

중등영어 읽기능력 평가문항 난이도 예측 변인 연구

정희정*, 김지영**

용인송담대학교, 중앙대학교

Jung, Hee-Jung & Kim, Jie-Young. (2008). Predicting the variables for difficulty of English reading proficiency test items for middle school students. *Modern English Education*, 9(1), 307-323.

This study explored the variables that affect the difficulty of English reading test items within the English Proficiency Test for middle school students. The variables in this study were hypothesized on the basis of the previous literature and the item development framework of the English Proficiency Test. Correlation analysis was conducted to arrive at variables that could predict the difficulty of test items. Six variables from the 12 hypothesized ones were found to be correlated with the proportion of correct responses to the reading questions in the English Proficiency Test. Regression analysis of the six variables showed that those six variables explained 93% of the difficult questions within the English Reading Proficiency Test.

[item difficulty/reading proficiency test/문항 난이도/읽기 능력 평가]

I. 서론

난이도는 피험자(test taker)가 갖는 각각의 다양한 영어능력 수준을 타당한 근거로 평가할 수 있게 해 주는 평가 요소이다. 난이도가 일정하여 타당도 있는 문항으로 평가를 치루는 피험자는 영어능력 평가를 신뢰할 수 있게 되고, 따라서 피험자는 자신의 영어능력을 객관적으로 평가한 결과를 바탕으로 부족한 영어능력을 보충할 수 있는 후광효과(washback effect)도 기대할 수 있다.

따라서 본 연구의 목적은 첫째, 영어능력 평가문항의 난이도에 영향을 미치는 변인을 찾아내는 데에 있으며, 둘째, 그런 변인을 찾아내어 차후 시행될

* 제1저자

** 교신저자

영어능력평가의 난이도를 예측함으로써 궁극적으로 타당성 있는 영어능력평가를 개발하는 데에 있다. 이에 영어능력 평가가 일회성으로 끝나는 것이 아니라 수차례에 걸쳐서 평가를 시행하여 피험자가 자신의 영어 능력을 객관적으로 판단할 수 있는 기준을 마련하는 것이다.

II. 이론적 배경

1. 난이도의 개념

문항 난이도(item difficulty)란 피험자가 문제를 맞히는 비율을 측정하여 한 문항의 어려운 정도를 나타내는 것이다. Kyoung-Ae Jin과 Chung Park(2004)에 따르면 평가에서 문항 난이도는 그 문항이 측정하고자 하는 구인 요소와 피험자의 능력과 연관이 있는 것으로 중요한 요소이고 특히 대학수학능력시험과 같이 시험결과가 피험자에게 중요한 영향을 미치는 고부담 시험(high-stakes test)의 경우 문항 난이도는 중요한 사항이 된다고 언급했다. 신상근(2005)은 시험의 난이도를 정확히 예측하는 것이 쉬운 일은 아니지만, 타당도 검증 측면에서 보면 문항 난이도 예측은 시험 개발자가 가정한 지식이나 능력이 시험 결과에 얼마나 영향을 미쳤는지 밝혀낸다는 점에서 구인 타당도를 확인할 수 있는 중요한 증거가 된다고 보았다. 문항 난이도를 구하는 식은 다음과 같다.

$$P = \frac{R}{N}$$

(P : 문항 난이도, R : 정답자 수, N : 총 피험자 수)

문항 난이도에 따른 문항 평가의 절대적 기준은 없으나 Cangelosi(1990)가 제시한 문항 난이도 평가기준에 따라서 0.25미만이면 매우 어려운 문항, 0.25-0.75미만이면 적절한 문항, 그리고 0.75이상이면 쉬운 문항으로 평가한다. Cangelosi(1990)는 문항 난이도에 따른 평가기준을 표 1과 같이 제시하였다.

표 1
Cangelosi의 문항 난이도 평가 기준

문항 난이도	문항평가
.25 이하	어려운 문항
.25~.75	적절한 문항
.75 이상	쉬운 문항

2. 읽기능력 평가의 난이도

1) 개념

Rowe, Ozuru와 McNamara(2006)는 전형적인 읽기능력 평가는 짧은 지문을 읽고 나서 그 지문을 기반으로 하는 선다형 문제에 대답하는 것으로 보았다. 즉, 정답을 고르는 것은 지문과 질문을 이해하고 지문 안에 내포된 잠재적인 답을 선택지에서 고르는 것인데, 이러한 활동은 읽기 활동에 포함되기 때문에 지문을 기반으로 하는 선다형 문제를 통해 읽기 능력을 측정할 수 있다고 보는 것이다. 따라서 읽기 평가의 난이도는 지문이 다루는 내용의 난이도와 그 지문을 바탕으로 제시되는 문항의 어려운 정도가 모두 작용하여 읽기 평가의 전체적인 난이도에 영향을 주게 된다.

2) 영향 변인들

(1) 언어 형태

Alderson(2000)은 질문을 제시하는 언어가 지문 자체보다 이해하기 어렵다면 피험자가 느끼는 문항에 대한 난이도는 훨씬 어려워질 것이라고 했고, Shohamy(1984)는 선다형 문제의 질문이 모국어로 제시되었을 때 같은 문제를 제2언어로 제시한 것보다는 쉽다는 것을 연구를 통해 밝혔다.

