

외국어 시험의 유용성에 관한 연구: 평가도구의 목적과 사용을 중심으로

신상근
(이화여자대학교)

Shin, Sang-Keun. (2004). An exploratory study of the usefulness of foreign language tests in Korea. *Foreign Language Education Research*, 7, 57-70.

This paper examined the usefulness of widely used foreign language tests in Korea, focusing on their purposes. The analyses of secondary data showed that the tests are used for purposes for which they have not been validated, which call into question the appropriateness of interpretations test users make of test scores. The results also indicated that the reliability of test scores could be undermined because of test security violations. Finally, the lack of correspondence of test tasks and real life tasks also question the meaningfulness of the interpretations that test users make on the basis of test scores.

Even though the present study was only exploratory in nature, its findings are hoped to contribute to an understanding of the usefulness of language tests for a given purpose.

Like medicines, language tests need clear labels giving details of measurement error and validation, which should appear on the face of score reports rather than in manuals read only by experts (Spolsky, 1995).

의사소통 외국어 시험에 관한 지금까지의 논의는 평가도구 자체의 특성에

관한 분석이 주를 이루고 있다(Bachman, 2000). 예를 들어 요인 분석이나 문항반응이론 같은 양적(quantitative) 연구 방법을 사용해서 문항의 난이도나 변별도를 분석하거나 시험 점수의 신뢰도를 추정해 보는 연구는 많이 이루어진 반면에 외국어 평가 분야의 또 다른 중요한 영역인 평가도구의 개발이나 목적, 그리고 평가 결과의 해석과 사용에 관한 연구는 상대적으로 많이 이루어지지 않는 편이다. 그 결과 언어평가 연구를 통계분석과 동일시 하는 잘못된 견해가 자리잡고 있기도 하다. 외국어 평가가 결국 수험자에 대한 각종 결정을 내리기 위해서 실시되고 시험의 결과가 개별 수험자나 시험의 사용자, 나아가 사회에 크고 작은 영향을 미친다는 점을 고려해 볼 때 의사소통 능력 평가가 어떤 목적으로 개발되어 실시되고 있고, 시험 결과가 어떤 식으로 사용되고 있는지에 대한 연구의 중요성은 아무리 강조해도 지나치지 않다. 아무리 정교한 통계 분석에 기초해서 개발된 시험이라 하더라도 적절하지 않게 사용된다면 수험생과 시험을 사용하는 사용자 모두에게 부정적인 영향을 미칠 수 밖에 없기 때문이다. 본 논문에서는 한국 사회에서 외국어 의사소통능력 시험이 어떤 식으로 사용되고 있는지 신문기사와 인터넷 자료 분석을 통해서 알아보려고 한다.

I. 이론적 배경

1. 평가의 목적

언어평가를 제작하거나 사용할 때 우리가 던져야 할 첫 번째 질문은 평가의 목적(purpose)이라고 할 수 있다. 어떤 결정을 내리기 위해서 평가도구가 사용될지가 분명하지 않다면 측정하고자 하는 구인(construct)을 정의하거나 과제(task)를 개발하는 일이 어려워질 수 밖에 없기 때문이다. 따라서 평가의 목적은 그 시험의 구인과 과제가 제대로 정의되고 설계되었는지를 알아보는 타당도 검사의 근거가 된다고 볼 수 있다(Chapelle 외 2인, 2003; Shepard,

1993). 평가도구를 선택해서 사용하고자 하는 경우에도 평가도구가 제작된 목적이 가장 중요한 고려사항의 하나이어야 한다. 두말 할 필요 없이 사용하고자 하는 평가도구가 제작된 목적을 살펴보아야 자기에게 주어진 목적에 부합되는 시험인지 판단할 수 있기 때문이다. 예를 들어 토플(Test of English as a Foreign Language) 시험의 경우 미국으로 유학오는 비원어민 유학생들이 영어로 진행되는 수업을 제대로 소화해 낼 만한 영어 실력이 있는지를 알아보기 위해 제작된 시험이다(ETS, 2004). 따라서 토플시험을 이와는 다른 목적으로 사용했을 때 이 시험이 수험자들의 언어 능력을 제대로 측정한다고 보기는 어려울 것이다.

