

외과학 임상실습 평가방법의 변경이 실습성적에 미치는 영향

연세대학교 원주의과대학 외과학교실

김명수 · 김익용 · 배금석 · 한애리 · 김대성 · 유수영 · 강성준 · 노병선

= Abstract =

Effect after Change of Evaluation Method in Surgical Clerkship Program

Myoung Soo Kim, MD, Ik Yong Kim, MD, Keum Suck Bae, MD, Airi Han, MD,
Dae Sung Kim, MD, Soo Young Yoo, MD, Seong-Joon Kang, MD, Byung Sun Rhoe, MD

Department of Surgery, Yonsei University Wonju College of Medicine, Wonju, Korea

Purpose: Since 2002, Department of Surgery, Yonsei University Wonju College of Medicine attempted to revise the evaluation method of surgical clerkship program. The purpose of this study was to analyze the effect after change of evaluation method.

Methods: Major changes in current revisions were summarized as follows; (1) multiple evaluator, (2) student's attitude evaluated by rating scale method, (3) attendance and procedure evaluation were measured in numbers according to the fixed criteria, (4) addition of clinic-based items such as problem-solving items in the final written examination. We compared the measurement of revised evaluation at 2002 with those at 1999. And the correlation or simple linear regression analysis between score of clerkship and student academic grade score were taken.

Results: The clerkship score of students at 2002 showed more powerful discrimination between high performance group and poor group than that of 1999's. Even though the subject score of clerkship was significantly correlated with grade score regardless of evaluation method of clerkship, the subject score of clerkship at 2002 was more closely correlated with a grade score than 1999's. The score of knowledge-based items from final written examination is not correlated with clerkship score, but on the other hands, there was significant correlation between score of clinic-based items and clerkship score.

Conclusion: In conclusion, variable measurements under the exact guideline and principle are more reliable method in evaluation of surgical clerkship.

Key Words: Surgical clerkship program, Evaluation, Measurement

교신저자: 김명수, 연세대학교 원주의과대학 외과학교실
강원도 원주시 일산동 162번지
Tel: 033)741-1306 / 0573, Fax: 033)742-1815
E-mail: ysms91@wonju.yonsei.ac.kr

서 론

의학교육은 크게 기초의학과 임상의학으로 나눌 수 있으며, 임상의학은 다시 이론교육과 실습교육으로 나눌 수 있다. 최근 의학교육의 경향은 조기 임상교육과 교과내용의 임상화 등으로 요약될 수 있다. 즉 교육과정중 임상실습에 많은 시간을 할애하여 지식 위주의 강의보다는 실습 위주의 교육을 통하여 의사로서 갖추어야 할 태도, 가치관 및 수기를 연마케 하려는 것이다(김영명, 1990). 이론교육은 필기시험을 통하여 비교적 객관적으로 평가될 수 있는 반면, 임상실습교육은 평가의 객관성을 확보하기가 어려우며 평가에 대한 검증이 쉽게 이루어 질 수 없는 단점이 있다. 실제로 장성호 등(1993)이 보고한 임상실습에 관한 설문조사를 살펴보면, 임상실습 평가는 그 평가기준이 없으며 부적절하다고 판단한 응답자가 적지 않다고 보고되고 있다. 따라서 최근에는 피교육자인 학생들의 평가에 근거하여 실습프로그램에 대한 수정이 시행되고 있으나(하은희, 1996) 이러한 프로그램의 개선이 실습평가결과에 어떠한 영향을 미치는지에 대하여서는 보고는 적다.

연세대학교 원주의과대학에서는 의학과 5학년(본과 3학년)에 주과목을 위한 임상실습이 2학기에 걸쳐서 시행되고 있으며, 이 중 외과학실습은 총 6주간 시행되고 있다. 과거의 외과학 실습교육은 학생 담당 수석전공의와 담당교수가 6주간의 출석, 태도, 수기 등의 항목을 항목점검표(checklist) 방법으로 평가한 실습점수와 학기말에 시행하는 필기시험 성적으로 평가되어 왔다. 그러나 실습점수의 점수화 과정에서 평가자의 주관적인 기준에 의하여 평가점수에 차등이 발생할 위험성이 있으며, 반대로 평가자가 학생의 실습교육 성취도와 상관없이 일률적인 점수를 부여해서 변별력이 낮아지는 위험성도 발생할 수 있었다. 이에 외과학교실에서는 2000년부터 임상실습평가에 대하여 계량화 작업을 포함한 개선 작업을 지속하여 왔으며, 본 연구에서는 이러한 평가방법의 변경이 외과학실습 평가결과에 어떠한 영향을 미쳤는지 확인하고자 한다.

