

CBT 시스템에서의 신뢰도 높은 문제 은행 구축을 위한 문항 분석 방법

정희영[†] 윤애선[†], 손건태[‡]
부산대학교 언어정보 연구실[†], 통계학과[‡]

Study of Item Analysis to Improve the Reliability of Item Bank in CBT System

Heeyoung Jeong[†] Aesun Yoon[†], Keontae Sohn[‡]
Language Information Lab[†], and Dept of Statistics[‡],
Pusan national University university

요 약

대규모 수험자를 대상으로 하는 경우, CBT 방식은 PBT 방식에 비해 비용이나 처리 방법에 있어 효율적이라는 장점을 갖고 있다. 하지만, CBT 평가가 제대로 운영되려면 많은 문항으로 구성된 대규모의 문제 은행이 있어야 하며, 기 구축된 문항은 적합성과 공정성이 반드시 보장되어야 한다. 이에 본 연구는 4,000여 명 이상의 수험자를 대상으로 8000문제가 넘는 문제 은행이 구축되어 실제 CBT 시스템을 개발·운영하고 있는 P 대학교 CBLT 테스트 시스템의 문항들을 분석할 방법을 제안한다. 연구 결과 CBLT 테스트 시스템의 문제 은행의 개별 문항들이 출제자 1인의 주관적 판단에 달려있어 문항에 대한 질 관리가 어렵다는 문제점을 해결하기 위해서는 고전검사이론이 적합하다고 판단된다. 그러나 매 시험이 동일한 문항 세트에 구성되기 위해서는 문항에 대한 충분하고 객관적인 자료가 절대적으로 필요하다. 이를 위해서는 CBLT 테스트 시스템과 같이 응답자 수가 2~300여 명 전후이며, 40문항의 문항 세트에 대한 분석은 2모수 정도가 적당하다고 판단된다.

1. 서론

잘 짜여진 CBT(Computer-Based Testing) 시스템에 의한 평가가 유의미한 결과를 도출하기 위해서는 CBT 시스템에서 사용하게 될 검사지의 각 문항에 대한 평가가 이루어지지 않으면 안 된다. 즉, 출제자에 의해 제작된 검사 문항들이 양질의 문항인지 어떤지 평가되고, 계속적으로 피드백 되어 문항의 수정 및 보완이 이루어져야 한다. 뿐만 아니라 모든 응시자들에게 언제나 공정한 평가가 이루어지도록 문제 은행으로부터 추출되는 검사지가 동일한 문항으로 구성되기 위해서는 각 문항에 대한 풍부한 정보가 필요하다.

문항들이 응답한 자료를 가지고 문항에 대한 정

보를 얻을 수 있는 검사이론에는 고전검사이론(Classical test theory)과 문항반응이론(Item response theory)이 있다. 고전검사이론은 1920년대 이후 개발되어 많은 이론적 발전과 더불어 응용되어 왔고, 현재까지 사용되고 있다. 이에 비추어 문항반응이론은 1940년대 이론적 전개를 이루어 1970년대와 1980년대 이르러 많은 발전을 가져왔으며, 국제적으로 실용적 측면에까지 활발한 응용 수준에 있다(성태제, 1998).

이에 본 연구는 실제 CBT 시스템을 개발·운영하고 있는 P대학교 언어교육원의 CBLT(Computer-Based Language testing) 시스템¹⁾의 문항의 질 관리를 위한 문항 분석 방법을

1) 본 연구팀(부산대학교 언어정보 연구실)은 초기부터 이 시스템

고찰해 보고자 한다. H교과목의 경우 8,000문제가 넘는 문제 은행이 구축되었고 2000년 10월부터 시행되어 현재까지 지속적으로 운영해오고 있으므로 시험결과에 대해 상당히 많은 양의 누적 자료를 확보 할 수 있다.

본 연구는 이러한 데이터를 검사이론을 통해 분석 해 봄으로써 CBT 시스템에서의 문항의 질 관리를 위한 문항 분석 방법을 고찰해 보고자 한다.

