의과대학 시험의 문항 특성 사이 상관관계

한림대학교 의과대학 기생충학교실 및 의학교육연구소

허 선

= Abstract =

Correlation among Item Characteristics of Examinations in College of Medicine

Sun Huh

*Department of Parasitology, College of Medicine and Institute of Medical Education,
Hallym University, Chuncheon 200-702, Korea*

After item analysis of examinations in College of Medicine, the correlation among characteristics were examined for the better understanding of their meaning. The 78 subjected examinations in College of Medicine, Hallym University, Korea from March 1999 to October 2000 were analyzed. Discrimination indexes (D) by the method of extreme group were positively correlated with item-total correlation (ITC) with mean correlation coefficient $r=0.8506$ ranged between 0.6430 and 0.9520. Number of items of each examination was positively correlated with Cronbach coefficient alpha reliability index ($r=0.7920$) whereas negatively correlated with standard deviation ($r=-0.5691$), odd-even split reliability index ($r=-0.8767$) and mean ITC ($r=-0.4079$). Thereafter, the standard deviation was positively correlated with odd-even split reliability index ($r=0.5072$) and mean ITC ($r=0.6166$). There was negative correlation between Cronbach coefficient alpha reliability index and odd-even split reliability index ($r=-0.7385$). Above results suggested that the number of items in each examination was most powerful factor affecting to other item analysis characteristics. The appropriate number of items should be considered for better result of item analysis characteristics. Odd-even split reliability index is not appropriate for the estimation of the reliability among item, since it decreased according to the increase of number of items. Positive and high correlation between D and ITC means that both methods are appropriate to interpretate the discriminating power of the items.

Key Words: Item analysis, Evaluation, Discrimination index, Reliability index

* 교신저자: 허선, 한림대학교 의과대학 기생충학교실

*이 연구는 한림대학교 의학교육연구재단 연구비의 지원을 받아 수행함(2000).

서 론

시험을 치르고 난 뒤, 시험의 난이도, 분별도, 문항추측도(오답분석)와 같은 문항분석 특성을 파악하면 앞으로 시험 문항을 개선하는 데 유용하게 활용할 수 있다. 이런 문항분석은 되먹이기에 활용하는 것 뿐 아니라 시험의 목적을 달성하였는지, 즉 평가하려고 한 것을 제대로 평가하였는지를 파악하는 데 유용하다. 현재 우리 나라 의학 교육에서 활용하는 문항분석의 특성은, 총점, 평균, 표준편차, 신뢰도, 난이도, 분별도, 답지 반응 분포, 문항추측도(예상오답선택수) 등이다. 이 특성 중 난이도, 분별도, 문항추측도는 대부분 고전검사이론에 따라 계산한다. 이 이론은 검사에 의한 관찰점수는 진점수와 오차점수에 의하여 합성된다고 가정한다. 또한 피험자의 진점수를 알 수 없기 때문에 이론적으로 동일검사를 동일 피험자에게 무한히 반복 실시하여 얻은 점수들의 평균점수로 추정한다. 이 고전검사이론에 따른 특성 계산은 이해하기 쉽고 계산도 쉽다는 장점이 있지만, 피험자 집단의 특성에 따라 특성 값이 바뀐다.¹ 피험자 집단에 따라 특성 값이 바뀌지 않는다는 것을 가정한 문항반응이론에 의한 계산법이 있으나, 대부분의 의대 시험에서는 계산하기 쉽고, 이해하기 쉬운 고전검사이론에 의한 해를 구하여 활용하고 있다. 이 해를 구할 때, 난이도 지수는 구하는 식에 이론의 여지가 없으나, 분별도 지수는 양극집단법과 문항총점상관계수 두 가지 방법으로 구할 수 있고, 신뢰도 역시 몇 가지 다양한 방법 중, 반분법과 크론바하 알파값을 구하는 내부신뢰도 검사법 두 가지를 사용할 수 있다. 이 고전검사이론에 의한 해는 이미 그 의미가 잘 알려져 있으나 우리나라 의학 교육 현장에서 어떤 값이 주로 나오는지, 그리고 그 값 사이에 상관관계가 있는 지, 또한 그 상관관계의 의미는 무엇인지에 대하여 많은 수의 시험 결과를 분석하여 자세히 밝힌 것이 드물다. 특성 사이의 상관관계에 대한 연구로는 분별도 계산에서 쓰이는 양극집단법 분별도지수와 문항총점상관계수 사이의 상관관계를 1996년도 시행한 네 종류의 전문의 시험 필답고사에서 분석하여 상관계수

