

듣기 평가의 문항 난이도 예측

신 상 근
(이화여자대학교)

Shin, Sang-Keun. (2005). The prediction of item difficulty in a listening test. *Language Research* 41(3), 683-692.

The present study examined the construct validity of emphasis items in the listening section of Michigan English Language Assessment Battery by assessing the extent to which relevant factors were responsible for test takers' performance on the items. Three relevant variables were identified, and the item difficulties were regressed on these variables. The results showed that prosodic prominence was the only significant predictor of item difficulty, thereby supporting the interpretation that these items measure test takers' ability to process the meaning conveyed by prosodic aspects of test input. Alternative approaches to examining construct validity issues are briefly described.

Key words: listening test, item difficulty, construct validity

1. 서론

언어평가를 개발하거나 사용하는 사람들은 종종 본인들이 가정한 지식이나 능력이 수험자들의 시험 결과에 실제로 관련되어 있는지 의문을 품게 된다. 만약 특정한 평가에서 측정하고자 한 구인(construct)과 관련된 요소가 평가 문항의 난이도에 영향을 주었다거나 또는 주지 않았다는 사실을 밝혀내지 못한다면 시험 점수에 근거해서 내리는 각종 해석이나 결정의 적절성 또한 지시되기 어려울 것이다 (Bachman, 1990). 따라서 시험 문항의 난이도 예측은 구인타당도 (construct validity)를 검증할 수 있는 중요한 증거의 하나이다.

듣기 기능의 분류(taxonomy of listening subskills)를 시도한 여러 학자들은 대부분 제 2언어 학습자가 연결된 담화 속의 강세를 파악하는 능력이 중요하다는 사실을 지적하고 있다(Hughes, 1989; Munby, 1978; Richards, 1983; Rost, 1994). 예를 들어 Richards (1983)는 듣기 능력의 하위 구성요소에 억양과 강세의 기능과 역할을 인식하는 능력을 포함시켰고 Munby (1978)의 의사소통기능 분류에도 담화 속에서 정보 단위를 표현하거나 강조나 대조를 위해 사용되는 강세의 역할을 인식하는 능력이 포함되어 있다. 실제로 강세와 억양은 분절적(segmental) 소리보다도 단어를 인식하는 데 더 중요하고, 청자가 담화의 의미를

해석하는데에도 직접적인 영향을 미친다(전상범, 1995; 한중임, 2001; McNerney & Mendelsohn, 1992). 이처럼 운율요소(prosodic element)가 듣기에서 차지하는 중요성에도 불구하고 수험자에게 이런 능력이 있는지를 알아보려는 연구나 시도는 지금까지 별로 이루어지지 않았다. 언어평가의 관점에서 보면 평가에서 이러한 중요한 능력을 측정하지 않을 경우 Messick (1989)이 우려한 구인 과소 표본(construct underrepresentation) 문제를 낳을 수 있다.

본 논문에서 분석한 Michigan English Language Assessment Battery (MELAB)의 듣기 영역에 나오는 강조형 문항(emphasis items)은 담화 속의 운율 정보를 파악할 수 있는 능력을 측정하려는 첫 번째 시도라고 볼 수 있다. 그러나 이 문항들이 실제로 측정하고자 하는 능력을 제대로 측정하는지는 아직까지 밝혀지지 않았는데 본 연구에서는 구인타당도를 검증할 수 있는 여러 가지 방법 중에서 문항 난이도 예측을 통해서 강조형 문항의 구인타당도를 조사해 보았다. 본 연구에서 문항 난이도 예측 방법을 선택한 이유는 강조형 문항이 학습자가 화자의 발화 속의 운율적 요소에 의해 전달되는 의미를 처리하는 능력이 있는지를 측정하기 위해 제작되었기 때문에 수험자의 운율적 요소를 파악하는 능력이 강조형 문항들을 푸는데 실제로 영향을 미쳤는지 알아보는 것이 구인타당도와 관련된 가장 중요한 질문 중의 하나이기 때문이다. 타당도 검증 방법에 대한 구체적인 실례가 부족한 현실을 고려할 때 본 연구가 문항의 난이도 예측을 통해서 평가 문항의 구인타당도를 검증하고자 하는 사람들에게 유용한 모델을 제시할 수 있을 것으로 기대된다.

