

임상술기 수행능력에 대한 의과대학생들의 자가평가와 객관구조화진료시험 (OSCE) 결과의 비교

단국대학교 의과대학 가정의학교실

박 일 환

= Abstract =

Comparison of Self-assessment and Objective Structured Clinical Examination (OSCE) of Medical Students' Clinical Performance

Eal Whan Park, MD, PhD

Department of Family Medicine, College of Medicine, Dankook University, Cheonan, Korea

Purpose: The purpose of this study was to assess the clinical experiences (in internal medicine) of medical students, and compare the self-assessment of their competency in clinical skills with the assessment of instructors through the OSCE.

Methods: The frequency of clinical experiences had by medical students and the self-assessments of their confidence in clinical skills were assessed by questionnaires administered before the OSCE. 52 medical students participated in the OSCE. After completing the OSCE, they re-evaluated their own performance of 9 clinical skills examined in the OSCE by self-assessment questionnaires. The scores of these self-assessments were compared with the scores given by the instructors.

Results: For 19 of 28 clinical skills items listed in the questionnaires, self-assessment of competence by medical students correlated with the frequency of performance ($p < 0.05$). For 9 clinical skills in the OSCE, the self-assessment scores did not correlate significantly with the instructors' assessment scores ($p > 0.05$). However, these 9 clinical skills scores from the instructors correlated significantly with the clerkship evaluation scores and the final semester grade point averages ($p < 0.05$).

Conclusion: There was a significant association between frequency of performance and self-assessed competency, which did not correlate significantly with the instructors' evaluation through the OSCE.

Key Words: Self-assessment, OSCE, Medical Students, Clinical Performance

교신저자: 박일환, 단국대학교 의과대학 가정의학과, 충남 천안시 안서동 16-5

Tel: 041)550-3997, Fax: 041)565-6167, E-mail: ewpark@dku.edu

* 이 연구는 2006학년도 단국대학교 대학연구비의 지원으로 연구되었음.

서 론

의과대학생들의 임상수행능력을 평가하기 위하여 객관구조화진료시험 (Objective Structured Clinical Examination, OSCE)이나 임상수행능력 평가시험 (Clinical Performance Examination, CPX) 등 객관적 평가 도구가 많이 개발되어 사용되고 있다. 이러한 평가도구들은 타당성의 측면에서 개선해야 할 점이 있겠지만, 의과대학생들의 임상수행능력을 평가하기 위한 정확하고 신뢰성 있는 도구로 받아들여지고 있다 (Park *et al.*, 1998). 그러나 객관적 평가도구 이외에 학생들이 임상수행능력을 스스로 평가할 수 있는 주관적인 자가 평가 방법도 의학교육 평가 영역에서 중요시되어 왔다.

주관적 자가평가도구는 교수를 포함한 외부의 평가자들에 의한 객관적 평가에 비하여 의과대학생들이 학습해야 할 부족한 점을 보다 정확히 평가할 수 있는 것으로 여겨지고 있다 (Woolliscroft *et al.*, 1993). 또한 주관적 자가평가도구는 의과대학생 스스로 자신의 임상수행능력을 얼마나 잘 수행할 수 있는가에 대한 자신감을 평가하게 하는 장점을 갖고 있다 (Gordon, 1991). 의과대학생들은 주관적 자가평가 도구를 활용함으로써 자신의 부족한 부분이 무엇인가 인식하게 되어, 이런 결과를 토대로 임상술기에서 어떤 부분의 향상이 필요할 것인가를 향후 교육에 반영할 수 있게 되고, 결과적으로 자기주도적 학습 기술을 습득하게 될 것이다.

의학교육에서 자가평가 방법의 중요성이 강조되고 있지만, 실제로 의학교육의 현장에서 객관적 평가도구로는 잘 활용되고 있지 않다. 그 이유는 자가평가 결과가 객관성과 신뢰성이 부족하다고 여겨지기 때문일 것이다. 문제바탕교육(problem-based learning, PBL)에서도 학생들이 스스로를 평가하게 하는 것이 바람직한 것으로 여겨져서 과정 평가 내용에 포함되기도 하지만, 교수나 동료의 평가와 자가평가 결과가 잘 일치하지 않아 학습능력을 평가하는 변별력이 있는 평가도구로 활용되고 있지는 못하고 있다 (Sullivan *et al.*, 1999).

지금까지의 몇몇 외국의 연구들이 학생들의 임상

수행 능력에 대한 객관적 평가 도구에 의한 평가결과와 학생들의 자가평가 결과의 일치성에 대해 연구하였으며, 객관적 평가 도구에 의한 평가결과와 학생들의 자가평가 결과가 잘 일치하고 있지 않다고 보고하고 있다. (Sclabassi & Woelfel, 1984; Arnold *et al.*, 1985; Reghr *et al.*, 1996; Gruppen *et al.*, 1997; Morton & Macbeth, 1997; Vivenkananda-Schmidt *et al.*, 2007).

