「언어연구」, 제24권 제1호(2007), 1～23

영어 말하기 평가 도구 IELTS와 MATE 비교분석[1]

이은혜 (동아대)　　김성아 (경희대)

Eun-Hye Lee, Sung-A Kim. 2007. A Comparative Analysis of English Speaking Proficiency Tests: IELTS and MATE. *Linguistic Research* 24(1), 1-23. The purpose of this study is to compare two English speaking proficiency tests: IELTS and MATE. Thirty intermediate college students participated in ten-week 'English Public Speaking Program' and their speaking performances were video-taped twice: before and after the program. Two native speakers of English evaluated the Korean learners' speaking proficiency by using IELTS and MATE. The null hypothesis for the present study was that there would be no significant differences between the evaluation results of the two English speaking proficiency tests. Statistical results showed that a significant positive correlation was obtained in general. Yet, no significant positive correlations were obtained in the areas of 'accuracy' and 'pronunciation,' which may suggests that native speakers of English feel difficulty in scaling non-native speakers' pronunciations.

Keywords English evaluation, IELTS, MATE, non-native speakers' pronunciation

1. 서론

제6차 교육과정이후 영어교육에서 의사소통 능력의 중요성이 강조되어오기는 하였으나 진학과 취업에서 다양한 방식의 영어 면접의 중요성이 더욱 강조되는 등, 요즘처럼 자신의 의사를 확실히 전달할 수 있는 기능적 말하기의 필요성에 대한 사회적 요구가 극대화 된 적은 거의 없었다고 해도 과언이 아닐 것이다. 이러한 사회적 요구에 부응하여 영어 말하기 능력을 지도하거나 평가하는 교과목과 관련 연구들이 증가일로에 있지만, 여전히 대부분의 논의가 추상적인 총론 수준에 머물고 있으며 한국 영어교육 현장에 적용 가능한 말하기 능력을 과학적으로 검증한 측정 도구의 개발은 미진한 것이 사실이다.[2] 실증적인 연구 검증 없이 연구자나 교수자의 경험과 직관

[1] 영어 말하기 도구 중에 국내에서 개발된 MATE와 국외에서 개발된 IELTS를 직접 비교하는 것에는 다소간 무리가 있을 수 있다는 익명의 심사자의 지적에 감사하며 그럼에도 불구하고 두 말하기 평가도구를 선택한 점은 논문의 본문에서 상술하였음을 밝힌다.

에 의존하여 단기성 과제를 위한 등급 기술표를 개발하거나 국외에서 개발된 말하기 능력 측정 등급 기술표에 일방적으로 의존하는 것이 현재 한국 영어교육계의 현실이다(신동일 2001a). 본 연구는 이 평가도구들의 평가자들 간의 신뢰도나 평가의 타당성에 대한 비판적인 관심을 고취하고자 현재 국내 영어 교육계에서 널리 사용되고 있는 말하기 능력 평가 도구들을 비교하고자 한다. 본 연구에서 선택된 영어 말하기 능력 평가 도구로는 국외에서 개발된 IELTS와 국내에서 개발된 MATE가 선정되었다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 현재 국내 영어교육계의 평가 현장에서 사용되어지는 평가도구들에 관한 간단한 소개와 더불어 문제점을 지적한다. 3장에서는 비교 대상으로 선택된 IELTS와 MATE를 소개하고 4장에서는 실험 방법을 기술하며 5장에서는 분석 결과를 논의한다. 결론에서는 본 연구 결과의 시사점과 더불어 한국적 실정에 맞는 말하기 능력 평가도구를 개발하기 위해 어떤 부분이 강조되어야 하는지에 대해 언급해 보기로 한다.

2. 국내외의 말하기 평가도구의 개발 동향: ACTEL-SOPI 중심으로

현재 국내에서 가장 널리 사용되는 영어 말하기 능력 평가도구는 ETS (Educational Testing Service)에서 개발된 TSE(Test of Spoken English)와 TOEIC과 짝을 이루는 Spoken English Proficiency(SEPT)가 있다. 불행히도 이 두 시험은 모두 가격, 행정도 등에 있어서 수요자인 한국 영어 학습자들의 불만을 사고 있는 실정이다. 이외에, ESPT(English Speaking Proficiency Test)라는 이름으로 국내에서 개발되어 시행되고 있는 평가도구도 있다. 이것은 인터넷 기반 멀티미디어 시스템을 통해 진행되는 영어 구사능력 전문 평가시험으로 영어를 모국어로 사용하지 않는 사람들을 대상으로 하여 컴퓨터 동영상을 이용해 인터뷰 형식으로 영어 회화 능력을 측정하는 방식을 취한다. 또한 숙명여자대학교, 강남대학교, 서울대학교 등에서 영어 말하기 능력 시험을 개발하는 데 관심을 기울이고 있으며 TEPS 시험의 일부로 간주될 수 있는 TOP(Test of Oral Proficiency)와 숙명여자대학교에서 개발된 MATE(Multimedia Assisted Test of English)등

---

2 최근에 들어 몇몇 연구자들에 의해 문항반응이론 모형 중 하나인 Rasch 측정모형(Rasch measurement model)을 적용한 영어과 평가도구 개발 및 연구가 몇 차례 시도되기는 했지만(신동일, 2001b, 2001c, 2002; 최인철, 2000), 이외로는 이영식(2000), 정행(2000), 정택진(2002) 등이 있지만 실질적인 자료에 근거한 영어 말하기 쓰기 평가 연구가 활발하지 못하다.

이 있다. 간략히 말해, 국내의 영어 말하기 평가도구들은 개발을 기대하는 수요자들의 요구에 호응하기 위해서 다양한 기관에서 개발 중이거나 개발을 시도하고 있으나 국내 영어 학습자들의 신뢰를 받는 주도적인 평가도구는 부재한 형편이다.

국외에서는 미국과 영국이 주도적인 역할을 담당하고 있는데, 미국의 대표적인 평가 도구로는 ETS가 평화봉사단의 업무 수행을 돕기 위해 개발한 OPI(Oral Proficiency Interview)와 그것의 변형인 SOPI(Simulated Oral Proficiency Interview)가 있다. 또한 샌디에고 주립대 부설 ITSC (International Testing Service Center)가 주관하여 개발한 G-TELP가 있다. 영국에서는 캠브리지 대학교 교육평가원(UCLES)에서 개발한 IELTS, KET, PET, CAE, CPE, CYLET, BEC, BULAT와 런던 상공회의소시험국에서 개발하여 시행하는 EFB, PBE, SEFIC, WEFT 등이 있다.

