

문제은행 시스템의 문항 차분 난이도 및 변별도를 기반으로 한 문항 적합성 검증

전병호

공주대학교 영상정보공학부

Item Difference Difficulty & Item Discrimination based Item Suitability Verification for Test Bank System

Byeong-Ho Jeon

Division of Media Image Computer Information,
Kongju National University

요 약

문제은행 시스템은 피험자의 조건에 따라 데이터베이스에서 문항을 추출하여 가상공간에서 평가를 수행한다. 가상 공간에서의 평가는 피험자에게 적용하는 경우, 출제 빈도에 따라 문항의 난이도 및 변별력에 영향을 주게 된다. 출제 빈도에 따라 난이도나 변별력이 낮아지는 문항은 출제를 제한하는 기준이 필요하다. 본 논문에서는 문항 사후 난이도와 문항 변별도를 기반으로 하여, 문항 차분 난이도를 주기적으로 측정하고 난이도 차이가 일정 수준 이상이 되는 문항에 대해 출제를 제한하는 방안과, 전체 피험자에 대한 문항의 변별력을 측정하여 변별력이 떨어지는 문항을 출제자에게 문항을 수정하게 하거나 삭제하도록 하는 방안을 제안한다.

1. 서론

인터넷을 기반으로 하여 웹을 통해 문제를 제공하는 서비스가 늘어나고 있다. 웹 상의 문제은행 시스템은 가상시험의 형태로 존재하고 있다. 문제은행 시스템의 장점은 피험자의 평가 정보를 분석하여 피험자에게 적합한 평가를 개별적으로 수행할 수 있다. 수준별 학습을 하도록 하기 위해서는 우선 피험자의 정보를 가지고 있어야 한다. 또한 각 문항의 수준 정보를 가지고 있어, 피험자의 수준과 문항의 수준을 맞추어 평가함으로써 학업 성취도를 향상시킬 수 있도록 유도할 수 있다[1].

이러한 장점 때문에 웹 상의 문제은행 시스템이 속속 개발되어 가고 있다. 문제은행 시스템의 특징은 첫째, 언제, 어디서나, 그리고 누구나 문항을 만들어 지정된 문제은행시스템으로 전송하여 문제은행 데이터베이스 구축이 용이하다는 점이다. 물론 문항을 출제하는 사람은 제한적이지만, 자격을 가진 여러 사람이

분산환경에서 문항을 개발하게 입력하게 된다. 둘째, 문제은행시스템은 피험자의 조건에 따라 문항을 임의로 추출하여 출제하게 된다. 이 때 피험자의 조건은 여러 가지 상황이 될 수 있다. 예를 들어 수준별 학습을 위한 조건, 모의 시험을 위한 조건, 온라인 가상 시험을 위한 조건들이 있다. 셋째, 이러한 조건에 따라 추출된 문항은 그 출제 빈도에 따라 사후 난이도가 상당히 변하게 되고, 이에 따라 변별력이 영향을 받게 된다. 출제 빈도가 많은 문항은 피험자가 자주 접하게 되기 때문에 실제 난이도가 초기에 정한 난이도보다 낮아지게 된다. 이렇게 되면, 이 문제은행 시스템의 평가 신뢰도가 떨어지게 된다.

문항의 난이도 차이를 주기적으로 측정하여 난이도 차이가 일정 수준 이상으로 되는 문항은 출제를 제한할 필요가 있다. 전체 피험자에 대한 문항의 변별도를 측정하여 변별도가 떨어지는 문항은 출제자에게 문항을 수정하게 하거나 삭제하도록 하는 기준이 필요하다.

본 논문에서는 이러한 시스템의 특징과 피험자의 성향을 고려하여 각 문항의 난이도 차이와 문항 변별도를 적용하여 문항의 적합성을 주기적으로 검증하는 방안을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 문항 적합성에 대해서 설명 설명하고, 3장에서는 각 평가 지수들에 대한 검토 및 분석을 살펴본다. 마지막으로 4장에서는 결론으로 마무리한다.

2. 문항 적합성

문항의 적합성은 문항 차분 난이도와 문항 변별도에 의해 정해진다.

