

DINA 모형에서 응시생 분류 정확성에 영향을 미치는 요인 탐구 : 응시생 분류방법을 중심으로

김지효^{1*}

¹충남대학교 교육학과

A Study on the Factors Affecting Examinee Classification Accuracy under DINA Model : Focused on Examinee Classification Methods

Ji-Hyo Kim^{1*}

¹Division of Education, Chungnam National University

요약 본 연구의 목적은 DINA(deterministic-input, noisy “and” gate)모형에서 최대우도(maximum likelihood: ML), 최대사후확률(maximum a posteriori: MAP), 사후기대(expected a posteriori: EAP)방법들의 분류 정확성이 어느 정도인지를 알아보는 것이다. 연구 목적을 달성하기 위하여 다양한 모의실험 조건들[인지요소의 수(K= 5, 7), 응시생 능력분포(고능력, 중간능력, 저능력 집단), 검사 길이(J= 15, 30, 45)]에 따라 모의자료를 생성했다. 응시생 분류 정확성을 평가하기 위한 준거로 참 인지요소(true α)와 ML, MAP, EAP방법으로 추정된 인지요소가 어느 정도 일치하는지를 계산했다. 본 연구의 주요결과를 요약하면 다음과 같다. 첫째, 본 연구에서 설정한 검사 조건에서 ML, MAP방법보다 EAP방법의 정확일치도 평균이 높았다. 둘째, 다른 검사 조건이 동일할 때, 인지요소의 수가 증가하면 ML, MAP, EAP방법 모두에서 정확일치도 평균이 낮아졌다. 셋째, 동일한 검사 길이에서 사전분포로 고능력, 중간능력, 저능력 집단을 각각 가정했을 때 ML, MAP방법보다 EAP방법의 정확일치도 평균이 높았다. 넷째, 동일한 응시생 능력분포에서 검사 길이가 증가하면 ML, MAP, EAP방법 모두에서 정확일치도 평균이 높아졌다. 인지요소의 수에 따라 응시생을 정확하게 분류하기 위한 적절한 검사 길이를 보면, 인지요소의 수가 5, 7개이고 이에 대응하는 검사 길이가 각각 30, 45문항일 때 본 연구에서 설정한 높은 분류 정확성 기준에 부합하는 것으로 나타났다.

Abstract The purpose of this study was to examine the classification accuracies of ML, MAP, and EAP methods under DINA model. For this purpose, this study examined the classification accuracies of the classification methods under the various conditions: the number of attributes, the ability distribution of examinees, and test length. To accomplish this purpose, this study used a simulation method. For the simulation study, data was simulated under the various simulation conditions including the number of attributes (K= 5, 7), the ability distribution of examinees (high, middle, low), and test length (J= 15, 30, 45). Additionally, the percent of agreements between true skill patterns(true α) and skill patterns estimated by the ML, MAP, and EAP methods were calculated. The summary of the main results of this study is as follows: First, When the number of attributes was 5 and 7, the EAP method showed relatively higher average in the percent of exact agreement than the ML and MAP methods. Second, under the same conditions, as the number of attributes increased, the average percent of exact agreement decreased in ML, MAP, and EAP methods. Third, when the prior distribution of examinees ability was different from low to high under the conditions of the same test length, the EAP method showed relatively higher average in the percent of exact agreement than those of the ML and MAP methods. Fourth, the average percent of exact agreement increased in all methods, ML, MAP, and EAP when the test length increased from 15 to 30 and 45 under the conditions of the same the ability distribution of examinees.

Key Words : Cognitive diagnosis, DINA model, Classification accuracy, Classification methods.

본 논문은 저자의 학위논문의 일부를 발췌하여 수정 보완한 것임.

*Corresponding Author : Ji-Hyo Kim(Chungnam National Univ.)

Tel: +82-10-7264-0010 email: yujin8403@hanmail.net

Received July 24, 2013

Revised August 6, 2013

Accepted August 7, 2013

1. 서 론

교육평가의 목적은 학생의 능력을 파악할 뿐만 아니라 학습의 결과가 기대했던 수준에 도달하지 못하였다면, 그 원인이 무엇인지를 파악함으로써 적절한 진단적 정보를 제공하는 것이다[22]. 기존의 고전검사이론이나 일차원 문항반응이론에 의해 산출된 점수(총점 또는 능력)는 전반적인 성취수준에 대한 정보를 제공한다. 그러나 이러한 총점기반 점수는 학생이 어떠한 학습요소를 성취하는데 어려움이 있는지에 대한 구체적인 진단적 정보를 제공하지 못하는 한계점이 있다. 반면, 인지진단이론은 상기 두 이론과 같이 점수를 이용하는 것이 아니라 문항에 정답을 하는데 필요한 특정 인지요소(attribute)에 대하여 응시생이 숙달했는지를 알려준다[10, 12, 28]. 인지요소숙달 여부에 대한 구체적인 정보는 학생에 대한 정확한 진단뿐만 아니라 학생이 지닌 장·단점을 파악하여 교수·학습에 직접적으로 활용할 수 있다[10, 28]. 인지진단모형은 여러 학자들에 의해서 약62개 이상이 개발되었지만[11], DINA모형은 다른 인지진단모형에 비해 추정해야 할 모수 수가 적고 모수의 해석이 용이하기 때문에 여러 연구들에서 사용되고 있다[6,8,14,15,18,19]. [14]는 DINA모형을 이용하여 응시생 분류 방법의 분류 정확성을 연구하였으나 검사 조건을 더욱 상세화하여 알아볼 필요가 있다. 따라서 본 연구에서는 DINA모형을 이용하기로 한다.

인지진단모형에서 응시생을 잠재체층에 분류하는 방법으로 최대우도(maximum likelihood: ML), 최대사후확률(maximum a posteriori: MAP), 사후기대(expected a posteriori: EAP)방법이 있다[14,26]. 응시생 분류방법들간 응시생의 인지요소숙달 여부가 일치하면 어떠한 분류 방법을 선택하여 분석하더라도 응시생은 동일한 결과를 제공받게 된다. 그러나 응시생 분류방법에 따라 일부 응시생의 인지요소 숙달여부가 다르게 판정되기도 한다[1]. 따라서 어떠한 분류방법이 응시생을 상대적으로 더 정확하게 분류하는지에 대해 알아보는 것은 실제적으로 매우 중요하다. 선택한 분류방법의 분류 정확성이 낮을 경우 응시생은 자신이 갖고 있는 인지요소숙달 여부와는 다르게 분류될 수 있다. 이렇게 잘못 분류된 응시생은 사실상 불필요한 교정교육을 받거나 필요한 교정교육을 받지 못할 수 있다. 또한, 실제 상황에서 검사 조건은 매우 다양하다. 그러므로 어떠한 검사 조건에서든 간에 분류 정확성이 높은 우수한 분류방법이 있는지, 아니면 검사 조건에 따라 분류 정확성이 높은 분류방법이 달라지는지에 대해 알아보는 것은 중요하다. 따라서 다양한 모의실험 조건에서 분류방법들의 분류 정확성이 어느 정도인가에 대하여 알아본다면 이에 대한 정보를 얻을 수 있을 것이다.

인지진단모형에서 응시생 분류 정확성을 평가하는 모의실험 연구가 진행되었다[3,4,9,14,25]. [3]은 DINA모형과 K-means군집분석, HACA에서 인지요소의 수($K=3, 4$), 검사 길이($J=20, 40, 80$), 표본크기($N=100, 500$), 문항모수 값(작음, 큼)에 따른 분류 정확성을 제시했다. [3]은 검사 길이에 따른 분류 정확성을 알아보았으나 비교적 인지요소의 수가 적을 때에 대해서만 제시하고 있다. 그러나 인지진단모형을 적용하여 자료를 분석할 때, 인지요소의 수는 더욱 다양할 수 있다. 실제로 인지진단모형을 적용한 선행연구에서 인지요소의 수는 3~14개[20,21,27]로 다양하게 나타났다. 또한, [5]는 인지요소의 수가 많을 때, 인지요소의 숙달여부를 정확하게 추정하기 위해서 더 많은 검사 길이가 필요하다고 하였다. 실제적으로 인지진단모형을 적용하는데 유용한 정보를 제공하기 위해서는 인지요소의 수에 따른 적절한 검사 길이를 탐색하는 것이 필요하다. 따라서 인지요소의 수를 다양하게 고려하고 인지요소 수에 따라 응시생을 정확하게 분류하는 검사 길이를 알아보려고 한다.

