영어어문교육 *제22권* 4호 2016년 겨울

대학생의 개인적 특성이 영어 말하기 시험의 타당성에 미치는 영향

주 미 진
(강원대학교)

Joo, Mi-jin (2016). The effect of individual characteristics of university students on validity of English speaking tests. *English Language & Literature Teaching, 22*(4), 315-337.

This study examined the effect of individual characteristics (introversion, confidence in English, English ability, gender) of students on validity of English speaking tests through the comparison of perceptions and performance of different groups of students in an Interview Oral Test (IOT), a Pair Oral Test (POT), and a Group Oral Test (GOT). Twenty-eight participants took the three tests and responded to the questionnaire assessing their perceptions of and reactions to the tests. Main findings of the study were as follows. First, introversion, English ability, gender did not influence the perceptions of validity, reliability, practicality, backwash effects of each test. However, the students who had confidence in English tended to have more positive attitudes to validity and practicality of IOT. In test preference, less introverted, more confident, or male students were more favorable to IOT and more introverted, less confident, or female students were more positive to POT. Lastly, introversion, gender, and confidence in English were not related to speaking performances on IOT, POT, GOT, but English ability had significant relationship with those of all of the tests.

[individual characteristics/validity/speaking test, 개인적 특성/타당성/말하기 시험]

Ⅰ. 서론

영어 말하기시험은 일반적으로 학생, 평가과제, 질문자, 채점자의 상호작용으로 이루어지므로 타당성을 저해하는 요소는 매우 다양하다. 그 중에서 학생의 개인적 특성은 시험에 대한 인식과 수행에 영향을 미치는 중요한 요인들 중 한 가지로 알려져 있다. 학생의 수행뿐만 아니라 인식 (또는 반응)에도 관심을 갖는 이유는 이들의 인식이 시험의 표면 타당성(face

validity)을 결정짓는 큰 부분을 차지하기 때문이다(Bensoussan & Zeidner, 1989; Brown, 1991; Scott, 1986; Zeidner & Bensoussan, 1988). 학생 관점에서의 시험유형에 대한 인식조사는 표면 타당성을 조사하는데 중요한 역할을 하고 있으며, 시험의 공정성(fairness)을 유지하고 학생의 불만족을 줄이는데도 유용한 도구가 되고 있다. 그러므로 본 연구는 학생의 개인적 특성이 영어말하기 시험의 타당성에 영향을 미치는지 연구하기 위하여 (일대일) 인터뷰시험(Interview Oral Test), 페어시험(Pair Oral Test), 그룹시험(Group Oral Test)에 대한 학생들의 인식과 수행능력이 그들의 특성(내향성, 영어자신감, 영어능력, 성별)에 따라 변화하는지 조사하고자 한다.

본 연구를 위한 영어말하기 시험으로서 POT (Pair Oral Test), GOT (Group Oral Test)를 선택한 이유는 주로 타당성과 신뢰성은 높으나 실용성이 떨어지는 것으로 알려진 IOT (Interview Oral Test)[1]의 대안으로서(Joo, 2008a) 이들의 타당성을 IOT와 비교하여 살펴보고 교실에서의 사용을 고려해 보기 위함이다. 일반적으로 2명으로 구성되는 POT와 4명으로 구성되는 GOT는 시간과 노력면에서 경제적이고 학생들간의 상호작용만으로 대화가 이루어지므로 채점자는 채점에만 집중할 수 있어 채점 신뢰성을 높일 수 있을 것으로 본다. 또한, IOT보다 좀 더 실제적인 자연스러운 대화 상황을 재현할 뿐만 아니라 친근하고 편안한 분위기 속에서 시험 수행이 가능하여 시험에 대한 긴장(anxiety) 또는 떨림을 줄여주므로 학생의 능력발휘에 도움을 줄 것으로 보이며 결과적으로 시험의 타당성을 높이는데 기여할 것으로 본다. 대학생의 개인적 특성이 IOT, POT, GOT의 타당성에 미치는 영향을 조사하기 위하여 다음의 연구문제를 설정하였다.

1. 개인적 특성이 시험에 대한 인식에 영향을 미치는가?
2. 개인적 특성이 시험 수행에 영향을 미치는가?

Ⅱ. 선행연구

1. 개인적 특성과 인식에 관한 연구

학생의 시험에 대한 인식 조사는 시험의 (표면) 타당성과 관련하여 그 중요성이 오래 전부터 강조되어 왔으므로, 이에 대한 연구는 다수 이루어져 왔다(Brown, 1993; Brutch, 1979; Clark, 1988; Joo, 2008b; Madsen & Murray, 1984; Savignon, 1972; Scott, 1980; Scott & Madsen, 1983; Shohamy, 1982;

[1] 이후의 문장부터는 편리를 위하여 인터뷰시험은 IOT, 페어시험은 POT, 그룹시험은 GOT로 표기하고자 한다.

Shohamy, Donitsa-Schmidt, & Waizer, 1993; Stansfield, 1991; Stansfield, Kenyon, Paive, Doyle, Ulsh, & Cowles, 1990). 이러한 연구들은 시험유형(test format), 길이(length), 시간(time constraints), 시험환경(testing environment), 시험유형과의 익숙도(familiarity), 시험 타당성(validity) 또는 신뢰성(reliability)에 대한 인식, 어려움(difficulty), 지시(instruction)의 명확성, 긴장(anxiety) 등과 같은 매우 다양한 요인들이 학생의 정의적 반응 또는 인식에 영향을 줄 수 있다고 보고하고 있다.

학생의 개인적 특성이 인식에 미치는 영향에 관한 연구는 상대적으로 그 수가 매우 적었지만, 다음과 같은 연구들을 찾을 수 있었다. Scott(1980)은 스페인과 일본인 학생들의 일대일 인터뷰, 읽기시험(불필요한 어휘 찾아 지우기), 문법시험(듣고 두 번째 어휘적기)의 첫 번째 시험과 일대일 인터뷰, 읽기시험(그림보고 true-false 문제 풀기), 듣기 시험(질문을 듣고 테입에 답하기)으로 구성되는 두 번째 시험에 대한 수행능력을 조사하였다. 그 결과, 스페인 학생들이 일본인 학생들보다는 시험들에 대하여 좀 더 긍정적인 태도를 보임으로서 사회문화적 배경이 학생의 시험상의 수행에서 뿐만 아니라 시험에 대한 태도 또는 인식에도 영향을 미친다는 것을 발견하였다. Zeidner(1988)는 서양(western)과 아시아학생들의 학업성취도시험(scholastic ability exam)에 대한 인식의 차이를 조사하였고, 그 결과 사회문화적인(sociocultural) 배경 변수들은 일반적으로 시험에 대한 인식에 큰 영향을 미치지 않는 것으로 결론내렸다. Zeidner와 Bensoussan(1988)은 남성이 여성보다 유의미하게 쓰기시험(written test)보다 말하기시험(oral test)에 좀 더 긍정적으로 응답하였고 여성이 시험 수행 시 더 긴장(anxiety)하였다고 보고하였다. 하지만, 전체 그룹뿐만 아니라 성별그룹에 따른 인식과 수행간에도 유의미한 상관관계를 보이지는 않았다. Bradshaw(1990)은 문법과 독해력(reading comprehension) 그리고 C-test로 구성된 배치시험(placement test)에 대한 인식조사에서 성별에 따른 반응의 차이는 없었다고 보고하였다. Brown (1991)도 일본어 말하기와 듣기 시험에 대한 인식에서 성별에 따른 유의미한 차이는 없었다고 결론내렸다. Scott(1986)은 낮은 언어능력의 학생들은 6명으로 구성되는 그룹시험(GOT)을 선호하고 높은 언어능력의 수험자는 2명으로 구성되는 페어시험(POT)을 선호할 것으로 가정하였지만 연구결과 유의한 인식의 차이를 발견하지 못하였다.