(2) 질문 형태

질문의 형태에 따라서도 문항의 난이도에 영향을 미칠 수 있는데, Pearson과 Johnson(1978)이 세 가지로 질문의 형태를 구분했다. 명시적 질문(textually explicit question)이란 질문에 대한 정보와 정답을 한 문장에서 찾을 수 있는 경우를 말하고, 암시적 질문(textually implicit question)은 여러 문장을 전반적으로 읽고 나서 정보를 조합해야 정답을 고를 수 있는 것이다. 마지막으로 함축적 질문(script-based implicit question)은 지문에 제시된 정보와 피험자가 가지고 있는 배경지식을 통해서 지문 자체에 나타나 있지 않은 정답을 찾아야 하는 것이다. 이러한 질문의 형태를 기반으로 Davey와 Lasasso(1984)가 실험을 한 결과 명시적 질문이 암시적 질문보다 쉽다는 결과가 나왔다.

(3) 지문에 대한 배경지식

많은 학자들이 읽기 능력과 배경지식이 지문의 난이도에 미치는 영향에 대해 연구한 결과, 두 변인 모두 영향을 주지만 읽기 능력이 더 영향을 준다는 결과(Hock, 1990)도 있고, 상대적으로 쉬운 지문에서는 언어 능력만 필요로 하지만 어려운 지문을 읽을 때에는 배경지식과 읽기 능력이 모두 필요하기도

하였다(Alderson & Urquhart, 1985). 또한 Clapham(1996)에 따르면 읽기능력 수준이 낮은 피험자들은 자신의 전문분야에 관한 지문도 어려워했지만, 읽기능력 수준이 높은 피험자들은 자신의 전문 분야가 아니더라도 지문을 읽는데 어려움을 덜 느낀다고 한다.

(4) 어휘에 대한 지식

Qian(2002)은 19개국의 학생들 217명에게 2000년 시행된 TOEFL을 바탕으로 읽기 문항과 피험자가 가지고 있는 어휘지식의 정도를 재는 기준인 The Depth of Vocabulary Knowledge Measure로 상관관계를 분석한 결과 .77($p < .01$)로 높은 상관관계를 보였고, 어휘가 읽기문항의 난이도를 60% 설명할 수 있다고 하였다. Schmitt(1999)는 TOEFL의 읽기 문항 중 어휘관련 문항을 영어를 제 2언어로 하는 학생들 30여명을 대상으로 풀게 한 후 어휘 문항에 대한 구인 타당도를 알아본 결과 어휘를 몰랐다고 대답한 인터뷰 83건 중 53건이 어휘 문항의 정답을 맞힌 사실을 발견했다. 피험자 스스로가 단어를 아는지에 대한 인터뷰 문항에 ‘잘 모르겠다’라고 대답했지만, 자신도 모르게 알고 있었던 단어일 경우도 있고, 정답을 제외한 선택지에서 힌트를 얻을 수도 있기 때문이다. 이것은 Johnston(1984)이 지문의 내용과 연관된 어휘 평가문항은 어휘 자체에 대한 지식보다 기존에 알고 있던 지식에 의존하는 경우가 많다는 연구 결과와도 상당히 관계가 있다.

(5) 문장의 길이

Kyoung-Ae Jin과 Chung Park(2004)은 2003년에 시행된 대학수학능력시험의 영어 영역 문항과 시험을 치른 학생들의 응답률을 바탕으로 문항 난이도를 46개의 변인을 가정하여 상관관계 분석과 다중회귀분석을 실시하였는데 여기에서 흥미로운 점은 지문을 구성하는 단어의 수 즉, 지문의 길이가 긴 것은 문항을 어렵게 만들지 않지만, 지문의 길이에 관계없이 한 문장 안에 단어 수가 많을수록 문항이 어렵다는 결과가 나온 것이다.

이 외에도 평가문항 외적 요소인 평가 장소, 평가 진행 방법, 피험자의 태도가 평가의 난이도에 영향을 줄 수 있는데, 김현진(2004)에 따르면 언어에 대한 불안감이 학습자의 실제 언어 능력을 평가할 때 큰 영향을 미치지 않는다고 하였다. 따라서 비슷한 조건의 장소에서 동일한 평가 진행 방법으로 평가를 치를 때 평가 난이도에 영향을 가장 많이 주는 요소로 문항 난이도를 들 수 있겠다.

3. 문항 난이도 예측 선행연구

Freedle과 Kostin(1991)은 SAT 읽기 문항의 난이도 예측 변인을 문항 관련변인(item variables), 지문 관련 변인(text variables), 문항·지문 관련변인(item/text variables)의 세 가지로 분류하였다. 지문 관련 변인은 지문의 특성과 관련된 변인으로 지시어, 의문문과 부정어의 빈도, 문단 전개 방식, 어휘의 복잡성, 문장의 수와 길이, 내용의 구체성, 제대 생소성과 지문의 응집력 변인이 문항 난이도 예측과 상관이 높고, 단서의 위치가 난이도 예측과 상관관계가 있다고 했으며 문항 관련 변인에서는 추론을 요구하는 문항일수록 어렵다고 했다. 또한 문항과 지문에 중복되어 관련된 변인에 의해 난이도에 영향을 받을 수 있다고 했다. 즉, 문제를 풀기 위해서는 질문을 읽고 정답과 관련된 문장이나 핵심 단어를 찾기 위해 지문을 읽게 되는데, 정답과 관련된 단서가 지문에서 어디에 위치하는지에 따라 문항의 난이도가 달라질 수 있다는 것이다.