2. 검사 이론

2.1 고전 검사 이론

고전검사이론(classical test theory)은 검사도구의 총점에 의하여 분석되는 이론으로 검사에 의한 관찰점수(observed score)는 진점수(true score)와 오차점수(error score)에 의하여 합성됨을 가정한다. 또한 피험자의 진점수를 알 수 없기 때문에 이론적으로 동일 검사를 동일 피험자에게 무한히 반복 실시하여 얻은 점수들의 평균점수로 추정한다고 가정한다.

고전검사이론에 구할 수 있는 문항에 대한 정보로 첫째, 문항난이도(item difficulty)를 들 수 있다. 문항 난이도는 문항의 쉽고 어려운 정도를 나타내는 지수로서, 총 피험자 중, 답을 맞춘 피험자의 비율, 즉 확률이 된다. 난이도 지수를 계산하기 위해서는 어떤 테스트 문항에 정답한 수험자의 수를 수험자의 총수로 나누기를 한다. 난이도의 백분율이 높으면 쉬운 문항이고, 낮으면 어려운 문항이다²⁾.

다음으로, 문항변별도(item discrimination)는 문항이 능력에 따라 피험자를 변별하는 정도를 나타내는 지수를 말한다. 능력이 높은 피험자가 문항의 답을 맞히고 능력이 낮은 피험자가 문항의 답을

의 설계 및 개발에 직접 참여하였고, 수험생의 성적 결과 자료에 대한 접근 권한을 갖고 있다. 단, 본 연구의 분석 대상은 실제로 운용되는 교과목이나, 개인정보 보호원칙에 따라 본 고에서는 익명으로 칭한다.

2) Cangelosi(1990)는 문항난이도에 따른 평가기준을 다음과 같이 제시 하였다.

문항난이도	문항평가
.25이하	어려운 문항
.25-.75	적절한 문항
.75이상	쉬운 문항

맞히지 못하였다면 이 문항은 학습자의 능력을 바르게 변별하지 못하는 문항으로 분석된다³⁾.

한편, 진위형 문항이나 선다형 문항에서 문항의 답을 맞힌 피험자 중 추측에 의하여 문항의 답을 맞힌 피험자도 있게 된다. 틀린 문항에 별점을 주지 않는 경우 추측은 검사에서 일어날 수 있는 행위이므로 문항추측도 역시 문항분석의 요소이다.

끝으로 오답지의 매력도가 있다. 선다형 테스트 문항의 경우, 선택지 즉 수험자에게 제시되어 선택해야 할 문항과 유사한 오답 즉 정답이라고 생각되는 오답 선택지가 있다. 오답 선택지는 수험자가 바른 답을 모르는 경우에 정답으로 수험자의 주의를 끄는 역할을 한다. 따라서 선다형 문항에서 답지에 대한 분석은 문항의 질을 향상시키는 중요한 작업이 된다. 뿐만 아니라, 각 선택지의 응답 비율도 문항의 질 향상을 위한 좋은 단서가 된다.

2.2. 문항반응이론

문항반응이론의 특징은 학습자의 능력 모수(ability Parameter)는 그 학습자에게 실시된 특정 테스트 항목과 떨어져 독립해서 구해질 수 있다는 점이다. 문항반응이론에서 요구하는 값을 이용하면 학습자마다 응답하는 테스트 문항이 달라도 상호 비교 가능한 모수를 얻을 수 있다.

즉 학습자 능력 모수는 그것을 추정하기 위해 사용되는 테스트 문항에 따라 그 특수 값으로 결정되는 것이 아니라, 말하자면 불변적으로 결정할 수 있다. 물론 이러한 것이 가능하기 위해서는 Baker(1985)에서 서술하듯이 2개의 조건을 갖추지 않으면 안 된다. 그 하나는 모든 테스트 문항은 학습자 동일 잠재 특성을 측정한다는 것이며, 또 하나는 모든 문항 모수치(Item Parameter)는 동일

3) 문항변별도 지수에 의하여 문항을 평가하는 절대적 기준은 없으나 검사도구의 신뢰도와 관련한 Ebel(1979)의 기준은 다음과 같다.