위와 같이, 학생의 개인적 특성이 시험의 인식에 미치는 영향을 조사한 연구는 몇몇 찾아 볼 수 있었지만, 학생의 개인적 특성이 POT, GOT에 대한 인식에 미치는 영향을 조사한 연구는 Scott(1986)을 제외하고는 찾아 볼 수 없었다. 그리고 Scott(1986)의 연구에서 조차도 IOT와의 비교는 이루어지지 않은 점에서 본 연구와는 구별된다.

2. 개인적 특성과 수행에 관한 연구

Bachman(1990)은 다양한 특성을 가진 각 학생 (또는 수험자)은 다른 유형의 시험형태에서 다른 수행을 할 수 있다고 주장하였다. 그리고 그러한 수행의 차이는 그들의 능력으로부터 기인된 것이 아니므로 시험의 타당성을 위협하게 된다. 그러므로 능력 이외에 수행에 영향을 주어 시험의 타당성을 감소시키는 변수들은 찾아서 그 영향을 최소화해야 한다. 학생의 특성과 수행에 관련한 논문은 다수 있었는데, 그 중 Swinton과 Powers(1980)은 아프리카어, 아랍어, 중국어, 페르시아어, 독일어, 일본어, 스페인어의 7개 언어그룹을 대상으로 연구한 결과 유럽인과 비유럽인들 사이에 TOEFL(Test of English as a Foreign Language)에서의 수행의 차이를 발견하였다. 마찬가지로, Alderman과 Holland(1981)도 비슷한 언어능력을 가졌지만 다른 모국어를 가진 학생들간의 TOEFL 시험문항에서의 수행의 차이를 발견하였고 이러한 차이는 학생의 모국어에 따라 상대적으로 이익 또는 불이익을 받을 수 있다는 것을 의미한다고 보고하였다. Politzer와 McGroarty(1985)은 아시아계와 히스패닉계 언어그룹간의 수행의 차이는 언어능력(linguistic competence), 듣기능력(auditory comprehension), 전반적인 말하기 능력(overall oral proficiency), 의사소통능력(communicative competence)에 걸쳐 유의미하게 나타난다고 보고하였다. Farhady(1982)은 성별, 학교에서의 지위, 전공, 국적에 따라서 여러 영어시험에서 수행의 차이를 보임을 발견하였다. Spurling과 Ilyin(1985)은 나이, 모국어, 고등학교의 수료여부가 클로즈(cloze) 시험, 읽기시험, 언어구조(structure) 시험, 듣기시험 상에서의 수행에 영향을 준다고 결론내렸다. Porter와 Shen(1991)은 다양한 국적을 가진 각 14명의 남성과 여성을 대상으로 성별, 직위, 상호작용 스타일이 구술시험상의 수행에 미치는 영향을 조사하였다. 그 결과, 직위는 수행에 영향을 주지 않았고 성별은 수행에 영향을 미쳤는데 면접관이 여성이었을 경우 좀 더 높은 점수를 받았음을 알 수 있었다. O'Loughlin(2002)은 IELTS 인터뷰상에서의 각 8명의 남녀 학생의 수행 차이를 조사했는데 성별간의 수행에는 차이가 없음을 발견하였다.

학생의 특성에 따른 POT와 GOT에 관한 연구도 있었는데 살펴보면 다음과 같다. Davis(2009)은 POT상에서 대화자간의 영어능숙도가 수행에 미치는 영향을 조사해 본 결과, 영어능숙도가 비슷한 학생들간 뿐만 아니라 다른 학생들간 시험에서도 수행에 유의미한 영향을 주지 않는 것으로 보고하였다. Liski와 Puntanen(1983)은 수다스러움(talktiveness)의 정도에 따른 수행의 차이를 조사하였다. 그 결과, 좀 더 수다스러운 학생들이 덜 수다스러운 학생들보다 오류가 적다는 것을 발견하였다. Bonk와 Van Moere(2004)은 수줍음(shyness)이 GOT 성적에 미치는 영향을 조사하였다. 연구 결과, 수줍음이 큰 학생일수록 GOT 성적이 낮게 나타났다.

위와 같이 학생의 개인적 특성이 시험형태의 인식 또는 수행에 미치는 연구결과는 다양하다. 선행연구들을 통하여 알 수 있는 것은 학생의 특성이 시험 타당성에 영향을 미칠 수 있음을 의미하며, 이러한 우려는 당연한 것으로 보인다. 하지만, 학생의 다양한 특성(내향성, 자신감, 영어능력, 성별)을 IOT, POT, GOT의 비교를 통하여 이들 시험상의 점수에 미치는 영향에 대해 조사한 논문은 찾을 수 없었다. 더욱이, 이러한 학생의 특성이 IOT, POT, GOT에 대한 인식에 영향을 주는지에 대한 연구는 이루어지지 않고 있었다. 그러므로 본 논문의 결과는 관련 연구자, 교육자, 평가자들에게 많은 시사점을 제시할 것으로 본다.

Ⅲ. 연구방법

1. 연구 대상 및 절차

본 연구는 한국 K대학의 영어학 전공 2학년 여학생 7명과 남학생 21명으로 구성된 총 28명의 학생들이 연구에 참여하였다. 참여 학생들의 영어능력은 제출된 토익 성적으로 측정되었고 학생들의 평균 (읽기와 듣기로 구성된) 토익 점수는 461점으로 높지 않은 편이었다. 한 학기 동안 학생들은 각 한달 간격으로 세 가지 시험(IOT, POT, GOT)을 모두 치르도록 하였으며, 각 시험상의 수행은 비디오녹음 되었다. 세 가지 시험 후 학생들은 시험에 대한 인식 및 태도에 대한 설문지를 작성케 하였다. IOT, POT, GOT상의 학생들의 수행비디오는 다년의 말하기 평가경험을 가진 두 명의 채점자가 학생들의 대화를 들으면서 동일 평가표를 참고하여 각 학생의 발음, 유창성, 문법, 어휘, 응집력에 대하여 개별 평가를 하였다. 채점 평가표는 O'Loughlin(2001: 217-219)의 채점표를 일부 수정 및 보완하여 사용하였다.

2. 연구도구

1) 설문지

IOT, POT, GOT에 대한 학생들의 인식과 특성을 조사하고자 총 35문항의 설문지를 작성하였다. 그 중 학생들의 인식을 조사하고자 주미진과 김양희(2015)의 설문지 문항 중 타당성, 신뢰성, 실용성, 역류효과에 관해 물어보는 문항과 선호도에 관한 선택형 문항을 이용하였다. 문항에 대한 자세한 내용은 다음과 같다.

