

TEPS 20년의 역사 및 그 전사

이용원¹ · 전희성[†]

¹서울대학교

A Twenty-Year History of the TEPS and Its Prehistory

Yong-Won Lee¹ and Heesung Jun^{1†}

¹Seoul National University

ABSTRACT

In commemoration of its 20th anniversary, we look back on the brief history of the TEPS as well as the social and historical contexts that prompted research efforts to develop the TEPS in this article. Towards this goal, we first review, as the prehistory of the TEPS, the tradition of foreign language testing at the Language Education Institute (LEI) of Seoul National University. Second, we survey a series of research projects undertaken to develop the TEPS. Third, we examine several milestones in the history of the TEPS and its family of tests, which symbolize important milestones of further growth, expansion, and transformation in various aspects of test development, production, and administration. Fourth, we discuss major research studies conducted to improve and validate the TEPS. Finally, we conclude the article with future prospects of language assessment and TEPS in Korea.

Keywords: history of language testing, TEPS history, SNULT, Language Research Institute, Language Education Institute

1. 서 론

현대의 언어 평가(혹은 외국어 평가) 발전과정은 언어학 이론의 발전뿐만 아니라 언어교육 및 심리측정 역사와 긴밀히 맞물려 발전해왔다. Spolsky(1977, 1978)는 일찍이 이러한 관련 학문 간의 연계성을 감안하여 현대언어평가의 역사를 ‘과학 이전의 시기’(the pre-scientific period), ‘심리측정 및 구조주의의 시기’(the psychometric-structural period), 그리고 ‘심리언어학 및 사회언어학적 시기’(the psycholinguistic-sociolinguistic period)의 3개의 시기로 구분하였다. Spolsky(2008) 자신을 포함해서 언어 평가 연구자들 사이에서는 최근 이러한 3단계 시기 구분이 아직도 얼마나 유용한가에 의문이 제기되고 있기도 하다. Lee(2008)의 경우에는 이러한 Spolsky(1977)의 시기 구분 방식에 Weir(1990)가 제안한 ‘의사소통적 시

[†] Corresponding author: hsjun@snu.ac.kr



기'(the communicative period)를 추가하고, '심리언어학 및 사회언어학적 시기'를 '심리측정 및 준의사소통적 시기'(the psychometric & semi-communicative period)로 명칭을 변경하여, 총 4개의 시기로 확대한 시대 구분 방식을 제안하였다. 여기에 한 발 더 나아가 이러한 새로운 틀을 통해 한국의 현대영어평가의 발전단계를 구분하고 각 시기에 개발된 영어시험의 주요 특징들을 분석하려고 시도한 바 있다.

물론 현대 이전에도 언어 평가는 장구한 역사를 가지고 있다. 현재 기록으로 남아 있는 가장 오래된 언어평가 시험의 예로는 기독교 성경인 구약의 사사기(Judges; BC 1473-1117년경) 12장에 등장하는 히브리어 단어 발음 시험인 쉬볼렛 시험(Shibboleth test)이다. 그 당시 이스라엘 지역에서 길르앗인(Gileadites)과 에브라임인(Ephraimites) 간에 전쟁이 있었고 전쟁에 패한 에브라임인들이 도망을 하자, 길르앗인들이 요단강의 수심이 낮아 도강이 가능한 길목을 미리 선점한 후, 도강하려 하는 자들 중에 에브라임인을 색출하기 위해 쉬볼렛(shibboleth: 시냇물)이란 단어를 발음을 해보도록 시켰다. 만약 이를 길르앗인들이 발음하는 것처럼 '쉬볼렛'으로 정확히 발음하면 살려주었지만, 에브라임인들이 발음하는 것처럼 '시볼렛'으로 발음하면 모두 죽여버렸는데, 실제로 현장에서 살해된 자들 중에는 가해자의 동족인 길르앗인들도 상당 수 있었다고 한다. 이는 극단적인 언어평가의 예이기는 하지만 오늘날에도 언어능력 평가 시험의 오남용의 예와 함께 언어평가의 사회적인 책무나 윤리적인 지침들을 논의할 때 종종 언급되기도 한다(McNamara & Roever, 2006; Park & Shin, 2015).