2. 사전연구

여러 가지 변인이 시험의 난이도에 영향을 미치기 때문에 시험의 난이도를 정확하게 예측하는 것은 쉬운 일이 아니다. 그러나 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 증거의 하나이다. 간단히 말해서 시험 결과에 근거해서 내리는 각종 해석이나 결정이 유의미하기 위해서는 평가도구를 개발하거나 사용하고자 하는 사람들은 평가 도구가 측정하고자 하는 능력, 즉 구인과 관련된 요소(element)가 문항 난이도에 실제로 연관되어 있다는 사실을 증명해 보일 수 있어야만 한다.

시험 개발 측면에서도 난이도를 예측해 내는 일이 필수적이라고 볼 수 있는데 (Choi, 1994) 그 이유는 시험의 난이도에 영향을 미치는 요소를 더 잘 이해할수록 주어진 목적에 부합되는 문항을 개발할 수 있는 가능성이 그만큼 더 커지기 때문이다. 안면타당도(face validity) 측면에서도 시험의 난이도 예측은 중요할 수 있는데, 우리나라에서 대학수학능력시험을 개발할 때 난이도 예측이 가장 중요한 고려사항 중의 하나이고 실제 시험의 난이도가 해마다 관심사가 되는 게 그

좋은 예라고 할 수 있다. Jin과 Park (2004)은 시험의 결과가 중요할수록 문항 난이도가 수험자들에게 더욱더 중요한 문제가 된다고 지적하였다.

언어평가 분야에서는 읽기 시험과 듣기 시험의 난이도 예측과 관련된 연구가 많이 이루어져 왔다(Buck & Tatsuoka, 1998; Choi, 1994; Drum, Calfee, & Cook, 1981; Freedle & Kostin, 1994, 1996, 1999; Jin & Park, 2004; Kostin, 2004; Nissan, DeVincenzi, & Tang, 1996). 이 분야의 대표적인 연구라 할 수 있는 읽기 시험의 난이도를 예측해 본 논문에서 Drum, Calfee와 Cook (1981)은 열 가지 읽기 하위 기능이 객관식으로 개발된 읽기 시험 난이도 변산의 94%를 설명했다고 보고하였다. Freedle과 Kostin (1993)은 Test of English as a Foreign Language (TOEFL) 시험의 읽기 영역 문항의 난이도를 예측하는 연구를 진행했는데, 7가지 유형의 문항 특성이 문항 난이도 변산의 58%를 예측해 내었다고 발표하였다. Freedle과 Kostin (1996)이 실시한 TOEFL 듣기 영역의 강의형(mini talk) 문항의 난이도를 예측한 연구에서는 듣기 텍스트와 관련된 14개의 변수가 37문항의 난이도 변산의 35%를 설명하는 것으로 드러났다. 국내에서 이루어진 연구로는 대학원 입학시험의 난이도를 분석한 김충배와 신명신의 연구(2001), 그리고 대학수학능력시험의 문항난이도에 미친 영향을 분석한 Jin과 Park (2004)의 연구가 있다. 김충배와 신명신(2001)은 대학원 입학 영어 시험 문항의 난이도를 예측하고 조정된 다음 실제 시험 결과 분석을 통해 예측된 난이도의 적중여부를 문항별로 살펴보았다. 그 결과 10% 범위 내에서 50개 문항중 27개 문항(54%)의 난이도가 적중된 반면에 나머지 문항의 경우 20% 범위를 넘게 예측이 빗나갔다. 또한 진경애와 박정(2004)은 2003학년도에 실시된 대학입학 수학능력시험 외국어 영역의 문항 난이도와 관련된 변인을 상관관계 분석과 중다회귀분석을 실시하여 알아보았다. 문항의 유형과 관련된 28개의 변인, 지문의 특성과 관련된 12가지 변인, 그리고 지문과 문항 사이의 상호작용과 관련된 6가지 변인 중에서 상관관계 분석 결과 14개의 변인이 문항 난이도와 통계적으로 유의미한 상관관계가 있는 것으로 밝혀졌다. 이 변인들로 중다회귀분석을 실시한 결과 이 14개의 예측변수가 문항난이도 변산의 49%를 설명하는 것으로 드러났다.

3. 연구 방법

3.1. 자 료

본 연구에서는 동형(parallel form)으로 제작된 MELAB A형과 B형에 출제된 20개의 강조형 문항의 난이도를 분석하였다. 본 연구에서 분석한 자료는 1999년 5월에서 2001년 9월 사이에 실시된 실제 시험 결과에서 무작위로 추출되었는데, A형의 경우 1793명, 그리고 B형의 경우 1410명의 시험 결과 자료를 사용하였다.