이러한 외국의 연구들은 대부분 객관구조화진료시험에 의한 임상술기의 평가와는 다르게 임상실습을 마친 의과대학생들의 지식, 기술, 태도를 평가서를 이용하여 포괄적으로 평가하여 비교하는 연구였으며, Vivenkananda-Schmidt *et al.*(2007)의 연구만이 객관구조화진료시험에 의한 평가와 자가평가의 상관성을 연구한 것이었다. 그러나 이 연구도 학생들의 근골격계 진찰 기술에 대한 연구에 국한하였다. 국내에서는 의과대학에서 객관구조화진료시험이 보편화되고 있으나 객관구조화진료시험 평가결과와 자가평가 결과를 비교하는 연구가 아직 이루어지지 않고 있다. 이에 본 저자는 의과대학 3학년 학생들을 대상으로 객관구조화진료시험 평가와 자가평가 결과의 상관성을 연구하기 위하여 본 연구를 수행하였다.

본 연구의 목적은 첫째로, 의과대학생들이 내과학 임상실습을 마치고 난 후, 실습지침서에서 요구하고 있는 흔한 임상술기를 얼마나 자주 경험하였는가를 조사해 보고, 흔한 임상술기에 대한 학생들의 경험빈도와 자가평가한 임상술기 수행능력과의 상관성을 조사해보는 것이었다. 이러한 조사는 객관구조화진료시험 평가결과와 자가평가결과를 비교하는 연구를 수행하기에 앞서서 임상실습을 처음 시작한 학생들이 임상술기를 얼마나 많이 경험하였으며, 이러한 경험에 대한 자신의 수준을 어떻게 평가하는가에 대한 기초적 정보를 얻고자 한 것이었다. 둘째로, 객관구조화진료시험 직후에 시험문항으로 제시된 임상술기에 대한 의과대학생들의 주관적 자가평가 결과와 교수들의 객관적 객관구조화진료시험 평가 결과를 비교하여 주관적 평가 결과와 객관적 평가 결과의 상관성을 조사해 보는 것이었다.

대상 및 방법

의학과 3학년 1학기를 종료한 학생들을 대상으로 본 연구를 시행하였다. 의학과 3학년 학생들은 의학과 3학년 1학기에 4주간의 PBL 블록 학습에 이어서 12주간의 내과학 실습을 하였으며, 내과학 실습이 종료된 직후에 내과학 실습 과정에서 경험했을 것으로 기대하는 흔한 임상술기의 수행평가를 위하여 객관구조화진료시험을 시행하였다. 학생들의 12주간의 내과학 임상실습은 대학병원 내에서만 이루어졌으며, 학생들은 9명을 단위로 1~2주씩 내과의 각 분과를 소그룹으로 순환하면서 실습을 경험하였고, 실습평가는 각 분과의 교육담당 책임교수에 의해 이루어졌다.

객관구조화진료시험을 시행하기 전에 연구자는 의과대학생들로 하여금 본교의 내과학 실습지침서에서 경험하도록 제시한 흔한 임상술기 28가지에 대한 자신의 경험 수준을 평가하게 하였고, 임상술기에 대한 자신의 수행능력을 스스로 평가하도록 하였다. 조사방법으로는 객관구조화진료시험에 대한 예비 설명을 하는 시간에 연구 대상 학생들의 임상술기에 대한 자신의 경험 수준을 자가 평가 설문지로 조사하였다. 조사는 관련된 임상수기 항목 28가지를 열거하고 '3회 이상 경험', '1~2회 경험', '참관', '경험 없음' 등으로 이전의 경험 빈도 및 수준을 표시하게 하였고, 수행능력에 대한 자가 평가는 항목마다 '매우 잘 할 수 있다', '잘 할 수 있다', '보통이다', '잘 못할 것이다', '매우 잘 못할 것이다' 등 5단계 척도로 표시하도록 하였다. 평가를 할 수 없다고 판단할 경우에 '모르겠다' 항목에 표시하게 하여 분석에서 제외하였다.

객관구조화진료시험에 참여한 학생은 총 52명이었으며, 평가는 9개의 문항으로 각각의 스테이션에서 이루어졌다. 평가 문제 중 3문제는 모의 환자를 활용한 면담술을 평가하는 것으로, 혈변, 두통, 고열의 문제를 가진 환자와의 면담이었고, 다른 3문제는 복부 진찰, 심잡음 청취, 강직성 척추염 환자의 진찰법을 평가하는 것이었다. 나머지 3문제는 흉수 천자, 비위관 삽입, 동맥혈 천자의 임상술기를 평가하

는 것이었다. 객관구조화 진료시험 당일에 시험이 끝난 직후에 새로운 평가 설문지를 사용하여 시험 문항 9개에 대하여 의과대학생들로 하여금 자신이 시험문제를 얼마나 잘 수행하였는지를 스스로 평가하도록 하였다. 수행결과에 대한 자가 평가는 항목마다 '매우 잘했다', '잘했다', '보통이다', '잘 못했다', '매우 잘 못했다' 등 5단계 척도로 표시하도록 하였고, 평가를 할 수 없다고 판단할 경우에 '모르겠다' 항목에 표시하게 하여 분석에서 제외하였다.

