

Developing and Testing an Evaluation Tool to Measure Clinical Teaching

Eun-Jung Im

Department of Medical Education and Office of Medical Education, Kyungpook National University School of Medicine, Daegu, Korea

임상실습 수업평가를 위한 평가도구 개발 및 검증

경북대학교 의학전문대학원 의학교육학교실, 의학교육실

임은정

Purpose: Teaching evaluation tools can be used to provide feedback to medical educators and help them improve their teaching skills. The purpose of this study was to develop a clinical teaching evaluation tool and test its efficacy.

Methods: The draft tool was developed based on clinical education literature and was modified by the stakeholder groups. To examine the quality of the instrument, content and construct validity, as well as reliability and fitness were analyzed. Content validity was tested by the stakeholder groups using a survey, and construct validity was verified by confirmatory factor analysis using LISREL 8.8. Internal consistency of items was assessed through Cronbach α estimation. Rasch analysis using Winstep 3.65 was performed to estimate the fitness of the tool.

Results: The resulting tool consisted of 4 large categories, 25 small categories, and 43 items. According to the test results, the average importance of all 43 items was 4.03 (3.63 to 4.29). Cronbach α was 0.9689, and the correlation coefficients between the items were high. With regard to construct validity, 10 items needed minor modifications in the category setting. As the infit (0.76 to 1.23) and outfit (0.75 to 1.40) indices show, 42 items were fit to the item response theory.

Conclusion: The clinical teaching evaluation tool that has been developed in this study is valid and reliable and fits the item response theory. It can be used as an evaluation method in a variety of clinical teaching settings.

Key Words: Educational measurement, Program evaluation, Faculty, Teaching

서론

수업은 의학교육자로서의 전문성이 최종적으로 실현되는 교육적 장면이다. 의과대학 교수 개개인이 진행하는 수업의

질이 의학교육의 질을 결정짓는다고 해도 과언이 아닐 정도로, 의학교육에서 수업이 차지하는 비중은 매우 크다고 할 수 있다. 이러한 수업의 중요성에 대한 인식이 확산되면서, 의학 교육 분야에서도 임상실습 수업을 개선하기 위한 다양한 시도와 연구들이 지속적으로 이루어져 오고 있다[1,2,3,4]. 그러

Received: October 9, 2010 • Revised: November 10, 2010 • Accepted: November 12, 2010

Corresponding Author: Eun-Jung Im

Department of Medical Education, Kyungpook National University School of Medicine, 101 Dongin-dong, Jung-gu, Daegu 700-422, Korea

Tel: +82.53.420.4904 Fax: +82.53.423.1369 email: eunjim@knu.ac.kr

This study was presented at The 26th Medical Education Conference in 2010.

Korean J Med Educ 2011 Mar; 23(1): 49-59.

doi: 10.3946/kjme.2011.23.1.49.

pISSN: 2005-727X eISSN: 2005-7288

© The Korean Society of Medical Education. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

나 문제는 수업 개선을 위한 이러한 노력들에도 불구하고, 임상실습 교육방법 및 평가도구의 미흡 등 임상실습 수업과 관련된 문제점들이 계속하여 제기되고 있다는 점이다[5,6,7,8].

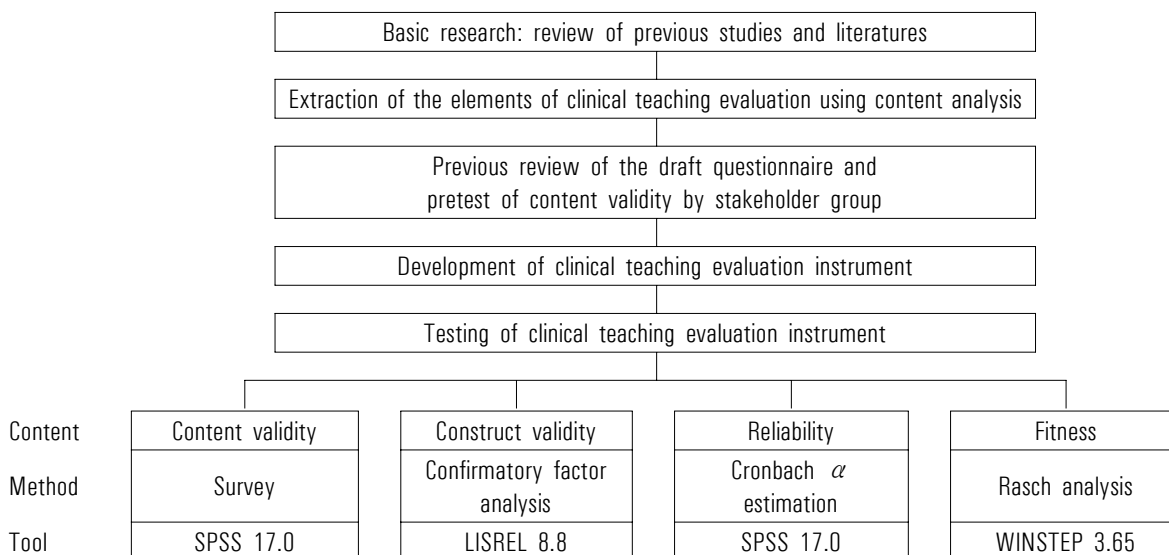
Stronge [9]는 수업 개선을 위한 다양한 노력들이 성과 없이 끝나 버리는 이유 중의 하나가 바로 수업평가의 부재 때문이라고 지적한다. 새로운 수업 개선안을 찾으려는 노력에 비해 현재 진행되고 있는 수업에 대한 분석과 평가는 그만큼 충실히 이루어지지 않았다는 의미이다. 수업에 대한 정확한 분석과 평가 없이는 실천력 있는 개선안도 나올 수 없으며, 교수들 역시 좋은 수업을 위한 체계적인 방법을 찾기 어렵다는 점을 감안해 볼 때, 수업평가는 의학교육의 질적 향상을 위해 반드시 이루어져야 할 중요한 작업임에 틀림이 없다.

