

# 다지선다형 수학평가문항의 채점방식과 문항작성에 관하여

장 경 윤(단국대)

## I. 서 론

수학교육에 있어서 평가는 학습의 진단을 위해, 수업개선을 위한 피드백으로, 성적을 매기기 위해, 일반화된 수학적취도의 측정을 위해, 또는 프로그램평가 등을 목적으로 이루어지며, 평가의 목적에 따라 평가에 사용되는 질문의 종류, 실시방법 등이 달라진다(NCTM, 1989). 어떤 경우이든 수학평가문항은 신뢰롭고 타당성이 있어야 한다. 즉, 유사한 조건 아래서 학생들의 반응이 일관성있게 나타날 수 있어야 하며, 또 문항에서 요구하는 지식이나 능력이 수학교육목표로 제시된 것과 동일한 것이어야 한다.

일반화된 수학적취도를 측정하려고 할 때, 가장 널리 사용되고 있는 것이 다지선다형문항 (multiple-choice item)이다. 이는 “연역적인 추론과 합리적인 이해를 측정하거나, 또는 학생들을 올바르게 판단하여 변별해 내는데 다른 어느 검사문항보다 분명히 효과적이다”(Stanley & Hopkins, 1972, p.236). 또한 다지선다형 검사문항은 채점상의 신속성과 객관성, 그리고 비교적 포괄적인 문항포집이 가능하다는 장점 때문에 현재 대학수학능력검사를 비롯하여 학교수학에서의 주요 평가도구문항으로 활용되고 있다. 그러나 다지선다형 문항은 몇 개의 답지가 제시되어 있어, 그 문제를 해결할 능력이 전혀 없는 경우에도 추측으로 정답을 하게 될 가능성이 높다는 데 그 약점이 있다. Weitzman 등은 다지선다형 검사의 점수는, 다른 어떤 요인들보다도, 얼마나 문항에 제시된 오답을 기술적으로 가려내는가에 따라 크게 좌우된다고 하였다(Stanley & Hopkins, 1972).

본 고에서는 수학학력검사에서 추측요인의 영향을 다룬 본인의 연구를 통하여 다지선다형문항에서 신뢰도와 타당도에 관한 논의를 하고, 아울러 수학교육목표에 타당한 다지선다형문항을 작성하려는 시도들을 소개하려고 한다.

## II. 본 론

다지선다형검사에서 임의적 추측은 眞變량의 증가없이 誤差變량만 증가시키므로 신뢰도를 감소시키며(Mattson, 1966), 재고자하는 특성과 관련이 없는 변량만 증가시켜 타당도를 감소시킨다(Lord, 1963, 1964). 추측점수 矯正에 대한 Mattson과 Lord 등의 이론적 근거에 따라, 교정공식을 사용한 추측점수의 교정이 검사의 신뢰도를 높일 수 있을 것인가를 검증하기 위한 연구들이 이루어져 왔다. 교정효과에 대한 경험연구들은 거의 국외에서 이루어졌으며, 검사의 內容과 실시條件, 답지의 數와 형태, 검사

의 難易度와 길이 등에 따라 상이한 결과들을 보고하고 있다. 이에 대한 경험적 연구 결과들에 관한 상세한 논의는 장경운(1983)을 참조하라. 추측점수의 교정효과연구에 사용된 검사들은 어휘력 또는 기능(skill)검사가 대부분이고, 약간의 산수추리력검사가 사용되었으나 수학학력검사를 사용한 연구는 거의 없는 실정이다.

### 1. 채점방식과 수학학력검사의 신뢰도

추측요인에 관한 전통적인 교정공식은  $S=R-W/(k-1)$ 로 표현되는 減點공식이다. 여기서 S는 교정점수, R은 정답수, W은 오답수, k는 답지수를 나타낸다. 참고로 미국의 대학입학을 위한 SAT는 두 가지 형태의 다지선다형 문항들로 구성되어 있으며 틀린 문항에 대하여는 감점을 하며 채점하는데, 94학년도 부터는 약간의 단답형 문항이 포함되고 있다(장경운, 1993). 이 외에 무응답문항수(0)에 보상을 해주는  $S=R+O/k$  로 표현되는 報償공식, 선택한 답지에 따라 달리 비중을 두어 채점하는 比重채점공식들이 있다. 95학년도 부터 대학수학능력검사에서 문항별로 두가지의 배점을 한다고 하는데, 이는 여기에서 말하는 比重채점방식과는 상이한 것이다.