객관구조화진료시험을 마친 후에 교수가 직접 평가한 객관구조화진료시험 성적과 시험 직후 학생들의 자가 평가 점수와의 상관성을 분석하였다. 객관구조화진료시험은 각 문제당 10점 만점으로 조정되었으며, 총점은 각 스테이션에서의 평가 점수의 합(90점 만점)으로 하였다. 학생들의 자가평가 점수도 각 문제당 10점씩 배점하여 합산하였다. 자가 평가 점수의 배점은 '매우 잘했다' 10점, '잘했다' 8점, '보통이다' 6점, '잘 못했다' 4점, '매우 잘 못했다' 2점으로 하였다.

분석 방법으로는 9개의 객관구조화 진료시험 문제에 대하여 OSCE 채점 교수의 평가 점수와 학생 자신의 자가평가 점수의 상관성을 비모수 상관분석 통계방법의 하나인 Kendall's tau-b 계수를 이용하여 분석하였고, 어떤 문항에서 과대 평가 및 과소 평가가 이루어졌는지 조사하였다. 과대평가와 과소평가는 교수의 평가 점수체계를 2점의 등간격으로 구분하고, 교수의 평가점수를 기준으로 하여 학생들의 자가평가 점수를 대응하여 비교하였다. 교수의 평가 점수보다 높은 경우는 과대평가로, 교수의 평가점수보다 낮은 경우는 과소평가로 구분하였다.

또한, 객관구조화진료시험 결과와 학기말 내과학 실습 시험 평가 점수와 학생들의 1학기말 평점과의 상관성을 Pearson 상관계수를 이용하여 비교하였고, 자가평가 점수와 내과학 실습시험 평가점수와 1학기 평점과의 상관성을 Kendall's tau-b 상관계수를 이용하여 비교하였다. 객관구조화진료시험의 각각의 문항에서 객관구조화진료시험 점수와 자가평가 점수의 평균을 비교하기 위하여 Wilcoxon 부호순위 검정을 시행하였다. 통계 분석은 SPSS version 12를

이용하였다.

결 과

가. 연구대상자의 일반적 특성

연구에 참여한 학생들의 성별은 남학생이 37명 (71.2%), 여학생이 15명 (28.8%)이었으며, 연령은 평균연령 25세 (표준편차 2.8), 연령범위 22~35세이었다 (Table I). 연구대상자들은 이전 학년에 객관구조화진료시험을 경험해 본 적은 없었다. 학생들의 의학과 3학년 1학기 내과 실습점수는 평균 84.5점 (표준편차 7.2점), 범위 71~98점이었으며, 평점 평균은 3.24점 (표준편차 0.62점), 범위 1.94~4.47점이었다.

나. 임상술기에 대한 학생들의 자가 평가 결과

1) 임상술기에 대한 학생들의 경험빈도

내과학 임상실습을 마친 학생들이 흔히 경험할 수 있는 임상 술기 28개 항목을 3학년 내과학 실습지침서에서 추출하여 나열하고, 학생들이 임상실습 기간 중 자신의 경험 빈도 및 수준을 스스로 자가 평가하도록 하였다. 48명의 학생이 설문에 응답하였다. 응답 결과로는 병력청취는 평균적으로 3회 이상 경험했다고 평가했으며, 흉부진찰, 복부진찰, 심전도 판독, 단순흉부촬영판독, 안저검사, 혈압측정, 혈당측정, 손씻기, 정맥혈 채혈 등은 학생들이 실습기간 중 평균 1~2회 경험하였다고 평가하였다. 갑상선 촉진, 혈액배양 (무균채혈), 심잡음 청진, 동맥혈 채혈, 신경학적 진찰, 정맥 주입, 강직성 척추염진찰, 복수천자, 흉수천자, 직장수지검사 등은 참관하

는 수준이었으며, 도뇨관 삽입, 이경검사, 비위관삽입, 성인기관삽관, 세침흡입검사, 척수액천자, 중심정맥관 삽입, 관절 천자 및 주사 등은 경험한 적이 없었다고 평가하였다 (Table II).

2) 임상술기 능력에 대한 학생들의 자가 평가

학생들에게 경험이 요구되는 28항목의 임상술기에 대하여 학생들이 자신의 수행능력을 평가하게 하였다. 자가 평가결과는 평균적으로 볼 때, 손씻기는 잘 할 것이라고 평가하였고, 병력청취, 혈당측정, 혈압측정, 정맥혈 채혈, 갑상선 촉진, 흉부 진찰, 복부 진찰, 강직성 척추염 진찰, 혈액 배양 (무균 채혈), 심전도 판독 등은 보통 정도로 할 수 있다고 평가하였다. 단순흉부촬영 판독, 안저검사, 정맥주입, 심잡음청진, 직장수지 검사, 동맥혈 채혈, 신경학적 진찰, 복수 천자, 도뇨관 삽입, 흉수천자, 중심정맥관 삽입, 비위관 삽입, 관절 천자 및 주사, 성인기관삽관, 척수액 천자, 이경검사 등은 잘 못할 것이라고 평가하였다 (Table II).