[14]는 DINA모형에서 응시생 분류방법에 따른 분류 정확성을 알아보았다. 즉, 인지요소의 수($K=4, 6$), 문항모수 값(작음, 큼), 응시생 능력분포(균일분포, 고능력 집단)를 달리한 검사 조건들에서 응시생을 상대적으로 더 정확하게 분류하는 방법을 탐구했다. 그러나 [14]는 사전분포로 균일분포와 고능력 집단만을 가정하였는데 한계가 있다. 실제 교육현장에서 시행되고 있는 검사들은 검사 목적에 따라 응시생의 능력분포가 편포를 이룰 수도 있다. 또한, 일반적으로 사전분포로 정규분포를 가정하는데, 검사 종류에 따라 응시생 능력 수준을 고려하여 사전분포를 가정하는 것이 요구된다. 따라서 본 연구에서는 고능력, 중간능력, 저능력 집단을 고려하고 응시생 능력 분포에 따라 분류 정확성이 높은 분류방법을 알아보려고 한다.

본 연구의 목적은 DINA모형에서 인지요소의 수, 응시생 능력분포, 검사 길이를 달리하였을 때 ML, MAP, EAP방법의 분류 정확성이 어느 정도인가에 대하여 알아보는 것이다. 본 연구의 목적에 따른 연구문제를 구체적으로 진술하면 다음과 같다.

첫째, 본 연구에서 설정한 검사 조건에서 분류방법들의 분류 정확성은 어느 정도인가?

둘째, 인지요소 수의 변화에 따른 분류방법들의 분류 정확성은 어느 정도인가?

셋째, 응시생 능력분포의 변화에 따른 분류방법들의 분류 정확성은 어느 정도인가?

넷째, 검사 길이의 변화에 따른 분류방법들의 분류 정확성은 어느 정도인가?

2. 이론적 배경

2.1 DINA모형

DINA모형[8,15]은 결합적 규칙을 따르는 비보상적 모형이다. 즉, 문항이 요구하는 인지요소를 빠짐없이 숙달한 학생은 정답을 할 것이라고 기대하지만, 요구하는 인지요소들 중 하나라도 숙달하지 못하면 오답할 것으로 기대한다. 그러므로 각 인지요소들이 동등한 비중을 지니는 검사자료에 적용하는 것이 적절하다.

DINA모형에서 응시생이 정답을 맞힐 확률을 구하기 위해서는 3가지 구성요소를 요구한다. 첫 번째 구성요소는 잠재변수인 η_{ij} 이며 식 2.1과 같이 정의된다.

$$\eta_{ij} = \prod_{k=1}^K (\alpha_{ik})^{q_{jk}} \quad (\text{식 2.1})$$

η_{ij} 는 응시생 i 가 문항 j 에 요구되는 모든 인지요소들을 숙달했는지 그렇지 않은지에 대한 이분법적 값을 의미한다. 모든 인지요소를 숙달했다면 $\eta_{ij}=1$ 이 되고, 인지요소들 중에서 하나라도 숙달하지 못했다면 $\eta_{ij}=0$ 이 된다. 그러므로 $\eta_{ij}=0$ 인 집단에 속하는 응시생들이 서로 다른 인지요소패턴을 갖고 있을지라도 변별이 불가능하다. q_{jk} 는 문항 j 에 정답을 맞히기 위해서 인지요소 k 의 숙달이 요구되면 1, 아니면 0이 된다. α_{ik} 는 응시생 i 의 인지요소 k 에 대한 숙달 여부로, 인지요소 k 를 숙달하였으면 $\alpha_{ik}=1$ 이 되며, 인지요소 k 를 숙달하지 못하였으면 $\alpha_{ik}=0$ 이 된다. 즉, 식 2.1은 응시생 i 가 문항 j 의 정답을 하려면 그 문항에 정답을 하는데 필요한 모든 인지요소를 갖추고 있어야 하며, 그 가운데 하나라도 숙달하지 못할 경우 오답을 한다는 것을 의미한다.

하지만 학생의 실제문항반응은 잡음요인(noisy)으로 인하여 기대되는 반응과 다른 결과로 나타날 수 있다. 즉, 추측하여 문항에 정답을 하거나 실수로 문항에 오답을 하는 경우가 있을 수 있다. DINA모형은 추측모수(g_j)와 부주의 오류모수(s_j)을 고려하여 정답을 할 확률을 구한다. 추측모수는 문항이 요구하는 모든 인지요소들을 숙달하지 않았는데 정답을 맞힐 확률과 Q행렬에 포함되지 않은 또 다른 인지요소를 사용하였을 가능성까지 포함한다 [8]. 반면에 부주의 오류모수는 문항 j 가 요구하는 모든 인지요소들을 숙달했음에도 불구하고 오답을 한 경우를 말한다. 다음 식 2.2와 2.3은 DINA모형에서 추측모수와 부주의 오류모수를 수식으로 나타낸 것이다.

$$g_j = P(X_{ij} = 1 | \eta_{ij} = 0) \quad (\text{식 2.2})$$

$$s_j = P(X_{ij} = 0 | \eta_{ij} = 1) \quad (\text{식 2.3})$$

다음 식 2.4는 응시생 i 의 잠재변수 η_{ij} 와 문항 j 의 추측모수, 부주의 오류모수를 동시에 고려할 때 문항 j 에 정답을 할 확률을 나타낸다. X_{ij} 는 문항 $J(j=1, \dots, J)$ 에 대하여 i 번째 응시생의 응답반응이며, 정·오답 반응은 1 또는 0으로 표시된다.

$$P(X_{ij} = 1 | \eta_{ij}) = g_j^{(1-\eta_{ij})} (1-s_j)^{\eta_{ij}} \quad (\text{식 2.4})$$

만약에 모든 인지요소를 숙달하였다면($\eta_{ij} = 1$), $(1-s_j)$ 가 정답을 맞힐 확률이 되며, 하나 이상의 인지요소가 결핍을 보일 경우($\eta_{ij} = 0$), g_j 가 정답을 맞힐 확률이 된다.

2.2 응시생 분류방법

응시생을 잠재계층에 분류하는 방법은 ML, MAP, EAP방법이 있으며[14,26], DINA모형에서 응시생 분류방법에 대해 기술하면 다음과 같다. 아래의 내용은 [14]의 내용들을 정리한 것이다.

2.2.1 ML방법

DINA모형에서 응시생이 인지요소패턴 α 를 갖는다고 가정할 때, 응시생 i 의 1부터 $J(j=1, \dots, J)$ 까지 문항에 대한 응답반응패턴의 우도[2]는 식 2.5와 같다. 인지요소의 수가 K 라고 했을 때, 가능한 인지요소패턴은 $L = 2^K$ 개가 되며 인지요소패턴 각각을 $\alpha_l (l=1 \dots L)$ 이라고 했을 때, 응시생은 가능한 인지요소패턴 중 하나의 α_l 을 갖게 된다. 이때, 응시생이 참(true)인지요소패턴을 모르기 때문에 추정된 인지요소패턴 $\hat{\alpha}$ 가 할당된다.

$$L(X_i | \alpha) = \prod_{j=1}^J [(1-s_j)^{X_{ij}} s_j^{1-X_{ij}}] \eta_{ij}^{X_{ij}} [g_j^{X_{ij}} (1-g_j)^{1-X_{ij}}]^{1-\eta_{ij}} \quad (\text{식 2.5})$$

ML방법은 각 인지요소패턴 α_l 에 대하여 응시생의 응답반응벡터를 이용하여 우도를 산출하며 응시생은 우도가 최대가 되는 추정된 인지요소패턴 $\hat{\alpha}_{MLE}$ 에 할당되어 분류된다.

$$\hat{\alpha}_{MLE} = \arg\max L(X_i|\alpha_i) \quad (\text{식 2.6})$$

2.2.2 MAP방법

이전의 검사 결과로부터 각 인지요소패턴에 대하여 응시생들이 갖는 비율에 대한 아이디어가 있을 수 있는데, 이를 사전분포로 가정할 수 있다. MAP방법은 일종의 베이저안 추정법으로 응시생별로 인지요소패턴의 사후분포를 구할 때 사전분포를 우도에 결합하여 사후분포를 최대화하는 모수치를 찾는 방법이다.