현대 이전의 좀 더 본격적이고 체계적인 언어평가 시험의 예들은 일찍이 중국과 동아시아의 고대 국가들에 도입되었던 과거제도(科擧制度)에서 찾을 수 있을 것이다. 과거제도는 처음 도입된 것은 수나라(581-618)가 통일 왕조를 이룬한 얼마 되지 않은 시기인 587년이다. 중국을 통일한 수문제 통치 시기 지역 호족과 군벌을 약화시키고 유능한 국가 관료를 선발하여 중앙집권적 왕권을 강화하려는 목적을 가지고 도입하였다. 우리나라의 과거 제도 역시 유사한 목적을 달성하기 위해 고려 초인 광종 9년(958)에 중국으로부터 도입하여 시행하였다. 이러한 과거제도의 핵심적인 일부인 문관(文官) 시험에는 중국의 고전이나 유교 경전에 대한 지식과 이해도에 대한 평가(明經業, 生員試)뿐만 아니라, 한자에 기반한 시(詩)·부(賦)·책(策) 등의 문학적·논술적 글짓기 능력의 평가(製述業, 進士試)가 포함되었다. 특히 조선의 과거제도는 왕조 초기부터 문과, 무과, 잡과라는 3 대 주요 시험 영역으로 정착되어 1 차 갑오개혁 이전까지 시행되었는데, 초시를 통과한 문관들이 본인이 중요한 국가 정책에 관하여 논술한 글(策)을 가지고 왕에 앞에 나아가 질문에 대답하는 책문(策問)이라는 면접 구술 시험도 있었다. 과거 시험의 여러 종류와 분야 중 오늘 날의 외국어 능력 평가 시험과 직접적으로 연계 될 수 있는 시험은, 국가 간 외교 활동 과정에 주로 통역, 교섭, 통상 업무를 담당했던 역관(譯官)의 선발 시험인 역과(譯科) 시험을 들 수 있다. 조선시대 역관의 양성은 주로 국가기관인 사역원(司譯院)에서 담당했고 역관 지망자들은 잡과의 일부인 역과(譯科) 시험에 응시해 합격함으로써 역관으로 선발되어 중하위 기술직에 진출할 수 있었다(Lee, 2004).

1999년부터 지난 20년간 시행되어온 TEPS(Test of English Proficiency developed by Seoul National University)는 주로 영어를 외국어나 제2언어로 배운 한국인 영어학습자의 영어능력을 평가하는 표준화된 지필 영어능력시험이다.¹⁾ TEPS 시험은 의사소통 능력과 의사소

통적 영어학습 및 교수법이 각광을 받던 시기인 1990년대 초부터 본격으로 개발 연구가 시작되었고, 시험의 설계와 개발에 의사소통적 언어평가의 틀을 활용하였으며, 당시 최신의 심리측정학 이론 중의 하나였던 문항반응이론(item response theory)을 고전적 이론(classical test theory)과 함께 문항 분석, 채점, 척도 설정, 점수 산출에 활용하는 방식을 채택하였다. 이러한 전통은 지금까지 계속 이어져 내려오고 있고, 그동안 컴퓨터 기반 영어능력 평가 시험(i-TEPS)과 컴퓨터 기반 말하기 쓰기 시험(TEPS Speaking & Writing)도 개발하여 시행하고 있다. 좀 더 긴 역사적 안목으로 보면, 이러한 TEPS 시험과 그 자매 시험들은 1990년대 말-2000년 대 초 갑자기 출현한 시험이라기보다는 그보다 훨씬 이전 시기(1967)부터 어학연구소(언어교육원의 전신)에서 시행해온 외국어 능력 평가 시험인 SNULT(Seoul National University Language Test) 외국어 시험 제작과 시행의 축적된 경험과 성과의 토대 위에서 탄생했다고 볼 수도 있을 것이다.²⁾