문항변별도 지수	문항평가
.40 이상	상당히 좋은 테스트 문항
.30에서 .39	보통 수준의 테스트 문항이지만 개선을 필요로 하는 문항
.20에서 .29	겨우 허용할 수 있는 정도로 보통은 개선이 필요한 문항
.19이하	불완전한 테스트 문항으로 사용중지를 하든지 개정을 필요로 하는 문항

척도에서 만들어졌다고 하는 것이다.

또한 문항반응이론을 이용한 분석에서는 동일 테스트 문항을 다른 학습자 집단에게 주어도 그 문항이 가지고 있는 특성에 변화가 없다고 하는 결론을 얻을 수 있다. 즉 문항난이도나 문항변별력은 그 문항이 문항반응이론에서 사용되고 있는 모델에 일치하면 그것을 수험하는 학습자 집단에 관계없이 하나의 불변적인 값으로서 나타낼 수 있다.

2.3. 고전검사이론과 문항반응이론의 비교

양적 평가에서 객관적으로 측정할 수 없는 것들(인지적, 정의적 능력 등)을 검사를 통해 측정하고자 할 때 검사 문항이 개발된다. 그리고 이러한 검사 문항들의 특성을 분석하는 이론을 검사 이론이라 하는데, 고전검사 이론과 현대적 이론인 문항반응이론을 들 수 있다.

고전 검사 이론에서의 문항 난이도와 문항 변별도 등의 분석은 학습자 집단의 능력 범위와 분포에 의해 좌우된다. 반면 문항반응이론은 글자 그대로 문항에 대한 반응의 결과를 가지고 그 문항에 대한 특성을 분석하는 것으로 학습자 집단의 능력 수준을 안다는 가정 하에서 난이도와 변별도에 해당하는 문항 모수치를 추정하는 이론이다. 따라서 문항반응이론은 어떤 한 문항에 대한 변별도와 난이도, 추측도가 통계적 방법으로 계산되기에 집단에 따라 이러한 수치가 변하는 것이 아니라는 장점을 가지고 있다. 즉 어떤 한 문항이 문항반응이론의 기본 가정(일차원성, 지역독립성)을 만족한다면 그 문항은 언제 어디서나 그 문항 고유의 난이도 변별도 추측도 등을 가지고 있다. 다만, 사전에 한 문항 한 문항에 대한 대규모 데이터(학습자 반응 관련 데이터)가 있었어야 하는 단점이 있다.

3. 검사이론을 이용한 문항 분석

CBLT 시스템은 월요일부터 금요일까지 하루 7시간씩 35 분반을 운영하고 수험자는 인터넷을 통해 시험 1주 전에 시험시간과 장소를 예약 후 테스트에 응하게 된다. 이 때 각 분반에 따라 문항 세트는 평가 범위, 평가 영역, 문항 유형, 출제자가 부여한 난이도에 따라 동등한 값을 갖도록 문항을

추출, 분반 별 검사지를 만든다.

즉, CBLT 시스템에서의 문제 은행은 개별 문항의 난이도 및 문항의 적절성이 전적으로 출제자 1인의 주관적 판단에 달려 있다. 따라서 각 분반에 따라 추출되는 문항 세트 역시 난이도와 변별도 등이 고려된 동질 문항 세트라고 보기 어렵다.

이에 CBLT 시스템에서의 문항의 질 관리를 위해서 문제 은행을 구축하고 있는 문항들에 대한 분석이 있어야 하며, 이 분석을 위한 가장 적절한 방법을 고찰하고자 한다.

3.1 중복 문항을 이용한 기초 분석

기초 분석으로 학습자 집단에 따라 문항의 특성이 달라질 수 있는 고전검사이론의 단점 때문에 문항반응이론을 이용하여, 응답 데이터를 분석하려고 하였다. 그러나 매 시간 CBLT에 응한 응답자 수가 50명 내외로 문항반응이론을 이용할 수 없다는 결론을 내렸다. 이에 35개 전체 분반에서 7회 이상의 응답 데이터가 있는 중복 문항을 선별하여 고전 검사이론을 이용 분석하였다.