일반적으로 수업평가(teaching evaluation)는 교수들의 수업활동에 대한 평가를 통해 교육의 질 향상을 위한 정보를 제공해 주는 데 궁극적인 목적이 있다[10,11]. 따라서 수업평가가 제대로 잘 이루어지지만 한다면, 수업평가 결과는 교수들의 수업능력을 진단해 주고 수업의 성공 여부를 가능하게 해 주는 중요한 정보가 될 수 있다. 그러나 현재 국내 의과대학 및 의학전문대학원에서 시행되고 있는 수업평가가 과연 이러한 목적을 달성하고 있느냐에 대해서는 의문의 여지가 많다. 과거에 비해 수업평가를 실시하고 있는 의과대학 및 의학전문대학원들이 많이 증가하기는 하였으나, 교수들은 수업평가

에서 많은 정보를 기대하지 않을 뿐만 아니라 수업평가 자체를 신뢰하지 않는 교수들도 많다.

특히 수업평가 도구의 문제는 가장 큰 문제점으로 인식되고 있다. 선행연구[12,13]에서 지적된 바와 같이, 타당성과 신뢰성을 갖춘 평가도구의 부족으로 인해 수업에 대한 평가와 피드백이 제대로 이루어지지 않고 있으며, 연구자가 본 연구의 사전 연구(pilot study) 차원에서 실시한 전국 의과대학 및 의학전문대학원의 임상실습 수업평가 현황 조사에서도 객관적인 평가도구의 문제가 임상실습 수업평가의 가장 큰 문제점으로 지적된 바 있다. 조사결과에 따르면, 대부분의 의과대학 및 의학전문대학원들이 수업평가를 실시하고는 있었지만, 강의식 수업과 임상실습 수업을 동일한 평가문항으로 평가하거나(62.2%), 본교에서 공통적으로 사용되는 수업평가 설문지와 같이 의과대학의 특수성을 충분히 반영하지 못한 설문지를 그대로 차용하는 경우도 많았다(35.1%). 그 결과, 수업평가가 정기적으로 실시되고 있음에도 불구하고, 교수들은 정작 자신의 수업에서 무엇이 잘되고 못되고 있는지에 대한 정보는 물론 더 나은 수업을 위한 피드백을 제대로 받지 못하고 있는 것으로 밝혀졌다. 간단한 몇 개의 문항만으로 실시된 수업평가 결과를 보고 자신의 교육자적 자질이나 수업능력을 검증해 보는 교수는 거의 없는 것으로 나타나, 결국 교수들에게 있어서 수업평가는 형식적인 관례나 행정적인 절차에 지

Fig. 1. The Overall Process of This Study



나지 않는 것으로 확인되었다.

수업평가가 교수들의 수업능력에 대한 구체적인 정보를 제공하고 수업 개선을 위한 도구로 활용되기 위해서는 타당성과 신뢰성을 갖춘 평가도구가 전제되어 있어야만 하며, 이러한 평가도구가 구비되어 있을 때 평가결과의 교육적 활용도 훨씬 유의하게 이루어질 수 있을 것이다. 이 연구는 강의식 수업평가에 비해 임상실습 수업평가가 제대로 시행되지 않고 있으며, 특히 임상실습 수업평가를 위한 평가도구 개발 및 평가도구에 대한 경험적 검증은 이루어지지 않고 있다는 사전 연구 결과에 근거하여 임상실습 수업평가를 위한 평가도구를 개발하고 검증하는 데 연구의 목적을 두었다.

대상 및 방법

이 연구는 크게 임상실습 수업평가를 위한 1) 평가도구 개발이라는 질적 연구와, 2) 평가도구 검증이라는 양적 연구로 나누어 진행되었으므로, 대상 및 방법도 두 부분으로 구분하여 기술하고자 한다. 전체적인 연구의 과정은 Fig. 1과 같이 진행되었다.

1. 평가도구 개발

평가도구 개발을 위한 첫 번째 단계는 임상실습 수업평가에 포함되어야 할 요소들을 추출해 내기 위한 기초연구의 단계로서, 선행연구들에 대한 이론적 고찰을 중심으로 이루어졌다. 먼저 임상실습 수업평가와 관련된 연구문헌 및 PubMed, MEDLINE, KSME, RISS4U 등의 데이터베이스를 분석대상으로 삼아 바람직한 임상교수의 요건 및 효율적인 임상수업의 요인, 임상실습 수업평가 등을 주제로 다루고 있는 연구물 128개를 1차적으로 추출해 낸 다음, 연구문헌들에 대한 분석을 통해 수업평가 도구 개발에 실질적인 도움을 줄 수 있는 연구들을 2차적으로 선별해 내는 과정을 거쳤다. 예컨대, 시론이나 단신 수준의 이론적 연구들은 제외하였고, 수업평가 도구와 관련된 연구들 가운데서는 최소한 2회 이상 타당성을 검증받은 연구들을 중심으로 자료의 범위를 축소시켜 나갔다.

다음 단계는 2차적으로 선별된 38개 연구들에 대한 내용분석(content analysis)을 통해 임상실습 수업평가 요소들을 실

제로 추출해 내는 단계로서, 최소한 5개 이상의 연구에서 공통적으로 포함하고 있는 임상실습 수업평가의 요소들을 발췌하여 가설적인 수업평가 설문지를 제작하였다. 각 연구에서 제시하고 있는 임상실습 수업평가의 범주들을 하나의 표로 통합하여 정리한 뒤, 5개 이상의 연구에 해당되는 범주만을 따로 선별해 내어 임상실습 수업평가 문항의 요소로 포함시켰다. 특히 많은 의과대학에서 임상실습 수업평가의 준거로 활용되고 있는 스탠퍼드 교수 개발 프로그램(Stanford Faculty Development Program)의 임상실습 수업의 범주[14]와 클리블랜드 클리닉 재단(Cleveland Clinic Foundation)의 임상교수 수업평가 도구[15], 그리고 임상교수 평가의 틀을 제안한 Litzelman et al. [16]의 연구와 좋은 임상교수의 특성과 관련된 선행연구들을 종합 분석해 놓은 Sutkin et al. [17]의 연구 등이 임상실습 수업평가의 요소를 결정하는 데 중요한 이론적 근거를 제공해 주었다.

