

# 한국어능력시험 읽기 영역 난이도 설정에 대한 비판적 검토

- 제25회~제27회 중급 읽기 영역을 중심으로 -

함정식\* · 김명광\*\*

## || 차례 ||

1. 서론
2. 연구 동향 및 연구 방법
3. 한국어능력시험의 중급 읽기 영역 문항 분석
4. 결론

### 【국문초록】

이 글은 제25회~제27회 한국어능력시험 중급 읽기 영역의 문항 양호도 분석과 오답 매력도가 높은 문항에 대한 내용 분석을 통해 시험의 안정성을 평가하고 평가틀 개선을 위한 방향을 탐색하기 위하여 수행되었다. 이 연구에서 사용된 자료는 제25회~제27회 한국어능력시험 국내 지역 응시자의 중급 읽기 영역 문항 반응 결과였으며 회차별로 사례수는 제25회 8,461건, 제26회 9,029건, 제27회 10,106건이었다. 이들 자료에 대해 고전검사이론에 의한 문항 분석을 실시하기 위해 TestAn 1.0을 사용하였다. 제25회~제27회 한국어능력시험 중급 읽기 영역의 문항 분석 결과를 요약하면 3회 시험 모두 안정적인 평균 난이도와 변별도를 보이며 검사 신뢰도 역시 양호한 것으로 파악되었다. 그러나 문항 단위 분석 결과, 4급 합격자와 불합격자로 상하위 집단을 구분하였을 때, 동 시험이 가진 문항 난이도의 문제가 분명히 드러났다. 또 오답 매력도가 높은 문항에 대한 내용 분석 결과 동 시험의 평가 기능 중 일부는 난이도를 높이기 위해 해당 수준에 맞지 않거나

\* 서울시 교육연구정보원 교육연구사

\*\* 대구대학교 국제한국어교육과 교수

지나치게 작위적인 방법으로 문항을 출제하는 것으로 드러났다. 이상의 분석 결과를 바탕으로 해당 수준에 맞지 않는 너무 쉽거나 어려운 문항을 배제해야 하며, 출제진의 전문성과 출제 체제의 개선을 위해 노력할 필요가 있고, 회차 간 동등화 방안의 마련이 필요하다는 점 등을 차후 평가틀 개선을 위한 시사점을 제시하였다.

주제어 : 한국어능력시험(TOPIK), 중급, 읽기 영역, 평가틀, 양호도 분석

## 1. 서론

이 연구의 목적은 제25회~제27회 한국어능력시험 중급 읽기 영역의 문항 양호도 분석과 오답 매력도가 높은 문항에 대한 내용 분석을 통해 시험의 안정성을 평가하고 평가틀의 개선 방법을 모색하는데 있다.

일반적으로 평가틀(test framework)이란 검사가 측정하고자 하는 내용을 효과적이고도 정확하게 평가할 수 있도록 여러 평가 요소를 정선하고 범주화한 것을 의미한다. 따라서 평가틀은 문항의 여러 속성-주제나 소재, 기능, 난이도, 문항 유형 등을 규정함으로써 문항을 제작하는 중요한 지침의 구실을 하므로 사실상 평가의 성패를 좌우한다고 볼 수 있다. 이에 비해 시험 체제란 평가틀을 포함하면서 등급이나 수준 설정, 동등화 방법, 합격 기준이나 점수 기준을 포함하는 것을 말한다.<sup>1)</sup>

1) 평가틀은 test specification(검사 명세서)라고도 할 수 있는데, 검사에 포함되는 문항들의 속성을 구체화 하는 것을 말한다. Millman과 Greene(1989)는 평가틀 제작을 위해 외부 맥락적 요인들인 검사의 목적, 응시자 속성, 검사 횟수, 검사 관리 등을 검토해야 한다고 주장하였다. 또 평가틀 제작을 위한 내부 맥락적 요인들로는 검사가 다루어야 할 내용의 범위, 문항 유형, 검사 내 문항 수, 문항에 요구되는 심리측정학적 특성, 문항의 평가와 선제 방법, 문항의 배열이나 개발 방법, 문항과 검사 점수 체제, 검사가 요구하는 심리적 특성 등이 규정되어야 함을 주장했다.

한국어능력시험에서는 김하수 등(1997)의 연구나 김정숙 등(2004), 이해영(2006)의 연구에서 한국어능력시험의 시험 체제와 평가들에 대한 논의가 있었다. 특히 현행 한국어능력시험의 평가들과 시험 체제를 사실상 규정하고 있는 이해영(2006)의 연구에서는 한국어능력시험의 종류와 등급제, 합격 기준 등의 시험 체제와 함께 영역별 평가들을 규정하고 있는데, 이는 현행 한국어능력시험의 평가들의 토대로 작동하고 있다는 데에 주목할 만하다.

한국어능력시험(S-TOPIK)의 시험 체제는 1997년 제1회 시험이 한국학술진흥재단(현 한국연구재단) 주관으로 시행된 이래, 크고 작은 변화를 거치면서 기존의 6종 6개 등급 체제에서 3종 6개 등급으로 정착되어 시행되고 있다. 그 과정에서 여러 차례의 크고 작은 시험 체제와 평가들의 변화가 있었는데, 최종적인 평가들의 변화는 2011년 국립국제교육원으로 이관된 후 출제진을 중심으로 진행되었던 쓰기 영역 중심의 변화였다. 그러나 한국어능력시험에서 이루어진 크고 작은 여러 차례의 시험 체제 변화는 사실상 시험의 신뢰도나 타당도보다 실용도에 초점이 두어졌다고 볼 수 있다. 즉 평가들의 개선이 내용 타당화나 문항의 질적 개선보다는 시행과 채점의 용이성 및 운영비용의 감소에 초점을 두고 이루어졌다는 비판을 면키 어렵다. 이러한 이유로 한국어능력시험에 대한 비판적 검토 연구들은 한결 같이 평가들에서의 문제점을 지적하곤 한다.

평가들에 대한 이론적 분석 연구들로는 김정숙 등(2004)과 김정숙 등(2011)의 연구들이 있다. 이들 연구에서 읽기 평가들은 주제와 소재, 평가 기능, 텍스트의 종류, 문항의 유형 등으로 구체화되고 있다. 이론적 분석 연구들은 읽기 영역 평가들의 체계화에 기여했다는 긍정적 판단을 내릴 수는 있으나 이론에 대한 실증은 다소 부족하였다고 평가될 수 있다.

한국어능력시험이 국립국제교육원으로 이관된 후, 국립국제교육원에서

는 시험의 시행으로 확보된 문항반응결과를 이용하는 정책연구를 직접 발주하거나 문항반응 데이터를 제공하여 외부 연구진의 연구를 지원하고 있다. 이러한 지원의 일환으로 실시된 정책연구에서 양길석 등(2012)은 구조방정식 모형을 통해 영역 설정의 적절성과 평가 영역별 하위 범주의 적절성을 평가하고, 중급과 고급시험에서의 다차원성 문제와 영역별 하위 범주의 차별성에 대해 이론적 근거가 부족함을 지적하였다. 특히 읽기 영역과 관련하여 이상린(2011)은 20회~22회 고급 읽기 영역 문항 반응 결과를 IRT(문항반응이론)에 입각하여 분석하고, 난이도의 안정성 부족, 일부 문항들에서의 난이도 변동성 문제 등을 밝혀냈다.

이러한 실증적 분석 연구들을 통해 한국어능력시험의 문항 난이도와 변별도, 검사의 신뢰도 등이 기대에 미치지 못한다는 밝혀졌지만, 이에 대한 원인 분석은 새로운 국면의 연구를 기대하게 한다. 환언하자면 검사 또는 문항 난이도의 변동의 문제는 표피적인 문제일 뿐이고, 이러한 문제가 발생하게 된 데에는 시험체제(점수 체제, 동등화)나 평가틀(속성 분류의 부정확성, 평가 기능상의 오류 등), 출제 방식(출제진 구성, 검토 과정), 참조 자료(어휘 목록) 등의 문제가 복합적으로 작용할 수 있다.

이에 본고에서는 한국어능력시험 중급 읽기 영역의 난이도를 분석하여 회차 간 난이도 차이가 발생하고 있는지를 확인하고자 한다. 아울러 문항 수준에서의 난이도를 합격자와 불합격자의 정답률과 견주어 봄으로써 현행 한국어능력시험의 평가틀이 갖는 문제점을 비판적으로 검토해 보고자 한다.

## 2. 연구 동향 및 연구 방법

### 2.1 연구 동향

한국어능력시험에 대한 문항 반응 결과 분석은 여러 연구자들에게 흥미로운 주제였다. 일찍이 이관식(2000)은 제3회와 제4회 한국어능력시험의 어휘 평가 연구를 통해, 한국어능력시험의 어휘 평가가 쓰기와 명확히 구분되지 않는 등의 문제를 가지고 있다고 지적하였다. 또한 조인욱(2007)은 제12회 한국어능력시험의 수험자 반응을 분석하여 언어권별 차별성을 추출한 바 있다. 안주호(2008)는 90명의 임의표집 응시자를 대상으로 7회 한국어능력시험 2급 표현 영역을 시행하여 정답률과 오답률을 분석하고 교수 학습상의 개선점을 모색하였다. 최근 들어 이상린(2011)은 읽기 영역 문항의 난이도와 변별도를 문항반응이론(IRT)에 기반하여 분석함으로써 연구자나 교육자의 입장이 아니라 응시자의 입장에서 노출된 문제점을 파악하고자 하였다. 또한 양길석 등(2012)의 연구는 고전검사이론(CTT)에 기반하여 분석하고 회차 간 난이도와 변별도가 상당히 안정적인 시험이라고 긍정적으로 평가하였다.