다음은 2000학년 2학기 CBLT를 이용하여 평가한 H교과목 중복 출제 문항 분석 개요이다.

<표 1> 2000학년도 2학기 H교과목 중복 출제 문항 분석 개요

개요
· 2000학년 2학기 H교과목 5차시 총 63회 2520문항 출제
· 출제 문항 범위 8180~8855(675문항)
· 실제 출제 문항 552문항
· 평균 중복 문항 횟수 448회
· 최고 중복 문항 횟수는 13회 - 고유번호 8297번
· 중복 횟수 7회 이상인 73문항 분석
· 분석 프로그램 고전검사이론에 기초한 TestAn 사용

분석 결과 중복 문항 73문항은 크게 7가지 패턴으로 나눌 수 있었는데, 패턴 및 문항수에 대한 요약은 아래와 같다.

<표 2> 문항 패턴 및 문항 수

패턴	문항수
1. 난이도 적절하고 변별도 높은 질 좋은 문항	13문항(18%)
2. 비교적 쉽지만 변별력 있는 문항	11문항(15%)
3. 비교적 어렵지만 변별력 있는 문항	0문항(0%)
4. 비교적 쉽고 변별도가 떨어지는 문항	18문항(25%)
5. 비교적 어렵고 변별도가 떨어지는 문항	1문항(1%)
6. 난이도는 적절하지만 변별력이 떨어지는 문항	17문항(24%)
7. 난이도도 변별도도 좋지 못한 나쁜 문항	13문항(17%)

2000학년도 2학기 데이터 분석은 고전검사이론으로 분석하였으며 이때 시험에 응시한 학습자 집단의 특성, 예를 들면 자연계열과, 인문계열, 그리고 서양어 전공계열과, 예체능계열 학습자들에 대한 고려를 하지 않았다.

때문에 이 분석이 CBLT 문제 은행의 문항에 대한 완전한 문항 특성 분석이라고는 말하기 어렵다

하지만, 분석 결과에서도 알 수 있듯이 33%를 제외한 67%의 문항에 문제점이 지적되었다. 물론 문항의 오답지를 수정하는 정도의 가벼운 문제점이 대부분이긴 하지만 17%의 문항에 질적 문제의 의문스러운 부분이 있다는 것은 염려스러운 부분이 아닐 수 없다.

요약하면, 40%의 문항이 난이도 면에서 쉬운 것으로 나타났으며, 33% 문항만이 변별력 있는 문항으로 나타났다. 즉, 대부분의 문항이 쉬운 경향이 있으며, 따라서 패턴 3으로 분류하기에 적당한 문항은 없었다. 또한 패턴 4, 5, 6번으로 판정된 문항은 문항의 오답지 수정 등을 통한 보완이 요구되며, 특히 7번과 같은 문항은 검사에서 제외하는 것이 바람직하다.

3.2 검사이론을 이용한 문항 분석

고전검사 이론에 의한 중복 문항을 바탕으로 분석한 결과 질적 문제가 염려스러운 문항이 상당히 있었다. 그러나 앞서도 설명하였지만, 고전검사이론은 검사집단에 영향을 받아, 문항의 난이도 및 변별도가 집단에 따라 달라질 수 있다.

이에 2001학년도 2학기 H교과목은 동일 시간대에 최대한 많은 학생들이 응시할 수 있도록 조절

하여, 고전검사이론 뿐만 아니라 문항반응이론을 이용하여 문항 분석을 하였다.

다음은 2001학년도 2학기 CBLT를 이용한 H교과목 문항 분석 개요이다.

<표 3> 2001학년도 2학기 H교과목 문항 분석 개요

개요
· 2001년 2학기 H교과목 5차시 중 수행자수가 가장 많은 시간대 선정
· 수행자수 464명
· 출제 문항 40문항
· 분석 프로그램 고전검사이론에 기초한 TestAn 사용
· 분석 프로그램 1모수 문항반응이론에 기초한 RaschAn 사용
· 분석 프로그램 2모수 문항반응이론에 기초한 BayesAn 사용

3.2.1 고전 검사이론에 근거한 분석

분석된 결과, 4문항을 제외하면 비교적 난이도 면에서 적절하거나 쉬운 문항들로 구성되었음을 알 수 있다.