수업평가 도구에 포함될 요소가 결정되고 이에 따른 평가 문항 초안이 개발된 다음에는 예비조사 차원에서 의학교육학 교실 교수 3명(겸무교수 포함)과 임상강사 2명, 전공의 2명, 대학원생 2명, 교육학 전공자 1명에게 가설적으로 개발한 수업평가 문항 초안에 대한 검토를 의뢰하였다. 수업평가 문항이 임상실습 수업의 핵심적인 요소들을 고루 평가하고 있는지의 여부와 각 문항이 임상실습 수업평가 문항으로서 지니는 중요도를 표시하도록 요청하였으며, 필요한 경우에는 개별 면담을 실시하였다. 이들의 의견을 종합하여 중요도가 떨어지는 2개의 문항은 삭제하였고, 문항의 의미가 모호하거나 이해가 어려운 문장은 표현을 일부 수정하여 최종적인 임상실습 수업평가 도구를 완성하게 되었다.

2. 평가도구 검증

새롭게 개발한 임상실습 수업평가 도구에 대한 검증은 내용타당도, 구성타당도, 신뢰도, 적합도로 나누어 실시되었다. 모든 통계자료의 분석은 통계 전문가의 협조를 얻어 진행되었으며, 검증 내용별로 사용된 분석도구와 분석방법을 정리해 보면 다음과 같다.

1) 내용타당도 검증

평가도구의 타당도를 검증하는 방법에는 내용타당도(content validity), 구성타당도(construct validity), 신뢰도

(reliability), 적합도(fitness) 등 여러 가지 방법이 있으나, 이 연구에서는 내용타당도와 구성타당도를 중심으로 평가도구의 타당도를 분석해 보았다. 내용타당도란, 임상실습 수업평가 도구가 임상실습 수업에서 평가하고자 하는 내용을 얼마나 적절하게 평가하고 있는지를 검증하는 것으로서, 그 분야 내용 전문가들의 분석을 통해 결정되는 것이 일반적이다. 이 연구에서는 새롭게 개발한 임상실습 수업평가 도구의 내용타당도 검증을 위하여 경북대 의학전문대학원 임상교수와 전임의, 전공의, 3, 4학년 학생들(n=538)을 대상으로 평가문항의 중요도를 묻는 설문조사를 실시하였다. 각 문항들이 임상실습 수업평가 문항으로서 어느 정도 중요성을 지니는가를 5점 척도에 표시하게 하여 개발된 문항 내용의 타당도를 분석하였고, 평가문항에 포함되지 않은 중요한 영역이 있는지를 확인하기 위하여 자유 기재식의 주관식 서술형 문항을 별도로 첨가해 놓았다. 설문조사는 2010년 2월부터 3월에 걸쳐 실시되었으며, 임상교수 77명, 전임의 45명, 전공의 90명, 학생 157명 등 총 369명이 최종 응답에 참여하였다. SPSS version 17.0 (SPSS Inc., Chicago, USA)으로 각 문항의 평균, 표준편차를 산출하여 문항별, 범주별, 평가도구 전체의 내용타당도를 각각 분석해 보았다.

2) 구성타당도 검증

구성타당도는 임상실습 수업평가 도구가 임상실습 수업의 요소들로 구성되어 있는가를 검증하는 작업이다. 특히 평가도구가 몇 개의 범주와 그 범주 안에 포함된 세부 문항들로 구성되어 있을 경우, 요인분석이나 상관관계수 분석을 통해 구성타당도를 검증할 수 있다. 이 연구에서 개발한 평가도구도 사전에 이미 몇 개의 범주와 문항으로 분류되어 제작되었기 때문에, 구성타당도 분석을 위해서 확증적 요인분석(confirmatory factor analysis)을 실시하였다. 통계 프로그램으로는 LISREL 8.8 프로그램(Scientific Software International, Chicago, USA)을 사용하였으며, 적합지수(goodness of fit index)와 수정모형 적합지수(adjust goodness of fit index) 분석을 통해 설문지에서 연구자가 임의적으로 설정해 놓은 문항들의 범주가 적절한지를 통계적으로 검증해 보았다.

3) 신뢰도 검증

신뢰도란, 평가도구가 평가하고자 하는 것을 얼마나 오차

없이 정확하고 일관성 있게 평가하고 있는가를 검증하는 것으로서, 평가결과 간의 상관관계 정도나 평가문항에 대한 응답자 반응의 일관성으로 추정하는 것이 일반적이다. 신뢰도의 종류에는 검사-재검사 신뢰도, 동형검사 신뢰도, 내적 일관성 신뢰도 등이 있으나, 이 연구에서는 SPSS 17.0 프로그램으로 Cronbach α 계수를 산출한 뒤 문항 간의 일관성을 측정하는 내적 일관성 신뢰도를 중심으로 평가도구의 신뢰도를 분석해 보았다.

4) 적합도 검증

적합도는 어느 문항이 문항반응이론(item response theory) 모형에 잘 부합하는지를 통계적으로 검증해 보는 것으로서, 문항 적합도를 분석해 주는 통계 프로그램으로는 BILOG-MG, WINSTEP, MULTILOG, TESTFACT 등이 있다. 문항반응이론은 고전검사이론과는 달리, 일차원성(하나의 검사는 하나의 특성만을 측정)과 지역독립성(한 문항에 대한 응답이 통계적으로 독립적)을 기본 가정으로 삼기 때문에, 검사 집단의 특성과는 상관없이 문항 고유의 특성을 측정할 수 있다는 장점을 지닌다. 따라서 이 연구에서도 문항반응이론에 기초한 라쉬(Rasch) 모형을 적용하여 임상실습 수업평가 문항의 적합도를 분석해 보았다. WINSTEP version 3.65 프로그램(Winsteps, Chicago, USA)을 사용하여 내적합도(infit)와 외적합도(outfit) 지수, 점이연 측정상관계수(PTMEA correlation)를 산출한 다음, 내적합도와 외적합도 지수는 0.5~1.5 사이를, 적합도 제곱 평균치는 1.4 이하를, 점이연 측정상관계수는 0.4 이상을 만족시키는 경우를 적합한 문항으로 판별하였다.