3.2. 강조형 문항

MELAB 기술 보고서는 강조형 문항의 구인을 화자의 발화 중에서 초분절적인 요소나 운율적인 요소를 통해 전달되는 의미를 파악하는 능력이라고 정의하고 있다(University of Michigan, 1996). MELAB 듣기 영역에는 두 가지 유형의 강조형 문항이 있는데 두 유형의 듣기 자료 모두 수험자가 발화 속의 강세 정보를 이용해야만 정확한 의미를 파악할 수 있도록 의도적으로 애매하게 제작되어 있다. 아래 보기 (1)에 제시된 바와 같이 첫 번째 유형은 특정한 단어가 강조된 문장이나 질문을 듣고 화자가 계속해서 이야기 할 말을 찾아내는 문항이다.

- (1) Tom said he was going to drive to Chicago next week,
 a. not last week.
 b. not next month.
 c. not fly.

두 번째 문항 유형은 아래 보기 2처럼 특정한 단어가 강조된 질문을 듣고 질문에 가장 적합한 답을 고르는 문항이다.

- (2) Do you have john's keys?
 a. No, but Jane does.
 b. No, I have Jim's.
 c. No, only his bags.

3.3. 예측변수(predictor variables)

문항의 난이도를 예측하기 위해서는 난이도에 영향을 미쳤을 것으로 판단되는 변수를 먼저 찾아내어야 한다. 본 연구에서는 먼저 듣기 시험의 난이도를 연구한 선행연구에서 문항 난이도와 관련된 것으로 밝혀진 변수를 분류한 다음 그 중에서 강조형 문항의 난이도에 영향을 미쳤을 가능성이 높은 아래의 세 가지 변수를 선정하였다(Freedle & Kostin, 1994, 1996, 1999; Nissan, Devincenzi, & Tang, 1996). 물론 이 변수들 이외에도 다른 많은 변수가 듣기 문항의 난이도와 관련되어 있다는 점에 주의할 필요가 있다. 예를 들어 화자의 발음 속도도 듣기 문항의 난이도에 영향을 미쳤다고 가정해 볼 수 있다. 그러나 강조형 문항의 경우 동일한 성우가 모든 문항을 같은 속도로 녹음하였기 때문에 본 연구에서는 예측변수로 포함하지 않았다. 본 연구에서 예측변수로 선정한 세 변수의 조작적 정의는 다음과 같다.

변수 1: 듣기자료(input)의 길이- 듣기 자료의 길이가 길면 길수록 수험자가 처리해야 할 정보의 양이 많기 때문에 더 어려운 문항이 될 가능성이 많

다. 따라서 변수 1은 강조형 문항의 난이도가 듣기 자료의 길이와 관련 되었다고 가정하고 있다. 듣기자료의 길이는 Pitchworks 프로그램을 사용해서 초 단위로 측정되었다.

변수 2: 운율적 탁립(prosodic prominence) - 수험자가 강조형 문항을 풀기 위해서는 문장에서 강조된 표현을 찾아야 하므로, 강조된 말이 더 두드러지면 질수록 문항이 더 쉬워질 가능성이 크다. 이 가설을 검증하기 위해서 강조된 단어와 선행 단어의 강세 음절의 강도(intensity)의 비율을 계산하였다. 음향학적인 강도가 귀로 듣는 소리의 크기(loudness)에 해당되므로(Ladefoged, 2001), 각각의 문장을 디지털화한 다음 각 강세 음절의 강도를 Pitchworks 프로그램을 사용하여 측정하였다. 그림 1에 두 음절의 강도의 비율이 어떻게 측정되었는지 예를 보여주고 있다.

변수 3: 내용어(content words)의 수 - 세 번째 변수는 문장 속에 내용어가 많으면 많을수록 처리해야 하는 정보의 양이 늘어나기 때문에 더 어려운 문항이 된다고 가정하고 있다. 듣기자료의 정보량과 난이도 사이의 관련을 알아보기 위해 문장 속에 있는 내용어의 숫자를 계산하였다.

