creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

교육학석사 학위논문

수학 서술형 문항에 대한 인지진단평가 적용 연구

2023년 2월

서울대학교 대학원

수학교육과

최 병 홍

수학 서술형 문항에 대한 인지진단평가 적용 연구

지도교수 유 연 주

이 논문을 교육학석사 학위논문으로 제출함

2023년 1월

서울대학교 대학원

수학교육과

최 병 홍

최병홍의 석사 학위논문을 인준함

2023년 1월

위 원 장 이 경 화 (인)

부위원장 윤 상 균 (인)

위 원 유 연 주 (인)

국문초록

수학 서술형 문항에 대한 인지진단평가 적용 연구

인지진단평가는 검사 문항 해결에 요구되는 인지요소를 상정하고, 검사 문항에 대한 피험자의 응답 데이터를 활용하여 피험자의 인지요소별 숙달/미숙달 여부를 추정하는 평가 방법이다. 인지진단평가를 위해 문항과 인지요소 간의 관계에 기반하여 문항을 맞힐 확률에서 고려하는 요인에 따라 다양한 모형이 존재하는데, 이를 인지진단모형이라 한다. 과거에는 선다형 문항이나 이분 반응 문항에 관한 인지진단모형이 주로 연구되었으며, 서술형 문항에 관한 인지진단모형 연구는 비교적 최근부터 이루어지고 있다. Diagnostic Tree Model(DTM)은 풀이 과정에서 2개 이상의 단계와 전략을 가진 서술형 문항에 대한 피험자의 반응을 분석할 수 있는 인지진단모형으로, 피험자의 문제해결 과정에서 나타나는 여러 전략과 전략별 풀이 단계를 모두 고려하는 모형이다. 이 연구에서는 DTM 모형으로 분석할 수 있는 수학 서술형 평가를 설계하고, 실제 학생 대상으로 적용하여 그 결과를 분석한다. 또한 DTM 모형 분석에 영향을 미치는 여러 요인들의 영향을 알아보기 위해 모의실험을 수행한다.

먼저 DTM 기반 인지진단평가의 설계를 위해 수학 서술형 문항의 풀이 과정을 구조화할 수 있는 분석틀을 제안하였다. 서술형 문항의 풀이를 인지요소, 단계, 반응의 세 가지 요소를 이용해 분석할 수 있으며, 이러한 과정을 통해 문제해결에 필요한 인지요소의 개수, 문제 풀이 과정의 단계의 복잡도, 채택한 해결전략에 의해 달라지는 반응 개수를 파악할 수 있다. DTM 기반 인지진단평가를 설계하기 위해서는 검사를 위한 인지요소와 관련된 후보 문항들의 예상 풀이를 이와 같은 틀로 분석해 본 후, 인지요소에 대한 식별성과 검사 시간 등을 고려하여 검사 문항을 선정할 수 있다. 또한 평가 이후의 결과해석에도 이 분석틀을 사용하여 학생의 이해를 진단할 수 있다.

다음으로, 모의실험을 통해 피험자 수, 검사의 문항 수, 인지요소 개수,

문항별 풀이 단계 수, 문항별 반응 종류, 문항 반응 오류확률의 크기 등이 DTM 기반 인지진단평가에 미치는 영향을 살펴보았다. 모의실험 결과, 피험자 수가 많을수록 인지요소별 숙달 여부 추정의 정확도가 대체로 높게 측정되었으나, 피험자 수가 적고 오류확률이 클 때는 오히려 피험자 수가 많을 때보다 적을 때 인지요소별 숙달 여부 추정 정확도가 높게 측정되는 경우가 있었다. 이런 경향은 문항 수, 인지요소 개수, 문항별 풀이 단계 수, 문항별 반응 종류가 많은 검사일수록 더 빈번하게 나타났다. 한편, 인지요소 숙달 패턴 추정의 정확도는 피험자 수와 상관없이 문항 수, 인지요소 개수, 문항별 풀이 단계 수, 문항별 반응 종류가 많은 검사일수록 낮았다. 문항 모수 회복력은 피험자 수가 많거나 문항 오류확률이 작을수록 좋았다.

DTM 모형을 기반으로 절대부등식의 증명에 관한 서술형 문항 인지진단평가를 설계하고, 과학영재고등학교 학생을 대상으로 평가를 실시하였다. 학생의 평가 수행 데이터에 대해 피험자의 반응이 다르게 관찰될 수 있는 풀이 단계별 응답 모형을 여러 가지로 채택하여 DTM 모형 분석을 진행해 본 결과, 문제를 풀기 위해 관련된 모든 인지요소의 숙달을 요구하는 DINA 모형을 채택한 경우가 절대적합도의 기준을 만족하면서도 상대적합도가 가장 좋았다. 학생들의 인지요소 숙달 패턴 추정 결과, 검사에 참여한 학생들은 산술-기하 평균 부등식보다 코시-슈바르츠 부등식을 상대적으로 어려워하였으며, 치환을 통해 부등식을 변형하는 것을 가장 어려워하였다. 문항 분석 결과, 일부 문항에서 학생들이 채택하는 전략에 따라 문제해결의 난이도가 다르게 나타나는 양상을 풀이 경로별 노드의 문항 모수 추정값을 통해 파악할 수 있었다.

마지막으로, DTM 모형의 수형도 구조(tree structure)를 활용하여 피험자의 반응을 다각도로 분석하였으며, 평가를 마친 피험자들에게 인지요소 숙달 패턴 추정 이외의 추가적인 피드백 제공 방법을 제안했다. 문항의 풀이 과정 분석틀에 기반하여 피험자의 풀이 과정을 분석할 수 있으며, 문항의 단계별 인지요소와 단계별 응답 비율, 문항별 피험자의 반응 · 피험자가 답할 가능성이 가장 큰 정답 반응 · 추가 인지요소 습득을 통한 정답 반응의 비교를 제시하여 피험자의 학습경로를 추천할 수 있다.

주요어 : 인지진단평가, 인지진단모형, 진단수형도모형, 서답형 문항, 서술형 문항
학 번 : 2020-21008

목 차

국문초록 ····· i
목차 ····· iv
표 목차 ····· vi
그림 목차 ····· viii
Ⅰ. 서론 ····· 1
1. 연구의 필요성 ····· 1
2. 연구의 목적 및 연구문제 ····· 6
Ⅱ. 이론적 배경 ····· 8
1. 인지진단평가 ····· 8
2. 여러 가지 인지진단모형 ····· 12
2.1. DINA, DINO model ····· 12
2.2. ACDM model ····· 13
2.3. G-DINA model ····· 14
3. 서술형 평가를 위한 인지진단모형 ····· 16
3.1. sG-DINA model ····· 16
3.2. Diagnostic Tree Model ····· 18
Ⅲ. 수학 서술형 문항의 풀이 과정 구조화 ····· 24
1. 연구방법 ····· 24
2. 연구결과 ····· 24
2.1. 수학 서술형 문항의 풀이 과정 분석틀 ····· 26
2.1.1. 인지요소(attribute) ····· 26
2.1.2. 단계(step) ····· 26
2.1.3. 반응(response) ····· 27
2.2. DTM 모형의 인지진단평가 문항 풀이 과정 구조화 28
Ⅳ. 다양한 구조의 서술형 문항을 포함한 DTM 모형 평가에 대한 모의실험 ····· 31

1. 연구방법 ····· 31
1.1. 실험 설계 ····· 31
1.2. 분석 방법 ····· 36
2. 모의실험 결과 ····· 39
Ⅴ. 수학 서술형 평가에 대한 DTM 인지진단의 적용 45
1. 연구방법 ····· 45
1.1. 연구 설계 ····· 45
1.1.1. 연구 수행 절차 ····· 45
1.1.2. 평가를 위한 인지요소 선정 ····· 46
1.1.3. 문항 및 평가지 구성 ····· 47
1.1.4. 피험자 집단 선정 ····· 51
1.1.5. 피험자 모집 및 동의 ····· 52
1.1.6. 실험 수행 과정 ····· 52
1.1.7. 문항에 대한 학생 반응의 분류 ····· 53
1.2. 연구 설계를 위한 모의실험 ····· 54
1.2.1. 실험 설계 ····· 54
1.2.2. 분석 방법 ····· 54
1.3. 학생 데이터 분석 방법 ····· 55
1.3.1. DTM 모형을 통한 분석 ····· 55
1.3.2. 학생 피드백 방법 ····· 56
2. 연구결과 ····· 57
2.1. 모의실험 결과 ····· 57
2.2. DTM 모형을 통한 결과 분석 ····· 59
2.3. DTM 분석 결과를 이용한 피드백 ····· 69
Ⅵ. 결론 ····· 72
1. 논의 및 결론 ····· 72
2. 제한점 및 제언 ····· 76
참고문헌 ····· 78
<부 록> ····· 83
Abstract ····· 100

표 목 차

<표 Ⅱ-1> Q행렬 예시 ····· 10
<표 Ⅱ-2> $3\frac{1}{6}-\frac{5}{6}$ 문항의 두 가지 해결전략 (Ma, 2019, p. 62) ····· 20
<표 Ⅱ-3> $3\frac{1}{6}-\frac{5}{6}$ 문항에 관한 Q행렬 ····· 21
<표 Ⅱ-4> DTM 모형 인지진단평가에서 사용되는 Q행렬 ····· 21
<표 Ⅲ-1> 방정식 $7(3x+1)-5(6x+2)=9$를 푸는 두 가지 전략 ····· 27
<표 Ⅲ-2> $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조, Q행렬, 해결전략 ····· 28
<표 Ⅲ-3> $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조, Q행렬, 분석틀 분석 ····· 30
<표 Ⅳ-1> 문항 유형별 수형도 구조, Q행렬, 분석틀 분석 ····· 31
<표 Ⅳ-2> 모의실험 평가 문항 세트 구성 ····· 36
<표 Ⅳ-3> 평가 A, B, C의 PCA, PCV ····· 39
<표 Ⅳ-4> 평가 A의 $RMSE_j$ ····· 40
<표 Ⅳ-5> 평가 B의 $RMSE_j$ ····· 41
<표 Ⅳ-6> 평가 C의 $RMSE_j$ ····· 42
<표 Ⅴ-1> 인지진단평가에 사용된 인지요소 ····· 47
<표 Ⅴ-2> 실험에 사용된 문항의 수형도 구조, Q행렬, 분석틀 분석 ·· 49
<표 Ⅴ-3> 연구 참여자 수와 학년별 분포 ····· 53
<표 Ⅴ-4> 학생 인지진단평가 문항 세트에 대한 PCA, PCV ····· 57
<표 Ⅴ-5> 문항 질이 낮을 때 $RMSE_j$ ····· 58
<표 Ⅴ-6> 문항 질이 중간일 때 $RMSE_j$ ····· 58
<표 Ⅴ-7> 문항 질이 높을 때 $RMSE_j$ ····· 59
<표 Ⅴ-8> 문항별 반응분포 ····· 60
<표 Ⅴ-9> DINA, DINO, ACDM, G-DINA 모형에 대한 상대 적합도 63
<표 Ⅴ-10> DINA, DINO, ACDM, G-DINA 모형에 대한 부트스트랩 모형 적합도 ····· 63
<표 Ⅴ-11> 인지요소별 숙달 피험자 비율 ····· 64

<표 Ⅴ-12> 인지요소 숙달 패턴의 분포 ····· 65
<표 Ⅴ-13> 각 노드의 guess와 slip ····· 65
<표 Ⅴ-14> 8번 문항의 수형도 구조와 문항 모수 ····· 66
<표 Ⅴ-15> 10번 문항의 수형도 구조와 문항 모수 ····· 67
<표 Ⅴ-16> 12번 문항의 수형도 구조와 문항 모수 ····· 68
<표 Ⅴ-17> 13번 문항의 수형도 구조와 문항 모수 ····· 68
<표 Ⅴ-18> 피험자 A의 인지요소 숙달 패턴 추정 결과 ····· 69
<표 Ⅴ-19> 분석틀을 이용한 피험자 A의 풀이 과정 분석 ····· 69
<표 Ⅴ-20> 피험자 B의 인지요소 숙달 패턴 추정 결과 ····· 71

그 림 목 차

[그림 Ⅱ-1] 인지진단평가의 원리 ····· 9
[그림 Ⅱ-2] 수형도 구조 예시 (Ma, 2019, p. 65) ····· 19
[그림 Ⅱ-3] G-DINA 모형, sG-DINA 모형, DTM 모형의 비교 ····· 23
[그림 Ⅲ-1] $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조, Q행렬 ····· 25
[그림 Ⅲ-2] 수형도 구조 예시 ····· 29
[그림 Ⅴ-1] 연구 수행 절차 ····· 45
[그림 Ⅴ-2] 인지진단평가에 사용된 문항 세트 ····· 47
[그림 Ⅴ-3] 문항의 단계별 인지요소 및 응답 비율 ····· 61
[그림 Ⅴ-4] 12번 문항의 단계별 인지요소 및 응답 비율 ····· 70
[그림 Ⅴ-5] 8번 문항의 수형도 구조를 활용한 피드백 ····· 71

Ⅰ. 서론

1. 연구의 필요성

인지진단평가(Cognitive Diagnostic Assessment, CDA)는 검사 문항을 해결하는 데 요구되는 인지요소를 상정하고, 검사 문항에 대한 피험자의 응답 데이터를 활용하여 피험자의 인지요소별 숙달/미숙달 여부를 통계적 방법으로 추정하는 평가 방법을 의미한다(Jang, 2008). 인지진단평가를 통해 추정된 학생의 인지요소 숙달 여부는 학생에게 학습 내용의 성취 정도를 파악할 수 있게 해주며, 학생을 지도하는 교사에게는 학생의 부족한 부분을 면밀하게 파악하여 적절한 피드백을 제공할 수 있게 해준다(Lee et al., 2011; Leighton & Gierl, 2007).

인지진단평가에서 평가 문항에 대한 피험자 집단의 반응으로부터 각 피험자의 인지요소 숙달/미숙달 여부 추정을 위해 문항과 인지요소 간의 관계 설정이나 문항을 맞힐 확률 계산 등에서 다양한 방식의 가정과 통계적 기법이 사용된다. 이와 같이 인지요소 숙달 여부 추정의 방법에 따라 다양한 모형이 존재하는데, 이를 인지진단모형(Cognitive Diagnostic Models, CDM)이라 한다. 최근에 인지진단모형은 많은 관심을 받아왔으며, 다양한 모형이 개발되었다. 대표적으로 DINA(Deterministic input, noisy “and” gate model; Haertel, 1989), DINO(Deterministic input, noisy “or” gate model; Templin & Henson, 2006), ACDM(Additive CDM; Torre, 2011) 등이 있다. DINA 모형은 피험자가 문항과 관련된 모든 인지요소를 숙달해야 문항을 맞힐 수 있다고 가정한 모형이며, DINO 모형은 피험자가 문항과 관련된 여러 인지요소 중 하나만 숙달해도 문항을 맞힐 수 있다고 가정한 모형이다. ACDM 모형은 문항을 해결하는 데 요구되는 각 인지요소 숙달 여부에 따라 문항을 해결할 확률이 달라진다고 가정한 모형이다. 이외에도 G-DINA(Generalized DINA; Torre, 2011), LCDM(Log-linear CDM; Henson, Templin, & Willse, 2009) 등 다양한 모형이 있다.

그런데 기존의 많은 인지진단모형 연구는 선다형 문항(multiple choice

item)이나 이분 반응 문항(dichotomous response item)과 같은 제한적인 가정하에 이루어졌다(Kuo, Chen, Yang & Mok, 2016). 이는 교육적 측면에서 서술형 문항(constructed response item)이 가지는 장점이 간과된 것이다. 박정과 박경미(2004)는 서답형 문항이 선다형 문항보다 학생들의 능력을 추정하는 데 더 많은 정보를 제공한다고 했다. 서답형 문항은 단답형, 서술형, 논술형 등으로 구분되는데 이중 서술형은 답안에 기술해야 할 요소가 한 문장 이상으로 설정되어 있어서 분석력, 문제해결력, 창의력 등의 고등사고 능력을 측정하기 적합하다.[1)]

서술형 문항은 답안에 학생의 사고 과정이 드러나기 때문에 문제해결 과정을 몇 개의 단계로 나누고 단계마다 채점 요소를 세워서 채점할 수 있다. 그러므로 서술형 문항은 다분 반응(polytomous response)을 가정한 모형으로 연구할 필요가 있다. 지금까지 개발된 다분 반응을 가정한 인지진단모형은 PC-DINA(Partial credit DINA; Torre, 2010), pGDM(GDM for graded responses; von Davier, 2008), NRDM(Nominal Response Diagnosis Model; Templin, Henson, Rupp, Jang, & Ahmed, 2008), polytomous LCDM(Hansen, 2013), sG-DINA(Sequential G-DINA; Ma & Torre, 2016) 등이 있다.

특히 sG-DINA 모형은 문제해결 과정을 일련의 단계로 나누고, 다음 단계를 수행하기 위해서는 이전의 모든 단계를 성공적으로 마쳐야 한다고 가정하는 방식으로 피험자의 반응을 분석한다. 보통 수학 문제에 대한 답과 그 이유를 서술형으로 답하는 것은 여러 명제를 나열하는 것이며, 각 명제는 그에 앞서 서술된 명제 혹은 사전에 알려진 수학적 원리나 정리에 의해 참임이 보장된다. 만약 앞서 서술된 명제가 뒤따르는 명제를 함의하지 않거나 중간에 거짓인 명제가 사용된다면, 그 이후로 얻어지는 추론은 수학적으로 옳다고 할 수 없다. 따라서 문제해결 과정을 일련의 단계로 나누어 피험자의 반응을 분석하는 sG-DINA 모형은 수학 서술형 문항에 대

1) 박정과 박경미(2004)는 수행형 문항과 선다형 문항의 수학적 능력 추정 효율성을 비교하였다. 여기서 수행형 문항은 광의로 해석되어 단답형 문항과 서술형 문항을 모두 포괄한다고 했다. 본 고에서는 선행 연구에서 사용된 수행형 문항이란 표현을 서답형 문항으로 표현하였다.

한 반응을 분석하기에 어느 정도 적합한 모형이라고 할 수 있다.

이처럼 DINA, DINO 모형을 시작으로 서술형 문항에 적용하기 적합한 sG-DINA 모형에 이르기까지 많은 종류의 모형이 개발되었다. 하지만 앞에서 제시한 모형들은 문항을 해결하는 과정에서 모든 피험자가 동일 전략을 사용할 것이라고 가정했다는 점에서 실제 서술형 평가의 응답에 적용하는 데 제한적이다. 학생들은 문제해결에서 다양한 해결전략을 사용하며, 서로 다른 해결전략은 다른 인지요소를 필요로 할 수 있다(Mislevy, 1996). MS-DINA(Multiple-Strategy Deterministic input, noisy "and" gate; Torre & Douglas, 2008) 모형은 다양한 해결전략을 가정하기는 하지만 서로 다른 해결전략이 동일 추측 모수(guess parameter)와 실수 모수(slip parameter)를 가지도록 설정됐다는 점에서 제한이 있다.

DTM(Diagnostic Tree Model; Ma, 2019)은 2개 이상의 단계와 전략을 가진 서술형 문항에 대한 피험자의 반응을 분석할 수 있는 인지진단모형이다. 피험자의 문제해결 과정에서 나타나는 여러 전략과 전략별 풀이 단계를 모두 고려한다. 평가 문항에 대한 피험자의 반응에서 나타날 수 있는 여러 전략과 전략별 풀이 단계는 수형도 구조(tree structure)로 추상화할 수 있다. 각 문제해결전략에 대해서는 sG-DINA 모형의 구조와 유사하기 때문에 DTM 모형은 sG-DINA 모형을 일반화한 것이라고 볼 수 있으며, 현재로서는 서술형 문항에 적용하기에 적합한 최신 모형이다.

Ma(2019)는 2007년 수학·과학 성취도 평가 추이 국제 비교 연구(Trends in International Mathematics and Science Study 2007, TIMSS 2007)에서 사용된 문항과 응답 데이터를 사용하여 DTM 모형, sG-DINA 모형, MS-DINA 모형의 모형 적합도를 각각 구하고 비교했다(Ma, 2019). 새로운 인지진단모형을 만들어서 그 효과를 검증할 때마다 새로운 검사문항을 만들고 검사대상을 선정하고 평가하고 채점하는 일련의 과정을 겪는 것은 큰 비용과 노력이 요구된다. 그래서 실제로 인지진단모형에 관한 많은 연구에서 TIMSS 데이터가 사용된다(박지현과 김수진, 2015; 이현숙과 고호경, 2014; Choi et al., 2015; Dogan & Tatsuoka, 2008; Lee et al., 2011). 이미 공개되어있는 TIMSS의 문항은 내용 전문가들의 충분한 검토

를 거쳐 설계되었으며, 피험자의 응답 데이터는 응답 표본이 크고 다양한 국가의 학생 정보를 지녔다. 하지만 TIMSS 데이터를 사용하여 DTM 모형의 효과를 확인하는 것에는 제한점이 존재한다. TIMSS의 수학·과학 성취도 평가는 수학·과학 교수학습을 개선하기 위한 정보 제공을 목적으로 60여 개국의 4학년과 8학년의 보통 학생들을 대상으로 시행된다(National Center for Education Statistics(NCES), 2021). 따라서 평가에 사용되는 문항은 대체로 목표에 이르는 길을 파악하기 쉽고, 해결을 위해 거쳐야 하는 단계가 길지 않으며, 다양한 전략이 사용되기 어렵다.

Polya(1962)는 문제해결이란 문제해결을 위해 극복해야 하는 장애물에서 벗어나거나 우회할 방법을 찾는 것이며, 즉시 달성할 수 없는 목표를 달성하는 것이라고 했다. 황혜정 외(2012)는 좋은 문제란 목표는 분명하지만, 목표에 이르는 길이 즉각적으로 주어져 있지 않으며, 여러 단계를 거쳐야 해결할 수 있고, 해결 과정에서 다양한 종류의 개념, 원리, 기능이 사용되며, 다양한 해법의 반응이 존재하는 것이라고 했다. DTM 모형은 여러 단계를 거쳐야 해결할 수 있고, 해결 과정에서 다양한 종류의 개념·원리·기능이 사용되며, 여러 해법이 존재하고, 다양한 반응이 도출된다고 가정할 수 있는 서술형 문항의 풀이 과정을 분석하는 데 사용할 수 있도록 정교하게 설정된 모형이다. 하지만 Ma(2019)의 연구에서 사용된 모의실험 문항과 TIMSS 2007의 문항은 해결 단계도 짧고, 적은 수의 개념·원리·기능이 사용되며, 해법이나 반응도 다양하지 않다. 실제로 Ma(2019)는 DTM 모형에 관한 모의실험에서 5개의 인지요소 추정 정확도를 구하기 위해 두 개의 전략을 지닌 동일 문항 6개, 이분반응 문항 5개로 이루어진 단순한 구조의 문항 세트를 활용했다. 그리고 TIMSS 2007 평가 문항으로 구성한 문항 세트를 통해 숙달 여부를 알아내고자 한 인지요소는 3가지로 매우 적었다.

따라서 DTM 모형을 적용한 인지진단평가를 통해 모형의 적합도를 확인하거나 인지요소 숙달/미숙달 여부 추정 정확도 등을 분석할 때는, 다양한 해법의 반응이 존재하고 여러 단계를 거쳐야 해결할 수 있고 해결 과정에서 다양한 종류의 개념·원리·기능이 사용되며 TIMSS 2007의 8학년

대상 문항보다는 목표에 이르는 길이 즉각적으로 주어지지 않은 수학 서술형 문항을 사용할 필요가 있다. 그리고 위에서 제시한 특징의 수학 서술형 문항을 사용하여 DTM 모형을 통한 인지진단평가를 시행하고자 한다면, 위에서 제시한 특징의 문항과 그렇지 않은 문항을 구분하는 일이 선행되어야 한다. 따라서 주어진 서술형 문항에 대하여 풀이 전략이 다양한지, 해결의 단계가 많은지, 해결 과정에서 다양한 개념과 원리가 사용되는지와 관련하여 피험자의 풀이 과정을 분석하는 분석틀이 필요하다. 분석틀을 이용해 예상 풀이 반응을 분석하면, 더 많은 단계를 요구하고 다양한 전략이 사용되는 문항들로 이루어진 서술형 평가를 인지진단평가로 구성할 수 있으며, 이를 활용하여 Ma(2019)의 모의실험보다 더 구조화된 모의실험을 진행할 수 있다.

모의실험과 더불어 피험자 집단을 모집하여 서술형 인지진단평가를 시행하여 DTM 모형에 관한 시사점과 교육적 시사점을 도출할 필요가 있다. 실제로 진행되는 인지진단평가는 인지요소, 문항, 문항 난이도, 시험환경, 피험자 집단 특성, 피험자 수 등에 따라 그 결과와 해석이 달라질 수 있는데, 모의실험에서는 이와 같은 현실적인 맥락을 반영하기 어렵다. 따라서 우리나라 중학교 혹은 고등학교 학생들을 연구 참여자로 모집하여 인지진단평가를 시행하고 그 결과를 분석해볼 필요가 있다. 추가로 문항 풀이 과정 분석틀, 인지진단평가의 Q행렬, DTM 모형의 수형도 구조(tree structure)를 통해 평가 문항의 특징과 구조를 다각도로 분석할 수 있다. 이를 통해 평가 설계자가 분석틀, Q행렬, 수형도 구조 등을 활용하여 설계 과정을 종합하고 평가 결과를 다각도로 해석하며 피험자에게 양질의 피드백을 제공하는 방법을 제시할 수 있다.

2. 연구의 목적 및 연구문제

본 연구의 목적은 수학 서술형 문항의 풀이 과정을 구조화하는 분석틀을 제시하고, 제시된 분석틀에 기반을 두어 수학 서술형 문항 세트에 대한 DTM 모형 모의실험과 학생 대상 인지진단평가를 실시하여 인지진단평가와 교육에 관한 시사점을 얻는 것이다. 구체적인 연구문제는 다음과 같다.

1. DTM 모형을 적용한 인지진단평가를 설계하기 위해 수학 서술형 문항의 풀이 과정을 어떻게 구조화할 것인가?

2. 다양한 구조의 수학 서술형 문항을 포함한 평가에 대해 DTM 모형으로 분석하는 모의실험을 진행했을 때, 인지요소 숙달/미숙달 여부 추정 정확도와 문항 모수 회복력은 어떠한가?

3. 고등학교 학생을 대상으로 수학 서술형 문항으로 구성된 인지진단평가를 실시하고 DTM 모형으로 분석하였을 때, 모형 적합도와 인지요소 숙달 양상은 어떠한가? 또한 학생들에게 평가 결과에 대한 피드백을 어떤 방법으로 제공할 수 있는가?