0.68로 양의 상관관계가 있었다는 보고가 있다.² 그러므로 이번에는 한림대학교 의과대학에서 시행한 여러 시험의 답지에서 고전검사이론에 따른 문항분석 특성 값을 구한 뒤, 이런 여러 특성간의 상관관계를 살펴보았다. 상관분석 결과 유의한 상관관계가 있는 각 문항 특성 사이의 관계를 기술하고 여러 특성 사이의 상관관계의 의미가 무엇인지 찾아보려고 하였다.

연구 대상 및 방법

1999년 3월부터 2000년 7월까지 3학기 동안 한림대학교 의과대학에서 시행한 시험 중, 광관독지에 답을 표시하여 광관독기로 채점한 과목 78종을 대상으로 문항분석을 시행하였다. 의학과 임상 실습 과목 31종, 졸업시험 겸 의사국가고시 모의고사 성격으로 치르는 임상종합시험 19종, 기초임상 통합교육 23종, 그리고 기생충학 및 여성건강간호학, 재활 의학 등 기타 5종이었다. 이 시험 수는 일부 기초의학 과목을 제외한 대부분의 시험을 포함한다. 문항분석은 광관독기에서 아스키 파일로 읽은 결과를 Analyst for Web 3.0을 사용하여 분석하였다.³ 이 결과는 다시 윈도우용 통계 프로그램인 DBSTAT 3.0에 데이터베이스 파일로 변환하여 입력한 뒤, 상관 분석을 하였다.⁴

문항분석에서 구한 특성 및 그 계산법은 다음과 같다.^{2,5}

- ① 총점: 시험 문제의 총 수
- ② 평균: $m = \text{각 수험생의 점수의 합} / \text{문제수}$
- ③ 표준편차

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

σ = 표준편차

X = 각 수험생의 점수

$\bar{X}$ = 평균 점수

N = 수험생의 수

- ④ 홀짝반분법 신뢰도계수 (odd-even split method discrimination index): 시험문제를 홀수 짝수

그림 1. 78종 시험의 평균 난이도 (100점 만점 시 평균 성적)의 도수 분포.

문제로 나누어 시험치를 때, 얼마나 결과가 일치하는 지를 계산한다. 두 표본의 비율 상관관계분석을 사용하여 상관계수를 구한다.

- ⑤ 내부신뢰도 (크론바하 알파 계수, Cronbach coefficient alpha)

$$\alpha = \frac{\kappa}{\kappa - 1} \left(1 - \frac{\sum \rho_{ii}^2}{\rho_y^2} \right)$$

κ = 측정 항목수

ρ_{ii}^2 = 측정 항목의 총 분산

ρ_y^2 = 측정항목의 분산

- ⑥ 난이도 (Difficulty Index) (p):

p = 정답자수/응시자수

- ⑦ 양극집단법 분별도지수 (Discriminating Index by method of extreme group) (D): 상위군 27%, 하위군 27%에서 얼마나 맞았는 지를 계산한다.

$$\text{분별도 (D)} = \frac{(\text{상위군의 정답자 수}) - (\text{하위군 정답자수})}{(\text{상위군 응시자 수} + \text{하위군 응시자수}) / 2}$$

- ⑧ 문항총점상관계수를 이용한 분별도지수 (Item

total correlation)

$$r = \frac{\sum (z_x \times z_y)}{N}$$

$$z_x = \frac{X - \bar{X}}{\rho_x}$$

x = 어떤문항에서 얻은 점수

y = 시험의 점수

ρ = 표준편차

우선 양극집단법 난이도지수와 문항총점상관계수 사이의 상관관계 및 반분법신뢰도계수와 크론바하 알파 값 사이의 상관관계를 구하였다. 문항분석에서 나온 해를 가지고, 처리하여 상관분석을 한 변수는 다음과 같다.