주어진 목적에 충실한 평가도구를 개발해야 하고, 주어진 목적에 적합한 시험을 선택해서 사용해야 한다는 사실은 너무나 당연해서 재론의 여지가 없어 보인다. 아무리 잘 만들어진 평가도구라 하더라도 원래 의도된 목적이 아닌 다른 목적으로 사용될 때 잘못된 결정이 내려질 가능성이 크고 이는 개인적으로나 사회적으로도 바람직하지 못한 결과를 낳을 수 밖에 없다. 이런 이유로 해서 정도의 차이는 있지만 평가도구가 수험자나, 학교 교육, 더 나아가 사회 전체에 미치는 영향도 평가도구의 타당도 논의의 중요한 한 부분이어야 한다고 주장하는 사람들도 많다. 그러나 각종 평가도구들이 원래 제작된 의도와 같은 목적으로 사용되고 있는지에 대한 연구는 별로 이루어지지 않았다.

2. 시험의 유용성

의사소통 외국어 시험을 평가하는 여러가지 척도가 있을 수 있다. 예를 들어 대표적인 준거로 신뢰도와 타당도를 들 수 있고, 최근 들어 시험의 영향도 시험의 중요한 척도로 자리잡고 있다. 본 논문에서는 Bachman과 Palmer (1996)가 제안한 시험의 유용도(utility)를 분석의 틀로 이용하고자 한다. 그들은 각 상황에서 어떤 시험이 얼마나 유용한지를 크게 여섯가지의 요소,

즉 구인 타당도(construct validity), 신뢰도(reliability), 영향(impact), 진정성(authenticity), 실용성(practicality), 그리고 상호작용성(interactiveness) 측면에서 살펴볼 것을 제안했다.

이들이 제안한 평가모델은 시험의 유용성을 분석할 수 있는 구체적인 가이드라인을 제공했다는 측면에서 볼 때, 타당도를 실제로 어떻게 검증할 수 있는지에 대한 설명이 부족하다는 Messick(1989)의 타당도 모델이 갖고 있는 한계를 극복했다는 데서 의미가 있다고 하겠다. 본 연구에서는 지면의 한계상 신뢰도, 구인 타당도, 진정성, 그리고 영향의 관점에서 한국 사회에서 사용되고 있는 의사소통능력 시험의 유용성에 관해 살펴보도록 한다.

II. 연구결과

외국어 의사소통능력 시험과 관련된 신문기사와 인터넷 자료를 분석한 결과는 다음과 같다.

1. 신뢰도

신뢰도란 측정의 일관성(consistency)을 뜻한다(Bachman, 1990). 예를 들어 동일한 수험자가 동일한 시험을 두 번 보았을 때 비슷한 점수를 얻고, 어떤 작문을 두 명의 채점자가 채점을 했을 때 비슷한 채점 결과가 나올 때 우리는 그 시험 결과의 신뢰도가 높다고 말할 수 있다. 또 다른 각도에서 본다면 신뢰도란 시험 점수에서 차지하는 오류의 비율을 뜻한다. 신뢰도가 높을수록 시험 점수에서 수험생의 실력이 차지하는 비율이 크고 오류가 차지하는 비율이 낮아지게 된다. 따라서 신뢰도는 평가도구 자체의 특성이라기 보다는 시험점수의 특성이라고 볼 수 있다. 한국 사회에서 실시되고 있는 외국어 평가 시험의 시험의 신뢰도를 저하시킬 수 있는 한 가지 요인이 아래에 소개된 토플시험 후기에 잘 나타나있다.

오늘 섬랐습니다.

LC 24...--; 전 거의 30점 받은 줄 알았습니다. 보니까 45차 시험 7-8 시험 후기를 달달 외우고 갔는데. 거의 거기서 다 나오더군요. bural님이 정리하신거... 슷 1개, 렉처 1개 빼놓고.