Ⅱ장에서는 연구문제의 해결을 위한 초석을 다지는 것을 목표로 인지진단평가에 관한 전반적인 설명과 인지진단모형 DINA, DINO, ACDM, sG-DINA, DTM의 개념 및 원리에 대해 검토한다. Ⅲ장에서는 DTM 모형을 적용한 인지진단평가를 설계하기 위한 수학 서술형 문항의 풀이 과정 분석틀을 제안하여 연구문제 1에 답하고자 한다. Ⅳ장에서는 분석틀에 기반을 두어 모의 문항 세트를 만들고, 모의 피험자 데이터를 생성하여 DTM 모형을 통한 인지요소 숙달/미숙달 여부 추정 정확도와 문항 모수 회복력을 확인하는 모의실험을 진행하여 연구문제 2에 답하고자 한다. Ⅴ장에서는 본 고의 저자가 제작한 수학 서술형 문항 세트를 활용하여 고등학교 학생들을 대상으로 시험을 시행하고, 그 반응 데이터를 DTM 모형으

로 분석하는 인지진단평가를 진행하여 연구문제 3에 답하고자 한다. 또한 인지진단평가를 통해서 학생의 인지요소 숙달 여부만을 추정하는 것이 아니라, DTM 모형의 수형도 구조와 관련지어 학생의 반응을 다각도로 분석해보고 그 결과를 활용하여 효과적인 피드백 방법을 제안해보고자 한다. 마지막으로 Ⅵ장에서는 이상의 논의를 요약하며 본 연구를 통해 얻은 시사점과 본 연구의 제한점을 밝히고, 후속 연구를 위한 제언을 하고자 한다.

Ⅱ. 이론적 배경

1. 인지진단평가

인지진단평가(Cognitive Diagnostic Assessment, CDA)는 검사 문항을 해결하는 데 요구되는 인지요소를 상정하고, 검사 문항에 대한 피험자의 응답 데이터를 활용하여 피험자의 인지요소별 숙달/미숙달 여부를 통계적 방법으로 추정하는 평가 방법을 의미한다. 인지진단평가를 통해 추정된 학생의 인지요소 숙달 여부는 학생에게 학습 내용의 성취 정도를 파악할 수 있게 해주며, 학생을 지도하는 교사에게는 학생의 부족한 부분을 면밀하게 파악하여 적절한 피드백을 제공할 수 있게 해준다. 인지진단평가에서 평가 문항에 대한 피험자 집단의 반응으로부터 각 피험자의 인지요소 숙달/미숙달 여부 추정을 위해 문항과 인지요소 간의 관계 설정이나 문항을 맞힐 확률 계산 등에서 다양한 방식의 가정과 통계적 기법이 사용된다. 이와 같이 인지요소 숙달 패턴 추정의 방법에 따라 다양한 모형이 존재하는데, 이를 인지진단모형(Cognitive Diagnostic Models, CDM)이라 한다.

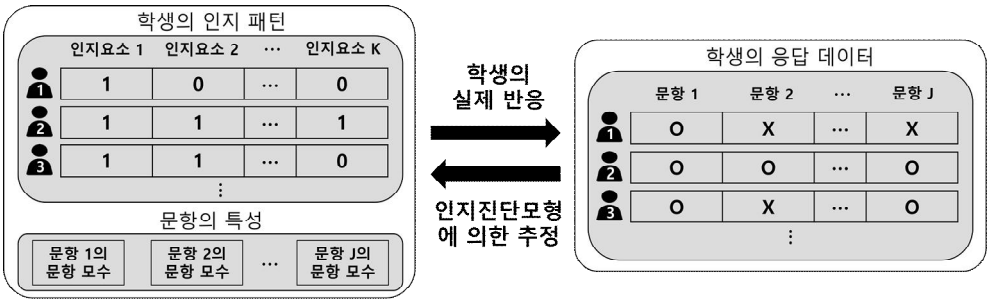
인지진단모형에서 피험자의 상태는 여러 인지요소의 숙달/미숙달 여부로 구성되어 있다고 가정한다. 인지진단평가에서 확인하고자 하는 피험자의 인지요소가 K개라면 피험자 i의 인지요소 숙달 패턴은 다음과 같이 벡터로 표현된다.

$$\boldsymbol{\alpha}_i = [\alpha_{i1},\ \alpha_{i2},\ \ldots\ldots, \alpha_{iK}]$$

$\alpha_{ik}(1 \le k \le K)$는 피험자 i의 k번째 인지요소 숙달여부를 의미하며, 숙달한 경우는 1, 숙달하지 않은 경우는 0으로 표현한다. 이 경우 피험자 i의 가능한 인지요소 숙달 패턴은 2^K개다. 예를 들어, 5개의 인지요소 숙달 여부를 파악하는 인지진단평가에서 피험자 i의 인지요소 숙달 패턴이 $\boldsymbol{\alpha}_i = [1,\ 1,\ 0,\ 0,\ 1]$이라면 첫 번째, 두 번째, 다섯 번째 인지요소는 숙달했으

나 세 번째와 네 번째 인지요소는 숙달하지 않은 것이다. 이때 피험자 i의 가능한 인지요소 숙달 패턴은 2^5개다.

[그림 Ⅱ-1]은 인지진단평가의 원리를 도식화한 것이다. 피험자들이 지닌 K개의 인지요소 숙달 여부는 잠재적인(latent) 상태이므로 어느 누구도 피험자 A의 인지요소 숙달 여부의 패턴을 관찰을 통해 직접적으로 알 수 없다. 이때 피험자들이 인지요소 숙달 패턴에 따라 K개의 인지요소와 관련된 J개의 문항에 응답한 결과가 주어지면 인지진단모형을 가정하여 피험자의 인지요소 숙달 패턴과 문항의 특성에 관한 문항 모수(parameter)가 추정된다.



[그림 Ⅱ-1] 인지진단평가의 원리

인지진단모형을 이용해 피험자들의 인지요소 숙달 패턴을 추정하기 위해서는 문항과 인지요소 사이의 관계가 설정되어야 하는데, 이때 Q행렬(Q-matrix)이 사용된다. J개의 문항과 K개의 인지요소로 이루어진 평가에 관한 Q행렬은 행렬 $(q_{jk})_{J \times K}$로 표현된다. 이때 각 성분q_{jk}는 문항 j를 해결하는 데 k번째 인지요소가 관련되어 있으면 1, 그렇지 않으면 0으로 설정한다. 예를 들어, <표 Ⅱ-1>은 4개의 문항과 5개의 인지요소로 이루어진 평가에 관한 Q행렬을 나타낸 것이다. 이 경우 첫 번째 문항은 인지요소 1만 관련이 있는 문항이고, 네 번째 문항은 인지요소 2, 4, 5와 관련이 있는 문항이다.

문항	인지요소 1	인지요소 2	인지요소 3	인지요소 4	인지요소 5
1	1	0	0	0	0
2	0	1	1	0	0
3	1	0	0	1	0
4	0	1	0	1	1

<표 Ⅱ-1> Q행렬 예시

인지진단모형에서 학생의 인지요소 숙달 패턴을 추정하는 방법에는 대표적으로 Maximum Likelihood Estimation(MLE), Maximum A Posteriori(MAP), Expacted A Posteriori(EAP)가 있다(Huebner & Wang, 2011).

인지진단평가에서 J개의 문항에 대한 피험자 i의 응답 벡터를 $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \cdots, X_{iJ})$라 하자. 그러면 인지요소 숙달 패턴 $\boldsymbol{\alpha}$를 지닌 피험자의 반응에 대한 우도(likelihood of response)는 다음과 같다.

$$L(\boldsymbol{X}_i|\boldsymbol{\alpha}) = \prod_{j=1}^{J} P(X_{ij}|\boldsymbol{\alpha})$$

여기서 $P(X_{ij}|\boldsymbol{\alpha})$는 인지진단모형에 따라 다양한 형태를 가진다. 실제 피험자의 인지요소 숙달 패턴 $\boldsymbol{\alpha}$는 정확히 알 수 없으나 추정된 인지요소 숙달 패턴 $\hat{\boldsymbol{\alpha}}$은 K개의 인지요소와 $l=1, \cdots, L(L=2^K)$에 대하여 $\boldsymbol{\alpha}_l$로 주어진다. MLE 방법에서는 반응에 대한 우도를 최대로 하는 인지요소 숙달 패턴 $\hat{\boldsymbol{\alpha}}_{MLE} = \arg\max_l \{L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)\}$이 채택한다. 한편, MAP 방법에서는 베이지안 이론에 의해 사후확률(posterior probability) $P(\boldsymbol{\alpha}_l|\boldsymbol{X}_i)$이 사용된다.

$$P(\boldsymbol{\alpha}_l|\boldsymbol{X}_i) = \frac{L(\boldsymbol{X}_i|\boldsymbol{\alpha}_l)P(\boldsymbol{\alpha}_l)}{\sum_{m=1}^{L} L(\boldsymbol{X}_i|\boldsymbol{\alpha}_m)P(\boldsymbol{\alpha}_m)}$$

이때 인지요소 숙달 패턴은 $\hat{\boldsymbol{\alpha}}_{MAP} = \arg\max_l \{L(\boldsymbol{\alpha}_l | \boldsymbol{X}_i)\}$이 채택된다.

EAP 방법은 문항반응이론(Item response theory)에서 피험자의 능력 θ (examinee's θ level)에 대한 확률밀도함수(population density function) $P(\theta)$와 능력 θ를 지닌 피험자 i의 반응에 대한 우도 $L(\boldsymbol{X}_i | \theta)$가 사용되며 θ의 추정은 다음과 같다.

$$\mathrm{EAP}(\theta) = \frac{\int_{-\infty}^{\infty} L(\boldsymbol{X}_i | \theta) P(\theta) \theta d\theta}{\int_{-\infty}^{\infty} L(\boldsymbol{X}_i | \theta) P(\theta) d\theta} \text{ 또는 } \mathrm{EAP}(\theta) \approx \frac{\sum_{all\ h} L(\boldsymbol{X}_i | \theta_h) P(\theta_h) \theta_h}{\sum_{all\ h} L(\boldsymbol{X}_i | \theta_h) P(\theta_h)}$$

피험자 i와 인지요소 $k = 1, 2, \cdots, K$에 대한 주변확률(marginal probability) $\widetilde{\alpha_{ik}}$는 k번째 인지요소를 습득한 인지요소 숙달 패턴의 사후확률의 합으로 정의된다.

$$\widetilde{\alpha_{ik}} = \sum_{l=1}^{L} P(\boldsymbol{\alpha}_l | \boldsymbol{X}_i) I(\alpha_{l,k} = 1),$$

$$I(\alpha_{l,k} = 1) = \begin{cases} 1, & \boldsymbol{\alpha}_l\text{의 } k\text{번째 인지요소가 1인 경우} \\ 0, & \text{그 외} \end{cases}$$

이때 $\widetilde{\alpha_{ik}}$가 0.5보다 크면 피험자 i는 k번째 인지요소를 습득한 것으로 추정하고 그렇지 않으면 습득하지 않은 것으로 추정한다(Huebner & Wang, 2011).

2. 여러 가지 인지진단모형

2.1. DINA, DINO model

DINA 모형(Deterministic input, noisy “and” gate model)과 DINO 모형(Deterministic input, noisy “or” gate model)은 가장 단순한(parsimonious) 인지진단모형에 해당한다(de la Torre, 2011; Haertel, 1989; Junker & Sijtsma, 2001; Templin & Henson, 2006). DINA 모형은 피험자가 문항과 관련된 인지요소를 모두 습득하고 있어야 해당 문항을 해결할 수 있다고 가정한 것이고, DINO 모형은 피험자가 문항과 관련된 인지요소 중 하나만 습득하고 있어도 문항을 해결할 수 있다고 가정한 것이다.

DINA 모형과 DINO 모형은 문항별로 추측 모수인 guess와 실수 모수 slip을 가진다. guess는 피험자가 주어진 문항을 맞히기 위해 필요한 인지요소 숙달 패턴을 갖추지 못하였는데 문항을 맞힐 확률을 의미하고, slip은 피험자가 주어진 문항을 맞히기 위해 필요한 인지요소 숙달 패턴을 갖추었는데 문항을 틀릴 확률을 의미한다. DINA 모형과 DINO 모형에서 K개의 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l = [\alpha_{l1}, \alpha_{l2}, \ldots\ldots, \alpha_{lK}]$을 가지는 피험자가 문항 j를 맞힐 확률을 나타내는 문항 반응 함수는 다음과 같다.

$$\mathrm{P}(X_{lj} = 1 | \eta_{lj}, s_j, g_j) = (1 - s_j)^{\eta_{lj}} g_j^{1-\eta_{lj}}$$

X_{lj}는 인지요소 숙달 패턴이 $\boldsymbol{\alpha}_l$인 피험자의 문항 j에 대한 반응을 정답은 1, 오답은 0으로 나타낸 것이다. s_j는 문항 j의 slip이고, g_j는 문항 j의 guess다. η_{lj}는 인지요소 숙달 패턴이 $\boldsymbol{\alpha}_l$이 피험자가 문항을 맞히기 위한 특정 조건을 갖추었으면 1, 갖추지 못했으면 0의 값을 가진다. DINA 모형과 DINO 모형에서 η_{lj}는 각각 다음과 같이 정의된다.

$$\eta_{lj} = \prod_{k=1}^{K_j^*} \alpha_{lk}^{\ q_{jk}} \quad \text{(DINA)}$$

$$\eta_{lj} = 1 - \prod_{k=1}^{K_j^*} (1-\alpha_{lk})^{q_{jk}} \quad \text{(DINO)}$$

K_j^*는 문항 j와 관련된 인지요소의 개수를 의미한다. 즉, $K_j^* = \sum_{k=1}^{K} q_{jk}$이다. 일반성을 잃지 않고, 문항 세트의 K개의 인지요소 중 처음 K_j^*개의 인지요소를 문항 j와 관련된 인지요소라 가정할 수 있다. α_{lk}는 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$의 k번째 인지요소 숙달 여부를 의미한다. 여기서 l은 1, ..., $2^{K_j^*}$이다. 위의 η_{lj}에 관한 등식에 의해 DINA 모형에서는 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$이 문항 j와 관련된 인지요소를 모두 습득하고 있어야 $\eta_{lj}=1$이며, 그 외의 경우는 $\eta_{lj}=0$이다. DINO 모형에서는 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$이 문항 j와 관련된 어느 인지요소도 습득하지 않아야 $\eta_{lj}=0$이며, 그 외의 경우는 $\eta_{lj}=1$이다.

2.2. ACDM model

ACDM(Additive CDM; Torre, 2011) 모형은 문항을 해결하는 데 요구되는 각 인지요소의 숙달 여부에 따라 문항을 해결할 확률이 달라진다고 가정한 모형이다. ACDM 모형에서 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$을 가지는 피험자가 문항 j를 맞힐 확률을 나타내는 문항 반응 함수는 다음과 같다.

$$\mathrm{P}(X_{lj}=1|\boldsymbol{\alpha}_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk}$$

$\boldsymbol{\alpha}_{lj}^*$는 피험자의 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$을 문항 j와 관련된 인지요소로만 나

타낸 부분 인지요소 숙달 패턴이다. δ_{j0}은 모든 α_{lk}의 값이 0일 때 피험자가 정답 반응을 보일 확률을 의미한다. 이를 문항 j의 인터셉트(the intercept for item j)라 한다. δ_{jk}는 문항 j와 관련된 k번째 인지요소에 의한 효과를 의미하며, α_k에 의한 주효과(the main effect due to α_k)라 한다. ACDM 모형에서는 문항과 관련된 인지요소를 습득함에 따라 δ_{jk}에 의해 문항을 맞힐 확률이 가산적으로 변한다. 그리고 서로 다른 인지요소의 숙달 여부는 정답 반응 확률에 서로 독립적으로 영향을 준다. ACDM 에서는 문항 j에 대한 모수는 $(\delta_{j0}, \ldots, \delta_{jK_j^*}$ 로 K_j^*+1개다.

2.3. G-DINA model

G-DINA(Generalized DINA; Torre, 2011) 모형은 DINA, DINO, ACDM 모형의 설정을 확장한 것이다. ACDM 모형에서는 문항과 관련된 각 인지요소에 의한 주효과만 고려했는데, G-DINA 모형에서는 그것과 더불어 인지요소 간의 상호작용효과까지 고려한다. 따라서 ACDM 모형은 G-DINA 모형의 특수한 경우라 할 수 있다. G-DINA 모형에서 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$을 가지는 피험자가 문항 j를 맞힐 확률을 나타내는 문항 반응 함수는 다음과 같다.

$$\mathrm{P}(X_{lj}=1|\boldsymbol{\alpha}_{lj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*}\delta_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K_j^*}\sum_{k=1}^{K_j^*-1}\delta_{jkk'}\alpha_{lk}\alpha_{lk'} + \cdots + \delta_{j12\cdots K_j^*}\prod_{k=1}^{K_j^*}\alpha_{lk}$$

δ_{j0}은 인터셉트, δ_{jk}는 α_k에 의한 주효과를 의미하며, ACDM 모형에서의 모수와 동일 역할을 한다. G-DINA 모형에서 추가된 모수는 $\delta_{jkk'}$, ..., $\delta_{j12\cdots K_j^*}$이다. $\delta_{jkk'}$는 α_k와 $\alpha_{k'}$의 상호작용 효과(the interaction effect due to α_k and $\alpha_{k'}$), ..., $\delta_{j12\cdots K_j^*}$는 α_1, α_2, $\cdots$, $\alpha_{K_j^*}$의 상호작용 효과(the interaction

effect due to α_1, $\cdots$, $\alpha_{K_j^*}$)를 의미한다.

G-DINA 모형 문항 반응 함수에서 몇 개의 모수에 조건을 부여하면면 DINA, DINO, ACDM 모형을 얻을 수 있다. DINA 모형의 경우, δ_{j0}, $\delta_{j12\cdots K_j^*}$이외의 모든 모수가 0이면 된다. 이때 guess는 δ_{j0}, slip은 $1-\left(\delta_{j0}+\delta_{j12\cdots K_j^*}\right)$가 된다. DINO 모형의 경우, $k=1, 2, \cdots, K_j^*$, $k'=1, \cdots, K_j^*-1$, $k''>k'$, $\cdots, K_j^*$에 대하여 $\delta_{jk}=-\delta_{jk'k''}=\cdots=(-1)^{K_j^*+1}\delta_{j12\cdots K_j^*}$이면 된다. 이때 guess는 δ_{j0}이며, slip은 $1-\left(\delta_{j0}+\delta_{jk}\right)$이다. ACDM 모형의 경우, δ_{j0}, $\delta_{jk}(1\le k\le K_j^*)$이외의 모든 모수가 0이면 된다(Torre, 2011; 왕효원, 2018).

3. 서술형 평가를 위한 인지진단모형

위에서 설명한 DINA, DINO, ACDM, G-DINA 모형은 문항에 대한 학생의 반응을 정답과 오답 두 가지로 분류하는 이분 반응 모형이다. 이어서 설명할 Sequential G-DINA(sG-DINA; Ma & Torre, 2016)와 Diagnostic Tree Model(DTM; Ma, 2019)은 학생의 반응을 세 가지 이상의 반응으로 분류할 수 있도록 가정된 다분 반응 모형이며, 이는 서술형 문항에 관한 인지진단평가를 실시할 때 사용된다.

3.1. sG-DINA model

sG-DINA 모형은 문제해결 과정을 일련의 단계로 나누고, 다음 단계를 수행하기 위해서는 이전의 모든 단계를 성공적으로 마쳐야 한다고 가정하는 방식으로 피험자의 반응을 분석한다. 한 문항에 대해 피험자가 첫 번째 풀이 단계에서 실패하면 범주(category) 0, 1단계는 해결했으나 2단계에서 실패하면 범주 1로 할당되며, 이와 같은 방식으로 문항 반응의 범주를 나눈다. H개의 단계로 이루어진 문항이면 범주 0부터 범주 H까지 $H+1$개의 범주가 주어진다. 범주별로 요구하는 인지요소의 결합이 존재하는데, 범주가 다르다고 해서 요구하는 인지요소가 다를 필요는 없다. 보통 풀이 과정을 여러 단계로 서술해야 하는 서술형 문항 답안의 경우, 앞에서 서술된 명제가 다음 단계의 명제를 뒷받침하기 때문에 앞의 명제에 오류가 있으면 이어지는 모든 서술이 틀리게 되는 경향이 있다. 따라서 해당 모형은 서술형 문항에 적용하기 적절하게 설정되었다고 볼 수 있다.

sG-DINA 모형의 핵심이 되는 Sequential processing function을 소개하고자 한다. 일단 $S_j(h|\boldsymbol{\alpha}_l)$는 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$을 가지는 피험자가 문항 j를 풀 때 h단계를 성공적으로 해결할 확률을 의미한다. 이론적으로 0단계에는 아무런 인지요소가 요구되지 않으므로 $S_j(\mathrm{X}_j=0|\boldsymbol{\alpha}_l)=1$이며, 문항 j가 H_j의 단계로 이루어져 있다면 H_j+1단계부터는 존재하지 않으므로 $S_j(H_j+1|\boldsymbol{\alpha}_l)=0$이다. $P(X_j=h|\boldsymbol{\alpha}_l)$는 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$을 가지는

피험자가 문항 j에서 범주 h에 할당될 확률을 의미하며 다음이 성립한다.

$$P(X_j = h|\boldsymbol{\alpha}_l) = \left[1 - S_j(h+1|\boldsymbol{\alpha}_l)\right] \prod_{x=0}^{h} S_j(x|\boldsymbol{\alpha}_l)$$

그리고 문항 j를 해결한 피험자가 할당될 수 있는 범주는 0, 1, ..., H_j이므로 다음이 성립한다.

$$\sum_{h=0}^{H_j} P(X_j = h|\boldsymbol{\alpha}_l) = 1$$

위에서 제시된 Sequential processing function과 G-DINA 모형의 원리를 결합하여 sG-DINA 모형이 정의된다. $S_j(h|\boldsymbol{\alpha}_l)$은 G-DINA 모형의 문항 반응함수와 같은 원리로 다음과 같이 정의된다.

$$S_j(X_j = h|\boldsymbol{\alpha}^*_{ljh}) = \phi_{jh0} + \sum_{k=1}^{K^*_{jh}} \phi_{jhk}\alpha_{lk} + \sum_{k'=k+1}^{K^*_{jh}} \sum_{k=1}^{K^*_{jh}-1} \phi_{jhkk'}\alpha_{lk}\alpha_{lk'} + \cdots + \phi_{jh12\cdots K^*_{jh}} \prod_{k=1}^{K^*_{jh}} \alpha_{lk}$$

K^*_{jh}는 문항 j의 범주 h에 필요한 인지요소 숙달 패턴 개수이다. l은 1, ..., $2^{K^*_{jh}}$이며, $\boldsymbol{\alpha}^*_{ljh}$는 문항 j의 범주 h에서 축약된 피험자의 인지요소 숙달 패턴이며, $\boldsymbol{\alpha}^*_{ljh} = \left[\alpha_{l1}, ..., \alpha_{lk}, ..., \alpha_{lK^*_{jh}}\right]$와 같이 표현되는 벡터이다. ϕ_{jh0}는 인터셉트, ϕ_{jhk}는 α_{lk}의 주효과, $\phi_{jhkk'}$는 α_{lk}와 $\alpha_{lk'}$의 상호작용 효과를 의미한다.

3.2. Diagnostic Tree Model

DTM 모형은 2개 이상의 전략을 가진 서술형 문항에 대한 피험자의 반응을 분석할 수 있는 인지진단모형이다. 피험자의 문제해결 과정에서 나타나는 여러 전략과 전략별 풀이 단계를 모두 고려하며 모든 전략과 전략별 풀이 단계를 수형도 구조(tree structure)로 추상할 수 있다. 각 문제해결전략에 대해서는 sG-DINA 모형의 구조와 유사하기 때문에 DTM 모형은 sG-DINA 모형을 일반화한 것이라고 볼 수 있다.

Ma(2019)는 모의실험에서 가상 피험자의 응답 데이터를 토대로 DTM 모형, sG-DINA 모형, MS-DINA 모형에 의한 인지요소 숙달/미숙달 예측 정확도를 구하고 비교하였다. 예측 정확도의 판단기준으로는 PCA(proportion of correctly classified attributes), PCV(proportion of correctly classified attribute vectors)가 사용되었다. 모든 영역에서 DTM 모형의 인지요소 숙달/미숙달 예측 정확도가 sG-DINA 모형, MS-DINA 모형의 것보다 높게 나타났다. 또한 DTM 모형을 통해 문항 풀이 과정의 각 단계별 문항 모수 회복력(item parameter recovery)을 구하고 비교하였다. 문항 모수 회복력의 판단기준으로는 RMSE_{nx}(root mean square error for the pseudo-item response function for non-zero internal outcome x of pseudo-item n)가 사용되었다. 피험자 수가 많을수록, 문항 질이 높을수록, 풀이 과정의 초기 단계일수록 RMSE_{nx}가 낮게 나왔다. 그리고 수렴형 수형도 구조(convergent tree structure)보다 발산형 수형도 구조(divergent tree structure)의 RMSE_{nx}가 낮게 나왔다.[2)]

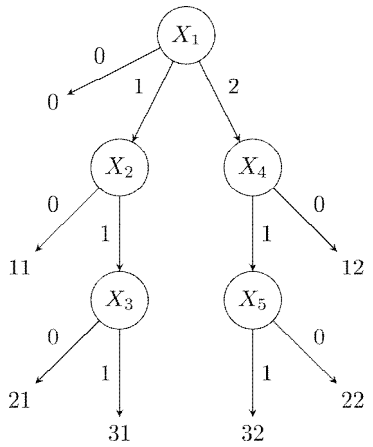
Ma(2019)는 DTM 모형을 적용하는 실제 데이터를 활용한 연구도 진행했는데, 2007년 수학·과학 성취도 평가 추이 국제 비교 연구(Trends in International Mathematics and Science Study 2007, TIMSS 2007)에서 사용된 문항과 응답 데이터를 사용하여 DTM 모형, sG-DINA 모형, MS-DINA 모형의 모형 적합도를 각각 구하고 비교했다. 모형 적합도의

---

2) RMSE_{nx}가 낮게 나올수록 문항 모수 회복력이 좋은 것이다. RMSE_{nx}에 관한 자세한 설명은 Ⅳ장에서 확인할 수 있다.

판단기준으로는 부트스트랩 방법(Bootstrap method)을 통한 χ^2(Pearson's chi-square) 통계량, Read & Cressie(1988)의 $CR(2/3)$ 통계량이 사용되었으며, 두 통계량 모두 DTM 모형의 p-value가 나머지 모형의 p-value보다 크게 측정되었다. 즉, TIMSS 2007의 시험 결과를 분석하는 데 DTM 모형이 sG-DINA 모형, MS-DINA 모형보다 더 적합하다는 결론을 얻었다.