문항반응 분석과 관련하여 상당한 시사점을 얻을 수 있는 연구는 양길석 등(2012)의 연구이다. 이 연구는 22회~24회 한국어능력시험 반응 결과를 다각도로 분석하여 평가 체제상의 많은 시사점을 제공하였다. 이 연구에서 읽기 영역은 타 영역과의 상관 분포가 0.70~0.82에 이르며 총점과의 상관은 0.87~0.94에 이르는 것으로 분석되었다. 이는 영역 고유의 특성이 있으면서 총점과의 상관도 높아 한국어능력시험의 영역 구분의 유의미성을 제시하는 것이라고 볼 수 있다. 그러나 영역과 총점, 영역 간 상관은 해당 문항에서 응시자들의 성적을 바탕으로 하는 것으로, 상관 분석만으로 영역

구분의 적절성을 완전히 설명해 주지는 못한다. 오히려 한국어능력시험 영역 구분의 적절성을 정확히 보여주는 분석은 같은 연구에서 시행된 구조방정식 모형 분석이었다. 이 분석을 통해 연구진은 영역별 각 문항을 모두 합하여 요인 분석을 실시하고, 각 급별 선다형 문항들이 급 전체에서 2~4개 요인으로 묶여지고 있음을 밝혔다. 이는 현행 한국어능력시험의 4개 영역 문항들은 상당히 자의적으로 구분되고 있으며, 각 영역의 일부 문항들은 사실상 영역 특수한 속성을 가지고 있지 않을 가능성이 있음을 보여준다. 한편 읽기 영역과 관련하여 이상린(2011)은 20회~22회 고급 읽기 영역 문항 반응 결과를 IRT(문항반응이론)에 입각하여 분석하고, 난이도의 안정성 부족, 일부 문항들에서의 난이도 변동성 문제 등을 밝혀냈다. 고급 수준의 읽기 영역에서 출제 의도와 문항 유형을 분석한 결과, 이상린(2011)은 제20회~22회 한국어능력시험에서 의도했던 난이도와 실제 난이도 결과에 지속적인 차이가 발생하고 있음을 발견하였다.

이상린(2011)은 더 나아가 출제 의도와 구현된 결과물의 분석을 통해 현행 한국어능력시험의 평가틀을 유추하고 읽기 영역 평가틀이 ‘양상(aspect)’ 요소만을 적용한 일원적 구성이라 비판하였다(이상린, 2012). 이는 양길석 등(2012)의 연구에서도 동의하는 바이고, 이런 문제점을 해결하기 위해 국립국제교육원에서는 평가틀을 포함한 한국어능력시험의 시험 체제를 개선시키고자 하는 비공개 연구를 진행한 바 있다.

평가틀 이외의 부분에서 특히 한국어능력시험이 비판받는 요소는 검사의 동등성 문제이다. 검사의 동등성이란 어떤 응시자가 능력이 일정한 상태에서 여러 번에 걸쳐 시험에 응시했을 때, 동일한 성적을 얻도록 하는 평가 기술을 의미한다. 일반적으로 연간 다회차로 진행되는 숙달도 평가들에서는 동등성을 확보하는 방안의 마련이 중요한데, 아직까지 한국어능력시험에서는 측정 평가분야에서 제안하는 동등화 방안을 채택하고 있지 않

다. 즉 현행 한국어능력시험에서는 검사의 난이도 유지를 위해 원점수 기반에서 출제진의 전문성에 의존하며, 회차 간 유사한 출제계획으로 문항을 출제한다.

## 2.2 분석 대상 및 분석 방법

이 연구에서 사용된 분석의 대상은 제25회~제27회 한국어능력시험 중급 읽기 영역의 문항 반응 결과이다. 한국어능력시험은 초급, 중급, 고급의 3종 시험으로 각 시험은 원점수 기준의 성적에 따라 1급과 2급, 3급과 4급, 5급과 6급의 합격 여부로 나뉜다. 2010년부터 한국어능력시험은 1년 4차례 시행이 시작되었으며, 연 2회 시험은 국내와 국외에서 시행되고, 연 2회의 시행은 국내에서만 실시된다.

〈표 1〉 제25회~제27회 국내 한국어능력시험 중급 응시 현황

구분	25회 중급	26회 중급	27회 중급
지원자	9,245	10,182	11,169
응시자	8,461	9,029	10,106
합격자	3,759	5,570	6,304
합격률(%)	44.43	61.69	62.38

※ 18대 국회 제출 요구 자료에서 인용

본 연구에서 사용된 자료 중 26회 시험은 국내외에서 시행이 되었고 25회 시험과 27회 시험은 국내에서만 시행이 되었다. 본 연구에서는 응시자 집단 특성의 일관성 유지를 위해 국내 응시자만을 대상으로 분석을 시행하였다. 동 시험에 응시한 국내 지원자들의 기초통계는 <표 1>과 같다. 분석 대상 문항은 각 회차별 선다형 문항 30문항이며, 고전검사이론에 근거하여 문항

을 분석하기 위해 EXCEL 2010과 TestAn 1.0프로그램을 사용하였다.

한국어능력시험 읽기 영역은 2교시에 듣기 영역 이후에 실시되며, 30개의 문항으로 100점 만점으로 책정되어 있다. <표 2>의 중급 읽기 영역 문항 구성표에서 보는 바와 같이, 읽기 영역 문항은 주제와 중심 생각 파악, 맥락 및 분위기 파악, 세부 내용 파악, 유추 또는 추론 기능 등의 크게 4가지 기능에 대하여 평가한다. 또 문항 유형 역시 평가 기능과 궤를 같이 한다.

<표 2> 한국어능력시험 중급 읽기 영역 문항 구성

대문항	소문항	배점	출제 의도	문항유형
31-34	31	3	주제 파악 여부를 측정한다.	읽고 주제 찾기
	32	3		
	33	3		
	34	3		
35-38	35	3	세부 내용 이해 여부를 측정한다.	세부 내용 파악하기
	36	3		
	37	4		
	38	4		
39-42	39	4	중심 생각 파악 능력을 측정한다.	중심 생각 파악하기
	40	4		
	41	4		
	42	4		
43-45	43	3	맥락 이해 능력을 측정한다.	순서 파악하기
	44	3		
	45	3		
46-48	46	4	맥락 안에서 필요한 어구를 찾는 능력을 측정한다.	상황 맥락 활용하기
	47	3		
	48	3		
49-50	49	3	중심 내용 파악 능력을 측정한다.	제목 붙이기
	50	3	세부 내용 이해 여부를 측정한다.	세부 내용 파악하기
51-52	51	4	중심 생각 파악 능력을 측정한다.	중심 생각 파악하기
	52	3	글을 읽고 유추하는 능력을 측정한다.	상황 맥락 활용하기



53-54	53	3	글쓴이의 태도를 파악하는 능력을 측정한다.	글쓴이의 태도 파악하기
	54	3	세부 내용 이해 여부를 측정한다.	세부 내용 파악하기
55-56	55	4	저술 의도 유추 능력을 측정한다.	이유, 목적, 근거 파악하기
	56	3	글을 읽고 유추하는 능력을 측정한다.	상황 맥락 활용하기
57-58	57	3	중심 생각 파악 능력을 측정한다.	중심 생각 파악하기
	58	4	세부 내용 이해 여부를 측정한다.	세부 내용 파악하기
59-60	59	3	중심 생각 파악 능력을 측정한다.	중심 생각 파악하기
	60	3	글의 전개 양상 이해를 측정한다.	논리적 흐름 파악하기

※ 한국어능력시험 홈페이지(www.topik.go.kr)에서 인용

이 연구에서는 한국어능력시험 중급 읽기 영역의 난이도를 추정하기 위해 고전검사이론(CTT)을 사용하였다. 고전검사이론은 관찰 점수가 진점수와 오차 점수의 합계( $X=T+e$ )라고 보고 진점수와 오차 점수의 상관은 0이라 가정한다. 따라서 무수히 많은 시험을 시행할 경우, 오차는 0이 되어 관찰점수와 진점수가 같아진다. 고전검사이론에 의해 문항을 분석할 경우, 응시자의 특성(능력)에 따라 결과가 좌우된다는 단점이 있으나, 사례 수가 충분하다면, 문항반응이론(IRT)에서 요구하는 복잡한 가정이나 조건이 필요치 않기 때문에 전체적인 경향성 추정에 용이하다고 할 수 있다. 이에 본 연구에서는 문항반응이론 대신 고전검사이론에 의해 분석을 하되, 회차별 집단 차이에 따른 영향을 줄이기 위해 회차별 분석 대상 자료의 출처를 국내 응시자로 제한하였다. 따라서 국내 응시자 집단으로 데이터의 취사범위를 제한한다 하더라도 회차별 집단간의 동질성을 확보할 수 없는 점은 이 연구의 제한점이다.

### 3. 한국어능력시험의 중급 읽기 영역 문항 분석

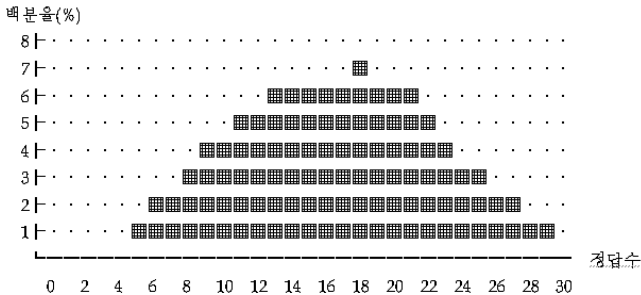
#### 3.1 기술통계 및 신뢰도

25회~27회 한국어능력시험의 중급 읽기 영역의 기술통계와 검사의 신뢰도(Cronbach  $\alpha$ )는 <표 3>과 같다. 이 표에서 보는 바와 같이 25회~27회 중급 읽기 영역의 응시자 수는 8,000명 이상으로 안정적인 결과 산출이 가능한 사례 수이며, 최고점은 100점, 최저점은 0점인 시험이었다. 세 차례 시험에서 평균과 표준편차는 안정적이었고, 문항내적합치도로 추출된 검사 신뢰도 역시 0.8 이상으로 양호하였다.

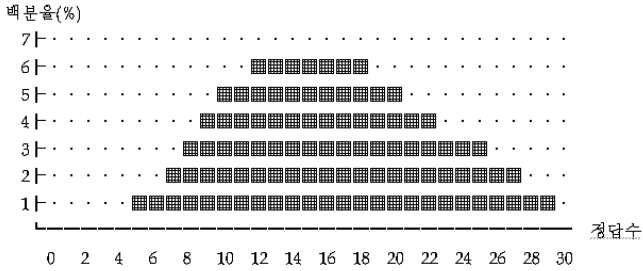
<표 3> 제25회~제27회 한국어능력시험 중급 읽기 영역 기술통계

회차 \ 구분	사례수	최소값	최대값	평균	표준편차	신뢰도( $\alpha$ )
25회	8,461	0	100	55.43	19.13	0.829
26회	9,029	0	100	54.78	18.87	0.822
27회	10,106	0	100	56.26	19.17	0.832

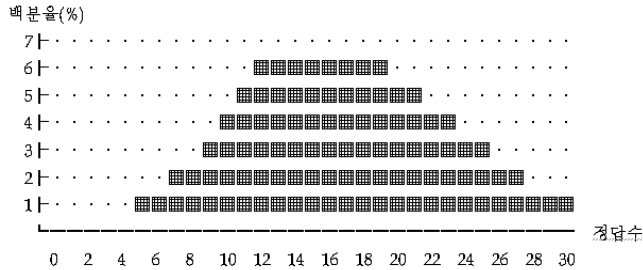
<그림 1> 제25회 한국어능력시험 중급 읽기 영역 점수 분포



〈그림 2〉 제26회 한국어능력시험 중급 읽기 영역 점수 분포



〈그림 3〉 제27회 한국어능력시험 중급 읽기 영역 점수 분포



기술통계치로는 확인할 수 없는 회차별 점수(정답수) 분포 모습은 [그림 1], [그림 2], [그림 3]에 나타나 있다. 기술통계에서는 세 차례의 읽기 영역 시험이 상당히 유사하고 안정적인 시험이었던 것으로 드러났지만, 이 그래프들을 보면 25회 읽기 영역 시험은 부적 편포로 전체적인 문항의 난이도가 쉬운 편이었던 것으로 보여진다. 반면, 26회와 27회 시험은 이와는 달리 정적 편포로 전체적인 문항의 난이도가 약간은 더 어려웠던 것으로 나타나고 있다.

