

복수의 답 선택을 허용하고 오답에 감점을 주는 객관식 시험

이창영*

동서대학교 정보시스템공학부, 부산광역시 617-716

Objective Test Allowing Multiple Choice for the Correct Answer with Penalty for Incorrect Choices

Chang-Young Lee*

Division of Information-System Engineering, Dongseo University, Pusan 617-716, Korea

요 약

현행 객관식 시험 방식에 내재되어 있는 결점을 보완하기 위하여, 수험자가 문항의 답을 둘 이상 복수로 선택할 수 있도록 하고, 틀린 답을 선택한 경우에는 감점을 주는 객관식 시험 방식인 MCP (Multiple-Choice with Penalty) 평가 방법을 고려했다. 제안된 방법의 효과를 검증하기 위하여, 두 수험자 A, B 그룹에 대해 각기 두 번씩의 시험을 실시하였다. 첫 번째는 객관식 시험으로, A 그룹은 현행의 오지선다 평가 방법에 의해, B 그룹은 본 논문에서 제안한 MCP 평가 방법에 의해 시험을 치렀다. 두 번째 시험은 첫 번째 시험에서의 답 선택이 문항에 대한 이해와 지식을 근거로 이루어졌는지를 파악하기 위함으로, 답을 선택한 근거를 상세히 기술토록 하였다. 현행 오지선다와 MCP 평가 방법의 호불호를 비교하기 위해 몇 가지 분석을 수행한 결과, MCP 평가 방법이 여러 가지 측면에서 오지선다 평가 방법에 비해 우수함이 확인되었다.

주제어 : MCP 평가, 객관식 시험, 오지선다, 감점, 성적의 뒤바뀜, 변별도

서 론

현재 치러지고 있는 많은 시험들은 사지선다 또는 오지선다 방식을 채택하고 있다. 수험자는 주어진 넷 또는 다섯의 답지들 중 하나를 고르는 것이다. 공정한 경쟁이 보장됨을 근거로 1969년 대입 예비고사에서 사지선다가 채택되었고, 행정고시 등 각종 공무원 시험에도 그 방식이 적용되었다(김태익, 2010). 이 흐름은 현재까지 이어져, 대학수학능력시험· 교원 임용고사· 토익과 같은 공인 시험들을 비롯하여 중등학교 학생들의 내신을 평가하는 많은 시험들이 선다형 방식에 의해 치러지고 있다. 많은 수험자들의 답안지를 신속하고 공정하게 처리하기 위해, 이러한 시험 방식은 유용한 평가 수단으로 자리잡아 왔고 사료된다.

선다형 시험의 명백한 결점은 운이 개입된다는 것이다. 즉, 문항에 대한 지식이 전혀 없는 수험자라도 정답을 맞힐 수 있는 것이다. 사지선다의 경우, 임의로 한 답지를 선택했을 때 그 결과가 정답일 확률은 25%로 높다. 운에 의한 맞힘을 줄이기 위해, 제 5의 답지로서 '정답 없음(None of the above)' 또는 '모두 정답(All of the above)'이 추가된 방식이 시도되기도 하는데, 이 처방의 효과 및 교육적 영향에 대해서는 논의의 여지가 있다(성태제, 2009). 또한, 답지 수를 다섯으로 늘려도 운으로 문제를 맞힐 확률은 20%로서 여전히 작다고 할 수 없다.

한 문제의 맞고 틀림에 따라 중대한 갈림이 생겨나는 사례들을 굳이 거론하지 않더라도, 시험에서 운이 개입되지 말아야 할 이유는 아무리 강조해도 지나치지 않다. 운으로 맞힐 기회가 수험자들에게 균등하게 주어진다고 해서, 그리고 이러한 시험의 결과에 대해 수험자들이 이의를 제기하지 않는다고 해서 그 시험이 공정하다는 것을 의미하지는 않는다. 다만 '불공정하

* 교신저자 이메일 주소 : seewhy@dongseo.ac.kr

• 2011년 1월 4일 접수, 2011년 1월 28일 수정, 2011년 1월 31일 통과

지는 않다'는 소극적 공정의 견지에서 통용되고 있을 뿐이다. 운에 의한 점수 등락을 둔화시키는 수단으로서 스태나인(Stanine) 9등급제(권대훈, 2008)가 도입되는 사례도 있으나, 이는 운의 개입을 해결하는 처방과는 거리가 있을 뿐더러 수험자들에 한테도 환영을 받지 못하고 있다. 대학수학능력시험에 9등급제가 적용된 학생들로부터는 '저수받은'이라는 용어가 등장하기도 하였다(진명선, 2007).

2010년 9월, 미국 교육부는 3900억 원을 들여 SAT를 비롯한 선다형 시험을 폐기하는 학력평가 개혁에 착수했다. 미국이 선다형 시험을 없애면, 경제협력개발기구(OECD) 국가 중 선다형으로 대학입시를 치르는 나라는 한국 일본만 남는다(송홍근, 2010). 하지만 미국의 야심찬 '학력평가 2.0' 프로젝트의 성공 여부는 아직 미지수이다.

피평가자들의 상대적인 서열을 매기는 규준지향평가(Norm-Referenced Evaluation)에서는 다양한 평가 방법들이 고려되고 비교 분석되어야 한다. 예를 들어, 현행 선거에서는 투표자가 한 사람의 후보자에게 표를 주고 최다 득표 후보자가 당선이 되는 단순다수제를 채택하고 있는데, 이 방식에는 후보자에 대한 '거부' 의견은 (그 비율이 아무리 높을지라도) 반영될 수 없는 문제가 있다(강원택, 2005). 운동 경기에서의 '승리/무승부/패배'를 통한 순위매김에도 다양한 방법이 가능하며(이준목, 2009), 세심한 분석을 통하여 점차 합리적인 방법이 도입되고 채택되는 추세이다. 본 논문에서는 선다형 시험 방식에 내재되어 있는 결점을 보완·개선하기 위하여 새로 제안되는 MCP 평가 방법에 대해 고찰해 보고자 한다.