수험자의 시험에 대한 불안감을 해소시키고 시험 문제 형식에 익숙하지 못한데서 오는 오류를 줄일 수 있다는 측면에서 시험에 대한 정보를 수험자에게 제공하는 일은 매우 중요하다. 그러나 시험에 대한 단순한 정보를 제공하는 차원을 넘어서 실제 시험 문제를 암기해서 인터넷에 공개하는 이런 시험 후기는 결국 평가도구의 신뢰도를 저하시키게 된다. 심지어 수험 교재와 강사 소개란에서도 이런 비슷한 내용들이 발견된다.

고득점 비법 확실히 있죠 50여번 시험...평점 980점 테이프 한 달 100번은 반복. 시험후기 만들어 단점 정리. 한국에서 가장 토익을 잘 가르친다고 소문이 났는데 인기 비결은? 토익시험의 요령을 파헤치기 위하여 6년 전부터 매달 토익시험을 보면서 항상 수강생의 입장에서 생각했다. 영어 공부 때문에 외국에 나간 적 없이 순수하게 우리나라에서 공부해 토익 시험에서 만점을 많이 받아 '비법'이라 공개할 수 있었다. 심지어 시험장에 연필은 문득하게 쥐고 가기, 시험장에 가서는 책상과 의자가 몸에 편한지 확인할 것. 스피커는 어느 학교가 좋다고 알려 주었다.

결국 이러한 활동들은 여러가지 바람직하지 않은 결과로 이어질 수 밖에 없고 시험을 개발한 곳이나 수험자, 그리고 시험 점수를 이용하고자 하는 사용자에게 직·간접적인 영향을 미치게 된다. 예를 들어 다음 기사에 보도된 것처럼 시험을 보는 방식 자체가 달라지기도 한다.

GRE시험] 10월부터 종이시험으로 바뀐다

입력 : 2002.08.08 05:02 57'

미국 대학원 입학에 위한 GRE시험이 오는 10월부터 종전의 컴퓨터 방식에서 종이시험으로 바뀐다. 토플(TOEFL), GRE 시험 등을 주관하는 ETS(Educational Testing Service)는 7월 또는 10월 1일을 기해 한국, 중국(홍콩 포함), 대만 등 아시아 3개국에서 시행하는 GRE 일반시험(General Test)을 현행 컴퓨터 방식에서 지필식으로 변경한다고 공고했다. ETS측은 최근 이들 나라에서 컴퓨터 방식으로 GRE 시험을 치른 수험생들이 문제와 답을 웹사이트에 띄우는 등 시험성적에 의심이 간다는 GRE위원회의 지적에 따라 시험의 보안과 신뢰도를 높이기 위해 일시적으로 시험방식을 바꾸기로 했다고 밝혔다.

시험의 유용성 측면에서 볼 때 이런 식으로 시험 문제를 공개하는 일은 결국 시험 점수의 신뢰도를 저하시키는 결과를 낳는다.

2. 구인 타당도

구인 타당도란 우리가 시험 점수에 근거해서 내리는 결정이 얼마나 적합한지를 의미하기 때문에(Messick, 1989), 타당도란 시험의 특성이 아니라 시험 점수 해석의 적합성 문제라고 볼 수 있다. 따라서 같은 시험이라고 할지라도 어떤 상황에서 어떤 목적으로 사용되느냐에 따라서 그 적합성이 달라진다. 예를 들어 성인용으로 제작된 시험을 어린 아이들에게 실시할 경우 그 평가도구의 적절성은 낮아질 수 밖에 없다.

타당도와 관련되어 우리사회에서 제기되는 가장 큰 문제점은 다음 신문 기사들이 보여주듯이 많은 시험들이 원래 개발된 목적과는 다른 목적으로 사용되고 있다는 점이다.

영어교사 토플 600점 이상에 임용 가산점

입력 : 2001.12.16 19:45 11'

내년 서울지역 중고교 영어교사 신규임용 때 토플 고득점자에 대한 가산점 기준이 강화되고, 초등교사 신규임용 시험에서는 영어 인터뷰가 도

입된다. 유인종 서울시 교육감은 영어 교육 강화를 위해 토플 성적 500 점(신 토플 173점) 이상 지원자에게 차등 가산점을 주던 것을 내년부터 600점(신 토플 250점) 이상 지원자로 상향 조정하고, 초등교사 임용 때는 현행 영어 듣기 평가 대신 영어 인터뷰를 실시한다고 16일 밝혔다.

