

컴퓨터화 변형 선다형 시험 방식에서 감점제가 시험 점수와 반응 포기에 미치는 영향*

송민해

박주용[†]

서울대학교 심리학과 & 심리과학연구소

정보통신 기술을 활용한 평가가 미래의 교육 평가를 주도할 가능성이 높은 가운데 이에 대한 국내 연구가 활발하지 않다. 컴퓨터를 이용하여 평가를 하게 되면 비용을 줄일 수 있을 뿐만 아니라 이전에는 가능하지 않은 방식으로 학생들의 수행 능력을 측정할 수 있다. 이런 맥락에서 본 연구에서는 현재 우리나라 교육 현장에서 가장 광범위하게 사용되는 선다형 방식의 문제점을 극복하는 방안을 제시하고 그 효과를 검증하였다. 답지가 문제와 함께 제시되는 선다형 시험 방식은 채점이 쉽지만, 학생들이 정답을 몰랐다가 답지를 보고 정답을 찾아낼 수 있다는 단점을 가지고 있다. Park(2005)은 컴퓨터의 상호작용성을 이용하여, 문제만 먼저 제시하고, 답지는 학생이 답을 먼저 생각하게 하고 나서 요청하면 짧게 제시되도록 하고 그 때 선택하도록 하는, 컴퓨터화 변형 선다형 시험 방식을 개발하였다. 본 연구는 이 방식에 감점제를 추가하면 학생들이 잘 모르면서도 답지를 보고 정답을 찾고자 할 가능성을 낮출 수 있는지를 알아보기 위해 수행되었다. 반응을 포기하면 감점이 없지만, 반응을 해서 틀릴 경우에는 감점이 이루어진다는 지시와 함께, 116 명의 대학생을 대상으로 상식 시험을 보도록 하였다. 조작된 변인은 감점 비율과 시험 방식이었다. 높고 낮은 비율과 전통적 선다형과 변형 선다형 방식이 교차된 4 실험 조건이 사용되었다. 종속측정치는 시험점수, 반응을 포기한 문항수, 그리고 자기보고 점수였다. 이들에 대한 이원 분산분석 결과, 반응을 포기한 문항수는 컴퓨터화 변형 선다형 방식에서 유의미하게 많았고 시험 점수는 높은 감점 조건에서 유의미하게 낮아짐을 발견하였다. 논의에서는 실제 교육 현장에서 변형 선다형 시험방식에 감점제를 적용한 평가가 적용될 가능성과 함께 컴퓨터화 평가 도구 개발에 대한 인지과학적 연구의 활성화를 촉구하였다.

주제어 : 선다형, 컴퓨터화 시험, 변형 선다형, 감점제

* 이 논문은 2017년도 서울대학교 학부생을 위한 사회과학 Research Grant의 지원을 받아 수행된 연구임.

† 교신저자: 박주용, 서울대학교 심리학과, (08622) 서울 관악구 관악로 1

연구분야: 인지 심리학

Tel: 02-880-9050, E-mail: jooyoung@snuc.ac.kr

세계적 수준의 인터넷 기반 시설이 있음에도, 정보통신 기술을 활용하는 평가 (e-assessment)에 대한 연구는 국내에서 그리 활발하지 않다. 호주에서는 학생들로 하여금 자기 컴퓨터를 가져와서 시험을 보도록 하는 연구가(<http://www.transformingexams.com/>) 정부의 지원을 받아 여러 대학이 동시에 진행되고 있다. 유럽과 미국에서는 평가의 전 과정을 디지털화하는 시스템을 이미 상용화하였다. 덴마크의 wiseFlow(<https://europe.wiseflow.net/>)와 미국의 QuestionMark(<http://www.questionmark.com>) 등은 그 좋은 예들이다. 이러한 기업들은 과제나 평가의 계획 실시 결과 보고를 하나의 플랫폼에서 이루어질 수 있도록 하여, 평가로 인한 행정 부담을 덜어주고 자료를 반영구적으로 보관할 수 있게 한다. 이런 맥락에서, 본 연구에서는 현재 국내에서 가장 널리 쓰이는 선다형 방식의 한계를 개선하기 위한 방안을 탐색하고 그 효과를 검증하였다.

답지가 문제와 함께 제시되는 선다형 시험 방식은, 110년의 역사를 자랑한다. 이 방식은 1907년에 당시 미국의 하버드 대학에서 비교 동물 심리학을 연구하던 로버트 여키스 (Robert M. Yerkes)에 의해 만들어진 것으로 알려져 있다 (Trewin, 2007). 선다형 시험 방식에서 하나의 문항은 발문 (stem)과 답지(choices)로 구성된다. 답지는 다시 정답과 방해자극(distractors)으로 나뉜다. 학생들은 제시된 발문을 읽고 여러 방해자극 가운데 제시된 정답을 찾아내어 반응해야 한다.

