목표가 확실하지 않을 수 있고, 관찰을 통한 정성적인 평가이므로 평가자마다 변이가 커서 신뢰성에 문제가 있을 수 있다.

본 연구자들은 구술 발표의 전체적인 능력을 상, 중, 하로 나누어 점수를 매기는 포괄적 평가(global ratings) 방법을 이용하여 왔는데, 구술 발표시 습득해야 할 능력을 좀 더 구체적으로 평가하기 위해 체크리스트 평가를 동시에 시행하였다. 본 연구의 목적은 포괄적 평가와 체크리스트 평가 방법을 비교해보고, 체크리스트로 평가한 항목들에 대해 평가자간에 차이가 있는지 알아보고자 하였다.

재료 및 방법

1. 연구 대상 및 방법

연구대상은 2005년 1월부터 2006년 10월까지 이대목동병원에서 진단검사의학 실습을 수행하였던 본과 3학년과 4학년 의과대학생 중, 3명의 교수 평가자가 모두 구술발표 평가를 시행하였던 21개조를 포함하였다. 실습조는 한 조에 4~5명의 학생으로 구성되었고, 실습은 2주 과정으로 진행되었다. 환자 증례는 실습 첫날 교수가 선택하여 학생들에게 주었으며, 실습 마지막 날 실습 조원 중 대표 1명이 구술 발표하였다. 평가자인 교수는 3명으로 구성하였으며, 평가자 A는 교수로써 교육경력이 25년 이상, 평가자 B는 15년 이상, 평가자 C는 5년 이상이었다.

구술 발표에 대한 평가는 2가지 방법으로 시행하였다. 먼저 교수 3명이 각자 독립적으로 체크리스트 항목에 평가하였으며, 평가자들은 각자의 체크리스트 평가 결과를 서로 알지 못하도록 하였다. 체크리스트는 5개 문항으로 구성하였고(Table 1), 각 문항에서 가장 잘한 경우 3점으로 평가하였으며 총 15점이 만점이었다. 체크리스트 평가를 마친 후, 교수 3명이 다시 모여 토론을 통해 종합한 포괄적 평가 점수를 산출 하였다. 즉, 전체적인 구술 발표 능력을 상, 중, 하로 나누고 점수를 매겼는데, '상'인 경우 13~15점, '중'인 경우 10~12점, '하'인 경우 9~11점을 주었다.

2. 통계분석

통계는 SPSS 11.0 통계 패키지를 이용하여 분석하였으며, P값이 0.05 이하인 경우 유의한 것으로 보았다. 포괄적 평가 점수와 체크리스트 점수의 차이는 짝지은 t-test를 이용하여 유의성 검증을 하였고, 상관성을 알아보기 위해 스피어만 순위 상관계수(Spearman rank order correlation coefficient)를 구하였다. 또한 3명 평가자 각각의 체크리스트 산출 점수를 ANNOVA test를 이용하여 비교하였다. 평가자간 차이를 보기위해 신뢰도의 분석방법으로 급간내 상관계수(intraclass correlation coefficient, ICC)를 이용하였으며, 결과 판정 기준은 다른 문헌을 참조하였다(ICC <0.20, poor agreement ; 0.21~0.40, fair agreement ; 0.41~0.60 moderate agreement ; 0.61~0.80 substantial agreement ; >0.80, almost perfect agreement)[3)4)].

Table 1. Checklist scoring form used for assessing a student's oral presentation

평가항목	평 가 영 역	점 수		
1	발표 내용의 문제를 잘 파악하고 이해했으며, 설명이 타당하였다	3	2	1
2	전체적으로 발표의 흐름이 원활하였다.	3	2	1
3	발표 내용은 복잡하지 않고 이해하기 쉽게 표현하였다.	3	2	1
4	발표자 목소리 크기, 속도, 발음은 적절하였다.	3	2	1
5	발표를 열심히 성실하게 준비하였다.	3	2	1

결 과

1. 체크리스트 평가와 포괄적 평가 비교

3명의 평가자가 체크리스트를 기반으로 산출한 점수의 평균은 토론후 산출한 포괄적 평가 점수보다 낮았다(Table 2). 포괄적 평가 점수와 체크리스트 점수간에 상관성은 r=0.82(P<0.01)이었다(Fig. 1).

2. 평가자간 체크리스트 평가 비교

체크리스트의 총 점수는 평가자 A, B, C에 따라 차이가 있었다. 평가자 A의 점수가 가장 높았으며, 평가자 A와 C사이에 총 점수는 유의한 차이가 있었다(Table 3). 체크리스트 총 점수의 평가자간 일치도를 보기 위한 급간내 상관계수는 0.750이었고, 체크리스트 각 항목별로는 급간내 상관계수가 0.350~0.753 범위였다(Table 4).