식 2.7에서 $P(\alpha_i|X_i)$ 는 응시생 i 의 모든 가능한 인지요소패턴의 사후분포를 나타내며, 오른쪽 항은 각 인지요소패턴의 우도에 사전분포 $P(\alpha_i)$ (단, $\sum_{l=1}^L P(\alpha_l) = 1$)를 곱한 것을 모든 가능한 인지요소패턴의 우도와 사전분포를 각각 곱한 것 모두를 합한 것으로 나눈 것이다.

$$P(\alpha_i|X_i) = \frac{L(X_i|\alpha_i)P(\alpha_i)}{\sum_{m=1}^L L(X_i|\alpha_m)P(\alpha_m)} \quad (\text{식 2.7})$$

식 2.8을 통해 사후분포가 최대가 되는 인지요소패턴 $\hat{\alpha}_{MAP}$ 에 할당되어 분류된다.

$$\hat{\alpha}_{MAP} = \arg\max P(\alpha_i|X_i) \quad (\text{식 2.8})$$

ML방법과 MAP방법은 인지요소별 속달여부를 추정하지 않고 모든 가능한 인지요소패턴 중에서 우도 또는 사후분포가 최대가 되는 인지요소패턴을 찾는다. 두 분류방법의 차이점은 ML방법은 사전분포를 고려하지 않지만 MAP방법은 사전분포를 고려한다는 점이다. 다시 말해, 균일분포 또는 아무런 정보가 없는 사전분포를 사용하는 경우 ML방법과 MAP방법에서 각각 얻은 모수치는 같다 [14]. ML방법과 MAP방법은 각 인지요소들에 대하여 직접적인 확률을 추정해주지 못하기 때문에 각 인지요소에 대한 해석을 하는데 어려움이 있다. 반면, EAP방법은 응시생들에게 각각의 인지요소들에 대한 확률을 제공해줄 수 있는 장점이 있다[26].

2.2.3 EAP방법

ML방법은 인지요소패턴 중에서 우도의 최빈치를 찾아내며, MAP방법은 인지요소패턴의 사후분포의 최빈치를 찾아낸다. 그러나 EAP방법은 인지요소패턴이 아닌 각 인지요소 α 에 대한 확률을 제공해 주는 방법이다. EAP방법에서 사후확률 $P(\alpha_i|X_i)$ 는 MAP방법으로 모든 인지

요소패턴($l=1...L$)에 대하여 산출된 것이며, 각 인지요소의 주변확률(marginal probability)은 식 2.8에 따라 산출된다. 이때, 모든 인지요소패턴의 사후확률의 합은 1이 된다[14]. $I(\alpha_{l,k} = 1)$ 는 l 번째 인지요소패턴의 k 번째 인지요소가 1이면 '1' 그렇지 않으면 '0'으로 이분법적 지표로 나타난다. 예를 들어, 1, 3, 7, 9번째 인지요소패턴만 인지요소1에서 1이고, 나머지 인지요소패턴은 인지요소1에서 0이면, 1, 3, 7, 9번째 인지요소패턴의 사후확률 값만을 더한다. 이 값이 바로 인지요소1의 주변확률이다[1]. 즉, 주변확률 $\tilde{\alpha}_k$ 는 인지요소 k 를 속달한 인지요소패턴의 사후확률 $P(\alpha_i|X_i)$ 을 모두 합한 것이다.

$$\tilde{\alpha}_k = \sum_{l=1}^L P(\alpha_l|X_i)I(\alpha_{l,k} = 1) \quad (\text{식 2.9})$$

$$\hat{\alpha}_{EAP} = 1 \begin{cases} \tilde{\alpha}_k \geq .5 \\ \tilde{\alpha}_k < .5 \end{cases} \quad (\text{식 2.10})$$

식 2.10와 같이, 각각의 주변확률 $\tilde{\alpha}_k$ 는 인지요소속달 여부를 판정하기 위해 추가로 속달여부의 기준이 되는 값을 정해야 한다. 속달여부를 판단하는 절대적인 기준 값은 없지만 일반적으로 .5를 기준으로 인지요소의 주변확률이 .5보다 크면 인지요소를 속달했다고 판정하며, .5보다 작으면 인지요소를 속달하지 못했다고 결론을 내린다.

3. 연구방법

3.1 모의실험 조건

본 연구에서 설정한 검사 조건은 인지요소의 수, 응시생 능력분포, 검사 길이이다. 표본크기는 문항모수 추정 정확성을 고려하여 1,000명[9]으로 하였다.

인지요소를 추출하는 과정에서 불필요한 인지요소를 지양하고 최소한의 인지요소만을 추출하여 모수의 개수를 최소화하는 것은 인지진단모형의 간결성 원칙을 충족시킬 수 있다[13]. 예컨대, 인지요소가 너무 많으면, 과도한 정보로 인하여 학생, 교사, 학부모에게 혼란을 줄 수 있으며 이는 향후 지도에 어려움으로 연결된다. 반대로 인지요소의 수가 너무 적다면 기존에 제공받았던 교육과정에 근거한 내용 분류와 차별성을 갖지 못하며, 학생이 어려움을 겪고 있는 하위 학습내용에 대한 구체적인 정보를 얻지 못한다. 따라서 본 연구에서는 인지진단모형을 활용한 선행연구[20,21]에서 사용된 인지요소의 수를 고려하고 교육현장에서 시행되고 있는 검사 유형에서 일반

적으로 추출될 수 있는 인지요소의 수를 고려하여 5, 7개 두 가지로 조건화하였다.

[14]는 사전분포로 균일분포와 고능력 집단을 가정한 검사 조건에서 분류방법에 따른 분류 정확성에 대하여 연구하였다. 그러나 실제 교육현장에서 시행되고 있는 다양한 검사(국가수준 학업성취도 평가, 영재판별 검사, 진단평가 등)를 볼 때, 응시생 능력수준을 고려하여 사전분포를 정하고 이에 따른 분류 정확성을 알아볼 필요가 있다. 또한, 일반적으로 응시생 모수를 추정할 때 사전분포로 정규분포를 가정하는데 검사의 종류에 따라 응시생 집단의 특정한 능력으로 편포될 수 있다. 예를 들어, 영재판별 검사는 응시생 대부분의 능력이 높다. 따라서 본 연구에서는 사전분포를 고능력, 중간능력, 저능력 집단으로 조건화하였다.

[5]는 인지요소의 수가 많을 때 인지요소의 숙달여부를 정확하게 추정하기 위해서는 더 많은 검사 길이가 필요하다고 하였다. 그러나 교육현장에서 시험시간은 제한되어 있고, 제한된 시간 안에 풀 수 있는 문제는 한정되어 있다. 그러므로 본 연구에서는 교육현장에서 시행될 수 있는 검사의 길이를 고려하여, 검사 길이를 15, 30, 45 문항으로 조건화하였다.

본 연구는 18개의 검사 조건[인지요소의 수(K= 5, 7) × 응시생 능력분포(고능력, 중간능력, 저능력 집단) × 검사 길이(J= 15, 30, 45)]에 따라 DINA모형을 분석하였다. 생성된 자료의 수는 1,800(1개 모형 × 2개 인지요소 수 × 3개 응시생 능력분포 × 3개 검사 길이 × 100회 반복)개이다.