선다형 방식이 개발되기 전까지의 평가는 논술이나 구술시험이었다. 이 방법들은 나름대로 오랜 전통을 갖고 있지만, 평가에 걸리는 시간은 물론 채점 부담이 크다는 한계가 있었다. 다행히 학생들이 많지 않았기에 그렇게 해도 평가가 유지될 수 있었다. 하지만 평가를 해야 할 대상이 수 백, 수 천, 수 만 명으로 늘어나게 되면 평가로 인한 시간과 부담이 엄청나게 커진다. 선다형 시험은 이런 문제를 간단히 해결할 수 있었다. 정답 칸에 구멍을 뚫어 각 답지 위에 올려놓고 맞은 숫자만 세면되기 때문이다. 오늘날에는 광학 판독기를 이용하면 수십 만 명의 결과를 한 두 주일이면 알 수 있게 되었다. 선다형 시험 방식의 또 다른 장점은 평가 범위에 크게 영향을 받지 않는다는 점이다. 많이 가르치면 그만큼 평가 문항수를 늘리면 된다. 논술의 경우 여러 문항을 제시하면 시험 시간이 길어져야 하지만 선다형의 경우에는 비교적 짧은 시간에 많은 문항을 포함시킬 수 있다. 따라서 넓은 범위에 걸쳐 평가를 해야 할 경우 그 범위 내의 모든 내용이 골고루 망라될 수 있다 (Wainer & Thissen, 1993).

물론 선다형에는 장점만 있는 것은 아니다. 답지가 주어지 있는 구조적 특성으로 답을 모르더라도 제한된 답지를 보고 맞힐 수 있는 역행 풀이로 인해 정확한 평가가 이루어지기 힘들 수 있다 (박주용, 2009; Park, 2009). 다시 말해, 정답을 골랐더라도 제대로 알고 골랐는지 아니면 모르고도 답지를 활용하거나 아니면 우연한 추측으로 맞추었는지를 구별할 수 없다. 또한 수험자들이 선다형 시험에서 좋은 결과를 얻을 수 있도록 교수가 지식이 아닌 문제를 푸는 스킬에 초점을 맞추어 가르치게 되어 중요한 지식을 소홀히 다루게 된다는 단점이 있다 (Agrawal, 2004). 선다형 문항에 대한 또 다른 비판은 일상적인 장면에서는 거의 등장하지 않는 인위적인 상황이라는 것이다(Veloski, Rabinowitz, Robeson, & Young, 1999). 일상 장면에서는 답지가 주어지는 재인

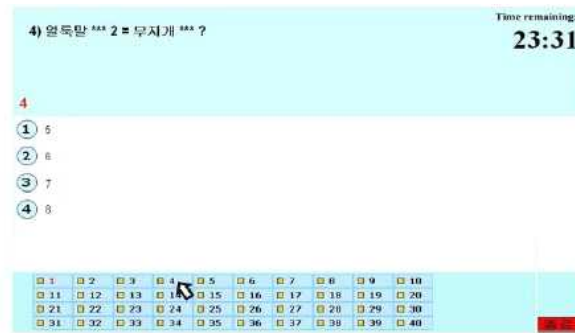
(recognition)보다는 답을 스스로 만들어내야 하는 회상(recall)이 더 흔하다. 이 밖에도 그럴 듯하고 매력적인 방해자극을 제작하기가 쉽지 않아 문항 제작에 많은 시간이 소요된다는 점도 지적될 수 있겠다. 시험에 따라서는 제한된 문항을 가진 문제 은행을 이용하는데 이 경우 시험 종료 후 답을 공개할 수 없어, 시험에 대한 피드백을 받기 어렵다(Roediger & Marsh, 2005)는 점도 약점으로 언급될 수 있다.

이런 여러 선다형 방식의 단점을 극복하는 동시에, 장점을 살리기 위해 여러 시도가 이루어졌다. 확장된 답지를 이용하는 방법과 그리드(grid) 문항은 그 좋은 예이다. 확장된 답지를 이용하는 방법은, 주로 의대에서 보는 시험에서 많이 사용되는데, 수험자에게 수천 개의 항목이 순서대로 배열된 책자를 제공하고 이를 이용하여 답을 하게 한다. 예를 들어 정답이 심장인 경우 심장 앞에 붙여진 #3537을 써넣도록 하고, 정답이 전두엽일 경우는 전두엽 앞에 붙여진 #1372번을 쓰면 된다. 그렇지만 미리 이런 책자를 만들어 제공하는 일은 그 자체로 쉽지 않아 의과대학 외에는 많이 사용되지 않는다 (Schuwirth, & van der Vleuten, 2003). 그리드 문항은 사실상 답지가 없다. 통상 각 문항마다 한 줄로 1~5번으로 구성된 답지가 사용되는데, 그리드 문항에서는 한 문항 당 예를 들면 3줄이 할당되어 있고 각 줄마다 0 ~ 9 와 / 와 같이 11개의 칸이 제시된다. 이들을 이용하면 예를 들면 1/3과 같은 답을 학생들이 입력할 수 있다. 첫 칸에는 1 두 번째 칸에는 / 3번째 칸에는 3을 채우면 된다. 최근에는 직접 숫자를 써넣게 하고 이를 컴퓨터로 읽는 방식을 취하기도 한다. 이상의 방법을 사용하면 답지를 제시하지 않아도 되지만, 숫자에 국한된 반응만 가능하다는 한계가 있다.