대상 및 방법

1999년 이후부터 2002년까지의 외과학실습 평가는 크게 실습점수와 필기시험점수로 대별된다. 필기시험은 객관식문항에 의한 평가로 학년말에 일괄적으로 시행하였으며, 실습점수는 항목점검표(checklist) 방법에 의한 실습기간중의 수시평가이었다. 본 연구에서는 2002년 외과학 실습평가결과를 1999년 외과 실습평가결과와 비교하였는데 이는 2000년에 발생한 의료계 파업상태로 인하여 2000년과 2001년 학사일정이 비정상적으로 진행되었기 때문에 해당년도의 성적을 비교대상으로 설정하기가 어려웠기 때문이었다.

가. 외과학실습 평가방법의 변경내용

1999년과 비교한 2002년 외과학 실습평가방법의 특징은 (1) 태도 및 기본적인 자세에 대한 평가를 다수의 평가자가 시행하며, 평가점수는 2점 간격의 비율척도(rating scale)방법으로 평가하였으며, (2) 수기 및 출석에 대한 항목점검표방법(checklist method)으로 확인된 항목의 숫자에 따라서 차등점수를 부여하였으며, (3) 임상 예에 대한 증례발표(case report)를 신설하고 이를 점수화 하였으며, (4) 필기시험 시험문제에 임상실습 위주의 실습형 문제(자료제시형 문제도 포함됨)를 추가하였다. 구체적인 평가내용의 변동사항은 Table I에 정리하였다.

나. 연구대상

2002년 외과학실습을 수강한 107명의 5학년 학생을 연구군으로 하였으며, 1999년 외과학실습을 수강한 102명을 대조군으로 설정하였다. 연구군과 대조군에 대하여 외과학실습 과목성적 및 항목별 점수(1999년에는 실습점수와 필기시험점수로, 2002년에는 실습점수, 필기시험점수는 물론 실습점수에 대한 세부항목점수와 필기시험점수 중 실습형 문제점수와 지식형 문제점수로 세분하였다)를 수집하였으며, 해당년도의 학년성적을 4.0점 만점의 평균평량으로 구하였다.

Table I. Scoring System of Surgical Clerkship; Comparison between Year 2002 and Year 1999

	Year 1999		Year 2002	
	Points	Scoring method	Points	Scoring method
Clerkship score (A)	100		100	
Attitude	20	Single evaluator No criteria for scoring	30	5 points * 6 weeks Multiple evaluator Rating scale method: 1-3-5
Attendance	40	←	30	Checklist method In proportion to attendance rate
Procedure checklist	20	No criteria for scoring	10	Checklist method Among 40 procedures, more 30: 10, 20-29: 8, 10-19: 5, less 10: 2 points
Case presentation	-		10	5 points * 2 times Rating scale method: 1-3-5
Mini examination	20	No learning objectives	20	Checklist method Limited 106 topics of learning objectives
Written examination score (B)	100	Knowledge based items	100	Clinic based items* Knowledge base items
Total score of clerkship	100	A : B = 3 : 7	100	A : B = 4 : 6

* Clinic-based items contains problem-solving items which were provided with figures or radiology films or laboratory results to simulate a real clinical setting

다. 외과학실습 평가의 비교방법 및 통계방법

각 점수는 100점 만점으로 환산하였으며, 기술통계량은 평균, 표준편차, 최고점수, 최하점수, 범위(range), 왜도(skewness), 첨도(kurtosis) 등을 표시하였다. 연구군과 대조군간 성적의 차이가 있는지는 SPSS 10.0을 이용한 Student's t-test로 검정하였다. 각 성적간의 상관관계를 확인하고자 단순상관분석을 SPSS 10.0을 이용하여 시행하였으며, 성적간의 상관도는 Pearson's 상관계수(correlation coefficients)로 표시하였다. 당년도 학년성적을 학생의 총체적인 학습 성취도라 가정하고 외과학실습 항목별 점수가 단독 혹은 병합하여 학년성적과 어떠한 상관관계가 있는지를 확인하기 위하여, 단순회귀분석과 다중회귀분석(입력방법, enter method)을 SPSS 10.0으로 시행하였으며, 유의수준이 0.05 이하인 항목별 점수

가 학년성적과 상관관계가 있는 것으로 검정하였다.

결 과

가. 기술통계량 및 점수분포의 차이; 1999년 성적과 2002년 성적의 비교

연구군(2002년)의 학년성적은 2.710±0.610으로 대조군(1999년)의 2.818±0.670과 통계학적으로 차이가 없었다. 마찬가지로 필기시험점수에 있어서도 두 군간에는 유의한 차이가 없었다. 그러나 실습점수는 1999년에는 평균 86.843±2.838, 2002년에는 88.187±4.464로 두 군간의 유의한 차이가 있었다(p=0.01)(Table II). 특히 2002년 실습점수는 1999년에 비하여 범위(range)가 증가하여 상·하위권간의 차이가 분명해 졌으며, 동점자가 일부점수대에 몰려 있는 현상도 완화되었다(Fig. 1).