이들 중에서 ACTFL-SOPI에 대한 비교적 자세한 언급을 할 것인데 이것은 본 연구에서 선정된 비교 대상 중의 하나인 MATE가 ACTFL-SOPI를 기반으로 개발된 평가도구이기 때문이다. ACTFL(American Council on the Teaching of Foreign Languages)은 구조주의와 청화식 언어교육이 주류를 이루던 시기에 미국 정부의 지원 하에 개발된 미국대학에서 외국어 능력의 등급화에 사용된 지침서이다. ACTFL에서 인정한 공인시험관에 의해 15분에서 30분간 1대1의 인터뷰가 실시되며 10등급으로 나뉘어져 점수를 받게 된다. OPI가 직접 인터뷰 방식이라면 SOPI 녹음 테이프를 매체로 학습자의 언어자료를 추출하기 위해 상용되는 말하기 능력 평가 방법을 지칭한다. ACTFL은 평가의 준거를 문어체 언어가 아닌 구어체에 둔다는 점에서 언어능숙도(language proficiency)평가에 좋다는 평가와 더불어 비전문가들도 이해하기 쉬운 평가 체계를 제공한다는 장점 때문에 오늘날까지도 미국 내에서 널리 사용되어지고 있다. 6차 교육과정이후 평가의 중요성이 부각되면서 한국의 영어교육전문가들의 눈길을 가장 먼저 잡아 끈 평가도구가 ACTEL이라고 할 수 있다. 그러나 1970년대 후반부터 미국 내에서는 ACTEL이 지나치게 언어의 사용 환경 그리고 언어 사용 능력을 단순화시켰고 그 결과 그 타당성과 신뢰성에 의심이 간다는 비판을 받아왔다(Shohamy, 1994). 평가의 역류 효과(backwash effect)가 상대적으로 더 큰 한국의 영어교육 상황에서, ACTEL에 근거한 영어 능숙도 평가는 중등학교에서의 수행 평가나 여러 기관의 직원들의 선발과 업무능력 진단에 사용되고 있기 때문에 한국적 상황에 맞추어 ACTEL의 내용과 구성을 조정하거나 ACTEL의 등급을 수정할 필요가 있음 또한 지적되어 왔다(김덕기 외 3인 1999).

한편 SOPI의 문제점으로는 오랫동안 평가자들을 훈련시켜야 한다는 점

을 들 수 있다. SOPI는 테이프에 녹음된 학습자의 말하기 샘플을 들으면서 학습자의 말하기 능력을 주어진 채점 기술표에 기준해 평가하는 방식이므로 직접 면접 방식인 OPI보다는 객관성을 확보할 수 있지만 영어 평가 영역의 인력이 충분하지 못한 한국의 상황에서, SOPI 방식을 바로 채용하는 것은 먼저 평가자들의 인력풀을 확보하고 이들은 전문적으로 훈련시켜야 하는 과정이 필수적으로 전제된다.

3. IELTS와 MATE

3.1 IELTS (International English Language Testing System)

IELTS는 IDP IELTS Australia와 캠브리지 대학 (University of Cambridge Local Examination Syndicate), 영국 문화원 (The British Council)이 공동 개발, 관리하는 영어 능력 평가 시험이다. IELTS는 영어권 국가로의 유학이나 이민, 취업을 희망하는 응시자의 영어 구사 능력을 평가하는 시험으로 영국, 호주 및 뉴질랜드의 학교와 호주, 뉴질랜드 이민성과 주한 캐나다 대사관에서 인정되고 있다. IELTS는 전 세계 105개국, 224개의 IELTS시험 센터(IDP와 영국 문화원)에서 시행되고 있으며 시험유형은 크게 학업용 영어능력 시험 (Academic Module)과 일반인 영어능력시험(General Training Module) 두 종류로 나뉜다.

학업용 영어능력 시험은 대학 정규 과정, 대학원 입학 등 학생들의 유학을 목적으로 널리 시행되고 있다. 학교 입학 전 영어의 4가지 영역 중 취약한 영역을 정확하고 쉽게 파악하고자 한다. 일반인 영어능력 시험은 호주와 뉴질랜드로 이민 희망자들의 영어능력을 평가하기 위한 목적으로 시행된다. 일반인 영어능력 시험은 전반적인 사회교육에 필요한 기본적인 능력에 중점을 두고 있다. 두 시험 모두 네 가지 영역 Listening(4영역, 40문항, 30분), General Reading/Academic Reading(3영역, 40문항, 60분), General Writing/Academic Writing(2문항-150& 250단어, 60분), Speaking (10-15). IELTS 시험 등급은 만점이 9점이며 각 테스트 영역별 점수와 평균 점수가 기록된다. 대학들은 대체적으로 평점 6.0~6.5를 입학조건으로 하고 있다. 이 시험에서 성적은 각 영역별로 1점에서 9점까지 주어지게 되며 전체 평점으로도 표시된다. 본 논문에서 다루고자 하는 것은 말하기 영역 평가이므로 IELTS의 영어 말하기 평가 영역 기준을 보다 자세히 소개하고자 한다.

① 말하기 평가 영역

말하기 영역은 수험자들이 영어 말하기 능력을 측정하도록 고안된 구술 면접시험으로 전공 영역에 관련한 내용 보다는 일반적인 말하기 능력 측정에 중점을 둔다. 말하기 능력에는 다음 네 가지 하위 영역별로 평가자가 1점에서 9점까지 점수를 부여한다. 점수 부여의 기준은 <표 1>에 나타나 있다.

a. Fluency and Coherence

유연하고 효과적으로 말을 할 수 있는 자신감과 상황의 변화 등에 효과적으로 대응하여 전체적인 의사소통이 자연스러운 정도를 측정한다.

b. Lexical Resource

적절한 어휘를 선택하고, 반복을 피하여 다양하게 어휘를 사용하는 능력을 측정한다.

c. Grammatical Range and Accuracy

문법적 범위와 정확성은 문법 사용의 오류가 최소화 될 수 있도록 하는 능력을 측정한다.

d. Pronunciation

개별적인 어휘의 발음과 말 전체의 억양이나 강약 등을 정확하고 효과적으로 낼 수 있는 능력을 측정한다.

② 말하기 영역 평가 구성

a. Part 1-Introduction and interview (4-5 minutes)

An examiner introduces him/herself and confirms a candidate's identity. The examiner interviews the candidates using verbal questions based on familiar topic frames.

The candidate answers questions about him/herself, his/her jobs/studies, his/her interest, and a range of similar familiar topic areas(ex. Tell me about the city you live in. or What do you like doing with your friends at weekend?)

b. Part 2-Individual long turn(3-4 minutes, 1 minute preparation time)

An examiner asks the candidate to speak for 1-2 minutes on a particular topic based on written input in the form of a general instruction and content-based prompts. The examiner asks one or two questions at the end of the long turn. The candidate is given a verbal prompt on a card and is asked to talk on a particular topic. The candidate has one

minute to prepare before speaking at length, for between one and two minutes. The examiner then asks one or two rounding-off questions (ex. Do you prefer celebrating family occasions at home or in a cafe? Why?).

c. Part3-Two way discussion(4-5 minutes)
An examiner invites the candidate to participate in discussion of more abstract nature, based on verbal questions thematically linked to Part 2 prompt. The examiner and the candidate engage in a discussion of more abstract issues and concepts which are thematically linked to the topic prompt in Part2 (ex. What kind of children activities are there available in your city?).