그리고 40문항 가운데 16문항, 이 가운데 특히 3문항이 변별력이 없거나 부족한 것으로 나타났으며, 24문항, 특히 8문항은 매우 변별력 높거나 있는 것으로 나타나, 비교적 변별력 있는 검사라고 판단된다. 즉, 40문항 가운데 변별력 있는 문항은 61%로 난이도가 적절한 문항은 56%로 나왔다. 오답지 매력도에 있어서는 문항 30의 오답지 수정이 요구되는 정도로 비교적 양호한 편이었으나, 난이도나 변별도 면에서 보면 30%의 문항이 오답지 수정 등을 통한 문항 보완이 요구되며, 7%로의 문항은 검사에서 제외하는 것이 바람직하고 결론지을 수 있다(표 4 참조).

<표 4> 문항 패턴 및 문항 수

패턴	문항수
1. 난이도 적절하고 변별도 높은 질 좋은 문항	13문항(33%)
2. 비교적 쉽지만 변별력 있는 문항	11문항(28%)
3. 비교적 어렵지만 변별력 있는 문항	0문항(0%)
4. 비교적 쉽고 변별도가 떨어지는 문항	3문항(7%)
5. 비교적 어렵고 변별도가 떨어지는 문항	1문항(2%)
6. 난이도는 적절하지만 변별력이 떨어지는 문항	9문항(23%)
7. 난이도도 변별도도 좋지 못한 나쁜 문항	3문항(7%)

3.2.2. 1모수 문항반응이론에 근거한 분석

문항반응이론에서 1모수 문항반응이론을 의미하는 것으로 문항 변별도와 문항 추측도를 고려하지 않고 문항난이도만 고려하는 모형이다. 이 모형에서 문항 변별도는 모든 문항이 동일하고 문항 변별도는 1, 문항추측도는 0이라고 간주한다.

1모수 모형을 이용한 검사의 분석 결과 13문항, 이 가운데 특히 3문항이 ‘매우 어렵다’, 혹은 ‘어렵다’였으며 나머지 27문항 특히 3문항이 ‘매우 쉽다, 쉽다, 적절하다’로 비교적 쉬운 문항으로 구성된 검사임을 알 수 있다.

3.2.3. 2모수 문항반응이론에 근거한 분석

문항반응이론에서 2모수 문항반응이론을 의미하는 것으로 문항 추측도를 고려하지 않고 문항난이도와 변별도만 고려하는 모형이다. 즉 모든 문항의 문항 추측도는 0이고 각 문항들의 난이도와 변별도가 다를 것이라는 가정을 전제로 한다.

<표 5> 문항 패턴 및 문항 수

패턴	문항수
1. 난이도 적절하고 변별도 높은 질 좋은 문항	4문항(10%)
2. 비교적 쉽지만 변별력 있는 문항	19문항(48%)
3. 비교적 어렵지만 변별력 있는 문항	0문항(0%)
4. 비교적 쉽고 변별도가 떨어지는 문항	6문항(15%)
5. 비교적 어렵고 변별도가 떨어지는 문항	8문항(20%)
6. 난이도는 적절하지만 변별력이 떨어지는 문항	2문항(5%)
7. 난이도도 변별도도 좋지 못한 나쁜 문항	1문항(2%)

2모수 모형의 분석에 따르면 본 검사는 8문항 특히, 6문항이 매우 어렵다, 어렵다 이고 그 외 32문항이 비교적 난이도 면에서 적절하거나 쉬운 문항이 많음을 알 수 있다.

변별도는 17문항, 특히 3문항에서 변별력 거의 없음, 낮음으로 나타났고 23문항에서 적절하거나 높음, 매우 높음으로 나타나 비교적 변별력 있는 검사임을 알 수 있다(표 5 참조).

이상의 분석 결과를 앞서 고전검사이론 및 1모수 문항반응이론을 이용한 결과와 비교하면 난이도 지수가 3가지 방법에서 거의 일치함을 알 수 있다. 하지만 1모수 모형은 2모수 모형에서 모든 문항의 변별도 모수를 1로 가정한 경우로 고전검사와 2모수에서의 다양한 변별도지수를 대변하지 못한다.