선다형 문항의 단점을 극복하는 또 다른 방법은 컴퓨터의 상호작용성을 이용하는 것이다. Park (2005, 2009)은 교사들이 쉽게 사용할 수 있으면서도 전통적 선다형 시험 방식의 문제를 해결할 수 있는 한 방안을 제시하였다. 이 방안은 컴퓨터를 이용한 변형 선다형 시험(Computerized Modified Multiple-choice Testing: CMMT)으로 명명되었다. CMMT는 기존의 컴퓨터를 이용한 시험 방식과 달리 컴퓨터와 수험자 간의 상호작용성을 활용하여 단답형 시험에서와 같이 문제를 풀도록 하되 반응은 선다형으로 하도록 하는 방식이다. 이를 단계별로 설명하자면 다음과 같다. 먼저 수험자에게 답지가 없이 문제만을 보여준다. 특정한 문제에 대해 답을 스스로 찾았다고 생각되면 마우스를 클릭하여 답지 제시를 요구한다. 답지는 정해진 시간만큼 짧게 제시되며, 수험자는 그 시간 내에 반응해야 한다. 그 시간이 지나면 문제와 답지 모두 화면에서 사라져 더 이상 반응을 할 수 없게 된다. CMMT에서의 답지는 단답형에서와 같이 스스로 생각해 낸 내용을 확인하는 용도로 사용하는 것이다.

이렇게 하면 단답형에서와 같이 인출의 난이도를 증가시켜 학습효과를 높일 수 있다 (Bjork, 1994; Park, 2005). 실제로 Park(2005)은 초등학생을 대상으로 컴퓨터를 이용하여 두 가지 방식으로 중간시험을 본 다음 며칠 후 최종 단답형 시험에서의 차이를 알아보았다. 한 집단의 학생들에게는 답지가 문제와 함께 주어진 전통적 선다형으로 시험을 보도록 하고, 다른 집단의 학생들

에게는 문제만 먼저 제시되는 CMMT로 중간시험을 보도록 하였다. 며칠 후 두 집단 모두에게 단답식으로 최종 시험을 보도록 하였다. 그 결과 CMMT 방식으로 중간 시험을 본 집단이 최종 시험에서 더 높은 점수를 받는다는 것을 관찰하였다. 이는 CMMT 방식으로 시험을 보는 것이 선다형으로 시험을 볼 때보다 기억을 향상시킴을 시사한다. 이런 결과는 학생들이 컴퓨터로 보는 중간시험을 집에서 보게 했을 때도 나타난다 (Park & Choi, 2008).



(a)



(b)

[그림 1] 선다형 방식과 변형 선다형 방식의 차이

(a) 선다형의 경우 1번 문항은 이미 풀었기 때문에 문제제시 박스의 1번 버튼이 붉은 색(혹백으로 출력시 회색으로 보임)으로 변화되었다. 마우스가 4번 버튼위에 놓여 있기 때문에 4번 문항의 발문과 함께 답지도 동시에 제시되어졌다. (b) 변형선다형의 경우 좌측화면은 1번 문항은 이미 풀었기 때문에 하단의 문제 제시 박스가 사라졌고, 마우스가 현재 4번 박스 위에 놓여 있기 때문에 4번 문항의 발문이 제시되었다. 하지만 답지는 제시되지 않았다. 우측 화면은 수험자가 마우스를 클릭하면 그때 답지가 제시되는데 답지는 4초 동안만 제시된다. 현재는 3초가 지난 상태이다. 수험자는 답지가 사라지기 전에 답지를 선택하여야 한다.

(Note. “컴퓨터를 이용한 변형 선다형 방식의 신뢰도 분석”. 박주용. (2009) 교육평가연구, 22, 743쪽.)

(그림 1) 변형선다형시험 방식의 구조

CMMT 방식으로 시험을 보면 선다형으로 볼 때보다 나중에 시험 점수를 더 향상시키는 현상은, 각 방식으로 시험을 보되 정답을 알려주지 않고 점수만 알려주더라도 나타난다 (Park, 2009). 언어 유추검사로 동일검사에 대한 검사-재검사 상관 비교에서도 CMMT의 신뢰도가 더 높은 것으로 나타났다 (박주용, 2009). 이상의 결과는 수렴적으로, 답지를 나중에 짧게 제시하기 때문에 인출 부담이 증가되는 동시에 답지를 이용한 추측이 어느 정도 배제되기 때문인 것으로 보인다. 따라서 웹을 이용한 시험이 평가에 더 많이 활용되게 되면, 선다형을 대신하여 CMMT가 더 널리 쓰일 가능성을 보여준다.

CMMT가 전통적 선다형의 한계인 역행포기를 효과적으로 줄이기는 하지만, 그렇다고 이를 완전히 배제하지는 못한다. 수험자들이 여전히 답지가 제시되는 짧은 시간을 이용하여 정답을 고를 수 있기 때문이다. 이 문제를 해결하는 한 방법은 감점제를 도입하는 것이다. 감점제란 답에 대한 확신이 없을 경우 응답 하지 않도록 하며 답을 틀릴 경우 일정한 점수를 빼는 방식이다. 감점제의 효과로 인해 실제로 몇몇 시험에서 사용되고 있다. 미국에서는 GRE나 SAT 등에 이 감점제를 도입하여 시행하고 있고, 우리나라에서도 치의학전문대학원 입학에 위한 공간능력검사를 비롯한 몇몇 시험에서 감점 제도를 적용하고 있다. 그러나 감점제는 일반적으로 사용되지 않고 있다. 가장 큰 이유는 감점제를 적용하려면 반응을 포기한 문항수와 반응하였지만 틀린 문항수 등을 일일이 조사한 다음 감점 비율을 적용해야 하는데 이로 인한 추가 부담이 생기기 때문이다. 컴퓨터를 이용하여 시험을 볼 경우 이런 계산이 상대적으로 쉽다.