3) 임상술기의 각 항목에서 경험빈도와 자가평가 결과의 상관성 비교

열거된 28가지의 임상술기 각각의 항목에서 학생들이 해당 술기를 얼마나 자주 경험하였는가의 빈도와 임상술기 수행 능력에 대한 자가평가와의 상관성을 분석하였다. 분석결과 28개 술기항목 중 19개 항목에서 해당술기의 임상 경험 빈도와 수행능력 평가 간에 통계적으로 유의한 상관성을 보여, 해당 술기 항목에서 경험의 빈도가 높다고 평가한 학생일수록 그 술기를 잘 할 수 있다고 평가하였다 (Table II).

다. 학생들의 자가 평가 결과와 객관구조화진료시험 결과의 비교

자가평가 총점과 객관구조화진료시험 총점의 상관성 분석에서 Kendall's tau-b 상관계수 0.092로 학생들의 자가평가 성적과 교수가 평가한 객관구조화진료시험 성적과의 상관성은 통계적 유의성이 없었다 ($p>0.05$) (Table III).

Table I. General Characteristics of Medical Students

| Characteristics | n (%) | |
|-----------------|----------|-----------|
| Ages (yrs) | 20~24 | 34 (65.4) |
| | 25~29 | 15 (28.8) |
| | ≥30 | 3 (5.8) |
| Sex | Male | 37 (71.2) |
| | Female | 15 (28.8) |
| Total | 52 (100) | |

Table II. Relationship between Frequency of Medical Students' Clinical Experiences and Self-assessment of Competency

| Procedure | No. of times performed | | | | Self-assessment of competency | | | | | Kendall's tau-b coefficient | p value |
|---|------------------------|-------------|-----|----|-------------------------------|------|------|------|-----------|-----------------------------|---------|
| | none | observation | 1-2 | ≥3 | very poor | poor | fair | good | very good | | |
| History taking | 0 | 1 | 0 | 47 | 0 | 5 | 23 | 15 | 5 | 0.229 | 0.094 |
| Hand washing | 2 | 16 | 21 | 8 | 0 | 8 | 15 | 15 | 9 | 0.545 | 0.000 |
| Chest exam | 0 | 5 | 25 | 18 | 1 | 21 | 19 | 6 | 1 | 0.258 | 0.052 |
| Abdomen exam | 3 | 6 | 20 | 19 | 2 | 21 | 18 | 6 | 1 | 0.202 | 0.118 |
| Fundoscopy | 2 | 3 | 40 | 3 | 7 | 25 | 8 | 6 | 0 | 0.329 | 0.015 |
| Otoscopy | 41 | 2 | 4 | 0 | 16 | 16 | 2 | 2 | 0 | 0.465 | 0.004 |
| Neurologic exam | 19 | 11 | 17 | 1 | 13 | 20 | 8 | 3 | 0 | 0.308 | 0.021 |
| Blood pressure measurement | 4 | 12 | 21 | 10 | 2 | 10 | 20 | 13 | 3 | 0.504 | 0.000 |
| Starting intravenous line | 11 | 30 | 6 | 1 | 4 | 22 | 15 | 3 | 0 | 0.208 | 0.131 |
| Blood glucose measurement | 0 | 13 | 33 | 2 | 1 | 5 | 22 | 14 | 6 | 0.346 | 0.009 |
| Foley catheter insertion | 33 | 13 | 2 | 0 | 17 | 21 | 3 | 2 | 0 | 0.332 | 0.022 |
| Auscultating heart sounds | 14 | 11 | 18 | 5 | 6 | 29 | 8 | 2 | 0 | 0.282 | 0.033 |
| Interpret electrocardiogram | 1 | 10 | 18 | 19 | 6 | 17 | 21 | 3 | 1 | 0.322 | 0.012 |
| Interpreting simple chest X-ray | 1 | 16 | 9 | 22 | 2 | 21 | 19 | 4 | 0 | 0.325 | 0.016 |
| Venipuncture | 4 | 17 | 23 | 4 | 4 | 10 | 15 | 16 | 3 | 0.461 | 0.000 |
| Obtaining an arterial blood sample | 12 | 23 | 12 | 1 | 14 | 20 | 10 | 1 | 1 | 0.392 | 0.003 |
| Nasogastric tube insertion | 29 | 16 | 3 | 0 | 20 | 19 | 4 | 0 | 0 | 0.326 | 0.025 |
| Digital rectal examination | 29 | 13 | 6 | 0 | 10 | 23 | 7 | 3 | 1 | 0.409 | 0.003 |
| Abdominal paracentesis | 20 | 24 | 4 | 0 | 19 | 16 | 7 | 2 | 0 | 0.283 | 0.042 |
| Thoracentesis | 23 | 23 | 2 | 0 | 20 | 18 | 5 | 1 | 0 | 0.152 | 0.285 |
| Endotracheal intubation (adult) | 36 | 12 | 0 | 0 | 21 | 17 | 4 | 0 | 0 | 0.483 | 0.001 |
| Fine needle aspiration | 29 | 18 | 1 | 0 | 29 | 11 | 2 | 0 | 0 | 0.154 | 0.311 |
| Thyroid exam | 3 | 22 | 20 | 3 | 0 | 11 | 26 | 9 | 2 | 0.477 | 0.000 |
| Lumbar puncture | 38 | 10 | 0 | 0 | 25 | 9 | 2 | 3 | 1 | 0.278 | 0.068 |
| Central venous line insertion | 32 | 16 | 0 | 0 | 22 | 11 | 4 | 2 | 1 | -0.028 | 0.852 |
| Examining ankylosing spondylitis | 14 | 28 | 6 | 0 | 6 | 12 | 13 | 13 | 1 | 0.288 | 0.030 |
| Arthrocentesis & intraarticular injection | 36 | 12 | 0 | 0 | 25 | 11 | 3 | 2 | 1 | 0.191 | 0.199 |
| Obtaining blood culture | 31 | 15 | 2 | 0 | 4 | 18 | 16 | 5 | 3 | 0.297 | 0.028 |

