

오답감점형 객관식 시험에 따른 표준편차 분석

Standard Deviation Analysis According to Multiple-choice Test with Penalty Points for Incorrect Answers

최 현*

Hyun Choi*

〈Abstract〉

There are a variety of evaluation methods used to evaluate and provide feedback to students, recognize shortcomings, and establish future plans. The purpose of the academic achievement evaluation is to check whether the student understands the content taught, identify areas that the student is unfamiliar with or misunderstand, and then provide supplementary study to prevent learning deficits. Multiple choice has been the evaluation system in Korea since the early 1970s, influenced by Tyler's goal-centered evaluation model that evaluates achievement of educational goals and Bloom's educational goal classification system. The most important item in education is the evaluation of whether you have properly understood the lesson and how deeply you know it. Therefore, in this study, in order to minimize cases where test takers stochastically answer the correct answer in a multiple-choice evaluation, the change in the standard deviation of grades was analyzed by applying a grading method that assigns negative points to incorrect answers. As a result of the study, it was found that the average score of the incorrect response evaluation was 1 to 4 points lower, and the standard deviation was wide. This showed that the test taker's discrimination ability was high in the wrong answer appraisal type.

Keywords : Evaluation Methods, Tyler's Goal-centered Evaluation Model, Bloom's Educational Goal Classification System, Incorrect Response Evaluation

* 정회원, 교신저자, 경남대학교, 교수
E-mail: hchoi@kyungnam.ac.kr

* Dept. of Civil Eng., Kyungnam Univ., Professor.
E-mail: hchoi@kyungnam.ac.kr

1. 서론

학생을 포함한 교육을 받는 사람은 배운것에 대한 평가와 피드백 그리고 부족한 부분에 대해 인식하고 앞으로의 계획을 수립하기 위해 실시되는 평가 방식은 다양하다. 특히 우리나라에서는 초등, 중등교육과정에서 시행되는 '국가수준 학업성취도 평가'는 학생이 학교에서 배운 내용에 대한 이해능력을 측정하고, 교육 목표에 어느정도 도달했는지를 체계적으로 진단하기 위해 국가에서 실시하는 평가이다. 성취도 평가는 교육과정 및 교수·학습방법을 효과적으로 개선하고 교육 정책을 수립하는 기초 자료가 되며, 학교 현장의 평가방법을 개선 및 발전시키는 것이 이 시험의 목적으로 '기초 미달율'과 '보통 성적율'를 판별하는 지표로 활용되고 있다. 여러 가지 단점이 있음에도 불구하고 성취도 평가[1]는 수준높은 평가도구의 개발을 통해 일선 학교의 교수·학습 방법을 개선하고 평가방법을 선도하고 있음을 알 수 있다. 학업성취도 평가는 학생에게 가르친 내용의 인지 여부를 확인하여 잘 모르는 부분이나 잘못 알고 있는 부분을 파악한 후, 보충학습을 실시하여 학습 결손을 방지하는데 있다. 학업성취도 평가는 전술한 바와 같이 초등, 중등학생을 대상으로 교육 방법개선을 위한 자료로 활용하므로 경쟁시험이나 자격시험의 경우에는 부적합한 방법이라 할 수 있다.

우리나라에서 시행하는 자격시험의 대부분은 객관식으로 이루어져 있다. 객관식은 시험 유형 중 하나로, 보통 4~5개의 예시 후보군 중에서 정답을 고르게 하는 방식으로 진위형, 연결형(배합형), 선다형이 존재한다. 객관식은 교육목표 달성도를 평가하는 Tyler의 목표 중심 평가모형과 Bloom 등의 교육목표 분류체계에 영향을 받아서 1970년대 초반부터 객관성을 강조하는 선택형 객관식 평가가 우리나라의 평가체제로 지속되고 있다[2]. 이후로 우리나라