Table 2. Comparison of scores between global ratings and checklists

	Scores (mean±SD)	*P* value
Global*	12.6±1.7	0.00
Checklist	11.1±2.0	

* : consensus scores of general competence after discussion of three faculty raters

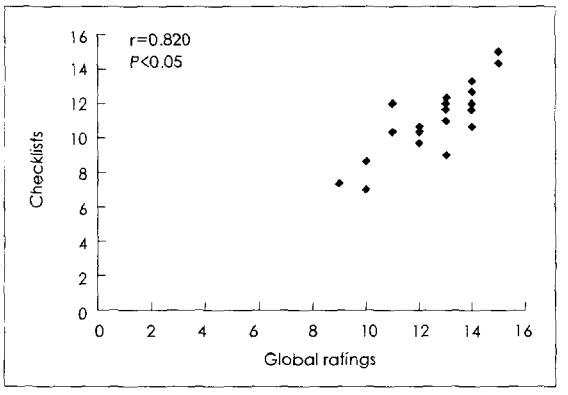

Fig. 1. Correlation between global ratings and checklist scores.

고 찰

평가 결과는 평가자간 차이, 관찰방법, 평가도구, 평가에 사용된 임상적 상황등에 의해 달라질 수 있으며, 평가의 신뢰성에 영향을 미친다. 따라서 평가결과에 영향을 미칠 수 있는 변수를 줄이기 위해 평가의 구체적 기준을 마련하고 이에 따라 객관적으로 평가하는 것이 필요하다.

포괄적 평가는 전공의나 의과대학생들의 임상 수행 능력을 평가하는데 유용하게 이용되어 왔다[5-7]. 이 방법은 구체적 체크리스트 항목 없이 전체적인 능력에 대한 평가로 이루어지므로, 특정 영역에서 탁월 또는 부족한 점들이 전체적인 평가에 영향을 미치고 편견이 개입될 수 있는 단점이 있다. 또한 본 연구자들이 시행했던 방법과 같이 평가자들이 토론후 포괄적 평가를 시행하는 것도 어느 평가자 한사람의 의견만 강하게 반영될 수 있는 문제점이 있을 수 있다. 따라서 저자들은 구술 발

Table 4. Inter-rater agreement of checklist scores

Checklist*	ICC	95% CI
Item 1	0.753	0.488~0.892
Item 2	0.599	0.171~0.825
Item 3	0.350	−0.344~0.716
Item 4	0.706	0.392~0.872
Item 5	0.383	−0.276~0.730
Overall	0.750	0.482~0.891

* : See Table 1
ICC : intraclass correlation coefficient, CI : confidence interval

Table 3. Comparison of checklist scores among three faculty raters

Checklist*	Scores (mean±SD)			*P* value
	Rater A	Rater B	Rater C	
Item 1	2.3±0.7	2.1±0.7	1.7±0.6	0.014
Item 2	2.4±0.7	2.1±0.7	2.2±0.5	0.224
Item 3	2.6±0.6	2.3±0.6	2.2±0.5	0.100
Item 4	2.6±0.5	2.2±0.6	2.1±0.6	0.031
Item 5	2.4±0.5	2.1±0.7	1.8±0.9	0.039
Overall	12.3±2.0	10.8±2.8	10.0±2.7	0.017

* : See Table 1

표시 습득해야 할 능력을 좀 더 구체적으로 평가할 수 있는 체크리스트를 만들어서 기존에 평가하던 포괄적 평가 방법과 비교하였다. 본 연구 결과 포괄적 평가 점수가 체크리스트 점수보다 유의하게 높았다(12.6±1.7 vs 11.1±2.0, P<0.001). 다른 연구 결과에 따르면, 표준화 환자를 이용한 수행 평가에서 포괄적 평가 점수가 체크리스트 평가 점수보다 더 높았는데[6], 상관성은 중등도(moderate) 정도로 보고하였다. 반면, 마취과 전공의들의 수행능력을 평가한 문헌에서는 포괄적 평가점수보다 체크리스트 평가 점수가 더 높았다[7]. 본 연구에서 포괄적 평가와 체크리스트 평가 점수는 유의한 차이가 있었으나 상관성은 높았다(r=0.82, P<0.05). 학술대회의 초록을 평가한 연구에서도 포괄적인 평가 점수와 체크리스트 점수간에 연관성이 높았다고 보고하였다[3].