③ 말하기 능력 기준표

<표 1> IELTS Speaking Band Descriptors(Public version)

Band	Fluency and Coherence	Lexical Resource	Grammatical Range and Accuracy	Pronunciation
9	-speaks fluently with only rare repetition or self- correction; any hesitation is content-related rather than to find words or grammar -speaks coherently with fully appropriate cohesive features -develops topics fully and appropriately	-uses vocabulary with full flexibility and precision in all topics -uses idiomatic language naturally and accurately	-uses a full range of structures naturally and appropriately -produces consistently accurate structures apart from 'slips' characteristic of native speaker speech	
8	-speaks fluently with only occasional repetition or self-correction ; hesitation is usually content-related and only rarely to search for language -develops topic coherently and appropriately	-uses a wide vocabulary resource readily and flexibly to convey precise meaning -uses less common and idiomatic vocabulary skillfully with occasional inaccuracies -uses paraphrase effectively as required	-uses a wide range of structures flexibly -produces a majority of error-free sentences with only very occasional inappropriacies or basic/non-systematic errors	easy to understand throughout, with L! accent having minimal effect on intelligibility -uses a wide range of phonological features to convey meaning effectively
7	-speaks at length without noticeable effort or loss of coherence -uses a range of connectives and discourse markers with some flexibility -may demonstrate language-related hesitation at times, or some repetition and/or self-correction	-uses vocabulary resource flexibly to discuss a variety of topics -uses some less common and idiomatic vocabulary and shows some awareness of style and collocation with some inappropriate choices -uses paraphrase effectively	-uses a range of complex structures with some flexibility -frequently produces error-free sentences, though some grammatical mistakes persist	

6	-is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation -uses a range of connectives and discourses markers correction or hesitation	-has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inappropriacies -generally paraphrases successfully	-uses a mix of simple and complex structures, but with limited flexibility -may make frequent mistakes with complex structures, though these rarely cause comprehension problems.	-can be understood throughout, though mispronunciation may occasionally cause momentary strain for the listener
5	-usually maintains flow of speech but uses repetition, self-correction and/or slow speech to keep going -may over-use certain connectives and discourse markers -produces simple speech fluently, but more complex communication causes fluency problems	-manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility -attempts to use paraphrase but with mixed success	-produces basic sentence forms with reasonable accuracy -uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems	
4	-cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction -links basic sentences but with repetitious use of simple connectives and some breakdowns in coherence	-is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice -rarely attempts paraphrase	-produces basic sentence forms and some correct simple sentences but subordinate structures are rare -errors are frequent and may lead to misunderstanding	-produces some acceptable features of English pronunciation but overall control is limited and there can be severe strain fo the listener
3	-speaks with long pauses -has limited ability to link simple sentences -gives only simple responses and is frequently unable to convey basic message	-uses simple vocabulary to convey personal information -has insufficient vocabulary for less familiar topics	-attempts basic sentence forms but with limited success, or relies on apparently memorized utterances -makes numerous errors except in memorized expressions	
2	-pauses lengthily before most words -little communication possible	-only produces isolated words or memorized utterances	-cannot produce basie sentence forms	-speech is often unintelligible
1	-no communication possible -no rateable language			
0	-does not attend			

<표 2> IELTS 말하기 능력 등급 판정표

band	Fluency and Coherence, Lexical Resource, Grammatical, Range Pronunciation
1	no communication possible
2	pause lengthily before most words
3	speaks with long pause
4	links basic sentences but with repetitious use of simple connectives and some breakdowns in coherence

5	usually maintains flow of speech but uses repetition, self-correction and/or slow speech to keep going
6	uses a range of connectives and discourse markers but noto always appropriately
7	speaks at length without noticeable effort or loss of coherence
8	speaks fluently with only occasional repetition or self-correction
9	develops topic fully and appropriately

3.2 MATE(Multimedia Assisted Test of English)

MATE는 숙명여자대학교가 개발한 멀티미디어 방식의 영어 말하기 및 쓰기 수행능력 평가도구이다. MATE는 수험자가 주어진 컴퓨터상에서 소리, 영상, 문자로 제시되는 과제를 헤드셋(말하기의 경우)과 컴퓨터 자판(쓰기의 경우)을 사용하여 답안을 입력하는 '멀티미디어 방식'의 영어 말하기/쓰기 평가이다. MATE Speaking은 총 3개영역, 8개의 과제로 구성되어 있으며, 소요시간은 약 25분이다. 컴퓨터를 통해 시험이 진행되며, 모니터를 통해 그림과 지문이 보여 지고, 정해진 응답 시간이 지나면 자동으로 다음 과제로 넘어간다. MATE의 말하기 시험의 구성을 보다 자세히 소개하면 다음과 같다.

a. 제1영역: 준비과제(warm-up)

수험자가 편안히 평가에 임할 수 있도록 고안된 것으로 신상에 관한 질문에 답하기, 일상적 상황에 간단히 응답하기 등 짧은 응답을 요구하는 과제로 구성되어 있다.

b. 제2영역: 그림중심과제(Picture-Based Tasks)

아이디어를 하나의 그림으로 보여주거나 이야기 거리를 여러 개의 연속적인 그림으로 제시하여 장소/활동/인물묘사하기, 길 안내하기, 서술하기 등 중급에서 고급에 이르는 과제를 수행하도록 구성되어 있다.

c. 제3영역: 주제 및 상황 중심과제(Topic-Based or Situation-Based Tasks)

어떤 주제에 관해 설명하거나 논쟁할 수 있도록 일상적 활동 묘사하기, 장점과 단점 설명하기, 의견 피력하기, 가상하기 등의 과제로 구성되어 있다. 내용은 개인적인 것에서 사회적 이슈에 이르기까지 매우 다양하지만 특별한 전문 지식이 없어도 수험자 모두가 다룰 수 있는 일반적 주제가 주어진다.

MATE Speaking의 능력 기준표에 근거하여 아래 4가지 영역을 분석하여 수험자의 말하기 능력을 평가한다. General Rating(Test A)는 Rudimentary (초급)에서 Expert(최상급)까지 12등급을 평가한다. Special Rating(Test B)는 Rudimentary(초급)에서 Moderate(중급 4단계)까지 6등급을 평가한다. 말하기 영역의 평가의 구체적인 내용은 다음과 같다.