2.1 문항 차분 난이도

문항의 수준은 출제자에 의해 1차적으로 정해진다. 출제자의 판단에 의해 2가지 관점에서 문항의 수준을 결정한다. 출제자는 문항의 중요도와 사전난이도를 정함으로써 문항의 수준을 정한다.

문항의 중요도는 사전난이도와 다른 지수이다. 문항의 중요도는 난이도처럼 문항의 어렵고 쉬운 정도를 나타내는 것이 아니라 문항의 학습에 대한 비중의 정도를 나타내는 것이다.

문항 중요도(Importance Degree of Item)

문항의 중요도를 나타내는 지수 C 로 문항을 출제하는 출제자에 의해 정해지며, 단계는 $1 \dots n$ 까지이다.

예를 들어 문항의 중요도를 3단계로 할 경우, 문항 중요도 값의 범위는 $[1, 3]$ 이 되며, 값이 높을수록 중요도가 높음을 의미한다.

사전 난이도(Pre-difficulty Degree of Item)

문항의 쉽고 어려운 정도를 나타내는 지수로 사전 난이도와 사후 난이도로 나누어진다[2].

사전난이도는 문항의 어려운 정도를 초기에 정해지는 지수 D_{pre} 로, 문항을 출제할 때 출제자에 의해 정해지며, 단계는 $1 \dots n$ 까지 이다.

예를 들어 문항의 사전 난이도를 3단계로 할 경우, 사전난이도 값의 범위는 $[1, 3]$ 으로 값이 높을수록 문항이 더 어렵다.

- 문항 수준도(The Level of Item)

문항 수준도는 각 문항별로 가로에 사전난이도 등급을 나열하고, 세로에 문항 중요도를 나열한 후 표 1에 따라 문항 수준도를 산출한다. 문항성취도는 좌상에서부터 우하까지 대각선으로 차례로 등급을 부여한다.

<표 1> 문항 수준도 평가

사전난이도 \ 문항중요도	1	...	$d_{j,k}$...	M
1	1	2	3
...	2	3
$d_{i,k}$	3	...	$L_{i,j,k}$...	$h-2$
...	$h-2$	$h-1$
N	$h-2$	$h-1$	h

문항 중요도 총 등급이 N, 사전 난이도 총 등급이 M일 때, k 번째 문항의 문항중요도는 $d_{i,k}$ 이고, 사전 난이도는 $d_{j,k}$ 이다. 이 때 k 번째 문항의 문항수준도 ($L_{i,j,k}$)는 식 1과 같다.

$$L_{i,j,k} = d_{i,k} + d_{j,k} - 1 \quad \text{--- 식 (1)}$$

이때 $L_{i,j,k}$ 는 $1 < L_{i,j,k} < h$ 의 범위를 갖는다.

사후 난이도(Post-difficulty Degree of Item)

정의 : 총 피험자 (M)중 답을 맞힌 피험자 (R)의 비율의 역수로 식 (2)와 같다[1].

$$D = \frac{R}{N}$$

$$D_{post} = |1 / D| \quad \text{--- 식 (2)}$$

피험자에게 문항을 적용한 결과 나타나는 가변적인 난이도로, 난이도 D 값의 범위 : $0 \leq D \leq 1$ 이다. 사후난이도는 사전난이도와 동일하게 하기 위해 난이도의 일반식에 역수를 취한 값을 사후난이도로 정의하였다. 또한 값의 범위를 $\max[C, D_{pre}]$ 으로 정규화한다. 정규화한 사후난이도 D_{post} 의 값의 범

위는 $[1, \max[C, D_{pre}]$ 가 된다. 사전/사후 난이도가 높을수록 어려운 정도가 높음을 의미한다.

문항 차분 난이도 (Item Difference Difficulty)

사후난이도는 양의 실수값을 갖게 된다. 이 사후 난이도값을 문항 수준도와 비교 평가하기 위해 문항 수준도 총 등급 (h)으로 나누어 등급을 산출한다.

문항 차분 난이도는 문항 수준도와 사후 난이도의 차이값으로 식 (3)과 같다.

$$D_d = D_{post} - L_{i,j,k} \text{ --- 식 (3)}$$

문항 차분 난이도 평가 기준

문항 차분 난이도 (D_d)가 음의 값을 가질 경우 피험자에 적용한 결과 난이도가 떨어짐을 나타낸다. 2등급 이상 차이를 나타내는 경우는 난이도의 급격한 하향으로 문항의 적합성 여부에 대해 검증하여야 한다.