TEPS시험 봐야 국장발령 내겠다;수원시 과장 2명에 400점 요구

발행일 : 2000-10-06 [사회] 기자/기고자 : 권상은

영어 실력이 되어 승진시켜 주겠다. 경기도 수원시가 40대 과장 2명을 국장에 발령하면서 3개월의 유예를 붙였다. 심재덕 수원시장은 지난 4일 이상윤(이상윤43) 건축과장을 도시계획국장으로, 주양원(주양원45) 도로과장을 건설교통국장으로 발령했다. 그러나 승진은 아니었다. 직위는 4급 서기관 자리이지만 직급은 아직 5급 사무관이기 때문에 '직무대리'라는 꼬리표를 달아야 했다.

심 시장은 이날 임용장을 주는 자리에서 앞으로 3개월 이내에 TEPS에 응시해 400점 이상의 성적표를 들고 와야 서기관 승진 발령을 해주겠다고 반드시 해낼 것으로 믿는다고 말했다. 두 40대 국장은 부담을 느끼는 눈치다. 한 국장은 공부를 열심히 하라는 뜻이겠지만 손을 놓았던 영어 공부를 다시 시작해야겠다고 말했다.

변리사 영어시험 토플토익으로 대체

[중앙일보 2004.05.04 18:10:25]

[중앙일보] 내년부터 변리사 1차시험의 영어과목이 토플(TOEFL)과 토익(TOEIC) 등 민간 영어능력 검정시험 성적으로 대체된다. 정부는 4일 고건 대통령 권한대행 주재로 국무회의를 열고 이런 내용의 변리사법 시행령 개정안을 의결했다. 영어 성적은 시험 통과 여부만을 가리는데 활용되며 총득점에는 합산하지 않을 예정이다. 시험별 통과기준 점수는 ?TOEFL(PBT) 530점?TOEFL(CBT) 197점?TOEIC 700점?TEPS 625점?FLEX 625점 등이다. 2005년 시험의 경우 2003년 1월 1일 이후 취득한 성적이어야 유효하다.

토익·토플시험대체로 司試 응시자 급감

[중앙일보 2004.02.09 18:29:46]

[중앙일보] 사법시험 1차 어학시험을 토익·토플 등 외부 영어시험으로 대체하는 첫 해인 올해 사법시험 응시자수가 지난해에 비해 40%나 준 것으로 나타났다. 법무부는 지난달 14일 마감한 2004년도 사법시험 및 군법무관 임용시험 접수 결과 1차 시험 응시자가 1만9천3백90명으로 집계돼 지난해 3만2천4백1명에 비해 40.2%나 감소했다고 9일 밝혔다. 2002년에는 3만24명이 응시했다.

2005학년도 특목고 전형 특징세부 입시요강] 영어듣기구술 면접이 당락변수

[국민일보 2004.02.09 16:19:00]

420명을 뽑는 명덕외고는 특별전형의 교과성적우수자 정원이 지난해 120명에서 60명으로 줄었다. 24명을 뽑는 전공우수자 전형 지원자격은 토플 성적이 250점 이상이거나 토익 930점 이상 또는 텡스 900점 이상이다. 모두 336명을 선발하는 일반전형은 단계별 전형이 폐지되고 내신 300점, 영어듣기 50점, 구술면접 50점을 반영한다. 교과성적 반영비율(일반특별전형 공통)은 2학년 1학기 30%, 2학년 2학기 30%, 3학년 1학기 40%이다.

예를 들어 유학생들의 영어실력을 측정하기 위해 제작된 토플시험이 한국에서는 법조인이나 변리사, 은행직원, 영어 선생님, 심지어는 외국어 고등학교 신입생을 선발하는데 사용되고 있다. 재론의 여지없이 이러한 평가도구의 선택과 사용은 주어진 목적에 따라 정도의 차이는 있겠지만 바람직하지 않다. 예를 들어 미국 대학교에서 영어로 진행되는 수업을 알아들을 수 있는 능력이 유학생에게 있는지 측정하기 위해 만든 토플시험을 우리나라 중·고등학교 영어 선생님을 뽑는 데 사용할 경우 그 시험의 적합도가 별로 높지 않다는 사실은 깊이 생각해보지 않더라도 누구나 쉽게 짐작할 수 있을 것이다.