Table II. Statistical Data of Clerkship Score; Comparison Between Year 2002 Score and Year 1999 Score

		Clerkship score (100)	Written examination score (100)	Total score of Clerkship (100)	Grade score (4.0)
Year	N	102	102	102	102
1999	Mean	86.843	68.402	73.934	2.710
	S.D.*	2.838	7.270	5.278	.610
	Maximum	94.00	84.00	85.50	3.97
	Minimum	82.00	48.00	61.00	1.53
	Range	12.00	36.00	24.50	2.44
	Skewness	.576	-.349	-.272	.300
	Kurtosis	.233	.073	-.093	-.887
Year	N	107	107	107	107
1999	Mean	88.187	69.284	76.845	2.818
	S.D.*	4.464	6.796	4.659	.670
	Maximum	98.00	83.33	86.70	4.16
	Minimum	77.00	37.50	56.58	.68
	Range	21.00	45.83	30.12	3.48
	Skewness	-.055	-1.018	-.815	-.354
	Kurtosis	-.215	3.437	2.298	-.007
	p-value [†]	.010	.366	<.001	.266

* S.D.: Standard Deviation

[†] p-value was calculated by Students t-test.

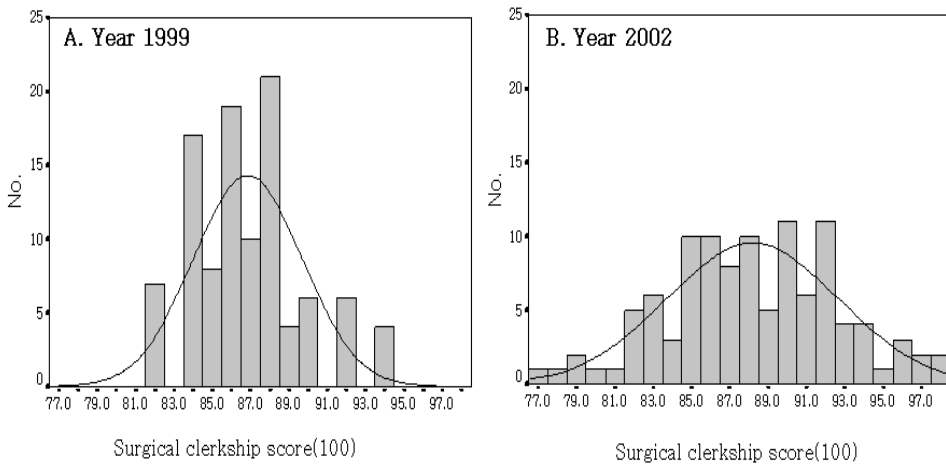


Fig. 1. Distribution of clerkship score; comparison between year 2002 score (B) and year 1999 score (A).

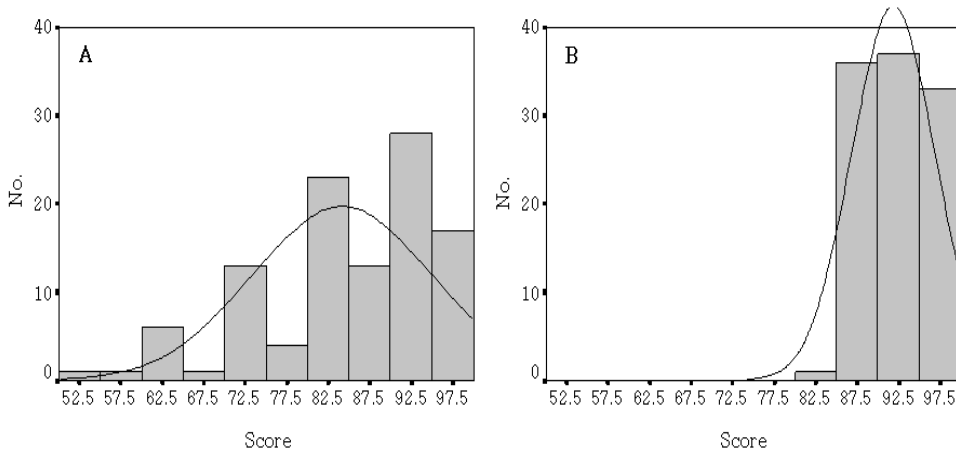


Fig. 2. Score distribution of year 2002 clerkship score evaluated by rating scale method(A) and checklist method(B).