### 3.2 문항 양호도 분석

제25회~제27회 한국어능력시험 중급 읽기 영역 문항의 양호도는 다음

<표 4>와 같다. 이 표에서 살펴볼 수 있는 바와 같이 회차별 난이도는 상당히 안정되어 있으며, 난이도의 평균과 표준편차도 그러한 모습을 보인다. 고전검사이론에서 개별 문항의 변별도는 대개 해당 문항의 점수와 총점과의 상관계수로 구하거나, 상하 27% 집단이 고른 정답 비율의 차로 분석한다. 본 연구에서는 상관계수에 의한 변별도를 구한 결과 변별도 역시 유형별로 회차에 따른 안정성을 보여준다는 결론을 내릴 수 있었다. 그러나 검사 전체가 아니라 문항별로 보았을 때는 일부 문항이 너무 쉽거나 너무 어려운 문항점이 드러나며, 변별도에서 문제를 보이는 문항도 파악할 수 있다.

문항 난이도를 평가하는 절대적인 기준은 없으나, 본고에서는 한국어능력시험이 급별로 2개 등급 이상의 능력 수준을 가지는 응시자를 판별한다는 점을 감안하여 관대한 기준을 채택하여, 0.75이상이면 매우 쉬운 문항, 0.30 미만이면 매우 어려운 문항이라고 평가하고자 한다.<sup>2)</sup> 이러한 기준에 따라 보았을 때, 25회 시험에서 31번, 32번, 37번 문항은 쉬운 문항이고, 58번과 60번 문항은 어려운 문항이라고 볼 수 있다. 26회 시험에서는 31, 33, 34, 35, 36, 39, 42번 문항이 쉬운 문항이었으며 어려운 문항은 없었다. 27회 시험에서는 32, 34, 35, 36, 37번 문항이 쉬운 문항이었고 어려운 문항은 없었다.

2) 성태제(1996)는 문항난이도를 평가하는 절대 기준은 없으나, .30 미만이면 매우 어려운 문항, .80이상이며 매우 쉬운 문항이라 평가할 수 있다고 언급하였다. 한편 Cangelosi는 CTT 분석에서 난이도의 평가 기준을 0.25 이하는 어려운 문항, 0.25~0.75는 적절한 문항, 0.75 이상은 쉬운 문항이라고 제시하였다(성태제, 1996에서 재인용). 한편 박도순 등(2007)은 .2 미만이 매우 어려운 문항, .2 이상 .4 미만은 어려운 문항, .4 이상 .6 미만은 중간 난이도 문항, .6 이상 .8 미만은 쉬운 문항, .8 이상은 매우 쉬운 문항이라는 평가 기준을 제시하였다.

〈표 4〉 제25회~제27회 문항 양호도 분석

구분 번호	난이도			변별도(상관계수)			해당 문항 제외 시 신뢰도		
	25회	26회	27회	25회	26회	27회	25회	26회	27회
31	0.82	0.86	0.74	0.39	0.36	0.43	0.825	0.818	0.826
32	0.88	0.68	0.76	0.40	0.39	0.45	0.825	0.817	0.826
33	0.60	0.84	0.67	0.42	0.32	0.48	0.824	0.819	0.824
34	0.62	0.81	0.87	0.44	0.45	0.43	0.824	0.815	0.826
35	0.33	0.82	0.78	0.44	0.43	0.39	0.824	0.816	0.828
36	0.65	0.85	0.80	0.42	0.40	0.38	0.824	0.817	0.828
37	0.82	0.56	0.87	0.46	0.51	0.40	0.823	0.812	0.827
38	0.70	0.58	0.74	0.45	0.38	0.43	0.823	0.818	0.826
39	0.74	0.75	0.65	0.43	0.41	0.52	0.824	0.816	0.823
40	0.63	0.64	0.43	0.44	0.51	0.45	0.824	0.812	0.826
41	0.47	0.37	0.58	0.45	0.32	0.49	0.823	0.820	0.824
42	0.65	0.75	0.52	0.53	0.42	0.43	0.820	0.816	0.827
43	0.45	0.55	0.74	0.47	0.35	0.40	0.823	0.819	0.827
44	0.69	0.41	0.59	0.38	0.40	0.42	0.826	0.817	0.827
45	0.63	0.39	0.59	0.41	0.27	0.42	0.825	0.822	0.827
46	0.65	0.40	0.52*	0.49	0.33	0.24	0.822	0.820	0.834
47	0.54	0.37	0.45	0.38	0.32	0.51	0.826	0.820	0.823
48	0.38	0.46	0.56	0.24	0.51	0.54	0.831	0.812	0.822
49	0.64	0.51	0.39	0.46	0.37	0.38	0.823	0.818	0.828
50	0.41	0.47	0.37	0.36	0.48	0.39	0.827	0.813	0.828
51	0.34	0.39	0.40	0.30	0.43	0.43	0.829	0.816	0.826
52	0.55	0.47	0.33	0.54	0.50	0.41	0.820	0.813	0.827
53	0.73	0.46	0.69	0.51	0.53	0.47	0.821	0.812	0.825
54	0.44	0.53	0.55	0.42	0.52	0.45	0.824	0.812	0.826
55	0.44	0.46	0.57	0.39	0.35	0.41	0.825	0.819	0.827

56	0.65	0.46	0.37	0.52	0.51	0.32	0.821	0.812	0.830
57	0.38	0.38	0.46	0.38	0.40	0.44	0.826	0.817	0.826
58	0.25	0.41	0.26	0.37	0.32	0.26	0.826	0.820	0.832
59	0.29	0.52	0.35	0.31	0.41	0.28	0.828	0.817	0.832
60	0.19	0.35	0.33	0.09	0.19	0.35	0.834	0.825	0.829
평균	0.55	0.55	0.56	0.41	0.40	0.41	회차별 검사의 신뢰도는 25회 0.829, 26회 0.822, 27 회 0.832임		
표준편차	0.18	0.16	0.17	0.09	0.08	0.07			

Ebel(1965)이 제시한 문항의 변별도(상관계수)를 평가하는 기준은 <표 4>와 같다. 영역 전체로 볼 때는 제25회~제27회 시험의 변별도 역시 회차 간에 안정성이 보이지만, 문항 단위에서는 적지 않은 문항이 문제가 있는 것을 알 수 있다. 25회 시험에서는 48번 문항이 변별도가 낮은 문항이었으며 60번 문항은 변별도가 매우 낮거나 없는 문항으로 평가할 수 있다. 26회 시험에서는 45번 문항이 변별도가 낮은 문항으로 보이며, 60번 문항은 변별도가 낮거나 매우 낮은 문항으로 평가될 수 있다. 27회 시험에서는 46, 58, 59번 문항의 변별도가 낮은 것을 볼 수 있다.

<표 5> Ebel(1965)의 문항 변별도 평가 기준

문항 변별도	문항 평가
0.4 이상	높은 문항
0.30 ~ 0.39	있는 문항
0.20 ~ 0.29	낮은 문항
0.10 ~ 0.19	매우 낮은 문항
0.10 미만	없는 문항

문항의 양호도에서 점검해야 할 또 다른 요소는, 해당 문항을 제외했을 때 검사의 신뢰도에 변화가 없거나 오히려 높아지는 문항의 존재 여부가

다. 검사 문항의 수를 늘리는 것은 측정하고자 하는 능력을 신뢰롭게 재고자 함이다. 그러나 검사 문항의 수가 늘어날수록 응시자의 수험 부담은 증가하므로 신뢰도와 수험 부담 사이에서 적절한 문항수가 결정된다. 이러한 검사 제작의 원칙에 따른다면, 검사의 신뢰도를 저해하거나 아무런 영향을 주지 못하는 문항은 검사지를 구성할 때, 제외되어야 한다(박도순 등, 2007). 더군다나 한국어능력시험과 같이 고정된 평가틀을 갖는 검사에서 회차를 거듭하여 일정한 문항 번호들에서 신뢰도 문제가 나타난다면, 이러한 문항들의 특성을 개선하는 일이 곧 평가틀을 개선하는 일이며, 궁극적으로는 검사의 완성도를 높이는 일이 된다고 볼 수 있다. <표 6>에서 보는 바와 같이, 25회 시험에서 검사 신뢰도를 저해하거나 영향을 주지 못하는 문항은 48번과 51번, 60번 문항이었다. 26회 시험에서는 45번과 60번 문항이, 27회 시험에서는 46번, 58번, 59번 문항이 검사 신뢰도를 저해하거나 영향을 주지 못하는 문항으로 나타났다.

<표 6> 오답의 매력도가 높은 문항

회차	문항 번호	정답	응답 비율			
			①	②	③	④
25회	58	①	0.25	0.29	0.25	0.19
	60	②	0.22	0.19	0.41	0.17
26회	60	③	0.09	0.35	0.34	0.19
27회	52	④	0.34	0.17	0.15	0.33
	58	④	0.15	0.39	0.19	0.25

※ 음영 부분은 정답보다 매력도가 높은 오답의 반응률을 의미함

문항의 양호도를 판별할 때 필요한 또 한 가지 요소는 오답의 매력도이다. 제25회~제27회 시험에서 정답보다 더 매력도가 높은 오답을 가진 문

항은 <표 6>과 같다. <표 6>에서 보는 바와 같이, 25회 시험에서는 58번 문항과 60번 문항, 26회 시험에서는 60번 문항, 27회 시험에서는 52번과 58번 문항에서 정답보다 매력도가 높은 문항이 나타났다.