- ① 시험 문항수

- ② 평균난이도- 100점 만점으로 환산할 때 평균 점수

- ③ 표준편차 -- 100점 만점으로 환산할 때 표준편차.

- ④ 크론바하 알파 신뢰도 계수

- ⑤ 홀짜반분법 신뢰도 계수

표 137. 문항분석에 의하여 구한 각 변수의 평균 및 최소, 최대값

변수	평균	최소값	최대값
문항수	57.0385	20	165
평균	68.0779	49.10	98.90
표준편차	8.7492	5.86	12.90
크론바하 알파	0.495	-0.11	0.92
홀짝반분법신뢰도	0.5888	0.29	0.93
양극단분별도평균	0.1297	0.0210	0.6260
문항총점상관계수평균	0.2201	0.1553	0.7000
양극단분별도-문항총점상관계수 사이의 상관계수	0.8506	0.6420	0.9520

표 2. 문항분석에 의하여 구한 각 변수 사이의 상관관계

	No.	Mean	SD	alpha	odd-even	D	ITC	R
No.	1.0000							
Mean	-0.1082	1.0000						
SD	-0.5691*	-0.2057	1.0000					
alpha	0.7920*	0.1145	-0.3181*	1.0000				
odd-even	-0.8767*	0.2477*	0.5072*	-0.7385*	1.0000			
D	-0.2615*	-0.3370*	0.3878*	-0.2625*	0.2223	1.0000		
ITC	-0.4079*	0.2483*	0.6166*	-0.0174	0.2965*	0.1338	1.0000	
R	-0.0004	-0.3491*	0.1064	-0.1192	-0.0391	0.1736	-0.1565	1.0000

숫 자: 두 변수간의 상관계수 (r) * 표시: $p < 0.05$

표본수: 78 자유도: $df = 76$ 유의수준 ($\alpha = 0.05$): $t = 1.9917$

No.: 문항수, Mean: 평균, SD: 표준편차, alpha: 크론바하 알파 odd-even: 홀짝반분법신뢰도,

D: 양극단분별도평균, ITC: 문항총점상관계수평균 R: 양극단분별도-문항총점상관계수 사이의 상관계수

- ⑥ 양극집단법 난이도 계수의 평균
- ⑦ 문항총점상관계 난이도 계수 평균
- ⑧ 양극집단법 난이도지수와 문항총점상관계수 사이의 상관계수

결 과

각 문제의 문항수는 실습 시험인 경우에는 20개의 적은 수가 많았고, 통합교육은 100개 전후, 그리

고 임상종합시험은 의사국가고시 문제수와 같은 20개에서 165개 사이로, 78종의 시험의 평균 문항수는 57개이고, 최소 20개, 최대 165개였다. 평균 난이도는 0.68로 최소 0.49, 최대 0.98로 매우 어려운 문제와 매우 쉬운 문제가 혼재하였다. 이 평균 난이도는 정규 분포를 보였다(그림 1). 표준편차는 100점 만점으로 환산 시 평균 8.7492로 5.86에서 12.9까지 차이가 있었다. 신뢰도 계수 중 크론바하 알파값은 평균 0.495, 홀짝반분법은 평균 0.5888이었다. 양극

그림 2. 문항수와 크론바하 알파 내부신뢰도계수의 상관관계. $r = 0.7920$, $p < 0.05$.

단 분별도의 평균은 0.1297, 문항총점상관계수의 평균은 0.2201 이었다. 각 문제의 양극단 분별도와 문항총점상관계수 사이의 상관관계의 평균은 0.8506 이고 최소값 0.6420, 최대값 0.9520 이었다(표 1). 신뢰도 계수 중 크론바하 알파값과 홀짜반분법 신뢰도계수와의 음의 상관관계가 있었다(표 2).