의 거의 모든 시험에서 선택형 객관식 문항이 당연히 평가하고 있는 것으로 인식되어 현재에 이르고 있다. 객관식은 쉬운 채점, 정량평가, 통계처리가 쉬우나, 출제자의 의도대로 생각하도록 유도하기 때문에 다양성이 인정되지 않고 있다. 객관식 평가의 가장 큰 단점으로는 수험생이 정답을 선택한 이유를 모르는 등 학생의 이해도를 측정하는 평가도구로는 활용에 제한이 있다. 이와 같은 문제를 해결하기 위해 변호사시험이나 5급 공무원 임용고시, 기술사 등에서는 주관식, 서술식 등을 추가하여 2차 평가를 시행하고 있다. 그러나 자격증, 임용 등의 시험에서는 여전히 채점이 간편하고, 통계처리가 쉬우며, 공정성이 확보된다는 이유로 객관식 평가시행하고 있다. 특히 대부분의 고등교육기관에서는 전문지식을 측정하기 위해서 주관식이나 서술식으로 시험을 치르고 있으며, 객관식 평가는 거의 이루어지지 못하고 있다. 교육에서 가장 중요한 항목은 수업을 제대로 이해하였는지, 얼마나 깊이 알고 있는지에 대한 평가이다. 따라서 본 연구에서는 객관식 평가에서 수험생의 정답 선택에 대한 학생의 이해도와 암기 등 지식기반의 정답 비율을 알고자 오답을 마이너스 점수를 부여하는 오답감점형 채점방식을 적용하여 성적 표준편차의 변화를 분석하였다.

2. 기본 이론 및 방법론

교육평가 방식에는 다양한 유형이 존재하고 있다. 평가 문항은 평가자의 개입에 따라서 분류하는 주관식과 객관식이 있으며, 학습자의 반응도에 따라서 선택형과 서답형이 있다. 객관식으로 불리는 선택형은 수험자가 제한된 선택지에서 표기하는 방식을 말한다. 선택형에는 진위형, 연결형 선다형이 있고, 서답형에는 단답형, 완성형 논술형이 있다.[3], [4] and [5]

2.1 선택형 유형

선택형 문항의 유형의 분류는 진위형(true-false type, Fig. 1)과 선다형(multiple choice type, Fig. 2) 등이 많이 쓰인다. 진위형은 수험생에게 질문지를 제시하고 정답을 고르게 하는 방식으로 수능시험, 각종 자격시험 등 우리나라에서 각종 시험제도에 가장 많이 이용되는 방식이다. 선다형은 하나의 질문지에 대하여 2개 이상의 선택지 중에서 정답을 고르게 하는 방법으로 선택형 문항에서 가장 많이 고르는 방식으로 수학능력시험, 공인중개사시험 등에서 난이도를 올리기 위해 많이 사용되는 방식이다.

20. 조선시대 양안에 대한 설명으로 가장 옳지 않은 것은?

- ① 토지를 측량하여 등록하는 오늘날의 지적공부로 볼 수 있다.
- ② 양안의 기재내용으로는 자호, 양전방향, 토지등급, 지형척수 등 다양한 내용이 있었다.
- ③ 양안의 명칭은 시대와 사용처 및 보관기관에 따라 다양하였다.
- ④ 토지의 위치를 나타낼 때 오늘날 지번의 형태로 나타내었으며, 양지 또는 양표라 하였다.

Fig. 1 true-false type

1. 그림 (가)는 전자기파 A, B를 이용한 예를, (나)는 진동수에 따른 전자기파의 분류를 나타낸 것이다.

전자레인지의 내부에서는 음식을 데우기 위해 A가 이용되고, 표시 창에서는 B가 나와 남은 시간을 보여 준다.

(가)

진동수(Hz)

라디오파 적외선 자외선 감마선

(나)

이에 대한 설명으로 옳은 것만을 <보기>에서 있는 대로 고른 것은?

<보 기>

ㄱ. A는 ㉠에 해당한다.
 ㄴ. B는 ㉡에 해당한다.
 ㄷ. 파장은 A가 B보다 길다.

① ㄱ ② ㄷ ③ ㄱ, ㄴ ④ ㄴ, ㄷ ⑤ ㄱ, ㄴ, ㄷ

Fig. 2 multiple choice type

선택형의 장점은 채점에 있어서 객관성과 신뢰성을 유지할 수 있고 학습한 내용을 전반적으로 출제 및 학생의 이해도를 파악할 수 있다. 단점으로는 단순한 기억력 평가만 할 가능성이 있어서 수험생의 표현이 제한되고 추측의 요인을 제거할 수 없다. Fig. 1은 2021년 공무원 임용시험에 기출된 문제로 4개의 경우를 두고 하나를 선택하게 하는 진위형의 가장 대표적인 방식이다.