김용석(2003)은 토익 시험의 읽기 부분에서 지문에 관한 변인과 문항에 관한 변인 또는 지문과 문항에 중복으로 나타나는 변인에 의해 문항난이도가 얼마나 적절하게 설명되는지에 관하여 연구하였다. 여기에서 토익시험의 문항난이도는 피험자의 영어능력과는 관계없이 지문과 관련한 변인보다는 문항과 관계한 변인에 더 영향을 받는 것으로 나타났다.

Kyoung-Ae Jin과 Chung Park(2004)은 2003년에 시행된 대학수학능력시험(이하 대수능)의 영어 영역 문항과 시험을 치른 학생들의 응답률을 바탕으로 문항 난이도를 46개의 변인을 가정하여 355,146명의 문과 지원자들의 시험 점수와 상관관계 분석을 실행한 결과 14개의 변인이 문항 난이도에 영향을 미치는 것으로 나타났고, 다중회귀분석한 결과 피험자의 정답률의 49%를 예측하였다.

장경숙(2005)은 대수능 중 외국어(영어)영역 읽기 문항의 난이도 관련 문항 검토 도구를 제작하여 인문계 상위 50% 수험생 집단을 기준으로 하여 정답률 예측 모형을 개발하였다. 안정성과 예측력에 있어서 탁월한 모형을 최종적으로 선정하여 그 모형의 설명력과 오답률에 대한 독립변인의 영향력도 비교해 보았다. 그 결과 대수능 외국어(영어)영역 읽기 문항 난이도 관련 변인으로 ‘추론 정도’¹, ‘문법성 판단’과 ‘선택지 매력도’ 변인들이 오답률에 대한 설명력이 있다는 결론을 얻게 되었고, 모형의 예측력을 확인해 보기 위해서 최적 예측 모형을 2004년 9월 모의평가의 정답률 평균을 예측하여 실제 정답률 평균과 비교해 본 결과, 예측 정답률은 65.99점이고 실제 정답률은 67.11로 1.12점 낮게 예상하였으나 근접한 값을 예측하고 있었다.

¹ 여기에서 추론 정도란 피험자들이 문제를 풀기 위해 요구되는 사고의 정도와 문제 해결을 위해 소요되는 노력의 정도를 말한다.

III. 연구방법

중등 읽기능력 평가문항 난이도 예측모형 개발은 다음과 같은 단계를 밟아 개발되었다.

첫째, 본 연구에서 사용되는 중등영어 읽기능력 평가문항이 믿을 수 있는 문항인지 알아보기 위해 통계적으로 유의미한 검증절차를 밟았다.

둘째, 본 연구에서 사용되는 중등영어 읽기능력 평가문항 30문항 각각에 대해 피험자 2,010명의 정답률을 알아보고, 난이도 관련 변인 연구에 대한 선행연구와 본 연구에서 사용된 평가문항의 개관을 통해 중등 읽기능력 평가문항의 난이도에 영향을 미쳤을 것으로 판단되는 변인을 찾아내었다.

셋째, 선행연구를 통해 추측된 변인들과 실제 피험자들의 시험결과 즉 정답률 사이의 상관관계 분석을 통해 난이도에 영향을 주는 변인을 선별하고, 난이도와 상관관계가 있는 변인들을 회귀분석 하였다. 각 결과는 다음에 기술하는 바와 같다.

1. 자료수집

1) 영어읽기 능력평가 결과

난이도 변인 연구를 위해 조기영어교육과 초·중등학생 중심의 영어교육을 담당하는 국내 A 사교육기관에서 2007년에 전국 중학생 1,2,3학년 2,010명을 대상으로 실시한 30회 영어능력평가 60문항 중 읽기능력평가에 대한 30문항의 결과를 수집하였다. 먼저 본 연구에서 사용된 읽기평가가 믿을만한 평가문항인지를 검증하기 위해 기술통계와 신뢰도, 타당도를 분석하였다. 읽기 영역의 총점 30점에서 평균은 14.88이고, 표준편차는 2.576으로 학생들 사이에 점수 차이가 많지 않다는 것을 보여주었다. 표 2는 영어 읽기영역에 대한 기술통계를 나타낸 것이다.

표 2
읽기영역 기술통계

참가인원	평균	표준편차	왜도	첨도
2,010	14.88	2.576	-0.098	1.311

또한 신뢰도를 분석함으로써 평가가 믿을만한 평가인지 알아보고, 평가가 측정하는 부분이 올바른 이론을 내포하고 있는지에 대한 타당성을 알아보았다. 표 3은 읽기능력 평가의 신뢰도 분석 결과이다.

표 3
평가 신뢰도 분석 결과

Cronbach 알파	표준 항목을 근거로 한 Cronbach 알파	응시자 수
.811	.810	30

이번 평가에서 읽기 영역에 대한 신뢰도를 Cronbach's alpha를 사용하여 분석한 결과, 표 3과 같이 0.811이 높은 수치로 나왔으므로 높은 신뢰성을 갖는다고 볼 수 있다.