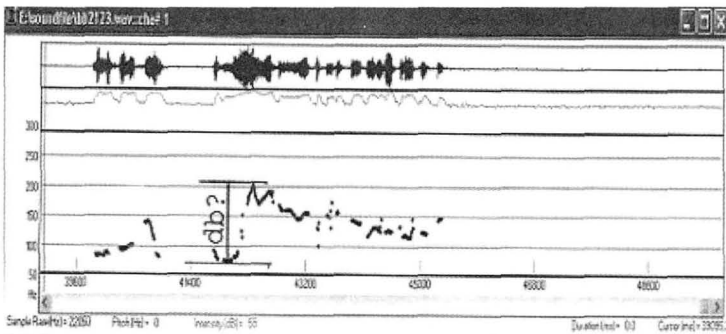


그림 1. 강도의 차이 계산

세 변수의 기술통계량이 표 1에 보고 되어 있다.

표 1. 기술통계량

변수	평균	표준편차	최소값	최대값
변수 1	3.67	.6	2.8	5.2
변수 2	1.10	.05	1	1.22
변수 3	6.15	1.35	4	8

강조형 문항들이 수험자가 초분절적인 요소를 통해 전달되는 발화의 뜻을 파악하는 능력을 측정한다는 구인 정의 측면에서 볼 때 세 예측변수 중에서 변수 2, 운율적 탁림이 제일 중요한 예측변수라고 볼 수 있을 것이다.

3.4. 자료 분석

먼저 문항 난이도는 각 문항을 맞춘 수험자의 숫자를 총 수험자의 수로 나누어서 계산하였다. 앞에서 보고한 세 변수가 문항 난이도를 예측하는 정도를 알아보기 위하여 중다회귀분석(multiple regression analysis)을 실시하였는데, 분석 자료가 회귀분석의 가정을 만족시키는지 알아보기 위하여 종속변수와 예측변수가 정상분포와 선형관계 가정을 만족하는지 알아보았다. 분석 결과 모든 잔차(residual)가 모든 변수 값에서 정규분포를 갖고 있었고 관찰된 잔차와 기대치 잔차의 누적치 분포도 거의 동일하였다. 종속변수와 세 예측변수 사이의 관계도 선형적인 것으로 드러나 회귀분석의 가정을 만족시키는 것으로 밝혀졌다. 회귀분석에서 더 많은 변수를 예측변수로 사용하면 사용할수록 더 높은 결정계수를 구할 수 있겠지만 본 연구에서는 난이도에 영향을 미쳤을 것으로 판단되는 세 변수만 분석에 포함시켰다. 따라서 변수와 관측자료의 비율이 변수 당 최소 10배에서 20배 이상 되어야 한다는 가정도 충족되었다. 자료 분석에는 SPSS 11이 사용되었다.

4. 결 과

4.1. 문항난이도

A형과 B형의 평균 점수는 각각 100점 만점에 75점과 80점으로 두 시험이 수험자 집단에게 비교적 쉬웠다는 사실을 알려준다. 두 시험의 Cronbach alpha 신뢰도 계수는 모두 .80으로 비교적 높았다. 각 문항의 난이도가 표 2에 보고되어 있는데 문항난이도는 .15에서 .95까지 분포되었다. 가장 어려운 문항은 B형 시험의 21번 문항으로 문항 난이도가 .15로 수험자의 15%만이 정답을 제시하였다. 그 반면에 B형의 17번 문항의 경우 문항 난이도가 .96으로 겨우 4%만이 오답을 선택한 제일 쉬운 문항이었다.