앞에서 언급된 현대 언어 평가 발전의 국내외적 시대 구분의 틀을 통해 볼 때에도, 능숙도 혹은 숙달도 시험(proficiency test)의 성격을 가지고 있는 TEPS의 20년 시행 역사와 그 이전 수년 간의 연구개발 및 사전 시행 기간 그리고 SNULT로 대표되는 30여년의 그 전사(前史)는 매우 흥미로운 역사 분석의 자료를 제공해주고 있다. 이는 우리 나라의 현대적인 언어능력 평가 방법이 어떻게 진화하고 발전해 왔는지를 보여 주고 있을뿐만 아니라, 그러한 발전 과정들이 어떻게 관련 분야의 국내외 추세들과 연동되어 발전해 왔는지에 대한 중요한 통찰을 제공해주고 있다. 이러한 측면에서 볼 때 지난 30여 년의 TEPS 연구개발과 시행의 역사 그리고 그 이전으로 더 거슬러 올라가 지난 반세기 동안에 언어교육원에서 이루어진 TEPS, i-TEPS, TEPS Speaking & Writing, SNULT를 포함하는 외국어능력 평가의 역사는 우리 나라 현대 언어평가의 발전사를 논할 때 빼 놓을 수 없는 귀중한 일부라고 할 수 있다. 뿐만 아니라 TEPS는 동아시아 지역의 언어 평가 및 영어 숙달도 평가의 역사와 발전과정을 논하는 데 있어서도 매우 중요한 자료를 제공해 주고 있다. 동아시아 역내의 국가들에는 TEPS와 유사하게 지역 내에서 자체적으로 개발되어 지역화된 특성을 가지면서도 TOEFL, TOEIC, IELTS 등 영어권 국가에서 개발되어 전세계적으로 사용되는 글로벌한 대단위 표준화 영어시험들과 지역 내에서 어깨를 나란히 하고 함께 사용되고 있는 여러 영어 시험들이 있다(예: 대만의 GEPT, 일본의 EIKEN, 중국의 CET).

본 논문에서는 1999년 TEPS 출시 이후의 20년간의 기존 TEPS 시험의 성장, 확대, 개선 및 변화 과정이나 다른 새로운 시험의 개발, 제작, 시행 및 안정화 과정뿐만 아니라 TEPS 출시 이전의 수년간의 TEPS 연구개발 과정 그리고 시간을 거슬러 더 올라가 영어를 포함하는

- 1) 한국인 영어학습자의 영어능력 평가라는 목표를 가지고 개발된 TEPS는 영어학습자가 얼마나 국제적인 무대에서 효과적으로 의사소통하는가를 평가하는 국제적인(international) 시험의 성격뿐만 아니라 한국인 영어학습자들의 영어학습의 어려움을 잘 분석해주고 그들의 필요를 잘 충족시켜 주는 지역화된(localized) 영어능력시험의 성격을 모두 갖고 있음을 의미할 수 있다.
- 2) SNULT 시험은 일찍이 1967년부터 서울대학교 언어교육원의 전신인 어학연구소에서 시행되었는데 그 당시에 영어가 중심적 언어 중의 하나였다. 이 SNULT 영어시험은 2016년까지 시행되다가, TEPS 시험과 중복적인 성격이 있고 응시자 규모도 크지 않았기때문에 2017년 중지되었다. 그러나 일본어, 중국어, 독일어, 불어, 러시아, 스페인어 시험은 계속 유지되고 있다.

SNULT 시험의 시작과 발전과정을 역사적 관점에서 탐구해 보고 그 주요 성과들을 정리 해보고자 한다. 아울러 TEPS 시험 개발, 개선, 개정 연구와 타당도 검증 연구들을 간략히 검토하고 요약함으로써 앞으로 TEPS를 포함한 외국어 평가 시험의 지속적인 발전과 개선을 위해 절실히 요구되는 미래의 연구 개발 주요 분야를 조망하는 기반을 마련하고자 한다.