3. 시험의 진정성

진정성이란 수험과제와 시험과제 사이의 일치를 의미한다(Bachman & Palmer, 1996). 진정성이 높은 시험과제를 만들기 위해서는 측정하고자 하는 영역에서 이루어지는 언어사용 과제를 조사 분석한 다음 그와 비슷한 평가 과제를 개발해야 한다. 예를 들어 영어 교사가 될 예비 교사들의 영어 실력을 측정하고자 한다면, 영어 교사들이 실제로 영어로 어떤 과제를 수행하는지 조사한 다음 그 과제들과 비슷한 평가 과제를 개발해야 할 것이다. 그런 측면에서 본다면 위에서 소개되었듯이 토플을 여러가지 다른 목적으로 사용하는 사실은 시험의 진정성 관점에서든 문제가 많음을 알 수 있다. 예를 들어 다음 신문 기사를 살펴보자.

광주은행 50명 채용 계약직사무원 등

김성현기자 shkim@chosun.com

입력 : 2004.10.28 10:38 05'

광주은행이 계약직 일반사원과 창구텔러 사무원 등 50명을 신규 채용한다. 계약직 응시자격은 4년제 대학 졸업자 또는 2005년 2월 졸업 예정자로 전산 관련학과와 인문사회계열 전공자. 76년 1월1일 이후 출생자로, 최종학교 전학년 평균성적이 B+ 이상이고, 토익(TOEIC) 750점 또는 토플(TOEFL) 550점, 텡스(TEPS) 650점 이상을 얻어야 한다.

이 시험 결과에 바탕해서 내려진 직원 선발 결정이 타당하기 위해서는 은행 직원들이 업무상 필요한 영어와 미국 대학에서 유학할 때 필요한 영어가 비슷하다는 전제가 성립할 때일 것이다. 그러나 구인 광고나 신문, 그리고 토플시험 매뉴얼 어디에도 그런 설명이나 해석은 보이지 않는다.

4. 시험의 영향

시험의 타당도를 살펴볼 때 시험이 미치는 영향을 어느 정도까지 고려해야 하는지에 대해서는 다른 견해가 존재하지만 시험이 최소한 부정적인 영향을

미쳐서는 안된다는 사실에는 대부분 동의한다. 지금까지 이루어진 시험의 영향에 관한 연구는 수업에 미치는 역류효과(washback) 연구에 집중되어 왔다고 볼 수 있다(Alderson & Wall, 1993). 그러나 최근에 들어 시험이 미치는 영향을 좀 더 광범위하게 살펴 보아야 한다는 주장들이 제기되고 있는데 우리나라에서는 아직까지 시험이 미치는 영향에 대한 관심 자체는 미미하다고 볼 수 있다. 그러나 시험이 개인의 삶에 직접적인 영향을 미치고 있다는 사실은 다음 기사들이 잘 보여준다.

보따리 싸는 '장수생' 신림동 고시촌 썰렁

[중앙일보 2004.07.16 06:29:54]

[중앙일보 이경용 기자] 10년째 서울 신림9동 고시촌에서 사법시험을 준비 중인 박모(37)씨는 요즘 시험공부를 계속해야 할지를 놓고 고민하고 있다. 함께 공부하던 '장수생'(장기수험생) 3명은 이미 짐을 싸들고 고시촌을 떠났다. 지난달 21일부터 4일 동안 2차 시험까지 치른 박씨가 고민하는 것은 달라진 시험제도 때문이다. 특히 올해 도입된 민간 영어시험에 그는 바짝 긴장하고 있다. 2차 시험에 낙방해 내년에 다시 1차 시험을 보려면 토익(700점).토플(530점).토픽스(625점) 가운데 하나를 선택해 일정한 점수 이상을 받아야 한다. 영어책만 달달 외우던 그에게 듣기평가가 들어 있는 새로운 시험은 두렵기만 하다. 박씨는 "고시원에 30세를 넘긴 고시생이 지난해 10명이었으나 지금은 3명밖에 없다"고 말했다.