연령군을 구별하여 자가평가 점수와 객관구조화진료시험 점수의 상관성을 분석하였으나 의미있는 상관성은 없었으며 (상관계수 24세 이하군 0.077, 25세 이상군 0.020, 각각 $p > 0.05$), 남학생군과 여학생군을 구별하여 자가평가 점수와 객관구조화진료시험 점수의 상관성을 분석하였으나 의미있는 상관성은 없었다 (상관계수 남학생 0.111, 여학생 0.038, 각

각 $p > 0.05$). 학기말 평점에 따른 상위 25%군과 하위 25%군에서 자가평가 점수와 객관구조화진료시험 점수와의 상관성을 분석하였으나 의미있는 상관성은 없었다 (상관계수 상위 25%군 -0.088, 하위 25%군 -0.013, 각각 $p > 0.05$) (Table III).

객관구조화진료시험 9개 문항 각각에 대하여 자가평가 점수의 평균과 교수들의 객관구조화진료시

Table III. Relationship between Self-assessment and OSCE

| Group | Subgroup | OSCE | | Students' self assessment | | Kendall's tau-b coefficient | p value |
|-----------|----------------|------|-----|---------------------------|------|-----------------------------|---------|
| | | Mean | SD | Mean | SD | | |
| Total | | 50.9 | 7.7 | 49.3 | 10.9 | 0.092 | 0.346 |
| Age (yrs) | 20~24 | 51.5 | 7.7 | 50.5 | 9.3 | 0.077 | 0.531 |
| | ≥25 | 50.0 | 7.8 | 47 | 13.4 | 0.020 | 0.909 |
| Sex | Male | 50.6 | 8.1 | 50.3 | 10.8 | 0.111 | 0.345 |
| | Female | 51.8 | 6.6 | 46.9 | 11.2 | 0.038 | 0.843 |
| GPA* | Upper quartile | 56.4 | 7.9 | 46.4 | 11.6 | -0.088 | 0.661 |
| | Lower quartile | 47.7 | 7.3 | 48.5 | 12.5 | -0.013 | 0.951 |

*grade point average

Table IV. Comparison of Mean Scores in Self-assessment and OSCE

| Procedure | Teachers' evaluation | | Students' self assessment | | Wilcoxon's Z | p value |
|----------------------------------|----------------------|-----|---------------------------|-----|--------------|---------|
| | Mean | SD | Mean | SD | | |
| History taking (melena) | 6.9 | 1.6 | 5.5 | 1.9 | -3.815 | 0.000 |
| History taking (headache) | 6.9 | 1.6 | 6.5 | 2.2 | -1.446 | 0.148 |
| History taking (high fever) | 5.4 | 2.0 | 6.0 | 1.9 | -1.744 | 0.081 |
| Abdomen exam | 7.7 | 2.2 | 5.7 | 2.1 | -5.347 | 0.000 |
| Examining ankylosing spondylitis | 4.3 | 2.8 | 4.0 | 1.8 | -0.373 | 0.709 |
| Auscultating heart sounds | 4.1 | 1.1 | 4.7 | 2.1 | -2.278 | 0.023 |
| Thoracentesis | 5.2 | 1.9 | 6.0 | 2.2 | -2.202 | 0.028 |
| Nasogastric tube insertion | 5.6 | 1.3 | 5.9 | 1.8 | -1.320 | 0.187 |
| Obtaining arterial blood sample | 5.0 | 1.3 | 5.5 | 2.0 | -1.802 | 0.072 |

험 점수의 평균을 비교하였다. 분석 결과 혈변환자의 병력 청취, 흉강천자, 복부 진찰, 심잡음 청진 등의 문항에서 통계적으로 의미있는 차이가 있어 자가평가 점수와 객관구조화진료시험 점수가 일치하지 않음을 의미하였다 (Table IV).