변형 선다형 시험방식은 처음에 문제의 내용만 제시되고 선택지가 제시되지 않으므로 선지를 보고 답을 추론하는 행위가 선다형 시험 방식에 비해 훨씬 줄어들 것으로 예상할 수 있다. 실제로 이런 예상을 바탕으로 이기은 (2010)은 감점제의 효과를 알아보는 세 개의 실험을 수행하였다. 실험 1과 2에서는 초등학생 290명을 대상으로, 한 집단으로 하여금 선다형 방식으로 다른 집단은 CMMT 방식으로 각각 시험을 보도록 하였다. 그 결과 감점제를 도입하면, 선다형보다 CMMT 방식에서 시험 점수와 이전의 학업성취도와 상관이 유의미하게 높음을 발견하였다. 이 결과는 학업성취도가 높은 학생들이 감점제로 인한 위험부담을 더 잘 파악한데서 비롯되었을 것으로 추측해볼 수 있다. 하지만 이 두 실험에서는, 한편으로는 감점제에 대한 이해 부족과 다른 한편으로는 감점으로 인한 점수 하락이 큰 의미가 없었기 때문에, 반응을 포기한 문항 수에 대해서는 두 집단 간에 차이가 나타나지 않았다. 실험 3에서는 40명의 대학생을 대상으로 감점제를 적용한 상태에서, 20명은 선다형으로 20명은 CMMT 방식으로 시험을 보도록 하였다. 그 결과 반응을 포기한 문항 수는 선다형보다 CMMT 방식에서 유의미하게 더 많다는 것을 확인하였다.

본 연구는 이기은(2010)의 연구를 반복 검증하는 동시에, 감점 비율을 다르게 하였을 때의 반응 양상을 알아보기 위해 수행되었다. 감점제의 효과에 대한 반복검증과 함께 감점 비율 효과를 알아보려는 이유는, 감점제가 학생들의 상위인지적 판단 능력을 강화하는데 도움을 줄 수 있기 때문이다. 또한 컴퓨터는 물론 태블릿 PC와 스마트폰의 보급으로 웹을 이용한 시험이 용이해짐

에 따라 감점제의 적용가능성이 높아지고 있기 때문이다.

실 험

본 실험은 시험 방식에 따른 감점제의 영향과 감점 비율의 효과를 알아보기 위해 수행되었다. 이를 위해, 시험 방식으로는 전통적 선다형과 CMMT 방식이 감점 비중은 높은 감점비율과 낮은 감점비율이 각각 조작되었다. 실험을 통해 확인하고자 하는 가설은, 1) 감점 비율이 낮을 때보다 높을 때 반응을 포기한 문항수가 많아지고, 2) 선다형에서보다는 CMMT 방식에서 반응을 포기한 문항수가 많을 것이라는 것이다. 본 연구에서는 각각 5 개의 답지가 제시되는 20개의 문항이 사용되었다. 낮은 감점 비율 조건에서는 반응을 하지 않으면 감점이 없지만, 반응을 해서 틀리면, 1점을 감점하였다. 이는 한 문항에 할당된 5점의 20%에 해당하는데, 우연 수준으로 정답을 맞히는 비율이었다. 높은 감점 비율 조건에서는 낮은 비율 조건의 2배인 2점을 감점하였다. 시험 방식과 감점 비율을 교차시킨 4 조건 각각에 30명의 실험 참여자가 할당되었다.

방 법

참가자

서울 소재 S 대학교에 재학 중인 학부생 116명을 대상으로 진행하였다. 실험 참가자의 평균 연령은 22.8세($SD=2.4$)였으며, 남자는 53명, 여자는 63명이었다. 이들은 실험 참가의 댓가로 소정의 보상을 지급받았다.

실험 도구

시험 문항

실험 자극은 일반 상식으로 예비 연구를 통해 선정된 20개의 5지 선다 문제가 사용되었다. 5명을 대상으로 한 예비연구 결과 전체 시험의 정답률은 50%였다.

시험 방식

실험에서 사용된 시험 방식은 2가지로 전통적 선다형과 CMMT 방식이었다. 총 시험 시간은 15분으로 동일하였다. 선다형의 경우 15분 동안 모든 문제를 풀거나 수정할 수 있었다. 하지만 변

형 선다형 시험 방식에서는 15분 동안 문제를 볼 수 있었지만, 일단 반응을 하고 나면 그 문제는 더 이상 수정할 수 없었다. 각 문항의 답지를 볼 수 있는 시간은 5초였다. 한 문제에 5점으로 20개 문제가 제시되었다. 본 연구에서의 감점제는 두 가지로 감점 비율이 높고 낮은 조건이다. 감점 비율이 낮은 조건의 경우 답을 추측해서 맞힐 기댓값만큼의 감점을 가한다. 즉 한 문항당 답지가 5개로 추측해서 답을 맞힐 확률인 0.2로 맞힐 기댓값은 5×0.2 로 감점 비율이 낮은 조건에서는 1점의 감점을 하였다. 감점 비율이 높은 조건에서는 낮은 조건의 2배 만큼의 감점을 하여 총 2점을 감점하였다. 문제를 모르는 것을 인정하고 반응하지 않을 경우 점수를 얻지 못하지만 잃지도 않았다. 모든 실험은 웹(web) 상에서 진행되었고 시험 직후 점수에 대한 피드백을 받았다.