결과

1. 평가도구 개발 결과

선행연구들에 대한 고찰과 개별 면담을 통해 개발된 임상실습 수업평가 도구는 총 4개의 대범주와 25개의 소범주, 43개 문항으로 이루어져 있으며, 5점 척도의 객관식 41문항과 자유 기재식의 주관식 2문항으로 구성되어 있다. Table 1에서 보는 바와 같이, 4개의 대범주는 A. 계획과 준비, B. 학습 환경,

Table 1. Clinical Teaching Evaluation Tool Developed by Kyungpook National University School of Medicine

A. Plan and preparation	
Preparation	1. The professor prepared for the clinical practice class thoroughly.
Medical knowledge	2. The professor knew the literature and latest trends related to his specialty. 3. The professor had broad medical knowledge beyond his specialty. 4. The professor lectured on the clinical practice content by connecting it with other subjects.
Clinical competence	5. The professor demonstrated clinical competence such as clinical reasoning skills, diagnostic competence, technical expertise, and managing patients.
Role model	6. The professor became a role model in fields such as patient-doctor relationship, communication skills and leadership. 7. The professor showed passion and enthusiasm for medicine.
Syllabus	8. The clinical practice protocol and syllabus were written in a specific and systematic manner. 9. The practice schedule was organized efficiently.
Orientation	10. The learning objectives were described with specific and clear terms. 11. The professor provided a sufficient length of orientation.
B. Learning climate	
Interaction and respect	1. The professor listened to the students' opinions and respected them during the practice period.
Learning climate	2. The professor promoted a comfortable atmosphere during the practice period. 3. I was able to join the practice without any stress or pressure during the practice period.
Encouragement of learning motivation	4. The professor motivated the students to have an interest of clinical practice. 5. The professor facilitated the students' self-directed study and practice. 6. The professor emphasized the students' active and voluntary participation.
C. Instruction	
Consistency with guideline	1. The professor proceeded with the clinical practice as written in guidelines. 2. The content of clinical practice was consistent with the learning objectives.
Learning objectives	3. The professor periodically mentioned the learning objectives during the practice.
Methods of practice	4. The professor used various practice methods appropriately, such as outpatient department observations, skill experiences, seminars, rounding, and PBL/PS etc. 5. There were enough opportunities to experience a variety of clinical cases. 6. There were enough opportunities to practice clinical skills with actual patients or standard patients (SP). 7. There were enough opportunities to see the professors' or residents' clinical skills.
Media and materials	8. The professor used various media and teaching materials. 9. Practice equipments were used appropriately (e.g., Simman model, clinical skill lab, etc.).
Teaching strategy 1): delivery	10. The professor provided a clear and detailed explanation during the clinical skill demonstrations. 11. The professor gave easy explanations for important content.
Teaching strategy 2): question and feedback	12. The professor answered the students' questions accurately. 13. The professor often provided immediate feedback during the practice period. 14. The professor gave specific feedback on the students' mistakes and fixed them.
Understanding of students	15. The professor carried out the practice based on the level of the students' understanding and skills.
Enthusiasm	16. The professor had a passion and enthusiasm for clinical teaching.
D. Evaluation	
Content of evaluation	1. The knowledge, skill, and attitudes of the students were evaluated evenly. 2. The evaluation contents reflected the content learned in practice.
Prior notice of evaluation	3. The evaluation methods, criteria, or checklists were noticed beforehand.
Evaluation criteria	4. The evaluation criteria were fair and objective.
Objectiveness of implementation	5. The evaluation was implemented reasonably and objectively.
Assignment	6. The amount of assignments and practice reports was appropriate.
Level of satisfaction	7. I am satisfied with the clinical practice course as a whole
Self-evaluation	8. I participated in the clinical practice with enthusiasm.
Overall evaluation (Subjective item)	9. What was especially good in this clinical practice? 10. What should be improved in this clinical practice?

Table 2. Results of Analysis on Content Validity (Item Importance Mean), Reliability (Cronbach α), and Fitness (Infit, Outfit, PIMEA Correlation) of the Clinical Teaching Evaluation Tool

Item category	Item no.	Item importance mean	Cronbach α	Infit		Outfit		PIMEA correlation
				MINSQ	ZSTD	MINSQ	ZSTD	
A. Plan and preparation	A1	4.10	0.968410	1.10	1.0	1.10	1.0	D 0.62
	A2	4.23	0.968488	1.02	0.3	0.94	-0.5	F 0.63
	A3	3.76	0.968348	1.06	0.6	1.09	0.9	E 0.69
	A4	3.76	0.968386	1.20	2.0	1.20	2.0	A 0.66
	A5	4.29	0.968515	1.00	0	0.95	-0.4	e 0.61
	A6	4.13	0.968135	0.88	-1.2	0.86	-1.3	c 0.69
	A7	4.12	0.968226	0.93	-0.7	0.96	-0.4	d 0.66
	A8	4.10	0.967829	0.72	-2.9	0.72	-3.0	a 0.75
	A9	4.03	0.968171	1.06	0.6	1.14	1.3	C 0.66
	A10	3.94	0.968087	0.83	-1.8	0.81	-2.0	b 0.74
	A11	3.80	0.968186	1.13	1.3	1.18	1.8	B 0.67
B. Learning climate	A	4.02	0.8968	0.99	-0.1	1.00	-0.1	
	B1	4.08	0.967996	0.88	1.2	0.90	1.0	a 0.75
	B2	3.86	0.968179	0.90	-1.0	0.87	-1.3	b 0.77
	B3	3.71	0.968699	1.13	1.4	1.13	1.3	A 0.74
	B4	4.11	0.968322	0.96	-0.4	0.98	-0.2	c 0.72
	B5	4.07	0.968204	1.03	0.4	1.03	0.3	C 0.71
	B6	4.19	0.968280	1.06	0.6	1.07	0.7	B 0.68
C. Instruction	B	4.00	0.8387	0.99	0	1.00	0	
	C1	3.96	0.968436	1.33	3.0	1.26	2.4	B 0.60
	C2	3.96	0.968070	1.09	0.9	1.09	0.9	E 0.67
	C3	3.63	0.968539	1.48	4.5	1.50	4.6	A 0.63
	C4	4.08	0.968159	0.95	-0.4	0.97	-0.2	H 0.68
	C5	4.10	0.968363	1.14	1.5	1.08	0.8	D 0.64
	C6	3.92	0.968241	1.17	1.7	1.27	2.3	C 0.65
	C7	4.04	0.967930	0.87	-1.2	0.87	-1.2	d 0.70
	C8	3.82	0.967953	0.94	-0.5	0.97	-0.2	G 0.71
	C9	3.92	0.967698	0.70	-3.0	0.69	-3.2	a 0.76
	C10	4.12	0.968046	0.85	-1.6	0.93	-0.5	h 0.69
	C11	4.18	0.968095	0.87	-1.3	0.84	-1.3	e 0.67
	C12	4.21	0.967968	0.78	-2.3	0.80	-1.8	b 0.70
	C13	4.00	0.967852	0.82	-2.0	0.82	-1.8	c 0.72
	C14	4.06	0.967806	0.85	-1.4	0.89	-1.0	f 0.70
	C15	4.04	0.967880	0.92	-0.8	0.88	-1.2	g 0.70
C16	4.08	0.967914	1.04	0.4	0.96	-0.3	F 0.65	
D. Evaluation	C	4.01	0.9372	0.99	-0.2	0.99	-0.1	
	D1	4.18	0.968233	0.96	-0.3	0.91	-0.8	c 0.73
	D2	4.15	0.967951	0.72	-3.2	0.69	-3.4	a 0.79
	D3	4.01	0.968252	1.10	1.0	1.24	2.1	A 0.70
	D4	4.17	0.968025	0.80	-2.2	0.75	-2.3	b 0.76
	D5	4.16	0.968203	0.92	-0.8	1.06	0.6	e 0.73
	D6	3.88	0.968258	1.08	0.8	1.08	0.8	E 0.73
	D7	3.87	0.967986	1.03	0.3	1.03	0.3	d 0.74
	D8	4.13	0.968267	1.22	2.0	1.18	1.6	B 0.67
	D9	4.02	0.968178	1.08	0.9	1.07	0.7	D 0.72
D10	4.14	0.968215	1.08	0.9	1.04	0.4	C 0.70	
	D	4.07	0.9133	1.00	-0.1	1.01	0	
Total (A+B+C+D)		4.03	0.9689	0.99	-0.1	1.0	0	