오답의 매력도가 높아지는 이유로서는 문항에서 정답의 오류, 복답이나 답없음의 가능성이 있다. 오답의 매력도가 정답보다 높은 문항은 반드시 문항의 내용 분석을 통해 정답이 바른 것인지, 복답이나 무답은 아닌 것인지를 확인해야 한다. 해당 문항에 대한 복답이나 무답 등의 가능성을 문항 별로 살펴보면 다음과 같다.

### (1) 가. 25회 58번 지문

※ 다음 글을 읽고 물음에 답하시오.

계란은 거의 모든 영양소를 함유하고 있는 완전식품이다. 그런데 계란 노른자에는 지방이 많다는 이유로 노른자 먹기를 꺼리는 사람들이 있다. 실제로 노른자는 지방 32.6%, 단백질 16.2%로 구성되어 있어 노른자에는 지방이 상대적으로 많이 함유되어 있다고 볼 수 있다. 그러나 이 정도의 지방은 계란을 하루에 네 개 이상 지속적으로 먹지 않는 한 건강에 아무 문제가 없다. 계란 노른자를 멀리하는 또 다른 이유는 혈액 순환 장애의 원인으로 알려진 콜레스테롤 때문이다. 그러나 계란 자체에 함유된 특별한 성분이 혈액의 흐름을 원활하게 해서 콜레스테롤의 흡수를 막아 준다. 따라서 계란 섭취와 혈액 속 콜레스테롤은 상관이 없다.

나. 25회 58번 문항

이 글의 내용과 같은 것을 고르십시오. (4점)

- ① 계란에는 피의 순환을 돕는 성분이 들어 있다.
- ② 콜레스테롤은 지방의 흡수를 막아 주는 효과가 있다.
- ③ 콜레스테롤은 계란의 섭취량에 따라 문제가 될 수 있다.
- ④ 계란 노른자에는 지방보다 단백질이 더 많이 함유되어 있다.

58번 문항은 세부 내용에 대한 이해 여부를 측정하는 문항이다. 문항 지문에서는 계란에 대한 오해와 그에 대한 과학적인 설명이 있다. 정답으로 제시된 ①번의 ‘계란에는 피의 순환을 돕는 성분이 들어 있다.’라는 내용은



지문 중 ‘계란 자체에 함유된 특별한 성분이 혈액의 흐름을 원활하게 해서 콜레스테롤의 흡수를 막아 준다.’는 설명에 의해 지지될 수 있다. 다만, ①번을 정답으로 하려면 지문의 내용으로 볼 때, ‘계란에 함유된 특별한 성분이 혈액의 흐름을 원활하게 해서 계란 노른자에 포함된 콜레스테롤의 흡수 정도는 문제가 되지 않는다’ 즉, ‘혈액의 흐름을 원활하게 하는 특별한 성분이 매우 많이 있다’라고 본문의 내용을 해석해야 할 것이다. 그러나 지문 전체의 내용은 그보다는 계란에 들어 있는 콜레스테롤의 문제는 계란에 함유된 혈액 순환을 돕는 특별한 성분 때문에 문제가 없다고 해석하는 것이 타당한 것으로 보인다. 즉, 계란에 포함된 콜레스테롤과 혈액의 흐름을 원활하게 하는 성분은 비슷한 정도의 해로움과 이로움을 준다는 뉘앙스로 보는 것이 타당할 것이다.

이 문제에서 오답의 매력도가 가장 높은 ②번 문항은 해석상의 오류라 해도, ③번 문항에 대해서는 검토해 볼 여지가 있다. 지문 내용 중 ‘이 정도의 지방은 계란을 하루에 네 개 이상 지속적으로 먹지 않는 한 건강에 아무 문제가 없다’에서 지방 대신에 콜레스테롤을 넣은 문장이다. 이는 세부 내용 파악이라는 문항에 충실하기 위해 지문에 분명히 드러나는 문장에서 일부만을 교묘히 바꾼 것이다. 또한 이 ③이 경우에 따라서는 오답이 아니라 정답일 수도 있다. 콜레스테롤이 계란을 섭취량에 따라 높아질 수도 있고 그렇지 않기도 하다는 내용이 나오기 때문이다. 곧 “그러나 이 정도의 지방은 계란을 하루에 네 개 이상 지속적으로 먹지 않는 한 건강에 아무 문제가 없다.”라는 본문의 내용은 4개 미만일 경우에 문제가 되지 않지만 4개 이상일 경우에는 문제가 될 수 있기 때문에, 콜레스테롤이 계란의 섭취량과 관련이 있음을 적극적으로 암시한다.

(2) 가. 25회 60번 지문

※ 다음 글을 읽고 물음에 답하시오.

이런 일은 어느 날 갑자기 닥친 것이 아니다. 에너지 전문가들은 이런 일이 실제로 벌어질 수 있음을 지속적으로 경고하였다. 전기를 만드는 발전소에서 사용하는 원료는 대부분 수입에 의존하고 있다. 그럼에도 불구하고 원가 이하의 전기 요금 때문에 사람들은 전기가 싼 에너지라는 인식을 갖게 되었다. 상대적으로 낮은 전기 요금이 오히려 에너지 낭비를 부추긴 것이다. 지난 수년간 물가가 오른 것을 감안하면 전기 요금은 거의 변하지 않았다. 지금이라도 사람들이 가장 민감하게 생각하는 전기 요금에 변화를 주어야 한다. 전기 절약 방법을 홍보하는 것보다 사람들이 스스로 에너지를 아껴 쓸 수 있도록 유도하는 것이 더 효과적이기 때문이다.

나. 25회 60번 문항

이 글 앞에 있었을 내용으로 가장 알맞은 것을 고르십시오.

- ① 전기를 만드는 원료가 부족한 상황
- ② 전기가 절대적으로 모자라는 상황
- ③ 전기 절약법을 널리 홍보하는 상황
- ④ 전기 제품 가격이 크게 오르는 상황

60번 문항은 글의 전개 양상을 이해하는 능력을 알아보는 문항이다. 즉 지문에 제시된 내용에 기반 하여 논리적 추론에 의해 선행하는 내용이 무엇이었는지를 추측케 한다. 지문 내용 중 앞글의 내용을 유추할 수 있는 단서는 ‘상대적으로 낮은 전기 요금이 오히려 에너지 낭비를 부추긴 것이다.’라는 문장이다. 이 문장을 중심으로 앞과 뒤의 내용에 전기를 만드는 원료를 수입한다든가, 전기를 절약할 수 있도록 가격을 높여야 한다는 내용이 있다. 이러한 힌트를 바탕으로 할 때, 정답인 ②번은 타당한 답이 될 수 있다. 그러나 지문 내용 후반부에 갑자기 ‘전기 절약을 홍보하는 것보다 사람들이 스스로 에너지를 아껴 쓸 수 있도록 유도하는 것이 더 효과적이다’라는 문장은 응시자들로 하여금 ③번의 ‘전기 절약법을 널리 홍보하는 상황’을 선택하게 하는 구실을 하였다. 물론 글 전체의 내용을 볼 때, 선행하는 글은 전기 부족으로 발생하는 문제가 될 수밖에 없지만, 지문에 제시

된 글의 논리적인 흐름을 볼 때, ‘전기 절약을 홍보하는 것보다’라는 어구는 앞부분에 ‘그런 내용이 있을지도 모른다’라는 의구심을 가지게 하였다고 할 수 있다. 즉 지문을 만들면서 문항을 어렵게 만들기 위해 오답으로 유인할 목적으로 앞 뒤 문맥과 상관없는 문구를 삽입함으로써 문항의 양호도에 문제를 발생시켰다고 볼 수 있다.

(3) 가. 26회 60번 지문

※ 다음 글을 읽고 물음에 답하시오.

그러나 이와는 전혀 다른 주장이 제기되었다. 남성과 여성의 운전 습관이 다르기 때문이 아니라, 자동차의 안전장치가 남성을 염두에 두고 설계되었기 때문에 교통사고 발생 시 여성이 더 크게 다친다는 것이다. 한 연구 결과에 따르면 동일한 차를 운전하고 똑같이 안전띠를 맸는데도 여성 운전자가 중상을 입을 가능성이 남성에 비해 47%나 높았다고 한다. 이러한 결과는 남성과 여성의 신체 구조 차이에서 비롯된 것이다. 여성은 상대적으로 남성에 비해 키가 작고 몸무게가 가볍다. 또 좌석에 앉았을 때 머리가 닿는 부분도 남성과 다르다. 운전자의 안전을 목적으로 하는 자동차의 모든 장치는 이 점을 고려하여 만들어져야 한다.

나. 26회 60번 문항

이 글 앞에 있었을 내용으로 가장 알맞은 것을 고르시오.

- ① 자동차 설계에 대한 운전자의 평가
- ② 성별에 따라 필요한 자동차 안전장치
- ③ 남녀의 운전 습관과 부상 정도의 관련성
- ④ 자동차 안전장치에 따른 교통사고 발생 비율

이 문항 역시 글의 전개 양상을 이해하는 능력을 알아보는 문항이다. 지문 내용 중 앞글의 내용을 유추할 수 있는 단서는 ‘그러나 이와는 전혀 다른 주장이 제기되었다. 남성과 여성의 운전 습관이 다르기 때문이 아니라,’이다. 따라서 정답으로 ③번을 선택하는 데는 큰 어려움이 없다. 다만 이 지문의 중심 내용은 자동차의 안전장치가 남성 편향적이라는 내용인데, 많은 응시자들은 ‘남성과 여성에게 필요한 안전장치의 종류가 다르다’라는

주장이 앞부분에 있었다고 생각하는 것으로 응답하였다. 이 문항에서 오답의 매력도가 정답보다 높았던 이유에 대해서는 논리나 추론 능력이 포함되는 문항이 한국어 사용 능력 이전에 논리적 사고력이 포함된 모국어 능력이 갖추어져야만 해결할 수 있는 문항이었기 때문에 나타난 문제가 아니었을까 추측된다.

(4) 가. 27회 52번 지문

※ 다음 글을 읽고 물음에 답하시오.

여러 사람이 차례로 자기 생각을 이야기할 때 마지막으로 말해야 할 경우가 있다. 이럴 때 우리는 자기가 할 말을 다른 사람들이 앞에서 모두 해 버릴 것이라고 걱정을 한다. 하지만 나중에 말하는 것도 좋은 점이 있다. 앞에서 말한 의견들을 잘 듣고 그것을 종합하여 내가 할 말을 준비할 수 있다는 점이다. 이 때 반드시 ( ) 덧붙이는 것이 필요하다. 아무리 말을 잘해도 자기 생각이 빠지면 좋은 말하기가 될 수 없기 때문이다. 가장 나중에 말하는 것은 좋은 점이 또 하나 있다. 많은 사람들이 제일 끝에 들은 이야기를 가장 오래 기억한다는 점이 바로 그것이다.