Fig. 2는 2023년 대학수학능력시험의 예시로 몇 가지 예를 두고 정답을 선택하는 선다형이라 할 수 있다.

2.2 선택형 문제의 답안 기재방식

전술한 바와 같이 선택형에서는 진위형과 선다형의 2가지 유형으로 분류가 가능하며, 질문에 대해 결과를 구한 다음, 선택지에서 해당하는 정답을 고르는 형태로 정답이 애매하거나 모를 경우에는 확률적으로 임의의 결과를 선택하지 않도록 하였다. 와 질문지와 선택지들 사이에서 어떤 선택지가 질문에 “정답이다.” 혹은 “정답이 아니다.”을 판단하여 그 중 하나의 선택지를 고르는 형태로 나눌 수 있다. 본 연구에서는 추측요인을 제거하기 위해서 질문지에 대한 정답을 정확히 알고 맞춘 수험생 수가 최소 0명 이상이 되기 위한 조건이 되어야 한다.

2.3 표준편차

정밀도를 표현하는데 가장 많이 사용되는 것이 표준편차와 분산이다. 표준편차는 관측값으로부터 최확값을 구하고, 측정자료를 평가하는 자료로 일반적으로 사용되며 측정값의 상호편차를 의미한다. 1회 측정에 대한 표준편차(ρ_s)는 식(1)과 같다.

$$\rho_s = \pm \sqrt{\frac{\Sigma v^2}{n-1}} \tag{1}$$

여기서, ρ_s

Σv^2 : 잔차의 제곱의 합

n : 관측수

$n-1$: 잉여관측수

표준편차가 클수록 측정값의 분포도가 크다는 것을 의미하고, 표준편차가 작으면 측정값의 평균에 가깝게 접근해 있음을 의미한다. 일반적인 관점에서 측정값 등의 오차분석에서는 표준편차가 작으면 데이터의 정확도 신뢰도가 높은 결과를 나타낸다. 성적처리와 같은 경우에는 표준편차가 크면 학생평가 변별력이 높다는 것을 알 수 있다. 따라서 표준편차와 평균을 통해 학생의 성적분포 파악이 유리하고 학습 효과나 난이도 평가가 가능하다.

3. 실험 및 고찰

3.1 실험 방법 및 시험조건

본 연구를 수행하기 위한 실험대상은 대학생 2, 3학년 2과목 2회 실시하였으며, 실험은 4지 선택 형으로 모를 경우 선택을 하지 못하게 하였으며, 확률적으로 맞춤을 방지하기 위해 오답의 경우 감점이 있음을 사전공지를 하였다.(Fig. 3)

교육과정에 대한 평가의 일관성 유지를 위해 타당성 높은 평가가 이루어져야 하고 이것은 관련된 타당도 즉 측정하고자 하는 것을 얼마나 정확하고 오차없이 측정해야 하며 관련된 신뢰도를 확보 해야 한다. 본 연구에서 실험 문항은 진위형으로 하였으며, 실험에 대한 타당도를 높이기 위해서 2학년 3학년, 세 과목으로 하여 표본집단은 A과목 16명, B과목 16명, C과목 20명 D과목 20명으로 설정하여 총 72명으로 하였다. 모집단을 크게 하면 신뢰성 높은 결과가 나올 수 있으나 교육환경의

4. 유의 사항
- 미응시 할 경우 학사규정에 의거 F
 - **15점부터 시작하며 문항당 1점, 오답체크시 1점 감점이니 모르는 문제는 표기하지 말 것**
- ex 1) 5문항 정답, 오답 3문항, 오답 미체크 7문항의 경우,
총 점수 = 15(기본점수) + 5(정답) - 3(오답) = 17점
- ex 2) 5문항 정답, 오답 미체크 10문항의 경우,
총 점수 = 15(기본점수) + 5(정답) = 20점
- ex 3) 1문항 정답, 오답 14문항의 경우,
총 점수 = 15(기본점수) + 1(정답) - 14(오답) = 2점