이론적 배경

운으로 답을 맞히는 것을 막으려면 부득이 오답에 감점을 주어야 한다. 임의로 한 답을 선택했을 때의 기댓값이 0이 되도록 감점을 주면 되므로, 현행 오지선다의 경우 정답을 맞혔을 때의 점수를 1이라 하면, 오답으로 틀렸을 때의 점수는 -0.25로 하던 된다. 감점이 있으면 수험자가 답을 선택하지 않는 경우가 생긴다. 즉, 감점을 도입함으로써, 현재의 <1 / 0 (맞음 / 틀림)>의 두 가지 점수는 <1 / 0 / -0.25 (맞음 / 모름 / 틀림)>의 세 가지 점수로 바뀌게 된다. 이제 문항에 대한 지식이 없는 수험자가 요행을 바라고 임의로 어느 한 답을 선택하는 일은 줄어들 것이다. 임의의 답지 선택에 따른 점수의 기댓값이 0이므로 수험자가 요행을 통해 얻을 이익은 없으며, 따라서 굳이 어느

한 개의 답을 임의로 선택하지는 않을 것이기 때문이다.

그런데, 이처럼 단순한 감점 도입 방식에는 불합리함이 내재되어 있다. 학생 A와 B가 한 문제를 푸는 가상적인 상황을 생각한다. 빈번하게 벌어지는 일로서, 학생 A는 "1번과 2번 중 하나가 답"이라는 결론에 도달하였다. 다섯 답지 중 둘로 정답의 범위를 좁혔으므로, 답을 선택하지 않고 0점을 받는 것보다는 그가 좁힌 1번과 2번의 둘 중 하나를 선택하는 것이 유리하다. 그의 생각대로 이 둘 중 하나가 정답이면 그의 점수 기댓값은 0보다 크기 때문이다. 갈등 끝에 그는 1번을 답으로 적었다. 그런데 정답은 2번이었고, 그 결과 감점을 당하여 -0.25점을 받았다. 한편, 학생 B는 이 문제에 대한 지식을 전혀 갖고 있지 못했다. 따라서 그는 답란을 빈칸으로 두어 0점을 받았다. 이제 두 학생의 점수를 비교해 보자. 문제를 상당히 많이 파악한 A가 아무런 지식을 갖고 있지 않던 B보다 낮은 점수를 받는 결과가 나타나고 말았다.

이러한 불합리함을 어떻게 해결해야 할까? 아마 학생 A는 "1번과 2번 둘 중 하나가 답입니다."라고 말하고 싶은 것이다. 하지만 현행 시험에서는 이러한 자유가 허용되지 않는다. 오직 하나만 정답으로 선택해야 하는 것이다.

평가 수단은 아니지만, 답변자를 곤혹스럽게 만드는 경우가 있다. 질문자가 답변자에게 "(예 또는 아니오)로만 대답하십시오!"라고 강요하는 경우가 그것이다. 답변자가 스스로에게 유리한 진술 또는 설명을 하지 못하도록, 질문자가 답변자를 구속한다. 답변자는 (예 또는 아니오)의 대답을 강요당하는 상황에서 부득이 하고 싶은 말을 삼켜야만 한다.

수험자가 얼마나 알고 있는가를 평가하기 위한 시험에서는 수험자를 구속할 이유가 없고, 따라서 바람직하지 않다. 도리어, 가능하다면, "이 문제에 대한 당신의 생각을 자세히 말해 보시오."라고 함으로써, 그의 생각을 충분히 표현토록 하는 것이 이상적이다. 다만 현실적으로 실현할 수 없기에 못할 뿐이다. 위의 학생 A에게 "다섯 중 오직 하나만 고르시오!"라는 구속은 바람직한 평가에 역행하는 제약인 것이다.

위에서 예로 든 두 학생의 시험 결과에서 나타난 불합리함은, 한 문항에 대해 둘 이상의 답을 복수로 선택할 수 있도록 함으로써 제거된다. 즉, 학생 A로 하여금 "1번과 2번 둘 중 하나가 답입니다."라고 표현할 수 있는 자유를 허용하는 것이다. 이는 '선다(選多)'의 영어 표현인 'Multiple-Choice'의 답백한 이행이다.

위 두 가지 사항을 정리하면, 본 논문에서 제안하는 <복수의

답을 선택할 수 있게 하고, 틀린 선택에 대해서는 감점을 주는 객관식 시험) 즉 'MCP (Multiple-Choice with Penalty)' 평가 방법은 다음과 같다.

- 수험자가 둘 이상의 답을 선택할 수 있도록 한다.
- 오지선다의 경우, 옳은 선택에는 1점을, 틀린 선택에는 -0.25점을 부여한다.

이 방식으로 시험을 치르게 되면, 학생 A는 1번과 2번 두 개를 답으로 선택한다. 정답은 2번이므로 그 선택에 대해서는 +1점을, 틀린 선택 1번에 대해서는 감점으로 -0.25점을 받아, 그의 점수는 0.75점이 된다. 그는 이 문제를 75%만큼 파악했던 것이다. 정답 2번만을 선택한 학생의 +1점에는 못 미치지만, 문항에 대한 지식이 전혀 없던 학생 B의 0점보다는 높은 점수를 받는 합리적인 결과가 나타남에 주목할 필요가 있다. 만약 다른 학생 C가 '답지 4와 5는 정답이 아니다.'라는 추론을 했다면, 그는 1과 2와 3의 셋을 답으로 선택할 것이다. 그러면, 2번이 정답이고 1번과 3번은 오답이므로, 그의 점수는 $+1-2 \times 0.25 = 0.5$ 점이 된다. 학생 C는 이 문제를 50%만큼 파악한 것이다. 현행 시험에서는 학생 C가 운으로 맞힐 확률이, 답지 4번과 5번을 정답에서 배제시킴에 의해, 1/5에서 1/3로 높아지지만, 그 결과는 0점 또는 1점으로 나타난다. 하지만, MCP 방식에서는 4번과 5번을 배제시킨 그에게 0.5의 구체적인 점수가 주어지는 것이다.