장경운(1983)은 서울시내 중학교 2학년 남학생 12학급 807명을 대상으로 25문항으로 구성된 다지선다형 數學學力검사를 실시하여, 채점방식이 학력검사의 신뢰도와 타당도에 미치는 연구를 하였다. 검사는 1주일 간격으로 2회 일제히 실시되었다. 전체를 6학급씩 크게 두 집단(A, B)으로 나누고, 집단A에는 보통의 사지선다형의 답지를, 집단B에게는 동일한 검사문항에 "(5)모르겠다"를 답지로 첨가하여 제공하였다. 집단 A, B는 다시 3개의 부분집단들로 나누어 각기 정답채점, 감점공식, 보상공식을 적용하여 채점할 것이라는 지시를 주고, 학생들의 답안은 지시한 대로 채점하였다. 따라서 학생들은 자신의 답이 어떤 방식으로 채점될 것이라는 것을 알고 문항에 응답하였다. 두 차례의 검사결과 각 집단에 속한 학생들의 평균정답수와 무응답 수는 다음과 같았다.

답지\채점방식	정답채점	감점	보상
A	26.8(.54)	21.2(10.26)	17.3(10.82)
B	22.1(3.50)	22.2(8.50)	16.0( 7.75)

표1. 채점방식과 답지형태별 평균정답수와 평균무응답문항수  
(장경운, 1983, p. 47)

표1에서의 숫자는 두 번의 검사를 합한 50문항에 대한 평균정답수이고 괄호 안은 평균 무응답수를 나타낸다. 여기에서 보통의 답지를 사용하였을 때, 감점이나 보상

채점의 지시를 하는 경우에 무응답문항수는 10개 (20%) 이상으로 정답채점지시 아래서의 무응답문항수 .54 (1%)에 비하여 현저히 많은 숫자이며 이는 학생들이 평소에 추측으로 답을 하는 비율이라고 보아도 무방할 것이다. 그 이유는 한 문항에 대하여 전혀 지식이 없거나 약간의 부분지식이 있는 경우에 응답하지 않는다고 볼 수 있기 때문이다. 따라서 수학적능력검사에서 부분지식을 사용하거나 임의로 추측하여 답을 하는 경우가 상당량에 이른다는 사실을 알 수 있다.

각 집단에서 얻어진 신뢰도를 요약하면 표2와 같다.

답지\채점방식	정답채점	감점	보상	평균
A	.7103	.8472	.7969	.7906
B	.8307	.8348	.8248	.8302
평균	.7766	.8409	.8116	.8101

표2. 채점방식과 답지형태별 신뢰도계수(장경운, 1983, p. 35)

전체적으로 신뢰도가 높은 것 부터 감점채점, 보상채점, 정답채점의 순서로 나타났고 보통의 답지형태(A)보다는 '모르겠다'를 첨가한 답지형태(B)의 경우에 신뢰도가 높은 것으로 나타났다.

표2에 나타난 신뢰도의 차이를 통계적으로 검증한 결과, 감점공식과 보상공식을 적용한 경우가 정답채점의 경우보다 각기 1%와 5%의 유의수준에서 높은 신뢰도를 보이는 것으로 나타났으며 (.8409 > .7766\*\*, .8116 > .7766\*), 감점공식과 보상공식적용 사이에 신뢰도의 의미있는 차이는 발견되지 않았다. 즉, 감점공식이나 보상공식을 이용하여 점수를 교정하는 경우가 정답채점의 경우에 비하여 보다 일관성있는 점수를 얻게 하며, 두 가지 교정공식 가운데 통계적으로 유의한 수준은 아니지만 감점지시가 약간 효과적인 것으로 나타났다.