표 2의 변수 사이의 상관관계 중 p 값이 0.05 이하로 유의한 차이가 있는 것 중 r 값이 +0.4 이상 또는 -0.4 이하인 것을 비교적 높은 양과 음의 상관관계가 있는 것으로 보면, 시험 문항수가 크론바하 알파값과 양의 상관관계가 있고, 표준편차, 홀짜반분법 신뢰도계수 및 문항총점상관계수 평균과는 음의 상관관계가 있다(그림 2, 3). 또한 표준편차는 홀짜반분법신뢰도계수 및 문항총점상관계수의 평균과 양의 상관관계가 있다(그림 4, 5).

고 찰

한림대학교 의과대학에서 치른 78종 시험 문제의 평균 난이도는 0.68 이라는 것은 0.3~0.7을 허용범위로 함에 비추어 적절한 수준이다.⁵ 평균 난이도가

0.99로 매우 높아 너무나도 쉬운 문제로 대부분이 수험생이 만점을 받은 경우가 있어서 이런 시험은 문항이 이미 공개된 것을 그대로 출제하는 것일 가능성이 높으므로 점검하여 볼 필요가 있다. 신뢰도 중 크론바하 알파값은 평균 0.4950으로 -0.11부터 0.92까지 매우 낮은 문제부터 높은 문제까지 분포하였다. 홀짜반분법에 신뢰도계수는 0.5888이 평균으로 크론바하 알파값보다 높았고, 0.29~0.93의 범위였다.

분별도 중 각 시험의 양극단분별도 평균은 평균 0.1297, 문항총점상관계수 평균은 0.2201로, 모든 시험의 경우에서 문항총점상관계수의 평균값이 양극단분별도의 평균값보다 높았다. 각 시험의 양극단분별도와 문항총점상관계수 사이의 상관관계는 p 값이 0.0021, 0.0008, 0.002, 0.0003인 네 종을 제외하고 모두 0.0000으로 매우 유의한 양의 상관관계가 있었고, r 값의 평균은 0.8506으로 높은 상관관계를 보였는데, 이것은 황인홍(2000)의 전문의 시험 4과목에서의 r 값 0.68보다 높은 값이다. 의학 교육에서 난이도가 양극단에 있는 문항에 대한 분별도를 측정할 때는 문항총점상관계수를 구하는 것이 유리하나,

그림 3. 문항수와 홀짜반분법에 의한 신뢰도의 상관관계. $r = -0.8767$, $p < 0.05$.

대부분의 경우에 양극단분별도와 문항총점상관계수 어느 하나를 사용하여도 무방함을 할 수 있다.

각 변수간의 상관관계에서 크론바하 알파값이 시험 문항수와 양의 상관관계가 있고, 표준편차, 홀짜반분법 신뢰도계수 및 문항총점상관계수 평균과 음의 상관관계가 있다는 점이 흥미롭다. 내부신뢰도를 높이려면 문항수를 늘리고, 문항총점상관계수에 의한 분별도지수를 높이려면 문항수는 줄이는 게 필요함을 알 수 있다. 신뢰도 중 크론바하 알파를 구하는 내부 신뢰도는 문항수와 양의 상관관계가 있으므로 내부 신뢰도는 문항을 어떻게 개발하느냐 보다 문항수를 늘리면 상승한다(그림 2). 이 내부 신뢰도가 올라가므로 문항 개발에서 시간안정성, 문항표본집단의 적절성, 문항의 동질성과 같은 요인은 크게 중요하지 않아서 이 내부신뢰도는 문항분석의 특성으로서의 기능을 상실한 것인지, 그렇지 않다면 대부분의 문항수가 많은 문제에서는 문항이 문항모집단을 대표할 수 있는 적절히 선택한 문항이고, 문항간에 동질성이 매우 높았으며 문항수가 적은 문제에서는 그렇지 않았다고 해석할 수 있으나, 이 문제는 모든 문항에 대한 동질성에 대한 검토를 하여