3.2 자료 생성

3.2.1 Q행렬 생성

Q행렬은 인지요소의 수(K=5, 7)와 검사 길이(J=15, 30, 45)의 가능한 조합에 대하여 다음의 두 가지 조건을 고정하여 생성하였다. 첫 번째 조건은 각 문항에 정답을 하는데 요구되는 인지요소의 수를 고정하였다. 이때, 한 개의 문항에 5개 이상의 인지요소가 연결되면 인지진단 모형의 분석이 복잡해져 추정의 오차가 커질 수 있으므로 [17] 한 개의 문항에 연결되는 인지요소의 수는 5개 미만으로 하였다. 즉, 인지요소의 수가 5개일 때는 각각의 문항이 요구하는 인지요소의 수가 1, 2, 3개인 인지요소 패턴을 할당하였다. 인지요소의 수가 7개일 때는 각각의 문항이 요구하는 인지요소의 수가 1, 2, 3, 4개인 인지요소 패턴을 할당하였다. 두 번째 조건은 각각의 인지요소를 측정하는 문항의 수를 고정하였다. 즉, 검사 길이가 15, 30, 45문항일 때, 이에 대응하는 각각의 인지요소를 측정하는 문항의 수를 7, 14, 21개로 고정하였다. Table 1은

인지요소의 수가 5개이고 검사 길이가 15문항일 때, 모의 생성된 Q행렬을 제시하였다.

[Table 1] Q-Matrix for the 15 Item Test

item	Attribute				
	A1	A2	A3	A4	A5
1	0	1	0	0	0
2	0	0	1	0	0
3	1	0	0	0	0
4	0	1	0	0	1
5	1	0	0	1	0
6	0	0	1	1	0
7	0	1	1	0	0
8	1	1	0	1	0
9	0	1	1	0	1
10	1	0	1	0	1
11	1	1	0	1	0
12	0	0	1	1	1
13	1	0	0	1	1
14	0	1	1	0	1
15	1	0	0	1	1

3.2.2 문항모수 생성

[9]는 작은 문항모수(추측모수, 부주의 오류모수) 값의 범위는 .05~.15로 큰 문항모수 값의 범위는 .20~.30으로 보고하였다. [6]은 DINA모형에서 덧셈·뺄셈검사 자료가 적합할 때, 추측모수와 부주의 오류 모수는 각각 .105, .138로 나타났다고 하였다. [14]는 작은 문항모수의 값을 .05~.30사이로 정하고 큰 문항모수의 값을 .20~.45사이로 정하였으며, 문항모수 값이 작을수록 분류 정확성이 높았다고 하였다. [6,9]의 연구에서 제시된 문항모수 값은 매우 작아 실제상황에서 극히 드물게 나타날 수 있다. 따라서 본 연구에서는 문항모수 값을 [14]가 제시한 낮은 문항모수 값인 .05~.30사이로 정하고 이 범위에서 문항모수를 무선 할당하였다. 본 연구에서 설정한 검사 조건에서 문항모수의 평균은 Table 2와 같다.

[Table 2] Mean of item parameters(guess, slip)

Population	Test length	K = 5		K = 7	
		guess	slip	guess	slip
High	15	.1801	.1964	.1620	.1860
	30	.1917	.1912	.1935	.1906
	45	.1578	.1708	.1715	.1674
Middle	15	.1848	.1817	.1478	.1787
	30	.1789	.1919	.1645	.1598
	45	.1650	.1643	.1834	.1652
Low	15	.1469	.1734	.1759	.1719
	30	.1667	.1904	.1593	.1669
	45	.1634	.1653	.1941	.1671

3.2.3 응시생 능력분포 생성

인지요소가 5, 7개일 때, 응시생 능력분포(고능력, 중간능력, 저능력 집단)에 따른 인지요소패턴의 비율은 Table 3과 같다. 응시생 능력분포는 인지요소패턴의 숙달된 인지요소 수를 기준으로 비율을 달리하여 생성하였다. 예컨대, 인지요소의 수가 5개일 때, 고능력 집단은 숙달된 인지요소의 수가 많은 인지요소패턴일수록 높은 비율을 갖도록 생성하였다. 각 인지요소패턴의 비율을 보면 숙달된 인지요소의 수가 0, 1개인 인지요소패턴 각각의 비율인 .0208을 기준으로 인지요소의 수가 2, 3개인 인지요소패턴 각각의 비율은 $.0208 \times 1.5 = .0312$ 이며, 인지요소의 수가 4, 5개인 인지요소패턴 각각의 비율은 $.0208 \times 2 = .0416$ 이다. 중간능력 집단은 최대한 대칭분포가 되도록 생성하였다. 숙달된 인지요소의 수가 0, 1, 5개인 인지요소패턴 각각의 비율이 .0208로 가장 낮았다. 숙달된 인지요소의 수가 2, 4개인 인지요소패턴 각각의 비율은 $.0208 \times 1.5 = .0312$ 이며, 숙달된 인지요소의 수가 3개인 인지요소패턴 각각의 비율은 $.0208 \times 2 = .0416$ 이다. 저능력 집단은 숙달된 인지요소가 적은 인지요소패턴일수록 높은 비율을 갖도록 생성하였다. 각 인지요소패턴의 비율을 보면, 숙달된 인지요소의 수가 4, 5개인 인지요소패턴 각각의 비율인 .0208을 기준으로 인지요소의 수가 2, 3개인 인지요소패턴 각각의 비율은 $.0208 \times 1.5 = .0312$ 이며, 인지요소의 수가 4, 5개인 인지요소패턴 각각의 비율은 $.0208 \times 2 = .0416$ 이다. Table 3에서 제시한 바와 같이, 고·저능력 집단과 중간능력 집단은 응시생 비율을 할당하는 기준인 인지요소패턴의 숙달된 인지요소 수에 차이가 있어, 중간능력 집단의 숙달된 인지요소 수는 괄호 안에 제시하였다.

3.3 자료생성 및 분석 절차

본 연구를 수행하기 위한 모의실험 절차는 다음과 같다.

1단계: 모의실험의 조건에 따라 Q행렬을 모의생성하고, 균일분포에서 문항모수(추측모수, 부주의 오류모수)를 .05~.30사이에서 무선적으로 생성한다. 이를 참(true) 문항모수 값으로 여긴다.

2단계: 각 응시생에 대해 무선 생성된 참 인지요소패턴(true α)을 부여한다.

3단계: 참 인지요소패턴(true α)과 Q행렬을 사용하여 η_{ij} 값을 구한다.

4단계: η_{ij} 값과 문항모수를 사용하여 첫 번째 문항에 대하여 정답확률을 구한다.

5단계: 정답확률을 버노울리 시행(Bernoulli trial)을 통하여 정답반응 (0, 1)으로 생성한다.

6단계: 3단계에서 5단계까지 모든 문항에 대해서 반복한다. 2단계부터 6단계까지를 각 개인별 모의자료로 생성한다.

7단계: ML, MAP, EAP방법으로 각 모의인의 인지요소에 대한 숙달여부를 판정한다.

8단계: 응시생 분류 정확성 통계치(정확일치도)를 계산한다.

9단계: 1단계부터 8단계까지를 다른 인지요소의 수, 응시생 능력분포, 검사 길이를 달리하여 반복한다.

위의 절차는 R-software version 2.15.1(R Development Core Team, 2012)을 사용하여 분석하였다.

[Table 3] Generating Probabilities for the High, Middle and Low Ability Examinee Populations for Simulation

Population		the number of attributes mastered								
		4, 5(0, 1, 5)			2, 3(3)			0, 1(2, 4)		
		rate	number	total	rate	number	total	rate	number	total
K=5	High	.0416	6	.2496	.0312	20	.6240	.0208	6	.1248
	Middle	.0208	7	.1456	.0416	10	.4160	.0312	15	.4680
	Low	.0208	6	.1248	.0312	20	.6240	.0416	6	.2496
Population		the number of attributes mastered								
		5, 6, 7(0, 1, 6, 7)			3, 4(3, 4)			0, 1, 2(2, 5)		
		rate	number	total	rate	number	total	rate	number	total
K=7	High	.0104	29	.3016	.0078	70	.5460	.0052	29	.1508
	Middle	.0052	16	.0832	.0104	70	.7280	.0078	42	.3276
	Low	.0052	29	.1508	.0078	70	.5460	.0104	29	.3016

4. 연구결과

4.1 응시생 분류방법의 정확성

인지요소의 수가 5, 7개일 때, 응시생 능력분포와 검사 길이의 변화에 따른 세 가지 분류방법의 인지요소별 정확일치도와 전체 인지요소의 정확일치도 평균을 Table 4와 Table 5에 제시했다. 검사 조건에 따라 정확일치도 평균이 높은 분류방법을 살펴보면 다음과 같다. Table 4와 Table 5에서 알 수 있듯이, 인지요소의 수가 5, 7개이고 고능력, 중간능력, 저능력 집단이며 검사 길이가 15, 30, 45문항일 때, 모든 인지요소에서 ML, MAP방법보다 EAP방법의 정확일치도가 높은 것으로 나타났다.