평가자간 일치율을 높이는 것은 평가의 신뢰도에 있어 필수적인 사항이다. 한 문헌에 따르면 구술 발표 평가시 평가자간 점수에 차이가 있었으며[1], 전공의의 구술 시험 평가에서도 평가자간 일치율이 낮았다[8]. 본 연구 결과 평가자간 체크리스트 점수는 유의한 차이가 있었으며(P=0.017), 교육 경력이 가장 길었던 평가자 A의 점수가 가장 높았고, 교육 경력이 가장 짧은 평가자 C의 점수가 가장 낮았다. 그러나 체크리스트 총 점수의 평가자간 급간내 상관계수는 0.750으로써 높은 일치율(substantial agreement)을 보였다.

본 연구에서는 포괄적 평가를 평가자마다 시행하지 않았고, 체크리스트 평가 후 다시 모여 토론을 통해 평가자 3명이 종합한 포괄적 평가를 하였으므로, 포괄적 평가와 체크리스트 평가에서 평가자간 차이를 비교할 수 없었다. 다른 문헌에서는 표준화 환자를 이용하는 임상 수행 능력 시험에 있어서 포괄적 평가보다는 체크리스트 평가가 신뢰성과 효율성이 더 높다고 보고하였다[9][10]. 이와 같이 체크리스트 평가는 포괄적 평가보다 구체적이고 객관화 되어 있으므로 평가자간 신뢰성이 증가할 것으로 기대할 수 있는데, 다른 연구에서는 포괄적 평가가 체크리스트 평가보다 평가자간 신뢰도가 더 높았다고 보고하였다[11][12]. 표준화 환자가 학생들을 직접 평가했던 연구에서도 포괄적 평가가 더 신뢰성 있는 결과를 보여주었다[13]. 위의 결과들을 종합할 때 포괄적 평가와 체크리스트 평가 중 어느 방법이 구술 발표 능력을 평가하기에 더 적절한 지 결정하기 어렵다. 그러나 학생과 평가자에게 성취해야 할 목표를 더 분명하게 인식시킬 수 있고, 학생들의 구술 발표 능력 중 어느 부분이 취약하고 어느 부분이 강점인지 되먹임을 구체적으로 하기 위해서는 체크리스트 평가 방법이 더 유용할 것으로 생각된다.

평가자간 차이를 유발시키는 원인 중 하나는, 특정 상황을 관찰하는 평가자의 능력에 차이가 있을 수 있고, 이로 인해 평가 점수에 영향을 미칠 수 있을 것이다. 한 문헌에서는 포괄적 평가 또는 체크리스트 평가 방법에 상관없이 평가자간 일치율은 낮았다고 보고하였다[7][14]. 이는 여러 평가자가 같은 사물을 관찰해도 보는 관점이 다를 수 있어 객관적 평가 기준을 가지고 평가를 시행해도 평가자간 차이를 보일 수 있음을 시사한다. 본 연구에서도 체크리스트의 총 점수는 평가자간에 일치율이 높았으나(ICC, 0.750), 체크리스트 각 항목마다 일치율은 차이가 있었다(ICC, 0.350~0.753). 다른 문헌에서도 체크리스트 전체 점수는 평가자간 일치율이 높은데 반해, 체크리스트 각 항목마다는 평가자간에 일치율이 낮았다[3]. 이는 전체적인 최종 평가는 평가자간에 신뢰성이 있지만, 이를 판단하는 평가 기준은 다를 수 있다는 사실을 뒷받침한다. 본 연구에서 발표내용의 이해도와 설명의 타당성을 평가하는 체크리스트 1 항목이 평가자간 일치율이 가장 높았다. 반면, 체크리스트 항목 3과 5의 일치율이 낮았다. 항목 3의 경우 발표내용의 간결성과 표현 기술을 측정하는 항목이었고, 항목 5의 경우는 발표 준비의 성실성을 평가하는 항목이었다. 체크리스트 항목 5의 경우 주관적인 평가가 될 수 있는 항목으로 평가자간 일치율이 낮을 것으로 예측하였다. 항목 3의 경우는 다른 항목에 비해 비교적 객관적으로 평가할 수 있는 항목으로 생각되었으나, 오히려 평가자간 일치율이 가장 낮았다. 이는 각 평가자가 보는 관점이 다르고, 중요시 하는 관점에 따라 나머지 평가 항목들도 영향을 받았을 가능성이 있다. 본 연구의 평가자들은 구술 평가에 대한 특별한 교육과 훈련을 받은 경험이 없었는데, 교육경력이 가장 긴 평가자 A와 교육경력이 가장 짧은 평가자 C간에 차이가 많았다. 한 문헌에서 평가자 집단에 경험이 없는 평가자가 추가로 포함된 경우 평가자간 일치율이 감소하였다고 보고하였다[3]. 따라서 평가자간 차이를 줄이기 위해서는 평가자들에게 구술 발표를 평가하는데 필요한 교육과 훈련이 이루어져야 할 것이다.