타당성은 4문항('좋은 말하기 시험방식이다', '의사소통능력을 측정하는데 좋은 시험방식이다', '영어능력 발휘를 충분히 할 수 있는 시험방식 이다',

‘공평한 시험이다’), 신뢰성은 3문항(‘신뢰할 수 있는 시험이다’, ‘충분한 질문이 주어졌다’, ‘시험 외의 다른 요인이 시험을 방해하였다’), 실용성은 2문항(‘시험방식이 편리하다’, ‘앞으로도 교실에서 활용되었으면 좋겠다’), 역류효과는 3문항(‘시험 결과는 앞으로의 영어 학습에 도움이 될 것이다’, ‘영어 말하기 능력 향상에 도움이 되는 시험이다’, ‘영어 말하기 학습에 긍정적인 영향을 미친다’) 그리고 선호도 7문항(‘어떤 시험이 더 편안한가?’, ‘어떤 시험이 더 실용적인가’, ‘어떤 시험이 말하기 능력을 평가하기에 더 적절한가?’, ‘어떤 시험이 영어로 더 많은 말을 할 수 있도록 유도하는가?’, ‘어떤 시험이 더 공평한가?’, ‘어떤 시험이 더 좋은가?’, ‘어떤 시험이 향후 계속적으로 활용되었으면 좋겠는가?’)으로 이루어졌다. 타당성, 신뢰성, 실용성, 역류효과에 관한 문항에서는 4점 척도(매우 그렇지 않다~매우 그렇다)를 이용하였고 선호도에 관한 질문은 선택형 문항으로 세 가지 유형의 시험 중 하나만을 선택하도록 하여 가장 호의적으로 또는 긍정적으로 인식하는 시험유형이 무엇인지 조사하고자 의도하였다.

내향성, 영어에 대한 자신감, 성별 등의 개인의 특성에 대해 조사하기 위해서는 Bonk와 Van Moere(2004)의 수줍음(shyness)에 대한 설문지 문항을 참고하여 8문항(‘대화 시 실수할까봐 두렵다’, ‘소규모일수록 말을 더 많이 하는 경향이 있다’, ‘여러 사람 앞에서는 말하기가 부담스럽다’, ‘모르는 사람과 대화 시 긴장된다’, ‘대화 시 망설이다가 종종 말할 기회를 놓친다’, ‘말하는 것보다 남의 이야기를 듣는 것을 더 좋아한다’, ‘말수가 적은 편이다’, ‘내성적인 편이다’), 영어에 대한 자신감에 대한 7문항(‘교사/수와의 영어대화가 동료학생과의 대화보다 편안하다’, ‘한국인보다 원어민과의 영어 대화가 즐겁다’, ‘영어사용이 즐겁다’, ‘영어 대화 시 자신있게 말한다’, ‘여러사람 앞에서의 영어발표가 두렵다’, ‘영어 대화 시 실수할까 두렵다’, ‘영어로 질문을 받을까봐 두렵다’), 그리고 성별에 관한 1문항으로 구성하였다.

2) IOT, POT, GOT

IOT는 채점자 특성이 IOT상의 수행에 미치는 영향을 최소화하기 위하여 연구자가 직접 면접 질문자로 참여하여 개별 학생을 대상으로 실행하였다. 시험 문항은 실용영어회화 교재(English Firsthand 1)에서 보편적으로 다루어지는 주제(예. 직업, 가족, 일상 스케줄, 휴가, 주말활동, 취미, 외모)에 대한 질문들로서 묻고 답하기(question and answer)형식으로 진행되었다. 각 시험은 15분 정도 소요되었다.

POT는 2명이 한 조가 되어 교재 내에 있는 주제에 대해 자유롭게 질문하고 답할 수 있도록 구성하였다. 학생들은 책상 위에 있는 대화의 주제가 적혀있는 카드를 한 사람씩 순서대로 젖혀보고 카드를 젖혀 본 학생이 상대방 학생에게 카드에 쓰여진 주제와 관련하여 질문하고 필요 시 추가질문(following-up

questions)을 하도록 장려하였으며 상대학생은 그에 답하도록 하였다. IOT와 마찬가지로 질문-대답형식의 대화가 이루어졌다. 하지만 면접자가 주로 질문하고 학생이 답하는 IOT와 달리 POT, GOT는 학생들 상호간의 질문, 대답이 가능하다는 점에서 큰 차이점이 있다. GOT는 4명으로 구성되었으며 POT와 같은 방식으로 시험이 진행되었다. POT와의 유일한 차이점은 각 그룹을 구성하는 학생 수뿐이었다. POT, GOT 모두 각 20분 정도 소요되었다.

3. 분석방법

세 가지 유형의 시험에 대한 인식과 수행을 비교하기 위하여 SPSS version 19.0을 사용하였다. 본 연구의 참여자 수가 적은 관계로 정규성이 의심되므로 비모수적 통계분석(Nonparametric Analysis) 실행과 함께 유의수준은 0.1로 정하여 분석하기로 결정하였다. 학생 특성이 각 시험에 대한 인식과 수행에 영향을 미치는지 조사하기 위하여 기술통계(descriptive statistic), 크루스칼-월리스(Kruskal-Wallis) 검정, 스피어만 상관분석(Spearman correlation analysis), 맨 휘트니(Mann-Whitney) 검정, 교차분석(crosstabluation)을 실시하였다.

Ⅳ. 분석결과

1. 인식에 미치는 영향

각 시험에 대한 설문지 문항을 타당성, 신뢰성, 실용성, 역류효과로 나누어 이들의 평균값을 이용하여 학생들의 개인적 특성이 인식에 영향을 미치는지 알아보았다. 인식에 대한 총12문항 신뢰도는 Cronbach alpha = .938으로 매우 높았다. 기술통계를 통하여 각 인식에 대한 응답 평균을 살펴보면 다음과 같다(표1).

표 1

각 시험에 대한 학생의 인식

시험 유형	타당성		신뢰성		실용성		역류효과	
	평균	S.D	평균	S.D	평균	S.D	평균	S.D
IOT	3.22	.57	2.81	.36	3.10	.60	3.06	.58
POT	3.04	.65	2.69	.40	3.10	.58	2.96	.64
GOT	2.86	.63	2.57	.45	2.94	.57	2.94	.68

학생들은 타당성, 신뢰성, 실용성, 역류효과에 대한 인식에서 전반적으로 IOT, POT, GOT 순으로 긍정적으로 응답하였음을 알 수 있다. 그리고 학생들의 특성 중 내성적인 성격이 이러한 인식에 미치는 영향이 있는지 알아본 결과, 내향성 그룹 간에는 유의한 인식의 차가 없었다(표 2). 즉, 학생의 내향성이 타당성, 신뢰성, 실용성, 역류효과에 대한 인식에 영향을 미치지 않는 것으로 보인다(p>.05). Spearman 상관분석 결과에서도 내향성, 인식 간에는 관련성이 없음을 볼 수 있었다(p>.05).