답지가 다섯인 MCP 방식의 시험에서 최악의 점수는 정답만 빼고 나머지 네 개의 오답만 선택한 경우로, 그 점수는 $-0.25 \times 4 = -1$ 점이다. 최선의 점수는 정답 하나만을 선택한 경우의 +1점이다. 한 문항의 점수는 -1점에서 +1점까지 0.25점 단계로 총 이홉 단계의 점수 구분이 이루어진다. 이로써, 현행의 0점과 1점 두 가지 나뉘에 비해 점수가 세분화된다. 운이 개입될 여지가 없으며, 수험자가 문항을 파악한 정도만큼 점수가 주어진다.

수험자가 여러 답을 선택한다는 점에서 MCP 방법은 다답형(권대훈, 2008) 문항과 비슷해 보이지만, MCP에서는 수험자가 여러 답을 고르기는 하나 정답은 다섯 답지 중 하나만 있다는 차이가 있다. 다답형에서는 수험자의 문항에 대한 부분적 이해가 반영될 기회가 없고, 맞음 또는 틀림의 두 가지만 판정하므로 MCP와는 성격이 다르다.

현행 많은 시험은 부득이하게 경쟁시험 즉 기준지향평가를 수행하고 있다. 이 평가 방법이 내포하고 있는 단점들은 학생 개개인에게는 물론 사회적으로도 좋지 않은 영향을 미친다. 반

면 기준참조평가(Criterion-Referenced Evaluation)에서는 학습자가 무엇을 얼마만큼 알고 있는가에 관심을 둔다(Glaser, 1963). 미국에서도 1930년 이후부터 평가방법을 주관식에서 객관식으로 전환하여 기준참조평가를 사용하게 되었다(Mislevy, 1997). MCP 방법에 의하면, 순위매김을 목적으로 하는 기준지향평가와 학습자의 알고 있는 정도를 평가하는 준거참조평가의 기능이 동시에 수행될 수 있을 것으로 사료된다.

연구 방법

본 논문에서 제안한 시험 방식의 유효함을 검증하기 위해서는, 하나의 수험자 그룹을 대상으로 동일한 내용의 시험을 다음과 같이 세 차례 반복하는 것이 이상적이다.

- 현행 오지선다 방식에 의한 시험
- 본 논문에서 제안된 MCP 방식에 의한 시험
- 수험자가 지식과 이해를 바탕으로 문제를 맞혔는가를 파악하기 위해 수험자에게 상세한 풀이를 요구하는 시험

하지만 동일한 내용의 시험을 세 번 치르게 하는 것은 수험자들에게 불쾌감을 줄 것이고, 시간이 갈수록 누적되는 피로감은 수험자가 답안 작성에 최선을 다하지 않게 만듦으로써, 본 연구의 방향을 왜곡시킬 가능성이 있다. 그렇다고 해서 위 세 시험을 다른 세 그룹이 각각 치르게 하면 그 결과는 '참된 비교'와는 거리가 있을 것이다. 이 점들을 감안하여, 한 그룹과 세 그룹의 절충으로서, 본 연구에서는 표 1과 같이 수험자들을 두 개의 그룹으로 나누어 각 그룹마다 두 번씩 시험을 치르게 하였다. A1은 현행 오지선다 방식의 시험이고, B1은 본 논문에서 제안한 MCP 방식의 시험이다. 시험 A2[B2]는 A1[B1]의 시험을 치를 때 문항에 대한 이해와 지식을 근거로 답을 선택했는지, 아니면 임의로 선택했는지를 판단하기 위한 것이다.

문항은 정답형 · 최선답형 · 합답형 및 불완전 문장형을

표 1. 두 수험자 그룹에 대해 치러진 네 가지 시험.

수험자 그룹	시험 구분	시험 방법
A	A1	현행 오지선다
	A2	시험 A1과 같은 내용과 방법으로 치러진 답을 얻은 과정에 대한 상세한 풀이를 요구함
B	B1	본 논문에서 제안한 MCP 방법
	B2	시험 B1과 같은 내용과 방법으로 치러진 답을 얻은 과정에 대한 상세한 풀이를 요구함

표 2. 현행 오지선다 시험 A1과 MCP 시험 B1에서의 문항곤란도.

시험	현행 (A1)	MCP (B1)
문항곤란도(평균정답률)	0.53	0.38

고루 나누어, 공과대학 1학년 학생들을 대상으로 수학과 과학 지식을 묻는 내용으로 작성하였다. 시간이 모자라 풀지 못하는 일은 없도록 함으로써, 속도검사(Speed Test)가 아닌 역량검사(Power Test)가 되도록 하였다. 이것이 본 논문의 탐구 목적, 즉 <알고 있는 정도와 실제 획득하는 점수의 상관관계>를 파악하는 데 적절하기 때문이다. 다섯 개의 정답 항 개수도 대략 비슷해지도록 배치하였으며, 문항의 난이도는 평균 60점 정도가 얻어지도록 하였다. 총 문항에 대한 수험자들의 문항곤란도(박도순과 홍후조, 2006) 즉 평균 정답률은 표 2와 같이 나타났다. MCP 방식에서는 운에 의해 맞는 일이 배제되므로, 현행 방식에 비해 정답률이 낮게 나타남은 당연한 결과이다.

A2와 B2의 시험에서는 답을 선택하게 된 근거를 상세히 기술하도록 하였다. 답안지로부터 수험자가 문항을 이해하였는지 못했는지 판단하기 어려운 경우에는 일일이 면담과 구술을 통한 확인과정을 거쳤다. 주어진 답지 중 정답 또는 최선답을 고른다는 점에서 A2와 B2는 객관식 시험이지만, 채점 과정에서 채점자의 주관성이 개입될 수 있다는 점에서는 주관식 시험이라 할 것이다(박도순과 홍후조, 2006). 또한 수험자들이 답을 선택한 근거를 서술해야 하므로 선다형이 아닌 서답형 시험이다. 따라서, 문항 제작에 있어서 가능한 한 채점자의 주관성이 개입되지 않고 명료한 채점 기준이 마련될 수 있도록 문항 제작에 세심한 주의를 기울였다.