또한 체크리스트 항목 중 주관적인 요소가 포함된 항목과 문장이 애매하거나 구체적이지 못한 항목은 평가

자간에 차이가 많았고[15], 체크리스트 항목을 객관적인 문장으로 수정한 후 평가자간 일치도가 높아졌다고 보고하였다[3]. 따라서 본 연구자들의 체크리스트를 더욱 명확하고 구체적인 문장으로 수정 보완하면 평가자간 항목별 일치도를 높일 수 있을 것으로 기대된다.

또한 평가자간 불일치의 원인 중 하나는 여러번의 평가가 아니라 1회 증례 발표만으로 평가하여 학생들의 구술 발표 능력을 제대로 반영하지 못했을 수도 있다. 문헌을 살펴보면 평가자간 일치도는 학생들을 대상으로 했을 때보다 전공의를 대상으로 평가했을때 평가자간 일치도가 더 높았다[16-20]. 이는 학생들에 비해 전공의는 교수가 관찰할 수 있는 기회가 더 많아 평가자간 차이가 감소되었을 가능성이 있다. 다른 연구에서도 전공의들이 임상 수행 능력을 다수의 평가자가 여러번 측정한 경우 평가자간 차이가 감소되었다고 하였다[21)22)]. 본 연구자의 과목같이 2주의 짧은 실습 기간인 경우 위와 같은 방법으로 평가자간 차이를 감소시키기는 어려우나, 실습기간이 긴 과목의 경우는 여러명의 평가자가 각각 여러번 평가할 경우 평가자간 차이를 감소시키는데 도움이 될 것으로 생각된다.

결론적으로 의과대학생 구술 발표 평가를 위한 체크리스트 평가는 포괄적 평가와 상관성이 높았고, 평가자간 일치율도 높았다. 그러나 체크리스트 각 문항에서는 평가자간 일치율이 낮았던 항목이 있었다. 따라서 체크리스트 항목을 좀 더 명확하고 구체적으로 수정 보완하고, 평가자들의 구술 평가 기술에 대한 교육과 훈련이 이루어진다면, 체크리스트 평가는 더욱 유용한 평가 방법이 될 것이다.

요 약

배 경 :

환자 증례의 구술 발표는 의사와 의사간의 의사소통에 있어 중추적인 역할을 하며, 의과대학생은 실습 기간 중 환자 증례의 구술 발표 능력을 습득하는 것이 중요하다. 본 연구의 목적은 구술 발표에 대한 포괄적 평가와 체크리스트 평가 방법을 비교해보고, 체크리스트로 평가한 항목들에 대해 평가자간 차이가 있는지 알아보고자 하였다.

재료 및 방법 :

2005년 1월부터 2006년 10월까지 이대목동병원에서 진단검사의학 실습을 수행하였던 21개조를 연구대상에 포함하였다. 먼저 교수 3명이 5개 문항으로 구성된 체크리스트 항목을 각자 독립적으로 평가하였고, 다시 모여 토론 후 포괄적 평가를 하였다. 평가자간 차이를 보기 위해 신뢰도의 분석방법으로 급간내 상관계수(intraclass correlation coefficient)를 이용하였고, 수치가 1에 가까울 수록 일치도가 높은 것으로 해석하였다.

결 과 :

평가자가 토론후 산출한 포괄적 평가 점수가 체크리스트를 기반으로 산출한 점수의 평균에 비해 더 높았으며(12.6±1.7 vs 11.1±2.0, $P<0.001$), 두 평가 점수간에 상관계수는 r=0.82($P<0.01$)이었다. 체크리스트의 총 점수는 3명의 평가자에 따라 유의한 차이가 있었다(12.3±2.0, 10.8±2.8, 10.0±2.7, $P<0.05$). 급간내 상관계수는 체크리스트 총 점수의 경우 0.750(0.482~0.891)이었고, 체크리스트 5개 각 항목의 급간내 상관계수는 0.350~0.753 범위였다.

결 론 :

결론적으로 의과대학생 구술 발표 평가를 위한 체크리스트 평가는 포괄적 평가와 상관성이 높았고, 평가자간 일치율도 높았다. 그러나 체크리스트 각 문항 평가에서는 평가자간 일치율이 낮았던 항목이 있었다. 따라서 체크리스트 항목을 좀 더 명확하고 구체적으로 수정 보완하고, 평가자들의 구술 평가 기술에 대한 교육과 훈련이 이루어진다면, 체크리스트 평가는 더욱 유용한 평가 방법이 될 것이다.