Table 4와 Table 5에서 주목할 점은 인지요소의 수에 따라 분류방법에 따른 정확일치도 평균의 차이가 달랐다는 점이다. 본 연구에서 설정한 검사 조건에서 인지요소의 수가 5개일 때, 분류방법에 따른 정확일치도 평균의

차이는 .0005~.0088, 인지요소의 수가 7개일 일 때, 분류방법에 따른 정확일치도 평균의 차이는 .0016~.0274로 나타났다. 이러한 결과는 인지요소의 수가 증가하면 분류방법에 따른 정확일치도 평균의 차이도 커지는 것을 보여준다.

각 인지요소의 정확일치도간 차이는 인지요소의 수가 5개일 때보다 7개일 때 더 커진 것으로 나타났다. 예컨대, 동일한 검사 조건(고능력 집단, J=15)에서 인지요소의 수가 5개일 때, 인지요소간 정확일치도의 차이를 보면 ML방법은 .0016~.1035, MAP방법은 .0017~.1038, EAP방법은 .0001~.0985로 나타났다. 한편, 인지요소의 수가 7개일 때, 인지요소간 정확일치도의 차이를 보면 ML방법은 .0021~.2253, MAP방법은 .0014~.1877, EAP방법은 .0052~.1931로 나타났다. 이러한 결과는 인지요소의 수가 증가하면 인지요소간 정확일치도 평균의 차이가 커지는 것을 보여준다.

[Table 4] Percent of Exact Agreement Between MLE/MAP/EAP and True Skill Patterns, by Individual Skills and Total N*K Skills for K=5

()standard deviation

Popula tion	Test length	Method	A1	A2	A3	A4	A5	Total N*K
High	15	MLE	.8609 (.0404)	.8724 (.0328)	.8708 (.0335)	.7777 (.0248)	.7689 (.0217)	.8302 (.0306)
		MAP	.8610 (.0401)	.8727 (.0328)	.8710 (.0334)	.7737 (.0234)	.7689 (.0217)	.8295 (.0303)
		EAP	.8657 (.0388)	.8766 (.0312)	.8765 (.0317)	.7841 (.0231)	.7781 (.0217)	.8362 (.0293)
	30	MLE	.9127 (.0251)	.9392 (.0215)	.9345 (.0213)	.8292 (.0169)	.8226 (.0174)	.8877 (.0204)
		MAP	.9127 (.0252)	.9393 (.0215)	.9347 (.0211)	.8301 (.0171)	.8283 (.0165)	.8890 (.0203)
		EAP	.9193 (.0223)	.9425 (.0195)	.9389 (.0178)	.8414 (.0153)	.8376 (.0154)	.8959 (.0181)
	45	MLE	.9526 (.0162)	.9671 (.0109)	.9669 (.0116)	.8565 (.0112)	.8540 (.0115)	.9194 (.0123)
		MAP	.9527 (.0162)	.9671 (.0109)	.9669 (.0115)	.8645 (.0143)	.8657 (.0119)	.9234 (.0130)
		EAP	.9539 (.0159)	.9679 (.0103)	.9683 (.0106)	.8697 (.0136)	.8695 (.0111)	.9259 (.0123)
Middle	15	MLE	.8584 (.0406)	.8674 (.0331)	.8685 (.0364)	.7644 (.0230)	.7592 (.0250)	.8236 (.0316)
		MAP	.8574 (.0404)	.8655 (.0331)	.8680 (.0363)	.7663 (.0227)	.7654 (.0246)	.8245 (.0314)
		EAP	.8631 (.0393)	.8701 (.0312)	.8726 (.0355)	.7764 (.0215)	.7723 (.0228)	.8309 (.0301)
	30	MLE	.9145 (.0242)	.9342 (.0196)	.9340 (.0189)	.8205 (.0172)	.8195 (.0181)	.8845 (.0196)
		MAP	.9140 (.0243)	.9338 (.0198)	.9338 (.0188)	.8249 (.0184)	.8246 (.0170)	.8862 (.0197)
		EAP	.9187 (.0222)	.9381 (.0179)	.9380 (.0171)	.8383 (.0164)	.8332 (.0159)	.8933 (.0179)
	45	MLE	.9538 (.0155)	.9679 (.0116)	.9668 (.0124)	.8491 (.0126)	.8535 (.0138)	.9182 (.0132)
		MAP	.9537 (.0153)	.9680 (.0117)	.9669 (.0122)	.8644 (.0137)	.8661 (.0133)	.9238 (.0132)
		EAP	.9542 (.0154)	.9684 (.0120)	.9677 (.0118)	.8688 (.0137)	.8694 (.0124)	.9257 (.0131)
Low	15	MLE	.8342 (.0452)	.8464 (.0423)	.8382 (.0429)	.7492 (.0257)	.7395 (.0233)	.8015 (.0359)
		MAP	.8374 (.0444)	.8510 (.0403)	.8427 (.0399)	.7510 (.0254)	.7390 (.0233)	.8042 (.0347)
		EAP	.8424 (.0429)	.8551 (.0382)	.8497 (.0380)	.7544 (.0245)	.7427 (.0231)	.8088 (.0333)
	30	MLE	.8923 (.0332)	.9230 (.0254)	.9186 (.0251)	.8015 (.0207)	.8007 (.0164)	.8672 (.0242)
		MAP	.8965 (.0301)	.9249 (.0240)	.9212 (.0239)	.8043 (.0194)	.8003 (.0154)	.8694 (.0226)
		EAP	.9050 (.0262)	.9300 (.0210)	.9281 (.0206)	.8075 (.0185)	.8020 (.0158)	.8745 (.0204)
	45	MLE	.9450 (.0193)	.9585 (.0145)	.9581 (.0166)	.8318 (.0121)	.8295 (.0151)	.9046 (.0155)
		MAP	.9458 (.0189)	.9592 (.0143)	.9590 (.0162)	.8321 (.0118)	.8293 (.0152)	.9051 (.0153)
		EAP	.9481 (.0179)	.9605 (.0139)	.9610 (.0156)	.8331 (.0124)	.8300 (.0151)	.9065 (.0150)

4.2 인지요소의 수에 따른 분류 정확성

Table 4와 Table 5에서 제시한 바와 같이, 동일한 응시생 능력분포와 검사 길이에서 인지요소의 수가 증가하면 정확일치도 평균은 감소하는 것으로 나타났다. 예컨대, 고능력 집단이고 검사 길이가 15문항일 때, ML방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .8302, 인지요소의 수가 7개 일 때는 .7110으로 나타났다. MAP방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .8295, 인지요소의 수가 7개일 때는 .7244로 나타났다. EAP방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .8362, 인지요소의 수가 7개일 때는 .7340으로 나타났다. 고능력 집단이고 검사 길이가 30문항일 때, ML방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .8877, 인지요소의 수가 7개 일 때는 .7591로 나타났다. MAP방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .8890, 인지요소의 수가 7개일 때는 .7767로 나타났다. EAP방법의 정확일치도 평균은 인지요소의 수가 5개

일 때는 .8959, 인지요소의 수가 7개일 때는 .7863으로 나타났다. 고능력 집단이고 검사 길이가 45문항일 때, ML방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .9194, 인지요소의 수가 7개 일 때는 .7916으로 나타났다. MAP방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .9234, 인지요소의 수가 7개일 때는 .8143으로 나타났다. EAP방법의 정확일치도 평균은 인지요소의 수가 5개일 때는 .9259, 인지요소의 수가 7개일 때는 .8188로 나타났다. 이러한 현상은 중간능력, 저능력 집단이고 검사 길이가 15, 30, 45문항일 때도 유사하게 나타났다.