타당도 분석은 Pearson 상관관계 지수를 이용해서 측정하였다. 표 4는 영어 읽기능력 평가가 평가하고자 하는 영역을 제대로 평가했는지를 나타내 준다.

표 4
평가 타당도 분석 결과

	주제찾기	추론하기	의미파악	응집력	세부사항	유추
주제찾기	1					
추론하기	.330(**)	1				
의미파악	.301(**)	.421(**)	1			
응집력	.359(**)	.501(**)	.469(**)	1		
세부사항	.488(**)	.493(**)	.390(**)	.457(**)	1	
유추	.300(**)	.371(**)	.350(**)	.443(**)	.368(**)	1

상관관계 지수가 0.3 이상으로 나타나면 각 변인들이 어느 정도 서로 관계가 있다는 것을 보여 주는데, 표 4에서 나타나듯이 읽기 능력을 이룬다고 가정한 여섯 가지 영역들이 서로 긍정적인 상관관계를 보여주고 있다. 따라서 본 연구에서 사용된 중등영어 읽기능력 평가문항은 믿을만한 평가문항으로 검증되었다.

2) 읽기능력 문항별 정답률

본 연구에서 사용된 영어 읽기능력 30문항에 대해 피험자 2,010명의 평가 결과를 가지고 각 문항별 정답률을 알아보았다. 표 5에 제시된 각 문항별 정답률은 난이도 관련 변인에 관련된 문항과 상관관계를 분석할 때 필요한 자료이다. 표 5에서와 같이 31번 문항부터 60번까지 읽기 문항에서 각 문항별 정답률을 알아 본 결과로 31번 문항은 정답률 0.5로 피험자 2,010명 중 50%가 정답을 맞힌 것으로 설명된다.

표 5
영어 읽기능력 문항별 정답률

문항 번호	평균	표준편차	문항 번호	평균	표준편차
31번	.50	.500	46번	.29	.454
32번	.71	.455	47번	.63	.484
33번	.50	.500	48번	.46	.499
34번	.60	.489	49번	.32	.468
35번	.50	.500	50번	.63	.482
36번	.83	.372	51번	.58	.493
37번	.68	.467	52번	.48	.500
38번	.50	.500	53번	.31	.462
39번	.74	.438	54번	.63	.482
40번	.66	.474	55번	.27	.442
41번	.48	.500	56번	.47	.499
42번	.30	.456	57번	.25	.433
43번	.61	.487	58번	.32	.466
44번	.49	.500	59번	.46	.499
45번	.46	.499	60번	.26	.440

Cangelosi(1990)의 문항 난이도 평가 기준에 따라 31번 문항의 난이도는 0.50으로 0.30-0.80 사이에 속하므로 난이도 중의 적절한 문항이라고 볼 수 있다. 반면 36번, 39번은 정답률이 각각 0.83과 0.74로 쉬운 문항 즉, 난이도가 하인 문항에 속하고, 57, 60번 문항은 정답률이 0.25에 가까우므로 어려운 문항으로 난이도가 상에 속한다.

2. 난이도 관련 변인 추출 및 검토 도구 개발

영어 읽기능력 평가문항의 난이도에 영향을 주는 변인을 찾아내기 위해서, 기존의 선행연구들과 Freedle과 Kostin(1991, 1992)에 의해 사용된 변인들, 그리고 Kyoung-Ae Jin과 Chung Park(2004)에 의해 사용된 난이도 예측 변인을 응용하였다. A사에 소속된 7명의 평가 문항 개발자들은 A사가 개발한 평가문항 도구 설계에 기초하여 그에 맞는 각 영역별 문항을 개발하였다. 따라서 본 연구에서는 영어 읽기능력 평가문항 설계 개관에 맞도록 6개의 영역을 기반으로 문항의 종류별로 구분하여 문항 변인을 중심으로 난이도에 영향을 미치는 변인을 찾고자 하였다.

본 연구에서는 질문의 형태를 영어로 제시하였으나, 평가 대상인 중학교 영어 학습자들은 수업시간 뿐만 아니라 기본에 영어 평가의 경험이 있다고 가정하여 질문의 언어 형태에 대한 난이도 변인을 제외하기로 하였다. 표 6은 영어 읽기능력 평가문항을 영역별로 구분하여 난이도 관련 변인을 추출한 것이다.

표 6
영역별 난이도 관련 변인 추출

영역	변인	문항 번호
1. 주제(gist)	v1. 주제 찾기 (understanding main ideas)	31
		50
		54
		59
	v2. 제목 찾기 (understanding titles)	36
		v3. 분위기 이해 (understanding atmosphere)
2. 유추 (inference)	v4. 관계 유추 (relationships)	
		v5. 속담 유추 (proverbs)
	v6. 내포하는 의미 (understanding meaning from context)	
		53
3. 세부정보 파악 (details)	v7. 사실적 이해 (literal understanding)	58
		32
		35
		42
4. 단어나 글자 이해 (guessing meaning from the context)	v8. 단어 이해 (understanding words)	51
		60
		34
		38
5. 응집력(cohesion)	v9. 구삽입 (inserting a phrase)	44
		49
		56
	v10. 문장삽입 (inserting a sentence)	52
		v11. 글의 순서 (coherence)
6. 추론 (reference)	v12. 지시어 이해 (understanding references)	
		40
		43
		47

3. 자료 분석

중등영어 읽기능력 평가의 문항 난이도에 영향을 주는 변인을 찾기 위해 추출된 12개의 변인과 2,010명의 피험자의 정답률 사이의 상관관계를 알아보았다. 또한 문항 난이도에 영향을 주는 변인으로 회귀분석을 통해 난이도의

예측율을 알아보았다. 분석에 사용된 프로그램은 SPSS 12.0이다.