결과 및 논의

1. 성적 분포

그림 1은 표준점수(T점수)에 따른 수험자들의 빈도를 나타낸다. 현행 방식에 비해 MCP 방식에 의한 시험에서 수험자들의 점수는 보다 세분화되어 분포함을 볼 수 있다. 이는 MCP 방식에서 한 문항의 점수가 아홉 단계로 나뉘기 때문이다.

2. 동점자 수

수험자들의 상대적인 서열을 매기는 규준지향평가에서는 가

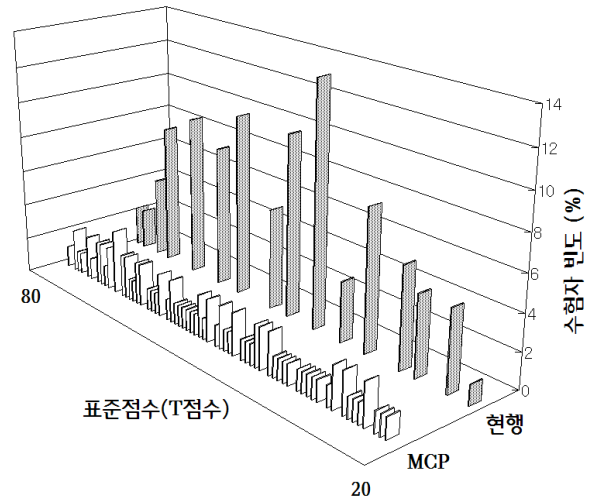


그림 1. 수험자들의 표준점수 분포.

급적 동점자 수가 적을수록 좋다. 하지만, 점수의 가짓수는 문항 수에 따라 한정적일 수밖에 없으며, 수험자 수가 점수의 가짓수보다 많으면 동점자는 불가피하게 나타나기 마련이다. 본 논문에서 제한한 방법의 효과들 중 하나는 동점자 수가 줄어드는 것이며, 이는 그림 1에서 이미 확인할 수 있다.

우선, 시험에서 동점자가 얼마나 많은가를 정량화할 필요가 있다. 점수에 따른 수험자들의 빈도 f 를 합산하여 동점자 정도를 수치화하는 것은 좋지 않다. 일례로, 어느 한 점수에 네 명이 동점을 받은 경우와, 다른 두 점수 각각에 두 명씩 동점을 받은 경우를 비교해 보자. 빈도의 합은 두 경우 공히 4이지만, 전자의 경우가 동점자 정도는 더 크다.

동점자 정도의 합리적인 정량화를 위해, 우선 다중도(Multiplicity) M 을 다음과 같이 정의한다.

$$M = \begin{cases} f, & \text{for } f \geq 2 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

이제 동점지수 Q 를 다음과 같이 정의한다.

$$Q = \frac{1}{E^2} \sum_s M^2(S) \quad (2)$$

식 (2)에서 S 와 E 는 각각 점수(Score)와 수험자 수(Examinee)를 나타낸다. $S=[0, 1, 2]$ 의 세 점수로 나뉘는 시험을 여섯 명($E=6$)이 치렀을 때, 세 경우의 동점지수 계산 예를 표 3에 나타내었다.

표 3. 동점지수의 계산 예.

점수(S)	경우 1		경우 2		경우 3	
	f	M ²	f	M ²	f	M ²
0	0	0	1	0	2	4
1	6	36	4	16	2	4
2	0	0	1	0	2	4
동점지수 Q	36/36 = 1		16/36 = 0.44		12/36 = 0.33	

일반적으로 모든 점수에 수험자들이 가장 고르게 분포할 때 동점지수는 가장 작으며, 따라서 규준지향평가에서는 이러한 경우가 가장 바람직하다. 문항 수를 P라 하면, 현행 시험에서 나타나는 점수의 가짓수 또한 P이므로, 가장 고르게 분포하는 경우의 다중도는 M* = (E/P)로 주어진다. 따라서 동점지수의 최소값은 다음과 같이 계산된다. 표 3의 '경우 3'이 이에 해당한다.

$$Q_{\min}|_{\text{현행}} = \frac{1}{E^2} \sum_{S=1}^P M^2(S) \Big|_{M=M^*} = \frac{1}{P} \quad (3)$$

한편, 본 논문에서 제안한 MCP 방법에서는 한 문항의 점수가 아홉 가지로 나뉘므로, 동점지수의 최소값은 다음과 같이 계산된다.

$$Q_{\min}|_{\text{MCP}} = \frac{1}{E^2} \sum_{S=1}^{9P} (E/9P)^2 = \frac{1}{9P} \quad (4)$$

즉, 동점지수가 현행 시험 방식에 비해 1/9로 줄어들 것이 기대되는 것이다.

본 연구에서 현행 방식과 MCP 방식 시험의 수험자 수는 각각 89명과 85명이었으며, 이에 대한 동점지수의 이론적 최소값은 각각 1.12%와 0.13%이다. 실제 시험에서는 이 이상적인 최소값들에 비해 높은 동점지수가 나타날 수밖에 없다. 표 4는 현행 방식과 MCP 방식에 의한 동점지수의 결과이다. MCP 방식에 의한 동점지수가 현행 방식에 비해 1/6 정도로 낮아졌음을 알 수 있다.

표 4. 현행 방식 시험과 MCP 방식 시험에서의 동점지수 결과.

	이론적 최저 동점지수	실제 동점지수
현행	1.12 %	7.98 %
MCP	0.13 %	1.29 %

문항 수를 늘림으로써 동점지수를 줄이려는 시도는 평가를 역량검사에서 속도검사로 만드는 결과를 초래할 수 있다. MCP 방법에 의하면 작은 문항 수로도 동점자 수를 크게 줄일 수 있는 효과가 있다.