2. TEPS 이전의 외국어 평가 역사

이미 앞에서 논의 한 것처럼 TEPS 시험은 1990년대에 갑자기 개발되어 출시된 시험으로 보기 보다는 그보다 훨씬 이전인 서울대학교 어학연구소 출범 초기부터 축적해 온 외국어 능력 평가 시험 시행 경험과 전문성, 연구 개발 역량, 대외적인 신뢰의 토양 위에서 탄생한 시험으로 보는 것이 합리적일 것이다. 실제로 서울대학교 어학연구소는 출범 초기인 1960년대부터 한국의 여러 정부기관과 기업이 요청이 있을 때 소규모로 외국어시험을 시행하였는데, 물론 영어가 중심적인 평가 대상 언어였고, 초기에는 외국의 시험을 구매해서 시행하기도 하였다(박남식, 대면 인터뷰, 2018년 11월 28일, 본 특별호 녹취록 참조). 그 이후에는 언어교육원에도 자체로 시험 문항을 출제하고 시험을 제작해서 실시하게 되었는데 이러한 경험들이 축적되면서 결국 SNULT 외국어 시험으로 결실을 맺게 된다.

실제 어학연구소의 SNULT 시험 시행의 역사는 1967년까지 거슬러 올라가는데, SNULT 시험에는 영어뿐만 아니라 프랑스어, 독일어, 스페인어, 일본어, 중국어, 러시아어와 같은 언어들을 일정한 시차를 두고 순차적으로 포함하게 되었다. 그 당시에는 외교부를 포함한 주로 외부 기관들의 요청에 따라 해외에 학업 또는 업무로 파견되는 한국 학생들과 공무원 및 직원들을 평가하기 위해 제작하고 시행하였다(박남식, 2018). 이렇게 비정기적 특별 시험 형태로 시행되던 SNULT 시험은 2008년 8월부터는 일반인을 대상으로 한 정기시험 형태로 전환되어 지금까지 시행되어 오고 있다. 1999년 이후에는 언어교육원에서 TEPS와 i-TEPS 같은 영어시험이 새로이 개발되어 시행되고 있었기에, SNULT와 이 시험들의 중복성 문제가 제기 되었고, 수험자의 숫자도 그리 많지 않아, SNULT의 영어시험은 2017년 폐지되었다. 그러나 일본어, 중국어, 독어, 불어, 러시아, 스페인어 시험은 계속 유지되고 있다.

SNULT 영어시험은, 2017년 시행을 중지하기까지, 지난 40여 년간 특별시험과 정기시험 형태로 SNULT를 구성하고 있는 다른 외국어 시험들과 함께 시행되었다. 어떤 의미에서 보면 이러한 SNULT 시험은 서울대학교 언어교육원(어학연구소) 외국어 평가의 전통을 상징할뿐만 아니라 1999년 출시되어 오늘날까지 시행되고 있는 TEPS 시험과 그 자매시험들의 탄생과 정착에도 중요한 토대가 되었다고 할 수 있다.

3. TEPS 시험의 연구와 개발

TEPS는 국제적인(international) 시험과 지역화된(localized) 시험의 특성을 둘 다 가지고 있다고 보는 것이 합당하다. 기본적으로 TEPS는 한국의 초·중·고교 교육과정이나 대학영어 교육과정을 직접 반영하는 성취 평가(achievement test)의 성격을 띠기 보다는, 영어학습자가 국제적으로 얼마나 효과적으로 의사소통이 가능한지를 평가하는 영어 숙달도 시험(proficiency test)으로서의 특성을 더 강하게 가지고 있다. 이러한 점은 영어권에서 제작되어 전세계적으로 사용되고 있는 TOEFL이나 TOEIC 같은 다른 국제적인 영어 숙달도 시험과 비교해 불만한 특징이다. 동시에 TEPS는 한국인 영어학습자의 일반적인 영어학습의 어려움을 잘 분석해주고 한국적 상황에서 그들의 필요를 잘 충족시켜 주는 지역화된 영어능력시험의 특성을 함께 염두에 두고 개발되었다(Park, 2018, Kim & Moon, 2018).