표 2

내향성이 시험에 대한 인식에 미치는 영향: Kruskal-Wallis 검정 결과

인식	시험 유형	내향성(N)	평균 순위	χ2	*df*	*p*
타당성	IOT	1 (7) 2 (11) 3 (8)	18.36 11.27 12.31	4.10	2	.129
	POT	1 (7) 2 (11) 3 (8)	13.43 11.86 15.81	1.28	2	.527
	GOT	1 (7) 2 (11) 3 (8)	15.00 12.86 13.06	.38	2	.827
신뢰성	IOT	1 (7) 2 (11) 3 (8)	13.57 12.85 12.69	.07	2	.967
	POT	1 (7) 2 (11) 3 (8)	11.07 12.05 15.88	2.08	2	.354
	GOT	1 (7) 2 (11) 3 (8)	14.36 10.85 14.50	1.52	2	.467
실용성	IOT	1 (7) 2 (11) 3 (8)	16.71 12.86 11.56	1.97	2	.373
	POT	1 (7) 2 (11) 3 (8)	14.14 13.55 12.88	.12	2	.944
	GOT	1 (7) 2 (11) 3 (8)	13.14 13.82 13.38	.04	2	.980
역류효과	IOT	1 (7) 2 (11) 3 (8)	12.43 12.77 15.44	.84	2	.658
	POT	1 (7) 2 (11) 3 (8)	11.71 13.05 15.69	1.33	2	.513
	GOT	1 (7) 2 (11) 3 (8)	14.79 12.18 14.19	.64	2	.726

*p<0.1, **p<0.05, ***p<0.01

내향성이 시험유형의 선호도에 영향을 주는지 조사하기 위하여 교차분석을 실시한 결과, 위의 Kruskal-Wallis 검정 결과와는 달리(표 2), 내향성이 약할수록, 즉 외향적인 학생일수록 IOT가 편안하다고 느꼈으며, 내성적인 학생일수록 POT가 편안하다고 느꼈다(X2=10.18, p=.038). 또한, 내성적인 학생일수록 POT가 말을 많이 하도록 유도하는 시험형태라고 생각하는 경향이 있었다(X2=8.54, p=.074). 이와 비슷하게, 내성적 성격이 약할수록 IOT를 선호하였고, 내성적 성격이 강할수록 POT를 선호한다고 응답하였다(X2=11.50, p=.021): 내성적 성격이 약한 71.4%(n=5), 강한 57.1%(n=4)의 학생들이 각각 IOT, POT 시험을 선호하였다.

영어에 대한 자신감은 IOT의 타당성과 실용성에 대한 인식에 영향을 미치는 것으로 보인다(p<0.1). 자신감이 약한 그룹과 강한 그룹간에는 IOT의 타당성, 실용성에 대한 인식에 차이를 보였는데, 자신감이 강한 학생이 좀 더 긍정적인 자세를 보였다(표 3). Spearman 상관분석 결과에서도 자신감과 IOT의 타당성, 실용성에 대한 인식 간에는 관련성이 있었다(p.<0.05). 자신감은 POT의 실용성과도 약한 상관관계를 보이기는 하였지만(p.<0.1), 그룹간의 인식에 있어 유의한 차이를 보이는 정도는 아니였다(표 4).

표 3

자신감이 시험에 대한 인식에 미치는 영향: Kruskal-Wallis 검정 결과

인식	시험 유형	자신감(N)	평균 순위	χ2	*df*	*p*
타당성	IOT	1 (10)	10.95	5.37*	2	.068
		2 (10)	12.73			
		3 (5)	20.30			
	POT	1 (10)	13.10	.82	2	.664
		2 (10)	12.64			
		3 (5)	16.20			
	GOT	1 (10)	11.50	1.14	2	.567
		2 (10)	14.73			
		3 (5)	14.80			
신뢰성	IOT	1 (10)	11.25	1.26	2	.532
		2 (10)	13.55			
		3 (5)	15.40			
	POT	1 (10)	11.20	1.79	2	.409
		2 (10)	15.25			
		3 (5)	12.10			
	GOT	1 (10)	10.60	2.43	2	.297
		2 (10)	15.55			
		3 (5)	12.70			

	IOT	1 (10) 2 (10) 3 (5)	10.60 13.23 19.90	5.35*	2	.069
실용성	POT	1 (10) 2 (10) 3 (5)	10.90 14.55 16.40	2.34	2	.311
	GOT	1 (10) 2 (10) 3 (5)	14.40 12.00 15.00	.83	2	.660
	IOT	1 (10) 2 (10) 3 (5)	12.80 12.86 16.30	.93	2	.630
역류효과	POT	1 (10) 2 (10) 3 (5)	13.05 13.14 15.20	.38	2	.827
	GOT	1 (10) 2 (10) 3 (5)	12.35 14.64 13.30	.51	2	.773

*p<0.1, **p<0.05, ***p<0.01

표 4

영어 자신감과 인식간의 관련성: Spearman 상관분석 결과

인식	시험유형	rho	p
타당성	IOT	.45**	.020
	POT	.13	.532
	GOT	.17	.415
신뢰성	IOT	.33	.108
	POT	.11	.619
	GOT	.23	.260
실용성	IOT	.48**	.014
	POT	.34*	.085
	GOT	.04	.867
역류효과	IOT	.14	.506
	POT	.10	.634
	GOT	.13	.517

*p<0.1, **p<0.05, ***p<0.01

자신감과 선호도에 대한 교차분석에서도 비슷한 결과를 보였다. 영어에 대한 자신감이 높은 학생일수록 IOT가 편안하고(X2=8.58, p=.072) 말하기 능력을

평가하기에 적절한 시험이라고(X2=7.89, *p*=.096) 느꼈다. 자신감이 높은 학생들의 100%(n=5)가 IOT가 가장 적절한 말하기 능력 평가 도구이고, 교실에서 활용하기에 가장 좋으며(X2=11.43, *p*=.022) 공평(X2=11.74, *p*=.019)한 시험이라 응답하였다. 그에 반해, 자신감이 낮을수록 POT가 공평하고, 교실에서 계속적으로 활용되기를 바라며, 말하기 능력을 평가하기에 적절하다고 느끼는 경향이 있음을 볼 수 있었다. 또한, 자신감이 낮을수록 GOT가 가장 편안하다고 느꼈다. 자신감이 낮은 각각 40%(n=4), 40%(n=4)의 학생들과 중간인 학생들 30%(n=3), 50%(n=5)가 POT 또는 GOT가 교실에서 계속적으로 활용되기를 원했다. 이러한 결과는 영어에 대한 자신감이 높은 학생일수록 IOT에, 낮은 학생일 수록 POT에 좀 더 긍정적인 태도를 보였음을 말해준다 – 자신감이 낮은 학생들은 POT 또는 GOT에 긍정적인 태도를 보였지만 POT에 대한 긍정성이 좀 더 높은 것으로 보여진다.