라. 자가평가 및 객관구조화진료시험 결과와 내과학 실습점수, 3학년 1학기 평점과의 상관성 비교

자가평가 점수와 내과학 실습성적의 상관분석에서 Kendall tau-b 상관계수 -0.035로서 통계적으로 유의한 상관성이 없었으며 (p>0.05), 3학년 1학기 전체 평점 평균과의 상관분석에서 Kendall's tau-b 상관계수 -0.05로 유의한 상관성은 없었다 (p>0.05).

객관구조화진료시험 성적과 내과학 실습성적의 상관분석에서 피어슨 상관계수 0.42로서 유의한 상관성을 보였으며 (p=0.01), 객관구조화진료시험 성적과 3학년 1학기 전체 평균 평점과의 상관분석에서 피어슨 상관계수 0.51로서 유의한 상관성을 보였다 (p=0.01) (Table V).

마. 학생들의 자가평가지 과대평가 및 과소평가 내용

9개 시험 문항별로 학생들의 시험 직후의 자가평가 결과와 채점교수의 평가결과를 비교하여 채점교수의 평가결과를 황금기준으로 할 때 학생들의 과대평가율과 과소평가율을 조사하였다.

학생들의 자가평가 결과 비위관 삽입, 혈변환자 병력청취, 고열환자 병력청취 등에 대하여 과소평가율보다 과대평가율이 높았다. 복부진찰의 경우에는 과소평가율과 과대평가율이 동일하였다. 학생들의 자가평가 결과 흉강천자, 강직성 척추염 진찰, 심잡음청진, 동맥혈 채혈 등에서 과대평가율보다 과소평가율이 상대적으로 높았으며, 두통환자 병력청취 문항에서도 과소평가율이 과대평가율보다 높았다 (Table VI).

고 찰

본 연구를 통하여 연구자는 의과대학생들의 임상술기에 대한 경험 빈도 및 수준에 대해 조사하였고, 임상수행능력에 대한 자가평가 결과와 객관구조화

Table V. Relationship between OSCE, Self-assessment, Clerkship Scores, and GPA

| | OSCE | Student self assessment |
|------------------|--------|-------------------------|
| Clerkship scores | 0.423* | -0.035* |
| GPA | 0.509* | -0.050 [†] |

*Pearson correlation coefficient, $p < 0.05$, [†]Kendall's tau-b coefficient, $p > 0.05$
GPA: grade point average

진료시험 결과와의 상관성을 분석하였다.

본 연구결과에서 학생들의 임상술기 경험에 대한 자가평가에서 28가지 임상술기 중 19가지에서 학생들이 경험 회수가 많을수록 자신의 수행능력을 높게 평가하였는데, Fincher & Lewis (1994)의 연구에서도 16가지의 임상술기 중 13가지에서 통계적으로 유의하게 학생들이 임상실습 기간 중에 경험해 본 횟수가 많을수록 자신의 수행능력을 높게 평가하였다고 보고하여 본 연구결과와 유사하였다. 그러나 Fincher & Lewis (1994)의 연구와 비교하여 본 연구에서는 28가지 임상술기 중 9가지로 비교적 많은 항목에서 학생들의 경험 빈도와 수행능력 자가평가 간에 통계적 유의성이 없는 것으로 나타났는데, 이러한 결과는 자가 설문조사에서 '모르겠다'로 응답한 학생들을 분석에서 제외함으로써 대상수가 감소하였고, 일부 설문 문항에서는 학생들이 2~3개의 응답항목에만 집중적으로 응답하여 전혀 응답하지 않은 응답항목들이 존재함으로써 통계적 검정력이 감소했기 때문으로 생각된다.

본 연구결과에서 학생들의 자가평가 결과와 교수들의 객관적 평가결과가 잘 일치하지 않고 있으며, 이러한 결과는 기존의 외국 연구들의 결과에서도 동일하였다. Sclabassi & Woelfel (1984)은 마취과학 임상실습을 마친 의학과 3학년 학생들의 임상 지식과 기술에 대한 자가 평가서를 이용한 평가 결과와 동일한 평가서를 이용한 임상강사의 평가 결과를

Table VI. Underestimation and Overestimation of Self-assessment

| Procedure | Underestimation | Overestimation | N |
|----------------------------------|-----------------|----------------|----|
| | n (%) | n (%) | |
| History taking (melena) | 14 (26.9) | 15 (28.8) | 52 |
| History taking (headache) | 16 (30.8) | 15 (28.8) | 52 |
| History taking (high fever) | 12 (23.1) | 20 (40) | 50 |
| Abdomen exam | 19 (36.5) | 19 (36.5) | 52 |
| Examining ankylosing spondylitis | 25 (48.1) | 9 (17.3) | 52 |
| Auscultating heart sounds | 28 (56) | 9 (18) | 50 |
| Thoracentesis | 35 (68.6) | 6 (11.8) | 51 |
| Nasogastric tube insertion | 14 (26.9) | 22 (42.3) | 52 |
| Obtaining arterial blood sample | 43 (82.7) | 2 (3.8) | 52 |