Table III. Statistical Data of Clerkship Score* by Subclass, Year 2002

Subclass	Scoring method	Mean	S.D.	Max	Min	Range	Skew-ness	Kurtosis
Clerkship score (A)	Score 1:2:3 = 30:10:60	88.19	4.46	98.00	77.00	21.00	-.055	-.215
Score 1 ; Attitude	Rating scale method	81.25	11.15	100.00	53.33	46.67	-.450	-.457
Score 2 ; Case presentation	Rating scale method	87.10	16.99	100.00	40.00	60.00	-1.228	.778
Score 3 ; Skill/ Attendance/ Mini examination	Checklist method	91.84	5.03	100.00	80.00	20.00	.508	-.745
Written examination Score (B)	B1:B2 =50:70	69.29	6.80	83.33	37.50	45.83	-1.018	3.437
Clinic based items (B1)	MCQ [†]	68.28	8.07	84.00	42.00	42.00	-.533	.300
Knowledge based items (B2)	MCQ [†]	70.00	7.81	85.71	34.29	51.43	-.874	3.054
Total score of clerkship	A:B = 40:60	76.84	4.66	86.70	56.58	30.12	-.815	2.298

* Full score of all presented score is 100

† MCQ: Multiple Choice Question

나. 2002년 실습성적의 세부분석

항목점검표방법으로 평가된 행위 (procedure), 출

석 및 지식분야에 관한 점수는 학생간의 큰 차이를 보이지 않아 최고점과 최하점의 차이가 20점에 불과하였다 (Table III). 반면에 비율척도법으로 평가된

Table IV. Correlation Analysis Among the Clerkship Score, Written Examination score, Total Score of Clerkship and Grade Score; Comparison between Year 1999 and Year 2002

Subclass	Subclass	Clerkship score	Written exam. score	Total score of clerkship	Grade score
Clerkship score	γ *	1.000	.130	.497	.328
	p-value	1.000	.183	.000	.001
Written exam. score	γ *	.143	1.000	.925	.785
	p-value	.152	1.000	.000	.000
Total score of clerkship	γ *	.299	.987	1.000	.813
	p-value	.002	.000	1.000	.000
Grade score	γ *	.315	.717	.742	1.000
	p-value	.001	.000	.000	1.000

Upper diagonal data (with shadowed background) mean data of year 2002 and lower diagonal data mean data of year 1999.

* γ means Pearson correlation coefficients

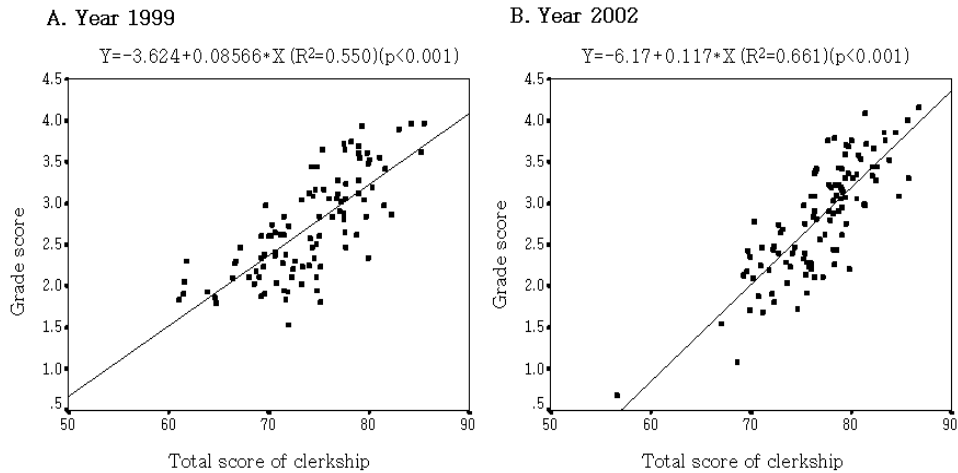


Fig. 3. Simple linear regression analysis between grade score and total score of clerkship at year 1999(A) and year 2002(B).

전반적인 태도나 증례발표의 경우에는 최고점과 최하점간의 차이가 각각 46.67점과 60점으로 상하위권간의 차이가 분명하였다. 마찬가지로 평가측정방법에 따른 실습점수 분포를 비교하여 보면, 비율적도방법에 의하여 측정된 세부점수가 보다 학생들의 우열을 구분하기가 용이하였다 (Fig. 2).

다. 외과학실습 과목성적과 학년성적간의 상관관계

단순상관분석에서 2002년은 물론 1999년에서도 외과학실습 과목성적은 학년성적과 유의한 상관관계를 가졌다. 1999년 외과학실습 과목성적의 학년성적과의 상관계수는 0.742인 반면 2002년에는 상관계수가 0.813으로 증가하여, 2002년 외과학실습

과목성적이 보다 학년성적에 근접한 상관관계를 가졌다 (Table IV). 산점도 (scattergram)와 단순회귀분석에서도 2002년 외과학실습 과목성적은 1999년 보다 학년성적에 근접한 상관관계를 가짐을 알 수 있다 (Fig. 3).