TEPS 개발을 위한 연구는 1990년 초에 어학연구소에 시작된다. 그 당시의 국내 정치적, 사회적, 교육적 상황을 살펴보는 것이 TEPS 개발의 전반적 배경을 이해하는 데 도움이 될 듯 하다. 대통령 직선제를 통한 평화적인 정권교체와 1988년 서울 올림픽의 성공적인 개최로 인해 민주화와 국제화의 분위기가 활발하게 진작되던 시기였다. 1989년 정부는 국민들의 해외 여행 자유화 조치를 시행했고, 외교적인 면에서 보면 소련(1990), 중국(1992), 그리고 여러 동구권 공산국가들과의 외교 관계를 정상화하고 남북이 동시에 유엔이 가입하는(1991) 변화가 일던 격동의 시기였다. 1992년 대통령 선거를 통해 당선된 김영삼 대통령은 선거시기부터 소위 ‘세계화 정책’에 대한 공약을 발표하고 실제로 임기 중에는 선진국들의 국제 조직인 OECD(Organization for Economic Co-operation and Development)에도 가입하게 된다.

이러한 세계화에 대한 강한 열망과 정책이 실현되고 있던 1990년대 초에 서울대학교 어학연구소에서 학생들의 전반적인 영어 실력을 촉진하는데 적합하고 영어 숙달도 시험으로서의 타당도도 갖춘 시험의 개발 연구가 시작되었다(최인철, 서면 인터뷰, 2019년 1월, 본 특별호 인터뷰 전문 참조). 그 당시에도 이미 TOEFL이나 TOEIC 같은 미국의 평가기관에서 개발된 영어시험이 국내에서도 주로 사용되고 있었다. TOEFL의 경우에는 미국이나 캐나다를 포함한 북미나 유럽의 영어권 대학에 유학하기 위해서는 거의 필수적으로 보아야 되는 시험이었지만, TOEIC의 경우에는 그렇지 않음에도 국내의 대학과 기업에서 많이 사용하고 있었다. 그 당시에도 두 시험 모두 수험료가 비싼 편이었고 여러 번 시험에 응시해야 했기 때문에 국부 유출에 대한 논란이 제기되었다. 아울러 국내에도 이러한 대단위 영어능력 평가 시험을 자체 개발하고 시행하는 역량을 갖추어 언어평가시험의 주권을 지키는 것이 필요하다는 생각들이 영어관련 학자들과 연구자들 사이에서 논의되기도 했다(Park, 2018).

현재의 TEPS 시험은 그 개발 초기부터 TEPS라는 명칭으로 불리지는 않았다. 당시 개발 중이던 시험은 잠정적으로 ‘서울대학교 절대평가 영어 능숙도 시험’이라는 의미를 가진 SNUCREPT(Seoul National University Criterion Referenced English Proficiency Test)라는 영문 명칭이 주로 사용되었다(Choi, 1993, 1995)였다. 시험 개발 프로젝트가 진행될 당시 연구팀은 어학연구소의 연구원들과 서울대학교 관련 학과의 교수들로 이루어졌다.

연구의 첫 단계로 한국 상황에서 대규모 영어능력평가 시험을 제작하여 시행하는 계획의

실현가능성을 점검하기 위한 노력이 있었다. 먼저 주요 대기업의 인사 담당자들과 대학 교수들을 대상으로 요구 분석(needs analysis)을 시행하였다. 이를 통해 우리 나라에서 대학 졸업 후 필요한 영어 의사소통능력의 수준과 실체를 파악해 보려고 시도하였고, 기존 영어시험의 미진한 점을 보완할 수 있는 보다 타당한 영어시험을 개발하기 위해 목표언어사용(target language use; TLU)의 상황과 과제를 정의하고 규정하는 실제적 틀을 마련하였다. 이러한 요구 분석 결과, 국내의 영어교육 분야의 변화 추세, 그리고 최신 언어평가 이론을 잘 종합하고 접목하여 전체 시험의 평가틀(문항의 수, 구조, 유형, 시험시간 등)을 마련하였고, 기존 영어시험에는 없던 새로운 시험 방식(청해 시험 방식, 문법 및 어휘 시험 방식 등)도 도입하였으며, 십여 차례의 파일럿 시험을 개발하여 양적, 질적 연구 방법을 사용하여 타당성 검증 연구를 수행하였다. 또한 여러 전문가로부터 시험의 구성 및 내용에 대해서 전반적인 피드백과 자문을 받았고, 이러한 결과를 바탕으로 평가틀을 세밀히 조정하는 과정을 거쳤다.