야 하므로 설명하기가 쉽지 않다. 그러므로 이 내부 신뢰도는 문항 표본 집단의 적절성이나 문항의 동질성을 나타내는 지표로 쓰기에는 적절한지 더 연구할 필요가 있다. 또한 홀짜반분법에 의한 신뢰도와 시험 문항수는 음의 상관관계로 문항수가 적을수록 신뢰도가 올라간다(그림 3). 이 홀짜반분법 신뢰도계수가 크론바하 알파값과 정반대로 문항수가 적은 것은 문항모집단을 잘 대표하고, 문항간 동질성이 높았고, 문항수가 많은 것은 그 반대라는 가정이 이 78종의 시험에서 성립한다고 하기 전에는 별 의미가 없다. 즉, 문항수에 달려 있는 신뢰도는 그 결과를 해석하는 것이 큰 의미가 없다. 앞으로 이 의학교육의 문항분석에서 신뢰도의 해석은 다시 하여야 하고, 문항모집단의 대표성과 문항간 동질성을 측정할 수 있는 적절한 다른 도구를 개발하는 것이 필요하다. 또한 문항수와 각각 다른 두 측정방법의 신뢰도계수가 각각 양, 음의 상관관계를 보였으므로 두 가지 종류의 신뢰도계수 사이에도 음의 상관관계를 보인다. 내부 신뢰도가 신뢰도 추정에서 더 적절한 방법으로 생각한다면, 홀짜반분법 신뢰도 계산에 적절치 않다. 표준편차가 홀짜반분법신뢰

그림 4. 표준편차와 홀짝반분법 신뢰도와 상관관계. $r = 0.5072$, $p < 0.05$.

도와 양의 상관관계를 보이고(그림 4), 문항수와 음의 상관관계를 보이므로, 크론바하 알파값과는 낮은 정도의 음의 상관관계를 보인다. 표준편차가 문항수와 음의 상관관계를 보이므로 이 두 가지 방법의 신뢰도가 문항수와 상관관계가 있어 나타나는 결과이다. 표준편차는 문항수가 많을수록 적다는 것은 문항수가 많을수록 성적이 평균이 몰린다는 것으로 많은 수의 문항이 수험생의 공부한 내용을 광범위하게 측정하면 그 집단의 평균에 측정점수가 몰린다는 것이다. 같은 내용으로 무한대 또는 여러 번의 시험을 보아 측정하는 진점수에 측정점수가 더 가까워진다고 해석할 수 있다.

각 시험의 양극집단분별도 평균 및 문항총점상관계수의 평균 사이의 상관관계는 없으나, 문항총점상관계수 평균과 표준편차 사이에 0.6166이라는 비교적 높은 상관관계는 흥미로운 결과이다(그림 5). 즉 분별도지수의 측정에서 문항총점상관계수를 사용하면 표준편차가 큰 시험에서는 분별도가 커진다는 것으로 문항수가 표준편차 및 문항총점상관계수의 평균과 음의 상관관계이므로 두 변수가 양의 상관관계를 보일 것이라는 것은 쉽게 짐작할 수 있다.

이러한 문항분석 결과는 해석할 때 특별히 주의할 기울이지 않던 부분에 대한 흥미로운 결과를 제공한다. 즉, 더 많은 시험을 가지고 점검하거나, 의사국가시험 등과 같이 문항의 모집단의 표본성과 문항사이 동질성을 사전 검토하고 출제하는 문제를 대상으로 추가 작업이 필요하겠으나, 이 결과는 고전검사이론에 입각한 문항분석의 특성 값이 상당히 문항수와 양 또는 음의 상관관계가 있다는 점을 잘 보여 준다. 표준편차, 두 가지 방법의 신뢰도계수 및 문항총점상관계수를 이용한 분별도가 이런 상관관계가 있다면, 시험 문항수 결정이 이런 분석결과에 큰 영향을 미치므로, 우선 문항수가 어느 수준이 적정한지를 정하는 것이 우선일 것이다. 분별도를 높으려면 지나치게 많은 문항수는 도움이 되지 않는다는 것을 알 수 있다. 두 종류의 신뢰도계수는 음의 상관관계이고 문항수에 달려 있으므로 역시 문항수를 늘리면 내부신뢰도는 늘어나게 마련이다.