3. 석차의 뒤바뀜

본 논문에서는 '수험자가 갖고 있는 지식의 정도'와 '시험을 치른 결과로 얻는 성적' 사이의 상관관계를 파악하는 데 가장 큰 주안점을 두고 있다. 이 조사를 수행하기 위해 동일한 수험자 그룹으로 하여금 두 번의 시험을 치르게 한 것이다. 우선 '석차 이득'을 다음과 같이 정의한다.

$$\text{석차 이득} \equiv (A2|B2\text{에서 얻은 석차}) - (A1|B1\text{에서 얻은 석차})$$

표 1의 수험자 그룹 A가 치른 시험을 예로 들어 설명하면, 시험 A1에서 수험자가 얻은 석차는 현행 오지선다 시험의 결과이다. 시험 A2에서 얻은 석차는, 답의 선택에 대한 상세한 설명이 요구되므로, 문항에 대한 이해와 지식을 평가하는 시험의 결과이다. A1과 A2의 시험은 각각 '선택 기반 시험'과 '지식 기반 시험'으로 표현될 수 있다. 만약 A1에서 얻은 석차는 5인데 A2에서 얻은 석차가 9라면, 그 수험자는 운에 의한 맞힘을 통해 +4의 석차 이익을 얻은 것이다. 그림 2의 가로축은 석차 이득을, 세로축은 석차 이득에 따른 수험자 빈도를 나타낸다.

A1과 A2를 비교한 결과, 현행 오지선다 방식에 의한 시험에

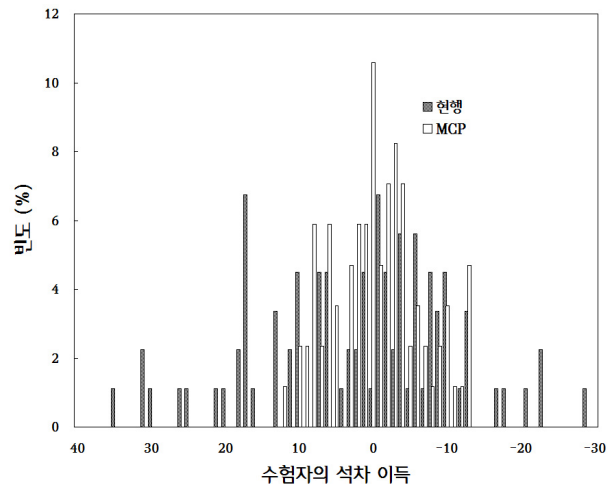


그림 2. 현행 오지선다 방식 시험과 MCP 방식 시험에서 석차 이득의 빈도.

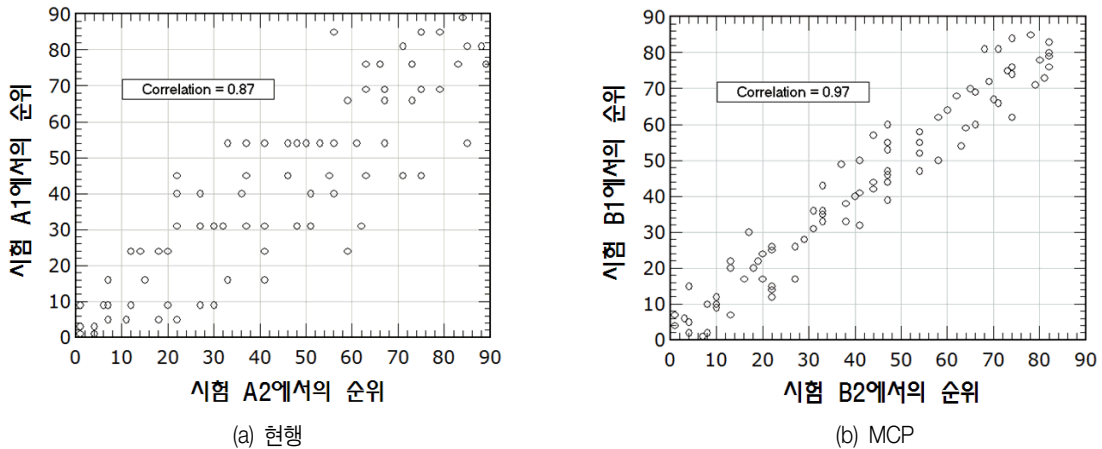


그림 3. 선택기반 시험과 지식기반 시험의 상관관계.

서 가장 큰 이득은 +35, 가장 낮은 이득은 -29로 나타났다. 89명의 수험자 수를 고려할 때 매우 큰 뒤바뀜이다. 반면, B1과 B2를 비교한 결과, MCP에 의한 시험에서는 석차의 이득이 ±15의 범위 안에 머물렀다. 그림 2에서 현행 방식에 비해 MCP에 의한 방식에서 수험자들의 석차 이득은 0 근처에 상대적으로 더 많이 분포되어 있다.

이론적으로는 MCP 방식의 시험에서 석차 이득은 모두 0일 것이 기대된다. 요행을 바라고 답을 선택하지 않고, 알고 있는 만큼 점수를 얻을 것이 예상되기 때문이다. 하지만 실제 실험 결과에서는 이 시험 방식에서도 다소의 석차 이득 및 손실이 나타났다. 그 원인은, (1) B1의 시험에서도 (기댓값이 0임에도 불구하고) 수험자가 요행을 바라고 답을 선택함, (2) B2의 시험 즉 ‘지식 기반 시험’에서 최선을 다하지 않거나 답 선택 근거에 대한 표현력 부족, (3) 시험 B2의 채점 과정에서 채점자의 주관적 개입 동일 것으로 추정된다.