라. 실습점수 및 필기시험점수와 학년성적과의 상관관계

1999년과 2002년 모두에서 외과학실습 과목성적을 구성하는 실습점수와 필기시험점수간에는 유의한

상관관계를 발견할 수 없었다 (Fig. 4). 단순상관분석에서도 1999년 실습점수와 필기시험점수간의 상관계수는 0.143 (p=0.152)에 불과하였으며, 2002년에도 상관계수는 0.13 (p=0.183)으로 실습점수와 필기시험점수간에는 유의한 상관관계는 없었다 (Table IV). 그러나 실습점수와 필기시험점수 각각은 단순상관분석과 단순회귀분석에서 학년성적과 유의한 상관관계를 가진다 (Fig. 5). 이는 임상실습에 임하는 태도, 출석상황 및 수기의 숙련도를 측정하는 실습점수가 의학적 지식 수준을 측정하는 필기시험과는 서로 독립적이기도 하지만 동시에 학년성적에도 영향을 미치고 있음을 보여 주고 있다. 학년성적에 영향을 미치는 요인에 관한 다중회귀분석상에서도 실습점수와 필기시험점수는 독립적으로 유의한 영향을 미쳐, 1999년도 성적에서는 [학년성적]=-5.269+0.217*[실습점수]+0.686*[필기시험점수] (R²=0.560, p<0.001)의 회귀방정식을, 2002년도 성적에서는 [학년성적]=-5.393+0.230*[실습점수]+0.755*[필기시험점수] (R²=0.669, p<0.001)의 유의한 회귀방정식을 얻을 수 있었다. 2002년도 실습점수와 필기시험점수는 1999년도에 비하여 학년성적에 대한 설명력 (R²)이 증가한 것으로 보아 (56.0% → 66.9%), 2002년도 외과학실습 과목성적이 보다 학년성적과 근접한 상관관계를 가짐을 알 수 있다.

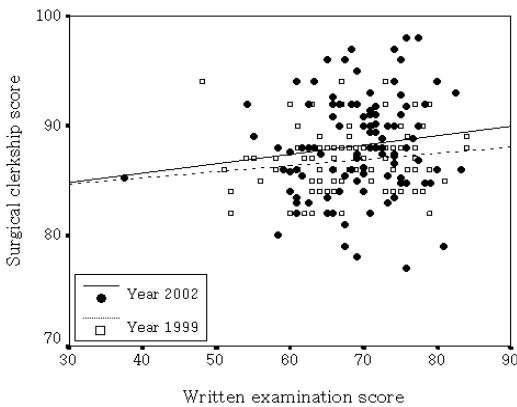


Fig. 4. Scattergram between written examination score and clerkship score. The written examination score was not correlated with the clerkship score (p>0.05).

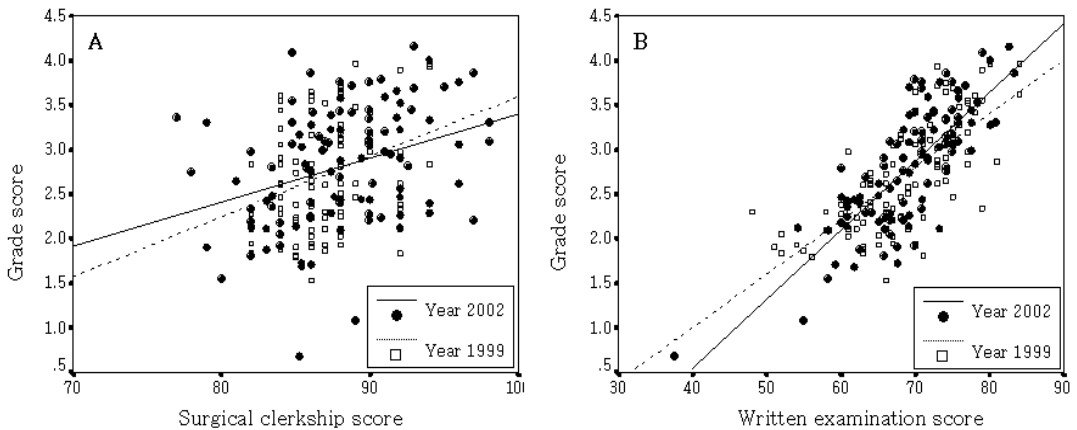


Fig. 5. Scattergram between clerkship score(A) / written examination score(B) and grade score

외과학 임상실습 평가방법의 변경이 실습성적에 미치는 영향

Table V. Correlation Analysis Between Clerkship Score and Written Examination Score at 2002