Fig. 3 Notice to examinees

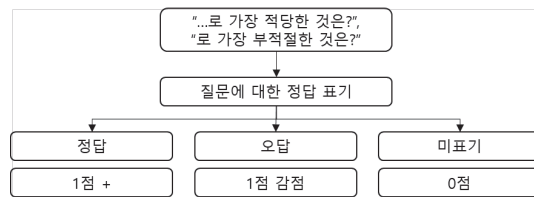


Fig. 4 Scoring method implemented in the study

특성상 한계가 있었다. 추후 시험결과와 자료가 확보되면 의미있는 결과를 얻을 수 있을 것으로 판단된다. 시험문항은 15문항으로 하였으며, 기본 15점을 전제로 15문항 전체를 맞출 경우 30점, 전체가 틀릴 경우 0점이 되도록 설정하였다.(Fig. 3)

3.2 실험 결과 및 분석

Table 1은 실험군 A, B, C, D 집단의 오답을 기재하더라도 감점이 없는 객관식 시험결과에 따른 성적의 수험인원, 최고 점수, 최저 점수, 평균과 표준편차를 나타낸 것이다. 오답을 기재할 경우 감점이 있음에도 일부 학생을 제외한 학생은 오답에 체크를 많이 하였다. 학생 대부분이 많은 학생들이 점수를 기둘 중의 하나의 기록한 것이다. 정답과 오답을 하나도 기재하지 않은 학생을 살펴

보면 A 실험군에서는 16명 중에서 3명, C 실험군에서는 20명 중에서 2명, D 실험군에서는 20명 중에서 5명으로 나타났으며, B실험군에서는 0명이었다. 정오답을 기재하지 않은 학생의 성적은 객관식과 오답감점형 모두 하위권을 차지 하였다. A 실험군의 경우 객관식에서는 하위권을 차지 하였으나 오답감점순위에서는 중위권을 보였다. A 실험군의 경우에는 사전에 공지하였음에도 불구하고, 오답에 마크를 기재하여 감점사항이 발생한 학생이 많이 분포하였기 때문이다.

Table 2은 실험군 A, B, C, D 집단의 오답감점형으로 성적을 기재한 수험인원, 최고 점수, 최저 점수, 평균과 표준편차를 나타낸 것이다. 오답을 기재할 경우 감점이 있음에도 기록한 것이다. 실험결과 표준편차는 집단 A, B, D에서는 오답감점형에서 편차가 크게 나타났으며, C 집단에서는 각각 2.48, 2.23으로 0.25가량 객관식으로 평가했을 때가 높게 나타났다.

일반적으로 과목평균이 높을 경우, 표준편차가

크다는 것은 상위권 학생과 하위권 학생의 원점수 차이가 큼을 의미하며 시험난이도가 쉬움을 알 수 있다. 반대로 표준편차가 작음은 상위권 학생과 하위권 학생의 원점수 차이가 적다. 즉, 학생의 학업 능력이 우수함을 알 수 이가 있으나 제대로 변별력을 갖추었다고는 할 수 가 없다. 반대로 과목평균이 낮을 경우에서의 큰 표준편차는 시험 난이도가 높거나, 학생 간 실력 차이가 많음을 의미하며, 표준편차가 적음은 시험 난이도가 높으나, 학생의 수준이 비슷하다고 할 수 있다. 우수한 평가가 되기 위해서는 타당도, 신뢰도 그리고 실용도가 높아야 한다. 타당도와 신뢰도를 높이기 위해서는 주관식이나 서술식이 가장 적절하나 여기에는 많은 시간과 비용이 발생하게 된다. 따라서 실용성이 높은 객관식을 선호하는 것은 당연한 결과일 수도 있다. 교육학 연구에서는 제한된 객관식 평가만으로는 학습을 제대로 측정할 수 없으므로 주관식, 서술형 평가를 함께 사용해야 한다는 주장이 존재하고 있다. 그러나 학생의 이해능력과 변별력과는 상이한 결과로 인하여 객관식의 단점을 보완하는 방법으로 오답감점형 객관식을 연구하였다. 교육은 학습목표를 학습자들에게 의도적으로 인지시키는 일련의 과정으로 평가의 방식은 교육목표에 따른 실현 정도를 파악하는 것은 매우 중요한 일이다.