① 말하기 평가 영역

a. 과제 수행 및 기능(Functions or Global Tasks)

수험자의 유창성, 어휘, 문법의 개별적 영역의 평가가 아니라 실생활 속에서 언어를 사용하여 수행할 수 있는 일을 전반적으로 살펴봄으로써 실제 언어 사용 능력을 평가한다.

b. 적절한 맥락과 내용(Context and Content)

수험자가 언어를 사용하고 있는 상황에 얼마나 잘 대처하는가를 평가한다. 보통 과제 수행 능력이 높을수록 신상 정보나 개인적 경험 같은, 자기 주변에서 일어나는 친숙하고 일상적인 상황 대처보다 자기 자신에서 벗어난 다양하고 복잡한 주제를 다루며 공식적 상황에도 대처할 수 있는 능력을 보여준다.

c. 정확성(Accuracy)

수험자가 상대방에게 자신의 메시지를 얼마나 잘 전달하고 이해시키느냐를 평가한다. 화자의 발음과 유창성, 어휘와 문법, 사회적 실용적 능력이 중요한 역할을 한다. 구체적으로 '발음(pronunciation)'은 강세 및 억양. '유창성(Fluency)'은 말의 속도와 문장 연결 능력. '어휘(Vocabulary)'는 화자가 사용 가능한 단어의 범위와 적합성. '문법(Grammar)'은 단어 형태에 대한 지식과 문장 구성 능력, '사회적 실용적 능력(Socio-pragmatic competence)'은 영어 사용 문화권에 대한 지식을 말한다.

d. 구문 형태(Text Type)

구문의 양과 구성적 측면을 평가한다. 초급 수준에서는 '단어'나 '구문', 중급에서는 '문장', 고급에서는 '단락', 최상급의 수준에서는 여러 단락의 논리적 '구성력'을 평가한다. 이러한 구문 형태는 수험자가 다룰 수 있는 과제의 종류와 밀접한 관계가 있다. 몇 가지 활동을 소개하거나 간단한 묘사를 하는 과제에서는 문장의 나열 형태로, 지난 일을 서술할 때에는 문장을 연결한 단락의 형태로, 추상적인 사회문제에 대한 의견은 논리적으로 연결된 여러 단락으로 표현할 수 있다[3].

③ 말하기 능력 기준표

MATE의 말하기 능력 기준표는 세부적인 영역이 나누어져 있지 않고 각 단계별로 기준이 기술되어 있다. 본 연구에서는 IELTS와 비교 분석을 위해 제시되어 있는 기준표를 가지고 그대로 네 가지 영역(유창성, 어휘, 정확성, 발음)으로 나누었다.

<표 3> MATE Speaking Band Descriptors (Public version)

Band	Fluency and Coherence	Lexical Resource	Grammatical Range and Accuracy	Pronunciation
12	communicate fully and effectively in the language with ease, and accuracy. They are further able to speak concretely as well as abstractly. They explain complex matters in detail, using extended discourse coherently without unnatural or lengthy hesitation to make their point. They explain their opinions, such as social and political issues, and provide structured argument to support their opinions.		communicate in the language with fluency	
11	basically function at the Expert level, but are not able to fully sustain performance in that level. They are able to consistently explain in detail and narrate fully and accurately in all time frames. They can provide structured argument of strings of paragraphs to support their opinions. They may even construct hypotheses. They can discuss some topics abstractly; however, they perform better when discussing topics concretely.			
10	-participate actively in most informal and some, but limited, formal exchanges on a variety of concrete topics. -they provide a complete account. Their paragraphs are quite lengthy and are filled with good links and details. -They speak with precision and clarity and, thus, are clearly understood although they might make a few errors.		Commanding High speakers are able to fully control their speech while narrating and describing in the past, present, and future tense.	
9	-are able to fully control their speech -The links provided to make their utterances cohesive, logical, will tend to be fairly simplistic. They will have more and simpler patterns of errors.		-narrating and describing in the past, present, and future tense -cohesive linguistic	-sound in structure will not be as tight as those of commanding high speakers
8	participate actively in most informal and a limited number of formal conversations. They demonstrate the ability to narrate and describe in all major time frames in paragraph length discourse, but control may be lacking at times. Their language may be marked by substantial, albeit irregular, flow. It is typically somewhat strained and tentative		-with noticeable self-correction and a certain grammatical roughness.	
7	narrate and describe in major time frames using connected discourse of paragraph length. However, their performance of these Commanding level tasks will exhibit one or more	appropriateness of vocabulary,	the misuse of cohesive devices,	

3 단 본 연구에서는 실험집단, 통제 집단의 평가지도가 개인의 스피치이므로 MATE 평가 영역에서 a, b 측정이 불가능하여 c, d 평가 영역에 중점을 둔 제한점이 있다.

	features of breakdown, such as the failure to maintain the narration or description semantically or syntactically in the appropriate major time frame, the disintegration of connected discourse, a reduction in breadth and or a significant amount of hesitation.			
6	tend to function reactively by responding to direct questions or requests for information. They are capable of asking a variety of different questions to obtain information to satisfy basic needs. They generally make more complex sentences They also are able to link their sentences and usually do so. Their vocabulary shows some peaks, but usually cannot be described as good or complete.	tacking phrases and other clauses onto the main clause.		
5	not as long or complex as those of Moderate High speakers, tend to be fairly complex and somewhat linked. Their speech might still be filled with hesitancies, In general, they are trying to make more complex and linked sentences, but they are not able to control this well and might be forced to fall back, from time to time, to simple, distinct sentences which are not linked logically.		false starts, and inaccuracies. not linked, either linguistically or logically.	
4	is restricted to concrete social exchanges. They can only deal with predictable topics necessary for survival in the target language culture. They are primarily reactive and struggle to answer direct questions or requests for information. In addition, they are also able to ask a limited number of appropriate questions.			sound shaky and unsure of themselves while they are doing so.
3	basically able to handle a variety of tasks pertaining to the Moderate level, but are unable to sustain performance at that level. They are able to successfully manage a number of uncomplicated communicative tasks in straightforward social situations. They are able to ask only a very few formulaic questions when asked to do so.	The test-taker's native languages may strongly influence their vocabulary	The test-taker's native languages may strongly influence their syntax	The test-taker's native languages may strongly influence their pronunciation,
2	When responding to direct questions, they may utter only two or three words at a time or an occasional stock answer. attempt to recycle things they have learned. They often show hesitations,	communicate minimally and with difficulty by using a number of isolated words and memorized phrases. They pause frequently as they search for simple vocabulary or lack of vocabulary,	inaccuracies, or failure to respond appropriately.	
1	have no real functional ability Given adequate time and familiar cues, they may be able to exchange greetings, give their identity, and name a number of familiar objects from their immediate environment	Their vocabulary is severely limited and may be restricted to a few dozen words and phrases.		because of their pronunciation, they may be unintelligible.