분류 정확성으로 정확일치도 평균 이외의 표준편차는 분류방법간 분류 정확성이 얼마나 안정적인가를 평가하는데 중요한 지수이다. 왜냐하면 반복되는 추정으로부터 분산이 큰 것은 분류방법간 정확일치도 평균에 대한 변화가 크다는 것을 의미하기 때문이다. 그러므로 표준편차를 작게 추정하는 검사 조건이 무엇인지를 알아보는 것이 필요하다. Table 4와 Table 5에서 제시한 바와 같이,

[Table 5] Percent of Exact Agreement Between MLE/MAP/EAP and True Skill Patterns, by Individual Skills and Total N*K Skills for K=7

Test length	Method	()standard deviation							Total N*K		
		A1	A2	A3	A4	A5	A6	A7			
H	15	MLE	.6104 (.0162)	.8357 (.0482)	.6660 (.0223)	.7439 (.0297)	.6526 (.0184)	.6547 (.0189)	.8135 (.0499)	.7110 (.0291)	
		MAP	.6500 (.0211)	.8377 (.0492)	.7095 (.0269)	.7443 (.0299)	.6562 (.0197)	.6548 (.0184)	.8182 (.0498)	.7244 (.0307)	
		EAP	.6564 (.0189)	.8495 (.0433)	.7163 (.0260)	.7543 (.0287)	.6662 (.0179)	.6610 (.0185)	.8346 (.0455)	.7340 (.0284)	
	30	MLE	.6349 (.0184)	.9032 (.0291)	.6916 (.0178)	.8039 (.0228)	.6941 (.0183)	.6957 (.0178)	.8904 (.0364)	.7591 (.0229)	
		MAP	.6898 (.0163)	.9042 (.0285)	.7520 (.0209)	.8046 (.0224)	.6975 (.0166)	.6972 (.0181)	.8919 (.0366)	.7767 (.0228)	
		EAP	.6987 (.0175)	.9124 (.0246)	.7642 (.0184)	.8159 (.0205)	.7105 (.0175)	.7013 (.0174)	.9007 (.0301)	.7863 (.0209)	
	45	MLE	.6465 (.0158)	.9464 (.0185)	.7143 (.0164)	.8496 (.0178)	.7231 (.0158)	.7275 (.0142)	.9336 (.0236)	.7916 (.0174)	
		MAP	.7242 (.0155)	.9466 (.0186)	.7918 (.0164)	.8495 (.0178)	.7263 (.0172)	.7277 (.0138)	.9343 (.0233)	.8143 (.0175)	
		EAP	.7266 (.0159)	.9503 (.0171)	.7947 (.0165)	.8540 (.0176)	.7363 (.0163)	.7307 (.0143)	.9389 (.0222)	.8188 (.0171)	
	M	15	MLE	.6221 (.0206)	.8252 (.0550)	.6774 (.0291)	.7190 (.0330)	.6273 (.0209)	.6243 (.0203)	.8156 (.0585)	.7015 (.0339)
			MAP	.5799 (.0160)	.8275 (.0546)	.6326 (.0226)	.7212 (.0320)	.6308 (.0205)	.6267 (.0195)	.8188 (.0584)	.6911 (.0319)
			EAP	.6286 (.0207)	.8431 (.0481)	.6863 (.0297)	.7341 (.0308)	.6380 (.0185)	.6302 (.0210)	.8342 (.0531)	.7135 (.0317)
30		MLE	.6609 (.0177)	.8973 (.0297)	.7290 (.0266)	.7828 (.0276)	.6653 (.0152)	.6675 (.0161)	.8676 (.0404)	.7529 (.0248)	
		MAP	.6058 (.0169)	.8980 (.0292)	.6647 (.0205)	.7841 (.0268)	.6692 (.0162)	.6683 (.0158)	.8697 (.0393)	.7371 (.0235)	
		EAP	.6708 (.0180)	.9069 (.0250)	.7422 (.0216)	.7958 (.0251)	.6783 (.0155)	.6722 (.0156)	.8818 (.0346)	.7640 (.0222)	
45		MLE	.6921 (.0159)	.9427 (.0173)	.7665 (.0177)	.8262 (.0150)	.6899 (.0166)	.6938 (.0157)	.9248 (.0262)	.7908 (.0178)	
		MAP	.6139 (.0172)	.9429 (.0173)	.6823 (.0156)	.8264 (.0148)	.6964 (.0158)	.6938 (.0158)	.9254 (.0261)	.7687 (.0175)	
		EAP	.6947 (.0160)	.9472 (.0159)	.7712 (.0170)	.8310 (.0152)	.7033 (.0156)	.6949 (.0154)	.9303 (.0244)	.7961 (.0171)	
L		15	MLE	.6182 (.0207)	.8203 (.0563)	.6725 (.0301)	.7070 (.0283)	.6330 (.0218)	.6279 (.0231)	.8094 (.0592)	.6983 (.0342)
			MAP	.6219 (.0202)	.8270 (.0538)	.6756 (.0292)	.7143 (.0267)	.6372 (.0208)	.6308 (.0221)	.8196 (.0565)	.7038 (.0328)
			EAP	.6295 (.0199)	.8394 (.0491)	.6885 (.0279)	.7309 (.0270)	.6427 (.0198)	.6371 (.0206)	.8325 (.0517)	.7144 (.0309)
	30	MLE	.6571 (.0204)	.8873 (.0325)	.7162 (.0243)	.7638 (.0234)	.6713 (.0177)	.6677 (.0179)	.8723 (.0414)	.7480 (.0254)	
		MAP	.6661 (.0175)	.8897 (.0328)	.7251 (.0214)	.7710 (.0216)	.6772 (.0169)	.6737 (.0154)	.8769 (.0392)	.7542 (.0236)	
		EAP	.6765 (.0167)	.9020 (.0270)	.7390 (.0185)	.7893 (.0198)	.6822 (.0157)	.6796 (.0149)	.8879 (.0327)	.7652 (.0208)	
	45	MLE	.6909 (.0119)	.9370 (.0196)	.7611 (.0142)	.8133 (.0144)	.6982 (.0108)	.6975 (.0103)	.9307 (.0242)	.7898 (.0151)	
		MAP	.6927 (.0117)	.9381 (.0191)	.7634 (.0142)	.8152 (.0142)	.6992 (.0105)	.6987 (.0104)	.9323 (.0237)	.7914 (.0148)	
		EAP	.6979 (.0108)	.9427 (.0178)	.7698 (.0128)	.8227 (.0136)	.7011 (.0105)	.7013 (.0099)	.9363 (.0224)	.7960 (.0140)	

다른 검사 조건이 동일할 때, 인지요소의 수가 5, 7개로 증가하면 정확일치도 평균의 표준편차는 증가하는 것으로 나타났다. 예컨대, 고능력 집단이고 검사 길이가 30문항일 때, EAP방법의 정확일치도 평균의 표준편차는 인지요소의 수가 5개일 때는 .0181, 인지요소의 수가 7개일 때는 .0209로 나타났다.

4.3 응시생 능력분포에 따른 분류 정확성

인지요소의 수와 검사 길이를 고정시키고 사전분포를 고능력, 중간능력, 저능력 집단으로 달리했을 때, 분류방법에 따른 인지요소별 정확일치도와 전체 인지요소의 정확일치도 평균은 Table 4와 Table 5와 같다. 다른 검사 조건이 동일할 때, 사전분포에 따라 정확일치도 평균이 상대적으로 더 높은 분류방법을 살펴보면 다음과 같다. 인지요소의 수가 5개이고 검사 길이가 15, 30, 45문항일 때, 고능력, 중간능력, 저능력 집단 모두에서 정확일치도 평균이 가장 높은 분류방법은 EAP방법으로 나타났으며, ML, MAP방법의 정확일치도 평균은 거의 차이가 없는 것으로 나타났다. 한편, 인지요소의 수가 7개이고 검사 길이가 15, 30, 45문항일 때, 응시생 능력분포에 따라 정확일치도 평균이 높은 분류방법은 달라졌다. 즉, 다른 검사 조건이 동일할 때, 고능력 집단과 저능력 집단은 EAP, MAP, ML방법 순으로 정확일치도 평균이 높았으나, 중간능력 집단은 EAP, ML, MAP방법 순으로 정확일치도 평균이 높은 것으로 나타났다.