문항변별도(Discrimination of Item)

문항이 능력에 따라 피험자를 변별하는 정도를 나타내는 지수[2]로, 능력이 높은 피험자가 문항의 답을 맞히고, 능력이 낮은 피험자가 문항의 답을 맞히지 못하면 이 문항은 변별력이 높은 문항이 된다. 문항 변별도는 다음 식(4)에 의해 산출한다.

$$r = \frac{N\sum XY - \sum X\sum Y}{\sqrt{N\sum X^2 - (\sum X)^2} \cdot \sqrt{N\sum Y^2 - (\sum Y)^2}} \text{ 식(4)}$$

N : 총 피험자 수
 X : 각 피험자의 문항 점수
 Y : 각 피험자의 총 점수

예) 5명의 피험자가 3문항으로 구성된 검사를 실시하여 얻은 응답 결과가 아래와 같을 때 각 문항의 변별도는 다음과 같다.

표 2. 문항변별도의 예

문항 피험자	문항 점수(X)			총점(Y)
	1	2	3	
A	1	1	1	5
B	1	1	0	4
C	0	0	1	3
D	0	0	0	2
E	1	0	1	1

문항변별도 평가기준

문항 변별도의 평가 기준은 표 3과 같다[3].

표 3. 문항변별도의 평가기준

문항변별도	평가	비고
0.40 이상	변별력이 높다	
0.30 - 0.39	변별력이 있다	
0.20 - 0.29	변별력이 낮다	
0.20 - 미만	변별력이 없다	

3. 검토

3.1 문항 수준도

문항의 수준을 나타내기 위해 문항의 중요도라는 개념을 도입하였다. 문항의 난이도만으로 문항의 수준을 나타낼 수 없기 때문이다. 문항이 어렵다고 문항의 수준이 높다고 판단할 수 없다. 그래서 문항의 수준을 나타내는 요소로 문항 중요도를 도입하여 난이도와 함께 수준을 설정하는 방안을 제시하였다. 문항의 중요도가 '상'이고, 문항의 난이도도 '상'인 경우 문항의 수준을 '최상'으로 설정할 수 있다. 반대로 문항의 중요도가 '하'이고, 난이도도 '하'이라면 이 문항의 수준은 '최하'로 설정한다. 이렇게 함으로써 문항의 수준을 중요도와 난이도를 고려하여 등급을 세분화할 수 있다.

3.2 사후난이도

몇 명의 피험자가 얼마나 정답을 하였는지를 나타내는 지수로 출제자의 사전 난이도와 비교 평가할 수 있다. 가상공간에서 출제자에 의해 정해진 난이도가 피험자들은 어렵게 느끼는지 쉽게 느끼는지 판단할 수 있게 된다.

$$r_1 = \frac{5(7) - (3)(8)}{\sqrt{5(3) - 3^2} \cdot \sqrt{5(18) - 8^2}} = 0.88$$

$$r_2 = \frac{5(5) - (2)(8)}{\sqrt{5(2) - 2^2} \cdot \sqrt{5(18) - 8^2}} = 0.72$$

$$r_3 = \frac{5(6) - (3)(8)}{\sqrt{5(3) - 3^2} \cdot \sqrt{5(18) - 8^2}} = 0.48$$

난이도 지정을 출제자의 판단만으로 정해지는 것이 아니고, 피험자가 어떻게 받아 들여지고 있는지를 난이도 지수에 반영함으로써 피험자의 상황을 반영할 수 있다. 또한 출제자 관점에서는 출제자의 난이도를 사후에 검증받을 수 있다.

3.3 문항차분난이도

문항차분난이도는 문항 수준도와 사후 난이도 차이로 가상공간에서 피험자의 평가를 거친 후 난이도 변화량을 측정하는 것이다. 사후난이도 반영은 피험자의 현 평가 상태를 적용하기 때문에 문항의 수준이 피험자의 동향에 의해 결정되어 진다. 출제자가 문항의 중요도와 사전난이도가 아무리 높게 설정한 문항이라고 피험자가 이 문항을 자주 접하게 되면 답할 확률이 높아지게 된다. 이렇게 되면 문항의 난이도는 낮아져야 되나, 출제자의 난이도는 쉽게 변경할 수 없는 문제이다. 그러나 현재의 피험자 상태는 반영되어야 한다. 이러한 상황을 고려할 수 있도록 문항차분난이도를 정의하였다.