IV. 연구 결과 및 논의

1. 상관관계 분석결과

영어 읽기 능력 변인별 상관관계를 살펴본 결과, 12개의 변인이 모두 읽기 능력 평가문항의 난이도와 상관관계를 갖는 것으로 나타났다. 다음에 제시된 표 7은 난이도에 영향을 주는 변인들과 정답률과의 상관관계를 나타내고 있다.

표 7

난이도 관련 변인과 읽기 영역 정답률 간 상관관계

	정답률	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11
v1	.442 (**)											
v2	.442 (**)	.229 (**)										
v3	.335 (**)	.098 (**)	.057 (*)									
v4	.413 (**)	.172 (**)	.233 (**)	.079 (**)								
v5	.429 (**)	.054 (*)	.099 (**)	.151 (**)	.127 (**)							
v6	.499 (**)	.108 (**)	.125 (**)	.159 (**)	.140 (**)	.184 (**)						
v7	.747 (**)	.211 (**)	.233 (**)	.248 (**)	.216 (**)	.313 (**)	.281 (**)					
v8	.682 (**)	.218 (**)	.305 (**)	.117 (**)	.244 (**)	.247 (**)	.236 (**)	.414 (**)				
v9	.727 (**)	.219 (**)	.262 (**)	.207 (**)	.258 (**)	.289 (**)	.332 (**)	.473 (**)	.425 (**)			
v10	.582 (**)	.228 (**)	.254 (**)	.172 (**)	.232 (**)	.247 (**)	.195 (**)	.365 (**)	.355 (**)	.376 (**)		
v11	.330 (**)	.080 (**)	.106 (**)	.119 (**)	.112 (**)	.085 (**)	.145 (**)	.189 (**)	.159 (**)	.159 (**)	.164 (**)	
v12	.713 (**)	.245 (**)	.355 (**)	.159 (**)	.279 (**)	.260 (**)	.239 (**)	.412 (**)	.431 (**)	.408 (**)	.432 (**)	.195 (**)

주. v1. 주제 이해, v2. 제목 찾기, v3. 분위기 이해, v4. 관계 유추, v5. 속담 유추, v6. 내포하는 의미, v7. 사실적 이해, v8. 단어 이해, v9. 구삼입, v10. 문장삼입, v11. 글의 순서, v12. 지시어 이해.

표 7에서와 같이 v7. 사실적 이해와 정답률이 $r=.747(p<.01)$ 로 가장 상관성이 높았으며 그 다음으로 v9. 구삽입($r=.727, p<.01$), v12. 지시어의 이해($r=.713, p<.01$)의 순이었으며 다른 변인들과도 $r=.330(p<.01)$ 에서부터 $r=.682(p<.01)$ 까지의 정적인 관계를 보였다. 그 중에서 6개의 변인이 난이도를 쉽게 만드는 데 영향을 주거나 어렵게 만드는 변인으로 나타났다. 즉, v7. 사실적 이해, v8 단어 이해, v9. 구삽입, v12. 지시어의 이해와 관련된 변인들이 문항 난이도를 쉽게 만들었다고 볼 수 있다. 반면 문항 난이도를 어렵게 만든 문항으로는 v3. 분위기 이해와 v11. 글의 순서와 관련된 변인이라고 볼 수 있다.

문항 난이도를 쉽게 만드는 변인을 통해 문항의 형태를 분석해 본 결과, v7. 사실적 이해(.747)에 해당하는 문항의 형태는 지문을 읽고 선택지에서 본문의 내용과 일치하는 것을 고르거나 사실이 아닌 것을 고르는 문제는 명시적 질문에 의해 답을 찾는 문제이므로 난이도가 쉬워지는 데에 영향을 미친 것으로 보인다. 이는 명시적 질문이 암시적 질문보다 난이도를 쉽게 한다는 선행연구와 같은 결과로 나왔다(Bensoussan, Sim & Weiss, 1984; Pearson & Johnson, 1978;).

v9. 구삽입(.727) 변인은 지문을 읽고 빈 칸에 알맞은 연결어를 넣어서 문맥이 자연스럽게 연결되도록 하는 문항이었다. Jonz(1987)는 미국 대학에서 원어민과 비원어민을 대상으로 두 가지 cloze test를 실시한 결과 평가를 치루기 전에 원문을 읽게 해준 그룹에서는 원어민과 비원어민의 평가 결과가 비슷했지만, 원문을 보여 주지 않고 단어를 넣거나 지시어를 넣는 평가에서는 비원어민이 예측 가능한 연결어를 넣을 수 있는 능력이 원어민보다 훨씬 낮았다. 이 연구로 미루어 보아 EFL환경에서 응집력에 관한 평가 문항은 난이도에 영향을 줄 수 있는 변인이 될 수 있지만 본 연구에서는 정답률과 상관관계가 높은 것으로 보아 문항의 난이도를 쉽게 만드는 변인으로 볼 수 있다. 이는 연결어에 관한 문항에 대해 피험자들이 많이 접하고 연습을 통해 예측 타당도를 높이는 결과를 보인 것으로 보인다.