DTM 모형에서는 문항에 대한 피험자의 풀이 과정을 수형도 구조(tree structure)를 통해 도식화하게 된다. 수형도 구조는 노드(node), 뿌리(root), 잎(leaf), 가지(branch)로 이루어졌다. 노드(node)는 풀이 과정에서 반응이 갈라질 수 있는 단계를 의미하며, 뿌리(root)는 가장 먼저 수행해야 할 노드를 의미한다. 잎(leaf)은 문항에 대한 피험자의 최종적인 반응에 해당하며 문항의 잎들은 서로 상호 독립적이다. 가지(branch)는 뿌리(root), 노드(node), 잎(leaf)을 서로 연결하는 경로를 나타낸다. [그림 Ⅱ-2]에서 X_1는 뿌리, X_1, X_2, X_3, X_4, X_5는 노드, 0, 11, 21, 31, 12, 22, 32는 잎, 화살표는 가지를 의미한다. 각 노드에서는 2개 이상의 결과가 도출된다. 0은 해당 노드에서 수행해야 할 물음에 아무런 성취도 보이지 못한 경우를 의미한다. 각 결과는 가지로 연결되어서 잎이나 다음 노드에 도달한다. 더이상 해결해야 할 물음이 없는 경우는 잎에 도달하게 되고, 이어서 해결해야 할 다음 물음이 남아있을 때는 다음 노드로 연결된다.



[그림 Ⅱ-2] 수형도 구조 예시 (Ma, 2019, p. 65)

Mislevy(1996)는 분수 뺄셈에 관해 $3\frac{1}{6}-\frac{5}{6}$를 계산하는 문항을 분석하였으며, 이 문제를 푸는 두 가지 전략을 제시했다. <표 Ⅱ-2>는 피험자가 $3\frac{1}{6}-\frac{5}{6}$를 계산할 때 나올 수 있는 반응을 전략별로 정리한 것이며 이는 [그림 Ⅱ-2]의 수형도 구조로 표현된다.

<표 Ⅱ-2> $3\frac{1}{6}-\frac{5}{6}$ 문항의 두 가지 해결전략 (Ma, 2019, p. 62)

전략 A		전략 B	
풀이의 단계	반응 분류	풀이의 단계	반응 분류
$=\frac{19}{6}-\frac{5}{6}$	11	$=2\frac{7}{6}-\frac{5}{6}$	12
$=\frac{14}{6}$	21	$=2\frac{2}{6}$	22
$=2\frac{1}{3}$	31	$=2\frac{1}{3}$	32

DTM 모형에서 하나의 문항은 여러 전략과 단계로 이루어졌기 때문에 하나의 문항에 대한 Q행렬은 전략별, 단계별 인지요소 정보를 담고 있어서 보통 두 개 이상의 행으로 표현된다. <표 Ⅱ-3>은 $3\frac{1}{6}-\frac{5}{6}$ 문항에 관한 Q행렬을 나타낸 것이다. 여기서 α_1, α_2, ..., α_7은 해당 문항을 포함한 문항 세트에서 숙달 여부를 추정하고자 하는 인지요소이며, α_1은 분수의 뺄셈, α_2는 분수를 대분수로 변환하거나 약분하기, α_3는 분수에서 정수를 분해하기, α_4는 대분수의 정수를 하나 빼서 분모에 더하기, α_5는 정수를 분수로 변환, α_6는 대분수를 분수로 변환하기, α_7는 뺄셈에서 받아내림을 의미한다(Mislevy, 1996).

<표 Ⅱ-3> $3\frac{1}{6}-\frac{5}{6}$ 문항에 관한 Q행렬

노드	범주	α_1	α_2	α_3	α_4	α_5	α_6	α_7
1	1	0	0	0	0	0	1	0
1	2	0	0	0	1	0	0	0
2	1	1	0	0	0	0	0	0
3	1	0	1	0	0	0	0	0
4	1	1	0	0	0	0	0	0
5	1	0	1	0	0	0	0	0

첫 번째 열의 노드와 두 번째 열의 반응 범주에 관한 인지요소는 세 번째 열부터 나열된다. 뿌리(첫 번째 노드)의 반응 범주 0에 관한 인지요소는 고려할 필요가 없고, 뿌리가 아닌 노드의 반응 범주 0은 해당 노드로 연결된 이전 노드의 반응 범주에 관한 인지요소 정보가 Q행렬에 이미 존재하므로 이 경우도 인지요소를 고려할 필요가 없다. 각 노드의 반응 범주는 항상 1부터 순서대로 입력하면 되므로 실제 R 프로그래밍 언어의 GDINA pakage에서 사용되는 Q행렬은 첫 번째 열에 문항 번호를 표시하고 범주에 관한 열은 입력하지 않는다(Ma et al., 2022). 문항 세트에서 해당 문항의 번호가 3이었다면 Q행렬은 <표 Ⅱ-4>와 같다.

<표 Ⅱ-4> DTM 모형 인지진단평가에서 사용되는 Q행렬

문항 번호	노드	α_1	α_2	α_3	α_4	α_5	α_6	α_7
1	1	1	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3	1	0	0	0	0	0	1	0
3	1	0	0	0	1	0	0	0
3	2	1	0	0	0	0	0	0
3	3	0	1	0	0	0	0	0
3	4	1	0	0	0	0	0	0
3	5	0	1	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20	4	0	0	1	0	0	0	1

DTM 모형, sG-DINA 모형, G-DINA 모형은 모두 동일 구조의 G-DINA 모형 문항 반응 함수를 가진다. 인지요소 숙달 패턴 $\boldsymbol{\alpha}_l$을 가지는 피험자와 문항 j, 노드 n, 반응 범주 x에 대하여 $\pi_{jnx}(\boldsymbol{\alpha}_l)$는 피험자가 해당 반응에 할당될 확률을 의미하고, $h(\delta_{jnx}, \boldsymbol{\alpha}_l)$는 문항 반응 함수를 의미한다.

$$\pi_{jnx}(\boldsymbol{\alpha}_l) = \frac{\exp\left[\mathrm{h}(\boldsymbol{\delta}_{jnx}, \boldsymbol{\alpha}_l)\right]}{\sum_{x=0}^{M_{jn}} \exp\left[\mathrm{h}(\boldsymbol{\delta}_{jnx}, \boldsymbol{\alpha}_l)\right]}$$

$$\begin{aligned}\mathrm{h}(\boldsymbol{\delta}_{jnx}, \boldsymbol{\alpha}_l) &= \mathrm{h}(\boldsymbol{\delta}_{jnx}, \boldsymbol{\alpha}^*_{ljnx}) \\ &= \delta_{jnx0} + \sum_{k=1}^{K^*_{jnx}} \delta_{jnxk}\alpha_{lk} + \sum_{k=1}^{K^*_{jnx}-1} \sum_{k'>1}^{K^*_{jnx}} \delta_{jnxkk'}\alpha_{lk}\alpha_{lk'} \\ &\quad + \cdots + \delta_{jnx12\cdots K^*_{jn}} \prod_{k=1}^{K^*_{jnx}} \alpha_{lk}\end{aligned}$$

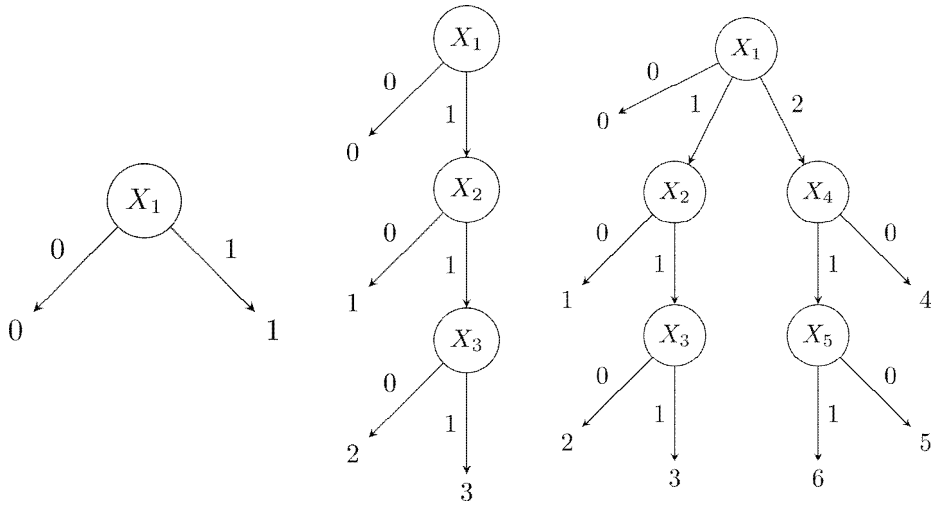
K^*_{jnx}는 문항 j, 노드 n, 반응 범주 x에 필요한 인지요소 숙달 패턴 개수이고, l은 1, ..., $2^{K^*_{jnx}}$이며, $\boldsymbol{\alpha}^*_{ljnx}$는 문항 j, 노드 n, 반응 범주 x에서 축약된 피험자의 인지요소 숙달 패턴이며, $\boldsymbol{\alpha}^*_{ljnx} = \left[\alpha_{l1}, ..., \alpha_{lk}, ..., \alpha_{lK^*_{jnx}}\right]$와 같이 표현되는 벡터다. δ_{jnx0}는 인터셉트, ϕ_{jnxk}는 α_{lk}의 주 효과, $\delta_{jnxkk'}$는 α_{lk}와 $\alpha_{lk'}$의 상호작용 효과를 의미한다.

Y_{ij}를 문항 j에 대한 피험자 i의 관찰된 최종 반응, ν_j를 문항 j에서 가능한 모든 반응으로 구성된 표본 공간이라 할 때, $y \in \nu_j$에 대하여 $P(Y_{ij} = y|\boldsymbol{\alpha}_l)$는 $\pi_{jnx}(\boldsymbol{\alpha}_l)$의 곱으로 표현된다. S_{jy}는 문항 j에서 최종 반응이 y인 모든 경로의 개수를 의미한다. N_j는 문항 j의 노드의 개수를 의미하고, M_{jn}는 문항 j의 노드 n에서 가능한 범주의 개수를 의미한다. $t^{(s)}_{jnxy}$는 0 또는 1의 값을 가지는데, 노드 n에서 반응 범주 x로의 대응이 경로 s에서

일어나면 1의 값을 가지고, 그렇지 않으면 0의 값을 가진다. DTM 모형에서 최종적인 문항 반응 함수는 다음과 같이 표현된다.

$$P(Y_{ij}=y|\boldsymbol{\alpha}_l)=\sum_{s=1}^{S_{jy}}\prod_{n=1}^{N_j}\prod_{x=0}^{M_{jn}}\left[\pi_{jnx}(\boldsymbol{\alpha}_l)\right]^{t_{jnxsy}^{(s)}}$$

DTM 모형과 sG-DINA 모형의 각 노드의 반응에 관한 문항 반응 함수는 G-DINA 모형의 것과 동일하다. 즉, DTM 모형은 sG-DINA 모형의 확장이며, sG-DINA 모형은 G-DINA 모형의 확장이라 할 수 있다. [그림 Ⅱ-3]은 수형도 구조를 사용하여 세 모형을 비교한 것이다.



G-DINA 모형 sG-DINA 모형 DTM 모형

[그림 Ⅱ-3] G-DINA 모형, sG-DINA 모형, DTM 모형의 비교

Ⅲ. 수학 서술형 문항의 풀이 과정 구조화

1. 연구방법

이 장에서는 DTM 모형을 적용한 인지진단평가를 설계하기 위해 수학 서술형 문항의 풀이 과정 구조화 방법을 제안하고자 한다. 수학 서술형 문항에 대한 피험자의 반응을 분류하는 Ma & Torre(2016)의 방법과 Ma(2019)의 방법을 참고하여 수학 서술형 문항 풀이 과정 분석틀을 고안하였다.

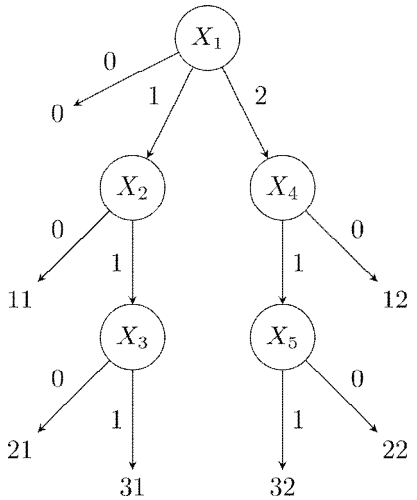
2. 연구결과

Ma & Torre(2016)는 수학 서술형 문항에 대한 인지진단모형으로 sG-DINA를 제안하였다. sG-DINA 모형에서 서술형 문항의 풀이 과정은 일련의 단계로 이루어져 있으며, 피험자는 N 번째 단계를 수행하기 위해서는 (N-1) 번째 단계를 성공적으로 마쳐야 한다고 가정한다. 보통 수학 문제에 대한 답과 그 이유를 서술형으로 답하는 것은 여러 명제를 나열하는 것이며, 각 명제는 그에 앞서 서술된 명제 혹은 사전에 알려진 수학적 원리나 정리에 의해 참임이 보장된다. 만약 앞서 서술된 명제가 뒤따르는 명제를 함의하지 않거나 중간에 거짓인 명제가 사용된다면, 그 이후로 얻어지는 추론은 수학적으로 옳다고 할 수 없다. 따라서 문제해결 과정을 일련의 단계 이행으로 보는 sG-DINA 모형의 가정은 합리적으로 보인다. 하지만 sG-DINA 모형에서는 각 문항이 단일한 풀이 방법으로 해결됨을 가정했다는 점에서 한계를 가진다. 일반적으로 하나의 수학 문제에 대한 해결전략은 단일하지 않다. 게다가 학생들은 문제해결 과정에서 다양한 해결전략을 사용한다.

Ma(2019)는 Ma & Torre(2016)가 제안한 sG-DINA 모형을 확장하여 2개 이상의 풀이 전략이 존재하는 수학 서술형 문항의 풀이 과정을 분석할 수 있는 DTM 모형을 제안하였다. 평가자와 피험자는 수형도 구조(tree

structure)를 통해 주어진 문항에서 몇 개의 해결전략이 존재하는지 알 수 있으며, 전략별로 최종 결론에 도달하기 위해 몇 개의 단계를 거쳐야 하는지 알 수 있다.

하지만 수형도 구조만으로 수학 서술형 문항의 풀이 과정을 분석하는데 한계가 있다. 수형도 구조는 도식이기 때문에 문항에 대한 피험자의 예상 반응 개수, 문제해결을 위해 거쳐야 하는 단계의 길이 등을 수치적으로 비교·분석하기 어렵다. 그리고 문항의 전략별 풀이 단계마다 요구되는 인지요소 정보는 수형도 구조를 통해 파악할 수 없으며 DTM 모형의 Q행렬을 통해 파악할 수 있다. [그림 Ⅲ-1]은 $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조와 Q행렬을 나타낸 것이다. 문항 $3\frac{1}{6}-\frac{5}{6}$에 대한 피험자 예상 반응 개수와 문제해결을 위해 거쳐야 하는 단계의 길이를 도출하려면 노드(node)의 개수를 세서 수치화하는 과정이 필요하다. 또한 풀이의 단계마다 요구되는 인지요소 정보를 얻기 위해서는 수형도 구조와 Q행렬을 동시에 분석해야 한다.

수형도 구조	Q행렬
	$\begin{bmatrix} j\,1\,0\,0\,0\,0\,0\,1\,0 \\ j\,1\,0\,0\,0\,1\,0\,0\,0 \\ j\,2\,1\,0\,0\,0\,0\,0\,0 \\ j\,3\,0\,1\,0\,0\,0\,0\,0 \\ j\,4\,1\,0\,0\,0\,0\,0\,0 \\ j\,5\,0\,1\,0\,0\,0\,0\,0 \end{bmatrix}$

[그림 Ⅲ-1] $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조, Q행렬

위에서 언급한 어려움을 해소함과 동시에 수학 서술형 문항에 대하여 문항과 관련된 인지요소 개수, 문항 풀이 과정의 단계 길이, 문항 해결을 위한 풀이 전략 개수를 수치로 분석할 수 있는 분석틀을 제안하고자 한다.

2.1. 수학 서술형 문항의 풀이 과정 분석틀

이 장에서 수학 서술형 문항이란 해결 과정에서 다양한 종류의 개념 · 원리 · 기능이 요구될 수 있고, 해결을 위해 여러 단계를 거쳐야 할 수 있고, 다양한 해법의 반응이 존재할 수 있는 것을 말한다. 따라서 수학 서술형 문항의 풀이 과정 분석틀은 인지요소(attribute), 단계(step), 반응(response) 3가지 차원을 가진다. 수학 서술형 문항의 풀이 과정 분석틀의 각 차원에 대해 살펴보겠다.

2.1.1. 인지요소(attribute)

문제해결 과정에서 사용되는 인지요소를 의미한다. 이는 한 문항을 해결할 때 얼마나 다양한 종류의 개념 · 원리 · 기능이 요구되는지 설명해준다. 인지진단모형에서는 Q행렬을 통해 확인할 수 있으며, DTM 모형의 수형도 구조를 통해서는 확인할 수 없다. 문항 j와 관련된 인지요소 개수를 K_j라 하면, 문항 j를 통해서 구별할 수 있는 인지요소 숙달 패턴은 2^{K_j}개 이하다. 문항 j에 대하여 K_j가 클수록 문항에 대한 피험자의 반응(response)이 다양할 수 있다.

2.1.2. 단계(step)

문제해결을 위해 거쳐야 하는 모든 과정을 세부적으로 나눈 것의 한 단위를 의미한다. 예를 들어, 방정식 $x+1=3$은 좌변의 1을 우변으로 이항하는 하나의 단계만 거치면 해를 구할 수 있다. 반면에 방정식 $7(3x+1)-5(6x+2)=9$는 분배법칙을 사용하여 식을 전개하고, 동류항끼

리 식을 더하고, 숫자로만 구성된 식을 우변으로 이항한 후 x의 계수로 양변을 나누어야 한다. 즉, 네 단계를 거쳐야 한다.

한 문항을 완벽하게 해결하는 전략은 항상 단일하지 않으며, 서로 다른 전략은 단계 길이가 항상 같지 않다. 예를 들어 방정식 $7(3x+1)-5(6x+2)=9$를 푸는 두 가지 전략은 <표 Ⅲ-1>과 같이 단계 길이가 서로 다르다.

<표 Ⅲ-1> 방정식 $7(3x+1)-5(6x+2)=9$를 푸는 두 가지 전략

단계	전략 A	전략 B
문제상황	$7(3x+1)-5(6x+2)=9$	$7(3x+1)-5(6x+2)=9$
1	$21x+7-30x-10=9$	$7(3x+1)-5\times2(3x+1)=9$
2	$-9x-3=9$	$(7-10)(3x+1)=9$
3	$-9x=12$	$3x+1=-3$
4	$x=-\frac{4}{3}$	$3x=-4$
5		$x=-\frac{4}{3}$

이처럼 한 문항을 완벽하게 해결하는 데 필요한 단계의 많고 적음은 문항을 해결하는 전략마다 다를 수 있다. 따라서 한 문항에 대한 단계 길이는 각 해결전략의 단계 길이의 평균으로 정의할 수 있다. 위 문항의 경우, 주어진 두 방법이 전부라면 단계 길이는 4.5다.

2.1.3. 반응(response)

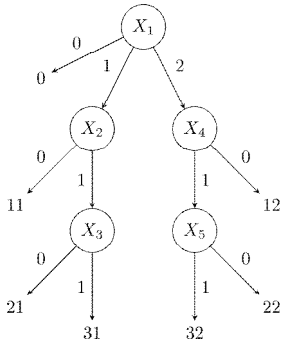
피험자가 문항을 풀었을 때 나올 수 있는 다양한 반응을 여러 개로 분류한 것의 한 유형을 의미한다. 반응 개수는 주어진 문항에 대한 피험자의 반응이 얼마나 다양한지를 말해준다. 한 문항에 대한 모든 반응(response)

의 집합은 수형도 구조의 잎(leaf)들의 집합에 일대일 대응된다. 그래서 반응 개수는 문항에 관한 Q행렬의 행의 개수보다 한 개 많다. 위의 인지요소(attribute)에서 언급했듯이 문항과 관련된 인지요소 개수가 많을수록 피험자의 반응이 다양할 수 있다. 하지만 항상 그런 것은 아니다. 문항에 따라 인지요소가 많이 요구되지만, 반응이 두 개뿐일 수 있고, 인지요소가 적게 요구되지만, 다양한 반응을 가질 수 있다.

2.2. DTM 모형의 인지진단평가 문항 풀이 과정 구조화

수학 서술형 문항의 풀이 과정 분석틀을 활용하면 DTM 모형의 인지진단평가 문항 풀이 과정을 구조화할 수 있다. <표 Ⅲ-2>는 $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조, Q행렬, 두 가지 해결전략을 나타낸 것이다.

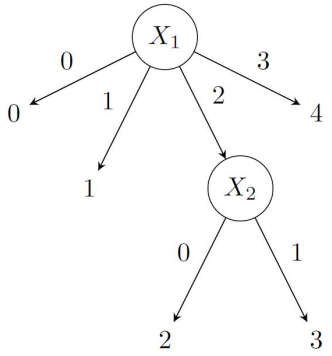
<표 Ⅲ-2> $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조, Q행렬, 해결전략

수형도 구조	Q행렬	전략 A		전략 B	
		풀이의 단계	반응	풀이의 단계	반응
	$\begin{bmatrix} j\,1\,0\,0\,0\,0\,0\,0\,1\,0 \\ j\,1\,0\,0\,0\,1\,0\,0\,0 \\ j\,2\,1\,0\,0\,0\,0\,0\,0 \\ j\,3\,0\,1\,0\,0\,0\,0\,0 \\ j\,4\,1\,0\,0\,0\,0\,0\,0 \\ j\,5\,0\,1\,0\,0\,0\,0\,0 \end{bmatrix}$	$=\frac{19}{6}-\frac{5}{6}$	11	$=2\frac{7}{6}-\frac{5}{6}$	12
		$=\frac{14}{6}$	21	$=2\frac{2}{6}$	22
		$=2\frac{1}{3}$	31	$=2\frac{1}{3}$	32

세 번째 열부터 아홉 번째 열까지의 일곱 개의 열은 주어진 문항을 포함한 문항 세트에서 확인하고자 하는 인지요소가 각 반응과 관련이 있는지

를 나타낸다. 다섯 번째, 일곱 번째, 아홉 번째 열의 모든 성분이 0인데, 이는 주어진 문항이 세 번째, 다섯 번째, 일곱 번째 인지요소와 관련이 없다는 의미다. 즉, 0 벡터가 아닌 열이 4개 있으므로 이 문항과 관련된 인지요소 개수는 4다.

<표 Ⅲ-2>의 수형도 구조를 보면 31과 32가 문항을 완벽하게 해결하여 나올 수 있는 반응인 것으로 예상될 수 있다. 하지만 수형도 구조만을 확인하여 문항의 단계 길이를 정확히 알 수 없다. 다시 말해서, 수형도 구조의 뿌리(root)에서 0이 아닌 가지(branch)를 통해 가장 먼 노드(node)를 거쳐 0이 아닌 가지로 연결된 잎(leaf)이 항상 정답 반응일 필요는 없다는 것이다. 예를 들어 [그림 Ⅲ-2]에 나타난 수형도 구조의 문항은 마지막에 도달할 수 있는 반응이 1, 3, 4로 세 가지인데, 각 반응이 문항을 완벽하게 해결하여 나온 반응인지 아닌지는 수형도 구조에 드러나지 않으며, 이는 모범답안을 확인해야 알 수 있다.



[그림 Ⅲ-2] 수형도 구조 예시

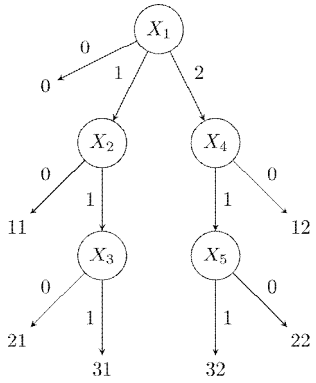
<표 Ⅲ-2>의 두 개의 해결전략 A, B에서 풀이의 단계별 반응을 대응시켜보면, 7개의 반응 0, 11, 21, 31, 12, 22, 32 중 31과 32만이 문항을 완벽하게 해결하여 나올 수 있는 반응임을 확인할 수 있다. 이처럼 한 문항을 완벽하게 해결하는 전략은 항상 단일하지 않다. 그리고 서로 다른 전략은 단계 길이가 항상 같을 수 없다. 따라서 한 문항에 대한 단계 길이는 각 해

결전략의 단계 길이의 평균으로 정의한다고 했다. 이 문항에서는 31과 32 모두 3개의 단계를 거쳐 나올 수 있는 반응이므로 이 문항 풀이 과정 단계 길이는 3과 3의 평균인 3이다.

문항의 반응 개수를 세보겠다. 문항의 반응 개수는 수형도 구조, Q행렬 각각에서 얻을 수 있다. 수형도 구조에서는 잎의 개수를 세면 된다. 잎은 0, 11, 21, 31, 12, 22, 32로 7개이므로 문항의 반응 개수는 7이다. Q행렬에는 문항의 풀이 과정에서 0의 반응에 해당하는 행이 없다. 따라서 Q행렬의 행 개수에 1을 더하여 문항의 반응 개수를 구할 수 있다. Q행렬은 6개의 행으로 이루어져 있으므로 문항의 반응 개수는 7이 된다.

위 내용을 종합하면, 수학 서술형 문항의 풀이 과정 분석틀을 통해 $3\frac{1}{6}-\frac{5}{6}$ 문항의 풀이 과정에 관한 정보를 분석할 수 있다. <표 Ⅲ-3>은 분석 결과를 수형도 구조, Q행렬과 함께 나타낸 것이다.