1995년부터는 언어교육원 내·외부 출제자들이 제출한 많은 양의 문항을 입력하고 편집하고 보관할 수 있는 문항 데이터베이스가 구축되었다. 그리고 시험의 안정성을 확보하면서도 보다 높은 측정 정확성과 변별력을 유지하기 위하여 문항반응이론(item response theory, IRT)을 도입하였는데, 당시 수험자의 채점된 문항 반응 데이터 분석에 문항반응이론을 실제 적용한 선행연구가 많지 않아서 측정 프로그램을 직접 개발하는 등의 어려움도 겪기도 하였다(최인철, 2019). 하지만 여러 차례의 파일럿 시험 자료의 분석과 시행착오 과정을 거쳐 완성도를 높였다.

그러다가 1990년대 말 갑자기 발생한 국가적 사태가 당시 개발 중이었던 TEPS 시험의 출시에도 일정 정도 영향을 주게 된다. SNUCREPT 시험 개발을 위한 연구가 한창 진행되던 중인 1997-1998년 시기 국내에서는 통칭 ‘IMF 사태’라 불리는 외환 위기가 발생하게 된다. 이러한 국가적 위기 상황은 TOEIC을 대체하여 외화의 해외 유출을 줄일 수 있는 독자적인 국산 시험 개발의 필요성을 더욱 증대시켰다. 이러한 사회적 요구에 부응하여 서울대학교 어학연구소는 1998년 10월에 연구 중이던 SNUCREPT의 이름을 TEPS로 개명하고 이듬해부터 본격적으로 시행하게 되었다.

이렇게 어학연구소에서 수 년간 지속적으로 이루어진 연구개발 프로젝트의 결과, TEPS는 1999년 1월 31일에 처음 시행되었다. 시험이 출시된 후 많은 정부 기관과 공기업 그 외의 민간 기관들이 TEPS 점수를 사용하기 시작하였다. 교육부, 외교부, 국방부, 경찰청, 육군본부 등의 기관과 함께 다수의 고등학교, 대학교, 대기업 등이 TEPS 점수를 활용하였다. 또한 어학연구소는 2001년 9월 지금의 언어교육원으로 명칭이 변경되었는데, 이는 어학 연구뿐만 아닌 언어 교육 및 평가를 아우르는 기관의 확장된 사명과 성격을 더 잘 반영하기 위함이었다.

4. TEPS의 출시, 성장과정, 그리고 새로운 도전

그동안 TEPS 사업은 이원화된 체제로 운영되고 있다. 서울대학교 언어교육원에 소속된 TEPS센터는 정기 및 특별 시험의 문항 출제 및 편집, 시험세트(검사형) 구성, 수험자 응답 채점 및 시험 데이터의 심리측정학적 분석을 담당하고 있다. 현재 정기 TEPS는 연 18회 시행하고

있는데, 요청하는 기관이 있을 경우 추가적인 특별 시험도 시행한다. 한편 재단법인 서울대학교발 전기금 소속의 TEPS관리위원회(사업본부)는 시험 등록, 시험 시행, 성적 공지 및 성적표 발송을 담당하고 있다.

TEPS가 처음 시행된 1999년 1월 이후 12년간 TEPS 사업은 급속한 성장과 확장을 이루었다. 1년 후인 2000년 1월부터 온라인 등록 시스템이 사용되었고, 곧이어 2003년에 민간자격 국가공인을 획득하였다(교육인적자원부 제2003-2호). 시험 출시 후 정기시험 횟수는 첫 해 연 6회에서 2005년에 12회로, 그리고 또다시 2010년에 16회로 증가하였으며, 연간 수험자 수도 첫째의 약 2만 7천 명에서 2010년에는 약 40만 이상에 도달할 정도로 급격히 증가하였다. 해외에서 개발한 영어능력시험인 TOEIC과 TOEFL 등이 이미 우세했던 한국의 상황에서 TEPS가 이런 이러한 업적은 놀랄만한 성과였다. 영어교육이 각광 받던 시대적 상황, 서울대 언어교육원이 외국어 평가 시행으로 수십 년간 쌓아온 전문적인 경험 및 대외적인 신뢰, 비교적 저렴한 수험료, 영어 고능력자 집단에서의 시험의 뛰어난 변별력 등의 여러 요인들이 이 시기 TEPS 성장에 크게 기여했다고도 볼 수 있다. 언어교육원은 지난 수년 간 한국의 수험자, 영어교육 및 평가 전문가, 그리고 기타 주요 이해당사자(stakeholders)의 요구와 의견을 바탕으로 개정 작업을 진행하여 개정된 TEPS(New TEPS)를 2018년 5월 첫 시행하였고, 2019년부터는 정기시험 횟수를 16회에서 2회 증가시켜 연 18회 시행하고 있다.