C. 수업, D. 평가로서, 수업을 계획하고 준비하는 초기 단계에서부터 마지막 평가의 단계에 이르기까지 수업의 전체적인 과정을 평가할 수 있도록 수업의 요소들을 순차적으로 배열시켜 놓았다. 4개의 대범주는 다시 몇 개의 소범주로 세분화되어 있는데, A. 계획과 준비는 수업준비, 의학지식, 임상능력, 역할모델, 수업계획서, 오리엔테이션이라는 6개의 소범주로, B. 학습 환경은 상호작용과 존중, 학습 분위기, 학습동기 유발이라는 3개의 소범주로, C. 수업은 임상실습 지침서와의 일치성, 학습목표, 교육(실습)방법, 수업매체 및 자료, 수업기법 1), 수업기법 2), 학생에 대한 이해, 수업에 대한 열정이라는 8개의 소범주로, D. 평가는 평가내용, 평가의 사전 공지, 평가기준, 평가시행의 객관성, 과제물, 만족도, 학생 자기평가, 전반적 평가(주관식)라는 8개의 소범주로 각각 구성되어 있다.

2. 평가도구 검증 결과

1) 내용타당도

임상교수와 전임의, 전공의, 3, 4학년 학생들을 대상으로 수업평가 문항의 중요도를 5점 척도로 조사한 결과에 따르면, 수업평가 도구의 대범주별 평균은 각각 A=4.02, B=4.00, C=4.01, D=4.07로 나타났으며, 43개 평가문항 전체의 평균은 4.03으로 나타나 평가도구가 임상실습 수업에서 평가하고자 하는 내용을 적절하게 평가하고 있는 것으로 확인되었다 (Table 2). 또한 임상실습 수업평가 내용으로서 중요도가 높게 나타난 문항은 A5. 교수는 임상교육자로서 가져야 할 충분한 임상능력을 겸비하고 있었다(m=4.29), A2. 교수는 자신의 전공분야에 대한 다양한 지식과 최신 동향들을 알고 있었다 (m=4.23), B6. 교수는 학생들의 적극적이고 능동적인 참여를 강조하였다(m=4.19), C11. 교수는 중요한 내용을 알기 쉽게 설명해 주었다(m=4.18), D1. 교수는 학생들의 지식, 술기, 태도를 고르게 평가하였다(m=4.18) 등이며, 반면에 B3. 나는 실습기간 동안 스트레스나 압박감 없이 실습에 참여할 수 있었다(m=3.71)와 C3. 교수는 학습목표를 주기적으로 반복하여 말해 주었다(m=3.63)와 같은 문항은 상대적으로 중요도가 떨어지는 것으로 나타났다.

2) 구성타당도

4개의 대범주와 25개 소범주, 43개 문항으로 구성되어 있

는 평가도구의 구성타당도를 확증적 요인분석을 통해 검증해 본 결과, 적합지수와 수정모형 적합지수가 각각 0.69와 0.66

Table 3. Results of Confirmatory Factor Analysis on Items of the Clinical Teaching Evaluation Tool

Item no.	A. Plan and preparation	B. Learning climate	C. Instruction	D. Evaluation
A1	0.60			
A2	0.60			
A3	0.64			
A4	0.29	0.35		
A5	0.60			
A6	0.69			
A7	0.67			
A8	0.78			
A9	0.68			
A10	0.75			
A11	0.67			
B1		0.74		
B2		0.73		
B3		0.61		
B4		0.64		
B5		0.65		
B6		0.63		
C1		0.60		
C2	0.67			
C3		0.57		
C4			0.67	
C5			0.61	
C6			0.64	
C7			0.74	
C8		0.25	0.52	
C9			0.82	
C10			0.76	
C11			0.74	
C12			0.77	
C13			0.79	
C14			0.80	
C15			0.51	0.29
C16	0.46			0.32
D1				0.80
D2				0.84
D3				0.69
D4				0.86
D5				0.83
D6		0.36		0.38
D7		0.50		0.28
D8		0.47		0.21
D9		0.44		0.26
D10			0.41	0.26
Goodness of fit index (GFI)		0.69		
Adjusted goodness of fit index (AGFI)		0.66		
Root mean square residual (RMR)		0.059		

으로 나타나 평가도구의 구성타당도가 보통보다 조금 높은 것으로 확인되었다. 그러나 Table 3에서 보는 바와 같이, 총 43개의 문항 중 8개 문항이 두 개의 요인에 동시에 적재되었으며, 2개 문항은 다른 요인에 소속되는 결과를 나타내 보여 이러한 문항들은 범주와의 관계 설정에 조정이 필요한 것으로 판단되었다.