Subclass		Clerkship score	Written exam. score	Clinic-based written exam. score	Knowledge-based written exam. score	Grade score
Clerkship score	γ *	1.000				
	p-value	-				
Written exam. Score	γ *	.130	1.000			
	p-value	.183	-			
Clinic-based written exam. Score	γ *	.199	.803	1.000		
	p-value	.040	.000	-		
Knowledge-based written exam. score	γ *	.047	.898	.460	1.000	
	p-value	.633	.000	.000	-	
Grade Score	γ *	.328	.785	.700	.654	1.000
	p-value	.001	.000	.000	.000	-

* γ means Pearson correlation coefficients

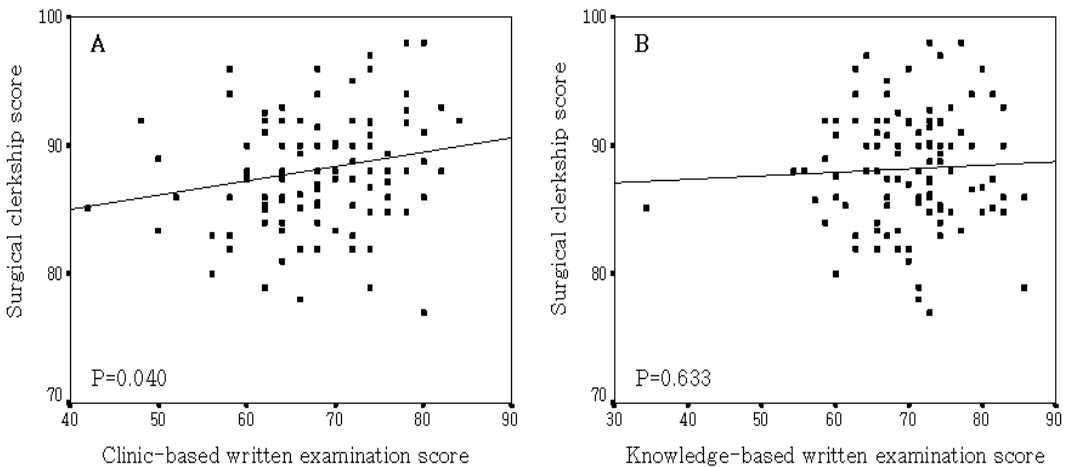


Fig. 6. Scattergram between clinic-based written examination score(A) / knowledge-based written examination score(B) and clerkship score at year 2002.

마. 2002년도에 시행한 실습형 문제점수와 실습점수와의 상관관계

2002년도에 시행한 필기시험에는 전체 120개의 문항중 50개 (41.6%)의 문항은 자료제시형 문제를

포함한 실습형 문제로 출제하였다. 이러한 실습형 문제는 임상실습에 충실하였던 학생들에게 익숙하도록 다발성 질환이나 자주 시행하는 수기 혹은 치료에 제한하여 출제하였다. 실습점수와 필기시험점수와의 유의한 상관관계가 없음에도 불구하고 실습

점수와 실습형 필기시험간에는 상관도는 낮으나 유의한 상관관계를 확인할 수 있었다(상관계수 = 0.199, $p=0.04$) (Table V) (Fig. 6A). 이는 실습형 필기시험 문제가 임상실습에 대한 학습정도를 평가할 수 있음을 입증하고 있다. 반면에 지식형 문제점수와 실습점수와는 유의한 상관관계가 없었다 (Fig. 6B).

고 찰

의학교육의 개선방안에 대한 기본 골격은 1차 의료 수행능력에 모아지고 있다(최삼섭, 1995). 이에 따라서 한국보건인국가시험원에서도 1995년 의사국가시험부터 수험생에 대하여 일차의료 수행능력을 평가하기 위한 임상문제를 출제하는 평가방법을 지향하고 있다(최삼섭, 1998). 임상실습, 특히 핵심과목의 임상실습교육이 의학과 교육과정중 중요하다고 인정되는 이유도 이러한 의학교육 기본골격의 변경에 기인한 것이다. 따라서 최근에는 피교육자인 학생들에 대한 설문조사에 근거한 실습프로그램의 평가에 관한 연구가 많았으며(하은희, 1996; 채수진 등, 2001; 황건, 2002), 이러한 평가결과에 따라서 실습프로그램을 수정하기도 하였다. 그러나 임상실습교육이 강조되는 반면 실습교육을 어떻게 평가할 것인가에 대한 연구나 임상실습 교육에 대한 평가가 적절한지에 대한 연구는 미약한 상태이다. 즉 임상실습 평가방법의 변경이 실제로 실습교육 평가결과에 어떻게 반영되는지에 대한 조사보고는 적은 편이었다. 본 연구에서는 임상실습교육의 평가방법이 변함에 따라서 평가결과에 어떠한 변동이 발생하였는지를 확인하고자 하였다.