<표 4> MATE 말하기 능력 등급 판정표

Expert	Expert	Speakers at the Expert level are able to communicate fully and effectively in the language with ease, fluency, and accuracy. They are further able to speak concretely as well as abstractly. They explain complex matters in detail, using extended discourse coherently without unnatural or lengthy hesitation to make their point.
	Emerging	basically function at the Expert level, but are not able to fully sustain performance in that level. They are able to consistently explain in detail and narrate fully and accurately in all time frames.
Commanding	High	participate actively in most informal and some, but limited, formal exchanges on a variety of concrete topics. commanding High speakers in the past, present, and future tense.
	Mid	are able to fully control their speech while narrating and describing in the past, present, and future tense. The paragraphs of commanding Mid speakers, while sound in structure will not be as tight as those of Commanding High speakers. The links provided to make their utterances cohesive, both linguistic and logical, will tend to be fairly simplistic.
	Low	participate actively in most informal and a limited number of formal conversations. They demonstrate the ability to narrate and describe in all major time frames in paragraph length discourse, but control may be lacking at times. Their language may be marked by substantial, albeit irregular, flow.
	Emerging	narrate and describe in major time frames using connected discourse of paragraph length. However, their performance of these commanding level tasks will exhibit one or more features of breakdown, such as the failure to maintain the narration or description semantically or syntactically in the appropriate major time frame, the disintegration of connected discourse, the misuse of cohesive devises, a reduction in breadth and appropriateness of vocabulary, or a significant amount of hesitation.
Moderate	High	tend to function reactively by responding to direct questions or requests for information. They are capable of asking a variety of different questions to obtain information to satisfy basic needs. They generally make more complex sentences by tacking phrases and other clauses onto the main clause. They also are able to link their sentences and usually do so.
	Mid	The sentences are not as long or complex as those Moderate High speakers, tend to be fairly complex and somewhat linked. Their speech might still be filled with hesitancies, false starts, and inaccuracies. In general, they are trying to make more complex and linked sentences, but they are not able to control this well and might be forced to fall back, from time to time, to simple, distinct sentences which are not linked, either linguistically or logically.
	Low	is restricted to concrete social exchanges. They can only deal with predictable topics necessary for survival in the target language culture. They are primarily reactive and struggle to answer direct questions or requests for information. In addition, they are also able to ask a limited number of appropriate questions. They often sound shaky and unsure of themselves while they are doing so.
	Emerging	basically able to handle a variety of tasks pertaining to the Moderate level, but are unable to sustain performance at that level. They are able to successfully manage a number of uncomplicated communicative tasks in straightforward social situations. They are able to ask only a very few formulaic questions when asked to do so.
Rudimentary	High	are able to communicate minimally and with difficulty by using a number of isolated words and memorized phrases. When responding to direct questions, they may utter only two or three words at a time or an occasional stock answer. They pause frequently as they search for simple vocabulary or attempt to recycle things they have learned.
	Mid	have no real functional ability and, because of their pronunciation, they may be unintelligible. Given adequate time and familiar cues, they may be able to exchange greetings, give their identity, and name a number of familiar objects from their immediate environment.

4. 연구 방법 및 절차

기초 실태 조사에서 영어 말하기 불안감이 높은 대학생 성인 학습자 30명 중15명을 실험집단으로 구분하여 10차시로 구성된 English Public Speaking Program에 참여하도록 하였다. 본 프로그램은 미국 Huston대학, Bloomsburg대학을 비롯한 국외 대학에서 Public speaking course에 사용되는 syllabus를 참조하여 본 연구자가 실험대상 집단의 실정에 맞게 채용한 것이다. 프로그램의 내용은 <표 5>에 나타 나 있다.

<표 5> 실험에 사용된 English Public Speaking Program

횟수	교육과목	교육목표	시간/분	실행일자
1	warm up	집단원간의 이해 및 신뢰감을 발달시킨다.	60	11.28
2	basic grammar (vocabulary)	영어로 말하기를 시도한다. 자신이 알고 있는 어휘, 문법을 말하기 할 때 사용할 수 있도록 한다.	60	12.5
3	picture description	unplanned speaking과 structured speaking의 차이를 알아본다. 준비된 스피치의 중요성을 안다.	60	12.12
4	discussion, debating speech	자신의 의견을 말한다. issue가 되는 쟁점들을 제시하여 영어로 토론하고, 논쟁하는 기회를 가진다.	60	12.14
5	informative persuasive speech	사실적 정보를 제공하고 자신의 주장과 강조점을 가지고 정보 전달 스피치, 설득조의 스피치를 시도한다.	60	12.19
6	Topic choosing brainstorming	Framing에 대한 이론을 알아본다. 자신이 선정한 주제에 대해 brainstorming을 하고 out-line을 설정한다.	60	12.21
7	Script writing	짜임새 있는 원고작성을 한다. 스피치 할 때 정확한 발음으로 구사하기 위한 연습을 한다.	60	12.26
8	Monitoring other speakers	다른 화자들의 사례를 보고 분석하며 자신의 스피치를 수행할 때 고려할 점을 살펴본다.	60	1.5
9	speech contest	자신이 준비한 원고를 가지고 숙지했던 여러 가지 주의사항들을 지키며 모의 스피치 대회를 가진다.	60	1.12
10	Evaluation	자신의 스피치가 녹화된 테이프를 보면서 자기평가, 타인 평가를 한다.	60	1.20

5. 분석 결과

5.1 사전 · 사후 영어 말하기 능력의 평가 결과

말하기 주제를 임의로 주어진 후, 실험에 참가한 학습자들의 개별 말하기 장면을 5분간 비디오 녹화를 하였다. 비디오 녹화는 실험에 참가하기 전과 실험에 참가한 후 두 번에 걸쳐 시행되었다. 영어 교육 분야에서 훈련을 받은 바 있는 평가자들은 실험 집단 8명과 통제 집단 7명의 비디오 녹화물을 각각 IELTS와 MATE 평가도구를 사용하여 평가하였다. 평가 도구간의 혼돈됨이나 짧은 간격으로 인한 평가자내 신뢰도를 떨어뜨리지 않게 하기 위해서 일주일의 간격을 두고 실시하였다. 결과 분석에는 SPSS 통계 프로그램(Version 12.0)이 사용되었다. 먼저 IELTS를 사용한 사전 · 사후 평가 결과가 <표 6>에 요약되어 있고 MATE의 결과는 <표 7>에 나타나 있다.

<표 6> IELTS의 사전 · 사후 결과 분석

	실험	학생 수(N)	평균(M)	표준편차	평균 차이	표준편차 간 차이	통계량 (U)	유의 확률(p)
실험 집단	사전	8	3.75	0.94	-0.96	0.26	-2.502	0.012(*)
	사후	8	4.71	1.20				
통제 집단	사전	7	6.48	2.62	-0.09	0.07	-0.317	0.751
	사후	7	6.57	2.55				

(*P 〈0.05, **P 〈0.01, ***P 〈0.001)

위 표는 비 모수 검정(윌콕슨의 순위 합 검정)의 결과로 편의상 평균과 표준편차를 제공하고 통계량 값은 원 자료의 순위를 매겨서 평균 순위로 비교하여 검정통계량(U)을 얻어, 그에 따른 유의확률 값을 제시한 것이다. <표 6>을 보면, 10주차 시 English Public Speaking Program을 제시했던 실험집단의 사전 · 사후의 대응표본 검정의 유의 확률 값이 0.012로 P 〈0.05의 유의수준에서 유의한 차이를 나타냈다. 반면에 통제 집단의 결과는 유의하지 않은 것으로 나타났다. 다시 말해 10 주 동안의 English Public Speaking Program을 거친 실험 집단의 영어 말하기 능력이 향상되었다는

결과를 얻은 것이다.