4.4 검사 길이에 따른 응시생 분류 정확성

인지요소의 수가 5, 7개일 때, 응시생 능력분포를 고정시키고 검사 길이를 15, 30, 45문항으로 달리하였을 때, 분류방법에 따른 인지요소별 정확일치도와 전체 인지요소의 정확일치도 평균은 Table 4와 Table 5에 제시되어 있다. Table 4와 Table 5에서 알 수 있듯이, 인지요소의 수가 5, 7개이고 응시생 능력분포가 고능력, 중간능력, 저능력 집단일 때, 검사 길이가 15, 30, 45문항으로 증가하면 정확일치도 평균은 증가했다. 예컨대, Table 4에서 알 수 있듯이, 인지요소의 수가 5개이고 고능력 집단일 때, 검사 길이가 15문항에서 45문항으로 증가하면 ML방법은 .8302에서 .9194, MAP방법은 .8295에서 .9234, EAP방법은 .8362에서 .9259로 증가했다. Table 5에서 알 수 있듯이, 인지요소의 수가 7개이고 고능력 집단일 때, 검사 길이가 15문항에서 45문항으로 증가하면 ML방법은 .7110에서 .7916, MAP방법은 .7244에서 .8143, EAP방법은 .7340에서 .8188으로 증가하는 것으로 나타났다. 이러한 현상은 인지요소의 수가 5, 7개이고 고능력, 중간능력,

저능력 집단일 때도 유사하게 나타났다. 이러한 연구 결과를 인지진단모형에 적용, 분석하는데 유용한 정보를 얻으려면, 조건에 따른 변화를 살펴보는 것뿐만 아니라 응시생을 정확하게 분류하는데 필요한 검사 길이를 판단하는 것이 중요하다. 검사 신뢰도의 경우 검사 전문가들에 의해 대략적으로 합의된 준거(표준화 검사는 .8이상, 교실검사는 .6이상)가 존재하지만, 심리측정이론을 통하여 계산된 분류 정확성의 경우 어느 정도의 값이 받아들여질 만만가에 대한 대략적 합의는 존재하지 않는 것으로 보인다[16]. DINA모형의 응시생 분류방법간 일관성 추정에서 [1]은 분류방법간 일관성을 인정하는 절대적인 기준은 없으나 방법간 일치도는 높아야 하므로 정확일치도는 .8이상을 제안하고 있다. 본 연구 또한 분류 정확성을 판단하는 기준을 정확일치도가 .8이상일 때로 정하고자 한다.

인지요소별로 검사 길에 따른 각 인지요소의 정확일치도를 살펴보면 다음과 같다. 인지요소의 수가 5개이고 고능력, 중간능력, 저능력 집단일 때 검사 길이가 15문항이면 인지요소 4, 5의 정확일치도가 .8미만으로 나타났으나 검사 길이가 30문항 이상이 되면 모든 인지요소의 정확일치도가 .8이상으로 나타났다. 인지요소의 수가 7개이고 고능력, 중간능력, 저능력 집단일 때, 검사 길이가 15, 30, 45문항으로 증가하면 모든 인지요소의 정확일치도는 증가하나, 검사 길이가 45문항일 때도 몇몇 인지요소의 정확일치도는 .8미만으로 낮은 것으로 나타났다.

5. 논의 및 결론

본 연구의 목적은 모의실험 연구를 통하여 DINA모형에서 인지요소의 수, 응시생 능력분포, 검사 길이를 달리했을 때, ML, MAP, EAP방법들의 분류 정확성 정도를 알아보는데 있었다. 본 절에서는 도출된 연구결과를 선행 연구와 관련지어 논의하기로 한다.

첫째, 인지요소의 수가 5, 7개일 때, 고능력, 중간능력, 저능력 집단이고 검사 길이가 15, 30, 45문항일 때, ML, MAP방법보다 EAP방법의 정확일치도 평균이 더 높은 것으로 나타났다. 이러한 결과는 [14]의 연구결과와 일치한다. 따라서 ML, MAP방법보다 EAP방법을 사용하기를 권장하나 국가 수준과 같은 대규모 시험에서 응시생에게 인지진단 정보를 제공하고자 할 때는 분류방법으로 한 가지를 선택했다 하더라도 다른 분류방법으로 함께 인지요소숙달 여부를 비교하는 것이 요구된다[1]. 응시생 분류방법에 따른 인지요소숙달 여부 판정이 불일치하는 응시생들에게는 측정하고자 하는 인지요소가 요구되는 문항을 추가로 제시하여 인지요소 숙달 여부를 재확인하는

작업이 필요하다. 이를 위해서는 인지진단모형에 적용이 가능한 진단적 정보를 갖추고 있는 문제는행을 구축할 필요가 있다.

인지요소의 수가 증가하면 인지요소별 정확일치도의 차이는 증가했다. 즉, 인지요소별 정확일치도의 차이는 인지요소의 수가 5개일 때는 13% 미만, 인지요소의 수가 7개일 때는 34% 미만으로 나타났다. 이러한 결과는 [14]의 연구에서 발견된 인지요소별 분류 정확성의 차이보다 큰 것으로 나타났다. 이는 본 연구에서 Q행렬을 모의 생성할 때, 인지요소의 수가 증가하면 문항에 정답을 맞히는데 필요한 인지요소의 수도 증가하도록 조건화하였기 때문으로 보인다. [13]은 한 개의 문항에 너무 많은 인지요소가 연결되면 인지진단모형의 분석이 복잡해져서 추정 오차가 커질 수 있다고 하였다. 따라서 한 문항에 연결되는 인지요소의 수가 분류 정확성에 어떠한 영향을 미치는지에 대한 추후 연구가 필요해 보인다.

둘째, 동일한 응시생 능력분포, 검사 길이에서 인지요소의 수가 증가하면 정확일치도 평균은 낮아졌다. 이러한 결과는 [3,4]의 연구결과와 일치한다. 따라서 인지요소의 수에 따라 어떠한 검사 조건에서 응시생을 정확하게 분류하는지를 고려할 필요가 있다. 반면, [14]는 인지요소의 수가 4개일 때보다 인지요소의 수가 6개일 때, 모든 인지요소에서 분류 정확성이 높았다고 보고하였다. 이들은 인지요소의 수가 증가하면 검사 길이도 증가하도록 Q행렬을 생성했기 때문에 인지요소의 수뿐만 아니라 검사 길이도 분류 정확성에 영향을 미친 것으로 판단된다.

응시생 분류방법에 따른 정확일치도 평균의 차이는 인지요소의 수가 증가할수록 커졌다. 분류방법에 따른 분류 정확성의 차이를 인지요소별로 살펴보면, 인지요소의 수가 5개일 때는 1% 미만, 인지요소의 수가 7개일 때는 3% 미만으로 나타났다. 따라서 인지요소의 수가 증가할수록 선택한 분류방법에 따라 분류 정확성의 차이가 커짐을 시사한다.

셋째, DINA모형에서 인지요소의 수가 5개일 때, 사전분포로 고능력, 중간능력, 저능력 집단을 각각 가정할 경우 대체로 ML, MAP방법보다 EAP방법의 정확일치도 평균이 높았다. 이러한 연구 결과는 [14]가 사전분포로 균일분포와 고능력 집단을 가정했을 때, ML, MAP방법보다 EAP방법이 응시생을 더욱 정확하게 분류한다고 보고한 결과와 일치한다.