분별도 구하는 두 가지 방법인 양극집단분별도지수와 문항총점상관계수는 매우 상관관계가 높음을 다시 한번 확인할 수 있어 두 특성이 난이도가 양극인 경우와 지수의 해석에서 차이가 있으나 두 값 모

그림 5. 문항총점상관계수 평균과 표준편차 사이 상관관계. $r = 0.6166$ $p < 0.05$.

두 문항분석에 의미가 있음을 뜻한다.

이 연구에서는 많은 수의 시험을 문항 분석한 결과를 가지고 각 특성간의 상관관계를 구하여 분석하려고 노력하였으나, 각 시험이 다양한 과목에서 치러져서 의사국가시험과 같은 좀더 통제하고 동질성을 유지하려고 노력하려고 투자를 많이 하고, 많은 수의 학생이 동시에 치르는 시험의 분석결과와 이번 결과와 비교하여 어떤 변화가 있는 지 알아보는 것이 필요할 것이다. 또한 이 분석은 한 학년 입학 정원이 80명인 의과대학의 예이므로 학생수가 다른 대학에서의 경우와 차이가 있을 것이므로 다른 대학의 결과와 비교하여 보는 작업이 필요할 것이다.

요 약

시험의 문항분석 특성 값을 구한 뒤, 분별도지수 중 양극집단법과 문항총점상관계수 사이의 상관관계, 신뢰도의 계산에 쓰이는 반분법신뢰도계수와 크론바하 알파 값 사이의 상관관계 및 시험문항의 수와 다른 특성과의 상관관계를 알아보았다. 1999년 3

월부터 2000년 7월까지 한림대학교 의과대학에서 치른 시험 과목 중, 광판독기로 채점한 과목 78종을 대상으로 문항분석 후 나온 특성 값의 상관관계를 분석하였다. 각 문제의 양극단분별도와 문항총점상관계수 사이의 상관계수의 평균은 0.8506으로 양의 상관관계였고, 최소값 0.6420, 최대값 0.9520 이었다. 시험 문항수가 크론바하 알파값과 양의 상관관계가 있고, 표준편차, 홀짜반분법 신뢰도계수 문항총점상관계수 평균과는 음의 상관관계가 있다. 또한 표준편차는 홀짜반분법 신뢰도계수 및 문항총점상관계수 평균과 양의 상관관계가 있다. 신뢰도 계수 중 크론바하 알파값과 홀짜반분법 신뢰도계수와는 음의 상관관계가 있었다. 두 방법의 신뢰도 및 문항총점상관계수를 이용한 분별도가 문항수와 이런 상관관계가 있다면, 시험 문항수 결정이 분석결과에 큰 영향을 미치므로, 우선 문항수가 어느 수준이 적정할지를 정하는 것이 우선일 것이다. 문항표본집단의 적절성이나 문항간의 동질성을 높이는 데 문항수 증가가 가장 중요한 요인임을 알 수 있었다. 홀짜반분법 신뢰도계수는 문항수가 적을수록 높아지므로 이 계수는 신뢰도계수의 특성을 파악하는 데

소용이 적다. 양극단분별도와 문항총점상관계수 사이의 높은 양의 상관관계는 두 방법의 분별도 구한 특성이 모두 의미가 있음을 뜻한다.

참 고 문 헌

1. 성태제: 문항제작 및 분석의 이론과 실제, 서울: 학지사, 1996
2. 황인홍: 분별도 지표로서 문항총점 상관계수의 활용. 한국의학교육 12:45-51, 2000
3. 박흥규, 허선. *Analyst for Web 3.0*, 2000. <http://analyst.hallym.ac.kr> 에서 구할 수 있는 문항분석 풀그림
4. 김수녕: 1 주 완성 윈도우용 통계 소프트, 서울: 탐진, <http://www.dbstat.com>에서도 구할 수 있음, 2000
5. 김용일: 의학교육평가: 시험문항의 개발과 분석, 서울: 서울대학교 출판부, 1985