나. 27회 52번 문항

( ) 안에 들어갈 가장 알맞은 말을 고르십시오.

- ① 다른 사람들의 의견을
- ② 자기가 기억하는 내용을
- ③ 마지막에 들은 이야기를
- ④ 자기만의 독창적인 생각을

이 문항은 글쓴이의 태도를 파악하는 능력을 평가하는 문항으로 글의 문맥을 고려하여 ( ) 안에 알맞은 어구를 고르는 문항이다. 이 문항에서 정답의 힌트로 사용된 것은 주어진 ( )의 바로 뒤에 오는 ‘아무리 말을 잘해도 자기 생각이 빠지면 좋은 말하기가 될 수 없기 때문이다.’이다. 즉 ‘자기 생각이 빠지면 좋은 말하기가 아니다’라는 이유가 왔으므로 앞의 문항에서는 그와 관련된 주장이 제기되어야 한다. 따라서 정답은 ④번이 된다. 이 문항에 오답의 매력도가 정답보다 높았던 보기 답은 ①번의 ‘다른

사람의 의견'이다. ①번 보기가 오답의 매력도가 높았던 이유는 문장 중간에 '이 때'를 기준으로 앞부분의 내용과 이어지는 내용에 차이가 있음에도 불구하고, 앞부분 내용의 잔상이 보기 답 선택에 영향을 끼쳤기 때문으로 생각된다. 이 문항에서 요구되는 한국어 능력은 다양한 편인데, '~ 때문이다.'가 이유로 작용하므로, 이에 대한 주장이 선행해야 한다는 어휘 지식이 필요하며, 이유와 주장간의 연관성을 파악하고 알맞은 주장을 선택할 수 있는 논리적 사고력도 필요하다. 또, '이때'라는 시간을 나타내는 부사어가 문단 중간에서 화제의 전환을 유도하는 역할을 하는 경우도 있다는 지식도 필요하다고 할 수 있다. 이런 점에서 본다면 이 문항은 상당히 다양한 한국어 능력을 요구하는 문항이라서 오답의 매력도가 정답의 그것을 능가하였다고 결론지을 수 있을 것이다.

(5) 가. 27회 58번 지문

※ 다음 글을 읽고 물음에 답하시오.

사람들은 화려한 색깔의 버섯을 독버섯이라고 생각하고 두려워한다. 독버섯을 먹으면 위험하기 때문이다. 그런데 독버섯 주변에 함께 살고 있는 나무는 아무런 문제가 없다. 독버섯이 나무에는 피해를 주지 않는 걸까? 사실 독버섯이라는 이름은 사람에게 독이 되기 때문에 붙여진 이름이다. 자연 속에서 독버섯은 다른 버섯과 마찬가지로 주변 동식물에 아무 피해도 주지 않고 살아간다. 독버섯은 몸체에 붙은 실처럼 긴 부분을 이용하여 땅속 깊은 데서 물을 빨아들인 후 그 물을 나무에 전해 준다. 그리고 나무로부터 영양분을 받아서 생활한다. 또한 독버섯은 달팽이같이 작은 동물에게는 귀한 먹이가 되기도 한다. 자연 상태에서 독버섯은 보통의 버섯과 별 차이가 없다.

나. 27회 58번 문항

이 글의 내용과 같은 것을 고르십시오. (4점)

- ① 자연 속의 독버섯은 일반 버섯과 다른 역할을 한다.
- ② 독버섯은 주변 나무에 자신이 만든 영양분을 나누어 준다.
- ③ 색깔이 예쁜 버섯 중에는 사람이 먹어도 되는 버섯이 많다.
- ④ 독버섯은 생활하는 데 필요한 물을 스스로의 힘으로 얻는다.

이 문항은 세부 내용의 이해 여부를 측정하는 문항으로 지문에 제시된 글의 일부분과 보기 지문의 내용을 견주어 바르게 판단하는 능력이 필요하다. 이러한 문항에서 보기 지문이 오답이 되는 경우는 제시된 글과 상반되거나, 제시된 글에는 없는 내용일 경우 오답이 된다. 이런 점에서 ④번은 제시문에 ‘독버섯은 몸체에 붙은 실처럼 긴 부분을 이용하여 땅속 깊은 데서 물을 빨아들인 후 그 물을 나무에 전해 준다’는 내용과 일치하므로 정답이 된다. 많은 응시자들은 정답인 ④번 대신 ②번을 선택했는데, 독버섯이 ‘물을 빨아들인 후 그 물을 나무에 전해 준다’는 문장에서 물 대신 영양분으로 바뀐 점을 파악하지 못하였기 때문으로 생각된다.

세부 내용의 이해 여부를 측정하는 문항에서 글의 내용과 상반되거나 글에는 없는 내용을 제시하거나 착각하기 쉬운 내용이나 단어만 바꿔치기 하는 방식은 오답의 매력도를 높이는 방식으로 사용된다. 이 중에서 착각하기 쉬운 내용이나 단어만을 바꿔치기 하는 방식은 실수에 의해 오답을 선택하게 한다는 측면에서 한국어능력을 측정하는 문제로 적합한지에 대해 좀 더 깊은 검토를 필요로 할 것이다.

일반적으로 문항의 양호도는 난이도와 변별도, 해당 문항을 제외했을 때 검사 신뢰도에 주는 영향, 오답의 매력도 등을 종합적으로 판단한다. 한 두 가지 문제가 제기되었다고 해당 문항에 오류가 있다고 판단하는 것은 무리가 있으며, 여러 가지 문제가 한 문항 내에서 중첩적으로 발생할 때, 특히 문항의 내용 분석까지 거쳐 문항에 문제가 있을 때, 문항의 양호도가 낮다고 보는 것이 바른 판단이다. 이러한 판단 방식으로 볼 때, 제25회~제27회 한국어능력시험 중급 읽기 영역의 문항 양호도는 대체로 높은 편이라고 볼 수 있다. 다만, 앞부분의 문제가 지나치게 쉽다거나, 뒷부분의 문제가 지나치게 어려울지 여부는 문항의 수준을 살펴보는 좀 더 다양한 방법에 의해 검토되어야 할 것이다. 이에 본고에서는 해당 시험의 합격자와 불합

격자가 각각의 문항들에 어떻게 반응했는가를 보고 문항과 그 문항을 출제하게 한 평가들에 대해 좀 더 검토해 보고자 한다.

### 3.3 합격/불합격 기준에 따른 문항의 난이도 수준 분석

한국어능력시험의 합격 기준은 <표 7>과 같다. 한국어능력시험에서는 해당 수준에서 얻은 성적의 4개 영역 평균 점수가 70점 초과이고 4개 영역에서 50점 미만인 점수가 없으면 상위 등급을 획득하게 된다. 또 영역 평균 점수가 50점 이상이고 4개 영역에서 40점 미만인 점수가 없으면 하위 등급을 획득하며, 영역 평균 점수가 50점 미만이거나 4개 영역에서 40점 미만인 영역이 하나라도 있으면 불합격이 된다.

<표 7> 한국어능력시험의 합격 기준

시험 수준	시험 등급	합격 점수	과락 점수
초급	1급	50점 이상	40점 미만
	2급	70점 초과	50점 미만
중급	3급	50점 이상	40점 미만
	4급	70점 초과	50점 미만
고급	5급	50점 이상	40점 미만
	6급	70점 초과	50점 미만

※ 국립국제교육원(2012)에서 인용

이러한 합격 기준을 통해 현행 한국어능력시험의 점수 체제는 원점수에 기반한 점수 체제이며, 회차 간 시험의 동등화를 위해서 문항 출제 과정에서 이전 회차와의 난이도 수준을 일치시키는 노력을 기울이고 있음을 알 수 있다. 이처럼 원점수 체제의 시험이 갖는 약점 중에 하나는 출제 과정에서 아무리 노력한다 하더라도 회차 간 동등화가 사실상 불가능하다는 것이

다. 또 점수 해석에 있어서 평균 90점이나 100점이 아무런 의미도 갖지 못한다는 것이다. 이러한 문제를 해결하기 위해 TOEIC이나 TOEFL 등의 시험들은 수준 설정, 점수 척도 개발, 회차 간 동등화 시행 등 고도의 동등화 절차를 수행하고 그에 따른 성적을 산출한다. 그러나 양길석 등(2012)은 제22회~제24회 시험의 회차 간 난이도와 변별도 비교를 통해 한국어능력시험이 원점수 기반 시험임에도 불구하고 상당히 안정적인 난이도와 변별도를 갖고 있음을 밝혔다. 물론 세 수준 내 영역 간 난이도와 변별도는 상당히 상이하였으나, 동일한 수준의 동일한 영역은 회차 간 차이가 별로 없는 안정적인 시험이었다. 이러한 경향은 본 연구에서도 유사하게 나타났는데, 본 연구의 앞 절에서 살펴본 바와 같이 제25회~제27회 난이도와 변별도 역시 상당히 안정적인 모습을 보이고 있다. 이는 현재의 한국어능력시험이 고정된 출제 계획-수준별 해당 영역에서 일정 문항 번호에는 동일한 소재와 기능, 텍스트의 종류, 문항 유형만을 출제하게 하는 방법을 사용하기 때문인 것으로 보인다. 이 때문에 현행 한국어능력시험 문항들은 어느 회차를 막론하고 이전 회차와 상당히 유사한 문항 속성을 갖는 문항들로 검사지가 구성된다.

이처럼 고정된 출제 계획에 의한 문항 출제는 시험의 안정성을 높인다는 장점이 있는 반면에, 시험이 측정하고자 하는 능력 중 일부만을 반복적으로 출제할 가능성을 높인다는 문제를 갖는다. 한국어능력시험과 같이 한국어 교육과 학습에 상당한 영향력을 행사하는 시험이 일부 능력만을 반복적으로 측정하게 되면 한국어 교육과 학습 자체를 파행으로 이끌 가능성이 높아진다. 실제로 말하기 영역이 제외되고 어휘와 문법이 포함되면서 한국어 교육에서 말하기 교육이 소홀해지고 문법 교육이 강화되었다는 비판은 그 동안 여러 차례 학계에서 제기된 바 있다.