<표 7> MATE의 사전·사후 결과 분석

	실험	학생수(N)	평균(M)	표준편차	평균 차	표준편차 간 차이	통계량(U)	유의확률(p)
실험집단	사전	8	3.06	1.39	0.37	0.04	-0.893	0.372
	사후	8	2.69	1.25				
통제집단	사전	7	5.50	3.18	-0.21	0.31	-0.046	0.963
	사후	7	5.71	2.87				

(*P 〈0.05, **P 〈0.01, ***P 〈0.001)

MATE를 사용한 사전·사후의 영어 말하기 능력을 평가한 결과는 <표 7>에 요약되어 있는 데, 이 경우에는 실험집단의 유의 확률 값이 0.372으로 P 〈0.05에서 유의미한 차이를 얻지 못했다. 즉, MATE 평가도구로 평가했을 경우, English Public Speaking Program을 거친 실험 집단이나 English Public Speaking Program에 참가하지 않은 통제집단이나 영어 말하기 능력에 변화가 없다고 판정되었다는 것이다.

이 시점에서 왜 이 두 평가 도구를 사용하였을 때 일치된 결과가 나오지 않았는지를 질문할 필요가 있다. MATE를 근거로 한 평가 결과가 나타내듯이 비교적 단기간에 시행된 English Public Speaking Program이 비교적 단기간에 시행된 실험이었기에 유의미한 결과를 얻지 못했을 가능성이 있다. 그러나 이 가능성을 확신할 수 없는 이유는 <표 6>에 나타난 IELTS의 결과에서는 왜 English Public Speaking Program 후에 실험 집단의 말하기 능력이 향상되었다는 결과를 얻었는가 하는 의문 때문이다. 이러한 질문에 대한 자연스런 귀결은 두 평가 도구 IELTS와 MATE에 말하기 능력 평가 항목에 있어 평가 신뢰도라든지 평가자들 간의 신뢰도에 문제가 발생했을 가능성을 또한 고려해볼 수 있다. 따라서 우리는 다음 절에서 각각의 평가자들의 IELTS 와 MATE의 평가 결과 간의 상관계수를 알아봄으로써 두 평가 도구간의 상이한 결과가 나온 이유를 고찰해 보고자 한다.

5.2 IELTS 와 MATE평가 간 비교분석

이상적인 평가 도구의 조건에는 타당도(validity), 신뢰도(reliability), 실용

도(practicality)등등이 포함된다. 언급된 네 가지 기준 중에서 본 논문에서는 IELTS와 MATE의 신뢰도를 비교해 보고자 한다.

평가의 신뢰도는 평가 결과의 안정성을 의미하는 것으로 동일 대상이 동일한 조건으로 시험을 반복하였을 때, 어느 정도 동일한 결과를 얻을 수 있는가의 문제를 살펴보기 위한 항목이다. 측정의 결과로 나온 성적이 다음에 동일한 조건과 동일한 대상에 다시 실시하여도 그 측정결과가 동일하게 나올 수 있다면 해당 평가는 신뢰도가 높다고 볼 수 있는 것이다. 평가 도구의 신뢰도를 검증하는 방법에는 크게 두 가지가 있는데, 그 하나는 평가 신뢰도(test reliability)이고, 또 다른 하나는 평가자 신뢰도(scorer reliability)이다. 평가 신뢰도는 재검사 신뢰도(test-retest reliability), 동형검사 신뢰도(parallel form), 내적 일관성 신뢰도(internal consistency reliability)로 나누어지며, 평가자 신뢰도는 다시 평가자내 신뢰도와 평가자간 신뢰도로 나누어진다.

5.2.1 평가자간 신뢰도

평가자가 신뢰도는 동일한 검사 결과에 대해 상이한 평가자들 간의 결과가 어떤 상관관계를 보이는가를 살펴봄으로써 해당 검사에 대한 평가자들의 결과가 일관성이 있는지 알아보는 방법이다. 지필고사의 경우와 달리 수행 평가에서 상대적으로 중요성을 가지는 것이 바로 평가자간 신뢰도이다. 두 평가자 결과의 상관관계 분석을 통해 얻은 동등성계수(coefficient of equivalence)가 높을수록 평가자간 신뢰도가 높다고 판정된다. 본 연구에서는 각각의 평가도구들에 대해 평가자들의 반응에 대해 상관 분석(correlational analysis)을 실시하였고 그 결과는 아래 <표 8>과 <표 9>에 요약되어 있다.

<표 8> IELTS 평가 결과에 대해 평가자들 간의 상관관계

평가방식 & 영역		평가자	평균	표준편차	평균차	표준편차차	t값	P값	상관계수
총괄적		A	8.0833	2.3620	1.7916	0.7443	1.770	0.098	.770*
		B	6.2917	1.6177					
분석	유창성	A	8.3333	2.7371	2.3333	1.1435	2.084	.061	.757*
		B	6.0000	1.5936					
	어휘	A	7.3333	2.7602	1.0000	0.6521	.814	.429	.742*
		B	6.3333	2.1081					

적	정확성	A	7.6667	2.1081	1.3334	0.3912	1.390	.186	.599
		B	6.3333	1.7089					
	발음	A	9.0000	2.1081	2.5000	-0.0813	2.326	.036	.426
		B	6.5000	2.1894					

(*P 〈0.05, **P 〈0.01, ***P 〈0.001)

<표 8>은 IELTS 평가 결과에 대해 평가자 A와 B가 부여한 점수간의 상관관계를 분석한 것으로 각각의 평가자와 평가 영역에 따라 다른 상관계수를 보여준다. 모든 항목의 결과를 종합한 했을 때, 평가자 A의 평균점수가 평가자 평균점수보다 1.7916만큼 높고, 표준편차는 0.7443만큼 커서 두 평가자간의 상관관계는 Pearson상관계수 r에 의해 0.770이 된다. 보다 구체적으로 말해, Pearson 상관계수 r이 0.770은 두 평가가자의 평가 결과 중 약 59% 겹친다는 것을 의미한다.[4] <표 8>의 결과 IELTS를 사용했을 때 두 평가자 간에 상관관계가 높은 것으로 나타났다고 볼 수 있다. <표 9>는 MATE평가도구를 가지고 평가자 A, B가 부여한 점수들 간의 상관관계를 분석한 것이다.