<표 Ⅲ-3> $3\frac{1}{6}-\frac{5}{6}$ 문항의 수형도 구조, Q행렬, 분석틀 분석

수형도 구조	Q행렬	인지 요소	단계	반응
	$\begin{bmatrix} j\,1\,0\,0\,0\,0\,0\,1\,0 \\ j\,1\,0\,0\,0\,1\,0\,0\,0 \\ j\,2\,1\,0\,0\,0\,0\,0\,0 \\ j\,3\,0\,1\,0\,0\,0\,0\,0 \\ j\,4\,1\,0\,0\,0\,0\,0\,0 \\ j\,5\,0\,1\,0\,0\,0\,0\,0 \end{bmatrix}$	4	3	7

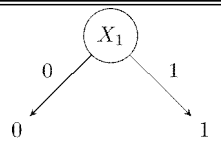
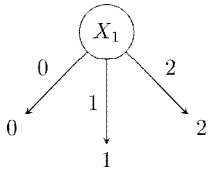
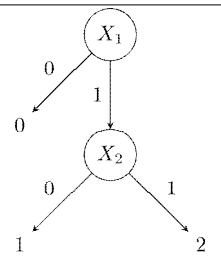
Ⅳ. 다양한 구조의 서술형 문항을 포함한 DTM 모형 평가에 대한 모의실험

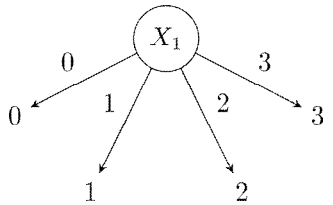
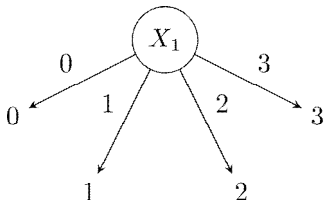
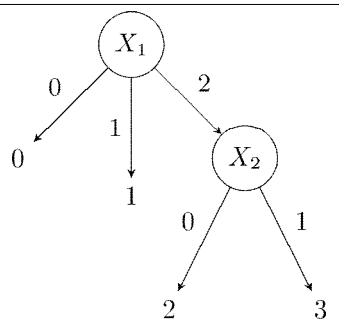
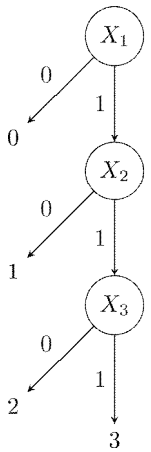
1. 연구방법

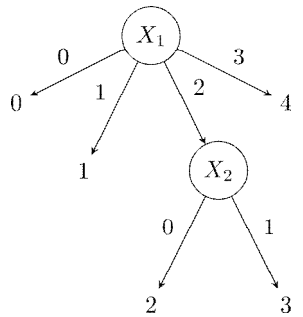
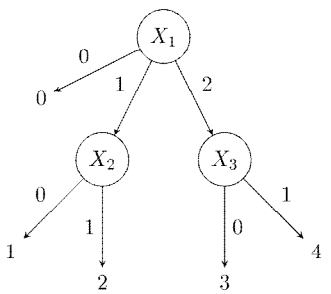
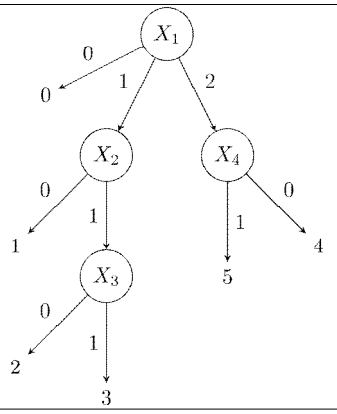
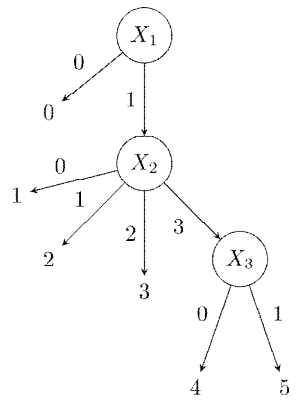
1.1. 실험 설계

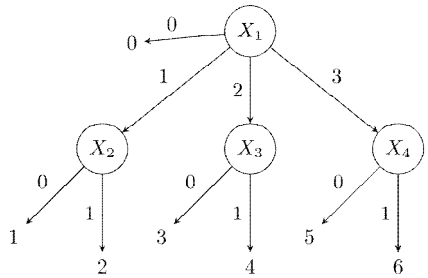
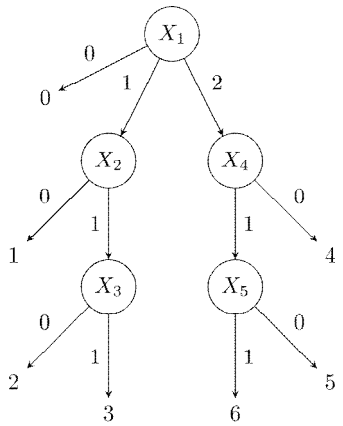
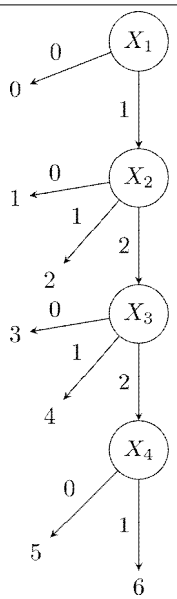
<표 Ⅳ-1>은 가상의 수학 서술형 문항 16개 유형을 나타낸 것이다. 각 문항 유형에 대한 수형도 구조, Q행렬, 분석틀 분석을 나타냈다. 수형도 구조는 실재하는 수학 서술형 문항이 지닐 법한 구조이며, Q행렬은 <표 Ⅳ-2>의 모의실험 평가 문항 세트 구성에서 사용되는 Q행렬을 나타낸 것이다. 또한 분석틀 분석의 문항별 인지요소 개수, 단계 길이, 반응 개수가 골고루 분포되도록 구성하였다.

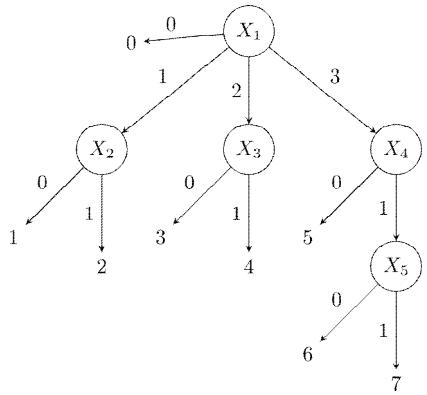
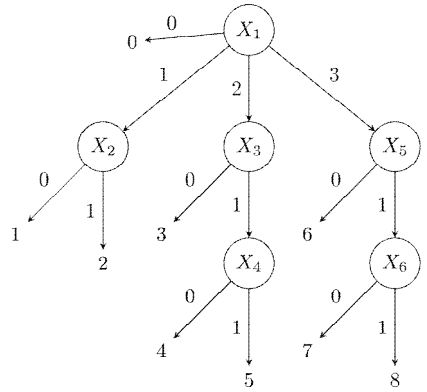
<표 Ⅳ-1> 문항 유형별 수형도 구조, Q행렬, 분석틀 분석

문항 유형	수형도 구조	Q행렬	인지요소	단계	반응
a		하나의 인지요소를 묻는 이분 반응 문항	1	1	2
b		$\begin{bmatrix} 4 & 1 & 0 & 1 & 0 \\ 4 & 1 & 0 & 0 & 1 \end{bmatrix}$	2	1	3
		$\begin{bmatrix} 6 & 1 & 1 & 0 & 0 & 0 & 0 \\ 6 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$	2		
		$\begin{bmatrix} 8 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 8 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	2		
c		$\begin{bmatrix} 5 & 1 & 0 & 1 & 0 \\ 5 & 2 & 1 & 0 & 0 \end{bmatrix}$	2	2	3
		$\begin{bmatrix} 7 & 1 & 0 & 0 & 1 & 0 & 0 \\ 7 & 2 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$	2		
		$\begin{bmatrix} 9 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 9 & 2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$	2		

d		$\begin{bmatrix}6&1&1&0&0\\6&1&1&1&0\\6&1&1&1&1\end{bmatrix}$ 3 $\begin{bmatrix}8&1&0&1&0&0&0\\8&1&0&1&1&0&0\\8&1&0&0&1&1&0\end{bmatrix}$ 3 $\begin{bmatrix}10&1&0&0&0&0&1&0&0\\10&1&0&0&0&0&1&1&0\\10&1&0&0&0&0&1&1&1\end{bmatrix}$ 3	1	4
e		$\begin{bmatrix}7&1&0&0&1\\7&2&1&0&0\\7&2&1&1&0\end{bmatrix}$ 3 $\begin{bmatrix}9&1&0&1&0&0&0\\9&2&0&0&0&1&0\\9&2&0&0&0&1&1\end{bmatrix}$ 3 $\begin{bmatrix}11&1&0&1&0&0&0&0&0\\11&2&0&0&1&0&0&0&0\\11&2&0&0&1&0&1&0&0\end{bmatrix}$ 3	2	4
f		$\begin{bmatrix}8&1&0&1&0\\8&1&0&0&1\\8&2&1&0&0\end{bmatrix}$ 3 $\begin{bmatrix}10&1&0&0&0&1&0\\10&1&1&0&0&0&1\\10&2&0&0&1&0&0\end{bmatrix}$ 4 $\begin{bmatrix}12&1&0&1&0&0&0&1&0\\12&1&0&0&1&0&1&0&0\\12&2&0&0&0&0&0&0&1\end{bmatrix}$ 5	2	4
g		$\begin{bmatrix}11&1&0&0&1&0&0\\11&2&0&1&0&0&0\\11&3&0&0&0&1&0\end{bmatrix}$ 3 ——— $\begin{bmatrix}13&1&0&0&0&1&0&0&0\\13&2&1&0&0&0&0&0&0\\13&3&0&1&0&0&0&0&0\end{bmatrix}$ 3	3	4

h		$\begin{bmatrix}12&1&0&1&0&0&0\\12&1&0&0&0&0&1\\12&1&0&1&0&1&0\\12&2&0&0&1&1&0\end{bmatrix}$ 4 $\begin{bmatrix}14&1&0&1&0&0&0&0&0\\14&1&0&1&1&0&0&0&0\\14&1&0&1&0&0&0&1&0\\14&2&0&0&0&0&1&0&0\end{bmatrix}$ 4	1.5	5
i		$\begin{bmatrix}13&1&0&1&0&0&0\\13&1&0&0&0&0&1\\13&2&0&0&1&0&0\\13&3&1&0&0&0&0\end{bmatrix}$ 4 $\begin{bmatrix}15&1&0&1&0&0&0&0&0\\15&1&0&1&1&0&0&0&0\\15&2&0&0&0&0&1&0&0\\15&3&0&0&0&0&0&0&1\end{bmatrix}$ 4	2	5
j		$\begin{bmatrix}14&1&1&0&0&0&0\\14&1&1&0&1&0&0\\14&2&0&1&0&0&0\\14&3&0&0&0&1&0\\14&4&0&0&0&0&1\end{bmatrix}$ 5 $\begin{bmatrix}16&1&1&0&0&0&0&0&0\\16&1&1&0&0&1&0&0&0\\16&2&0&1&0&0&0&0&0\\16&3&0&0&0&0&0&1&0\\16&4&0&0&0&0&0&0&1\end{bmatrix}$ 5	2.5	6
k		$\begin{bmatrix}15&1&0&0&1&0&0\\15&2&0&1&0&0&0\\15&2&0&1&0&1&0\\15&2&0&1&0&1&1\\15&3&1&0&0&0&0\end{bmatrix}$ 5 $\begin{bmatrix}17&1&0&0&1&0&0&0&0\\17&2&0&0&1&1&0&0&0\\17&2&0&0&0&1&1&0&0\\17&2&0&1&0&1&0&1&0\\17&3&0&1&0&0&0&0&1\end{bmatrix}$ 6	2.5	6

l		$\begin{bmatrix} 18 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 18 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 18 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 18 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 18 & 3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 18 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	6	2	7
m		$\begin{bmatrix} 19 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 19 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 19 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 19 & 3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 19 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 19 & 5 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$	6	3	7
n		$\begin{bmatrix} 20 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 20 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 20 & 2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 20 & 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 20 & 3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 20 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	6	4	7

o		$\begin{bmatrix} 21 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 21 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 21 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 21 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 21 & 3 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 21 & 4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 21 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$	7	$\frac{7}{3}$	8
p		$\begin{bmatrix} 22 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 22 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 22 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 22 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 22 & 3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 22 & 4 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 22 & 5 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 22 & 6 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$	7	$\frac{8}{3}$	9

<표 Ⅳ-1>에 설명된 각 문항 유형을 사용하여 <표 Ⅳ-2>와 같이 세 개의 문항 세트를 구성했다. 총 문항 수, 사용된 문항 유형의 종류 개수, 문항 세트 인지요소 개수, 문항별 인지요소 개수 평균, 문항별 단계 길이 평균, 문항별 반응 개수 평균 각각이 평가 A에서 평가 B, 평가 C로 갈수록 비슷한 값만큼 커지도록 설계했다. 상대적으로 평가 A는 단순하고 쉬운 문제들로 구성된 시험이며, 평가 C는 복잡하고 어려운 문제들로 구성된 시험이라고 할 수 있다.

<표 Ⅳ-2> 모의실험 평가 문항 세트 구성

문항 세트	평가 A	평가 B	평가 C
총 문항 수	8	15	22
사용된 문항 유형	a, a, a, b, c, d, e, f (6가지)	a, a, a, a, a, b, c, d, e, f, g, h, I, j, k (11가지)	a, a, a, a, a, a, a, b, c, d, e, f, g, h, I, j, k, l, m, n, o, p (16가지)
문항 세트 인지요소 개수	3	5	7
문항별 인지요소 개수 평균	2	2.667	3.455
문항별 단계 길이 평균	1.375	1.653	1.841
문항별 반응 개수 평균	3	3.6	4.363

1.2. 분석 방법

모의실험을 위해서 먼저 개별적인 인지요소 숙달 패턴을 지닌 가상의 피험자그룹을 생성했다. 각 피험자는 인지요소 숙달 패턴을 바탕으로 각 문항에서 보일 수 있는 반응이 주어진다. DTM 모형을 통해 평가 전체에서 피험자가 보인 반응을 바탕으로 피험자의 인지요소 숙달 패턴을 추정하고, 이를 피험자의 원래 인지요소 숙달 패턴과 비교하여 얼마나 정확히 추정되었는지 확인한다.

피험자 수, 문항 질(item quality)을 변화시켜가면서 정확하게 추정한 인지요소 비율(proportion of correctly classified attributes, PCA), 정확하게 추정한 인지요소 패턴 비율(proportion of correctly classified attribute vectors, PCV), 문항별 평균 제곱근 오차(root mean square error for item j, $RMSE_j$)를 구하고 그 결과를 분석한다. PCA와 PCV의 의미는 각각 다음과 같다.

$$\mathrm{PCA} = \frac{\sum_{r=1}^{R}\sum_{i=1}^{I}\sum_{k=1}^{K} 1_{[\alpha_{ik}=\hat{\alpha}_{ik}]}^{(r)}}{R \times I \times K}$$

$$\mathrm{PCV} = \frac{\sum_{r=1}^{R}\sum_{i=1}^{I} 1_{[\alpha_{i}=\hat{\alpha}_{i}]}^{(r)}}{R \times I}$$

여기서 R은 반복 횟수, I는 피험자 수, K는 한 평가에 관여하는 인지요소 개수를 의미한다. 모든 분석과정에서 $R=100$으로 고정했다. $1_{[\alpha_{ik}=\hat{\alpha}_{ik}]}^{(r)}$는 r번째 추정에서 i번째 피험자가 k번째 인지요소를 지녔는지의 여부에 대하여 추정된 것과 원래의 것이 같으면 1, 다르면 0의 값을 가진다. $1_{[\alpha_{i}=\hat{\alpha}_{i}]}^{(r)}$는 r번째 추정에서 i번째 피험자의 인지요소 숙달 패턴에 대하여 추정된 것과 원래의 것이 같으면 1, 다르면 0의 값을 가진다.

$\mathrm{RMSE_j}$의 의미는 다음과 같다.

$$\mathrm{RMSE_j} = \sqrt{\frac{\sum_{r=1}^{R}\sum_{x=1}^{M_j}\sum_{l=1}^{2^{K_{jx}^{*}}}\left(\hat{\pi}^{(r)}\left(\boldsymbol{\alpha}_{ljx}^{*}\right)-\pi^{(r)}\left(\boldsymbol{\alpha}_{ljx}^{*}\right)\right)^2}{R \times \sum_{x=1}^{M_j} 2^{K_{jx}^{*}}}}$$

M_j는 문항 j에서 가능한 반응 개수를 의미한다. K_{jx}^{*}는 문항 j, 반응 범주 x에 필요한 인지요소 개수이고, l은 1, ..., $2^{K_{jx}^{*}}$이며, $\boldsymbol{\alpha}_{ljx}^{*}$는 문항 j, 반응 범주 x에서 축약된 임의의 인지요소 숙달 패턴이며, $\boldsymbol{\alpha}_{ljx}^{*} = \left[\alpha_{l1}, ..., \alpha_{lk}, ..., \alpha_{lK_{jx}^{*}}\right]$와 같이 표현되는 벡터다. $\hat{\pi}^{(r)}\left(\alpha_{ljx}^{*}\right)$은 r번째 추정에서 l의 인지요소 숙달 패턴을 지닌 피험자가 문항 j에서 x의 반응을 나타낼 확률에 대한 DTM 모형을 통한 추정값을 의미한다. $\pi^{(r)}\left(\alpha_{ljx}^{*}\right)$는 r번째 추정에서 l의 인지요소 숙달 패

턴을 지닌 피험자가 문항 j에서 x의 반응을 나타낼 확률을 모의실험 상황에 맞게 설정한 값이다.

Ma(2019)는 RMSE_{nx}(root mean square error for the pseudo-item response function for non-zero internal outcome x of pseudo-item n)를 구하여 분석했다.

$$\mathrm{RMSE}_{nx} = \sqrt{\frac{\sum_{r=1}^{R}\sum_{j=1}^{J^*}\sum_{l=1}^{2^{K^*_{jnx}}}\left(\hat{\pi}^{(r)}\left(\boldsymbol{\alpha}^*_{ljnx}\right)-\pi^{(r)}\left(\boldsymbol{\alpha}^*_{ljnx}\right)\right)^2}{R\times\sum_{j=1}^{J^*}2^{K^*_{jnx}}}}$$

하지만 본 연구에서는 다양한 반응을 가정한 문항을 사용하여 인지진단평가를 진행하는 것에 의미를 두고 있으며, 비교의 대상은 문항 구조의 차이이므로 RMSE_j를 구해서 비교하는 것이 더 적절하다. DTM 모형을 통한 파라미터 추정은 R 프로그래밍 언어와 R의 GDINA package를 사용했다(Ma et al., 2022).

문항 질은 고, 중, 저로 구분했는데, 각각은 추측 모수인 guess와 실수 모수인 slip을 각각 $U[0.05, 0.15]$, $U[0.15, 0.25]$, $U[0.25, 0.35]$로 설정했을 때이다. guess는 $\pi\left(\alpha^*_{ljnx}=0\right)$를 의미하고, slip은 $1-\pi\left(\alpha^*_{ljnx}=1\right)$을 의미한다. 피험자수는 50, 100, 200, 400, 반복횟수는 $R=100$으로 설정하여 진행했다.

2. 모의실험 결과

평가 A, 평가 B, 평가 C 각각의 환경에서 피험자 수, 문항 질을 변화시켜가면서 정확하게 추정한 인지요소 비율(proportion of correctly classified attributes, PCA), 정확하게 추정한 인지요소 패턴 비율(proportion of correctly classified attribute vectors, PCV), 문항별 평균제곱근 오차(root mean square error for item j, $RMSE_j$)를 구하고 그 결과를 분석해보았다.

<표 Ⅳ-3> 평가 A, B, C의 PCA, PCV

피험자 수	문항 질	평가 A		평가 B		평가 C	
		PCA	PCV	PCA	PCV	PCA	PCV
50	저	**0.718**	**0.379**	**0.738**	**0.229**	**0.761**	**0.154**
	중	0.846	0.602	**0.878**	0.517	**0.888**	**0.429**
	고	0.958	0.878	0.964	0.832	**0.968**	**0.798**
100	저	0.717	0.363	0.729	0.210	**0.743**	**0.126**
	중	0.858	0.634	0.878	0.522	0.884	0.421
	고	0.959	0.881	0.967	0.844	0.967	**0.794**
200	저	0.724	0.381	0.739	0.216	0.740	0.121
	중	0.873	0.665	0.887	0.547	0.886	0.425
	고	0.962	0.891	0.972	0.868	0.968	0.794
400	저	0.738	0.402	0.755	0.242	0.750	0.131
	중	0.886	0.696	0.899	0.585	0.895	0.458
	고	0.964	0.897	0.973	0.872	0.972	0.816

<표 Ⅳ-3>은 평가 A, B, C의 PCA, PCV를 나타낸 것이다. 보통은 피험자 수가 많을수록 PCA, PCV 값이 크게 측정되었다. 하지만 <표 Ⅳ-3>의 진하게 표시한 부분은 피험자 수가 많을 때보다 적을 때 PCA, PCV가 크게 측정된 경우를 나타낸 것이다. 피험자 수가 50으로 작고, 문항 질이 좋지 않은 경우에는 항상 그러했고, 평가 A보다는 평가 B, 평가 B보다는 평

가 C에서 그런 경우가 많았다. 평가 세트별로 비교하자면 PCA의 경우, 피험자 수가 100 이하일 때는 평가 A, B, C 순서로 PCA가 작게 측정되는 경향을 보였다. 그런데 피험자 수가 200 이상일 때는 오히려 평가 B가 평가 C보다 PCA가 크게 측정되는 경향을 보였다. 반면 PCV의 경우, 피험자 수와 상관없이 평가 A, B, C 순서로 PCV가 크게 측정되었다.

<표 Ⅳ-4>, <표 Ⅳ-5>, <표 Ⅳ-6>은 각각 평가 A, B, C에서 문항 질이 저, 중, 고 일 때, $RMSE_j$를 나타낸 것이다. 문항 질이 높을수록, 피험자 수가 많을수록 $RMSE_j$가 낮게 측정되었다. 피험자 수가 가장 적고, 문항 질이 가장 낮을 때, 평가 A에서는 문항 6에서 0.246, 평가 B에서는 문항 6에서 0.198, 평가 C에서는 문항 10에서 0.217로 가장 큰 값이 측정되었다.

<표 Ⅳ-4> 평가 A의 $RMSE_j$

문항 번호	문항 질	피험자수			
		50	100	200	400
1	저	0.175	0.137	0.136	0.122
	중	0.130	0.101	0.083	0.060
	고	0.073	0.057	0.042	0.028
2	저	0.175	0.155	0.116	0.088
	중	0.116	0.087	0.061	0.044
	고	0.069	0.049	0.036	0.024
3	저	0.172	0.159	0.121	0.094
	중	0.119	0.092	0.064	0.042
	고	0.074	0.051	0.038	0.025
4	저	0.164	0.146	0.116	0.082
	중	0.130	0.088	0.066	0.040
	고	0.086	0.059	0.042	0.028
5	저	0.153	0.125	0.099	0.080
	중	0.118	0.089	0.066	0.046
	고	0.075	0.057	0.039	0.026
6	저	0.246	0.232	0.194	0.153
	중	0.220	0.157	0.114	0.069
	고	0.154	0.107	0.071	0.049
7	저	0.170	0.140	0.110	0.096
	중	0.133	0.101	0.070	0.046
	고	0.101	0.069	0.047	0.032

8	저	0.156	0.128	0.107	0.076
	중	0.119	0.089	0.063	0.042
	고	0.082	0.059	0.039	0.028

<표 Ⅳ-5> 평가 B의 $RMSE_j$

문항 번호	문항 질	피험자수			
		50	100	200	400
1	저	0.166	0.138	0.120	0.103
	중	0.128	0.100	0.075	0.050
	고	0.072	0.054	0.037	0.026
2	저	0.137	0.101	0.086	0.059
	중	0.099	0.076	0.047	0.034
	고	0.068	0.043	0.030	0.021
3	저	0.144	0.114	0.076	0.058
	중	0.097	0.071	0.048	0.033
	고	0.069	0.042	0.033	0.022
4	저	0.145	0.113	0.106	0.084
	중	0.111	0.086	0.067	0.041
	고	0.071	0.047	0.034	0.024
5	저	0.153	0.139	0.112	0.111
	중	0.128	0.094	0.079	0.049
	고	0.072	0.056	0.038	0.028
6	저	0.198	0.163	0.142	0.103
	중	0.151	0.116	0.080	0.051
	고	0.113	0.076	0.051	0.038
7	저	0.121	0.103	0.085	0.060
	중	0.104	0.072	0.052	0.035
	고	0.068	0.049	0.034	0.025
8	저	0.189	0.153	0.122	0.081
	중	0.135	0.100	0.071	0.048
	고	0.118	0.078	0.056	0.039
9	저	0.142	0.119	0.100	0.072
	중	0.120	0.098	0.063	0.042
	고	0.102	0.063	0.046	0.031
10	저	0.173	0.149	0.121	0.095
	중	0.150	0.104	0.077	0.048
	고	0.116	0.078	0.047	0.036

11	저	0.122	0.097	0.073	0.052
	중	0.090	0.069	0.045	0.032
	고	0.071	0.046	0.033	0.023
12	저	0.168	0.140	0.108	0.075
	중	0.131	0.093	0.070	0.046
	고	0.113	0.078	0.054	0.036
13	저	0.125	0.105	0.086	0.059
	중	0.108	0.077	0.054	0.039
	고	0.080	0.057	0.038	0.028
14	저	0.139	0.118	0.096	0.070
	중	0.108	0.082	0.058	0.039
	고	0.091	0.059	0.043	0.029
15	저	0.141	0.119	0.100	0.074
	중	0.125	0.096	0.065	0.044
	고	0.118	0.080	0.053	0.037

<표 Ⅳ-6> 평가 C의 $RMSE_j$

문항 번호	문항 질	피험자수			
		50	100	200	400
1	저	0.117	0.093	0.081	0.062
	중	0.096	0.081	0.056	0.036
	고	0.060	0.044	0.032	0.023
2	저	0.111	0.103	0.077	0.058
	중	0.091	0.071	0.047	0.035
	고	0.061	0.043	0.033	0.021
3	저	0.127	0.101	0.081	0.057
	중	0.095	0.067	0.049	0.037
	고	0.059	0.047	0.032	0.022
4	저	0.130	0.104	0.089	0.064
	중	0.096	0.082	0.049	0.033
	고	0.067	0.046	0.032	0.022
5	저	0.126	0.093	0.096	0.070
	중	0.102	0.080	0.062	0.045
	고	0.069	0.055	0.036	0.025
6	저	0.150	0.125	0.105	0.106
	중	0.123	0.097	0.080	0.060
	고	0.072	0.056	0.046	0.030

7	저	0.157	0.150	0.124	0.111
	중	0.113	0.093	0.081	0.053
	고	0.070	0.057	0.042	0.037
8	저	0.167	0.125	0.108	0.079
	중	0.135	0.094	0.069	0.046
	고	0.101	0.074	0.051	0.036
9	저	0.113	0.091	0.072	0.049
	중	0.094	0.068	0.050	0.033
	고	0.070	0.048	0.034	0.024
10	저	0.217	0.190	0.169	0.144
	중	0.184	0.142	0.113	0.081
	고	0.152	0.105	0.074	0.051
11	저	0.126	0.105	0.085	0.066
	중	0.112	0.081	0.060	0.039
	고	0.097	0.064	0.046	0.031
12	저	0.179	0.142	0.120	0.092
	중	0.146	0.109	0.080	0.057
	고	0.119	0.086	0.058	0.042
13	저	0.108	0.086	0.069	0.054
	중	0.096	0.070	0.045	0.032
	고	0.072	0.045	0.033	0.024
14	저	0.142	0.112	0.090	0.069
	중	0.118	0.090	0.064	0.045
	고	0.100	0.070	0.051	0.035
15	저	0.130	0.103	0.079	0.062
	중	0.106	0.071	0.055	0.038
	고	0.086	0.060	0.044	0.030
16	저	0.121	0.096	0.079	0.055
	중	0.105	0.074	0.054	0.038
	고	0.082	0.059	0.038	0.028
17	저	0.141	0.121	0.102	0.078
	중	0.128	0.097	0.068	0.046
	고	0.116	0.078	0.054	0.037
18	저	0.113	0.093	0.074	0.057
	중	0.100	0.072	0.054	0.038
	고	0.081	0.058	0.042	0.030
19	저	0.114	0.087	0.070	0.049
	중	0.096	0.069	0.049	0.033
	고	0.081	0.054	0.037	0.026

20	저	0.102	0.080	0.064	0.048
	중	0.090	0.066	0.046	0.031
	고	0.083	0.056	0.036	0.027
21	저	0.136	0.114	0.095	0.071
	중	0.115	0.089	0.064	0.045
	고	0.100	0.071	0.050	0.033
22	저	0.137	0.120	0.099	0.082
	중	0.125	0.093	0.068	0.050
	고	0.107	0.074	0.052	0.035

Ⅴ. 수학 서술형 평가에 대한 DTM 인지진단의 적용

1. 연구방법

1.1. 연구 설계

1.1.1. 연구 수행 절차

본 연구에서는 다양한 해법의 반응을 허용하는 수학 서술형 문항 세트를 활용하여 고등학교 학생들을 대상으로 시험을 시행하고, 그 반응 데이터를 DTM 모형으로 분석하는 인지진단평가를 진행하고자 한다. 연구 절차는 [그림 Ⅴ-1]과 같다.