아울러 원점수에 기반하고 고정된 출제 계획에 의해 출제되는 시험은



잘못된 출제 계획을 개선하는데 많은 노력과 어려움을 겪게 된다. 일부 문항에서 정해진 속성이 잘못된 것으로 드러나 문항의 속성을 변경할 경우 시험의 안정성이 크게 훼손될 가능성이 있다는 문제점이 있다.

일반적으로 상하위 집단 정답 비율차에 의해 변별도를 구하는 방식은 해당 영역에서 상위 27% 집단의 정답률과 하위 27% 집단의 정답률의 차이를 변별도로 정하는 방식이다. 상하위 집단 정답 비율차에 의한 변별도 산정 방식은 검사 내 문항의 완성도가 높고, 집단의 능력이 안정적이라는 가정이 필요하다. 또 이는 상하위 집단 정답 비율차라는 개념을 위해 상위 집단은 전체 성적에서 상위 27% 집단이라고 조작적으로 정의한 것이다. 그러나 한국어능력시험의 경우, 원점수에 기반한 불합격, 하위 등급 합격, 상위 등급 합격의 세 집단이 존재한다. 따라서 추상적이고 불안정한 개념인 상하위 27% 집단보다, 동 시험에 의해 규정된 하위 집단인 불합격자 집단, 상위 집단인 상급 합격자 집단의 정답률을 검토해 봄으로써 난이도에 대해 분석할 필요가 있다. 이에 이 글에서는 현재의 고정된 한국어능력 시험 출제 계획에서 정한 문항별 난이도에는 어떠한 문제가 있는가를 살펴 보기 위해 해당 회차에서 합격한 응시자와 불합격한 응시자로 나누어 문항의 난이도를 살펴보고자 한다.

### 3.3.1. 25회 문항의 난이도

앞에서 살펴 본 바와 같이, 한국어능력시험은 초급과 중급, 고급으로 구분되는 수준별 등급제 시험이다. 하나의 시험을 수준별로 나누는 실익은 해당 수준에 부합하는지 여부를 보다 정확하게 판별하기 위해서이다. 이상적인 형태의 수준별 시험이라면, 중급 시험의 각 영역에 속한 문항들이 불합격자에게는 어렵게, 합격자에게는 쉬운 문항이어야 한다는 것을 의미한

다. 그렇지 않고 불합격자에게 쉬운 문항이 있다거나 합격자들이 풀지 못할 정도로 어려운 문항이 포함되면, 너무 쉬운 문항이나 너무 어려운 문항들이 해당 등급을 판정하는데 있어 별다른 도움을 주지 못한다는 것을 의미한다. 이는 앞서의 신뢰도에서 언급했다시피 문항 수가 시험 부담을 최소화하고 신뢰도를 최대화하는 지점에서 결정되어야 함에도 쉬운 문항과 어려운 문항이 등급 결정에 도움이 되지도 못하면서 시험 부담은 늘리고 신뢰도는 저하시키는 역할을 하게 된다는 것을 의미한다.

이처럼 너무 쉽거나 어려워서 등급 판정에 도움이 되지 못하는 문항이 있는지 여부를 파악하기 위해 제25회 한국어능력시험 중급 읽기 영역 문항 분석을 합격자와 불합격자로 구분하여 나타낸 것이 <표 8>이다.

<표 8> 25회 중급 읽기 영역 4급 합격자와 불합격자의 난이도

번호	전체 응시자	4급 합격자	불합격자	번호	전체 응시자	4급 합격자	불합격자
31	0.82	0.97	0.73	46	0.65	0.96	0.48
32	0.88	1.00	0.81	47	0.54	0.81	0.44
33	0.60	0.85	0.48	48	0.38*	0.58	0.32
34	0.62	0.91	0.49	49	0.64	0.93	0.50
35	0.33	0.64	0.21	50	0.41	0.68	0.31
36	0.65	0.87	0.53	51	0.34*	0.55	0.28
37	0.82	0.99	0.70	52	0.55	0.90	0.39
38	0.70	0.93	0.56	53	0.73	0.97	0.59
39	0.74	0.91	0.63	54	0.44	0.78	0.31
40	0.63	0.89	0.49	55	0.44	0.71	0.34
41	0.47	0.83	0.34	56	0.65	0.95	0.49
42	0.65	0.94	0.48	57	0.38	0.70	0.28
43	0.45	0.79	0.31	58	0.25	0.53	0.18
44	0.69	0.87	0.59	59	0.29	0.50	0.22
45	0.63	0.85	0.52	60	0.19*	0.23	0.19

능력 수준에 따른 차이를 감안하기 위해서는 난이도의 평가 기준을 좀 더 강화할 필요가 있다. 상위 득점자-한국어능력시험 중급에서는 4급 합격자에게 어려운 문항은 중급에 응시한 모든 응시자에게 어려운 문항일 수밖에 없다. 반면 하위 득점자-한국어능력시험 중급에서 불합격자에게 쉬운 문항은 중급에 응시한 모든 응시자에게 쉬운 문항일 수밖에 없다. 이에 본고에서는 성태제(2002)와 박도순 등(2007)의 기준을 참고하여 4급 합격자에게 어려운 문항인 난이도가 .4 미만인 문항과, 불합격자에게 쉬운 문항인 .6 이상인 문항을 검토하여 한국어능력시험 중급 읽기 영역 난이도의 적정 수준을 살펴보고자 한다.

4급 합격자에게 0.4 미만의 난이도를 보이는 문항은 25회 시험에서 60번 문항이었다. 60번 문항은 4급을 합격한 응시생들 중 겨우 23%만 정답을 맞출 수 있는 문항이므로 4급인지 3급인지를 가려내는데 아무런 도움이 되지 못하는 문항이라고 할 수 있다. 반면, 31, 32, 37, 39번 문항은 불합격자들도 60% 이상이 정답을 맞힐 확률을 보이는 문항으로 불합격자를 걸러내는데 거의 도움을 주지 못하는 문항이라고 평가할 수 있을 것이다.

### 3.3.2. 26회차 문항의 난이도

제26회 한국어능력시험 중급 읽기 영역 문항의 난이도를 합격자와 불합격자로 구분하여 나타낸 것이 <표 9>이다. <표 9>에서 보는 바와 같이, 26회 시험 중급 읽기 영역의 문항 중 매우 쉬운 문항은 31, 33, 34, 35, 36번 문항으로 불합격자들이 60% 이상의 정답률을 보이는 문항이었다. 이들 문항은 전체 응시자의 문항 난이도에서 0.8 이상의 정답률을 보이는 문항이었다. 반면 매우 어려운 문항인 4급 합격자가 .40 미만의 정답률을 보이는 문항은 없었으나, 60번 문항은 4급 합격자라도 42%의 정답률을 보이는 꽤 어려운 문항이라고 볼 수 있다.

〈표 9〉 26회 중급 읽기 영역 4급 합격자와 불합격자의 난이도

번호	전체 응시자	4급 합격자	불합격자	번호	전체 응시자	4급 합격자	불합격자
31	0.86	0.98	0.72	46	0.40	0.64	0.27
32	0.68	0.86	0.50	47	0.37	0.56	0.25
33	0.84	0.95	0.73	48	0.46	0.80	0.22
34	0.81	0.98	0.62	49	0.51	0.75	0.36
35	0.82	0.96	0.64	50	0.47	0.80	0.27
36	0.85	0.98	0.68	51	0.39	0.73	0.25
37	0.56	0.87	0.29	52	0.47	0.82	0.26
38	0.58	0.81	0.39	53	0.46	0.82	0.21
39	0.75	0.95	0.56	54	0.53	0.87	0.30
40	0.64	0.92	0.36	55	0.46	0.68	0.32
41	0.37	0.58	0.24	56	0.46	0.82	0.25
42	0.75	0.92	0.54	57	0.38	0.68	0.26
43	0.55	0.76	0.42	58	0.41	0.63	0.32
44	0.41	0.67	0.26	59	0.52	0.77	0.36
45	0.39*	0.56	0.30	60	0.35*	0.42	0.30

### 3.3.3. 27회차 문항의 난이도

〈표 10〉 27회 중급 읽기 영역 4급 합격자와 불합격자의 난이도

번호	전체 응시자	4급 합격자	불합격자	번호	전체 응시자	4급 합격자	불합격자
31	0.74	0.93	0.54	46	0.52*	0.62	0.43
32	0.76	0.96	0.55	47	0.45	0.78	0.21
33	0.67	0.92	0.44	48	0.56	0.88	0.29
34	0.87	0.99	0.70	49	0.39	0.61	0.27
35	0.78	0.94	0.63	50	0.37	0.60	0.24
36	0.80	0.95	0.62	51	0.40	0.68	0.24

37	0.87	0.99	0.71	52	0.33	0.58	0.20
38	0.74	0.93	0.55	53	0.69	0.92	0.48
39	0.65	0.92	0.36	54	0.55	0.82	0.37
40	0.43	0.70	0.22	55	0.57	0.82	0.40
41	0.58	0.88	0.34	56	0.37	0.54	0.26
42	0.52	0.80	0.33	57	0.46	0.74	0.31
43	0.74	0.92	0.57	58	0.26*	0.39	0.21
44	0.59	0.83	0.40	59	0.35*	0.50	0.28
45	0.59	0.82	0.41	60	0.33	0.53	0.23

<표 10>은 제27회 한국어능력시험 중급 읽기 영역 문항의 난이도를 합격자와 불합격자로 구분하여 나타낸 것이다. 27회 시험 문항 중에서는 34, 35, 36, 37번 문항이 불합격자도 60% 이상의 정답률을 보이는 문항으로 매우 쉬운 문항이었다. 반면 58번 문항은 4급 합격자들이 39%의 정답률을 보이는 문항으로 매우 어려운 문항이었다고 할 수 있다.

### 3.4 문항 분석 결과에 대한 논의

이상에서 분석된 제25회~제27회 중급 읽기 영역의 문항 분석 결과를 요약하면 <표 11>과 같다. 일반적으로 볼 때, 쉬운 문항에서는 난이도 외의 문제가 발생하지 않지만, 지나치게 어려운 문항에서는 난이도의 문제에 더하여 변별도에 문제가 생기고 추측에 의해 답을 맞힐 가능성이 높아지게 된다.