비교하였으나, 두 평가 결과 사이에 유의한 일치성은 없었다. Woolloscroft *et al.* (1993)도 내과학 실습을 마친 3학년 학생을 대상으로 한 연구에서 의학적 면담술과 진찰술의 평가에서 교수들과 학생들간에 의미있는 일치가 없었다고 보고하였다. Reghr *et al.* (1996)은 정신과 임상실습을 도는 의학과 3학년 학생들을 대상으로 면담술에 대해 동일한 평가지를 사용하여 자가평가와 전문가 평가를 비교하였으나, 통계적으로 유의한 상관성은 없었다. Gruppen *et al.* (1997)도 의학과 4학년 학생들의 면담술 평가에서 학생들의 자가평가 결과와 모의 환자의 평가 결과 사이에 통계적으로 유의한 상관성을 발견하지 못했다고 보고하였다. Weiss *et al.* (2005)은 산부인과 임상실습을 마친 의과대학생들에 대한 연구에서 면담술에서는 학생들의 평가 점수와 전공의와 교수들의 평가 점수가 일치하지만, 임상 술기에서는 일치하지 않았다고 보고하였다.

Vivenkananda-Schmidt *et al.* (2007)은 근골격계 진찰술기에 대하여 15개 항목의 10점 척도의 자가평가서를 개발하여 자가평가와 객관구조화진료시험 결과를 비교하였는데, 자가평가 점수와 객관구조화진료시험 점수 간에 통계적으로 유의한 상관성을 관찰하지 못했다. 이 연구에서는 남학생 군에서는 객관구조화진료시험 평가와 자가평가 간에 유의한 차이가 없었으나 여학생군에서는 유의한 차이가 있었다고 보고하였으나 ($p < 0.05$), 본 연구에서는 성별에 따라 구분한 소그룹 분석에서 모두 유의한 차이가 없었다.

본 연구에서 관찰된 불일치의 원인으로서는 다음과 같은 요인을 고려해 볼 수 있다.

첫째로, 평가 도구의 평가 방법의 차이이다. 객관구조화진료시험 평가도구는 각각의 임상술기에 대하여 자세한 세부 평가 기준에 근거하여 평가하고 있는 데 반하여, 학생들의 자가평가 도구는 자신의 수행능력을 5점 척도로 포괄적으로 평가해 보게 하는 것이었다. 따라서 양쪽 평가 도구의 평가 범주 및 척도기준이 달라서 평가 결과의 불일치에 영향을 줄 수 있을 것이다. 또한, 객관구조화진료시험의 특성상 의과대학생들이 어떤 술기를 평가할 때에

어떤 행동을 잘 하였는가를 평가하기보다는, 어떤 행동을 하였는가 하지 않았는가에 초점을 두기 때문에 자가능력 평가 결과와 잘 일치하지 않을 수도 있다 (Kim, 2002; Vivenkananda-Schmidt *et al.*, 2007).

둘째로, 학생들의 자가평가 능력의 부족이다. 본 연구에서 많은 항목에서 학생들이 스스로를 과소평가하고 있다는 사실에 근거하여 학생들이 교수나 전공의들에 의해 자신의 임상수행능력에 대한 적절한 되먹임을 받은 경험이 부족했다는 것을 고려할 수 있다. 이러한 이유로 학생들은 자신들의 임상수행능력에 대하여 자신감을 갖지 못했다고 볼 수 있다.

Mavis (2001)는 의학과 2학년 학생들을 대상으로 객관구조화진료시험 평가전후에 실시한 자신의 임상수행능력에 대한 자신감을 묻는 설문 조사의 결과를 분석한 결과, 객관구조화진료시험 평가점수와 자신감에 대한 평가 점수가 통계적으로 유의한 상관관계를 보이지는 않았으나, 자신감이 높은 군과 자신감이 낮은 군으로 구분하여 비교하였을 때에 자신감이 높은 군에서 객관구조화진료시험 점수가 높게 나타났다.

셋째로, 평가도구의 평가 영역의 차이이다. 자가평가는 객관구조화진료시험 등의 객관적 평가와는 다르게, 자신에 대한 인식과 자신의 능력을 표현하는 과정과 연계되어 있다고 할 수 있다. 객관적 평가 방법은 학생의 학문적 혹은 임상적 성취도를 평가하며, 바람직한 의사로서의 수행능력을 평가하는 것이라고 할 수 있는 데 반해, 자가평가는 자신의 능력이 어떠한가에 대한 전반적 평가와 자신의 수행능력에 대한 자신감을 평가하며, 이러한 평가는 과거로부터 지금까지 자신의 능력이 어느 수준일 것이라는 자아 인식과 연결되어 있다고 하겠다 (Gordon, 1991).