4. 상관계수

다른 비교 방법으로서, ‘수험자가 문항을 이해하고 있는 지식의 정도’와 ‘시험에서 얻는 성적’ 사이의 상관관계를 살펴보기로 한다. 그림 3 (a)의 가로축은 문항에 대한 이해와 지식을 평가하는 시험 A2에서 수험자들이 얻은 순위를, 세로축은 현행 오지선다 방식의 시험 A1에서 얻은 순위를 나타낸다. 동 그래미로 나타내어진 데이터 수는 그룹 A의 수험자 수인 89이다. (b)의 가로축과 세로축은 각각 시험 B2와 B1에서 얻어진 순위이다. 가로축과 세로축의 두 순위가 정비례한다는 것은 수험자들이 실제 객관식 시험에서 얻은 순위가 문항에 대한 이해

와 지식에 따라 매겨졌음을 의미하고, 따라서 그 평가가 우수했음을 의미한다.

보다 먼밀한 정량적 비교를 위해 두 순위 사이의 적률상관계수(Produce-Moment Correlation)를 분석하였다(김대현과김석우, 2005). 그 결과는 표 5와 같다. MCP 방식 시험에서의 적률상관계수가 약 10% 더 크게 나타났는데, 이는 그림 3으로부터 짐작할 수 있는 바이다.

표 5. 시험 A1[B1]과 A2[B2] 사이의 적률상관계수.

시험	현행 (A1/A2)	MCP (B1/B2)
적률상관계수	0.87	0.97

5. 변별도

변별도는 문항의 질에 관련된 개념이므로 평가 방법에 관해 고찰하는 본 논문과는 관련이 없어 보인다. 하지만 평가방법이 달라지면 수험자들의 반응이 달라지므로 변별도도 달라지게 마련이다. 현행 시험에서는 문항에 대한 내용을 전혀 파악하지 못한 학생도 운에 의해 정답을 맞힌 학생에 포함될 가능성이 있지만, MCP 방식의 시험에서는 요행에 의한 맞힘이 상대적으로 적으므로 변별도 또한 높아지리라 기대된다. 즉, 동일한 문항으로도 평가 방법에 의해 변별도는 달라질 수 있는 것이다.

우수 학생과 비우수 학생을 문항이 얼마나 잘 구분했는가를 나타내는 척도로서

$$\text{변별도} = \frac{(\text{상위 절반의 정답자 수}) - (\text{하위 절반의 오답자 수})}{(\text{학생 절반 수})} \quad (5)$$

의 식이 사용된다. 상위 절반만 모두 맞히고 하위 절반은 모두 틀렸다면 변별도는 +1이 되어 가장 바람직한 문제로 여겨지는 것이다. MCP 방식의 시험에 식 (5)를 그대로 적용할 수는 없다. 한 문항의 결과가 맞거나 틀렸나로 양극화되지 않고 이혼 단계로 나뉘기 때문이다.

식 (5)는 다음과 같이 바뀌 쓸 수 있다.

$$\text{변별도} = \frac{(\text{상위 절반의 점수 합}) - (\text{하위 절반의 점수 합})}{(\text{문항당 점수}) \times (\text{학생 절반 수})} \quad (6)$$

이렇게 일반화된 식 (6)은 이제 MCP에도 적용될 수 있다. 표 6은 현행 방식 시험 A1과 MCP 방식 시험 B1에 이 식을 적용하여 계산한 변별도를 보여준다.

표 6. 현행 시험 방식 A1과 MCP 시험 방식 B1에서의 변별도.

시험	현행(A1)	MCP(B1)
변별도	0.31	0.43

MCP 방식에 의해 변별도가 크게 향상되었음을 알 수 있다. 그 이유를 파악하는 것은 어렵지 않다. MCP 방식에서는 문항을 파악하지 못한 수험자가 정답을 맞힐 가능성이 현행 시험에 비해 작다. 이론적으로는 0이다. 따라서 하위 절반 그룹의 점수 합은 현행 시험에 비해 상대적으로 낮을 수밖에 없으므로 변별도는 커지게 되는 것이다.

결론 및 제언

현재 보편적으로 시행되고 있는 선다형 시험의 결점을 보완하기 위하여, 본 논문에서는 MCP 평가 방식의 객관식 시험을 제안하였다. 그 내용은 다음 두 가지로 기술된다.

- 문항에 주어지는 답지들 중에서 수험자가 자유롭게 선택할 수 있게 한다. 보다 일반적으로는 각 답지가 정답일 가능성을 [0, 1] 범위의 숫자로 나타내도록 할 수도 있겠으나, 이는 현실적으로 실현하기 쉽지 않을 것이다.
- 답지의 개수가 N 인 경우, 정답 선택의 점수를 1이라 할 때 오답 선택에 대해서는 $-1/(N-1)$ 의 점수를 부여한다.

제안된 방법의 기대되는 효과는 두 가지이다. 하나는 운에 의해 정답을 맞힐 가능성이 배제된다는 것이다. 임의로 하나의 답을 선택했을 때 수험자가 얻는 점수의 기대값은 0이기 때문이다. 다른 하나는 점수의 세분화이다. 현행 오지선다에서는

한 문항의 점수가 둘로 나뉘는 반면, 본 논문에서 제안한 방법에서는 이혼 가지의 점수로 나뉜다. 문항에 대한 수험자의 부분적인 이해 정도를 파악하는 것이 가능해지며, 한 문항의 점수는 곧 수험자가 그 문항을 이해하는 정도와 일치한다.

제안된 방법의 효과를 검증하기 위하여 두 수험자 그룹으로 하여금 각각 두 번의 시험을 치르게 하였다. 한 수험자 그룹은 먼저 현행의 오지선다 방식 시험을 치르게 한 후, 두 번째 시험에서 문항에 대한 이해와 지식을 상세히 기술토록 하였다. '선택 기반 성적'과 '지식 기반 성적'의 상관관계를 비교하기 위함이다. 다른 수험자 그룹은 우선 본 논문에서 제안한 방식의 객관식 시험을 치르게 한 후, 두 번째 시험에서 문항에 대한 이해와 지식을 상세히 기술토록 하였다. 역시 '선택 기반 성적'과 '지식 기반 성적'의 상관관계를 비교하기 위함이다.