영어능력이 인식에 미치는 영향을 살펴보면, 영어능력이 상위에 속하는 학생들이 각 시험의 역류효과에 대하여 가장 긍정적으로 반응하긴 하였지만(표 5), 평균 순위를 살펴보면 상위, 하위, 중위 학생 순으로 긍정적이었으므로 영어능력에 따라 순차적으로 일어난 것은 아니었다. 그러므로 IOT, POT, GOT의 역류효과에 대한 인식이 영어능력 상위, 중위, 하위 학생 간에 유의한 차이가 있긴 하지만, 그 차이를 영어능력에 따른 영향으로 보기는 어렵다. 또한, 교차분석 결과 영어능력은 선호도를 나타내는 어떠한 문항과도 유의한 인식의 차를 보이지 않음으로써 특정 시험에 대한 선호도와도 관련이 없음을 나타내었다.

영어능력이 우수할수록 영어에 대한 자신감도 높을 것으로 예상하였지만 기대와는 달리 영어능력과 자신감은 상관분석 결과 관련이 없었다(rho=.007, *p*=.975). 그러므로 영어에 대한 자신감은 본인의 실제 영어능력보다는 내향성 또는 영어에 대한 인식 또는 태도와 관련이 있어 보였다.

표 5

영어능력이 시험에 대한 인식에 미치는 영향: Kruskal-Wallis 검정 결과

인식	시험 유형	영어능력(N)	평균 순위	X 2	*df*	*p*
타당성	IOT	1 (8)	16.00	2.84	2	.242
		2 (9)	10.44			
		3 (9)	15.56			
	POT	1 (8)	17.33	3.09	2	.214
		2 (9)	10.89			
		3 (9)	13.78			
	GOT	1 (8)	13.94	.05	2	.975
		2 (9)	13.61			
		3 (9)	14.44			

	IOT	1 (8) 2 (9) 3 (9)	14.19 12.28 14.11	.38	2	.826
신뢰성	POT	1 (8) 2 (9) 3 (9)	14.31 12.00 14.28	.59	2	.743
	GOT	1 (8) 2 (9) 3 (9)	13.88 10.61 16.06	2.48	2	.289
	IOT	1 (8) 2 (9) 3 (9)	16.17 8.88 13.56	4.52	2	.104
실용성	POT	1 (8) 2 (9) 3 (9)	16.56 11.00 11.00	3.74	2	.154
	GOT	1 (8) 2 (9) 3 (9)	13.56 10.69 14.69	1.40	2	.496
	IOT	1 (8) 2 (9) 3 (9)	16.72 7.31 14.50	8.37**	2	.015
역류 효과	POT	1 (8) 2 (9) 3 (9)	16.94 7.75 13.81	8.22**	2	.016
	GOT	1 (8) 2 (9) 3 (9)	14.67 8.38 15.75	5.22*	2	.074

*p<.1, **p<.05, ***p<.01

표 6

성별이 인식에 미치는 영향: Mann-Whitney U 검정 결과

종속변수		평균 순위 여성-남성	U	Z	p
타당성	IOT	14.57	73.00	-.03	.978
	POT	15.50	66.50	-.38	.704
	GOT	15.64	65.50	-.43	.668
신뢰성	IOT	11.86	55.00	-.87	.385
	POT	13.00	63.00	-.41	.683
	GOT	13.93	69.50	-.03	.977

	IOT	14.17	56.00	-.25	.800
실용성	POT	10.58	42.50	-1.13	.259
	GOT	14.50	54.00	-.38	.702
	IOT	13.58	56.50	-.03	.974
역류효과	POT	14.83	52.00	-.54	.587
	GOT	15.17	50.00	-.64	.525

*p<0.1, **p<0.05, ***p<0.01

Mann-Whitney U 검정 결과에 따르면, 성별 그룹 간의 인식에도 유의한 차는 없었다(p>0.05)(표6). 즉, 성별도 인식에 영향을 미치지 않는 것으로 나타났다. 반면, 성별에 따른 선호도의 차이를 살펴보면, 여성은 POT(n=4, 66.7%), 남성은 IOT(n=11, 57.9%)가 말하기 능력 평가 도구로서 가장 적절하다고 느꼈음을 알 수 있었다(χ2=7.96, p=.019). 또한, 여성의 66.7%(n=4)가 POT, 남성의 52.6%(n=10)가 IOT가 더 공평하다고 응답하였다(χ2=7.91, p=.019).

내향성, 자신감, 성별 그룹의 타당성, 신뢰성, 실용성, 역류효과에 대한 인식 조사를 위한 Kruskal-Wallis 검정과 Spearman 상관분석의 결과가 이들 특성들과 선호도에 대한 교차분석의 결과와 약간의 차이를 낳은 원인으로는 참여자 수가 적은 상황에서 타당성, 신뢰성, 실용성, 역류효과에 대한 인식을 물어보는 4점 척도에서 보다는 3가지 시험유형 중 하나만을 고르도록 하는 선택형 문항에서 학생의 반응과 태도가 더욱 분명히 나타났을 가능성이 높다. 각 인식 문항에 대한 응답 평균 수치를 살펴보면 외향적이거나 자신감이 강하거나 또는 남학생은 각 타당성, 신뢰성, 실용성, 역류효과에 대한 인식에서 IOT에 대한 응답 평균 수치가 가장 높고 (즉, 가장 긍정적으로 응답하였고), 내성적이거나 자신감이 낮거나 또는 여학생은 POT에 대한 응답 평균 수치가 가장 높았음을 볼 수 있었다. 이러한 수치는 외향적이거나, 자신감이 높거나, 남성인 경우는 IOT에 좀 더 긍정적이며, 반대로 내성적이거나, 자신감이 낮거나, 여성인 경우는 POT에 대해 좀 더 긍정적인 경향이 있음을 말해준다.

요약하면, 학생 특성 중 내향성, 영어능력, 성별은 각 시험의 타당성, 신뢰성, 실용성, 역류효과에 대한 인식에 영향을 미치지 않는 것으로 보인다. 하지만 영어 자신감은 인식에 영향을 미쳤는데 자신감이 강한 학생이 IOT의 타당성과 실용성에 대한 인식이 좀 더 긍정적임을 볼 수 있었다. 전반적인 선호도에서는 외향적이거나 자신감이 높은 학생일수록 IOT에 대해 긍정적이었고 내성적이거나 자신감이 낮은 학생일수록 POT를 선호하는 경향이 있음을 볼 수 있었다. 그리고 남성은 IOT, 여성은 POT에 대해 좀 더 긍정적인 자세를 보였다.

2. 수행에 미치는 영향

학생 특성이 수행에 미치는 영향을 조사한 결과, 내향성과 자신감은 각 시험상에서의 수행과는 관련이 없는 것으로 나타났다(표 7, 표 8). 즉, 내향성과 자신감에 따른 시험상의 수행에는 변화가 없었다.