인지요소 선정
⇩
문항 및 평가지 구성
⇩
피험자 집단 선정
⇩
피험자 모집 및 동의
⇩
검사 시행
⇩
피험자 반응 분류
⇩
결과 분석

[그림 Ⅴ-1] 연구 수행 절차

1.1.2. 평가를 위한 인지요소 선정

본 연구에서는 다양한 해법의 반응이 존재하고, 여러 단계를 거쳐야 해결할 수 있고, 해결 과정에서 다양한 종류의 개념 · 원리 · 기능이 요구되는 수학 서술형 문항을 사용하여 DTM 모형을 활용한 인지진단평가를 시행하고자 한다. 따라서 인지요소 선정 및 평가 문항 구성을 위한 기준은 다음과 같다. 첫째, Ma(2019)에서 사용된 TIMSS 2007 4학년, 8학년 수준의 수학 문항에 비해 풀기 어려운 문제여야 한다. 둘째, 우리나라 고등학교 교육과정에 포함되거나 관련된 개념, 원리를 사용하여 인지요소를 구성해야 한다. 셋째, 서술형으로 묻기 적절한 주제 혹은 문항이어야 한다. 이러한 기준을 종합하여 선택한 주제는 절대부등식(absolute inequality)의 증명이다.

부등식(inequality)이란 수 또는 식 사이의 대소 관계를 부등호로 나타낸 것이다. 부등식에 관한 문제 중에는 특정 조건의 변수로 이루어진 부등식이 왜 성립하는지 증명하는 것이 있다. 예를 들어 A가 B보다 큰 이유를 밝히는 문제라면, 그 이유를 직접 밝히거나 A보다 작고 B보다 큰 대상 C를 제시하고 수학적 개념이나 원리를 사용하여 이유를 설명해야 한다. 때때로 A와 B를 매개하는 대상이 두 개 이상일 수 있으며, 부등식의 성질, 곱셈공식, 절대부등식, 치환(substitution) 등 여러 개념 · 원리 · 기능이 요구되기도 한다. 절대부등식은 “부등식의 문자에 그 문자가 가질 수 있는 어떤 실숫값을 대입해도 항상 성립하는 부등식”을 의미한다(박교식 외, 2020, p. 198). 2015 개정 수학과 교육과정(교육부, 2015, pp. 51-52)에 의하면, 선택 중심 교육과정(공통과목) <수학>의 ‘수와 연산’ 영역에는 “절대부등식의 의미를 이해하고, 간단한 절대부등식을 증명할 수 있다.”라는 성취기준이 명시되어 있으며, ‘평가 방법 및 유의사항’에는 “증명 능력을 평가할 때에는 과정 중심 평가를 할 수 있다.”고 명시되어 있다. 그리고 2015 개정 수학과 교육과정이 반영된 대부분의 교과서에서 절대부등식을 증명하는 문제로 산술-기하 평균 부등식과 코시-슈바르츠 부등식을 다룬다. 이상의 논의를 종합하여 산술-기하 평균 부등식과 코시-슈바르츠 부등식에 관한 절대부등식의 증명이 위에서 제시한 세 가지 기준을 만족한다는 점에서 검사 영역으로 설정되었다.

<표 Ⅴ-1> 인지진단평가에 사용된 인지요소

기호	인지요소
A1	$a \geq b > 0$, $m > 0$인 실수 a, b, m에 대해 $a^m \geq b^m$이 성립함을 알고, 이를 이용하여 주어진 부등식과 동치인 다른 부등식을 도출할 수 있다.
A2	산술-기하 평균 부등식을 이용하여 문제해결에 도움이 되는 부등식을 얻을 수 있다.
A3	산술-기하 평균 부등식의 등호 성립 조건을 이용하여 주어진 부등식의 등호 성립 조건을 구할 수 있다.
A4	코시-슈바르츠 부등식을 이용해서 문제해결에 도움이 되는 부등식을 얻을 수 있다.
A5	코시-슈바르츠 부등식의 등호 성립 조건을 이용하여 주어진 부등식의 등호 성립 조건을 구할 수 있다.
A6	치환(substitution)을 이용하여 주어진 부등식과 동치인 다른 부등식을 도출할 수 있다.

본 연구의 인지진단평가에서 사용된 6개의 인지요소는 <표 Ⅴ-1>과 같다. A1은 다양한 절대부등식을 해결하기 위해 기본이 되는 성질에 관한 인지요소이다. A2와 A3은 산술-기하 평균 부등식에 관한 인지요소, A4와 A5는 코시-슈바르츠 부등식에 관한 인지요소이다. 피험자 집단으로부터 다양한 해법의 반응이 나올 수 있는 문항을 제작하고자 치환에 관한 인지요소 A6을 추가하였다.

1.1.3. 문항 및 평가지 구성

본 고의 저자가 직접 제작한 문항(1번부터 7번까지)과 Manfrino et al.(2009)에 수록된 문항(8번부터 13번)으로 [그림 Ⅴ-2]와 같이 13개의 문항으로 구성된 평가지를 제작했다.

1. $a \geq b > 0$, $m > 0$인 실수 a, b, m에 대하여 a^m과 b^m의 대소 관계를 부등식으로 나타내시오.

2. 산술-기하 평균 부등식을 이용하여, 양의 실수 x, y에 대하여 $x^2 + y^2 = 4$일 때 다음 부등식이 성립함을 보이시오.

$$xy \leq 2$$

3. 위 2번에서 $xy = 2$일 때, x와 y의 값을 구하시오.

4. 산술-기하 평균 부등식을 이용하여, 양의 실수 x에 대하여 $\left(x + \frac{2}{x}\right)\left(x + \frac{8}{x}\right)$의 최솟값을 구하고, 그 과정을 서술하시오.

5. 코시-슈바르츠 부등식을 이용하여, 양의 실수 x, y에 대하여 $x^2 + y^2 = 4$일 때 다음 부등식이 성립함을 보이시오.

$$4x + 3y \leq 10$$

6. 위 5번에서 $4x + 3y = 10$일 때, x와 y의 값을 구하시오.

7. 양의 실수 x, y에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\sqrt{x+y} \geq \frac{\sqrt{2x} + \sqrt{2y}}{2}$$

8. 실수 a, b, c, x, y, z에 대하여 x, y, $z > 0$ 일 때, 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\frac{a^2}{x} + \frac{b^2}{y} + \frac{c^2}{z} \geq \frac{(a+b+c)^2}{x+y+z}$$

9. 삼각형의 세 변의 길이가 각각 a, b, c일 때, 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$(b+c-a)(c+a-b)(a+b-c) \leq abc$$

10. 삼각형의 세 변의 길이가 각각 a, b, c일 때, 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\sqrt{a+b-c} + \sqrt{b+c-a} + \sqrt{c+a-b} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$$

11. 양의 실수 a, b, c에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\frac{a}{b}+\frac{b}{c}+\frac{c}{a} \geq \frac{c+a}{c+b}+\frac{a+b}{a+c}+\frac{b+c}{b+a}$$

12. 양의 실수 a, b, c에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\frac{a^3}{a^3+2b^3}+\frac{b^3}{b^3+2c^3}+\frac{c^3}{c^3+2a^3} \geq 1$$

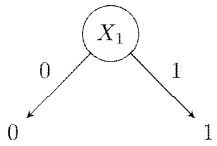
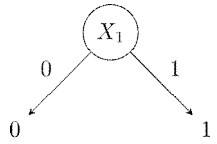
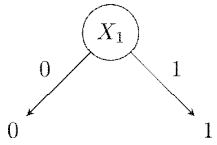
13. 양의 실수 a, b, c에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

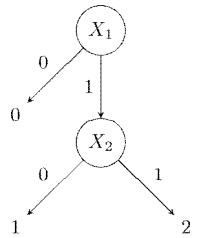
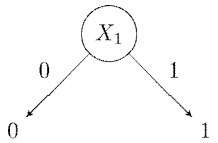
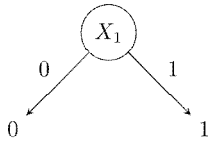
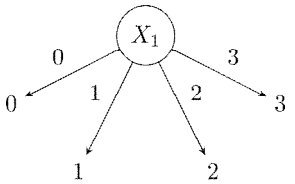
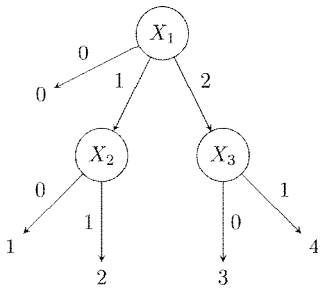
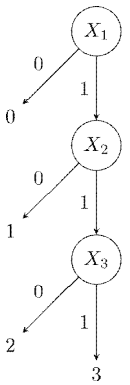
$$\frac{1}{3}\left(\frac{bc}{a}+\frac{ca}{b}+\frac{ab}{c}\right) \geq \sqrt{\frac{a^2+b^2+c^2}{3}}$$

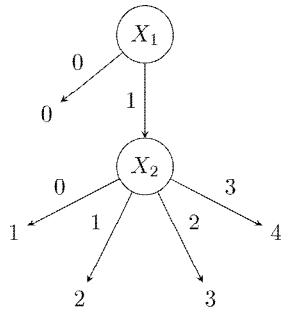
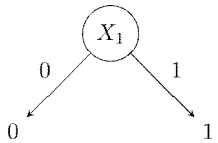
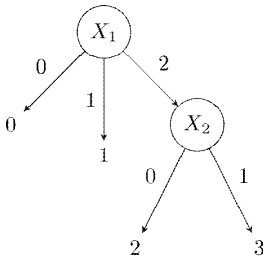
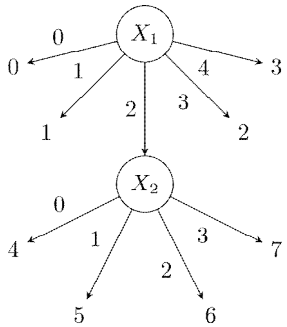
[그림 Ⅴ-2] 인지진단평가에 사용된 문항 세트

<표 Ⅴ-2>는 각 문항에 대한 수형도 구조, Q행렬, 풀이 과정 분석틀의 인지요소 개수, 단계 길이, 반응 개수를 나타낸 것이다.

<표 Ⅴ-2> 실험에 사용된 문항의 수형도 구조, Q행렬, 분석틀 분석

문항 번호	수형도 구조	Q행렬	인지 요소	단계	반응
1		$Q_c(1)=[1\,1\,1\,0\,0\,0\,0\,0]$	1	1	2
2		$Q_c(2)=[2\,1\,0\,1\,0\,0\,0\,0]$	1	1	2
3		$Q_c(3)=[3\,1\,0\,0\,1\,0\,0\,0]$	1	1	2

4		$Q_c(4) = \begin{bmatrix} 4\,1\,0\,1\,0\,0\,0\,0 \\ 4\,2\,0\,0\,1\,0\,0\,0 \end{bmatrix}$	2	2	3
5		$Q_c(5) = [5\,1\,0\,0\,0\,1\,0\,0]$	1	1	2
6		$Q_c(6) = [6\,1\,0\,0\,0\,0\,1\,0]$	1	1	2
7		$Q_c(7) = \begin{bmatrix} 7\,1\,1\,0\,0\,0\,0\,0 \\ 7\,1\,1\,1\,1\,0\,0\,0 \\ 7\,1\,1\,0\,0\,1\,1\,0 \end{bmatrix}$	5	1	4
8		$Q_c(8) = \begin{bmatrix} 8\,1\,0\,1\,0\,0\,0\,0 \\ 8\,1\,0\,0\,0\,1\,0\,0 \\ 8\,2\,0\,0\,1\,0\,0\,0 \\ 8\,3\,0\,0\,0\,0\,1\,0 \end{bmatrix}$	4	2	5
9		$Q_c(9) = \begin{bmatrix} 9\,1\,0\,0\,0\,0\,0\,1 \\ 9\,2\,0\,1\,0\,0\,0\,0 \\ 9\,3\,0\,0\,1\,0\,0\,0 \end{bmatrix}$	3	3	4

10		$Q_c(10) = \begin{bmatrix} 10\ 1\ 0\ 0\ 0\ 0\ 0\ 1 \\ 10\ 2\ 1\ 0\ 0\ 0\ 0\ 0 \\ 10\ 2\ 1\ 1\ 1\ 0\ 0\ 0 \\ 10\ 2\ 1\ 0\ 0\ 1\ 1\ 0 \end{bmatrix}$	6	2	5
11		$Q_c(11) = [11\ 1\ 0\ 0\ 0\ 0\ 0\ 1]$	1	1	2
12		$Q_c(12) = \begin{bmatrix} 12\ 1\ 0\ 0\ 0\ 1\ 1\ 0 \\ 12\ 1\ 0\ 0\ 0\ 0\ 0\ 1 \\ 12\ 2\ 0\ 1\ 1\ 0\ 0\ 0 \end{bmatrix}$	5	3/2	4
13		$Q_c(13) = \begin{bmatrix} 13\ 1\ 1\ 0\ 0\ 0\ 0\ 0 \\ 13\ 1\ 0\ 0\ 0\ 0\ 0\ 1 \\ 13\ 1\ 1\ 1\ 1\ 0\ 0\ 0 \\ 13\ 1\ 1\ 0\ 0\ 1\ 1\ 0 \\ 13\ 2\ 1\ 0\ 0\ 0\ 0\ 0 \\ 13\ 2\ 1\ 1\ 1\ 0\ 0\ 0 \\ 13\ 2\ 1\ 0\ 0\ 1\ 1\ 0 \end{bmatrix}$	6	3/2	8

1.1.4. 피험자 집단 선정

연구 참여 모집 대상은 과학영재학교인 A고등학교의 1, 2, 3학년 학생이다. 본 연구는 DTM 모형이 풀이의 단계가 많고 사용되는 전략이 다양한 수학 서술형 문항에 대해서도 학생들의 인지요소 숙달 패턴에 대한 진단에 실제적인 효과를 보이는지 확인하고자 설계되었다. 과학영재고등학교

학생들은 절대부등식의 증명에 관한 문항에 대해 다양한 전략을 사용할 가능성이 크고, 풀이의 단계를 분명하게 서술하는 경향이 높기 때문에, 수형도 구조의 복잡도가 높은 수학 서술형 평가의 효과를 확인하기 위한 피험자로 적절한 것으로 판단하여 목표 연구 참여자로 선정하였다. 또한, 모든 학년의 학생들을 연구에 참여하게 함으로써 학생들의 수준에 따라 인지요소 숙달 패턴이 다양하게 관찰될 수 있게 하였다.

1.1.5. 피험자 모집 및 동의

연구책임자가 정교사로 재직 중인 A 고등학교의 1, 2, 3학년 전교생 379명(1학년 128명, 2학년 124명, 3학년 127명)을 대상으로 연구 참여자를 모집했다. 학교 홈페이지, 학교 포털시스템의 공지사항 게시판에 모집문건을 게시하고. 모집문건을 통해 연구 참여를 홍보했다. 참여를 원하는 학생은 모집문건에 제시된 구글 설문지(Google Forms) 링크에 접속하여 참여 여부를 입력하고, 연구책임자는 응답 결과를 확인하여 연구 참여 희망자를 파악했다.

학교의 종례 시간을 이용하여 연구 참여 희망자들에게 연구 내용과 연구 참여자의 권리에 관해 설명했다. 연구 참여자와 법정대리인은 충분히 숙고할 시간을 가진 후, 연구책임자에게 동의서를 제출하도록 했다. 연구 참여자와 법정대리인이 모두 승낙한 학생들만 연구를 진행했다. 학생들의 인권과 안전을 보장하고 연구 참여 윤리를 준수하기 위하여 서울대학교 생명윤리위원회(Institutional Review Board)의 승인(승인번호 2203/001-005)을 받아 진행되었다.

1.1.6. 실험 수행 과정

시험은 100석 이상의 책상이 설치된 대강당에서 진행되었다. 연구 참여자마다 참여할 수 있는 시간이 서로 다를 것으로 예상됐다. 따라서 동일 시험을 4회 진행했으며 이를 사전에 계획하고 공지했다. 연구 참여자들은 방과 후, 자신이 원하는 시간에 시험 장소에 집결하여 5분 이내로 시험 유의 사항

에 대해 안내받고, 60분 이내로 제시된 서술형 수학 문항을 해결하고 풀이 과정을 답안지에 서술하여 연구책임자에게 제출했다. 아직 시험에 응시하지 않은 연구 참여자가 시험 문항에 관한 정보를 알 수 없도록, 시험에 응시한 연구 참여자에게 비밀 유지를 당부했다. 본 연구는 평가 결과를 바탕으로 학생들의 우열 관계나 합격-불합격 여부를 가리는 것이 목적이 아니며, 학생들의 다양한 풀이 답안을 수집하는 것이 목적이다. 이와 같은 연구의 취지를 연구 참여자에게 설명했으므로 학생의 대부분이 비밀 유지를 지켰으리라 예상된다. 문항에 대한 정보가 유출됐다고 하여도 연구 참여자마다 푸는 방법은 다를 수 있으므로 다양한 풀이 답안을 수집하는 본 연구의 목적에 위배되지 않는다. 연구수행과정에서 평가지 관리, 유의 사항 안내, 평가 감독, 평가지 회수 등 일련의 절차는 연구책임자가 혼자서 수행했다.

결과적으로 1학년 20명, 2학년 29명, 3학년 14명으로 총 63명이 연구에 참여했다. 학년 구성은 <표 Ⅴ-3>과 같다.

<표 Ⅴ-3> 연구 참여자 수와 학년별 분포

	1 학년	2 학년	3 학년	합계
인원	20	29	14	63
비율	0.32	0.46	0.22	1

1.1.7. 문항에 대한 학생 반응의 분류

피험자 개개인의 인지요소 추정을 위해서 제출된 모든 답안지는 미리 생성해놓은 문항 수형도 구조에 의해 반응이 분류되어야 한다. 세 명의 수학교육 전문가가 독립적으로 학생의 반응을 분류하였다. 의견이 불일치한 경우 협의를 통해서 신뢰도를 높였다. 연구책임자가 반응 분류 및 검토 결과를 종합하여 최종 반응 분류 데이터를 생성했다. 연구 참여자의 개인정보는 모두 가리고 진행했다. 이는 학생에 대한 사전정보가 학생 반응 분류에 영향을 주지 않도록 함과 동시에 연구 참여자의 개인정보를 보호하기 위함이다.

각 문항에 대한 피험자들의 풀이 과정과 반응은 수형도 구조로 표현되

며 문항을 개발하는 단계에서 학생들의 예상 반응을 추측하여 수형도 구조를 구성했다. 그런데 실제로 학생 반응 분류를 시행하였을때 예상했던 어떤 반응은 나타나지 않았으며, 예상하지 않았던 반응이 나타나기도 했다. 이런 경우, 기존의 수형도 구조를 수정하여 재확립한 후 반응 분류를 확정하였다. 반응 분류기준은 [부록 1]에 제시했다.

1.2. 연구 설계를 위한 모의실험

1.2.1. 실험 설계

학생 인지진단평가를 위한 문항 세트를 활용하여 모의실험을 진행하는 목적은 다음과 같다. 첫째, DTM 모형을 통해 피험자의 인지요소를 추정하는 데 제작된 문항 세트에 구조적 문제가 없는지 확인해야 한다. 둘째, 학생 대상 인지진단평가 피험자 집단의 인지요소 숙달 패턴 추정 결과를 어느 정도 정확도로 추정됐다고 볼 수 있는지에 대한 판단기준이 필요하다. 또한 본 연구의 학생 인지진단평가의 조건에서 피험자 수가 변한다면 추정 정확도가 어떻게 변할지 확인할 필요가 있다.

피험자 수, 문항 질을 변화시켜가면서 정확하게 추정한 인지요소 비율(proportion of correctly classified attributes, PCA), 정확하게 추정한 인지요소 패턴 비율(proportion of correctly classified attribute vectors, PCV), 문항별 평균 제곱근 오차(root mean square error for item j, RMSE_j)를 구하고 그 결과를 분석한다. 문항 질은 고, 중, 저로 구분했는데, 각각은 추측 모수인 guess와 실수 모수인 slip을 각각 $U[0.05, 0.15]$, $U[0.15, 0.25]$, $U[0.25, 0.35]$로 설정했을 때이다. 학생 대상 인지진단평가에서 63명의 피험자가 참여한 것을 고려하여 모의실험의 피험자 수는 30, 60, 90, 200으로 설정했다. 반복횟수는 $R=100$으로 설정하여 진행했다.

1.2.2. 분석 방법

피험자 수, 문항 질을 변화시켜가면서 정확하게 추정한 인지요소 비율

(PCA), 정확하게 추정한 인지요소 패턴 비율(PCV), 문항별 평균 제곱근 오차($RMSE_j$)를 구하고 그 결과를 분석한다.

1.3. 학생 데이터 분석 방법

1.3.1. DTM 모형을 통한 분석

평가의 각 문항에서 피험자들이 보인 반응은 사전에 설정된 반응 분류기준에 의해 분류되며, DTM 모형을 활용하여 피험자별 인지요소 숙달 패턴이 추정된다. 이를 통해 인지요소별 숙달 피험자 비율, 추정된 인지요소 숙달 패턴 분포 등을 구할 수 있다. 모형의 모수 추정에는 MMLE/EM(Marginal Maximum Likelihood Estimation using the Expectation-Maximization) 알고리즘이 사용되었다(Bock & Aikin, 1981). 학생들의 개별 인지요소 숙달 패턴을 추정할 때는 인지요소의 숙달 유무뿐 아니라 숙달 확률도 같이 수치화할 수 있는 EAP 방법을 사용하였고, 숙달 확률이 0.5 이상인 경우 인지요소를 숙달하고 있는 것으로 간주하였다(Huebner & Wang, 2011).

문항에 대한 반응 모형으로는 DINA, DINO, ACDM, G-DINA의 네 가지 모형을 고려하였고, 모형 적합도를 비교하여 가장 적합한 모형을 선택하고자 하였다. 네 모형을 비교하여 선택할 때는 상대적합도와 절대적합도의 두 가지를 모두 고려하였다. 상대적합도의 경우 AIC(Akaiki's Information Criterion)와 BIC(Bayesian Information Criterion)를 이용하였고, 절대적합도의 경우 모수적 부트스트랩(parametric bootstrap)을 시행하여 부트스트랩 χ^2과 부트스트랩 $CR(2/3)$의 값을 비교하였다. 대표적인 절대적합도인 χ^2과 $CR(2/3)$지수의 경우 데이터가 희소한(sparse) 형태를 띠고 있을 경우 적합하지 않기 때문에, 문항이 아닌 노드별 반응 범주가 기록되는 DTM 구조에 적용하기에는 어려움이 있다(Ma, 2019). 따라서 모수적 부트스트랩 방법을 이용하였고, 총 1,000번의 부트스트랩 샘플을 수집하여 절대적합도를 계산하였다.

인지진단모형에서 인지 패턴 추정을 위한 식별성(identifiability)을 보장하기 위해서는 인지요소를 하나만을 측정하는 문항이 각 인지요소마다 존재해야 하고, 해당 문항의 guess와 slip의 합이 1보다 작아야 한다(Xu & Zhang, 2016). 따라서 식별성을 보장하고 문항으로부터 유의미한 정보를 추출하기 위하여, 세 문항에 대해서는 guess와 slip의 합이 0.8을 넘지 않도록 제약조건을 추가하여 추정을 진행하였다. 또한, 모든 인지요소 k에 대하여 $\alpha_{lk} \leq \alpha_{l'k}$이면 $P(X_j = 1|\boldsymbol{\alpha}_l) \leq P(X_j = 1|\boldsymbol{\alpha}_{l'})$을 만족하도록 하는 단조성 가정(monotonicity assumption)을 추가하여 문항 모수를 추정하였다.

문항 모수를 추정한 후에는, 문항의 수형도 구조, Q행렬, 문항 모수 추정값을 이용하여 문항 분석을 진행하였다. DTM 모형에서는 문항을 해결하는 여러 전략이 수형도의 경로를 통해 표현되며, 경로에 포함된 노드를 기준으로 각 전략을 단계별로 분석할 수 있다. 또한, Q행렬에서 전략의 각 단계를 해결하기 위해 필요한 인지요소를 확인할 수 있으며, 추정된 문항 모수를 이용하여 인지요소의 숙달 여부에 따른 각 단계의 확률적인 특징을 알 수 있다. 이와 같은 정보를 활용하여 문항 해결전략을 단계별로 나누어 분석하고 서로 다른 전략들을 비교하였다.

데이터 처리, 문항함수의 모수와 인지요소 추정값 계산을 위해 R 프로그램을 사용하였다. DTM 모형을 적용하기 위해 R 패키지 GDINA(Ma & de la Torre, 2020)에서 제공되는 함수를 다양한 반응 모형 및 문항 모수에 대한 조건을 고려할 수 있도록 함수 내부 코드를 수정하여 사용하였다.

1.3.2. 학생 피드백 방법

평가를 마친 피험자들에게 DTM 모형을 통해 추정된 학생의 인지요소 숙달 패턴 이외에도 추가적인 정보를 제공하여 피험자의 학습을 도울 수 있다. 수학 서술형 문항의 풀이 과정 분석틀에 기반하여 피험자의 풀이 과정을 분석할 수 있으며, 문항의 단계별 인지요소와 단계별 응답 비율, 문항별 피험자의 반응 · 피험자가 답할 가능성이 가장 큰 정답 반응 · 추가 인지요소 습득을 통한 정답 반응의 비교를 제시하여 피험자의 학습경로를 추천할 수 있다.