자기보고

실험 참여자가 본 실험에서 모르는 문제에 대해 얼마나 추측했는지를 묻는 설문이 실시되었다. 설문은 5점 척도로 구성된 한 개의 문항이었다. 점수가 1점에 가까워질수록 답지를 보고 정답을 추측하는 역행 풀이를 하지 않았음을 의미하고 5점에 가까울수록 역행 풀이를 많이 했음을 의미한다.

실험 절차

실험 참여자들은 무작위로 시험방식과 감점의 방식에 따라 다른 4개의 집단으로 나뉘어 할당되었다. 실험은 외부와 차단된 실험실에서 개별적으로 진행되었으며 실험 진행 방법 및 내용에 대해 설명서를 읽은 후 실험 참가 동의서에 서명하였다. 그 다음 실험 참여자가 시험방식과 감점제에 익숙하도록 5분 동안 연습문제를 풀어보게 하여 작동되는 방법을 학습시킨 후 본 실험에 도입하였다. 본 실험에서 참여자들은 약 15분 간 총 20개의 5지 선다형 문제를 풀었다. 이 때, 실험 참여자들이 성실히 답을 응답하도록 동기를 부여하기 위해, 점수가 각 집단의 하위 20%에 해당되는 경우 재시험을 볼 수 있다는 눈속임 지시가 이루어졌다. 문제를 풀 이후 점수를 확인함과 동시에 통과되었다고 안내한 후 모르는 문제에 대해 얼마나 추측했는지 설문에 응답하도록 하였다.

실험 설계 및 분석

본 연구에서 시험의 종류와 감점의 크기에 따른 모르는 문제의 추측 행위에 미치는 영향을 알아보기 위해, 시험 방식과 감점 비율을 두 요인으로 하는 이원 변량분석을 실시하였다. 분석에 앞서, 실험에 성실하게 참여하지 않은 4명의 실험 참여자는 분석에서 제외시켰다. 분산 분석에 이어 각 시험 방식에서 감점 비율의 효과를 알아보기 위한 계획된 (planned) t 검증이 실시되었다.

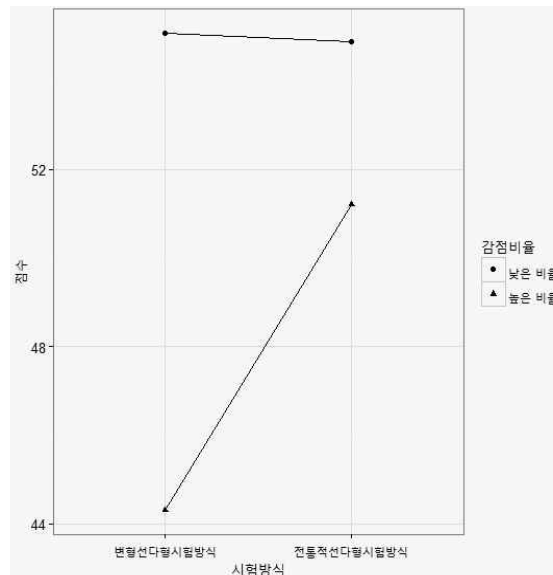
결 과

각 조건 내의 참가자들로부터 수집한 반응하지 않은 문항수, 점수 그리고 자기보고 점수의 평균과 표준편차는 표 1에 제시되었다. 시험 점수를 시험 방식과 감점 비율을 기준으로 그린 그래프는 그림 2에 제시되었다. 시험 점수를 변량분석 결과, 시험 방식에서는 조건 간에 차이가 없었고, 감점 비율에서만 통계적으로 유의미한 차이를 보였다 ($F(1,112) = 8.8, p < .01, d = .1$). 즉, 감점 비율을 높을 때의 점수가 낮을 때의 점수보다 낮았다 ($M = 47.8$ 대 55). 그리고 이 시험 방식과 감점 비율 간의 상호작용도 나타나지 않았다.