임상실습의 평가에서는 먼저 평가기준을 분명하게 설정하여야 한다. 즉 평가는 반드시 기준에 근거하여야 하며 기준별 수준(level)이 구체적으로 명시되어야 한다(김영명, 1990). 저자들은 항목점검표방법에 의거한 임상수첩을 1999년 이후 수정하면서 2002년도에는 40개의 임상실습 술기(skill)와 106개의 학습목표를 재설정하였다. 아울러 항목점검표에 점검된 항목의 수치에 따른 채점기준을 설정하였다 (Table I). 이는 다수의 학생과 평가자가 순환하는

임상실습교육의 실제 상황에서는 학생과 교육자 모두에게 평가의 기준을 공개하고, 이를 일정하게 적용하는 것이 실습평가의 기본이라 판단하였기 때문이다. 실제로 이러한 기준과 평가수준의 공개는 실습기간중 학생들의 학습행위를 유발시키는 효과는 있었으나, 학생간의 우열을 측정하기에는 어려웠다. Table III에서 보듯이 항목점검표 방법으로 측정된 점수에서는 학생들간의 점수차가 크지가 않았다. 전체 실습점수는 오히려 비올척도법으로 측정된 점수에서 차이가 났음을 알 수 있다. 이는 다수의 학생이 장기간에 걸쳐서 순환하는 임상실습교육과정상 항목을 점검할 평가자가 너그럽게 항목을 점검해주는 경향이 있기 때문이다. 따라서 항목점검표방법은 학생들의 우열을 평가하는데 제한이 있으며, 이런 분야에 대한 평가에서는 다른 평가방법이 적용되어야 할 것이다.

실습교육에 대한 평가는 태도나 임상술기와 같은 무형의 능력을 평가해야 하므로 적절한 수준의 평가기준이 구체적으로 규정되어야 함은 물론 적절한 평가방법이 적용되어야 함은 당연하다. 임상실습의 평가방법에 대하여서는 (1) 평가기준이 명확하고, (2) 직접 관찰하여 평가하여야 하며, (3) 가능한 많은 사람이 평가자가 되게 할 것을 요구하고 있다(김영명, 1990). 따라서 저자들은 학생들의 태도평가 주체를 학생교육담당 교수나 전공의에서 모든 교수나 전공의로 확대하였다. 즉 6주간의 순환실습과정중 일주일 간격으로 평가가 규칙적으로 이루어짐으로서 실제적으로는 6인의 평가자가 학생들의 실습태도를 평가한 것이다. 이는 많은 사람이 평가에 참여함으로써 실습점수가 한명의 평가자에 의하여 왜곡되는 것을 방지하기 위함이었다. 실제로 2002년도 실습점수중 태도항목점수는 다른 점수에 비하여 학생간의 우열의 차이를 분명히 나타내어 평가방법으로는 적절한 것으로 판단되었다.

실습점수와 필기시험점수로 구성된 외과학실습 과목성적을 분석한 본 연구결과에 의하면 2002년은 물론 1999년에서도 실습점수와 필기시험점수는 각각 외과학실습 과목성적에 영향을 미치지만, 이들 사이에는 어떠한 유의한 상관관계가 없었다. 이는

실습점수나 필기시험점수가 각각 다른 평가를 내리고 있으면서도 과목성적에는 독립적인 영향을 미치고 있음을 의미하고 있다. 따라서 필기시험점수가 실습평가에 근접하도록 개선하여야 필요성이 제기되고 있다. 저자들은 2002년부터 시작한 실습형 필기시험문제를 문제해결형문항으로 구성하면서, 임상실습과정중 자주 접하게 되는 임상증례의 사진, 그림, 방사선필름 및 검사결과를 제시한 자료제시형으로 시도하였다. 이러한 형식은 1995년 이후의 의사국가시험 문항의 형식이나 내용을 적용시킨 것이었다. 그 결과 2002년에 시행된 실습형 필기시험점수는 실습점수와 상관도는 낮으나 (상관계수=0.199) 유의한 상관관계를 보여주고 있다. 이는 시험의 형식이나 내용을 바꾼다면 객관식 형태의 필기시험도 임상실습평가의 도구로 사용될 수 있음을 의미하고 있다. 따라서 임상실습에 참여하는 학생을 대상으로 한 필기시험은 지식형 문제보다는 임상증례 위주의 문제를 개발함으로써 타당한 실습평가의 도구로서 활용할 수 있을 것이다.