표 2. 문항 난이도

문항번호	A형	B형
16	.26	.63
17	.47	.96
18	.35	.44
19	.55	.65
20	.37	.66
21	.69	.15

문항번호	A형	B형
22	.47	.69
23	.44	.85
24	.65	.52
25	.41	.34

4.2. 중다회귀분석

문항 난이도와 세 변수 사이의 관계를 알아보기 위하여 단계적 방법(stepwise method)을 사용하여 중다회귀분석을 실시하였는데 그 결과가 표 3에 보고 되어 있다. 회귀 분석에서 문항 난이도가 종속변수이었고 세 변수가 예측변수이었다. 회귀분석 결과 변수 2, 즉, 운율적 탁립(prosodic prominence)만이 문항 난이도에 통계적으로 유의미한 영향을 미친 것으로 밝혀졌다. 표 3에서 보는 것처럼 변수 2의 회귀계수만 통계적으로 유의미한 수준에서 0과 달랐고 R^2 계수는 .47로, 문항 난이도 변산의 약 47%를 설명하는 것으로 드러났다. 세 변수 중에서 변수 2가 문항 난이도를 성공적으로 예측한 유일한 변수였다는 사실은 강조형 문항의 구인 타당도를 지지하는 결과로 볼 수 있다. 또한 듣기평가의 난이도에 영향을 미칠 수 있는 여러 가지 시험방식양상(test method facet)(이완기, 2003; Bachman, 1990) 중에서 본 연구에서는 듣기 자료의 특성만 고려했음에도 불구하고 문항 난이도 변산의 47%가 설명되었다는 점도 주목할만한 결과로 판단된다.

5. 논의와 결론

본 연구에서는 MELAB 듣기 영역에 나오는 강조형 문항의 구인타당도를 문항 난이도 예측을 통해 알아보았다. 문항 난이도와 관련되었을 것으로 추정되는 세 가지 변수를 선정한 다음, 이 변수들이 문항난이도를 예측하는 정도를 알아보기 위하여 중다회귀분석을 실시하였다. 회귀분석결과 운율적 탁립이 난이도에 영향을 미친 유일한 변수로 밝혀졌고, 문항 난이도 변산의 약 47%를 설명하였다. 이 결과는 강조형 문항이 화자의 발화에서 운율적 정보를 처리할 수 있는 능력을 측정한다는 강조형 문항의 구인 정의를 지지하는 증거라고 볼 수 있다.

표 3. 회귀분석 결과

모델	R	R^2	추정치의 표준오차	
1	.685	.469	.157	
변수	B	표준오차	t	유의확률
상수	-2.42	.811	-2.984	.009
변수 1	.005	.063	.079	.983

변수	B	표준오차	t	유의확률
변수 2	2.697	.737	3.660	.002
변수 3	-.0016	.028	-.058	.955

타당도 검증이란 시험 그 자체가 아니라 시험 결과에 근거해서 우리가 내리는 각종 추론이나 결정의 타당도가 검증 대상이기 때문에 타당도 검증이란 여러 가지 증거를 모아 가는 과정이라고 볼 수 있다(Messick, 1989). 본 연구는 문항의 난이도 예측이 평가도구의 구인 타당도를 검증할 수 있는 유용한 방법이 될 수 있다는 사실을 보여주었다. 앞에서도 지적한 것처럼 만약 문항 난이도에 영향을 미치는 요소들을 밝혀낼 수 있다면 시험의 구인타당도를 검증할 수 있을 뿐만 아니라 구인타당도가 높은 문항을 제작하는 데도 도움이 되기 때문이다(Carroll, 1989; Embretson, 1985). 따라서 언어평가의 난이도를 예측하려는 시도가 앞으로 더 많이 이루어져야 할 것이다.

그러나 문항 난이도 예측은 시험 개발자가 의도한 요소들이 실제 문항 난이도에 영향을 미쳤는가 하는 문제에 대한 답을 제공할 뿐 문항이 측정하고자 하는 구인을 제대로 측정하고 있는가라는 큰 질문에 대한 모든 답을 다 제공하지는 못한다는 점에 유의할 필요가 있다. 실제로 문항 난이도는 문항 그 자체의 특성이라기 보다는 문항 특성과 수험자 특성 사이의 상호작용의 결과라고 보아야 할 것이다(Bachman, 2002). 또한 위에서 보고한 것처럼 본 연구에서는 문항 난이도의 47%를 설명해 내었는데 이 결과는 이 예측변수들이 강조형 문항 난이도 변산의 나머지 53%는 설명하지 못하고 있다는 사실을 의미한다. 따라서 강조형 문항의 구인타당도를 확인하기 위해서는 문항 난이도 예측 이외의 다른 종류의 증거도 모아야 할 것이다. 예를 들어, 수험자가 강조형 문항을 푸는 과정을 살펴서, 시험 개발자들이 이 문항들의 측정한다고 본 구인과 실제 수험자가 사용하는 언어 기능이 일치하는지 알아보는 연구를 진행할 수 있을 것이다.