결론적으로 자가평가는 학생들이 자신이 얼마나 임상술기를 잘 할 수 있겠는가에 대해 갖고 있는 자신감을 표현하는 것으로서, 임상수행 능력 자체를 평가하는 객관구조화진료시험 평가와는 다르다고 할 수 있다.

의과대학생들의 임상술기 교육에 있어서 학생들

의 임상수행능력의 부족한 부분을 교육시키는 것은 물론 중요하겠지만, 자신의 임상수행능력을 정확히 평가하는 기술도 교육해야 한다는 것을 고려해야 한다 (Barnsley *et al.*, 2004). 현재 의과대학생들이나 전공의들의 임상술기 교육은 자신의 능력에 대한 부정확한 인식을 바탕으로 자신의 행위에 대한 지도전문의의 적절한 되먹임에 의한 교정의 과정없이 단순히 반복수행함으로써 임상술기가 저절로 향상 될 것이라는 확신 속에서 이루어질 가능성이 높다 (Gordon, 1991). 의과대학생들이 임상수행능력을 공정하게 평가할 수 있는 교사의 역할 모델을 많이 경험하여 자신의 임상수행능력에 대하여 적절한 되먹이기를 받는다고 하면, 자신의 임상수행능력을 스스로 평가하는 기술을 갖추고 정확한 자가평가를 이룰 수 있을 것으로 생각된다 (Woolliscroft *et al.*, 1993)

그러나, 객관구조화진료시험을 위해 평가자를 훈련시키는 과정보다 자가평가를 위해 학생들의 평가 기술을 훈련시키는 과정은 더욱 복잡하며 많은 시간을 요한다고 하겠다. 현재로서는 비용, 효과적인 측면에서 객관적인 평가도구에 의한 피드백을 줄 수밖에 없겠지만, 자가평가의 여러 장점을 고려할 때에 정확한 자가평가를 위한 많은 준비가 필요할 것이다.

참 고 문 헌

Arnold, L., Willoughby, T.L., & Calkins, E.V.(1985). Self-evaluation in undergraduate medical education: a longitudinal perspective. *J Med Educ*, 60(1), 21-28.

Barnsley, L., Lyon, P.M., Ralston, S.J., Hibbert, E.J., Cunningham, I., Gordon, F.C., & Field, M.J. (2004). Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. *Med Educ*, 38(4), 358-367.

Fincher, R.M., & Lewis, L.A.(1994). Learning, experience, and self-assessment of competence of

third-year medical students in performing bedside procedures. *Acad Med*, 69(4), 291-295.

Gordon, M.J.(1991). A review of the validity and accuracy of self-assessments in health professions training. *Acad Med*, 66(12), 762-769.

Gruppen, L.D., Garcia, J., Grum, C.M., Fitzgerald, J.T., White, C.A., Dicken, L., Sisson, J.C., & Zweifler, A.(1997). Medical students' self-assessment accuracy in communication skills. *Acad Med*, 72(10), S57-S59.

Kim, S.H., Lee, Y.M., Lee, D.H., Choi, Y.S., Cho, K.H., & Hong, M.H.(2002). Assessing the clinical competence of family medicine clerkship students using the OSCE. *Journal of Korean Academy of Family Medicine*, 23(5), 583-592.

Mavis, B.(2001). Self-efficacy and OSCE performance among second year medical students. *Adv Health Sci Educ Theory Pract*, 6(2), 93-102.

Morton, J.B., & Macbeth, W.A.A.G.(1997). Correlations between staff, peer, and self assessments of fourth-year students in surgery. *Med Educ*, 11, 167-170.

Park, H.K., Kim, D.W., Kim, D.A., Choi, H.S., & Kim, K.T.(1998). An OSCE assessment of fourth-year medical students as a comprehensive evaluation of medical skills before graduation. *Korean Journal of Medical Education*, 10(1), 43-56.

Regehr, G., Hodges, B., Tiberius, R., & Lofchy, J.(1996). Measuring self-assessment skills: an innovative relative ranking model. *Acad Med*, 71(10S), S52-S54.

Scabassi, S.E., & Woelfel, S.K.(1984). Development of self-assessment skills in medical students. *Med Educ*, 18(4), 226-231.

Sullivan, M.E., Hitchcock, M.A., & Dunnington, G.L.(1999). Peer and self assessment during problem-based tutorials. *Am J Surg*, 177(3), 266-269.

Vivenkananda-Schmidt, P., Lewis, M., Hassell, A.B., Coady, D., Walker, D., Kay, L., McLean, M.J., & Rahman, A.(2007). Validation of MSAT: an instrument to measure medical students' self-assessed confidence in musculoskeletal examination skills. *Med Edu*, 41(4), 402-410.

Weiss, P.M., Koller, C.A., Hess, L.W., & Wasser, T.(2005). How do medical student self-assess

ments compare with their final clerkship grades? *Med Teach*, 27(5), 445-449.

Woolliscroft, J.O., Tenhaken, J., Smith, J., & Calhoun, J.G.(1993). Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Acad Med*, 68(4), 285-294.