<표 9> MATE 평가 결과에 대해 평가자들 간의 상관관계

평가방식 & 영역		평가자	평균	표준편차	평균차	표준편차차	t값	P값	상관계수
총괄적		A	6.5313	2.0373	0.4375	0.4397	.478	.640	.863**
		B	6.0938	1.5976					
분석적	유창성	A	6.7500	2.8157	0.7500	1.2122	.655	.523	.854**
		B	6.0000	1.6035					
	어휘	A	5.6250	1.9955	-0.5000	0.0361	-.506	.621	.854**
		B	6.1250	1.9594					
	정확성	A	6.7500	2.1213	0.5000	0.5402	.535	.601	.660
		B	6.2500	1.5811					
	발음	A	7.0000	1.6035	1.0000	-0.8359	.966	.350	.291
		B	6.0000	2.4494					

(*P 〈0.05, **P 〈0.01, ***P 〈0.001)

[4] 영어 능력 평가 분야에서 과학적인 방법론을 처음으로 제시한 미국의 Lado(1961) 또한 말하기 평가 상관계수가 .70~.90정도가 되어야 신뢰할 수 있는 평가라고 간주 할 수 있다고 보고 한 바 있다.

총괄적인 평가를 했을 때는 평가자 A의 평균점수가 평가자 B의 평균점수보다 0.4385만큼 높았고, 표준편차도 0.4397만큼 컸다. Pearson 상관계수 값은 0.863이 나왔는데 이것 역시 두 평가자들의 결과 점수의 약 74%정가서로 겹친다는 것을 의미한다. MATE의 결과 역시 IELTS 결과와 마찬가지로 비교적 높은 평가자 간 상관관계를 가지는 것으로 드러났다.

총괄적으로 보았을 때, <표 8>과 <표 9>에서 IELTS와 MATE를 사용한 평가 결과에 있어 높은 상관관계를 가진 것으로 나타나 일견 두 평가자 간에 평가 도구에 관해 이견이 없는 듯 보이지만, 각 항목 별 상관관계를 자세히 살펴보면 예상하지 못했던 한 가지 특징이 있다는 것을 발견할 수 있다. 주의를 기울여 살펴보아야 할 항목은 '정확성'과 '발음' 항목이다. MATE를 사용했던 IELTS를 사용했던 간에 두 평가자들 간의 유의미한 상관관계를 발견할 수 없는 영역이 바로 이 두 항목이다. 이것은 '유창성'과 '어휘' 항목에 관해서는 두 명의 평가자들 간에 높은 상관관계를 나타내는 것과 대조적으로 '정확성'과 '발음' 이 두 항목에 있어서 평가자들 간에 이견이 나타난 경우가 많다는 것을 암시하는 내용이다. 이에 관련한 내용은 결론에서 다시 논의하기로 하겠다.

5.2.2 재검사 신뢰도

평가 신뢰도 중 한 하위 항목은 재검사 신뢰도이다. 이는 동일한 검사 도구를 가지고 반복하여 동일한 대상을 측정하였을 때 어느 정도로 동일한 결과를 얻느냐에 기초를 둔 신뢰도를 말한다. 단, 동일한 검사 도구를 반복하여 시행하는 경우 피험자들이 평가내용을 기억할 수 있기 때문에 정확한 재검사 신뢰도를 측정할 수 없다는 한계가 있다. 본 연구에서는 성인 학습자들의 영어 말하기를 11월에 녹화하고 1월에 녹화하였는데, 동일한 자료를 가지고 11월에 평가 받은 것과 1월에 평가받은 것의 상관관계를 분석해 보았다.

<표 10> IELTS로 평가한 결과

		1월 평가
11월 평가	상관계수	.769(***)
	유의확률 값(P)	.000

(*P 〈0.05, **P 〈0.01, ***P 〈0.001)

<표 11> MATE로 평가한 결과

		1월 평가
11월 평가	상관계수	.705(***)
	유의확률 값(P)	.000

(*P 〈0.05, **P 〈0.01, ***P 〈0.001)

위의 표에서 알 수 있듯이 11월에 평가한 것과 1월에 평가한 것의 상관계수가 IELTS의 경우 0.769로 나왔고, MATE의 경우 0.705로 나와서 유의미한 확률 값과 강한 상관관계를 보이는 것으로 나타나 이 두 평가 도구는 재검사 신뢰도가 높은 것으로 추정할 수 있다.

5.2.3 평가자들의 피드백

여러 차례의 평가를 마치고 난 후에 평가자 A, B에게 각 평가도구에 대한 피드백을 받기위해 간단한 설문지를 작성하여 답변을 부탁했다. 제시한 7가지 질문에 대한 답변은 각 문항은 1점에서 5까지 반응할 수 있도록 구성하였다. 설문지에서 각 문항의 응답이 5평점 척도(Likert-type Scale)로 되어서 채점 방법은 '매우 그렇다(Yes, very)' 5점, '대체로 그렇다(Yes, probably)' 4점, '모르겠다(I don't know)' 3점, '대체로 그렇지 않다(Not really)' 2점, '매우 그렇지 않다(No, not at all)'에 1점을 배당하였다. 그 결과는 <표 12>와 <표 13>에 나타나 있다.

<표 12> 평가자 A의 피드백

	Questions	IELTS					MATE				
		5	4	3	2	1	5	4	3	2	1
1	Is the differential (description) clear?	●							●		
2	Is the statement unambiguous?		●					●			
3	Are the criteria specific (enough to decide the candidate's level)?	●						●			
4	Is having 9 levels (for IELTS) or 10 levels (for MATE) sufficient?	●						●			
5	Are there equal intervals between levels?		●					●			
6	Is it appropriate to apply these criteria for Korean students?			●					●		
7	Was it easy to choose your marks?	●								●	
8	Would you like to make any comment about these criteria?										
	MATE Layout difficult										

<표 13> 평가자 B의 피드백

	Questions	IELTS					MATE				
		5	4	3	2	1	5	4	3	2	1
1	Is the differential (description) clear?	●							●		
2	Is the statement unambiguous?	●								●	
3	Are the criteria specific (enough to decide the candidate's level)?	●							●		
4	Is having 9 levels (for IELTS) or 10 levels (for MATE) sufficient?	●					●				
5	Are there equal intervals between levels?		●						●		
6	Is it appropriate to apply these criteria for Korean students?	●						●			
7	Was it easy to choose your marks?		●							●	
8	Would you like to make any comment about these criteria?										
	The IELTS banding descriptions are more objective so more reliable										

<표 12>와 <표13>의 내용을 간단히 기술하자면, 두 평가자의 공통적인 대답은 IELTS로 평가했을 때는 네 가지 평가 기준이 명확하게 제시되어 어려움이 없었던 반면, MATE의 경우에는 전체적인 영어 능력 수준에 대한 기술만이 언급되어 있었다는 것이었다. 평가자들이 말하기 능력 평가에 다년간 참여한 경험이 있고 각각의 평가 도구의 기준표를 충분히 숙지했음에도 불구하고 여전히 평가자들은 평가도구의 많은 문장들이 애매하게 기술되어 있어 평가의 곤란함이 있었다고 언급했다. 그런 답변은 특히 MATE에서 상대적으로 많이 발견되었다. <표 12>와 <표 13>를 보면 IELTS의 경우에는 긍정적인 답변인 5번 쪽에 대체적으로 점수가 많이 분포한 반면에 MATE의 경우에는 그렇지 못함을 발견 할 수 있다. 이들의 주관적인 답변으로 IELTS가 MATE가 우월하다 혹은 MATE가 문제가 많다고 단정 지어 말 할 수 는 없지만, 전반적으로 평가 기준 진술에 있어서 모호하다는 것이 MATE에 관한 평가자들의 의견임 알 수 있다.