토하면서 토익·토플 공부하는 아이들

정순화 인턴기자·이화여대 4년·likemaria@empal.com

입력 : 2004.08.07 08:58 31' / 수정 : 2004.08.07 09:03 16'

토플을 공부하는 아이들은 '손익분기점' 뜻도 모른 채 'break even point(손익분기점)'란 단어를 기계적으로 외웁니다. 또 하루 4시간동안 대학교재로 공부하다 보면 다들 지쳐서 널브러지죠. 심지어 수업 도중에 토하는 아이도 있었어요. (서울 강남 A어학원 전직 영어강사 한모씨)

학부모들의 욕심에 의한 과도한 토익·토플 공부로 정신적·육체적 고통

을 겪는 초·중등생들이 많다. 성인 대상 영어평가지험인 토익·토플 점수는 특목고 특별전형시 응시자격으로 활용되고 있다. 따라서 많은 학부모들은 어린 자녀들에게 토익·토플 공부를 시키고 있다.

대학 졸업요건 강화 탈락 속출

발행일 : 2001-02-16 [사회] 기자/기고자 : 양근만대학들이 졸업요건을 강화하면서 이달 졸업예정자 가운데 탈락자가 속출했다. 15일 각 대학에 따르면 서울대학교 인문대학은 지난 95년부터 시행해오던 외국어자격시험 기준을 엄격히 적용, 이번 졸업예정자 중 11명에 대해 '졸업불가' 방침을 확정했다. 특히 졸업 전 텀스(TEPS)시험 의무화가 적용되는 99학번(2003년 2월 졸업예정자)부터는 졸업탈락자가 더욱 늘어날 것으로 학교측은 예상하고 있다. 오는 24일 졸업식을 앞둔 고려대는 경영대학에서 전체 졸업예정자 499명 중 6%에 달하는 30명이 학점불량이나 토익토플 점수 미달 등으로 졸업을 못하게 됐다. 2년 전부터 외국어사회봉사활동 컴퓨터 능력 등 3가지 분야에 대해 '3품제'를 실시해온 성균관대는 3품제에 걸린 졸업 탈락자가 지난해 17명보다 거의 3배 가까운 50명으로 급증했다.

물론 이런 제도 자체가 얼마나 적합한지부터 살펴보아야 할 것이다. 설령 이런 제도들의 도입 취지가 정당하다고 하더라도 주어진 목적에 적합한 평가도구가 사용되어서 공정한 결정이 내려지고 있는가는 별개의 문제이다.

예를 들어 사법시험에 토플 성적을 고려하는 정책의 경우 법조인들이 실제로 필요로하는 영어 실력을 측정하기 위해 만들어진 시험도 아닌 토플이나 토익 같은 시험 때문에 법조인이 되는 꿈을 접는 사람들이 있다는 사실은 문제가 있어 보인다. 이렇게 시험을 포기하거나 영어 시험 때문에 시험에 실패한 사람들에게 토플시험 성적 사용이 미래 법조인을 선발하는 적합한 준거가 될 수 있다는 근거를 제시할 수 있을 때 새로 도입된 규정이 정당화 될 수 있을 것이다.