문제은행은 많은 문항을 임의로 추출하여 피험자에게 출제하는 시스템이기 때문에 자주 출제된 문항은 피험자가 답할 확률이 높아지게 된다. 모든 피험자가 그 문항을 모두 잘 푼다면 그 문항은 당분간 출제되어서는 안된다. 이러한 상황을 문항차분난이도를 통해 해결할 수 있다. 즉, 문항수준도가 일정치 이하로 내려가면 문항 출제시 배제하도록 하면 된다.

3.4 문항변별도

문항 변별도 측정은 피험자의 변별력을 나타내야 하기 때문에 피험자의 고의적 실수를 최대한 제한할 수 있어야 한다. 문항 변별도는 피험자의 고의적 실수에 의해 치명적인 결과를 가져오기 때문에, 반드시 성적이 반영되는 시험 결과만 가지고 변별도를 측정하여야 한다.

피험자의 실수를 최소한 한 상태에서 측정된 변별력이 낮은 문항은 더 이상 문제은행에 존재할 필요가 없다. 변별력이 낮은 문항이 많은 문제은행은 그 만큼 피험자의 학업성취도 신뢰도가 떨어져 문제은행 시스템의 신뢰도도 같이 떨어지게 된다. 변별력이 낮은 문항은 출제자나 시스템 관리자에게 통보하여 문항의 지속 여부를 판단하도록 한다.

4. 결론

웹 상에서 문제은행을 구축하여 학습자의 학업성취도를 측정하는 사이트가 많이 등장하고 있다. 문제은행 시스템은 좋은 문항이 많을수록 훌륭한 시스템으로 인정받게 된다. 그러나, 가상공간에서 이루어지는 평가는 반복 횟수가 무제한적이고, 가변성이 많기 때문에 시스템 신뢰도 확보가 가장 시급한 과제이다.

문제은행 시스템 신뢰도를 확보하기 위해 문항을 자체적으로 평가하고, 적합성 여부를 검증하는 방안이 마련되어야 한다.

본 논문에서 제시한 문항차분난이도는 출제자의 난이도 기준에 피험자가 현재 어떤 난이도를 나타내는지를 반영한 것이다. 피험자들의 현 상태를 반영함으로써 피험자에 의한 문항 수준을 가변적으로 적용할 수 있게 하였다. 문항차분난이도를 주기적으로 검사하여 차이가 일정 수준 이상인 경우 일정한 기간 출제를 제한하는 기준으로 사용할 수 있다.

문항변별도는 피험자의 능력을 판단하는 주요한 지수로 반드시 문제은행 시스템 구축시 적용하여야 하는 지수이다. 문항 변별도는 문제은행 시스템 성능을 좌우하는 중요한 지수이다. 문항 변별도를 측정하는 방안은 여러 가지 제시되어 있으나 본 논문에서 인용한 방안이 시스템 구성이 용이하다고 판단하였다. 출제빈도가 높아질수록 변별도는 영향을 받게 되기 때문에 변별도의 급격한 변화가 일어나는 문항은 그 원인을 분석하여 출제자에게 문항의 적합성 여부를 판단하도록 정보를 제공한다.

가상공간에서 이루어지는 문항 구축은 사전 검토 없이 이루어지는 경우가 많기 때문에 문항의 적합성 검증 방안은 좀 더 연구되어야 하고, 문항 적합성 검증을 통해 문제은행 시스템의 신뢰도를 유지하는데 노력해야 할 것이다.

[참고문헌]

- [1] Carolina Paakkanen, M. Ed.(1996) *Hypermedia and the transfer of self-directed learning*. Proceedings of ED-MEDIA 96, p.551-556
- [2] 성태제(1996), "문항제작 및 분석의 이론과 실제", 학지사
- [3] Ebel. R. L, "Measuring Educational Achievement", Englewood Cliffs, NJ: Prentice-Hall