v12. 지시어 이해(.713) 변인과 관련된 문항으로는 문제의 형태가 모두 본문의 밑줄 친 대명사 중 다른 것을 지칭하는 것을 묻는 문제로 지문을 신중하게 읽어보면 나머지 셋과 다른 하나를 고를 수 있는 문제이다. 이런 문제의 형태는 기존에 많이 다룰 수 있는 문제이므로 피험자들이 문제에 익숙해져 있거나 지문 내용이 쉬었을 수도 있다. 이것은 Freedle과 Kostin(1991)의 연구에서 추론에 관련된 문항이 어렵다는 결과와는 다르게 추론에 관한 변인 중 v12. 지시어 이해 변인은 난이도를 쉽게 하여 정답률을 높이는 것으로 나타났다.

v8. 단어 이해(.682)에 관한 문항은 본문의 밑줄 친 단어와 밀접한 의미를 같은 단어, 즉 동의어를 찾는 문제로 이루어져 있다. 단어 이해와 관련된 변인은 난이도를 쉽게 만드는 또 다른 변인인 사실적 이해, 구삽입, 지시어 이

해관련 변인들이 나타내는 상관관계와 비교할 때 대체로 쉬운 문항이라고 볼 수 있는데 Schmitt(1999)와 Johnston(1984)의 연구에서와 같이 피험자가 주어진 어휘의 뜻을 잘 모르더라도 지문의 내용을 읽고 답을 유추해내는 과정을 거치기 때문이라고 볼 수 있다.

문항 난이도를 어렵게 만드는 변인으로 추출된 v3. 분위기 이해와 v11. 글의 순서를 살펴보면 두 변인에 관한 문항들의 지문의 내용이 피험자들에게 익숙하지 않거나 어휘가 어려운 경우일 수 있다고 보인다. v3. 분위기 이해 변인에서 대표되는 문항의 지문은 학교를 주제로 하지만 다루는 내용이 어렵고 문장의 구조가 복잡한 것이 문항의 난이도를 어렵게 만드는 이유일 수도 있고 또 다른 이유로는 선택지에 제시된 어휘의 생소한 정도가 난이도를 어렵게 만드는 원인이 될 수 있다. Kobayashi(2002)가 754명의 일본 대학생을 대상으로 지문의 구성과 형식이 영어능력평가에 미치는 효과에 대해 연구한 결과 지문의 구성과 피험자의 영어능력과는 유의미한 상관관계를 보인다고 하였다. 따라서 v11. 글의 순서에 관한 문항은 지문의 내용이 생소할 수 있고 지문을 구성하고 있는 문장의 구조가 복잡하거나 어휘의 생소한 정도가 난이도를 어렵게 만드는 원인으로 보인다.

Kyoung-Ae Jin과 Chung Park(2004)의 연구에서 문항 난이도를 쉽게 만드는 변인인 v1. 주제찾기와 v3. 분위기 파악에 관한 변인은 본 연구결과에서는 문항 난이도를 쉽게 만드는 변인이 아닌 것으로 나왔다. v1. 주제찾기는 정답률과의 상관관계가 .442로 적절한 수준의 난이도를 보여주었으며, v3. 분위기 이해에 관한 변인은 정답률과의 상관관계가 .335로 문항 난이도를 어렵게 하는 변인으로 분석되었다.

2. 회귀분석 결과

영어읽기 능력평가의 난이도에 영향을 미치는 변인으로 가정된 변인들과 피험자들의 정답률간의 상관관계에서 난이도가 어려워지는 변인으로 추출된 6개의 변인을 중심으로 난이도를 예측할 수 있도록 다중공선성(multicollinearity) 문제에 유의하여 회귀분석을 하였다.

표 8은 정답률(종속변인)과 6개의 변인(독립변인)을 회귀분석한 결과이다.

² 다중공선성이란 다중회귀분석에 채택된 독립 변수들간에 높은 상관관계가 나타나 현상을 가리키는 말이다. 다각도로 이루어진 다중공선성의 탐색을 바탕으로 공선성이 존재하고 있다고 판단이 될 경우, 가장 간단한 해결방법으로는 관련된 변수들 중 적당한 변수를 자료를 수집한 사람과 논의하여 회귀모형에서 제거하는 것이다(김충락, 강근석, 1999). 또는 다중공선성의 문제는 공차한계(Tolerance)와 분산팽창요인(VIF)의 값을 기준으로 해결할 수 있다. 먼저, 공차한계는 1 이하의 값, 분산팽창요인은 10 이하이며, 1에 가까울수록 다중공선성이 낮다고 할 수 있다

표 8

정답률과 변인별 회귀분석 결과

	비표준화 계수		표준화 계수	t	공선성 통계량	
	B	표준오차	베타		공차한계	VIF
(상수)	.070	.003		23.181		
v3	.043	.003	.100	15.837***	.922	1.084
v7	.040	.001	.327	43.933***	.669	1.495
v8	.043	.001	.258	35.653***	.708	1.412
v9	.047	.001	.299	40.492***	.679	1.472
v11	.040	.002	.108	17.197***	.940	1.064
v12	.047	.001	.308	42.673***	.712	1.404

F= 4168.010*** R제곱=.926

주. v1. 주제 이해, v2. 제목 찾기, v3. 분위기 이해, v4. 관계 유추, v5. 속담 유추, v6. 내포하는 의미, v7. 사실적 이해, v8. 단어 이해, v9. 구삽입, v10. 문장삽입, v11. 글의 순서, v12. 지시어 이해.