시험 방식에서 차이가 나타나지 않은 결과는 얼핏 보면 이기은 (2010)의 실험 3의 결과를 반복 검증하는데 실패한 것으로 보일 수 있다. 하지만 감점 비율을 고려하면 그렇지 않다. 이기은

〈표 1〉 반응을 포기한 문항수, 점수, 자기보고 점수의 평균과 표준편차

	CMMT *높은 감점	CMMT *낮은 감점	전통적 선다형 *높은 감점	전통적 선다형 *낮은 감점
실험참여자 수	30	28	30	28
반응을 포기한 문항수	4.3(2.0)	3.0(2.1)	2.9(1.7)	2.6(2.0)
점수	44.3(11.9)	55.1(13.7)	51.2(14.8)	54.9(11.9)
자기보고 점수	2.8(1.2)	3.1(1.1)	3.1(1.1)	3.1(1.1)

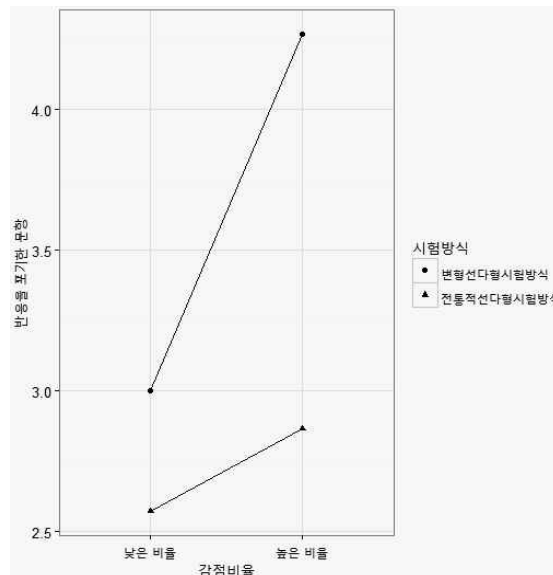


(그림 2) 점수에 대한 상호작용 그래프

(2010)의 실험 3에서의 감점 비율은 0.33으로 본 실험의 높은 감점 비율인 0.4에 더 가깝다. 그림 2에서 볼 수 있듯이 낮은 감점 비율의 경우 시험 방식 간에 차이가 없었다. 하지만 높은 감점 비율 조건에서는 두 집단 간에 유의미한 차이가 관찰되었다 ($t(55) = 2.0, p < .05, d = .5$). 이 결과는 이기은(2010)의 실험 3에서 얻어진 결과에 부합된다. 두 실험을 종합해보면, 감점제가 시험 방식에 따라 점수에 영향을 주기 위해서는 우연 수준보다 높은 감점 비율이 적용되어야 함을 시사한다.

반응을 포기한 평균 문항 수를 시험 방식과 감점 비율을 기준으로 그린 그래프는 (그림 3)에 제시되었다. 반응을 포기한 문항수에 대한 변량분석 결과, 시험 방식에 따라 반응을 포기한 문항수는 통계적으로 유의미한 차이를 보였다($F(1, 112) = 6.5, p < .05, d = .04$). 즉, 선다형 보다 CMMT 방식에서 반응하지 않은 경우가 더 많았다 ($M = 3.7$ 대 2.8). 이 패턴은 이기은 (2010)의 실험 3에서의 결과와 일치한다. 감점 비율의 주효과도 유의미하였다, ($F(1, 112) = 4.5, p < .05, d = .02$). 그렇지만 두 변인 간 상호작용 효과는 나타나지 않았다($F(1, 112) = 1.8, p > .05$). 각 시험 방식에서 감점 비율의 효과를 알아보기 위해, 사전에 계획된 t 검증을 실시하였다. 그 결과 CMMT 방식에서는 유의미한 차이가 있었지만 ($t(55) = 2.3, p < .05, d = .6$), 선다형 방식에서는 통계적인 차이가 관찰되지 않았다. CMMT에서만 감점 비율이 영향을 미친 이유는, 답지가 보이지 않는 상황에서 감점 비율이 더 큰 부담을 주었기 때문인 것으로 해석된다.

이상의 결과는 1) 감점 비율이 낮을 때보다 높을 때 반응을 포기한 문항수가 많아지고, 2) 선다형에서보다는 CMMT 방식에서 반응을 포기한 문항수가 많을 것이라는 본 연구의 가설을 지지



(그림 3) 반응을 포기한 문항수에 대한 상호작용 그래프

한다.

한 문항으로 이루어진 자기 보고 점수에 대한 분석결과로 부터는 유의미한 차이를 발견할 수 없었다. 다만 감점 비율이 높은 CMMT 조건의 실험참여자들이 다른 세 집단에 비해 자기보고 점수가 약간 낮은 경향성을 보였다. 즉 감점비율이 높은 CMMT 조건에서는 추측을 시도하지 않으려는 경향이 있었다.

논 의

선다형 시험 방식은 채점 부담이 적으며 넓은 범위에 걸친 성취도를 쉽게 측정할 수 있어, 전 세계적으로 광범위하게 사용되고 있다. 그러나 깊은 사고력을 측정하는 문항을 만들어내기 어렵고, 선지가 미리 제시되는 구조적 특성으로 인해 정확한 답을 모르더라도 답지를 이용하여 정답을 맞힐 수 있다는 단점이 있다. 단점을 보완하기 위해 고안된 여러 방법들 중, 본 연구에서는 CMMT 방식에 감점제를 추가하는 방안이 주목하였다. CMMT 방식은 문제만 제시하고 답지는 필요할 때 잠깐 보여주기 때문에, 선다형 시험 방식의 단점인 역행 풀이를 효과적으로 줄일 수 있다. 그러나 답지가 제시되는 짧은 시간을 이용하여 정답을 고를 수 있는 가능성은 여전히 남아 있기 때문에, 본 연구에서는 CMMT에 감점제를 추가하고자 하였다. 감점제는 모르는 문제에 대해 반응을 하지 않으면 감점이 없지만, 풀려고 반응을 해서 틀릴 경우 점수를 깎는다. 이렇게 하면 추측으로 정답을 맞힐 가능성을 줄여, 수험자의 수행 수준을 더 정확하게 파악할 수 있게 된다. 수험자에게는 자신이 아는 것과 모르는 것을 명확히 구분할 수 있도록 하여, 자신의 성취도에 대한 정확한 평가를 하도록 할 수 있다. 이런 장점에도 불구하고, 감점제는 아직 널리 활용되지 않고 있다.