또 "모르겠다"라는 답지를 첨가한 경우(B)가 보통의 사지선다형(A)보다 1%의 유의수준에서 전체적으로 신뢰도가 높은 것으로 보고되고 있다(.8302 > .7906\*\*). 이때 채점방식과 답지형태사이의 상호작용의 효과는 거의 없었다(장경운, 1983, pp37-38). 그리고 답지형태에 따라 신뢰도계수에 큰 차이를 보인 경우는 정답채점의 경우로 "모르겠다"라는 답지가 첨가된 경우에, 1%의 유의수준에서, 신뢰도가 크게 증가하였다. 그러나 감점이나 보상공식을 적용한 경우에는 "모르겠다"는 답지형태의 첨가가 신뢰도에 영향을 주지 않는 것으로 나타나고 있다. 즉, 정답채점의 경우에 한해서만 추측을 억제할 것으로 기대되는 답지의 효과가 있다는 것이다. 이는 학생들이 정답에만 점수를 준다는 것을 알고 있음에도 불구하고, "모르겠다"는 답지의 첨가는 학생들의 추측을

억제하는 효과가 있었음을 보여 준다. 그러나 이러한 효과는 일시적인 것으로 보이는데, 그 이유는 추측하여 임의로 답을 하여도 불이익이 없다는 것을 지속적으로 경험하게 될 때 “모르겠다”는 답지가 학생들에게 계속하여 추측을 억제하도록 영향을 미칠 것이라고 생각하기는 어렵다. 따라서 학교수학에서 정답채점방식을 적용하는 상황에서 “모르겠다”는 답지의 첨가는 별다른 의미를 주지 못할 것으로 보인다. 또한, 장경운(1983)의 연구에서 감점 또는 보상공식의 적용이나 답지형태의 수정으로 타당도가 감소되었다는 증거는 나타나지 않았다.

요약하면 추측을 억제할 것으로 기대되는 답지가 첨가되지 않은 보통의 사지선다형 수학기력검사에서는, 감점 또는 보상공식의 적용이 타당도의 감소시키지 않으면서 검사의 신뢰도를 의미있게 증가시키며, 감점공식의 적용할 때 다소 효과적인 것으로 나타났다.

## 2. 다지선다형 검사문항작성

다지선다형검사 문항을 작성할 때 중요한 것은 정답이 아닌 답지들이 동등한 매력을 갖도록 작성하는 일로서, 다지선다형검사 답지작성시의 일반적인 주의점들이 여러가지로 제시되고 있으나, 어떤 문제에 대한 지식이 전혀 없는 학생들에게 동등한 매력을 갖도록 답지들을 작성하는데는 한계가 있다. 그러므로 다지선다형검사문항에서 어느정도의 추측요소의 작용은 불가피한 것이라 하겠다. 되고 있다(Stanly & Hopkins, 1972, p.p.246-255). 표3는 Stanley등(1972, pp. 246-252)이 제시하는 다지선다형문항작성규칙으로 Millman의 21개 규칙에 몇가지를 덧붙인 것이다.

1. Stem(문제제시부분)은 중심문제와 모든 수식어들을 포함하여야 한다.
2. 각 문항은 가능한 한 간결해야 한다.
3. 문제제시부분에 부정적인 표현은 피하되, 만일 부정적 표현이 사용되는 경우에는 밑줄을 긋거나 이탤릭체로 눈에 띄게 해야 한다.
4. 문제제시부분은 다른 문항의 도움없이 그 문제를 충분히 진술하여야 한다.
5. 정답이외에 적어도 부분적으로 옳은 답지가 있는 경우에는, 최선의 답을 고르라 하든지 “가장”, “우선적으로” 같은 용어를 사용하라.
6. 빠진 부분을 채워 넣는 문제에서 빈칸은 문제 전반부에 나오지 않게 한다.
7. 문항에는 이해하기 쉬운 용어를 사용한다.
8. 각 문항은 다른 것을 검사하라. 즉, 유사한 것을 묻는 문항을 피해야 한다
9. 답지들을 작성하는데 논리적인 계열이 있다면- 예를 들면, 크기순 등- 이를 이용하라.