시험 결과는 다각도로 분석되었다.

첫째, 성적 분포를 조사한 결과 현행의 오지선다 방식에 비해 본 논문에서 제안한 방식에 의한 시험에서는 보다 많은 가짓수로 수험자들의 점수가 분포하고, 이로 인해 동점자 수가 확연히 줄어들 수 있었다.

둘째, 본 논문의 가장 중요한 사항으로서 '선택 기반 성적'과 '지식 기반 성적'의 상관관계를 조사하였다. MCP 평가 방식에 의하면 그 두 가지 성적이 뒤바뀌는 사례가 현행 오지선다 방식에 비해 현저히 줄어들어 확인되었으며, 두 성적의 정(正)의 상관관계가 현행 오지선다에 비해 높게 나타났다.

셋째, 문항 변별도가 향상됨을 확인할 수 있었다. 하위 그룹에 속하는 수험자가 운에 의해 문항을 맞히는 것이 배제되므로 문항 변별도는 커지는 것이다. 이처럼, 본 논문에서 분석된 몇 가지 결과는 제안된 방식의 시험이 현행 오지선다 시험 방식에 비해 우수함을 실험적으로 확인시켜 주었다.

본 논문에서 제안한 방법을 구현하는 데에는 큰 변화가 요구되지 않는다. 우선 문항 제작 단계에서는 아무런 변화가 필요치 않다. 수험자에게는 작은 변화가 주어진다. 주어진 문항에 대해 단 하나의 답을 선택해야 하는 구속에서 벗어나, 자유롭게 답을 선택할 수 있다. 가장 큰 변화는 채점 단계에 있다. 답지가 다섯인 경우, 문항 수가 N 이면 수험자는 최대 $5N$ 개의 답을 선택할 수 있다. 점수는 다음과 같이 계산된다.

$$\text{점수} = (\text{정답 선택의 개수}) - 0.25 \times (\text{오답 선택의 개수})$$

채점 과정이 다소 복잡하지만, 기계를 통한 자동 채점에서는

어렵지 않은 조작에 의해 본 논문이 제안하는 방식의 채점을 구현할 수 있을 것이다.

평가는 공정해야 한다. 현행 시험에서의 공정은 '기회의 공정'을 의미한다. 운에 의한 맞힘의 기회가 누구에게나 열려 있다는 의미이다. 하지만, 수험자가 아는 만큼 점수를 얻게 하는 '적극적 의미의 공정'이 바람직하다. 표준지향평가에서는 순위매김이 주된 목적이며, 그러한 평가에서 문항에 대한 이해와 지식의 실제 시험에서 얻는 순위와 높은 정(正)의 상관 관계를 가져야 할 이유는 매우 크다. 제안된 MCP 평가 방식은 또한 수험자가 무엇을 얼마만큼 알고 있는가를 평가하는 표준 참조평가의 역할도 수행할 것으로 사료된다.

Abstract

In an effort to remedy the problems inherent in the current objective test, we propose a new type of test that allows multiple choice for the correct answer and imposes appropriate penalty for the incorrect choices. To investigate the effect of the proposed method, two examinee groups were employed to take two kinds of examinations each. The one is to evaluate the examinee according to the choice of the answer. The other is to get the examinee's degree of knowledge and understanding for the given test by requiring them to expound the reason for the choice of the answer. By several estimations for comparison of the current and the proposed methods, it has been found that the MCP method proposed in this paper is superior to the current one in every respect examined.

Keywords: MCP Evaluation, Objective Test, Multiple Choice, Penalty, Rank Reversal, Degree of Discrimination

참고문헌

강원택 (2005) 한국의 정치개혁과 민주주의. 인간사랑.
 권대훈 (2008) 교육평가. 학지사.
 김대현, 김석우 (2005) 교육과정 및 교육평가. 학지사.
 김태익 (2010) 사지선다형 시험 퇴출. 조선일보 2010/9/5.
 박도순, 홍후조 (2006) 교육과정과 교육평가. 문음사.
 성태제 (2009) 교육평가의 기초. 학지사.

송홍근 (2010) 인터뷰: 교육공무원으로 변신한 대치동 학원가 전설 이범. 신동아.

이준목 (2009) 이준목 칼럼. 야구타임스 2009/11/19.

진명선 (2007) 내신의 배신... 끝모를 '08 번뇌'. 한겨레신문 2007/7/8.

Glaser R (1963) Instructional Technology and Measurement of Learning Outcome: Some Questions. American Psychologist 18: 519-621.

Mislevy RJ (1997) Assessing Student Learning, in Walberg HJ and Haertel GD (Eds.), Psychology and Educational Practice, MrCutrhan Publishing Corporation, pp 176-195.

부록: 본 논문의 연구를 위해 치러진 시험

1. $\sqrt{10}$ 에 가장 가까운 정수는?
 ① 2 ② 3 ③ 4 ④ 5 ⑤ 6

2. 다음 <보기>에서 옳은 것만을 모두 고른 것은?

<보 기> ㄱ. 0은 자연수이다. ㄴ. 6은 완전수이다. ㄷ. 17은 소수(素數, Prime Number)이다.

- ① ㄱ ② ㄴ ③ ㄱ, ㄴ ④ ㄱ, ㄷ ⑤ ㄴ, ㄷ

3. 윗놀이에서 도가 나올 확률은?

- ① $\frac{5}{16}$ ② $\frac{1}{4}$ ③ $\frac{3}{16}$ ④ $\frac{1}{8}$ ⑤ $\frac{1}{16}$

4. Korea는 IT 강국이다. 이 표현의 'IT'에서 'I'가 뜻하는 단어는?

- ① Interaction ② Information ③ Intelligence
 ④ Instant ⑤ International

5. 다음 보기 중에서 옳은 것만을 모두 고른 것은?