3) 신뢰도

새롭게 개발한 임상실습 수업평가 도구의 신뢰도를 분석해 본 결과는 Table 2의 Cronbach α 계수로 제시되어 있다. 일반적으로 Cronbach α 계수가 0.7 이상이면 평가도구의 신뢰도가 양호한 것으로 해석하고, 0.8~1.0 사이는 신뢰도가 매우 높은 것으로 판단하는데, 이 연구에서 개발한 임상실습 수업평가 도구의 Cronbach α 계수는 0.9689로 나타나 수업평가 문항 간의 내적 일치도는 매우 높은 것으로 검증되었다. 수업평가 도구의 4개 범주별 Cronbach α 계수도 각각 A=0.8968, B=0.8387, C=0.9372, D=0.9133으로 나타나, 각 범주별 평가 문항들 간의 내적 일치도 역시 높은 것으로 밝혀졌다.

4) 적합도

문항반응이론에 근거하여 각 평가문항들의 적합도를 분석해 본 결과, C3번 문항이 내적합도와 외적합도 지수가 각각 1.48과 1.5로 나타났고 제곱 평균치도 1.4 이상이 되어 적합도가 떨어지는 문항으로 판명되었다. 그러나 C3번을 제외한 모든 문항의 내적합도 지수는 0.76~1.23, 외적합도 지수는 0.75~1.4로 나타나 0.5~1.5의 범위 내에 위치하고 있었고, 점이연 측정상관계수의 값도 모두 0.4 이상을 만족시켜 C3번을 제외한 모든 문항이 문항반응이론 모형에 적합한 것으로 확인되었다. 따라서 이 연구에서 개발한 43개 수업평가 문항 중 총 42개의 문항이 문항반응이론의 기본가정인 일차원성과 지역독립성을 충족시키고 있는 것으로 판별되었다.

고찰

수업평가는 의학교육자들이 자신의 교육자적 자질과 수업능력을 검증해 보고 수업 개선을 위한 정보를 얻을 수 있는 유용한 수단이다. 그러나 현재 의과대학 및 의학전문대학원에서 시행되고 있는 수업평가는 교수 개개인의 수업에 대한

평가라기보다는 전체 강좌 단위의 평가가 대부분이어서, 교수들은 강좌평가 평점으로 전달되는 평가결과를 통해 유용한 정보를 얻기 어렵다. 이 연구는 수업평가가 수업 개선을 위한 교육적 수단으로 활용되기 위해서는 타당성과 신뢰성을 갖춘 평가도구가 전제되어 있어야 한다고 판단하고, 임상실습 수업평가를 위한 평가도구를 개발하여 이 도구의 적합성을 다각도로 검증해 보고자 하였다. 연구결과, 4개의 대범주와 25개 소범주, 43개 문항으로 구성된 평가도구가 최종적으로 개발되었으며, 이 도구는 구성타당도를 제외한 내용타당도 및 신뢰도, 적합도 분석에서 모두 만족할 만한 결과를 나타내 보였다. 이러한 연구결과를 토대로 수업평가 및 의학교육 개선을 위한 몇 가지 제언을 해 보고자 한다.

먼저 이 연구에서 임상실습 수업평가 도구를 개발하기 위해 가장 주안점을 둔 부분은 선행연구들을 토대로 한 이론적 고찰과 분석의 과정이었다. 임상실습 수업평가 도구가 내용타당도를 지니기 위해서는 임상실습 수업의 핵심 요인들을 제대로 평가하고 있어야 하는데, 그러기 위해서는 효율적인 임상수업의 요인들을 추출해 내는 것이 가장 시급한 작업이라고 판단하였기 때문이다. 외국의 경우에는 수업평가 준거를 추출하기 위한 기초연구의 일환으로 바람직한 임상교수의 자질이나 효과적인 임상수업의 요인을 탐색하는 연구들이 지속적으로 이루어져 오고 있는 데 반하여[8,12,14,15,16,17], 국내에서는 강의식 수업에서 좋은 수업의 요건을 분석한 연구[18]를 제외하고는 이러한 유형의 연구들이 전무한 상태이다. 물론 의과대학의 여건상, 수업평가 설문지 하나를 개발하기 위하여 장기간에 걸친 이론적 고찰의 과정을 거치고 많은 관계자들의 의견을 수렴하는 것은 결코 쉬운 일이 아니다. 그러나 임상실습 수업평가에서 무엇을 평가할 것인지에 대한 충분한 논의 없이 타 영역에서 사용되는 설문지를 차용하여 임상실습 수업평가를 실시하고 있다면, 이는 임상수업의 핵심요인은 간과한 채 부수적인 요인들만 평가하는 것이나 다름없으며, 수업평가를 정기적으로 실시하면서도 교수들에게 유용한 정보를 제공하지 못했던 이유도 이와 무관하지 않을 것이다. 임의적으로 개발된 설문지보다는 이론적 고찰과 체계적인 분석을 통해 개발된 설문지가 내용타당도를 지닐 가능성도 높으며, 의학교육 연구가 상식에 기초한(commonsense-based) 연구가 아니라 이론에 근거한(theory-based) 연구가 되어야 하는 이유도 바

로 여기에 있다. 이 연구에서 임상실습 수업평가 문항에 포함될 요소들을 결정하기 위해 장기간에 걸친 문헌고찰과 사전연구, 설문조사와 같은 다양한 방법을 동원한 것도 결국 평가도구의 내용타당도를 높이려는 노력과 직결된 것이었다.