4. 결론 및 향후 과제

이상 본 연구는 CBT 시스템에서의 문항의 질 관리를 위한 문항 분석 방법 고찰을 위해 실제 운영 중인 P대학교 CBLT 시스템의 문항 응답 데이터를 검사이론들을 통해 분석 비교하였다.

그 결과 CBLT 시스템의 문제 은행의 개별 문항들이 출제자 1인의 주관적 판단에 달려있어 문항에 대한 질 관리가 어렵다는 문제점을 해결하기 위해서는 고전검사이론이 적합하다고 판단된다.

즉, 고전 검사 이론에서의 난이도 및 변별도 등에 의해 극단적인 결과 값이 나오는 문항들은 삭제 및 수정을 위한 객관적 자료가 될 수 있다. 또한 고전 검사이론에 근거한 오답지의 매력도 측정 결과는 양질의 선택지 제작을 위한 중요한 자료가 될 수 있다.

그러나 CBLT 시스템과 같이 35분반의 35개의 검사지가 동질한 문항 세트로 구성되기 위해서는 문항에 대한 충분하고 객관적인 자료가 절대적으로 필요하다. 이를 위해서는 문항반응이론에 의한 문항 분석이 적합한데, 이 경우 문항을 분석하기 위해서는 ‘많은 수의 문항과 매우 많은 수의 응답자’가 필요하다. 그리고 이들 많은 수에 대한 구체적인 실험적 연구 결과는 현재 거의 이루어지지 않은 상태이다.

CBLT 시스템의 경우 한 검사지당 200~300여

명이 최대 응답자 수이며, 검사지의 문항은 40문항으로 IRT에서 요구되는 매우 많은 응답자수는 아니다.

따라서 객관식 문제에서 생길 수 있는 추측도를 고려한 3모수를 적용하기에는 응답자 수가 작다. 그러므로 앞서 연구 결과 등을 종합해 볼 때, CBLT 시스템과 같이 응답자 수가 2~300여 명 전후의 소규모 그룹이며, 40문항의 문항 세트에 대한 분석은 난이도와 변별도를 포함하는 2모수 정도가 적당하다고 판단된다.

끝으로 P대학교 언어교육 연구원에서는 CBLT 시스템에 문항 반응이론에 근거한 상용 프로그램인 BILOG를 연동시킬 계획에 있는데, 이에 따른 문항 분석 결과 및 동질 문항 추출방법 등은 이후 연구 과제로 남긴다.

참고 문헌

- [1] 윤애선·정희영. “컴퓨터기반 언어 테스트 시스템의 구축을 위한 요소 분석 및 그 적용에 관한 연구,” 정보통신부 학술 연구 과제, 2002.
- [2] 성태제. “문항제작 및 분석의 이론과 실제,” 학지사, 1998.
- [3] 성태제. “문항반응이론의 이해와 적용,” 교육과학사, 2001.
- [4] 大友賢二. “問項應答理論入門,” 大修館書店, 1996.
- [5] Baker, F. B. “*The Basics of Item Response Theory*,” Heinemann, 1985.
- [6] Brown, J. D. “*Testing in language programs*,” New Jersey, Prince Hall, pp.1-20, 1996.
- [7] Cangelosi, J. S. “*Designing tests for evaluating student achievement*,” New York, Longman, 1990.
- [8] Ebel, R. L. “*Essentials of educational measurement(3rd ed.)*” Englewood Gliffs, NJ: Prentice-Hall, 1979.
- [9] Hambleton, R. K. & Swaminathan, H. “*Item Response Theory : Principles and Applications*,” Kluwer-Nijhoff Publishing,

1985.

- [10] Ronald K. Hambleton, H. Swaminathan, H. Jane Rogers. “*Fundamentals of Item Response Theory*,” California : SAGE, 1991.

* 본 연구는 한국정보통신부 정보 통신 개발 연구 사업의 연구 결과 중 일부임