본 연구는 정보통신 기술의 발달로 컴퓨터화 시험이 점차 확산될 가능성을 염두에 두고, CMMT 방식에 감점제를 추가하였을 때 실제로 수험자들이 추측행동을 줄이는 지를 알아보고자 수행되었다. 대학생 120명을 대상으로 한 실험을 통해, 선행 연구의 결과를 반복검증하는 동시에, CMMT 방식에서 감점제가 효과적으로 활용되려면 감점 비율이 우연 수준 이상이 되어야 한다는 점을 새롭게 확인할 수 있었다.

구체적으로 감점제를 적용하되 감점 비율이 우연에 의해 정답을 맞힐 수 있는 수준에서는 시험 방식에 따라 시험 점수에서는 물론 반응을 포기한 문항수에서 차이가 없었다. 그렇지만 감점 비율이 높아지면 시험이 끝날 때까지 반응을 포기한 문항의 수가, 선다형에 비해 CMMT 방식에서 유의미하게 많아졌다. 선행 연구를 통해 CMMT 만으로도 답지를 이용한 역행 풀기를 유의미하게 감소시킬 수 있다는 것이 밝혀졌는데, 본 연구에서는 여기에 높은 비율의 감점제를 추가하면, 그 효과가 더 두드러질 수 있다는 점이 확인되었다. 한 문항만 사용되었기 때문인지 얼마만

큼 추측을 했는지에 대한 자기보고 점수에서는 조건 간에 유의미한 차이가 발견되지 않았다.

CMMT 방식을 이용하거나 여기에 감점제를 추가하는 이유는 학생들로 하여금 스스로의 수행 수준에 대해 정확한 판단을 내릴 수 있도록 돕기 위해서이다. 혹자는 답지를 이용하여 정답을 찾아내는 것도 하나의 능력으로 볼 수 있다고 주장할 수 있다. 하지만, 서론에서 언급된 것처럼, 선다형 시험 방식은 그야말로 평가를 위해 만들어진 인위적인 검사 방식이라는 점을 신중히 고려할 필요가 있다. 선다형에서 점수를 높이는 전략이나 능력은, 사실 평가의 예언 타당도를 낮출 수 있기 때문이다. 즉 시험 점수는 높인데, 정작 실제 장면에서의 수행이 형편없을 수 있다는 것이다. 이런 일이 벌어지지 않도록 하려면, 학생들이 자신이 정확히 아는 것과 모르는 것을 잘 구별할 수 있도록 도와야 한다. 기존의 선다형 방식은 답지를 활용하여 정답을 찾아낼 가능성이 있고 실제로 수험자들이 그런 행동을 하기 때문에 그로 인해 검사의 타당도가 낮아지는 문제가 있다. 본 연구는 바로 이런 맥락에서 CMMT 방식에 감점제를 도입하였고 이런 조작을 통해 수험자가 추측을 하지 않도록 하는데 어느 정도 효과가 있다는 점을 확인하였다.

본 연구는 CMMT 방식은 물론 감점제의 활용가능성을 높이는 한 실증적 근거를 제공한다. 하지만 실험적 연구가 갖는 한계를 포함하여, 실제 평가 장면에서 사용되기 전에 밝혀져야 할 여러 과제가 남아 있다. 우선 본 연구는 실험실에서 얻어진 결과라는 한계가 있다. 최소한의 변인 조작을 통해 그들 간의 인과 관계를 밝힐 수 있지만 주로 대학생 실험참여자를 대상으로 상식 문제를 사용하여 얻어진 결과이다. 따라서 더 넓은 범위의 실험참여자를 대상으로 하여 다양한 내용에 대한 검토가 필요하다. 더불어 감점 비율이 달라짐에 따라 반응 양상이 달라졌기 때문에 CMMT를 적용했을 때의 최적의 감점 비율을 연구할 필요가 있다. 아울러 실제 평가 장면에서 이 방식을 시범적으로 도입하는 연구도 필요하다. 기존의 평가 방식을 주로 사용하고 여기에 약간의 비중을 두어 CMMT와 감점제를 추가한 방식으로 시험을 보도록 하는 것이다. 물론 이렇게 사용할 경우 CMMT 방식에 대한 사전 훈련이 필요하다. 그리고 평가 후에 새로운 방식에 대한 수험자들의 반응을 수집하여 거부감을 줄일 수 있는 방안을 모색할 필요가 있다.