시험은 관련된 타당도 즉 측정하고자 하는 것을 얼마나 정확하고 오차없이 측정해야 하며 관련된 신뢰도를 확보해야 한다. 각종 평가에서의 시험은 대부분 객관식으로 이루어지고 있으며 수험생의 변별력을 높이기 위한 많은 방법들이 존재한다. 본 논문에서 제시한 방법은 확률적으로 정답을 맞추는 것을 최소화 하는 방식으로 각종 수험생에 대한 타당성 높은 평가가 이루어지고 있음을 알 수 가 있었다. 그러나 본 연구에서 제안한 방식은 적은 수의 모집단으로 인하여 신뢰의 정량화

Tbale 1. Standard deviation of multiple choice scores

집단	수험인원	최고 점수	최저 점수	평균	표준 편차
A	16	24	15	18.60	3.06
B	16	25	16	20.13	2.74
C	20	23	15	17.95	2.48
D	20	22	15	17.95	2.63

Table 2. Standard deviation of incorrect answer evaluation multiple choice score

집단	수험인원	최고 점수	최저 점수	평균	표준 편차
A	16	21	4	14.63	3.56
B	16	23	7	16.62	3.43
C	20	21	14	16.75	2.23
D	20	22	8	16.75	3.05

에는 한계가 있다. 아울러 모집단을 크게 하면 신뢰성 높은 결과가 나올 수 있으나 교육환경의 특성상 한계가 있었다. 특히 C집단과 같이 아무것도 기재 하지 않은 학생은 기본적으로 중간점수(30점 중에서 15점)를 받았을 경우 표준편차가 줄어들어 변별력이 감소하는 결과를 나타내는 문제점이 있었다. 추후 이러한 단점을 보완하는 평가방식을 갖춘다면 오답감점형 객관식 시험은 학생이나 수험생의 변별력을 높이는 수단으로 가치가 있을 것으로 판단된다.

4. 결론

본 연구는 오답감점형 객관식 시험에 따른 표준편차에 미치는 영향에 관한 연구로 객관식 평가에서 오답감점형 평가가 평균점수는 1-4점 정도 낮게 나왔으며, 표준편차는 넓음을 알 수 있었다. 이것은 오답감점형이 수험생의 변별력이 높음을 의미한다. 따라서 객관식으로 오답감점형으로 성적을 처리하게 되면 표준편차가 넓게 나오므로 수험생의 변별력을 높일 수 있을 것으로 판단이 된다. 정답과 오답을 하나도 기재하지 않은 학생은 72명 중 10명으로 약 13.89%로 나타났다. 10명을 대상으로 체크를 하지 않은 이유를 살펴보니,

감점이 있으므로 공부하지 않고도 중간을 할 수 있을 거라는 기대감이 대부분이었다. 그러나 A 집단에만 중간등급이 나왔으며 나머지 집단에서는 전부 최하위 등급이 나와서 이 부분에서는 지속적인 개선과 추가 연구가 필요한 부분이다.

참고문헌

- [1] Sang, K.A., 2021 “National academic achievement assessment results: High school”, Report 11, ISBN 979-11-6846-129-1 94370, Korea Institute for Curriculum and Evaluation, Chungcheongbuk-do, Republic of Korea, p. 166, (2022).
- [2] Kim, D.J., “Analysis of ground settlement due to circular shaft excavation”, Our secondary education, Vol. 106, pp. 158-163, (1998).
- [3] Lee, T.U. and Choi, H.J., “Informatics education”, Hanbit Publishing Network, p.404, (2016).
- [4] Kim, J.C. and Yang, G.S. “Curriculum and educational evaluation”, Kyoyook book, p.448, (2017).
- [5] Kwon, B. S., “Effect of guessing on the correct answer in a multiple choice”, The Journal of Korean Association of Computer Education, vol.23, no.1, pp.29 – 36, (2020).

(접수: 2024.09.25. 수정: 2024.10.07. 게재확정: 2024.10.10.)