2. 연구결과

2.1. 모의실험 결과

피험자 수, 문항 질을 변화시켜가면서 정확하게 추정한 인지요소 비율(proportion of correctly classified attributes, PCA), 정확하게 추정한 인지요소 패턴 비율(proportion of correctly classified attribute vectors, PCV), 문항별 평균 제곱근 오차(root mean square error for item j, $RMSE_j$)를 구하고 그 결과를 분석해보았다.

<표Ⅴ-4>는 학생 인지진단평가 문항 세트에 대한 PCA, PCV를 나타낸 것이다. 문항 질이 고정됐을 때 피험자 수에 따른 인지요소 추정 정확도의 차이는 크게 드러나지 않았다. PCA의 경우, 문항 질이 낮을 때는 약 0.7, 중간일 때는 약 0.83, 높을 때는 약 0.93으로 측정되었다. PCV의 경우, 문항 질이 낮을 때는 약 0.11 이상 0.15 이하, 중간일 때는 약 0.31 이상 0.35 이하, 높을 때는 약 0.66 이상 0.68 이하로 측정되었다.

<표 Ⅴ-4> 학생 인지진단평가 문항 세트에 대한 PCA, PCV

피험자 수	문항 질	PCA	PCV
30	저	0.724	0.150
	중	0.835	0.347
	고	0.937	0.679
60	저	0.710	0.132
	중	0.827	0.315
	고	0.932	0.661
90	저	0.704	0.120
	중	0.824	0.309
	고	0.933	0.664
200	저	0.698	0.115
	중	0.827	0.318
	고	0.935	0.669

<표 Ⅴ-5>, <표 Ⅴ-6>, <표 Ⅴ-7>은 문항 질이 각각 저, 중, 고 일 때, 문항별 평균 제곱근 오차($RMSE_j$)를 구한 것이다. 문항 질이 높을수록, 피험자 수가 많을수록 $RMSE_j$가 낮게 측정되었다. 문항 질이 낮고, 피험자 수가 적을 때 문항 7에서 약 0.266으로 가장 큰 값이 측정되었다.

<표 Ⅴ-5> 문항 질이 낮을 때 $RMSE_j$

문항 번호	피험자수			
	30	60	90	200
1	0.168	0.138	0.121	0.102
2	0.171	0.145	0.133	0.127
3	0.186	0.142	0.136	0.111
4	0.163	0.138	0.120	0.094
5	0.203	0.181	0.170	0.161
6	0.212	0.207	0.187	0.183
7	0.266	0.229	0.213	0.190
8	0.149	0.121	0.115	0.092
9	0.152	0.124	0.109	0.082
10	0.169	0.141	0.135	0.109
11	0.173	0.137	0.144	0.117
12	0.218	0.180	0.169	0.145
13	0.214	0.183	0.163	0.142

<표 Ⅴ-6> 문항 질이 중간일 때 $RMSE_j$

문항 번호	피험자 수			
	30	60	90	200
1	0.142	0.103	0.104	0.087
2	0.133	0.095	0.087	0.065
3	0.138	0.108	0.098	0.089
4	0.123	0.104	0.087	0.065
5	0.150	0.135	0.131	0.097
6	0.153	0.138	0.133	0.113
7	0.239	0.203	0.183	0.144
8	0.131	0.113	0.090	0.072
9	0.114	0.091	0.077	0.055

10	0.165	0.131	0.115	0.082
11	0.129	0.095	0.092	0.053
12	0.200	0.152	0.132	0.101
13	0.202	0.169	0.154	0.120

<표 Ⅴ-7> 문항 질이 높을 때 $RMSE_j$

문항 번호	피험자 수			
	30	60	90	200
1	0.094	0.075	0.066	0.057
2	0.088	0.064	0.053	0.036
3	0.091	0.080	0.061	0.049
4	0.086	0.065	0.062	0.041
5	0.091	0.075	0.071	0.059
6	0.098	0.076	0.076	0.064
7	0.217	0.163	0.130	0.091
8	0.103	0.077	0.065	0.052
9	0.101	0.072	0.052	0.037
10	0.153	0.113	0.097	0.061
11	0.089	0.058	0.047	0.032
12	0.169	0.122	0.096	0.070
13	0.186	0.147	0.122	0.083

2.2. DTM 모형을 통한 결과 분석

<표 Ⅴ-8>은 문항별 반응분포를 나타낸 것이다. 1, 2, 3, 5, 6은 정답률이 매우 높았으며, 문항 1, 2, 3은 전원이 옳은 반응을 보였다. 인지진단평가에서 인지요소 숙달 패턴 추정을 위한 식별성(identifiability)을 보장하기 위해서는 인지요소 하나만을 측정하는 문항이 각 인지요소마다 존재해야 하는데, 문항 1, 2, 3, 5, 6은 각각 인지요소 A1, A2, A3, A4, A5의 숙달 여부를 확인할 수 있으며, 문항 11은 인지요소 A6의 숙달 여부를 확인할 수 있다.

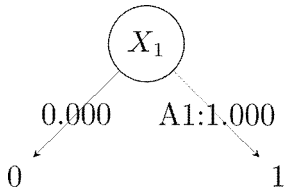
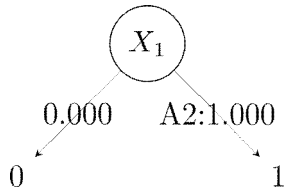
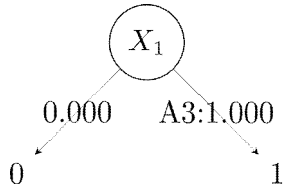
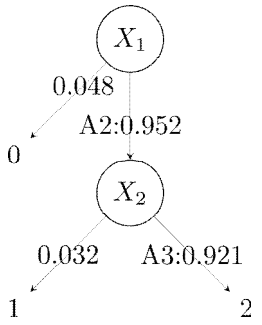
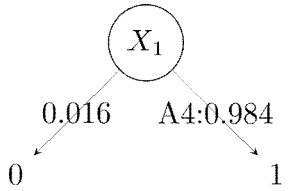
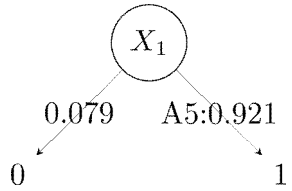
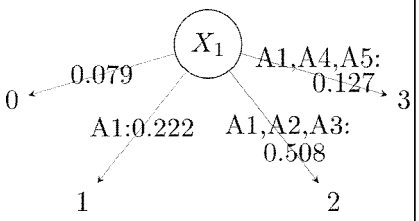
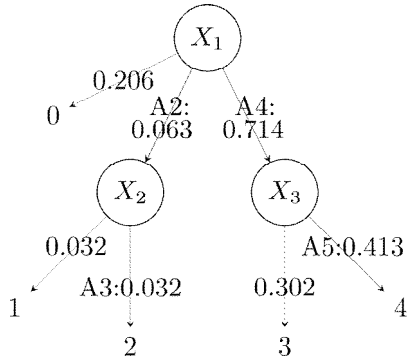
평가를 시행하면 평가 설계 과정에서 예측하지 못한 피험자의 반응을 얻을 수 있는데, 주어진 문제를 해결했거나 문제해결 과정에서 의미있는 반응을 보였으

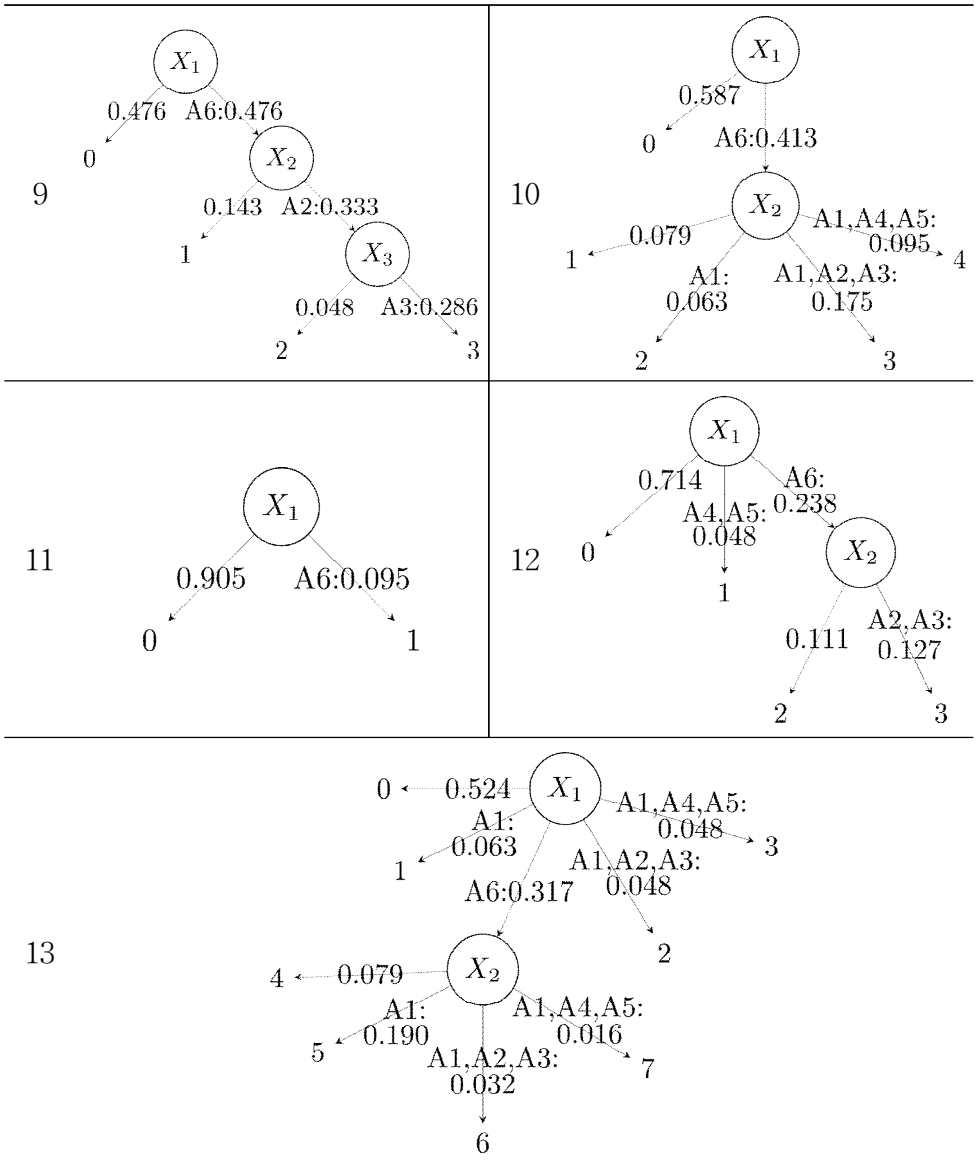
나 평가에서 계획된 인지요소를 사용하지 않은 경우가 그렇다. 이 경우는 분류 기준(0, 1, 2, 3, 4, 5, 6, 7)에 의해 반응을 나눌 수 없어서 결측치로 분류했다. R 프로그래밍 언어의 GDINA package에서 DTM 모형으로 피험자의 인지요소 숙달 여부를 추정하는 함수에는 결측치를 고려하는 옵션이 없다(Ma et al., 2022). 그래서 결측치를 0의 반응으로 분류하여 추정을 진행했다.

<표 Ⅴ-8> 문항별 반응분포

반응	문항												
	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	0	3	1	5	5	13	30	37	56	45	33
1	63	63	63	2	62	58	14	2	9	5	6	3	4
2				58			2	2	3	4		7	3
3							38	19	18	11		8	3
4								26		6			5
5													12
6													2
7													1
결측	0	0	0	0	0	0	4	1	3	0	1	0	0

[그림 Ⅴ-3]은 <표 Ⅴ-8>의 문항별 반응분포와 Q행렬의 정보를 활용하여 문항의 단계별 인지요소 및 응답 비율을 도식화한 것이다. [그림 Ⅴ-3]에 나타난 단계별 인지요소에 관한 도식은 <표 Ⅴ-2>의 수형도 구조와 Q행렬을 활용하여 생성할 수 있다. 수형도 구조에 관한 행렬 표현이나 Q행렬에 관한 배경지식이 없는 사람도 문항 세트의 구성을 한 눈에 확인할 수 있다는 장점이 있다. 평가 설계자는 [그림 Ⅴ-3]을 통해서 평가의 준비단계에서 문항 세트의 구성을 확인할 수 있고, 평가 결과해석 단계에서 피험자들의 문항별 반응분포를 문항의 구조와 함께 확인할 수 있다.

문항	수형도 구조	문항	수형도 구조
1		2	
3		4	
5		6	
7		8	



[그림 Ⅴ-3] 문항의 단계별 인지요소 및 응답 비율

우선 수집한 데이터에 대해 DINA, DINO, ACDM, G-DINA의 네 가지 반응 모형을 기반으로 한 DTM 모형으로 분석을 진행하였다. 상대적합도를 비교하기 위하여 각 반응 모형에 대한 AIC와 BIC 값을 확인해보았고, 그 결과는 <표 Ⅴ-9>와 같다. AIC와 BIC 모두에 대해 DINA 모형을 기반

으로 한 DTM 모형이 가장 낮은 값을 보였고, 따라서 가장 적합한 모형임을 확인할 수 있다. 이와 같은 결과는, DINA 모형과 같은 비보상적 모형의 경우 특정 인지요소에 대한 숙달이 다른 인지요소에 대한 미숙달을 보상하지 않는다는 특징을 가지고 있고, 따라서 문제를 해결하기 위하여 관련된 모든 인지요소가 필요하다는 점에서 수학 검사에 적합하다는 선행연구와도 일치하는 결과이다(Roussos et al., 2007; Wu, Liang et al., 2020). 다음으로, 절대적합도를 비교하기 위해 부트스트랩 χ^2과 부트스트랩 $CR(2/3)$의 p-value를 계산해 보았고, 그 결과는 <표 Ⅴ-10>과 같다. 유의수준 0.05에 대해 모든 모형이 충분히 높은 유의확률을 보여서 적합함을 확인할 수 있다. 절대적합도와 상대적합도를 모두 고려하여 모형을 선택할 때는, 절대적합도 기준을 만족하는 모형들에 대하여 상대적합도가 가장 좋은 모형을 선택하게 된다(Sen & Bradshaw, 2017). 따라서 DINA 모형이 절대적합도의 기준을 만족하고 상대적합도가 가장 좋다는 점에서 최종적인 모형으로 선택되었다.

<표 Ⅴ-9> DINA, DINO, ACDM, G-DINA 모형에 대한 상대 적합도

모형	-2LL	npar	AIC	BIC
DINA	891.38	127	**1145.38**	**1417.55**
DINO	942.71	127	1196.71	1468.89
ACDM	869.11	145	1159.11	1469.86
G-DINA	864.64	179	1222.61	1606.23

<표 Ⅴ-10> DINA, DINO, ACDM, G-DINA 모형에 대한 부트스트랩 모형 적합도

모형	부트스트랩 χ^2	부트스트랩 $CR(2/3)$
DINA	0.609	0.523
DINO	0.245	0.209
ACDM	0.741	0.703
G-DINA	0.709	0.657

수집된 데이터를 DINA 모형을 기반으로 한 DTM 모형으로 분석한 결과는 다음과 같다. <표 Ⅴ-11>은 각 인지요소를 습득한 피험자의 비율을 나타낸 것이다. A1의 경우 지수법칙과 관련된 간단한 인지요소이고 숙달 비율이 0.905로 대부분의 학생이 숙달하고 있는 것으로 나타났다. A2와 A3은 각각 산술-기하 평균 부등식과 등호성립조건에 대한 인지요소로, A2는 숙달 비율이 0.587이고 A3은 숙달 비율이 0.968로 나왔다. 과반의 학생이 A2를 숙달하고 있긴 하나 A3에 비해서 숙달 비율이 낮게 나왔고, 이는 학생들이 산술-기하 평균 부등식의 등호 성립 조건은 잘 알고 있으나 산술-기하 평균 부등식을 문제에 따라 알맞게 적용하는 것은 상대적으로 어려워한다는 것을 의미한다. A4와 A5는 각각 코시-슈바르츠 부등식과 등호성립조건에 대한 인지요소이고, 숙달 비율은 각각 0.556과 0.762로 나왔다. 산술-기하 평균 부등식과 마찬가지로 코시-슈바르츠 부등식 또한 등호 성립 조건은 잘 알고 있으나 이를 문제에 따라 알맞게 적용하는 것은 상대적으로 어려워함을 알 수 있다. 또한 A2보다 A4의 숙달 비율이 낮고, A3보다 A5의 숙달 비율이 낮다는 점을 통해, 학생들이 산술-기하 평균 부등식보다 코시-슈바르츠 부등식과 관련된 인지요소를 숙달하는 데 상대적으로 어려움이 있음을 알 수 있다. 마지막으로 부등식의 치환과 관련된 A6은 숙달 비율이 0.381로 나왔다. 이는 6개의 인지요소 중 가장 낮은 숙달 비율에 해당하며, 이는 학생들이 부등식의 치환을 가장 어려워함을 의미한다.

<표 Ⅴ-11> 인지요소별 숙달 피험자 비율

인지요소	A1	A2	A3	A4	A5	A6
숙달비율	0.905	0.587	0.968	0.556	0.762	0.381

추정된 인지 패턴의 빈도는 <표 Ⅴ-12>와 같고, 이를 통해 인지요소의 전반적인 숙달 비율뿐 아니라 구체적으로 학생들의 인지 패턴이 어떻게 분포하는지를 확인할 수 있다. <표 Ⅴ-11>의 전반적인 숙달 비율에서 예상할 수 있듯이, 숙달 비율이 상대적으로 높은 A1, A3, A5만을 습득하여

인지 패턴 101010에 해당하는 학생이 22명으로 가장 많았으며, 모든 인지 요소를 습득하고 있는 인지 패턴 111111에 해당하는 학생이 12명으로 두 번째로 많았다. 다음으로, A5와 A6만을 습득하지 못한 인지 패턴 111100에 해당하는 학생이 10명으로 나왔다. 코시-슈바르츠 부등식을 적용하는 A4가 코시-슈바르츠 부등식의 등호 성립 조건과 관련된 A5보다 습득하기 어려운 인지요소이긴 하지만, 반대로 A4를 습득하고 A5는 습득하지 못한 학생들도 많았음을 의미하고, 이는 A4를 바탕으로 A5를 숙달할 수 있는 학습 기회를 제공하는 것 또한 필요하다는 것을 의미한다.

<표 Ⅴ-12> 인지요소 숙달 패턴의 분포

인지 패턴	101010	111111	111100	111101	101110	111011	011111	기타	합계
인원	22	12	10	5	3	3	3	5	63

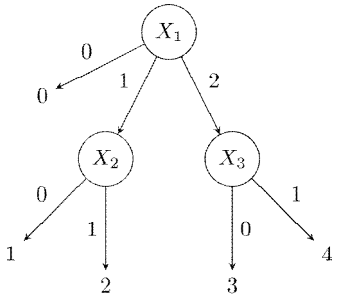
<표 Ⅴ-13> 각 노드의 guess와 slip

문항	노드	guess	slip	문항	노드	guess	slip
1	1	0.80	0.01	10	1	0.06	0.01
2	1	0.80	0.01		2	0.12	0.29
3	1	0.80	0.01		2	0.27	0.12
4	1	0.79	0.01		2	0.01	0.07
	2	0.36	0.01	11	1	0.06	0.84
5	1	0.96	0.01	12	1	0.03	0.87
6	1	0.85	0.06		1	0.18	0.64
7	1	0.01	0.33		2	0.01	0.20
	1	0.75	0.18	13	1	0.04	0.88
	1	0.01	0.15		1	0.19	0.31
8	1	0.01	0.39		1	0.01	0.78
	1	0.44	0.01		1	0.01	0.67
	2	0.01	0.50		2	0.71	0.29
	3	0.02	0.17		2	0.01	0.58
9	1	0.19	0.04		2	0.01	0.61
	2	0.01	0.06				
	3	0.86	0.14				

총 13개의 문항 중 단계와 전략이 모두 두 개 이상인 문항은 8번, 10번, 12번, 13번 문항이므로, 이 네 개 문항에 대한 문항 분석 결과를 제시하고자 한다. <표 Ⅴ-13>은 DTM 모형을 통해 추정된 전체 문항의 모수 추정 결과이며, <표 Ⅴ-14>, <표 Ⅴ-15>, <표 Ⅴ-16>, <표 Ⅴ-17>은 네 개의 문항에 대한 모수 추정 결과를 수형도 구조와 함께 제시한 것이다.

8번 문항의 경우 산술-기하 평균 부등식과 코시-슈바르츠 부등식 중 하나의 전략을 선택하여 해결하는 문항이다. 산술-기하 평균 부등식을 사용하는 전략의 경우 노드 1과 노드 2에서 모두 slip이 상대적으로 높고 guess는 낮은 것을 확인할 수 있다. 반면, 코시-슈바르츠 부등식을 사용하는 전략의 경우 노드 1과 노드 3에서 모두 slip이 낮고 노드 1에서의 guess가 0.44로 높은 것을 확인할 수 있다. 이는 두 전략 중 코시-슈바르츠 부등식을 사용하는 전략이 숙달한 인지요소를 적용하여 문항을 해결하기 수월한 전략임을 의미한다.

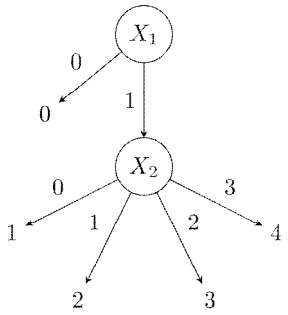
<표 Ⅴ-14> 8번 문항의 수형도 구조와 문항 모수

수형도	노드	범주	인지요소	guess	slip
	1	1	A2	0.01	0.39
	1	2	A4	0.44	0.01
	2	1	A3	0.01	0.50
	3	1	A5	0.02	0.17

10번 문항은 노드 1에서 치환을 이용하여 부등식을 변형한 뒤, 세 가지 방법 중 하나를 선택하여 해결하는 문항이다. 첫 번째 전략은 지수법칙만을 이용하여 해결하는 전략으로, 요구하는 인지요소는 A1 하나밖에 없으나 slip이 가장 높은 전략으로 나왔다. 두 번째 전략은 지수법칙과 산술-기하 평균을 함께 이용해야 하는 전략으로, 요구되는 인지요소는 3개로 가장

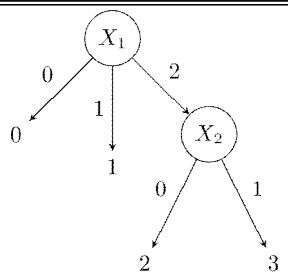
많으나 guess가 가장 높은 전략임을 확인할 수 있다. 이는 첫 번째 전략은 요구되는 인지요소는 적으나 인지요소를 적용하여 문제를 해결하는 것은 상대적으로 어려운 전략이며, 두 번째 전략은 요구되는 인지요소는 많으나 이를 적용하여 문제를 해결하는 것은 상대적으로 쉬운 전략임을 의미한다. 세 번째 전략은 코시-슈바르츠 부등식을 이용하는 전략으로, guess와 slip이 모두 낮게 나왔다.

<표 Ⅴ-15> 10번 문항의 수형도 구조와 문항 모수

수형도	노드	범주	인지요소	guess	slip
	1	1	A6	0.06	0.01
	2	1	A1	0.12	0.29
	2	2	A1, A2, A3	0.27	0.12
	2	3	A4, A5	0.01	0.07

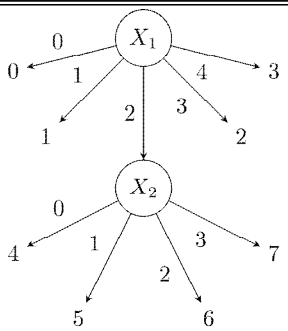
12번 문항의 경우, 노드 1에서 코시-슈바르츠 부등식을 활용하여 한 단계만에 해결하는 첫 번째 전략과, 노드 1에서 치환을 이용하여 부등식을 변형하고 노드 2에서 산술-기하 평균 부등식을 이용하는 두 번째 전략이 있는 문항이다. 코시-슈바르츠 부등식을 활용하는 첫 번째 전략의 경우 한 단계만에 해결할 수 있으나, slip이 매우 높은 것을 확인할 수 있다. 산술-기하 평균 부등식을 사용하는 두 번째 전략의 경우 두 단계를 거쳐야 하나, 첫 번째 전략에 비교했을 때 slip이 상대적으로 낮고, 노드 1에서의 guess가 상대적으로 높은 것을 확인할 수 있다. 즉, 산술-기하 평균 부등식을 사용하는 전략이 두 단계를 거쳐야 하지만 숙달한 인지요소를 이용하여 각 단계를 해결하는 것은 상대적으로 쉽다는 것을 알 수 있다.

<표 Ⅴ-16> 12번 문항의 수형도 구조와 문항 모수

수형도	노드	범주	인지요소	guess	slip
	1	1	A4, A5	0.03	0.87
	1	2	A6	0.18	0.64
	2	1	A2, A3	0.01	0.20

마지막으로 13번 문항의 경우 한 단계로 해결할 수 있는 전략이 세 가지 존재하며, 부등식의 치환을 이용하여 두 단계로 해결할 수 있는 전략이 세 가지 존재하는 문항이다. 한 단계로 해결할 수 있는 세 가지 전략은 모두 노드 1에서의 slip이 매우 높게 나오는 것을 확인할 수 있고, 이는 숙달한 인지요소로 문항을 해결하기가 쉽지 않은 전략임을 의미한다. 치환을 이용하여 두 단계로 해결하는 세 가지 전략 중에서는, 지수 법칙을 이용하는 전략이 guess는 가장 높고 slip은 가장 낮아서 쉬운 전략임을 확인할 수 있고, 산술-기하 평균 부등식이나 코시-슈바르츠 부등식을 이용하는 전략의 경우 guess는 낮고 slip은 높아서 상대적으로 어려운 전략임을 확인할 수 있다.

<표 Ⅴ-17> 13번 문항의 수형도 구조와 문항 모수

수형도	노드	범주	인지요소	guess	slip
	1	1	A1	0.04	0.88
	1	2	A6	0.19	0.31
	1	3	A1, A2, A3	0.01	0.78
	1	4	A4, A5	0.01	0.67
	2	1	A1	0.71	0.29
	2	2	A1, A2, A3	0.01	0.58
	2	3	A1, A4, A5	0.01	0.61

2.3. DTM 분석 결과를 이용한 피드백

수학 서술형 문항의 풀이 과정 분석틀, DTM 모형을 통한 추정 결과, 수형도 구조를 활용하여 피험자에게 다양한 방법으로 피드백을 제공할 수 있다. 첫째, 피험자의 인지요소 숙달 패턴을 추정할 수 있다. 예를 들면 가상의 피험자 A의 인지요소 숙달 패턴은 <표 Ⅴ-18>과 같다. 각 인지요소에 대하여 숙달한 것으로 추정되는 경우는 1, 숙달하지 않은 것으로 추정되는 경우는 0으로 표현했다.