표 8에서 보는 바와 같이 정답률의 예측변인을 살펴본 결과 전체적 회귀식은 $F=4168.010(p<.001)$ 로 나타나 유의한 회귀식이었으며 설명력은 92.6%를 나타내었다. 각 변인별로 정답률을 예측하는 변인을 살펴본 결과 분위기 이해, 사실적 이해, 단어 이해, 구삽입, 글의 순서, 지시어 이해 등이 난이도에 영향을 주는 것으로 나타났다. 특히 B (베타)값³의 크기순으로 보아 v7. 과 v12. 지시어 이해와 관련된 변인의 B 값이 높을수록 정답률이 높아지는 것을 볼 수 있다. 따라서 난이도를 쉽게 만드는 예측변인이 됨을 가장 잘 설명하였으며 그 다음으로 v9. 구삽입, v8. 단어이해, 그리고 v3. 분위기 이해와 v11. 글의 순서와 관련된 변인의 B 값이 낮을수록 정답률을 낮추는 것을 볼 수 있다. 즉, 분위기 이해하는 문항과 글의 순서를 묻는 문항이 난이도를 어렵게 만드는 요인으로 볼 수 있는 것이다.

V. 결론 및 제언

영어능력이 중요시 되는 현대 사회 풍조와 함께 자신의 영어능력을 평가하는 평가도구가 일정한 난이도를 유지하는 것은 중요한 문제로 대두되고 있다. 또한 문항 난이도는 평가가 재려고 하는 구인 타당성을 검증하는 데 있어서 중요한 역할을 한다(Kyoung-Ae Jin & Chung Park, 2004). 따라서 본 연구의 목적은 주기적으로 시행되는 영어 능력 평가 문항의 난이도에 영향을

³ 베타값이란 각각의 독립변수가 하나 증가할 때 종속변수(정답률)가 얼마나 증가하는지 예측하는 기준이 된다. 또한 여기서 t 값이 유의해야 영향을 미치게 된다.

미치는 변인을 찾아내어 난이도를 예측하여 평준화된 문항을 개발하는 데에 있으며, 나아가 피험자들이 타당도 있는 평가 문항을 통해 신뢰도 있는 평가 결과를 얻었을 때, 평가결과를 통해 자신의 취약점을 찾아 보완하고 자신의 영어능력을 향상시킬 수 있도록 하는 데에 있다.

Kyoung-Ae Jin과 Chung Park(2004)의 연구에 따르면 문항 난이도를 쉽게 만드는 변인인 주제찾기와 분위기 파악에 관한 변인은 본 연구결과에서는 문항 난이도를 쉽게 만드는 변인이 아닌 것으로 나왔으나, 오히려 분위기 이해에 관한 변인은 문항 난이도를 어렵게 하는 변인으로 분석되었고 Freedle과 Kostin(1991)의 연구에서 추론과 관련된 변인이 난이도를 어렵게 하는 것과는 달리 오히려 난이도를 쉽게 만드는 변인이 되었다. 회귀분석 결과에 따르면, 상관관계 분석을 통해 난이도에 영향을 미치는 변인인 v3. 분위기 이해, v7. 사실적 이해, v8. 단어 이해, v9. 구삽입, v11. 글의 순서, v12. 지시어 이해가 영어 읽기능력 평가의 정답률의 93%(R제곱=0.926)를 설명할 수 있었다.

이러한 연구 결과는 본 평가에서 사용된 평가도구를 가지고 문항을 개발할 경우 영어능력평가의 문항 난이도를 예측할 수 있다. 그러나 이 연구에서는 영어 읽기능력 평가문항에 대해서 난이도에 영향을 미치는 변인을 문항 관련 변인에서만 찾았기 때문에 향후 연구에서는 지문 관련 변인과 지문과 문항에 중복 관련된 변인을 찾아서 좀 더 정확한 난이도 예측이 될 수 있도록 해야 할 것이며 각 영역을 측정하는 문항의 수도 적절하게 배분해서 객관성을 높여야 할 것이다. 다시 말해서 지문 관련 변인이란 지문에 제시된 단어의 수, 문장의 수, 구로 시작하는 문장의 빈도수로 문항의 난이도를 예측할 수도 있고, 지문 안에 포함되는 세미콜론, 콜론, 대시의 개수 또는 지문에 제시된 부정어의 수와 같은 문장 구조의 복잡성과 관련된 변인을 찾아 문항의 난이도를 예측하는 변인을 찾을 수 있겠다. 지문에서 다루고 있는 소재의 생소한 정도나 지문을 구성하는 어휘나 숙어의 생소한 정도로 난이도에 영향을 주는 변인을 찾아낼 수도 있을 것이다. 문항과 지문에 중복 관련된 변인으로 문제 해결을 위해 필요한 단어가 지문에서 위치하는 정도를 통해서 문단의 도입부분에 위치할 경우 쉬운 문항이 될 수 있고, 문단의 중간에 위치할 경우 어려운 문항으로 구분할 수도 있겠다. 또한 난이도에 영향을 줄 수 있는 변인으로 피험자의 배경지식이나 피험자가 갖추고 있는 영어능력 그리고 피험자의 특성에 따라 난이도에 영향을 미칠 수도 있을 것으로 보이고, 선다형 문항으로 평가를 실시할 때 선택지가 4지 선다형이나 5지 선다형일 때 느끼는 난이도에 대해서도 연구할 필요가 있겠다. 피험자의 연령이나 수준에 따라 선택지의 개수가 달라질 수도 있겠지만 선다형 문항이 피험자의 영어 능력을 정확하게 측정할 수 있도록 하려면 무엇보다도 적절한 난이도를 유지할 수 있는 선택지를 개발하여 타당도 있는 문항을 개발해야 할 것이다. 영어능력을 평가하는 것이 언어의 4기능 중 어느 하나만을 측정해서 나올 수