10. 그 문항이 이해를 바로 하고 있는가를 측정하기 위한 것이라면, 誤답지들도 그럴 듯하게 매력적이어야 한다.
11. 문항의 답지들은 문제제시부분과 문법적으로 일관성이 있어야 한다.
12. 정답지의 길이, 명백성, 또는 답지의 전문적 사항이 다른 답지들과 달라서는 안된다.
13. 답지들은 교과 내용, 형태, 문법적 구조에 있어서 동질적이어야 한다.
14. 특정번호의 답지가 더 자주 정답이 되게 한다든지, 정답에 어떤 패턴이 있지 않도록 하라.
15. 한 문항의 답지수는 적어도 4개는 되게 하라.
16. 문제제시부분과 정답지 사이에서 소리나 단어, 句들이 詩적으로 반복되지 않도록 하라.
17. 교과서의 단어들이나 상투적인 표현을 피하라.
18. 한 문항의 문제제시부분이 다른 문항의 답에 단서를 주지 않도록 하라.
19. 답지들은 중복되거나, 다른 것에 포함되거나, 서로 다른 답지들과 동의어가 되지 않게 하라.
20. '항상', '결코' 등의 한정어를 피하라.
21. 용어나 개념의 이해를 검사하려면, 용어를 먼저 쓰고 일련의 정의나 서술을 쓰는 것이 좋다.
22. 문항의 배열을 교과서와 같은 순서로 하지 않도록 하라.
23. 한 검사의 문항수를 너무 많지 않게 하라. 그럴 경우에는 능력검사가 아닌 속도검사가 될 우려가 있다.
24. '정답없음'을 마지막 답지로 할 경우, 특히 수학에서 유용할 수 있다.
25. '위의 것 모두 정답임'이나 '정답이 두개 이상임'이라는 답지는 'I, II, III' 등으로 보다 구체적으로 기술하는 것이 좋다.

### 표3. 다지선다형검사문항 작성의 규칙들

#### 3. 다지선다형문항작성

신뢰도와 타당도라는 검사의 두가지 요건 가운데 신뢰도 보다 교육적으로 더욱 중요하게 고려되어야 할 것은 타당도이다. 위에 서술한 문항작성시의 규칙들이 추측요인의 효과를 최소화하기 위한 시도라고 한다면, 평가의 내용이 수학교육의 목표에 비추어 적합한가에 초점을 맞추어 검사문항을 개선시키려는 노력들은 타당도에 관심을 둔

것이라고 할 수 있을 것이다. 여기서는 다지선다형문항의 틀을 유지하면서 문항의 내용을 개선하려는 몇가지 시도를 제시하려고 한다(Marshall, 1988).

첫째, 학생들의 오류의 유형을 분석하여 각각을 답지로 제시하는 것이 유용하다. 이 경우에 학생들이 잘못된 지식을 가지고 체계적인 오답을 하는 경우에 선택한 답지에 따라 오류의 종류가 드러나므로 진단의 목적에 매우 적합하게 사용될 수 있다.

둘째, 기존의 다지선다형문항에서 질문의 본질을 수정함으로써 어떻게 학생들의 도식적 지식의 습득과 조직의 초점을 맞출 수 있으며, 이러한 문항들은 진단이나 수업개선을 위한 피드백을 위한 목적으로도 사용될 수 있다(Marshall, 1988).

#### 문제의 배후에 있는 구조를 인식하기

먼저 다음의 예와 같이 표면적으로 유사하나 다른 해법이 요구되는 몇 개의 문항들을 제시하거나, 또는 표면적 특성이 다르지만 동일한 방법으로 해결할 수 있는 문제들을 제시하여, 학생들이 문제의 배후에 있는 구조를 인식하는지의 여부를 알 수 있다. 다음의 두 문제를 보자.

#### 문제1.

계란 1개의 무게는 50g이다. 계란 10개의 무게는 얼마인가?