본 연구에서는 실습평가중 비교적 쉽게 실행이 가능한 평가방법을 다양하게 적용하여 2002년도 실습점수를 측정하였으며, 이를 해당학년의 학년성적과의 상관성을 조사하였다. 학생의 전체적인 학습성취도가 학년성적과 꼭 일치하는 것은 아니나, 주요 과목 (내과, 외과, 소아과, 산부인과, 정신과)의 실습이 집중적으로 이루어지는 본교의 학사일정에서는 5학년성적이 학생들의 1차 의료 수행능력이나 전체적인 학습성취도와 가장 근접하다고 추정할 수 있다. 따라서 외과학실습 과목성적은 학년성적과 높은 상관성이 있어야 할 것이다. 본 연구결과를 보면 2002년은 물론 1999년도에도 외과학실습 과목성적이 학년성적과 매우 높은 상관도(상관계수>0.7)를 가짐을 알 수 있었다. 그러나 2002년의 외과학실습 과목성적은 1999년보다 높은 학년성적과의 상관도를 보여, 2002년 외과학실습 과목성적이 보다 학년성적과 근접한 조건을 보이고 있음을 알 수 있었다 (Fig. 3). 특히 외과학실습 과목성적을 실습점수와 필기시험점수로 나누어 살펴본 바에서도 각각의 점수가 1999년도에 비하여 학년성적에 보다 높은 상

관도를 가짐을 알 수 있다. 이는 임상실습 측정방법을 다양한 방법으로 개선함으로써 실습평가가 변경전에 비하여 학생의 전반적인 학습성취도와 보다 근접한 평가를 내리고 있음을 의미하고 있다.

교육평가는 넓은 뜻으로는 교육기관에서 이루어지는 여러 측면의 판단절차를 다 포함하고 있다(백상호, 2001). 이제까지 발표된 실습평가에 관한 논문은 주로 프로그램 평가로 실습의 질을 높이고 실습의 문제점을 파악하는데는 매우 유용하였다(채수진 등, 2001; 황진, 2002). 아울러 재순환과정 (feedback)을 거쳐서 실습프로그램의 수정과 보완의 도구로 사용될 수가 있음을 강조하였었다(하은희, 1996). 그러나 설문지에 의한 프로그램 평가는 학생의 학습 성취도를 판단하는 학습평가와는 분명한 차이가 있다. 본 연구에서는 학생들의 학습 성취도의 기준을 학년성적이라 가정하고 두개의 다른 외과학실습 측정방법의 타당성을 비교하였는데 측정방법이 변경됨에 따라서 평가결과도 상이하였고 전반적인 학습 성취도(학년성적)와의 상관도에도 차이가 있음을 알 수 있었다. 이러한 시도는 교육평가의 내용중 무엇을 평가할 것인가라는 문제보다는 어떻게 평가해야 하는가에 대한 문제로서 교육평가에서 적절한 평가대상을 설정하는 것도 중요하지만 어떠한 평가도구를 사용하느냐도 중요한 문제라 할 수 있다(백상호, 2001). 따라서 학생의 전반적인 학습 성취도와 비례하는 실습프로그램의 개발을 위하여서는 프로그램 평가이외도 평가방법에 대한 검증도 반드시 병행되어야 할 것이다.

결론적으로 다양한 방법을 적용한 임상실습교육의 평가측정은 학생들의 전반적인 학습 성취도에 보다 근접한 실습평가를 내릴 수 있으며, 이러한 과정에서 얻어진 평가결과는 실습프로그램의 수정과 보완과정에 반영되어야 할 것이다.

참 고 문 헌

- 김영명(1990). 임상실습 내용을 평가할 수 있는가? - 수기 및 태도 평가 방법- *한국의학교육*, 2, 19-24.
- 백상호(2001). 학습평가의 뜻. 문항개발능력 향상을

- 위한 워크숍 자료집. 의학교육 연수원 및 한국의학교육학회.
- 장성호, 안덕선, 유세화, 김형규(1993). 의과대학 임상실습 개선에 대한 연구. *한국의학교육*, 5, 34-40.
- 채수진, 장기현, 강홍식, 김우선(2001). 설문지를 이용한 학생에 의한 실습수업평가-서울의대 방사선과학교실 학생실습 결과를 중심으로- *한국의학교육*, 13, 131-141.
- 최삼섭(1995). 의학과 교육과정의 현황과 개선방향. *한국의학교육*, 6, 1-25.
- 최삼섭(1998). 의사국가시험. *한국의학교육*, 10, 65-69.
- 하은희(1996). 실습평가를 통한 실습프로그램의 개선에 관한 연구-일개 의과대학 산업보건 실습을 중심으로- *한국의학교육*, 8, 97-114.
- 황건(2002). 학생들에 의한 성형외과학 실습평가. *한국의학교육*, 14, 43-49.