표 7

내향성이 말하기 수행에 미치는 영향: Kruskal-Wallis 검정 결과

시험 유형	평가 항목	내향성(N)	평균 순위	χ^2	*df*	*p*
IOT	발음	1 (7) 2 (11) 3 (8)	11.14 13.27 15.88	1.74	2	.420
	유창성	1 (7) 2 (11) 3 (8)	12.00 14.14 13.94	.40	2	.818
	문법	1 (7) 2 (11) 3 (8)	11.71 13.91 14.50	.59	2	.747
	어휘	1 (7) 2 (11) 3 (8)	12.07 13.09 15.31	.79	2	.674
	응집력	1 (7) 2 (11) 3 (8)	13.36 12.91 14.44	.21	2	.902
	합계	1 (7) 2 (11) 3 (8)	11.93 13.50 14.88	.56	2	.757
POT	발음	1 (7) 2 (11) 3 (8)	14.29 13.95 12.19	.53	2	.769
	유창성	1 (7) 2 (11) 3 (8)	12.71 14.59 12.69	.46	2	.796
	문법	1 (7) 2 (11) 3 (8)	12.07 14.55 13.31	.50	2	.777
	어휘	1 (7) 2 (11) 3 (8)	12.29 14.64 13.00	.55	2	.761
	응집력	1 (7) 2 (11) 3 (8)	13.00 15.59 11.06	2.50	2	.287
	합계	1 (7) 2 (11) 3 (8)	11.29 14.86 13.56	.96	2	.619

GOT	발음	1 (7) 2 (11) 3 (8)	14.86 14.27 11.25	1.19	2	.553
	유창성	1 (7) 2 (11) 3 (8)	15.00 13.73 11.88	.72	2	.698
	문법	1 (7) 2 (11) 3 (8)	11.43 15.27 12.88	1.28	2	.528
	어휘	1 (7) 2 (11) 3 (8)	13.29 15.14 11.44	1.32	2	.518
	응집력	1 (7) 2 (11) 3 (8)	13.21 14.91 11.81	.90	2	.636
	합계	1 (7) 2 (11) 3 (8)	13.79 14.41 12.00	.48	2	.787

*p<0.1, **p<0.05, ***p<0.01

표 8

자신감이 말하기 수행에 미치는 영향: Kruskal-Wallis 검정 결과

시험 유형	평가항목	자신감(N)	평균순위	χ2	*df*	*p*
IOT	발음	1 (10) 2 (11) 3 (5)	13.60 13.00 14.40	.14	2	.932
	유창성	1 (10) 2 (11) 3 (5)	13.70 12.45 15.40	.56	2	.754
	문법	1 (10) 2 (11) 3 (5)	13.70 13.05 14.10	.08	2	.960
	어휘	1 (10) 2 (11) 3 (5)	13.45 12.82 15.10	.33	2	.846
	응집력	1 (10) 2 (11) 3 (5)	12.00 13.95 15.50	.84	2	.657
	합계	1 (10) 2 (11) 3 (5)	13.25 13.09 14.90	.21	2	.900

POT	발음	1 (10) 2 (11) 3 (5)	11.65 13.09 18.10	3.67	2	.160
	유창성	1 (10) 2 (11) 3 (5)	11.30 14.41 15.90	1.74	2	.419
	문법	1 (10) 2 (11) 3 (5)	14.00 11.73 16.40	1.50	2	.472
	어휘	1 (10) 2 (11) 3 (5)	13.60 12.18 16.20	1.15	2	.563
	응집력	1 (10) 2 (11) 3 (5)	13.60 11.77 17.10	2.50	2	.286
	합계	1 (10) 2 (11) 3 (5)	13.10 12.68 16.10	.75	2	.688
GOT	발음	1 (10) 2 (11) 3 (5)	10.90 13.23 19.30	4.68	2	.097
	유창성	1 (10) 2 (11) 3 (5)	12.80 11.27 19.80	4.96	2	.084
	문법	1 (10) 2 (11) 3 (5)	13.95 11.09 17.90	3.06	2	.216
	어휘	1 (10) 2 (11) 3 (5)	13.30 11.59 18.10	3.02	2	.221
	응집력	1 (10) 2 (11) 3 (5)	13.50 11.00 19.00	4.40	2	.111
	합계	1 (10) 2 (11) 3 (5)	12.75 11.32 19.80	4.45	2	.108

*p<0.1, **p<0.05, ***p<0.01

학생들의 영어능력이 말하기 시험상의 수행에 영향을 미치는지 분석한 결과(표 9), 읽기, 듣기로 구성된 토익성적을 토대로 분류된 상위, 중위, 하위 학생들의 영어능력은 IOT, POT, GOT 성적과 유의한 관련이 있음을 볼 수 있었다. 가끔 하위 그룹이 중위 그룹보다 말하기 수행에서 조금 더 높은 점수를 받기도 하였지만 중위, 하위 그룹 간에는 유의한 점수 차가 나지 않았다. 그리고 상위 그룹의 말하기 수행 성적은 다른 두 그룹의 성적과는 차별화되게 높은 점수를 받았음을 볼 수 있었다. Spearman 상관분석 결과에서도 토익성적은 각 시험의 말하기 수행과 유의한 상관관계가 있음을 볼 수

있었다. 그리고 영어능력(즉, 토익성적)과의 관련성 정도는 세 시험 모두 비슷하였다(표 10).

표 9

영어능력이 말하기 수행에 미치는 영향: Kruskal-Wallis 검정 결과

시험 유형	평가항목	영어능력(N)	평균 순위	X 2	*df*	*p*
IOT	발음	1 (9)	18.56	5.61***	2	.006
		2 (9)	11.72			
		3 (9)	11.72			
	유창성	1 (9)	20.56	10.31***	2	.006
		2 (9)	9.83			
		3 (9)	11.61			
	문법	1 (9)	20.83	10.74***	2	.005
		2 (9)	10.17			
		3 (9)	11.00			
	어휘	1 (9)	20.22	9.59***	2	.008
		2 (9)	9.94			
		3 (9)	11.83			
	응집력	1 (9)	20.78	10.92***	2	.004
		2 (9)	10.61			
		3 (9)	10.61			
	합계	1 (9)	20.78	10.00***	2	.007
		2 (9)	10.11			
		3 (9)	11.11			
POT	발음	1 (9)	19.61	11.72***	2	.003
		2 (9)	10.50			
		3 (9)	11.89			
	유창성	1 (9)	18.39	5.04*	2	.081
		2 (9)	11.50			
		3 (9)	12.11			
	문법	1 (9)	20.44	10.35***	2	.006
		2 (9)	11.33			
		3 (9)	10.22			
	어휘	1 (9)	17.83	3.82	2	.148
		2 (9)	12.61			
		3 (9)	11.56			
	응집력	1 (9)	18.17	6.48**	2	.039
		2 (9)	12.56			
		3 (9)	11.28			
	합계	1 (9)	20.83	10.24***	2	.006
		2 (9)	10.61			
		3 (9)	10.56			

GOT	발음	1 (9)	18.17	4.33	2	.115
		2 (9)	11.72			
		3 (9)	12.11			
	유창성	1 (9)	18.22	4.93*	2	.085
		2 (9)	10.61			
		3 (9)	13.17			
	문법	1 (9)	21.22	12.73***	2	.002
		2 (9)	11.11			
		3 (9)	9.67			
	어휘	1 (9)	19.50	8.02**	2	.018
		2 (9)	11.17			
		3 (9)	11.33			
	응집력	1 (9)	19.44	8.02**	2	.018
		2 (9)	12.00			
		3 (9)	10.56			
	합계	1 (9)	20.17	8.45**	2	.015
		2 (9)	10.17			
		3 (9)	11.67			