다음 문제들 가운데 어떤 것이 위의 문제와 같은 방법으로 풀 수 있는가?

- 1) 영희는 90분에 16km를 달렸다. 영희가 1km를 달리는데 몇 분이 걸렸는가?
- 2) 연필은 1자루에 150원, 색연필은 1자루에 200원이다. 이 두가지를 다 사려면 얼마가 필요한가?
- 3) 오렌지쥬스 1잔에 250cc가 들어간다. 1500cc짜리 1병 으로 오렌지쥬스 몇 잔을 만들 수 있겠는가?
- 4) 노트 1권값은 200원이다. 노트 10권의 값으로 얼마를 지불하여야 하는가?

#### 문제2. (예)

다음의 문제들 가운데 나머지 문제들과 다른 것을 고르라.

- 1) 윤희는 사탕 12개를 가지고 있다. 이 사탕들을 3명의 친구에게 똑같이 나누어 주려면 한 친구에게 몇 개씩 줄 수 있겠는가?
- 2) 학교농구선수팀의 선수 5명이 모두 같은 수 만큼 득점하였다. 팀 전체의 득점수가 80점이었다면 1명의 선수는 얼마씩 득점하였는가?
- 3) 병수는 4마리의 강아지를 기른다. 한마리가 매일 300g 의 음식을

먹는다면 강아지들을 1주일 동안 먹으려면 음식을 얼마나 준비해야 하는가?

4) 사과 15개를 4800원에 샀다. 사과 한 개의 값은 얼마인가?

문제1의 답지 가운데 3개는 표면적으로는 문제제시부분에 있는 것과 유사하지만 해결방법은 다른 것들이다. 문제2의 답지 가운데 하나를 제외한 나머지 3개는 문제의 구조가 같고 동일한 해결전략을 요구하는 것들이다.

#### 여러가지 형태에서 동일한 정보를 인식하기

최근 NCTM(1989)은 수학적 의사소통을 유치원에서 고등학교까지 수학교육의 필수적인 목표의 하나로 제시하고 있다. 즉, 학교수학에서 수학적 용어나 기호를 사용하여 자신의 생각을 표현하고, 역으로 주어진 수학적 표현을 이해하고 재진술할 수 있는 능력이 요구된다. 개인이 문장제문제를 이해하고 이를 해결하는 정도는 문제가 어떤 형태로 주어졌는가와 밀접한 관련이 있다(Findell, 1990). 따라서 문제해결력의 부분능력의 하나로서 문제의 재진술을 묻는 문항을 다음과 같이 다지선다형문항으로 제시할 수 있다.

#### 문제3.

순영이는 옛날동전 몇 개를 가지고 있다.

철이는 순영이의 절반 만큼의 동전을 가지고 있다.

다음의 어느 문장이 위의 사실과 같은가?

- 1) 순영이는 철이의 반 만큼 동전을 가지고 있다.
- 2) 철이는 순영이의 두 배의 동전을 가지고 있다.
- 3) 순영이는 철이의 두배의 동전을 가지고 있다.
- 4) 철이는 순영이만큼 동전을 가지고 있다.

#### 4. 평가방법의 새로운 경향

다지선다형검사문항을 사용한 평가방법을 개선하기 위해 교정공식의 사용이라든지 비중채점방식의 도입, 문항작성의 수정등의 노력이 이어져 오고 있는데 반하여, 수학적 지식이나 능력을 평가하는데 “옳거나 그름”이라는 이분법에 의해 이를 점수화하는 전통적인 방법 이외에 사용되는 몇 가지 대안적인 방법들이 사용되고 있다.