중심 단어 : 구술발표 · 체크리스트.

References

1) Kim S, Kogan JR, Bellini LM, Shea JA : *A randomized-controlled study of encounter cards to improve oral case presentation skills of medical students. J Gen Intern Med 2005 ; 20 : 743-747*

2) Haber RJ, Lingard LA : *Learning oral presentation skills : a rhetorical analysis with pedagogical and professional implications. J Gen Intern Med 2001 ; 16 : 308- 314*

3) Rowe BH, Strome TL, Spooner C, Blitz S, Grafstein E, Worster A : *Reviewer agreement trends from four years of electronic submissions of conference abstract. BMC Med Res Methodol 2006 ; 6 : 14-19*

4) Bhandari M, Templeman D, Tornetta P 3rd : *Interrater reliability in grading abstracts for the orthopaedic trauma association. Clin Orthop Relat Res 2004 ; 423 : 217-221*
5) Daelmans HE, van der Hem-Stokroos HH, Hoogenboom RJ, Scherpbier AJ, Stehouwer CD, van der Vleuten CP : *Global clinical performance rating, reliability and validity in an undergraduate clerkship. Neth J Med 2005 ; 63 : 279-284*
6) Swartz MH, Colliver JA, Bardes CL, Charon R, Fried ED, Moroff S : *Global ratings of videotaped performance versus global ratings of actions recorded on checklists : a criterion for performance assessment with standardized patients. Acad Med 1999 ; 74 : 1028-1032*
7) Ringsted C, Ostergaard D, Ravn L, Pedersen JA, Berlac PA, van der Vleuten CP : *A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. Med Teach 2003 ; 25 : 654-658*
8) Jacobsohn E, Klock PA, Avidan M : *Poor inter-rater reliability on mock anesthesia oral examinations. Can J Anaesth 2006 ; 53 : 659-668*
9) Han JJ, Kreiter CD, Park H, Ferguson KJ : *An experimental comparison of rater performance on an SP-based clinical skills exam. Teach Learn Med 2006 ; 18 : 304-309*
10) Morgan PJ, Cleave-Hogg D, Guest CB : *A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. Acad Med 2001 ; 76 : 1053-1055*
11) Regehr G, MacRae H, Reznick RK, Szalay D : *Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998 ; 73 : 993-997*
12) Wilkinson TJ, Newble DI, Frampton CM : *Standard setting in an objective structured clinical examination : use of global ratings of borderline performance to determine the passing score. Med Educ 2001 ; 35 : 1043-1049*
13) Cohen DS, Colliver JA, Marcy MS, Fried ED, Swartz MH : *Psychometric properties of a standardized-patient checklist and rating-scale form used to assess interpersonal and communication skills. Acad Med 1996 ; 71 (1 Suppl) : S87-S89*
14) LaMantia J, Rennie W, Risucci DA, Cydulka R, Spillane L, Graff L, et al : *Interobserver variability among faculty in evaluations of residents' clinical skills. Acad Emerg Med 1999 ; 6 : 38-44*
15) Montgomery AA, Graham A, Evans PH, Fahey T : *Inter-rater agreement in the scoring of abstracts submitted to a primary care research conference. BMC Health Serv Res 2002 ; 2 : 8-11*
16) Haber RJ, Avins AL : *Do ratings on the American Board of Internal Medicine Resident Evaluation Form detect differences in clinical competence? J Gen Intern Med 1994 ; 9 : 140-145*
17) Maxim BR, Dielman TE : *Dimensionality, internal consistency and interrater reliability of clinical performance ratings. Med Educ 1987 ; 21 : 130-137*
18) Kwolek CJ, Donnelly MB, Sloan DA, Birrell SN, Strodel WE, Schwartz RW : *Ward evaluations : should they be abandoned? J Surg Res 1997 ; 69 : 1-6*
19) Davis JD : *Comparison of faculty, peer, self, and nurse assessment of obstetrics and gynecology residents. Obstet Gynecol 2002 ; 99 : 647-651*
20) Durning SJ, Cation LJ, Jackson JL : *The reliability and validity of the American Board of Internal Medicine Monthly Evaluation Form. Acad Med 2003 ; 78 : 1175-1182*
21) Durning SJ, Cation LJ, Markert RJ, Pangaro LN : *Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med 2002 ; 77 : 900-9004*
22) Norcini JJ, Blank LL, Arnold GK, Kimball HR : *Examiner differences in the mini-CEX. Adv Health Sci Educ Theory Pract 1997 ; 2 : 27-33*