*p<0.1, **p<0.05, ***p<0.01

표 10

영어능력과 말하기 수행간의 관련성: Spearman 상관분석 결과

평가항목	IOT rho(p)	POT rho(p)	GOT rho(p)
발음	.378* (.052)	.591*** (.001)	.386** (.047)
유창성	.475** (.012)	.289 (.143)	.356* (.069)
문법	.505*** (.007)	.497*** (.008)	.589*** (.007)
어휘	.400** (.039)	.317 (.107)	.515*** (.007)
응집력	.529*** (.005)	.429** (.026)	.515*** (.006)
합계	.458** (.016)	.467** (.014)	.465** (.015)

*p<0.1, **p<0.05, ***p<0.01

성별 그룹간과 말하기 수행능력 차이를 Mann-Whitney U 검정 결과를 통하여 알아본 결과, 성별 그룹간에 유의한 성적 차를 보이지 않았음을 볼 수 있었다(표 11). 즉, 성별차이는 인식과 마찬가지로 말하기 수행에도 영향을 미치지 않는 것으로 보인다.

표 11

성별 그룹간과 말하기 수행간의 Mann-Whitney U 검정 결과

종속 변수	평가 항목	평균순위 여성-남성	U	Z	*p*
IOT	발음	14.43-14.52	73.00	-.03	.977
	유창성	13.71-14.76	68.00	-.31	.761
	문법	13.71-14.76	68.00	-.30	.763
	어휘	15.36-14.21	67.50	-.34	.737
	응집력	14.50-14.50	73.50	.00	1.00
	합계	14.21-14.60	71.50	-.11	.92
POT	발음	13.29-14.90	65.00	-.57	.570
	유창성	15.36-14.21	67.50	-.35	.729
	문법	13.43-14.86	66.00	-.42	.672
	어휘	14.93-14.36	70.50	-.17	.863
	응집력	15.14-14.29	69.00	-.30	.764
	합계	14.50-14.50	73.50	.00	1.000
GOT	발음	17.43-13.52	53.00	-1.17	.241
	유창성	15.93-14.02	63.50	-.56	.572
	문법	12.86-15.05	62.00	-.64	.520
	어휘	12.64-15.12	60.50	-.76	.449
	응집력	13.86-14.71	69.00	-.26	.793
	합계	14.36-14.55	72.50	-.05	.957

*p<0.1, **p<0.05, ***p<0.01

간략히 요약하면, 내향성, 자신감, 그리고 성별의 차이는 각 시험상에서의 말하기 수행능력과는 관련이 없음을 볼 수 있었다. 하지만, 영어능력은 각 시험의 말하기 수행과 유의미한 관련이 있었다. 영어능력이 뛰어난 학생은 각 시험상의 말하기 수행능력도 높은 경향이 있었다.

Ⅴ. 결론

본 논문은 학생의 개인적 특성이 영어말하기 시험의 타당성에 미치는 영향을 연구하기 위하여 IOT, POT, GOT의 비교를 통하여 학생의 내향성, 자신감, 영어능력, 성별이 이들 시험에 대한 인식과 수행 점수에 미치는 영향에 대하여 조사하였고, 연구결과는 다음과 같다. 첫째, 학생들의 개인적 특성이 시험 유형에 대한 인식에 미치는 영향을 조사한 결과, 내향성, 영어능력, 성별은 각

시험의 타당성, 신뢰성, 실용성, 역류효과에 대한 전반적인 인식에 영향을 미치지 않는 것으로 보인다. 하지만 영어에 대한 자신감은 인식에 영향을 미쳤는데 자신감이 강한 학생은 IOT의 타당성과 실용성에 대한 인식이 좀 더 긍정적임을 볼 수 있었다. 그리고 영어능력은 영어에 대한 자신감과는 관련이 없음을 알 수 있었다. 타당성, 신뢰성, 실용성, 역류효과에 대한 인식과는 달리, 전반적인 선호도에서는 외향적이거나 자신감이 높은 학생일수록 또는 남성인 경우는 IOT에 대해 좀 더 긍정적이었고 내성적이거나 자신감이 낮은 학생일수록 또는 여성인 경우는 POT를 선호하는 경향이 있음을 볼 수 있었다.

IOT는 연구자가 직접 면접관 역할을 하였으므로 학생들이 상대적으로 더 떨리고 어렵다고 느꼈을 것으로 추측된다. 이러한 상황에서 성격이 외향적이고 영어에 대한 자신감이 높은 학생들은 IOT에 대해서 좀 더 긍정적인 반응을 보였을 것으로 생각되고 그에 반해 소극적이고 자신감이 없는 학생들은 평가자의 입장이기도 하는 교수 앞에서 더 소극적으로 처신하게 되고 시험 자체가 부담스럽게 느껴졌을 것이다. 그에 반해, 자신과 비슷한 실력을 가진 급우와의 대화는 IOT보다는 훨씬 부담감과 긴장감이 적고 자신의 실력을 좀 더 발휘할 수 있는 계기가 되었을 가능성이 높다. 하지만 POT와 마찬가지로 급우들 간의 대화로 이루어지는 GOT는 덜 선호하였는데 아마도 내성적이고 영어 자신감이 부족한 학생들은 급우들 속에서 선뜻 말할 기회를 찾기가 힘들었을 것이며 영어실력이 좀 더 뛰어난 학생들에게 주눅이 들고 소극적으로 행동되어 자신의 실력 발휘 뿐만 아니라 GOT에 대한 긍정적 인식도 상대적으로 낮았을 가능성을 배제할 수 없다.

둘째, 내향성, 자신감, 그리고 성별의 차이는 각 시험상에서의 말하기 수행능력과는 관련이 없음을 볼 수 있었다. 하지만, 영어능력은 각 시험의 말하기 수행과 유의한 상관관계를 보임으로서 토익의 읽기와 듣기 능력이 각 시험 유형의 말하기 능력과 관련성이 있음을 알 수 있었다. 즉, 토익성적이 뛰어난 학생은 각 시험상의 말하기 수행능력도 높은 편이었다. 그러므로 영어능력을 제외한 학생의 특성은 각 시험의 말하기 수행에 영향을 끼치지 않으므로 타당성을 위협하는 요인으로 작용하지 않았음을 볼 수 있었다.

결론적으로, 학생의 개인적 특성이 시험에 대한 인식과 말하기 수행에 부정적인 영향을 끼치거나 시험의 타당성을 해친다고 보기는 어려우므로, IOT보다 경제적일 뿐만 아니라 IOT의 단점을 보완할 수 있는 POT, GOT의 교실 사용을 적극 고려해 볼 만하다. 그리고 교실 사용 시에는 학생들의 영어 실력이 낮거나 소극적이라 판단될 때는 IOT보다는 POT가 더 적절할 것으로 보이며, 학생 수가 적으면서도 학생들의 영어에 대한 자신감이 높거나 적극적인 학생들일 경우는 IOT가 좀 더 효과적일 것으로 보인다. 그러므로 본 연구결과는 평가도구를 학생들의 특성, 평가 목적과 환경 또는 교실 상황에 따라 달리 선택해야 함을 말해준다.