이런 한계점을 갖고 있음에도 불구하고 우연 수준 이상의 감점을 도입한 CMMT는 기존의 선다형 시험 방식의 단점이었던 모르는 문제의 답을 추측하는 행위를 방지하여 정확한 평가를 할 수 있도록 도움을 줄 수 있다. 후속 연구를 통해 언급된 한계점이 보완되고 다양한 환경에서 적용할 수 있다는 점이 확인되면 교육이나 그 밖에 다른 평가 현장에서 지금보다는 다양한 방식의 평가를 가능하게 할 수 있다. 궁극적으로는 CMMT 방식이 학생들의 성취 수준을 정확히 평가하도록 하는 한편 학생 스스로도 아는 것과 모르는 것에 대한 구분을 잘 해내도록 돕는 도구로 사용될 것을 기대한다. 아울러 단지 선다형 방식의 개선에서 그치는 것이 아니라, 심리학자와 교육 평가 연구자는 물론 컴퓨터 공학자들의 협력을 통해, 효율적이고 혁신적인 컴퓨터화 평가 도구의 개발을 서둘러야 한다. 교육에서 평가의 중요성이 재발견되면서 (예, Roediger, Putnam, & Smith, 2012), 평가의 빈도는 물론 다양한 평가 도구에 대한 요구가 증대하고 있기 때문이다.

단지 이런 요구에 부응하는 것 이상으로, 한계 상황에 도달한 우리의 교육 제도의 돌파구를 새로운 평가 도구의 개발에서 찾을 수 있기에 이에 대한 인지 과학적 연구가 활발해지기를 기대해 본다.

참고문헌

- 박주용 (2009). 컴퓨터를 이용한 변형 선다형 방식의 신뢰도 분석. *교육평가연구*, 22, 737-756.
- 이기은 (2010). *컴퓨터화 변형 선다형 시험 방식에 감점제 적용시의 반응 특성고찰*, 세종대학교 박사학위 논문.
- Agrawal, M. (2004). Curricular reform in schools: the importance of evaluation. *Journal of Curriculum Studies*, 36(3), 361-379.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human being. In J. Metcalfe & A. P. Shimamura (Eds.) *Metacognition*, MIT Press.
- Park, J. & B. Choi (2008). Higher retention after a new take-home computerised test. *British Journal of Educational Technology*, 39(3), 538-547.
- Park, J. (2005). Learning in a new computerized testing system. *Journal of Educational Psychology*, 97(3), 436-443.
- Park, J. (2009). A computerized testing system that reduces backward reasoning in multiple-choice items. *Korean Journal of Cognitive Science*, 20 (3), 275-289.
- Roediger III, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1155.
- Schuwirth, L. W., & van der Vleuten, C. P. (2003). ABC of learning and teaching in medicine: written assessment. *BMJ: British Medical Journal*, 326(7390), 643.
- Trewin, S. (2007). Robert Yerkes' Multiple-Choice Apparatus, 1913-1939. *American Journal of Psychology*, 120 (4). 645-660.
- Veloski, J. J., Rabinowitz, H. K., Robeson, M. R., & Young, P. R. (1999). Patients don't present with five choices: An alternative to multiple-choice tests in assessing physician's competence. *Academic Medicine*, 74, 539-546.
- Wainer, H., & Thissen, D. (1993). Combining multiple choice and constructed response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education*, 6, 103-118.

송민해 · 박주용 / 컴퓨터화 변형 선다형 시험 방식에서 감점제가 시험 점수와 반응 포기에 미치는 영향

1차 원고 접수: 2017. 10. 02
1차 심사 완료: 2017. 11. 14
2차 원고 접수: 2017. 11. 30
2차 심사 완료: 2017. 11. 30
최종 게재 확정: 2017. 12. 09

(Abstract)

The Effect of Penalizing Wrong Answers Upon the Omission Response in the Computerized Modified Multiple-choice Testing

Min Hae Song

Jooyong Park

Department of Psychology & Institute of Psychological Science, Seoul National University

Even though assessment using information and communication technology will most likely lead the future of educational assessment, there is little domestic research on this topic. Computerized assessment will not only cut costs but also measure students' performance in ways not possible before. In this context, this study introduces a tool which can overcome the problems of multiple choice tests, which are most widely used type of assessment in current Korean educational setting. Multiple-choice tests, in which options are presented with the questions, are efficient in that grading can be automated; however, they allow for students who don't know the answer, to find the correct answer from the options. Park(2005) has developed a modified multiple-choice testing system (CMMT) using the interactivity of computers, that presents questions first, and options later for a short time when the student requests for them. The present study was conducted to find out if penalizing wrong answers could lower the possibility of students choosing an answer among the options when they don't know the correct answer. 116 students were tested with the directions that they will be penalized for wrong answers, but not for no response. There were 4 experimental conditions: 2 conditions of high or low percentage of penalizing, each in traditional multiple-choice or CMMT format. The results were analyzed using a two-way ANOVA for the number of no response, the test score and self-report score. Analysis showed that the number of no response was significantly higher for the CMMT format and that test scores were significantly lower when the penalizing percentage was high. The possibility of applying CMMT format tests while penalizing wrong answers in actual testing settings was addressed. In addition, the need for further research in the cognitive sciences to develop computerized assessment tools, was discussed.

Key words : test format, computerized testing system, multiple-choice format, penalizing wrong answers