둘째, 이 연구에서 개발한 임상실습 수업평가 도구는 선행 연구를 통해 추출해 낸 임상실습 수업의 다양한 요인들로 구성되어 있으며, 이러한 요인들의 타당도는 교수, 학생, 전공의 등을 대상으로 한 설문조사에서도 확인된 바 있다. 연구결과에서 나타난 바와 같이, 임상실습 수업이 다양한 요인들로 구성되어 있다는 것은 임상교육자가 임상실습 수업에서 얼마나 많은 요인들을 체계적으로 고려해야 하는가를 간접적으로 말해 주고 있다. 스승에서 제자로 이어지는 전통적인 도제식 임상실습 수업의 경우, 교수의 수업능력이나 가르치는 기술은 크게 문제가 되지 않았다. 교수는 자신의 지식이나 기술을 있는 그대로 재현해 주면 되었고, 학생들은 관찰과 수련을 통해 이를 전수받기만 하면 교육이 성립되었다. 그러나 이 연구에서 추출해 낸 임상실습 수업의 요인들 가운데는 수동적이고 간접적인 참여에 의한 도제식 수업과는 다소 상반되는 요인들, 예컨대 학생들의 자기 주도적인 학습(B5)이나 학생들의 능동적이고 자발적인 참여(B6)와 같은 요인들이 다수 포함되어 있었다. 이는 곧 임상교육자들이 자신의 기술을 재현해 주는 전통적인 역할에서 한걸음 더 나아가 효율적인 임상실습의 요인들이 무엇인지를 구체적으로 인식하고, 이를 자신의 수업에 적극 반영하기 위해 노력해야 함을 의미한다. 전통적인 도제식 수업보다 체계적 요인에 따른 표준화된 수업을 통해 학생들의 술기가 향상되었음을 경험적으로 입증한 연구들[19,20]은 임상실습 수업에서도 효율성의 요인이 분명 존재하며, 또 체계적으로 분석될 수 있는 가능성을 시사해 준다. 따라서 후속연구에서는 바람직한 임상교수의 특징이나 임상실습 수업의 효율성 요인들에 대한 경험적, 실증적 연구가 더욱 다양하게 이루어질 필요가 있다.

셋째, 임상실습 수업평가 내용의 중요도를 설문 조사한 결과에 따르면, A5. 교수는 임상교육자로서 가져야 할 충분한 임상능력을 겸비하고 있었다(m=4.29), A2. 교수는 자신의 전공분야에 대한 다양한 지식과 최신 동향들을 알고 있었다(m=4.23), B6. 교수는 학생들의 적극적이고 능동적인 참여를 강조하였다(m=4.19), C11. 교수는 중요한 내용을 알기 쉽게

설명해 주었다(m=4.18), D1. 교수는 학생들의 지식, 술기, 태도를 고르게 평가하였다(m=4.18), D4. 평가준거는 공정하고 객관적이었다(m=4.17) 등의 문항이 임상실습 수업평가 문항으로서 중요도가 높은 것으로 나타났다. 이들 문항 가운데서 일부(C11, D4)는 이미 강의식 수업에서도 중요한 요인들로 밝혀진 것들이지만[18], 나머지 문항들은 '임상실습 수업'에서 중요하게 평가되어야 할 요인으로 새롭게 제안된 것들이라는 점에서 차별성을 지닌다. 이러한 연구결과를 통해 우리는 '좋은 수업'이 되기 위한 보편적인 요건도 존재하지만 가르치는 내용이나 과목에 따라 '좋은 수업'의 요건이 조금씩 달라질 수도 있음을 유추해 볼 수 있다. 예컨대 철학 수업과 공학 수업이 다를 것이며, 기초의학을 강의하는 강의식 수업과 임상실습 수업이 동일하지는 않을 것이다. 수업유형별 특성을 무시한 채 동일한 설문지로 각 수업들을 평가하는 것은 국어 수업 평가지로 수학 수업을 평가하는 것과 크게 다르지 않다. 따라서 수업평가 도구가 내용타당도를 지니기 위해서는, 기초의학과 임상의학, 문제바탕학습, 인문사회의학 등 각 수업유형에서 다루어야 할 핵심적인 특성에 근거하여 평가도구가 개발되고 시행되어야 할 것으로 판단된다.

넷째, 수업평가 도구의 구성타당도를 검증해 본 결과, 10개 문항(A4, C2, C3, C8, C16, D6, D7, D8, D9, D10)이 범주와의 관계 설정에 조정이 필요한 것으로 확인되었다. 이 연구에서는 평가도구 개발 초기부터 4개의 대범주와 25개 소범주로 구분하여 평가문항을 제작하였기 때문에, 이렇게 구분된 범주와 문항 간의 관계가 타당한지를 통계적으로 검증해 보기 위한 방법으로 확증적 요인분석을 실시하였으나, 몇 개 문항이 사전에 소속된 범주와는 다른 범주에 적재되는 결과를 나타내 보였다. 따라서 후속연구에서는 사전에 범주를 미리 구분하지 않고 43개 문항 전체를 탐색적 요인분석(exploratory factor analysis)을 통해 검증한 다음, 각 문항들이 어떤 범주로 새롭게 묶일 수 있는가를 분석해 보고자 한다.

다섯째, 이 연구에서는 Cronbach α 계수를 산출하여 문항 간의 내적 일관성을 검증해 봄으로써, 새롭게 개발한 임상실습 수업평가 도구의 신뢰도를 분석해 보고자 하였다. 분석결과에 따르면, 임상실습 수업평가 도구의 Cronbach α 계수가 0.9689로 나타나 수업평가 문항 간의 내적 일치도는 매우 높은 것으로 밝혀졌다. 그러나 Cronbach α 계수의 경우, N

(설문 항목)값이 커질수록 분산은 감소하고 $N/(N-1)$ 이 증가하여 자연스럽게 신뢰성이 높다고 나올 가능성이 크기 때문에, Cronbach α 계수만을 근거로 평가도구의 신뢰도를 확증하기에는 부족함이 있을 수도 있다. 따라서 보다 정확한 신뢰도 추정을 위해서는 Revelle's β , McDonald's ω 나 일반화 가능도 계수와 같은 또 다른 신뢰도 분석의 틀을 사용하거나 문항반응이론에서의 정보함수 등을 이용하여 평가도구의 신뢰도를 다각도로 검증해 보려는 노력이 추가적으로 이루어져야 할 것이다.