6. 결론 및 제언

본 연구는 최근 들어 사회적 요구가 높아지고 있는 영어 말하기 능력 평가 도구들 중에서 국외에서 개발된 평가 도구 중의 하나인 IELTS과 국내에서 개발된 평가도구 중의 하나인 MATE를 선택하여 이들을 비교분석해보았다. 영어 말하기 불안감이 높은, 중간 수준의 영어 숙달도를 가진 30명의 대학생 학습자를 대상으로 보다 자신감 있게 영어 말하기를 시도할

수 있게 해주는 English Public Speaking Program을 삼 개월에 걸쳐서 10차시로 제공하고 English Public Speaking Program의 사전과 사후의 말하기 과제의 결과를 비디오 녹화 하였다. 두 명의 훈련된 영어 모국어 화자들은 녹화된 비디오 자료를 보고 각각 IELTS와 MATE를 사용하여 이들의 말하기 능력을 평가하였다.

그 결과, IELTS로 평가한 경우에는 실험집단의 경우에는 말하기 능력이 향상되었다는 ($P<0.05$ 수준)로 유의미한 차이를 얻었지만, MATE로 평가한 경우에는 실험집단과 통제집단간의 유의미한 차이를 얻을 수 없어 실험집단의 영어 말하기 능력이 향상되었다고 볼 수 없다는 상이한 결과가 나타났다. 이에 본 연구에서는 IELTS와 MATE의 평가 결과들 간의 동형 검사 신뢰도와 재검사 신뢰도를 검증한 결과, 총괄적으로는 두 평가 결과들 간에 높은 상관도를 보이지만, '정확성'과 '발음' 의 두 항목에 대해서만 평가자들 간의 이견이 있음을 발견하였다. 뿐만 아니라 평가자들에게 주관적으로 평가도구들에 대한 의견을 작성하도록 요구하였을 때, 평가자들 모두 MATE의 평가기준표가 상대적으로 애매하다고 답변하였다. 이 점은 MATE의 개발과 사용에 연관된 영어교육 관련자들이 주의를 요하는 점이라고 사려 된다.

국내외의 여러 말하기 평가 도구들 중에서 단 두 평가도구를 선정하여 비교하였을 뿐임에도 불구하고 우리는 말하기 능력 평가들 간의 상이한 결과가 나올 수 있음을 본 연구를 통해서 발견하였다. 더욱 중요한 점은 '발음'과 '정확성' 항목에서 보여 진 평가자들 간의 이견이다. '정확성'과 '발음' 항목은 '어휘'와 마찬가지로 상대적으로 평가하기 명료한 항목이라고 간주되는 것이 일반적이다. 영어교육에 관련된 전문가이든 비전문가이든 영어 모국어 화자가 아닌 한국인 누구나 쉽게 어느 화자의 영어 발음이 좋고 나쁨을 쉽게 판정할 수 있다고 여기는 것이 사실이다. 그러나 본 연구 결과에서는 이러한 일반적인 기대가 영어 모국어 화자의 경우에는 적용되지 않을 수 있음을 드러냈다. 다시 말해, '발음'과 '정확성' 항목에서 모국어 화자들의 판정이 용이하지 않을 수 있으므로 보다 자세한 기준이 마련되어야 할 필요가 있음을 암시하는 것이다.

연구 대상자가 30여명이며 평가자가 영어 모국어 화자 두 명이므로 본 연구 결과를 일반화하는 데 어려움이 있다는 것이 본 연구의 제한점임을 밝히며, 보다 많은 연구 대상자를 대상으로 더 많은 영어 모국어 화자와 영어 모국어 화자가 아닌 평가자들이 유사한 연구를 수행하였을 때 어떤 결과를 가져올 것인가는 추후 연구에서 관심 있게 살펴 보야 할 내용이다.

참고문헌

김덕기, 안병규, 오윤자, 김영규. 1999. 중등영어 수행능력 절대평가기준 제안: 표현 기능을 중심으로. *영어교육*, Vol.54, 175-200.

신동일. 2001a. ACTFL-SOPI체제의 영어 말하기 평가 현장 적용: 한국에서의 문제점과 현실적 대안. *영어교육*, Vol.56, 309-331.

신동일. 2001b. 일반화 가능도 이론 적용을 중심으로 한 말하기 평가도구 타당도 검증 연구. *응용언어학*, Vol. 17, 199-221.

신동일. 2001c. 채점 경향 분석을 위한 Rasch 측정모형 적용연구. *외국어교육*, Vol. 8, 249-272.

신동일. 2002. Rasch 모형을 이용한 고등학교 영어과 말하기 및 쓰기 능력 등급 기술표 개발. *영어교육*, Vol.57, 469-499.

신동일, 서인영. 2002. SLA 선행 연구에 근거한 말하기 평가척도 제안: 문법 영역 중심으로. *영어교육*, Vol.57, 423-450.

이영식. 2000. 영어작문 평가에 대한 채점자 훈련의 원리. *영어교육*, Vol. 55, 201-217.

정택진. 2002. 영어 말하기 능력 평가 방법에 대한 이론적 고찰. *동서대학교 동서논문집* Vol. 8, 133-141.

정 행. 2000. SOPI를 활용한 영어 구술능력 평가 개발과 결과 분석. *영어교육*, Vol. 55, 219-241.

최인철. 2000. 영어 의사소통능력의 모의 구슬 면접시험 방식 양상 차당성 검증. *응용언어학*, 16, 215-246.

ACTFL Proficiency Guidelines. 1986. Hasting-on-hudson, New York: ACTFL.

Lado, R. 1961. *Language teaching*. New York: McGraw-Hill.

MATE. 2001. http://sookmyung.ac.kr/~smu-mate.

SEPT. 2002. http://www.ybmsisa.com/sept.htm.

Shohamy, E. 1994. The validity of direct versus semi-direct oral test. *Language Testing*, Vol. 11, 99-124.

TOP. 2002. http://www.teps.or.kr/top_checkfactor.html.

TSE. 2002. http://www.toefl.org/tse.tseindx.html.

이은혜
부산광역시 사하구 하단2동 840번지
동아대학교 영어영문학부
E-mail: gracelee@hanmail.net
Fax: +82-51-200-7057

김성아
서울시 동대문구 회기동 1번지
경희대학교 교양학부
E-mail: sakim@khu.ac.kr
Fax: +82-2-961-9185

접수일자: 2007. 03. 31
게재일자: 2007. 06. 20