물론 듣기 능력의 하위 구성요소 하나하나를 독립적으로 평가하는 문항을 만드는 것은 현실적으로도 어렵고 하위 구성요소들이 대개 통합적으로 작용한다는 점에서 평가 문항의 진정성도 낮아질 위험이 크다(이완기, 2003; Buck, 2001). 그러나 서론에서도 밝혔듯이 듣기 하위 기능 중에서 초분절적 요소는 모든 차원에서의 듣기에 관련되기 때문에, 듣기 평가에서도 초분절적 요소를 파악하는 능력을 측정할 필요가 있다. 또한 분절음이 초분절음에 비해 의사소통에서 차지하는 비중이 낮음에도 불구하고(한중임, 2001; Dalton & Seidlhofer, 1994) 아직도 말하기 교육에서 초분절적인 요소보다는 분절적인 요소에 치중하고 있는 현실을 고려할 때 시험의 긍정적인 세환효과(washback effect)도 기대할 수 있을 것이다. 본 연구의 결과 강조형 문항이 수험자의 강조 및 대조 강세를 인식하는 능력을 측정할 수 있는 실용도가 높은 문항이라는 점이 밝혀졌다. 물론 구어 담화의 많은 특성들이 결여되어 있지만(Weigle, 2000) 실용도(practicality)가 높기 때문에 대규모 표준화 시험에서 사용될 수 있을 것으로 판단된다.

참고 문헌

- 김충배, 신명신. (2001). *영어학 영어교육의 제문제*. 서울: 한국문화사.
- 이완기. (2003). *영어과 평가론*. 서울: 문진미디어.
- 전상범. (1995). *영어 음성학 개론*. 서울: 을유 문화사.
- 한종업. (2001). *영어 음성학과 발음지도: 영어 교육학적 접근*. 서울: 한국문화사.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing* 19, 453-476.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Buck, G. & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing* 15, 119-157.
- Carroll, J. B. (1989). Intellectual abilities and aptitudes. In A. Lesgold and R. Glaser, eds., *Foundations of a psychology of education* (pp.137-197). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Choi, I-C. (1994). Content and construct validation of a criterion-referenced English proficiency test. *English Teaching* 48, 11-348.
- Dalton, C. & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.
- Drum, P. A., Calfee, R. C. & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly* 16, 486-514.
- Embretson, S. (1985). *Test design: Developments in psychology and psychometrics*. Orlando: Academic Press.
- Freedle, R. & Kostin, I. (1994). Can multiple-choice reading tests be construct valid? *Psychological Science* 5, 107-110.
- Freedle, R. & Kostin, I. (1996). *The prediction of TOEFL listening comprehension item difficulty for minitalk passages: implications for construct validity*. Princeton, NJ: Educational Testing Service Research Report 56.
- Freedle, R. & Kostin, I. (1999). Does text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing* 16, 2-32.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Jin, K-A. & Park, J. (2004). The prediction of English item difficulty in

- college scholastic ability test. *English Teaching* 59, 267-278.
- Kostin, I. (2004). *Exploring item characteristics that are related to the difficulty of TOEFL dialogue items*. Princeton, NJ: Educational Testing Service Research Report RR 79.
- Ladefoged, P. (2001). *A course in phonetics*. Fort Worth, TX: Harcourt College Publishers.
- McNerney, M. & Mendelsohn, D. (1992). Suprasegmentals in the pronunciation class: Setting priorities. In P. Avery and S. Ehrlich, eds., *Teaching American English pronunciation* (pp.185-196). Oxford: Oxford University Press.
- Messick, S. A. (1989). Validity. In R. L. Linn, ed., *Educational measurement* (pp.13-103). New York: American Council on Education/Macmillan.
- Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.
- Nissan, S., DeVincenzi, F. & Tang, K. L. (1996). *An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension*. ETS Research Report RR-95-37. Princeton, NJ: Educational Testing Service.
- Richards, J. C. (1983). Listening comprehension: approach, design, procedure. *TESOL Quarterly* 17, 219-240.
- Rost, M. (1994). *Introducing listening*. Harmondsworth: Penguin.
- University of Michigan. (1996). *MELAB technical manual*. Ann Arbor, MI: University of Michigan Press.
- Weigle, S. C. (2000). Test review: The Michigan English language assessment battery. *Language Testing* 17, 449-455.

신상근
120-750

서울시 서대문구 대현동 11-1 이화여자대학교 영어교육과
전자우편: sangshin@ewha.ac.kr

접수일자 : 2005. 6. 1

수정본 접수 : 2005. 8. 14

게재결정 : 2005. 9. 8