있는 결과가 아니기 때문에 읽기능력 평가문항 뿐만 아니라, 듣기, 쓰기, 말하기 능력 평가 문항에서도 각 평가에 알맞게 난이도 관련 변인을 찾는 연구를 기반으로 한 타당도 있는 평가를 통해 피험자들이 신뢰할 수 있는 평가도구가 개발되어야 하겠다.

참고문헌

- 계영아. (2006). *초등학교 6학년 학생들의 영어 듣기, 읽기 능력 평가 도구 개발*. 박사학위논문. 중앙대학교, 서울.
- 김용석. (2003). 토익읽기 선다형 문제의 구인타당성과 문항난이도 예측 연구. *영어교육연구*, 15(4), 191-214.
- 김충락, 강근석. (1999). *회귀분석*. 서울: 교우사.
- 김현진. (2004). 자각 언어 능력, 실제 언어 능력, 언어 불안감 간의 관계와 자기 평가 경향에 미치는 영향. *현대영어교육*, 5(2), 68-85.
- 신상근. (2005). 듣기 평가의 문항 난이도 예측. *어학 연구*, 41(3), 683-692.
- 장경숙. (2004). 대학수학능력시험 외국어(영어)영역 읽기 난이도 예측 모형 개발. *외국어교육* 11(1), 111-130.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., & Urquhart, A. H. (1985). The effect of student's academic discipline on their performance on ESP reading tests. *Language Testing*, 2(2), 192-204.
- Bensoussan, M., Sim, D., & Weiss, R. (1984). The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations. *Reading in a Foreign Language*, 2(2), 262-276.
- Cangelosi, J. (1990). *Designing tests for evaluating student achievement*. New York: Longman.
- Clapham, C. (1996). *The development of IELTS: A study of effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.
- Davey, B., & Lasasso, C. (1984). The interaction of reader and task factors in the assessment of reading comprehension. *Experimental Education*, 52, 199-206.
- Freedle, R., & Kostin, I. (1991). *The predication of SAT reading comprehension item difficulty for expository prose passages*. (ETS research report RR-91-29). Princeton, NJ: Educational Testing Service.

- Freedle, R., & Kostin, I. (1992). *The predication of GRE reading comprehension item difficulty for expository prose passages for each of three item types: Main idea, inferences and explicit statements*. (ETS research report RR-91-59). Princeton, NJ: Educational Testing Service.
- Hock, T. (1990). The role of prior knowledge and language proficiency as predictors of reading comprehension among undergraduates. In J. H. A. L. Jong & D. K. Stevenson (Eds.), *Individualizing the assessment of language abilities* (pp. 214-224). Clevedon, PA: Multilingual Matters.
- Johnston, P. (1984). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly*, *XIX*(2), 219-239.
- Jin, Kyoung-Ae, & Park, Chung. (2004). The prediction of English item difficulty in college scholastic ability test. *English Teaching*, *59*(1), 267-278.
- Jonz, J. (1987). Textual cohesion and second-language comprehension. *Language Learning*, *37*(3), 409-438.
- Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing*, *19*(2), 193-220.
- Pearson, M., & Johnson, D. (1978). *Teaching reading comprehension*. New York, NJ: Holt, Rinehart and Winston.
- Qian, D. (2002). *Validation the role of depth of vocabulary knowledge in assessing reading for basic comprehension tasks in TOEFL 2000*. Princeton, NJ: Educational Testing Service.
- Rowe, M., Ozuru, Y. & McNamara, T. F. (2006). An analysis of standardized reading ability test: What do questions actually measure? Proceedings of the 7th international conference on Learning Sciences. International Conference on Learning Science. *International Society of the Learning Science*, 627-633.
- Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing*, *16*(2), 185-219.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, *1*(2), 147-170.

정희정

송담대학교

경기도 수원시 팔달구 우만2동 월드메르디앙 101-405

Tel: (011) 9638-4442

Email: coolij032@hotmail.com

김지영
중앙대학교 외국어대학 영어학과
경기도 대덕면 내리
Tel: (031) 670-3154
Email: jk221@cau.ac.kr

Received in January, 2008
Reviewed in February, 2008
Revised version received in March, 2008