<보 기> ㄱ. $\sin 30^\circ = 0.5$ ㄴ. $\cos(-60^\circ) = -0.5$ ㄷ. $\tan 45^\circ = \frac{1}{\sqrt{2}}$

- ① ㄱ ② ㄴ ③ ㄴ, ㄷ ④ ㄱ, ㄷ ⑤ ㄱ, ㄴ, ㄷ

6. 다음 보기 중에서 옳은 것만을 모두 고른 것은?

〈보기〉

ㄱ. $\log(1) = 1$ ㄴ. $\log\left(\frac{1}{32}\right) = -5\log(2)$ ㄷ. $\log(2) + \log(3) = \log(6)$
--

- ① ㄴ ② ㄷ ③ ㄱ, ㄴ ④ ㄴ, ㄷ ⑤ ㄱ, ㄴ, ㄷ

7. 함수 $y = \sqrt{x}$ 에 대해 $\frac{dy}{dx} = 1$ 이 되는 x 의 값은 얼마인가?

- ① $\frac{1}{4}$ ② $\frac{1}{2}$ ③ 1 ④ 2 ⑤ 4

8. $\int_0^1 (2x-1)dx$ 의 값은 얼마인가?

- ① -2 ② -1 ③ 0 ④ 1 ⑤ 2

9. 방정식 $x^2 - 4x + 1 = 0$ 의 두 근을 α 와 β 라 할 때, $|\alpha - \beta|$ 의 값은?

- ① 0 ② $\sqrt{2}$ ③ $\sqrt{3}$ ④ $2\sqrt{2}$ ⑤ $2\sqrt{3}$

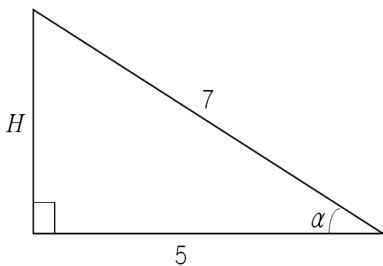
10. $\sum_{k=1}^3 k^2$ 의 값은 얼마인가?

- ① 6 ② 8 ③ 11 ④ 14 ⑤ 27

11. 퀴리부인이 발견하고 그녀 조국의 이름을 딴 원소는?

- ① 라듐 ② 우라늄 ③ 게르마늄
 ④ 폴로늄 ⑤ 플루토늄

12. 그림에서 H 의 값은 얼마인가?



- ① $\sqrt{10}$ ② $2\sqrt{3}$ ③ 4 ④ $2\sqrt{5}$ ⑤ $2\sqrt{6}$

13. GB(기가바이트)의 'G'와 '마이크로소프트'의 기호 μ 의 비 (G/μ)의 값은 얼마인가?

- ① 10^6 ② 10^9 ③ 10^{12} ④ 10^{15} ⑤ 10^{18}

14. 관성의 법칙과 관련된 속담으로 다음 보기 중에서 옳은 것만을 모두 고른 것은?

〈보기〉

ㄱ. 처녀가 배부르랴. ㄴ. 아니 땀 굴뚝에 연기 날 리 없다. ㄷ. 몸이 멀어지면 마음도 멀어진다.
--

- ① ㄱ ② ㄷ ③ ㄱ, ㄴ ④ ㄴ, ㄷ ⑤ ㄱ, ㄴ, ㄷ

15. 만유인력에 관한 설명으로 다음 보기 중에서 옳은 것만을 모두 고른 것은?

〈보기〉

ㄱ. 거리 제곱에 반비례해서 약해진다. ㄴ. 진공 중에 있는 물체들 사이에는 만유인력이 작용하지 않는다. ㄷ. 지구가 사과를 당기는 힘과 사과가 지구를 당기는 힘의 크기는 같다.

- ① ㄱ ② ㄴ ③ ㄱ, ㄷ ④ ㄴ, ㄷ ⑤ ㄱ, ㄴ, ㄷ

16. 물 1리터의 무게는 몇 뉴턴(N)인가?

- ① 0.1 ② 1 ③ 10 ④ 100 ⑤ 1,000

17. 뉴턴의 제 3 법칙은 작용-반작용의 법칙이다. 이를 잘 표현하는 다음 문장의 괄호 안에 들어갈 단어는?

We cannot kiss () being kissed.

- ① by ② to ③ that ④ except ⑤ without

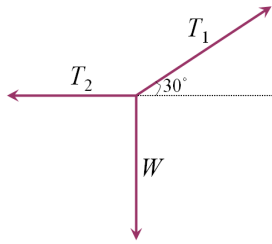
18. 다음 보기 중에서 옳은 것만을 모두 고른 것은?

〈보기〉

ㄱ. 원소기호는 양성자 수에 의해 결정된다. ㄴ. 양성자 수와 전자 수의 합을 원자량이라 한다. ㄷ. 중성자로만 구성된 원자핵을 가진 원자를 중성원자라 한다.
--

- ① ㄱ ② ㄴ ③ ㄱ, ㄴ ④ ㄱ, ㄷ ⑤ ㄴ, ㄷ

19. 그림에서 힘들의 가로 성분이 비기기 위한 식은 무엇인가?



- ① $T_1 \cos 30^\circ = T_2$
- ② $T_1 \sin 30^\circ = T_2$
- ③ $T_1 \tan 30^\circ = T_2$
- ④ $T_1 \cos 30^\circ = W$
- ⑤ $T_1 \sin 30^\circ = W$

20. 다음 보기 중에서 옳은 것만을 모두 고른 것은?

〈보 기〉

- ㄱ. 제동거리는 음주운전과는 무관하다.
- ㄴ. 공주거리와 제동거리의 합을 정지거리라고 한다.
- ㄷ. 주행속도가 두 배가 되면 제동거리는 네 배가 된다.

- ① ㄱ
- ② ㄴ
- ③ ㄴ, ㄷ
- ④ ㄱ, ㄷ
- ⑤ ㄱ, ㄴ, ㄷ