〈표 11〉 25회~27회 중급 읽기 영역 문항 분석 결과 요약

번호	25회				26회				27회			
	난이도	변별도	오답 매력도	제외 신뢰도	난이도	변별도	오답 매력도	제외 신뢰도	난이도	변별도	오답 매력도	제외 신뢰도
31	쉬움				쉬움							
32	쉬움											
33					쉬움					낮음		
34					쉬움				쉬움			
35					쉬움				쉬움			
36					쉬움				쉬움			
37	쉬움								쉬움			
38												
39	쉬움											
40												
41												
42												
43												
44												
45						낮음		높음				
49												
50												
51				높음								
52											특이	
53												
54												
55												
56												
57												
58			특이						어려움	낮음	특이	높음
59										낮음		높음
60	어려움	낮음	특이	높음		낮음	특이	높음				

추측에 의해 답을 맞힐 확률이 높아지면, 응시자의 능력과 무관하게 정답을 선택하게 되므로, 해당 문항을 제외했을 때 검사의 신뢰도가 높아진다. 문항 반응 분석을 실시해 보면 쉬운 문항에서는 난이도만의 문제로 나타나지만 어려운 문항은 낮은 난이도와 낮은 변별도, 정답보다 매력도가 높은 오답 보기 발생, 해당 문항을 제외했을 때 검사의 신뢰도가 높아지는 등의 문제가 중복하여 나타나게 된다. 이런 경향은 <표 11>에 잘 나타나 있는데, 따라서 <표 11>을 해석할 때는 문제 항목의 수만 가지고 어떤 문항이 문제가 있다고 판단해서는 안 된다. 즉, 쉬운 문항들은 난이도의 문제만 보일 경우라도 해당 난이도가 그 시험에 적합한 것인가를 판단해서 문제인지 여부를 검토해야 한다. 반면 어려운 문항들은 문항의 난이도가 변별도와 결부될 수밖에 없으므로 어려운 정도와 여타 평가 기준들을 함께 고려해야 할 것이다.

이러한 문항 분석의 유의 사항을 고려하여 <표 11>에 나타난 문제점을 분석하면 우선, 각 회차별로 불합격자도 쉽게 해결할 수 있는 쉬운 문항이 매우 많다는 점을 발견할 수 있다. 이들 문항들은 변별도나 해당 문항을 제외했을 때 신뢰도 등에서 문제를 보이지는 않겠지만, 해당 등급을 판가름하는데 도움이 되지 않는 문항이라고 볼 수 있을 것이다. <표 11>에서 발견할 수 있는 또 다른 문제점으로는 58번과 60번 문항에서 반복적으로 문제가 나타난다는 점이다. 이들 문항들은 검사의 후반부에 위치하여 문항의 난이도가 매우 낮기 때문에 문제들이 나타나는 것일 수도 있다. 그러나 앞서 오답의 매력도에 문제가 있는 문항들을 검토해 본 바와 같이, 58번 문항은 평가 기능 중 ‘세부 내용 파악’ 기능을 평가하는 문항이었다. 일반적으로 세부 내용 파악 기능은 문항을 만들면서 문제를 어렵게 만들기 위해 매우 작거나 간과하기 쉬운 부분에서 변형을 줌으로써, 피험자들을 혼란에 빠뜨리기 쉬운 문항이다. 더군다나 듣기와 읽기가 함께 진행되는 중급 2교

시 시험에서 시험에 대한 몰입도가 가장 악화되는 시험 끝부분에 세부 내용 파악 문항을 넣음으로 해서 시험에 대한 피로도도 문항의 난이도가 더욱 낮아지는 것일 수도 있다. 이런 점에서 볼 때, 세부 내용 파악 문항이 꼭 필요하다면, 검사의 중반으로 이동시키고, 문항의 난이도도 지나치게 낮아지지 않도록 전체적인 맥락에서 외떨어진 작은 부분을 묻는 것은 지양하는 것이 바람직할 것으로 판단된다.

이와 마찬가지로 60번 문항은 글의 전개 양상을 이해하는 능력을 측정하는 문항이다. 글의 전개 양상을 이해하는 능력이 한국어 중급에서 갖추어져야 할 능력이라고 해도, 지문의 수준이나 지문에 포함된 어휘, 지문의 종류 등이 중급 수준에 맞는 것인가를 재검토할 필요가 있다. 또 지문에 포함된 아주 사소한 부분을 제시하고 ‘앞부분에 왔어야 하는 내용’을 묻는 방식 보다는, 글의 자연스러운 흐름을 파악할 수 있는가를 평가하는 방식으로 문항 유형이 바뀌어야 할 것이다. 숙달도 검사는 실수가 적은 피험자를 찾아내는 시험이 아니라, 요구하는 기준에 부합하는 능력을 갖춘 피험자를 찾아내는 시험이기 때문이다.

제25회~제27회 한국어능력시험 중급 읽기 영역에 대한 문항 반응 결과에 대한 분석을 통해 동 시험의 중급 읽기 영역에 대한 세 가지 시사점을 얻을 수 있었다. 우선 한국어능력시험 중급 읽기 영역은 매 회차 별로 중급 수준을 판별하기에 적합하지 않은 너무 쉬운 문항이나 너무 어려운 문항이 발견된다. 중급 읽기 영역 30 문항 중, 매 회차 별로 부적절한 난이도의 문항이 5문항(16.67%)이었다. 그 중 매우 쉬워서 불합격자들조차 쉽게 풀 수 있는 문항은 제25회~제27회 중급 읽기 영역 이상 난이도 문항 15개 중 13 문항(회차 별 평균 4.3 문항)이었고, 4급 합격자에게도 상당히 어려웠던 문항은 2문항(회차 별 평균 0.67 문항)이었다. 회차 별로 통시적으로 볼 때, 읽기 영역 시험의 앞부분인 31번-36번까지는 대체로 필요보다 매우



쉬운 문항으로 구성되어 있으며, 58번-60번은 4급 합격자들도 풀기 어려운 매우 어려운 문항으로 구성되어 있음을 발견할 수 있었다.

두 번째로 지적할 사항은 회차 별 이상 난이도의 문항 번호에 일관성이 없다는 것이다. 한국어능력시험은 주관기관인 국립국제교육원이 밝히고 있는 바와 같이 규격화된 문항 구성표를 가지고 시험을 제작하고 있다. 이러한 문항 구성표(출제 계획표)의 특징은 매회차 일정한 번호에 들어가는 문제의 속성을 일치시켜 회차별로 시험의 동등성을 유지한다는 것이다. 이때, 정해지는 문항의 속성에는 주제, 소재, 평가 기능, 문항 유형, 텍스트의 종류, 문항의 난이도 등이 있다. 국립국제교육원이 공개한 평가틀(문항 구성표)에는 문항별 난이도가 공개되어 있지 않지만, 시험의 개발과 운영 과정에서 문항별 난이도를 고정하여 검사의 난이도가 일관성을 유지하도록 한다는 것은 당연한 일이다.

분석 결과로 드러난 바와 같이 문항별 난이도가 회차마다 크게 달라지는 점은 검사의 동등성을 저해하는 상당히 심각한 문제라고 볼 수 있다. 이렇게 문항 난이도가 달라지는 이유는 몇 가지로 예상되는데, 한 가지 이유는 일관된 문항 난이도를 유지시킬 만큼 출제진의 전문성이 확보되지 않을 가능성이 있다. 문항을 출제하는 출제위원이나 출제된 문항을 검토하고 평가하는 평가위원이 문항 난이도 유지를 위한 전문성이 부족할 경우 이러한 문제가 나타날 수 있다. 또 다른 이유는 출제를 위해 참조하는 자료 또는 준거에 문제가 있을 때에도 문항 난이도를 유지하는데 어려움을 겪을 수 있다. 예를 들어 수준별 어휘나 문법 목록이 현실의 언어 사용을 반영하지 못한 것일 때, 아무리 우수한 출제진이 출제와 검토를 시행한다 하더라도 문항 난이도 유지에 어려움이 발생할 수밖에 없다.

이에 더하여 현행 한국어능력시험이 갖는 점수 체제도 문항 난이도 유지를 곤란하게 만드는 요소일 수 있다. 즉, 원점수 체제에서는 제 아무리

우수한 출제진이 모여 문항을 개발한다고 하더라도 회차별로 문항 난이도를 유지시킨다는 것은 사실상 불가능하다. 이를 해결하기 위해, 여타의 많은 시험들에서는 동등화 방안을 강구하고 있으며, 수준 설정 등의 방법을 보완하여 사용하고 있다. 이런 점에서 문항과 검사의 난이도를 출제진에게만 의존하는 현행 한국어능력시험 출제 체제는 개선되어야 할 것이다.

세 번째 문제점은 문항 난이도 분포와 점수 체제와의 불일치성이다. 일반적으로 규준 참조 검사(norm reference test)는 평가 결과를 해석할 때 집단 내의 상대적인 위치를 기준으로 평가하는 방법으로 검사 점수의 분포가 정상분포가 될 것을 가정한다(박도순 등, 2007). 반면 준거 지향 검사(criterion reference test)는 정해진 수준을 달성하거나 충족하였는가를 평가하는 방법으로 운전면허시험, 간호사시험과 같이 응시 집단을 숙달/미숙달로 나누어 자격증이나 합격증을 발급하는 시험에 사용된다. 그러므로 한국어능력시험은 원점수에 기반하는 준거 지향 검사라고 볼 수 있다. 이러한 준거 지향 검사를 제작하기 위해서는 숙달/미숙달의 경계를 정확히 구분해 줄 수 있는 문항 제작이 필요하다. 즉, 한국어능력시험 중급 시험이 불합격자와 3급 합격자, 4급 합격자로 응시 집단을 구분해 주는 시험이 되기 위해서는 불합격자와 3급 합격자, 3급 합격자와 4급 합격자를 구분해 주는 문항들로 검사지가 구성되어야 한다.<sup>3)</sup>

그러나 앞서 살펴본 바와 같이 25회~27회 한국어능력시험 중급 읽기 영역의 문항들은 그러한 경계선을 구분해주는 문항들로만 이루어져 있다고 보기에는 어려움이 있다. 일부 문항들은 불합격자들도 쉽게 풀 수 있는

3) 이에 대해 성태제(2002)는 ‘규준 참조 검사에서는 매우 쉬운 문제부터 매우 어려운 문항까지 골고루 출제하여 피험자 점수의 폭을 넓혀서 서열화가 용이하도록 한다. 이에 비하여 준거 참조 검사의 경우 문항 난이도를 어떤 과업의 수행 준거와 유사하게 조정한다’라고 주장하면서, 수행 준거와 유사한 난이도를 갖는 문항들로 검사를 구성하는 것이 준거 참조 검사 문항 제작의 원칙이라고 설명하였다.