그 중의 하나가 지필 검사에 부분점수를 주는 방법(Charles & Lester, 1984; Schoenfeld, 1982)이다. 예를 들어 문제해결 과정을 주관식으로 쓰게 하여 그 학생이 완전히 해결한 경우에는 3점, 문제이해에 그친 경우에는 1점, 문제해결계획을 세운 경우에는 2점을 주는 방법이다. 또 숙련된 교사는 학생이 무엇을 알아야 하고 무엇을 할

수 있는 지에 관해 잘 알고 있으므로 교사가 객관적으로 타당한 평가를 할 수 있다는 전제하에, 모든 것을 교사에게 맡기는 방법도 사용되고 있다(Coladarci, 1986). 이 밖에 학생들의 능력을 발달단계등 특정한 범주의 것으로 분류하는 방법, 진단이나 처치에 관심을 가지고 지필검사 또는 면접을 통해 학생들의 취약점에 초점을 맞추는 방법이다. 또 한가지 방법은 학생의 풀이과정을 점수화하거나 어느 범주에 넣지 않고 단순히 서술하는 것이다. 이는 '소리내어 생각하며' 수행한 지필검사나 면접 등을 통해서 이루어 지는데, 이 때의 관심은 학생들의 수행능력을 평가하기 보다는 학생들의 사고의 진전상황들을 아는데 있다.

## V. 요약 및 결론

우리나라 남자 중학생들을 대상으로 다지선다형 수학학력검사를 실시하였을 때, 감점이나 보상공식을 적용하여 채점한 경우에 무응답문항수가 20% 정도에 이르며, 보통의 정답채점의 경우보다 일관성있는 점수를 얻을 수 있었다. 다지선다형을 포함하여 객관적인 양적(quantitative) 평가가 결과에 관한 평가이고 과정을 평가하기 어렵다는 단점이 있기 때문에, 이를 보완하기 위하여 관찰이나 면담등의 주관적 평가방법의 중요성이 인식되고 있다. 새로 개정된 우리나라의 제6차 교육과정에서도 주객관식 지필검사와 함께 관찰, 면담, 질의응답등의 방법을 사용할 것을 명시하고 있는 것도 이와 같은 관점에서이다. 그러나 이러한 평가방법은 소수의 학생들을 대상으로 할 경우에 가능하며, 이러한 방법들이 학교현장에 적용되려면 교사들에게 기술적 훈련들이 뒤따라야 할 것이다. 다지선다형검사는 실용성과 경제성등의 장점 때문에 학교수학교육에서 계속적으로 활용되어 질 것이다. 따라서 더욱 신뢰로운 문항, 교육목표에 비추어 타당한 검사문항의 개발을 위한 노력이 수반되어야 할 것이다. 더우기 현재 시행되고 있는 다지선다형의 대학수학능력검사의 중요성에 비추어 다지선다형 수학평가문항에 있어서 신뢰도와 타당도에 관한 논의는 재론의 여지가 있는 것으로 보여 진다.

## 참고문헌

- 장경운. (1983). 實施指示形態와 答脂形態가 수학학력검사의 동형검사信賴도와 準據妥當度에 미치는 영향. (교육학석사학위논문, 서울대학교대학원).
- 장경운. (1993). SAT와 미국 수학교육. 數學教育, 32(1), 91-99.
- Charles, R.I. & Lester, F.K. (1984). An evaluation of process oriented instruction program in mathematical problem solving in grades 5 and 7. Journal for Research in Mathematics Education, 15, 15-34.
- Lord, F.M. (1963). Formula and validity. Educational & Psychological



Measurement, 23, 663-672.

- \_\_\_\_\_. (1964). The effect of random guessing on test validity. Educational & Psychological Measurement, 25, 745-747.
- Marshall, S.P.(1988). Assessing problem solving: A short-term remedy and a long-term solution. In R. Charles & E. Silver. (Eds.), The teaching and assessing of mathematical problem solving, vol.3, (pp. 159-177). Reston, VA: NCTM.
- Mattson, D. (1966). The effects of guessing on the standard error of measurement. Educational & Psychological Measurement, 25, 723-730.
- NCTM. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: NCTM.
- Schoenfeld, A.H. (1982). Measures of problem solving performance and problem solving instruction. Journal for Research in Mathematics Education, 13, 31-49.
- Stanley, J.C. & Hopkins, K.D. (1972). Educational and psychological measurement and evaluation. Englewood Cliff, N.J.: Prentice Hall, Inc.