여섯째, 기존의 평가연구들이 고전검사이론에 입각한 평가 도구 검증에 집중되어 있었던 반면에, 이 연구에서는 문항반응이론에 기초한 적합도 분석을 통해 평가도구의 질을 검증해 보고자 하였다. 고전검사이론은 검사 총점에 의해 평가도구를 분석하는 방법으로서 검사 집단에 따라 결과가 달라지는 단점이 있지만, 문항반응이론은 검사 집단의 특성과는 상관없이 문항 고유의 특성을 측정하기 때문에 문항의 특성을 일반화할 수 있다는 장점을 지닌다. 따라서 이 연구에서 적합성을 인정받은 42개 문항(C3번 제외)들은 다른 의과대학이나 의학전문대학원의 임상실습 수업평가 문항으로 사용되어도 무방하며, 그럴 경우에도 문항의 적합성은 그대로 유지되는 것으로 판단하면 된다. 또한 문항반응이론에 근거하여 평가 문항을 검증할 경우에는 각 문항별 특성이나 정보의 비축이 용이하기 때문에, 수업평가 문항들만을 모아 놓은 문제은행 운영도 고려해 볼 수 있다. 문제은행에 비축된 수업평가 문항들 가운데서 비슷한 특성의 평가문항을 교대로 선택하여 설문지를 정기적으로 새롭게 재구성함으로써, 학생들이 반복되는 수업평가 문항에 대한 식상함으로 불성실한 응답을 할 가능성을 훨씬 감소시킬 수도 있다.

마지막으로, 이 연구에서 개발한 평가문항들은 임상실습 수업의 핵심 요인들을 추출하여 개발된 것으로서, 강의식 수업과 임상실습 수업의 구분 없이 사용되던 기존의 평가도구에 비해 임상실습 수업평가를 위한 적합한 도구라고 볼 수 있다. 그러나 이 평가도구 역시 강의나 문제바탕학습과 같은 모든 수업 영역을 평가하기 위한 도구로 설계된 것이 아니므로, 임상실습 수업을 평가하는 데 제한하여 그 결과를 해석하고 활용해야 할 것이다. 또한 이 연구에서는 전공과를 초월하여 일반적으로 활용될 수 있는 임상실습 수업평가 문항 개발에

초점을 맞추었기 때문에, 구체적인 실습과(예: 내과, 외과, 소아과 등)의 특수성을 충분히 포함시키지 못했을 수도 있다. 따라서 의학교육 현장에서는 이 연구에서 제시한 평가문항들을 임상실습 수업평가를 위한 일반적인 도구로 사용하되, 각 과별 특수성에 따른 평가문항은 추가로 포함시켜 융통성 있게 활용할 수 있기를 기대한다.

Acknowledgements: None.

Funding: This study was supported by the research encouragement award at the 25th Medical Education Conference in 2009 funded by the Korean Society of Medical Education.

Conflicts of interest: None.

REFERENCES

1. Yoo TW, Song YM. A student's evaluation of educational behavior of clinical preceptors in a teaching hospital. *Korean J Med Educ* 1991; 3: 57-65.
2. Yu SH. The contents and level of clinical education in medical schools. *Korean J Med Educ* 1992; 4: 55-68.
3. Kim YI, Lee YD, Oh JH, Choi IS, Chah H, Lee ED, Ham BJ, Lee K, Lee BK, Im YM, Park GH. Development of an evaluation baseline module for clinical clerkship: Gachon Medical School experience. *Korean J Med Educ* 2003; 15: 131-141.
4. Park JH, Kim S. Evaluating clinical clerkship in medical school. *Korean J Med Educ* 2004; 16: 157-167.
5. Ohr H, Yang EB, Chung MH, Lee MS. The study on the faculty evaluation system of teaching ability in Korea. *Korean J Med Educ* 1999; 11: 297-312.
6. Hwang K, Lee YM, Baik SH. Clinical performance assessment as a model of Korean medical licensure examination. *Korean J Med Educ* 2001; 13: 277-287.
7. Yang E, Suh DJ, Lee Y, Lee S, Kim S, Lee E, Chae G,

- Jo Y, Ahn D. Status of clerkship education and its evaluation in Korean medical schools. *Korean J Med Educ* 2007; 19: 111-121.
8. Graffam B, Bowers L, Keene KN. Using observations of clinicians' teaching practices to build a model of clinical instruction. *Acad Med* 2008; 83: 768-774.
 9. Stronge JH. *Evaluating teaching: a guide to current thinking and best practice*. Thousand Oaks, USA: Corwin Press; 1997. p 21-38.
 10. Won HH. *Understanding and application of teaching evaluation*. Seoul, Korea: Gyoyook Gwahaksa; 2002. p 8-132.
 11. Kwak YS. *Teacher evaluation, teaching evaluation*. Seoul, Korea: Wonmisa; 2005. p 20-338.
 12. Chandran L, Gusic M, Baldwin C, Turner T, Zenni E, Lane JL, Balmer D, Bar-On M, Rauch DA, Indyk D, Gruppen LD. Evaluating the performance of medical educators: a novel analysis tool to demonstrate the quality and impact of educational activities. *Acad Med* 2009; 84: 58-66.
 13. Chae SJ, Lim KY. An analysis of course evaluation programs at Korean medical schools. *Korean J Med Educ* 2007; 19: 133-142.
 14. Skeff KM, Stratos GA, Berman J, Bergen MR. Improving clinical teaching: evaluation of a national dissemination program. *Arch Intern Med* 1992; 152: 1156-1161.
 15. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000; 75: 161-166.
 16. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998; 73: 688-695.
 17. Sutkin G, Wagner E, Harris I, Schiffer R. What makes a good clinical teacher in medicine? A review of the literature. *Acad Med* 2008; 83: 452-466.
 18. Im EJ, Lee YC, Chang BH, Chung SK. Investigation of the requirements of 'good teaching' to improve teaching professionalism in medical education. *Korean J Med Educ* 2010; 22: 101-111.
 19. McLeod PJ, Steinert Y, Trudel J, Gottesman R. Seven principles for teaching procedural and technical skills. *Acad Med* 2001; 76: 1080.
 20. Junger J, Schafer S, Roth C, Schellberg D, Friedman Ben-David M, Nikendei C. Effects of basic clinical skills training on objective structured clinical examination performance. *Med Educ* 2005; 39: 1015-1020.