Ⅲ. 논의 및 결론

본 논문에서는 한국 사회에서 많이 사용되고 있는 외국어 평가의 유용성을 평가도구들이 개발된 목적과 실제로 사용되는 목적 사이의 일치를 중심으로 살펴보았다. 신문기사와 인터넷 자료를 분석해 본 결과 각종 평가도구들이 원래 개발된 것과는 다른 목적으로 많이 사용되고 있다는 사실을 알 수 있었다. 외국어 평가결과를 해석할 때 던져야 할 가장 중요한 질문은 시험 결과에 근거해서 내려진 결정이 적합하다고 어느 정도 확신할 수 있는가의 여부라고 볼 수 있다. 본 논문에서 소개된 것과 같은 시험의 부적절한 사용은 시험 결과에 근거해서 내려지는 많은 결정의 정당성에 의문을 제기한다. 예를 들어 토플시험 성적이 사법 고시의 한 전형요소로 사용되는 경우 특정 시험 점수를 받은 사람이 그 점수를 받지 못한 사람보다 더 훌륭한 법조인이 될 수 있다는 '해석상의 주장' (interpretive claims) (Kane, 1992)을 하는 셈인데 안타깝게도 토플시험이 원래 개발되어진 목적과 시험 문항을 고려해 볼 때 토플시험 성적을 기준으로 이러한 해석을 내릴 수 있는 근거는 별로 없어 보인다.

물론 본 논문에서 다룬 내용은 아니지만 이런 잘못된 외국어 평가도구의 사용은 몇가지 잘못된 오해에서 기인하는 것 같다. 많이 알려진 외국 평가기관에서 만들어진 시험의 유용성에 대한 지나친 과신이 그 하나이다. 전세계적으로 사용되는 표준화된 검사이므로 공정하고 믿을 수 있는 시험이고 따라서 다양한 목적으로 사용해도 괜찮다고 믿는 경향이 있다. 때로는 어쩔 수 없는 상황에서 이런 시험들이 최선의 선택이라는 주장도 있다. 즉 평가의 필요성은 있지만 주어진 평가 목적에 적합한 평가도구를 찾기 힘든 경우도 있고 기존의 평가도구의 질이 워낙 낮아서 차라리 잘 알려진 평가도구를 선택하는 것이 더 공정하다는 견해도 있을 수 있다. 그러나 그렇다고 해서 그들이 평가결과에 근거해서 내리는 주장에 정당성이 부여되지 않는다는 점을 간과해서는 안될 것이다. 또 다른 오해는 모든 직종에서 영어 실력이 필요하

다는 검증되지 않은 믿음이다. 평가의 목적과 관련해서 볼 때 단순히 지원자들을 걸러내기 위해 영어 평가를 사용하는 것은 아닌지 고민해 보아야 할 것이다. 정말 주어진 상황에서 외국어 실력이 필요한지 먼저 알아보아야 하고, 필요가 있는 경우에는 어떤 외국어 능력이 요구되는지에 대한 연구가 선행될 때 더 공정한 평가가 이루어질 수 있을 것이다.

지금까지의 논의를 종합해보면, 우리 사회에서 외국어 시험과 관련하여 제일 먼저 던져야 할 질문은 주어진 상황에서 외국어 시험이 정말로 필요한가의 여부이고 만약 이 질문의 답이 '예'인 경우 다음으로 던져야 할 질문은 측정하고자 하는 언어 능력이 무엇인가일 것이다. 이 두 질문에 대한 답이 명확해야만 목표에 부합되는 평가도구를 제작하거나 선택할 수 있을 것이다. 만약 적합한 시험이 없을 경우 시험을 안보는 것보다는 그래도 이런 시험이라도 보는게 더 좋다는 'something is better than nothing' 식의 쉬운 해결책을 선택해서 적절하지 않은 시험을 잘못 사용하기 보다는 적합한 평가도구를 제작해서 사용해야 할 것이다. Spolsky(1995)가 제약회사들이 약 병에 약 복용방법을 표시하는 것 처럼 언어 평가도구에도 평가 사용방법을 표시해야 한다고 지적했던 것처럼 언어 시험의 오남용을 막는 노력이 더 많이 이루어져야 할 것이다.

참 고 문 헌

Alderson, J. C., and Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115-129.

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

Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing*, 17, 1-42.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in*

practice. Oxford: Oxford University Press.

Chapelle, C., Jamieson, J, & Hegelheimer, V. (2003). Validation of a web-based ESL test. *Language Testing, 20*, 409-439.

Educational Testing Service. (2004). *Test of English as a Foreign Language*. Princeton, NJ: Educational Testing Service.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*, 527-535.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. New York: American Council on Education.

Shepard, L. (1993). Evaluating test validity. *Review of Research in Education, 19*, 405-450.

Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.