<표 Ⅴ-18> 피험자 A의 인지요소 숙달 패턴 추정 결과

인지요소	A1	A2	A3	A4	A5	A6
숙달 여부	1	1	1	0	0	0

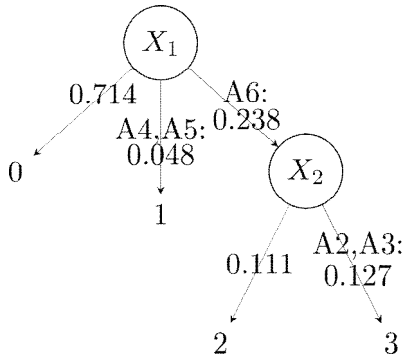
둘째, 수학 서술형 문항의 풀이 과정 분석틀에 기반을 두어 학생 반응 분석과 문항 분석을 비교하여 <표 Ⅴ-19>와 같이 제시할 수 있다. 10번 문항에서 '학생 인지요소 개수/문항 인지요소 개수'는 2/6인데, 10번 문항에서 사용할 수 있는 인지요소는 6개지만 피험자 A는 2개의 인지요소를 사용했기 때문이다. 13번 문항에서 '학생 단계 길이/문항 단계 길이'는 $2/\frac{3}{2}$인데, 이때 학생이 사용한 단계 길이는 문항 단계 길이보다 길다. 문항에 대한 단계 길이를 각 해결전략의 단계 길이의 평균으로 정의했기 때문에, 피험자가 평가에서 사용하지 않은 전략의 단계 길이가 사용한 전략의 단계 길이보다 짧으면 13번 문항과 같이 학생 단계 길이가 문항 단계 길이보다 길 수 있다.

<표 Ⅴ-19> 분석틀을 이용한 피험자 A의 풀이 과정 분석

문항 번호	학생 인지요소 개수/ 문항 인지요소 개수	학생 단계 길이/ 문항 단계 길이
1	1/1	1/1
2	1/1	1/1
3	1/1	1/1
4	1/2	1/2

5	0/1	0/1
6	1/1	1/1
7	3/5	$1/\frac{4}{3}$
8	2/3	2/2
9	2/3	2/3
10	2/6	$2/\frac{7}{3}$
11	0/1	0/1
12	1/5	$1/\frac{3}{2}$
13	2/6	$2/\frac{3}{2}$

셋째, 문항의 단계별 인지요소와 단계별 응답 비율을 수형도 구조를 사용하여 시각화한 것이다. [그림 Ⅴ-4]는 12번 문항을 예로 제시한 것이다. 피험자는 평가 문항, 모범답안 혹은 반응 분류기준과 함께 단계별로 필요한 인지요소를 확인하고, 다른 피험자들이 각 단계에서 응답한 분포를 확인할 수 있다.



[그림 Ⅴ-4] 12번 문항의 단계별 인지요소 및 응답 비율

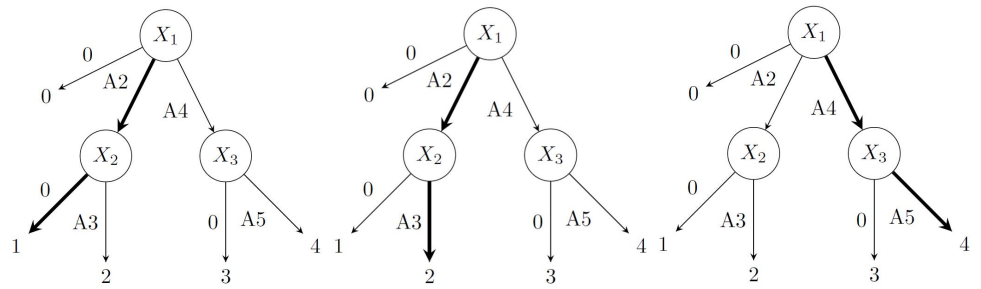
넷째, 피험자에게 학습경로를 추천해줄 수 있다. <표 Ⅴ-20>은 피험자 B의 인지요소 숙달 패턴 추정 결과를 나타낸 것이며, [그림 Ⅴ-5]는 피험자의 반응, 추천 경로, A4를 습득한 후 추천 경로를 나타낸 것이다. '추천 경로'에 관한 수형도는 학생의 현재 인지 상태를 고려하여 가장 적합한 해

결전략을 굵게 표시된 경로로 추천한 것으로, 학생의 인지 패턴을 고려하였을 때 정답에 해당하는 경로 중 가장 높은 진행 확률을 보이는 경로를 추천한 것이다. 현재 인지요소 습득 상황에서 문제를 푼다면 어떤 경로로 문제를 해결하는 것이 좋을지를 나타낸다고 해석할 수 있다. 예시에 나타난 학생의 경우 A2와 A3은 모두 습득하고 있으나 A4는 습득하지 못하고 있으므로, A4와 A5를 사용하여 코시-슈바르츠 부등식으로 문제를 해결하는 경로가 아닌, A2와 A3을 사용하여 산술-기하 평균 부등식으로 문제를 해결하는 경로를 추천받았음을 확인할 수 있다. 다음으로 'A4를 습득한 후 추천 경로'에 관한 수형도는 추가적인 인지요소의 습득을 고려하여 해결 전략을 추천하는 방법으로, 정답 확률이 높은 경로와 해당 경로를 위해 필요한 인지요소를 함께 추천하는 것이다. 해당 문제를 잘 해결하기 위해 어떤 인지요소를 우선적으로 습득해야 하는지를 알려줄 수 있으며, 학생은 이를 통해 학습에 대한 전략을 세울 수 있게 된다. 8번 문항은 A2와 A3을 사용하는 경로보다 A4와 A5를 사용하는 경로가 guess는 상대적으로 높고 slip은 낮아서 정답을 맞힐 가능성이 더 크다. 따라서 A4에 대한 학습을 진행하는 학습경로 추천이 적절하다는 결정을 할 수 있다.

<표 Ⅴ-20> 피험자 B의 인지요소 숙달 패턴 추정 결과

인지요소	A1	A2	A3	A4	A5	A6
숙달 여부	0	1	1	0	1	0

피험자 B의 반응	추천 경로	A4를 습득한 후 추천 경로



[그림 Ⅴ-5] 8번 문항의 수형도 구조를 활용한 피드백

Ⅵ. 결론

1. 논의 및 결론

본 연구의 목적은 다양한 전략으로 해결할 수 있는 수학 서술형 문항에 대한 DTM 모형 인지진단평가를 시행하여 그 결과를 분석하고 시사점을 찾는 것에 있다. 이를 위해 DTM 모형을 적용하여 인지진단평가를 설계할 수 있도록 수학 서술형 문항의 풀이 과정 분석틀을 제안했다. 분석틀을 이용하여 구조화된 문항 세트를 생성하였으며, 모의실험을 통해 생성한 문항 세트들을 사용하여 인지요소 추정 정확도에 영향을 미치는 요인을 확인하였다. 또한 절대부등식의 증명 영역의 서술형 평가 문항 세트를 개발하고, 경기도 소재 과학영재고등학교에서 연구 참여자를 모집하여 인지진단평가를 시행하여 결과를 분석하였으며, DTM 모형의 문항 풀이과정 분석틀을 활용한 피험자 대상 피드백 제공 방법을 제안했다. 연구 결과로부터 얻은 연구의 의의는 다음과 같다.

첫째, 수학 서술형 평가 문항 혹은 문항 세트 설계에 필요한 문항의 풀이 과정 분석틀을 제안했다. 본 고에서는 해결 과정에서 다양한 종류의 개념·원리·기능이 요구되고, 여러 단계를 거쳐야 해결할 수 있고, 다양한 해법의 반응이 존재하는 수학 서술형 문항과 그렇지 않은 문항을 구분하기 위하여 주어진 문항의 풀이 과정을 분석할 수 있는 분석틀을 제안했다. 분석틀은 인지요소(attribute), 단계(step), 반응(response) 3가지 차원을 가진다. 각 차원을 통해서 문항을 해결하기 위해 알아야 하는 개념·원리 개수, 문항을 해결하기 위해 거쳐야 하는 단계 길이, 문항을 해결하기 위해 사용할 수 있는 전략 개수를 파악할 수 있다. 각 차원의 개념을 수치로 제시할 수 있어서 평가 문항 혹은 문항 세트의 특징을 분석할 수 있고, 문항 간 혹은 문항 세트 간의 비교가 가능하다. 실제로 <표 Ⅳ-1>과 같이 문항 간 비교 분석이 가능하며, <표 Ⅳ-2>와 같이 문항 세트 간 비교 분석이 가능하다. 또한 분석틀은 DTM 기반 인지진단평가를 설계하는 데 활용될 수 있는데, 검사를 위한 인지요소와 관련된 후보 문항들의 예상 풀이를 이와

같은 틀로 분석해 본 후, 인지요소에 대한 식별성과 검사 시간 등을 고려하여 검사 문항을 선정할 수 있다.

둘째, 모의실험을 통해 DTM 모형을 적용한 인지진단평가 설계에 참고할 수 있는 문항 세트 구성의 예시와 실험 요인 등을 확인하였다. Ma(2019)는 DTM 모형을 적용한 인지진단평가에서 두 개의 전략을 지닌 동일 문항 6개, 이분반응 문항 5개로 이루어진 문항 세트를 활용하여 모의실험을 진행했다. 이는 다양한 전략으로 해결할 수 있는 수학 서술형 문항에 대한 인지진단평가에 사용하기에는 다소 단순한 문항들로 구성된 문항 세트였다. 반면에 <표 Ⅳ-2>에 제시된 문항 세트는 사용된 문항 유형, 문항 세트 인지요소 개수, 문항별 인지요소 개수 평균, 문항별 단계 길이 평균, 문항별 반응 개수 평균 등의 측면에서 Ma(2019)의 문항 세트보다 복잡하게 구성되어있다. 또한 <표 Ⅳ-1>에서 16개의 수학 서술형 문항 유형을 제시했는데, 주어진 유형을 적절히 조합하여 다양한 문항 세트를 개발할 수 있다.

셋째, 수학 서술형 문항을 활용한 학생 대상 인지진단평가의 계획, 실시 과정과 결과를 처음으로 제시했다. 본 고에는 인지요소 선정, 문항 및 평가지 구성, 피험자 집단 선정, 피험자 모집, 검사 시행, 피험자 반응 분류, 결과 분석 전체를 계획하고 진행하는 과정과 결과가 안내되어 있다. 분석 결과를 통해 학생들의 인지요소 숙달 상황을 개별적으로 또는 전체적으로 파악할 수 있으며, 학생들을 유형으로 분류하고 이를 피드백에 활용할 수 있음을 보였다.

넷째, DTM 모형을 이용하여 학생들의 반응 데이터를 분석한 결과를 바탕으로 절대부등식의 증명에 관한 인지진단 결과를 제시하였다. 각 인지요소의 숙달 비율에 대해 분석한 결과, 지수 법칙과 관련된 인지요소(A1)는 대부분의 학생이 습득하고 있는 것으로 나타났다. 산술-기하 평균 부등식과 관련된 인지요소(A2, A3) 및 코시-슈바르츠 부등식과 관련된 인지요소(A4, A5)의 숙달 비율을 통해서는, 학생들이 등호 성립 조건은 잘 알고 있지만 부등식을 문제에 따라 알맞게 적용하는 것은 상대적으로 어려워함을 알 수 있었고, 산술-기하 평균 부등식보다 코시-슈바르츠 부등식을 상대적으로 어려워함을 알 수 있었다. 부등식의 치환과 관련된 인지요소(A6)는

학생들이 가장 숙달하지 못하고 있는 인지요소로 나왔다. 관측된 인지 패턴의 빈도를 분석한 결과, 숙달 비율이 가장 높은 A1, A3, A5를 습득하고 있는 학생이 가장 많으며, 모든 인지요소를 습득하고 있는 학생이 두 번째로 많은 것을 알 수 있었다.

다섯째, 학생들의 반응 데이터를 DTM 모형으로 추정한 결과로부터 각 노드의 반응 범주별 문항 모수를 분석한 결과를 통해 경로별 사용된 전략들의 난이도를 비교할 수 있었다. 문항의 풀이 단계 노드의 범주별로 추측 모수(guess)와 실수 모수(slip) 값의 차이를 이용하여 사용된 전략들의 난이도를 비교한 결과, 일부 문항에서 학생들이 다른 인지요소를 요구하는 다른 전략을 채택함에 따라 문제해결의 난이도가 다르게 나타나는 양상을 관찰할 수 있었다. DTM 모형에서는 다분 문항 반응의 각 노드의 반응 범주별로 문항 반응 함수가 추정되기 때문에, 문제를 해결하는 각 단계와 전략별로 문항 특성이 어떻게 다르게 나타나는지를 비교해 볼 수 있다. 특히 추측 모수가 낮거나 실수 모수가 높은 범주의 경우는 학생들이 그 범주에 해당하는 전략을 채택하여 풀이를 진행할 때 어려움을 겪을 확률이 높은 것으로 해석할 수 있다. 이러한 분석 결과를 바탕으로 교사는 학생들이 채택하는 전략의 상대적 난이도를 그 학생의 인지요소 숙달 패턴을 기반으로 판단하여 그 학생이 전략을 바꾸도록 유도하거나 적절한 추가 학습을 통해 채택한 전략을 적절하게 적용하여 문제를 해결할 수 있게 도움을 줄 수 있다.

여섯째, 피험자의 반응을 다양한 방법으로 분석하여 효과적인 피드백 제공 방법을 제안했다. 이는 단순히 DTM 모형을 기계적으로 적용하여 피험자의 인지요소 숙달 패턴을 추정하는 데 그치는 것이 아니며, 수학 서술형 문항의 풀이 과정 분석틀에 기반을 두어, 평가에서 학생이 사용한 인지요소 개수, 풀이 과정 단계 길이를 문항의 것과 비교하여 제공하는 것이다. 그리고 문항의 단계별 인지요소와 단계별 응답 비율을 제공할 수 있으며, 수형도 구조를 활용하여 피험자의 반응, 피험자가 답할 가능성이 가장 큰 정답 반응, 추가 인지요소 습득을 통한 정답 반응을 제공할 수 있다. 또한 피험자가 습득하지 못한 인지요소가 있는 경우, 각 인지요소를 습득했다면

달라졌을 평가 결과를 보여주는 방법으로 피험자에게 학습경로를 추천해 줄 수 있다. 이를 통해서 각 피험자는 평가에서 자신이 보인 반응을 다각도로 이해하거나 다른 피험자들의 반응과 비교하여 이해함으로써 학습에 도움을 받을 수 있다.

2. 제한점 및 제언

본 연구 결과에 관한 몇 가지 제한점과 후속 연구를 위한 제언은 다음과 같다.

첫째, 학생 대상 인지진단평가에서 얻은 문항별 피험자 반응 중 결측치 정보를 인지요소 숙달 여부 추정에 사용하지 못했다. 인지진단평가를 시행하게 되면 평가 설계과정에서 계획하지 않은 인지요소를 사용하여 문제를 해결한 피험자 반응이 나올 수 있다. <표 Ⅴ-8>에 제시된 바와 같이 본 연구에서도 결측 반응이 존재했으며, 해당 학생은 0의 반응을 보인 것으로 간주하여 추정을 진행했다. 그런데 이런 피험자의 경우는 문항 세트와 관련되지 않은 인지요소를 사용함으로써 사전에 설계된 문항 구조에 관한 반응 정보를 제공하지 않은 것일 뿐, 해당 문제해결에 요구되는 인지요소를 숙달하지 않았다고 단정하는 것에는 무리가 있다. 현재 R 프로그래밍 언어의 GDINA package에서 DTM 모형을 통해 인지요소 숙달 여부를 추정하는 함수에는 결측치를 고려하는 옵션이 존재하지 않는다. 따라서 후속 연구에서는 인지진단평가에서 결측치로 분류된 학생 반응 정보가 인지요소 숙달 여부 추정에 이용될 수 있도록 GDINA package의 DTM 추정 함수를 개선하여 실험을 진행할 필요가 있다.

둘째, 본 연구에서 사용된 Q행렬은 양적 분석을 통한 타당성 검증이 이루어지지 않았다. 인지진단평가의 Q행렬이 오류(misspecificaion)를 가지면 문항 모수가 부정확하게 추정되며, 피험자의 인지요소 숙달 여부 추정 정확도가 떨어진다(Rupp & Templin, 2008). 이에 따라 양적 분석을 통해 정확한 Q행렬을 제작하기 위한 연구가 진행되어왔으며, Q행렬 타당성 지수(validation index)로 타당도를 수치화하여 주어진 Q행렬을 개선해나가는 연구가 이루어지고 있다(Torre, 2008; Feng, 2013; 안희상, 2016). 그런데 본 연구에서 사용된 Q행렬에 대해서는 타당성 지수를 활용한 Q행렬 개선 과정이 이루어지지 않았다. 게다가 기존 Q행렬 타당성 연구에서 다루어진 대부분의 인지진단모형 Q행렬과 달리 DTM 모형의 Q행렬은 하나의 문항에 대한 정보가 두 개 이상의 행에 입력되어있으며, 피험자는 Q행렬에 있는 모든 행에 대한 정보를 제공하지 않는 등의 차

이가 있다. 따라서 DTM 모형의 특성을 고려하여 DTM 모형에 관한 Q 행렬 타당성 지수와 Q행렬 개선 방법이 새롭게 제안될 필요가 있다. 이를 통해 수학 서술형 문항 세트의 Q행렬 개선 과정이 적용된 후속 연구가 이루어질 수 있다.

셋째, 학생 대상 인지진단평가의 연구 참여자 수가 충분하지 않아서 평가 결과 해석에 유의할 필요가 있다. 일반적으로 통계적 판단을 신뢰하기 위해서는 충분한 sample size가 보장되어야 한다. <표 Ⅴ-5>, <표 Ⅴ-6>, <표 Ⅴ-7>은 학생 대상 평가에 대한 모의실험에서 피험자 수(sample size)별 $RMSE_j$를 나타낸 것이다. 모든 문항, 모든 문항 질에서 피험자 수와 $RMSE_j$가 반비례하는 것을 확인할 수 있다. 그리고 본 연구에서는 부트스트랩 방법(Bootstrap method)이 사용되었다. Dziak et al.(2014)은 sample size를 각각 50, 100, 150으로 설정하여 부트스트랩의 검정 능력을(statistical power) 확인하였는데, sample size가 작을수록 검정 능력이 떨어진다는 결론을 얻었다. 본 연구의 학생 인지진단평가 피험자 수는 63명이므로 Dziak et al.(2014)의 관점으로 볼 때 본 연구 환경에서 이루어진 부트스트랩의 검정 능력은 좋지 않을 것으로 예상된다. 그러므로 <표 Ⅴ-9>, <표 Ⅴ-10>에 제시된 모형 적합도의 해석에 유의해야 한다. 따라서 본 연구처럼 학생 대상 인지진단평가를 시행하게 된다면 참여 대상 학생 선정, 연구 참여 모집 방법 계획 단계부터 더 많은 연구 참여자 모집을 위한 노력이 필요할 것이다.

넷째, 현재 개발된 DTM 모형은 오개념이나 오류의 유형까지 분류하는 모형으로 확장되지 않은 모형이다. 선다형 문항의 경우 답지의 선택에 따라 오개념을 진단하는 Bug-DINO, SISM과 같은 방법이 개발되어 있다(Kuo et al., 2018). 서술형 문항의 풀이에는 학생들의 오개념을 판단할 수 있는 추가 정보가 관찰될 확률이 높기 때문에, 이 정보를 활용할 수 있는 새로운 인지진단모형을 통해 더 정밀한 서술형 답안의 진단이 가능할 것으로 예상되므로 이와 관련된 후속 연구가 필요할 것이다.

참고문헌

교육부(2015). **수학과 교육과정**. 교육부 고시 제 2015-74호 [별책 8]. 서울: 저자

박교식, 이종희, 김진환, 남진영, 김남희, 임재훈, 유연주, 권석일, 김선희, 김종욱, 김경직, 윤형석, 고현주, 윤형주, 김영실, 김해성, 이경진, 조은수, 이정연, 양정은(2020). **고등학교 수학**. 서울: 동아출판㈜

박정, 박경미(2004). 수행형 문항과 선다형 문항의 수학적 능력 추정 효율성 비교. **수학교육논문집, 43**(2), 151-162.

박지현, 김수진(2015). 수학 인지적 속성에 따른 TIMSS 2011 8 학년 성취도 상위국 특성 비교. **수학교육학연구, 25**(3), 303-321.

안희상(2016). **인지진단 검사의 Q행렬 타당성 분석 방법 연구**, 서울대학교 석사학위논문.

왕효원(2018). **인지진단평가의 문항수준 모형적합도 연구**, 서울대학교 석사학위논문.

이현숙, 고호경(2014). 인지진단모형을 적용한 TIMSS 8 학년 수학 기하 영역의 성차 분석. **학교수학, 16**(2), 387-407.

황혜정, 나귀수, 최승현, 박경미, 임재훈, 서동엽(2012). **수학교육학신론**, 서울: 문음사.

B. Efron. (1979). Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics, Vol. 7, No. 1*, pp. 1-26.

Bor-Chen Kuo, Chun-Hua Chen, & de la Torre, J. (2018). A cognitive diagnosis model for identifying coexisting skills and misconceptions. *Applied Psychological Measurement, 42*(3), 179-191.

Bor-Chen Kuo, Chun-Hua Chen, Chih-Wei Yang, & Magdalena Mo Ching Mok (2016). Cognitive diagnostic models for tests with multiple-choice and constructed-response items, *Educational Psychology*, *36*(6), 1115-1133.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood

estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443 - 459.

Choi, K. M., Lee, Y. S., & Park, Y. S. (2015). What CDM can tell about what students have learned: An analysis of TIMSS eighth grade mathematics. Eurasia Journal of Mathematics, *Science and Technology Education, 11*(6), 1563-1577.

de la Torre, J. (2008). An Empirically Based Method of Q-matrix Validation for the DINA Model: Development and Applications, *Journal of Educational Measurement*, *45*(4), 343-362.

de la Torre, J. (2010). The partial-credit DINA model. *Paper presented at the international meeting of the Psychometric Society*, Athens, GA.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179-199.

de la Torre, J., & Douglas, J. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika, 73*, 595 - 624.

Dogan, E., & Tatsuoka, K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics, 68*(3), 263-272.

Feng, Y. (2013). Estimation and Q-matrix validation for diagnostic classification models. (Master's thesis). *University of South Carolina Scholar Commons.*

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*(4), 301-321.

Hansen, M. (2013). Hierarchical item response models for cognitive diagnosis. *Unpublished doctoral dissertation,* University of

California at Los Angeles.

Henson, R. A., Templin, J. L., & Wilse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika, 74*(2), 191.

Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement, 71*(2), 407-419.

Jang, E. E. (2008). A framework for cognitive diagnostic assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), Towards adaptive CALL: Natural language processing for diagnosis language assessment (pp. 117-131). Ames, IA: Iowa State University.

John J. Dziak, Stephanie T. Lanza & Xianming Tan (2014). Effect Size, Statistical Power, and Sample Size Requirements for the Bootstrap Likelihood Ratio Test in Latent Class Analysis, *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 534-552.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258-272.

Lee, Y. S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *International Journal of Testing, 11*(2), 144-177.

Leighton, J. P., & Gierl, M. J. (2007). Why cognitive diagnostic assessment? In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education: Theory and Applications* (pp. 3 - 18). New York: Cambridge University Press.

Ma, W. (2019). A diagnostic tree model for polytomous responses with multiple strategies. *British Journal of Mathematical and Statistical Psychology, 72*, 61-82.

Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology, 69*, 253 - 275.

Ma, W., de la Torre, J., Sorrrel, M., Jiang, Z. (2022). GDINA: The generalized DINA model framework [Computer software]. Retrieved from https://CRAN.R-project.org/package=GDINA

Mislevy, R. J. (1996). Test theory reconceived, *Journal of Educational Measurement, 33*, 379 - 416.

National Center for Education Statistics(NCES)[Website]. (2022.11.27.). URL:https://nces.ed.gov/timss/overview.asp

Polya, G.(1962). *Mathematical Discovery*. New York : John Wiley & Sons, Inc..

Radmila Bulajich Manfrino, Jose Antonio Gomez Ortega, Rogelio Valdez Delgado. (2009). *Inequalities: A Mathematical Olympiad Approach*. Birkhauser Verlag, Basel, Schweiz.

Read, T. R., & Cressie, N. A. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York, NY: Springer.

Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills diagnosis using IRT based latent class models. *Journal of Educational Measurement, 44*(4), 293-311.

Rupp, A. A., & Templin, J. (2008). The effects of Q-Matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68,* pp. 78-96.

Sen, S., & Bradshaw, L. (2017). Comparison of relative fit indices for diagnostic model selection. *Applied Psychological Measurement, 41*(6), 422-438.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological method,*

11(3), 287.

Templin, J. L., Henson, R. A., Rupp, A. A., Jang, E., & Ahmed, M. (2008). Cognitive diagnosis models for nominal response data. *Paper presented at the annual meeting of the National Council on Measurement in Education*, New York.

Von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of Psychological Research Online, 2*(2), 29 - 48.

Von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology, 61*, 287 - 307.

Wu, H., Liang, X., Yurekli, H., Becker, B. J., Paek, I., & Binici, S. (2020). Exploring the impact of Q-Matrix specifications through a DINA model in a large-scale mathematics assessment. *Journal of Psychoeducational Assessment, 38*(5), 581-598.

Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika, 81*(3), 625-649.

<부 록>

[부록 1] 문항에 대한 학생 반응 분류기준

1. $a \geq b > 0$, $m > 0$인 실수 a, b, m에 대하여 a^m과 b^m의 대소 관계를 부등식으로 나타내시오.

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A1	$a^m \geq b^m$이 성립한다.

2. 산술-기하 평균 부등식을 이용하여, 양의 실수 x, y에 대하여 $x^2 + y^2 = 4$일 때 다음 부등식이 성립함을 보이시오.

$$xy \leq 2$$

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A2	산술-기하 평균 부등식에 의해 $4 = x^2 + y^2 \geq 2xy$이 성립한다. 따라서 $xy \leq 2$가 성립한다.

3. 위 2번에서 $xy = 2$일 때, x와 y의 값을 구하시오.

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A3	산술-기하 평균 부등식 $x^2 + y^2 \geq 2xy$에서 등호가 성립하기 위한 조건은 $x = y$이다.

4. 산술-기하 평균 부등식을 이용하여, 양의 실수 x에 대하여 $\left(x+\frac{2}{x}\right)\left(x+\frac{8}{x}\right)$의 최솟값을 구하고, 그 과정을 서술하시오.