하지만, 다음의 연구 제한점을 가지므로 결과해석에 주의를 요한다. 첫째,

참여 학생 수가 특정지역의 28명의 학생들만을 대상으로 하였으므로, 통계적으로 충분한 표본수를 확보하지 못하였다. 둘째, 설문지 문항을 이용하여 학생들의 주관적인 의견만으로 내향성과 자신감을 판단하였다. 셋째, 편리를 위하여 참여학생들의 영어능력을 읽기, 듣기로 구성된 토익 성적만으로 측정하므로서 타당성이 의심될 수 있다. 그러므로 이러한 제한점을 고려한 반복 연구가 필요하다고 본다.

본 연구를 포함하여 다수의 POT, GOT에 대한 선행 연구에도 불구하고 여전히 이에 대한 다각적인 연구가 계속되어야 할 것으로 보인다. 예를 들면, 상대 학생, 즉 질문자의 다양한 특성이 학생에게 미치는 영향은 무엇인지 POT, GOT상에서의 수행 비교를 통하여 밝힐 필요가 있다. 또한, 이들 시험상에서의 상호작용 양상은 어떠한지, 평가 가능한 충분한 언어를 학생들로부터 도출해 낼 수 있는지, 그리고 POT, GOT 상에서 질적으로 같은 또는 다른 과제 유형 또는 언어를 생성하도록 하는지에 대한 연구와 함께 채점자 신뢰성도 조사되어 채점자 훈련에 활용되어야 한다. 이와 같은 연구는 상대 학생 또는 그룹 구성원이 어떤 방식으로 수행과 점수에 영향을 미치는지에 대한 이해를 넓히는데 기여할 것이므로 POT, GOT을 좀 더 적절히 구성하고 실행하는데 이바지 할 것이다.

참고문헌

주미진, 김양희(2015). 인터뷰, 페어, 그룹 구술시험에 대한 인식 및 수행 비교 조사. *영어교육연구. 27*(1), 137-156.

Alderman, D. L., & Holland, P. W. (1981). *Item performance across native language groups on the Test of English as Foreign Language* (TOEFL Research Report No. 6). Princeton, NJ: Educational Testing Service.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bensoussan, M., & Zeidner, M. (1989). Anxiety and achievement in a multicultural situation: The oral testing of advanced English reading comprehension. *Assessment and Evaluation in Higher Education, 14*(1), 40-54.

Bonk, W. J., & Van Moere, A. (2004). *L2 group oral testing: The influence of shyness/outgoingness, match of interlocutors' proficiency level, and gender on individual scores.* Paper presented at the Language Testing Research Colloquium. Temecula, CA.

Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing, 7*, 13-30.

Brown, A. (1991). *The role of test-taker feedback in the validation of a language proficiency test.* Unpublished master's thesis. University of Melbourne, Australia.

Brown, A. (1993). Test development process: test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing, 10*, 277-303.

Brutch, S. (1979). *Convergent/Discriminant validation of prospective teacher proficiency in oral and written production of French by means of MLA foreign language proficiency tests for teachers (TOP and TWP) and self ratings.* Unpublished doctoral dissertation. University of Minnesota, Minnesota, USA.

Clark, J. L. D. (1988). Validation of a tape-mediated ACTFL/ILR scale-based test of Chinese speaking proficiency. *Language Testing, 5*, 187-198.

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing, 26*(3), 367-396.

Joo, M. J. (2008a). The need for an alternative approach to oral testing. *English Language & Literature Teaching, 14*(1), 1-20.

Joo, M. J. (2008b). Face validity of direct and semi-direct oral tests from test taker perspectives. *The New Korean Journal of English Language & Literature, 50*(2), 285-310.

Liski, E., & Puntanen, S. (1983). A study of the statistical foundations of group conversation tests in spoken English. *Language Learning, 33*, 225-246.

Madsen, H. S., & Murray, N. (1984). *Retrospective evaluation of testing in ESL content and skills courses.* Unpublished manuscript. Brigham Young University, Provo, UT.

O'Loughlin, K. (2001). *The equivalence of direct and semi-direct speaking tests (studies in language testing 13)*. Cambridge: Cambridge University Press.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing* 19, 169–92.

Politzer, R. L., & McGroarty, M. (1985). An exploratory study of learning behaviors and their relationship to gains in linguistic and communicative competence. *TESOL Quarterly, 19*, 103-123.

Porter, D., & Shen, S. H. (1991). Gender, status and style in the interview. *The Dolphin, 21*, 117-128. Aarhus: Aarhus University Press.

Savignon, S. J. (1972). *Communicative competence: An experiment in foreign language teaching.* Philadelphia, Pennsylvania: The Center for Curriculum Development Inc.

Scott, M. L. (1980). *The effect of multiple retesting on affect and test performance*. Unpublished MA thesis. Brigham Young University, Provo, UT.

Scott, M. L. (1986). Student affective reactions to oral language tests. *Language Testing, 3*(1), 99-118.

Scott, M. L., & Madsen, H. S. (1983). The influence of retesting on test affect. In J. W. Oller (Ed.), *Issues in language testing research* (pp.270-279). Rowley, MA: Newbury House Publishers.

Shohamy, E. (1982). Affective considerations in language testing. *Modern Language Journal,*

66, 13-17.

Shohamy, E., Donitsa-Schmidt, S., & Waizer, R. (1993). *The effect of the elicitation mode on the language samples obtained in oral tests.* Paper presented at the 15th Language Testing Research Colloquium, Cambridge, UK.

Stansfield, C. W. (1991). A comparative analysis of simulated and direct oral proficiency interviews. In S. Anivan (Ed.), *Current developments in language testing* (pp. 199-209). Singapore: Regional Language Center.

Stansfield, C. W., Kenyon, D. M., Paive, R., Doyle, F., Ulsh, I., & Cowles, M. (1990). The development and validation of the Portuguese speaking test. *Hispania, 73*, 641-51.

Spurling, S., & Ilyin, D. (1985). The impact of learner variables on language test performance. *TESOL Quarterly, 19*, 283-301.

Swinton, S. S., & Powers, D. E. (1980). *Factor analysis of the test of English as a foreign language for several language groups* (TOEFL Research Report No. 6). Princeton, NJ: Educational Testing Service.

Zeidner, M. (1988). Sociocultural differences in examinees' attitudes toward scholastic ability exams. *Journal of Educational Measurement, 25*, 67-75.

Zeidner, M., & Bensoussan, M. (1988). College students' attitudes towards written versus oral tests of English as a foreign language. *Learning Testing, 5*, 100-114.

예시언어(examples in): English
적용가능 언어(Applicable Language): English
적용가능 수준(Applicable Level): Secondary

주미진
강원대학교 삼척캠퍼스
영어과
245-711 강원도 삼척시 중앙로 346
Email: ing1115@hanmail.net

Received in October 15th, 2016
Reviewed in November 26th, 2016
Revised version received in December 5th, 2016