DINA모형에서 인지요소의 수가 7개일 때, 동일한 검사 길이에서 사전분포로 고능력, 저능력 집단을 가정했을 때, 대체로 EAP, MAP, ML방법 순으로 정확일치도 평균이 높았다. 그러나 사전분포로 중간능력 집단을 가정했을 때, EAP, ML, MAP방법 순으로 정확일치도 평균이

높았다. 이러한 결과는 응시생의 인지요소숙달 여부를 세 가지 분류방법에서 비교할 때, 사전분포에 따라 상대적으로 정확한 분류방법을 선택하는데 유용한 정보로 제공될 수 있을 것이다. 즉, 인지요소의 수가 7개일 때, 사전분포로 고능력 집단 또는 저능력 집단을 가정했을 경우, 응시생의 인지요소숙달 여부는 ML방법보다 MAP, EAP방법이 더욱 정확할 것으로 보인다. 그러나 사전분포로 중간능력 집단을 가정했을 경우, MAP방법보다 ML, EAP방법으로 추정된 인지요소숙달 여부가 더 정확할 것으로 보인다.

넷째, DINA모형에서 인지요소의 수가 5, 7개일 때, 동일한 응시생 능력분포에서 검사 길이가 증가하면 분류 정확성은 높아졌다. 이러한 결과는 [3,5]의 연구결과와 일치한다. [5]가 언급한 바와 같이, 인지요소의 수가 증가하면 인지요소를 측정하는 문항의 수도 증가해야 모든 인지요소의 숙달 여부가 정확하게 판단된다는 것을 알 수 있다. 그러나 교육현장에서는 시간과 비용 등의 이유로 검사 길이를 늘이는 것 보다는 적절한 Q행렬을 개발하는 것이 중요해 보인다.

결론적으로 응시생 분류방법에 따라 분류 정확성에 차이가 있는데, 본 연구결과와 [14]의 연구 결과를 종합해 볼 때, 다른 두 방법이 비해서 EAP방법이 응시생을 정확하게 분류한다. 또한, 검사가 측정하는 인지요소의 수에 따라 적절한 검사 길이를 고려해야 한다. 본 연구 결과에 따르면 인지요소의 수가 5개일 때는 30문항 이상, 인지요소의 수가 7개일 때는 최소한 45문항 이상이 요구된다.

본 연구에서는 DINA모형을 활용할 때, 다양한 모의실험 조건에서 분류방법들의 분류 정확성을 알아보았다. 그러나 여러 다른 인지진단모형에 적용하여 분류방법들의 분류 정확성을 알아보는 것이 필요하다. 본 연구에서는 이분형 응답자료를 분석할 수 있는 인지진단모형을 사용하였지만 다분형 응답자료의 분석이 가능한 BIN, Full NC-RUM, GDM의 분류 정확성을 밝히는 연구가 필요하다. 본 연구에서는 응시생의 분류 정확성에 영향을 미치는 요인으로 인지요소의 수, 표본크기, 검사 길이, 응시생 능력분포에 대하여 알아보았다. 그러나 분류 정확성에 영향을 줄 수 있는 다른 요인들로 이루어진 다양한 검사 조건을 생성하고 응시생을 정확하게 분류하기 위한 검사 조건을 탐색하는 것이 요구된다. 즉, 분류 정확성에 영향을 미칠 수 있는 다른 요인으로 모형의 적합성, 다양한 인지요소의 관계(위계적, 고차원적), Q행렬의 적절성, 각 문항에 연결되는 인지요소의 수에 대하여 연구할 필요가 있다.

References

- [1] Ban, J. C., Kim, S. (2012). A Comparison of Skill Mastery Estimation Methods for DINA and Changes of Skill Masteries of Elementary Students. *Journal of Educational Evaluation*, 25(4), 721-744.
- [2] Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74(4), 619-632.
DOI: <http://dx.doi.org/10.1007/s11336-009-9123-2>
- [3] Chiu, C. Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74(4), 633-665.
DOI: <http://dx.doi.org/10.1007/s11336-009-9125-0>
- [4] Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49(1), 19-38.
DOI: <http://dx.doi.org/10.1111/j.1745-3984.2011.00158.x>
- [5] 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), 343-363.
DOI: <http://dx.doi.org/10.1111/j.1745-3984.2008.00069.x>
- [6] de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1), 115-130.
DOI: <http://dx.doi.org/10.3102/1076998607309474>
- [7] de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179-199.
DOI: <http://dx.doi.org/10.1007/s11336-011-9207-7>
- [8] de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333-353.
DOI: <http://dx.doi.org/10.1007/BF02295640>
- [9] de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47(2), 227-249.
DOI: <http://dx.doi.org/10.1111/j.1745-3984.2010.00110.x>
- [10] DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment*(pp. 361-390). Hillsdale, NJ: Lawrence Erlbaum Associates.
- [11] Fu, J., & Li, Y. (2007, April). An integrative review of cognitively diagnostic psychometric models. Paper presented at the annual meeting of the National Council on Measurement in Education. Chicago, IL.
- [12] Hartz, S. (2002). A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practice. Doctoral thesis, The University of Illinois at Urbana-Champaign.
- [13] Henson, R. A., & Templin, J. L. (2009, April). Q-matrix construction. Paper presented at 2009 NCME training session. San Francisco, CA.
- [14] Huebner, A., & Wang, B. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, 71(2), 407-419.
DOI: <http://dx.doi.org/10.1177/0013164410388832>
- [15] 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.
DOI: <http://dx.doi.org/10.1177/01466210122032064>
- [16] Kang, T. H., Park, C. H., & Kim, I. S. (2011). The effects of test length and performance-level number on classification consistency and accuracy. *Journal of Educational Evaluation*, 24(4), 1017-1038.
- [17] Kim, H. K., Han, J. A., Choi, S. G., Kim, B. M. (2012). The Application of Cognitive Diagnostic Model for Achievement Profile Analysis and Score Report(RRE 2012-7). Seoul: Korea Institute for Curriculum and Evaluation.
- [18] Kim, S. H., Kim, S. J., & Song, M. Y. (2008). Using Cognitive Diagnosis Theory to Analyze the Test Results of Mathematics. *Journal of Korea Society of Educational Studies in Mathematics School Mathematics*. 10(2), 259-277.
- [19] Kim, S. E., Park, Y. S., & Lee, Y. S. (2012). Application of Latent Class Model to Multiple Strategy CDM Analysis. *Journal of Educational Evaluation*, 25(1), 49-68.
- [20] Kim, S. H., Song, M. Y. (2011). Diagnosis of knowledge states using large scale assessments: An application of DINA model. *The Journal of Curriculum and Evaluation*, 14(1), 177-200.
- [21] Kunina, O., Rupp, A. A., & Wilhelm, O. (2008, June 29-July 2). *Convergence of skill profiles for cognitive diagnosis models and other multidimensional scaling approaches: An empirical illustration with a diagnostic mathematics assessment*. Paper presented at the annual international Meeting of the Psychometric Society, Druham, NH.

- [22] Lee, J, S (2009). Present Educational evaluation. Seoul: Kyoyookgwahaksa.
- [23] Lee, Y, S, Park, Y, S, Song, M, Y, Kim, S, E, Lee, Y, J, In, B, R (2012). Investigating Score Reporting of Attribute Profiles from the National Assessment of Educational Achievement using Cognitive Diagnostic Models. *Journal of Educational Evaluation*, 25(3), 411-433.
- [24] R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>.
- [25] Rupp, A. A., & Templin, J. L. (2008a). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68, 78-96.
DOI: <http://dx.doi.org/10.1177/0013164407301545>
- [26] Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and application*. NY: The Guilford press.
- [27] Song, M, Y, Lee, Y, S, & Park, Y, S (2011). Analysis and score reporting based on cognitive diagnostic models using the National Assessment of Educational Achievement(RRE 2011-8). Seoul: Korea Institute for Curriculum and Evaluation.
- [28] Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 435-354.
DOI: <http://dx.doi.org/10.1111/j.1745-3984.1983.tb00212.x>
- [29] Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*. 11(3), 287-305.
DOI: <http://dx.doi.org/10.1037/1082-989X.11.3.287>

김 지 효(Ji-Hyo Kim)

[정회원]



- 2010년 2월 : 숙명여자대학교 대학원 교육심리학 (교육학석사)
- 2013년 8월 : 충남대학교 대학원 교육학 (교육학박사)

<관심분야>
인지진단이론, 문항반응이론