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
X_2	1	A2	$x+\frac{2}{x} \geq 2\sqrt{2}$, $x+\frac{8}{x} \geq 2\sqrt{8}=4\sqrt{2}$ 이므로 $\left(x+\frac{2}{x}\right)\left(x+\frac{8}{x}\right) \geq 2\sqrt{2}\times 4\sqrt{2}=16$이 성립한다.
	2	A2, A3	$\left(x+\frac{2}{x}\right)\left(x+\frac{8}{x}\right)=x^2+\frac{16}{x^2}+10 \geq 2\sqrt{16}+10=18$ 이고, 등호 성립 조건인 $x^2=\frac{16}{x^2}$이 가능하다. 따라서 $\left(x+\frac{2}{x}\right)\left(x+\frac{8}{x}\right)$의 최솟값은 18이다.

5. 코시-슈바르츠 부등식을 이용하여, 양의 실수 x, y에 대하여 $x^2+y^2=4$일 때 다음 부등식이 성립함을 보이시오.

$$4x+3y \leq 10$$

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A4	코시-슈바르츠 부등식에 의해 다음이 성립한다. $100=4(4^2+3^2)=(x^2+y^2)(4^2+3^2) \geq (4x+3y)^2$ 따라서 $4x+3y \leq 10$가 성립한다.

6. 위 5번에서 $4x+3y=10$일 때, x와 y의 값을 구하시오.

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A5	코시-슈바르츠 부등식 $(x^2+y^2)(4^2+3^2) \geq (4x+3y)^2$에서 등호가 성립하기 위한 조건은 $\frac{x}{4}=\frac{y}{3}$이다. 즉, $x=\frac{8}{5}$, $y=\frac{6}{5}$

7. 양의 실수 x, y에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\sqrt{x+y} \geq \frac{\sqrt{2x}+\sqrt{2y}}{2}$$

<풀이>

전략 A

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A1	양변을 제곱하면 $x+y \geq \frac{2x+2y+2\sqrt{(2x)(2y)}}{4}$가 성립하며 다음과 같다. $\frac{4x+4y}{4}-\frac{2x+2y+2\sqrt{(2x)(2y)}}{4} \geq 0$ $\frac{1}{2}(\sqrt{x}-\sqrt{y})^2 = \frac{x+y-2\sqrt{xy}}{2} \geq 0$ ($x=y$일 때 등호가 성립한다.)

전략 B

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	2	A1, A2, A3	$x+y \geq \frac{2x+2y+2\sqrt{(2x)(2y)}}{4}$가 성립하며 다음과 같다. $2x+2y \geq 2\sqrt{(2x)(2y)}$ 산술-기하 평균 부등식에 의해 $2x+2y \geq 2\sqrt{(2x)(2y)}$이 성립한다. 산술-기하 평균 부등식 $2x+2y \geq 2\sqrt{(2x)(2y)}$에서 등호가 성립하기 위한 조건은 $x=y$이다.

전략 C

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	3	A1, A4, A5	코시-슈바르츠 부등식에 의해 다음이 성립한다. $\left((\sqrt{2})^2+(\sqrt{2})^2\right)\left((\sqrt{x})^2+(\sqrt{y})^2\right)$ $\geq \left(\sqrt{2x}+\sqrt{2y}\right)^2$ 따라서 $4(x+y) \geq \left(\sqrt{2x}+\sqrt{2y}\right)^2$가 성립한다. $2\sqrt{x+y} \geq \sqrt{2x}+\sqrt{2y}$가 성립한다. 코시-슈바르츠 부등식 $\left((\sqrt{2})^2+(\sqrt{2})^2\right)\left((\sqrt{x})^2+(\sqrt{y})^2\right)$ $\geq \left(\sqrt{2x}+\sqrt{2y}\right)^2$ 에서 등호가 성립하기 위한 조건은 $x=y$이다.

8. 실수 a, b, c, x, y, z에 대하여 x, y, $z>0$ 일 때, 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\frac{a^2}{x}+\frac{b^2}{y}+\frac{c^2}{z} \geq \frac{(a+b+c)^2}{x+y+z}$$

<풀이>

전략 A

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
X_2	1	A2	양변에 $(x+y+z)$를 곱한 후, 양변에 $(a^2+b^2+c^2)$를 빼서 다음과 같이 정리한다. $\left(\frac{y}{x}a^2+\frac{x}{y}b^2\right)+\left(\frac{z}{y}b^2+\frac{y}{z}c^2\right)+\left(\frac{x}{z}c^2+\frac{z}{x}a^2\right)$ $\geq 2ab+2bc+2ca$ 산술-기하 평균 부등식에 의해 다음이 성립한다. $\frac{y}{x}a^2+\frac{x}{y}b^2 \geq 2ab$ $\frac{z}{y}b^2+\frac{y}{z}c^2 \geq 2bc$ $\frac{x}{z}c^2+\frac{z}{x}a^2 \geq 2ca$ 따라서 주어진 부등식이 성립한다.
	2	A2, A3	(위 풀이에 이어서 다음을 서술) 등호가 성립하기 위한 조건은 $\frac{a}{x}=\frac{b}{y}=\frac{c}{z}$이다.

전략 B

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
X_3	3	A4	코시-슈바르츠 부등식에 의해 다음이 성립한다. $\left(\left(\frac{a}{\sqrt{x}}\right)^2+\left(\frac{b}{\sqrt{y}}\right)^2+\left(\frac{c}{\sqrt{z}}\right)^2\right)$ $\times\left((\sqrt{x})^2+(\sqrt{y})^2+(\sqrt{z})^2\right)\geq(a+b+c)^2$ 따라서 $\frac{a^2}{x}+\frac{b^2}{y}+\frac{c^2}{z}\geq\frac{(a+b+c)^2}{x+y+z}$가 성립한다.
	4	A4, A5	(위 풀이에 이어서 다음을 서술) 코시-슈바르츠 부등식에서 등호가 성립하기 위한 조건은 $\frac{a}{x}=\frac{b}{y}=\frac{c}{z}$이다.

9. 삼각형의 세 변의 길이가 각각 a, b, c일 때, 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$(b+c-a)(c+a-b)(a+b-c)\leq abc$$

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
X_2	1	A6	$a=y+z,\ b=z+x,\ c=x+y$로 치환한다. 부등식은 $2x\cdot 2y\cdot 2z\leq(y+z)(z+x)(x+y)$과 동치다. (혹은 다양한 방식의 치환을 통해 주어진 부등식과 동치인 부등식을 얻음)
X_3	2	A2	산술-기하 평균 부등식에 의해 다음이 각각 성립한다.

			$0 < 2\sqrt{yz} \le y+z,\ 0 < 2\sqrt{zx} \le z+x,$ $0 < 2\sqrt{xy} \le x+y$ 이들을 변과 변끼리 곱하여 다음 부등식을 얻는다. $2x \cdot 2y \cdot 2z = 2\sqrt{yz} \cdot 2\sqrt{zx} \cdot 2\sqrt{xy}$ $\le (y+z)(z+x)(x+y)$ 따라서 다음이 성립한다. $(b+c-a)(c+a-b)(a+b-c) \le abc$
	3	A2, A3	(위 풀이에 이어서 다음을 서술) 등호가 성립하기 위한 조건은 $x=y=z$이다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

10. 삼각형의 세 변의 길이가 각각 a, b, c일 때, 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\sqrt{a+b-c}+\sqrt{b+c-a}+\sqrt{c+a-b} \le \sqrt{a}+\sqrt{b}+\sqrt{c}$$

<풀이>

전략 A

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A6	$x=\dfrac{-a+b+c}{2},\ y=\dfrac{a-b+c}{2},\ z=\dfrac{a+b-c}{2}$로 치환하면 주어진 부등식은 다음과 동치다. $\sqrt{2x}+\sqrt{2y}+\sqrt{2z} \le \sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$ (혹은 다양한 방식의 치환을 통해 주어진 부등식과 동치인 부등식을 얻음)

X_2	2	A1	$(\sqrt{x+y})^2-\left(\frac{\sqrt{2x}+\sqrt{2y}}{2}\right)^2$ $\geq x+y-\frac{2x+2y+2\sqrt{(2x)(2y)}}{4}$ $\geq \frac{x+y-2\sqrt{xy}}{2}=\frac{1}{2}(\sqrt{x}-\sqrt{y})^2 \geq 0$ 이므로 $\sqrt{x+y}\geq\frac{\sqrt{2x}+\sqrt{2y}}{2}$가 성립한다. (7번 문제의 결과) 같은 방법으로 $\sqrt{y+z}\geq\frac{\sqrt{2y}+\sqrt{2z}}{2}$, $\sqrt{z+x}\geq\frac{\sqrt{2z}+\sqrt{2x}}{2}$이 성립한다. $\sqrt{2x}+\sqrt{2y}+\sqrt{2z}\leq\sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$이 성립하므로 주어진 부등식이 성립한다. $\sqrt{2x}+\sqrt{2y}+\sqrt{2z}\leq\sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$에서 등호가 성립하기 위한 조건은 $x=y=z$이다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

전략 B

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A6	$a=y+z$, $b=z+x$, $c=x+y$로 치환하면 주어진 부등식은 다음과 동치다. $\sqrt{2x}+\sqrt{2y}+\sqrt{2z}\leq\sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$ (혹은 다양한 방식의 치환을 통해 주어진 부등식과 동치인 부등식을 얻음)
X_2	3	A1, A2, A3	산술-기하 평균 부등식에 의해 $2x+2y\geq 2\sqrt{(2x)(2y)}$이 성립하고, 이는 다음과 같다.

			$x+y \geq \dfrac{2x+2y+2\sqrt{(2x)(2y)}}{4}$ 부등식은 $\sqrt{x+y} \geq \dfrac{\sqrt{2x}+\sqrt{2y}}{2}$과 동치다. 같은 방법으로 $\sqrt{y+z} \geq \dfrac{\sqrt{2y}+\sqrt{2z}}{2}$, $\sqrt{z+x} \geq \dfrac{\sqrt{2z}+\sqrt{2x}}{2}$이 성립한다. 따라서 다음이 성립한다. $\sqrt{2x}+\sqrt{2y}+\sqrt{2z} \leq \sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$ 산술-기하 평균 부등식 $2x+2y \geq 2\sqrt{(2x)(2y)}$에서 등호가 성립하기 위한 조건은 $x=y$이다. 그러므로 $\sqrt{2x}+\sqrt{2y}+\sqrt{2z} \leq \sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$에서 등호가 성립하기 위한 조건은 $x=y=z$이다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

전략 C

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A6	$a=y+z,\ b=z+x,\ c=x+y$로 치환하면 주어진 부등식은 다음과 동치다. $\sqrt{2x}+\sqrt{2y}+\sqrt{2z} \leq \sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$ (혹은 다양한 방식의 치환을 통해 주어진 부등식과 동치인 부등식을 얻음)
X_2	4	A1, A4, A5	부등식 $\sqrt{x+y} \geq \dfrac{\sqrt{2x}+\sqrt{2y}}{2}$은 $4(x+y) \geq (\sqrt{2x}+\sqrt{2y})^2$과 동치다.

			코시-슈바르츠 부등식에 의해 다음이 성립한다. $4(x+y)=\left((\sqrt{2})^2+(\sqrt{2})^2\right)\left((\sqrt{x})^2+(\sqrt{y})^2\right)$ $\geq(\sqrt{2x}+\sqrt{2y})^2$ 같은 방법으로 $\sqrt{y+z}\geq\dfrac{\sqrt{2y}+\sqrt{2z}}{2}$, $\sqrt{z+x}\geq\dfrac{\sqrt{2z}+\sqrt{2x}}{2}$이 성립한다. 따라서 다음이 성립한다. $\sqrt{2x}+\sqrt{2y}+\sqrt{2z}\leq\sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$ 코시-슈바르츠 부등식 $\left((\sqrt{2})^2+(\sqrt{2})^2\right)\left((\sqrt{x})^2+(\sqrt{y})^2\right)\geq(\sqrt{2x}+\sqrt{2y})^2$ 에서 등호가 성립하기 위한 조건은 $x=y$이다. 그러므로 $\sqrt{2x}+\sqrt{2y}+\sqrt{2z}\leq\sqrt{x+y}+\sqrt{y+z}+\sqrt{z+x}$에서 등호가 성립하기 위한 조건은 $x=y=z$이다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

11. 양의 실수 a, b, c에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\frac{a}{b}+\frac{b}{c}+\frac{c}{a}\geq\frac{c+a}{c+b}+\frac{a+b}{a+c}+\frac{b+c}{b+a}$$

<풀이>

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A6	$x=\dfrac{a}{b}$, $y=\dfrac{b}{c}$, $z=\dfrac{c}{a}$로 치환한다.

			그러면 부등식은 $\frac{1-x}{1+y}+\frac{1-y}{1+z}+\frac{1-z}{1+x}\geq 0$와 동치다. (혹은 다양한 방식의 치환을 통해 주어진 부등식과 동치인 부등식을 얻음)

실제 모범답안은 다음과 같으나, 실제 실험에서 나타난 학생들의 반응은 위의 경우뿐이었다.

<모범답안>

$x=\frac{a}{b}$, $y=\frac{b}{c}$, $z=\frac{c}{a}$로 치환한다.

그러면 부등식은 $\frac{1-x}{1+y}+\frac{1-y}{1+z}+\frac{1-z}{1+x}\geq 0$으로 변형되고, $xyz=1$이 성립한다.

이는 $x^2z+y^2x+z^2y+x^2+y^2+z^2 \geq x+y+z+3$과 동치다.

코시-슈바르츠 부등식에 의해 $(1^2+1^2+1^2)(x^2+y^2+z^2)\geq(x+y+z)^2$이 성립한다.

산술-기하 평균 부등식에 의해

$x^2+y^2+z^2 \geq \frac{x+y+z}{3}(x+y+z)\geq \sqrt[3]{xyz}(x+y+z)=x+y+z,$

$x^2z+y^2x+z^2y\geq 3\sqrt[3]{x^3y^3z^3}=3$이 성립한다.

즉, $x^2z+y^2x+z^2y+x^2+y^2+z^2 \geq x+y+z+3$이다.

등호가 성립하기 위한 조건은 $\frac{x}{1}=\frac{y}{1}=\frac{z}{1}$, $x^2z=y^2x=z^2y$인데,

이는 $x=y=z$이다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

12. 양의 실수 a, b, c에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\frac{a^3}{a^3+2b^3}+\frac{b^3}{b^3+2c^3}+\frac{c^3}{c^3+2a^3}\geq 1$$

<풀이>

전략 A

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A4, A5	코시-슈바르츠 부등식에 의해 다음이 성립한다. $(a^3(a^3+2b^3)+b^3(b^3+2c^3)+c^3(c^3+2a^3))$ $\times\left(\frac{a^3}{a^3+2b^3}+\frac{b^3}{b^3+2c^3}+\frac{c^3}{c^3+2a^3}\right)$ $\geq (a^3+b^3+c^3)^2$ 그리고, $a^3(a^3+2b^3)+b^3(b^3+2c^3)+c^3(c^3+2a^3)$ $=(a^3+b^3+c^3)^2$이므로 다음이 성립한다. $(a^3+b^3+c^3)^2\left(\frac{a^3}{a^3+2b^3}+\frac{b^3}{b^3+2c^3}+\frac{c^3}{c^3+2a^3}\right)$ $\geq (a^3+b^3+c^3)^2$ 양변을 $(a^3+b^3+c^3)^2$로 나누면 $\frac{a^3}{a^3+2b^3}+\frac{b^3}{b^3+2c^3}+\frac{c^3}{c^3+2a^3}\geq 1$이 성립한다. 등식의 등호가 성립하기 위한 조건은 $\frac{1}{a^3+2b^3}=\frac{1}{b^3+2c^3}=\frac{1}{c^3+2a^3}$이다. (또는 $a=b=c$이다.)

전략 B

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황

	1	A6	$a^3=x,\ b^3=y,\ c^3=z$로 치환한다. 그러면 주어진 부등식은 $\frac{x}{x+2y}+\frac{y}{y+2z}+\frac{z}{z+2x}\geq 1$으로 변형된다.
X_2	2	A2	주어진 부등식은 다음과 같다. $x(y+2z)(z+2x)+y(x+2y)(z+2x)$ $+z(x+2y)(y+2x)\geq(x+2y)(y+2z)(z+2x)$ 정리하면, $\frac{x^2y+y^2z+z^2x}{3}\geq xyz$와 같으며 이는 산술-기하 평균 부등식에 의해 성립한다.
	3	A2, A3	(위 풀이에 이어서 다음을 서술) 등식의 등호가 성립하기 위한 조건은 $x^2y=y^2z=z^2x$이다. (또는 $x=y=z$이다.) 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

13. 양의 실수 a, b, c에 대하여 다음 부등식이 성립함을 증명하고, 등호가 성립하기 위한 조건을 구하시오.

$$\frac{1}{3}\left(\frac{bc}{a}+\frac{ca}{b}+\frac{ab}{c}\right)\geq\sqrt{\frac{a^2+b^2+c^2}{3}}$$

<풀이>

전략 A

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	1	A1	주어진 부등식의 양변을 제곱하면 다음이 성립한다. $\frac{a^2b^2c^2}{9}\left(\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}\right)^2\geq\frac{a^2+b^2+c^2}{3}$

			$\Leftrightarrow \frac{1}{3}\left(\frac{1}{a^4}+\frac{1}{b^4}+\frac{1}{c^4}-\frac{1}{a^2b^2}-\frac{1}{b^2c^2}-\frac{1}{c^2a^2}\right)\geq 0$ $\Leftrightarrow \frac{1}{6}\left\{\left(\frac{1}{a^2}-\frac{1}{b^2}\right)^2+\left(\frac{1}{b^2}-\frac{1}{c^2}\right)^2+\left(\frac{1}{c^2}-\frac{1}{a^2}\right)^2\right\}\geq 0$ 등호가 성립하기 위한 조건은 $a=b=c$이다.

전략 B

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	2	A1, A2, A3	주어진 부등식의 양변을 제곱하면 다음이 성립한다. $\frac{1}{9}\left(\frac{bc}{a}+\frac{ca}{b}+\frac{ab}{c}\right)^2 \geq \frac{a^2+b^2+c^2}{3}$ $\Leftrightarrow \frac{1}{9}\left(\frac{b^2c^2}{a^2}+\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2}+2a^2+2b^2+2c^2\right)$ $\geq \frac{a^2+b^2+c^2}{3}$ $\Leftrightarrow \frac{b^2c^2}{a^2}+\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2}\geq a^2+b^2+c^2$ 산술-기하 평균 부등식에 의해 다음이 성립하므로 문제에서 주어진 부등식이 성립함을 알 수 있다. $\frac{1}{2}\left(\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2}\right)\geq a^2,\ \frac{1}{2}\left(\frac{b^2c^2}{a^2}+\frac{a^2b^2}{c^2}\right)\geq b^2,$ $\frac{1}{2}\left(\frac{b^2c^2}{a^2}+\frac{c^2a^2}{b^2}\right)\geq c^2$ 각각 $\frac{bc}{a}=\frac{ca}{b},\ \frac{bc}{a}=\frac{ab}{c},\ \frac{ab}{c}=\frac{ca}{b}$일 때 위 부등식의 등호가 성립한다. 따라서 등호가 성립하기 위한 조건은 $a=b=c$이다.

전략 C

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	3	A1, A4, A5	코시-슈바르츠 부등식에 의해 다음이 성립한다. $\left(\frac{b^2c^2}{a^2}+\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2}\right)\left(\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2}+\frac{b^2c^2}{a^2}\right)$ $\geq (a^2+b^2+c^2)^2$ 이는 다음 부등식과 동치다. $\Leftrightarrow \left(\frac{b^2c^2}{a^2}+\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2}\right)^2 \geq (a^2+b^2+c^2)^2$ $\Leftrightarrow \frac{b^2c^2}{a^2}+\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2} \geq a^2+b^2+c^2$ $\Leftrightarrow \frac{b^2c^2}{a^2}+\frac{c^2a^2}{b^2}+\frac{a^2b^2}{c^2}+2a^2+2b^2+2c^2$ $\geq 3(a^2+b^2+c^2)$ $\Leftrightarrow \left(\frac{bc}{a}+\frac{ca}{b}+\frac{ab}{c}\right)^2 \geq 3(a^2+b^2+c^2)$ $\Leftrightarrow \frac{1}{9}\left(\frac{bc}{a}+\frac{ca}{b}+\frac{ab}{c}\right)^2 \geq \frac{a^2+b^2+c^2}{3}$ 따라서 다음이 성립한다. $\frac{1}{3}\left(\frac{bc}{a}+\frac{ca}{b}+\frac{ab}{c}\right) \geq \sqrt{\frac{a^2+b^2+c^2}{3}}$ 등호가 성립하기 위한 조건은 $a=b=c$이다.

전략 D

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황

	4	A6	$x=\frac{bc}{a}$, $y=\frac{ca}{b}$, $z=\frac{ab}{c}$로 치환한다. 그러면 부등식은 $\frac{x+y+z}{3} \geq \sqrt{\frac{xy+yz+zx}{3}}$ 으로 변형된다. (혹은 다양한 방식의 치환을 통해 주어진 부등식과 동치인 부등식을 얻음)
X_2	5	A1	양변을 제곱하여 정리하면 $x^2+y^2+z^2 \geq xy+yz+zx$가 된다. 이는 $(x-y)^2+(y-z)^2+(z-x)^2 \geq 0$와 동치다. $x=y=z$일 때 등호가 성립한다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

전략 E

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	4	A6	$x=\frac{bc}{a}$, $y=\frac{ca}{b}$, $z=\frac{ab}{c}$로 치환한다. 그러면 부등식은 $\frac{x+y+z}{3} \geq \sqrt{\frac{xy+yz+zx}{3}}$ 으로 변형된다. (혹은 다양한 방식의 치환을 통해 주어진 부등식과 동치인 부등식을 얻음)
X_2	6	A1, A2, A3	양변을 제곱하여 정리하면 $x^2+y^2+z^2 \geq xy+yz+zx$가 된다. 산술-기하 평균 부등식에 의해 $\frac{x^2}{2}+\frac{y^2}{2} \geq xy$가 성립한다.

			같은 방법으로 $\frac{y^2}{2}+\frac{z^2}{2}\geq yz,\ \frac{z^2}{2}+\frac{x^2}{2}\geq zx$ 따라서 $x^2+y^2+z^2\geq xy+yz+zx$가 성립한다. 등호가 성립하는 조건은 $x=y=z$이다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

전략 F

과정	scoring rule	인지 요소	풀이 단계
X_1	0		문제상황
	4	A6	$a^2=x,\ b^2=y,\ c^2=z$로 치환한다. 그러면 주어진 부등식은 다음과 같다. $\frac{1}{3}\left(\sqrt{\frac{yz}{x}}+\sqrt{\frac{zx}{y}}+\sqrt{\frac{xy}{z}}\right)\geq\sqrt{\frac{x+y+z}{3}}$
X_2	7	A1, A4, A5	양변을 제곱하여 정리하면 다음의 부등식을 얻는다. $\frac{yz}{x}+\frac{zx}{y}+\frac{xy}{z}\geq x+y+z$ 이를 다시 한번 제곱하여 다음의 부등식을 얻는다. $\left(\frac{yz}{x}+\frac{zx}{y}+\frac{xy}{z}\right)^2\geq(x+y+z)^2$ $\Leftrightarrow\left(\frac{yz}{x}+\frac{zx}{y}+\frac{xy}{z}\right)\left(\frac{yz}{x}+\frac{zx}{y}+\frac{xy}{z}\right)$ $\geq(x+y+z)^2$ 코시-슈바르츠 부등식에 의해 위의 부등식이 성립한다. 등호가 성립하는 조건은 $x=y=z$이다. 원래의 문자로 변환하면 $a=b=c$일 때 등호가 성립한다.

Abstract

A study on the application of Cognitive Diagnostic Assessment for constructed-response items

Byunghong Choi

Department of Mathematics Education

The Graduate School

Seoul National University

Cognitive Diagnostic Assessment (CDA) is an evaluation method that assumes skills or attributes required to solve test items and estimates whether students mastered or not the skills or attributes by using the students' response data. There are various models according to the method of calculating the probability that students get the right answer based on the relationship between items and attributes for CDA, which is called Cognitive Diagnostic Model (CDM). CDMs have been extensively studied for multiple choice or dichotomous response items, but recent advancements have led to the development of CDMs for constructed-response items. One such CDM is the Diagnostic Tree

Model (DTM), which can analyze students' response to constructed-response items by considering the multiple strategies and steps they take to solve a problem. This study aimed to design mathematics assessments using constructed-response items that can be analyzed by the DTM and apply to actual students to analyze the result. Additionally, a simulation study was conducted to examine the impact of several factors on the DTM analysis.

First, this study proposed a framework for structuring the process of solving mathematics constructed-response items. The framework utilizes three factors - attribute, step, and response - to analyze the solution process and determine the number of required attributes, problem-solving complexity, and changes in the number of response caused by the strategy used. The framework can be utilized to design a DTM-based CDA by identifying the expected solution of candidate items related to attributes and selecting test items based on the identification of attributes and test time. Furthermore, the framework can also be used to diagnose students' understanding and interpret the results of the assessment.

Next, the simulation study aimed to investigate the impact of various factors on the performance of DTM-based CDA. The factors considered were the number of examinees, the number of items in the assessment, the number of attributes, the number of steps to solve each item, the type of response, and the size of the item response error probability. The results indicated that an increase in the number of examinees led to a rise in attribute classification accuracy. However, when the number of examinees was small and the error probability was high, the opposite trend was observed. This was particularly evident in assessments with a higher number of items, attributes, steps for each item, and types of response. On the other hand, regardless of the number of examinees,

the accuracy of the attribute mastery pattern was lower in assessments with a larger number of items, attributes, steps for each item, and types of response. The item parameter recovery was better when there were more examinees or a lower probability of item error.

A CDA was designed to evaluate the "absolute inequality" domain using constructed-response items, administered to students from a science high school for the gifted. The students' response data are obtained and analyzed using the DTM incorporating different types of CDM as the response model given a node that represents a subtask in the solution process of an item. The DTM with DINA model, which assumes that examinees should master every attributes required by a subtask in order to solve the task correctly, shows best relative model fit and acceptable absolute model fit. The findings from cognitive diagnosis of examinees indicate that students face more challenges in mastering the Cauchy-Schwarz inequality compared to the arithmetic-geometric mean inequality, with the greatest difficulty being in the application of the substitution method. The analysis revealed that in certain test items, the estimated guess and slip parameters vary depending on the nodes of the path selected by the chosen strategy, indicating the difficulty to solve the problem may depend on the strategy the students choose.

Lastly, students' responses were analyzed to various aspects using the tree structure of the DTM, so it provided the students who took the test an estimated attributes pattern as well as additional feedback. The framework enabled an analysis of students' problem-solving process. Students' learning path can be recommended by presenting attributes and response ratio for each step of problem-solving and a comparison of students' responses with the most likely response and response through addintional attributes acquisition.

keywords : Cognitive diagnostic assessment, Cognitive diagnostic model, Diagnostic tree model, constructed-response items
Student Number : 2020-21008