문항이므로 불합격자와 3급 합격자를 구분 짓는데 사용할 수 없는 문항이라 할 수 있다. 또 일부 문항들은 4급 합격자들에게도 어려운 문항으로 3급 합격자와 4급 합격자를 구분 짓는데 도움이 되지 못한다. 이처럼 등급간 구분이나 불합격자와 낮은 수준의 등급 합격자간의 구분에 도움이 되지 못하는 문항이 포함될 경우, 검사는 필요 이상의 문항을 가지게 되고 그로 인해 응시자의 검사 부담은 증가할 수밖에 없다. 따라서 현행 한국어능력시험 중급 읽기 영역의 문항 수는 30문항이지만, 실제로 등급 설정에 쓰이는 문항은 그보다 3~5문항 정도 적은 문항만이 쓸모 있는 문항이라고 판단할 수 있는 것이다. 이는 한국어능력시험의 평가틀과 시험 체제가 만들어지고 수정되는 과정에서 충분히 고려되지 못한 문제점으로 향후 한국어능력시험의 개선 과정에서는 등급제 시험의 성격에 부합하는 문항 난이도 설정이 필요하리라 생각된다.

#### 4. 결론

2011년 제21회 시험부터 한국어능력시험이 한국교육과정평가원에서 국립국제교육원으로 이관된 후, 유학생 유치관리 인증제의 도입 등으로 한국어능력시험은 비약적인 성장을 이룩하여 2013년 1월에 시행된 제29회 시험을 기점으로 100만의 응시자를 기록하였다. 이러한 양적 팽창은 질적 개선을 요구하는 바, 현행 한국어능력시험은 국내외 학계 및 교육계의 변화 요구에 직면하고 있다. 이런 상황에서 본고는 제25회~제27회 한국어능력시험 중급 읽기 영역의 문항 난이도를 비판적으로 검토하고 도출된 문제점이 평가틀과 어떤 관련이 있는가를 살펴보고자 하였다.

문항별로 분석해 보았을 때, 25회 시험에서 불합격자에게도 매우 쉬운

문항은 4개 문항이었고, 4급 합격자에게도 매우 어려운 문항은 1개 문항이었다. 같은 기준으로 볼 때, 26회 시험에서는 매우 쉬운 문항이 5개였고, 매우 어려운 문항은 없었다. 27회 시험에서는 매우 쉬운 문항이 4개, 매우 어려운 문항이 1개였다. 또 해당 문항을 제외했을 때 검사 신뢰도가 높아지는 문항은 25회에서는 3개, 26회서는 2개, 27회에서는 3개로 분석되었다. 아울러 오답의 매력도가 정답보다 높은 문항은 25회에서는 2개, 26회에서는 1개, 27회에서는 2개였다.

제25회~제27회 한국어능력시험 중급 읽기 영역의 문항 중 오답의 매력도가 정답보다 높은 문항들의 내용 분석을 시행한 결과 몇 가지 출제상의 시사점을 얻을 수 있었다. 우선 세부 내용 파악 문형을 출제함에 있어서, 문항을 어렵게 만들기 위해서 지나치게 세부적인 사항을 묻게 되면 한국어 능력을 측정하는 것이 아니라 실수가 적은가를 측정하게 된다. 따라서 세부 내용 파악 문형이 필요하다면, 난이도를 조정하여 문제의 중반으로 옮기는 것이 타당하리라 생각된다. 또, 글의 전개 양상을 파악하는 문제에 있어서도, 지문에 포함된 사소한 실마리를 제시하고 그 실마리를 이용해 앞부분에 어떤 내용이 왔어야 하는지를 따지는 방식은 실마리를 포착할 수 있는 능력을 측정하는 문항이 된다. 이러한 문형에서 피험자들은 시험 형식에 익숙해지기 전에는 자신들의 실력을 정확하게 내보이기가 어렵게 된다. 따라서 이러한 문형보다는 글의 자연스러운 흐름을 파악할 수 있는가를 평가하는 방식으로 문항 유형이 바뀌어야 할 것이다. 그리고 제시된 문항이 상당히 다양하고 종합적인 한국어 능력을 갖추어야만 풀 수 있는 문항도 있었다. 그러한 능력이 중급 수준의 한국어 능력이라는 직접적인 증거가 없고, 몇 차례의 시행을 통해 중급 응시자들이 그런 문항에 대해 어려워하는 것으로 드러났다면 해당 문형이 중급 문항으로서 적절한가에 대해 심도 있는 검토가 필요할 것이다.

이러한 분석 결과를 바탕으로 중급 읽기 영역의 평가틀에 대한 검토 결과와 개선점을 제시하면 다음과 같다. 우선 회차별로 매우 쉬운 문항과 매우 어려운 문항들이 나타나고 있는데, 이러한 문항들은 등급 판정에 도움이 되지 못하므로 차후 평가틀을 개선할 때, 해당 등급 판별에 적정한 난이도를 갖도록 개선할 필요가 있다. 두 번째로 매우 쉬운 문항과 매우 어려운 문항들이 회차별로 일관되지 못한 점은 출제진의 전문성 함양과 출제 참조 자료-수준별 어휘 목록이나 문법 목록 등의 개선을 통해 해결해야 할 문제로 판단된다. 끝으로 시험 체제와 관련하여 현행 원점수 기반의 시험 체제는 문항의 난이도 유지 여부를 출제진의 전문성에만 의존하므로 회차 간 동등성을 확보하기 위해 별도의 수준 설정이나 동등화 방안 마련 등이 필요할 것이다.

## 【참고문헌】

- 국립국제교육원, 『2012년 하반기 한국어능력시험 업무처리지침』, 국립국제교육원, 2012.
- 김정숙 외, 『한국어능력시험의 등급 부여 방식 및 평가 문항 유형 개선 방안 연구』, 한국교육과정평가원 연구보고서, 2004.
- \_\_\_\_\_, 『한국어능력시험 체제 개선 연구』, 국립국제교육원 정책연구, 2011.
- 김유정, 『한국어능력시험의 난이도 분석 연구: 제6회~제8회 시험을 중심으로』, 『한국어교육』17권 1호, 국제한국어교육학회, 2006, pp.21-46.
- 김하수 등, 『한국어 능력 평가 제도의 기본 모형 개발에 관한 최종 연구 보고서』, 한국교육과정평가원, 1997.
- 남명호 외, 『한국어능력시험 등급 기준 조정 및 문제은행 구축 방안 연구』, 한국교육과정평가원 연구보고서, 2000.
- 박도순 등, 『교육평가-이해와 적용』, 교육과학사, 2007.
- 성태제, 『문항제작 및 분석의 이론과 실제』, 학지사, 1996.
- \_\_\_\_\_, 『현대교육평가』, 학지사, 2002.
- 안주호, 『한국어능력시험 중 표현 영역 평가에 대한 연구』, 『새국어교육』79, 한국국어교육학회, 2008, pp.229-253.
- 양길석·민경석·박정진, 『한국어능력시험 구인 분석 연구』, 국립국제교육원 정책연구, 2012.
- 이관식, 『제3회, 제4회 한국어 능력시험 어휘 평가 연구』, 『인문학연구』4, 경희대학교 인문학연구원, 2000, pp.250-268.
- 이상린, 『한국어능력시험 읽기 영역에 대한 연구-제20회-제22회 고급 수준을 중심으로-』, 『교육종합연구』 제9권 2호, 교육종합연구소, 2011, pp.218-240.
- \_\_\_\_\_, 『한국어능력시험 읽기 영역 평가 틀에 대한 비판적 검토』, 『우리어문연구』43, 우리어문학회, 2012, pp.255-279.
- 이혜영, 『한국어능력시험 문항 유형 개발을 위한 기초 연구』, 한국교육과정평가원, 2006.
- 조인옥, 『한국어능력 평가 답안 분석-제12회 한국어능력시험 수험자 반응 연구-』, 『외국어로서의 한국어교육』32권, 연세대학교 언어연구교육원 한국어학당, 2007, pp.237-258.

한국교육과정평가원, 『한국어능력시험 15년사』, 한국교육과정평가원, 2011.

Millman & Greene, *The specification and development of tests of achievement and ability*, *Educational Measurement*, N. Y: Macmillan Publishing Company, 1989.

Ebel, R. L., *Measuring Educational Achievement*. Englewood Cliffs, N. J: Prentice-Hall, 1965.

**Abstract**

**Critical appraisal on setting difficulty of reading area in TOPIK**

- On the focus of intermediated level between the 25<sup>th</sup>  
and the 27<sup>th</sup> -

Hahm, Jeong-Sik · Kim, Myoung-Kwang

This study aims to evaluate test stability of TOPIK(Test of Proficiency in Korean) and explore ways to improve its evaluation frame through the quality analysis of test items and content analysis of highly attractive distractors in the area of reading test of TOPIK between 25<sup>th</sup> and 27<sup>th</sup> in the intermediate level. The results of item reactions were collected from the area of reading tests of TOPIK between 25<sup>th</sup> and 27<sup>th</sup> in the intermediate level and the sample size were 8,461, 9,029, and 10,106 cases for each test from 25<sup>th</sup> to 27<sup>th</sup>. TestAn 1.0 was used to implement item analysis by the classical test theory about these sample items. To summarize the results of item analysis of these materials, the all three tests from 25<sup>th</sup> to 27<sup>th</sup> showed the item difficulty, item discrimination and test reliability are overallly stable and adequate as a whole. When dividing test takers in the fourth level into two groups between successful ones(the high rank group) and unsuccessful ones(the low rank group), however, this test of this range(the fourth level) indicated an error in terms of item difficulty in the light of item-based analysis. Moreover, it was revealed that the test items were made irrelevant in terms of test takers' level or excessively artificial to increase the item difficulty when the highly attractive distractors were analyzed. The results of analyzing the test, therefore suggest that level-irrelevant test items should be avoided, professional qualified test item makers should be trained, test making systems should be improved and the each case of tests should be equated to improve the overall evaluation frame hereafter.



Key Word : TOPIK(Test of Proficiency in Korean), the area of reading, Item reaction theory, analysis of item, difficulty, discrimination, TestAn 1.0 program, intermediate level, quality analysis

함정식

소속 : 서울특별시 교육연구정보원 연구사

주소 : (100-873) 서울특별시 중구 소파로 46(회현동 1가 100-177)

전화번호 : 02-311-1232 / 010-2765-9757

전자우편 : hjs001@sen.go.kr

김명광

소속 : 대구대학교 국제한국어교육과 조교수

주소 : (712-714) 경상북도 경산시 대구대학교 국제학부 국제한국어교육과

전화번호 : 010-9059-0857

전자우편 : kimmk0857@daegu.ac.kr

이 논문은 2014년 2월 28일 투고되어  
2014년 3월 25일까지 심사 완료하여  
2014년 4월 8일 게재 확정됨.