한편 이 기간 동안에는 다른 시험들도 개발되어 출시되었다. 우선 1999년에는 영어 말하기 시험인 TOP(Test of Oral Proficiency)가, 2000년에는 영어 쓰기 시험인 TWP(Test of Written Proficiency)가 출시되었다. TOP과 TWP는 2008년에 TEPS-Speaking과 TEPS-Writing으로 각각 명칭이 변경되어 정기적으로 시행되어 왔다. 이 두 시험 모두 컴퓨터 기반 수행평가 형태를 띠고 있는데, 말하기 시험에는 총 6 개의 서답형(constructed response) 문항이 포함되어 있고, 쓰기 시험에는 4 개의 서답형 문항이 포함되어 있다. 뒤이어 4개 영역(듣기, 읽기, 말하기, 쓰기)을 평가하는 컴퓨터 기반 영어능력 시험인 i-TEPS가 2009년에 출시되었다. 기존 TEPS의 사지선다형 문항으로 구성된 4 개영역(문법, 어휘, 청해, 독해)의 문항 수를 줄이고 거기에 말하기 쓰기 영역을 추가한 형태를 취하고 있으며, 지필 시험인 TEPS와 달리 컴퓨터 기반 시험이다. 이 세 시험의 수험자 수는 TEPS와 비교해서는 다소 적었지만, TEPS 시험이 직접적으로 측정하지 못하는 말하기, 쓰기 능력을 추가적으로 평가하고, 컴퓨터 기반 영어 시험이라는 점에서 그 나름의 중요한 의미를 가지고 있다.

5. TEPS 관련 연구³⁾

이 장에서는 TEPS 점수의 해석 및 사용을 위한 주요 타당도 연구 과제를 세 시기로 나누어 요약하여 TEPS 연구의 역사를 보여주고자 한다. 그동안 학술지 논문으로도 TEPS의 타당도를 뒷받침할 연구 결과가 공개되나 있다(Choi & Lee, 2001a; Choi, 1993, 1994, 1997, 1999

3) 5장의 내용은 Lee, Jun, & Kim (in press)의 내용 일부를 재구성하였다.

등). 그러나 지난 20여년간 언어교육원에서 이루어진 대다수의 연구 과제의 결과는 내부 연구보고서의 형태로 집필되어 보관되고 있다. 1991년 이후 언어교육원 내에서 25개가 넘는 연구 프로젝트가 TEPS를 개발하거나, 타당도를 검증하거나, 개정 가능성을 모색하기 위한 목적으로 진행되었다.

5.1. TEPS를 개발하기까지의 연구

TEPS 연구의 첫 번째 시기는 1991년부터 1998년까지로 TEPS를 개발하기까지의 기간이다. 이 기간 동안 요구 분석, 시험 개발, 구인 타당도, 시험 방법 등에 대한 연구 과제가 수행되었다. 어학연구소를 중심으로 사실상 국내에서 최고의 영어 평가 및 교육 분야의 전문가들의 역량이 집결된 연구팀이 꾸려져, 국내 영어 숙달도 시험 개발의 필요성을 적극적으로 검토하게 되었고, 실제 이러한 목적으로 파일럿 시험과 필드 시험을 제작해 시행해보고, 그 시험 점수의 타당도 검증 연구를 수행하기도 하였다. 이러한 일련의 부단한 노력의 결과로 TEPS가 개발되었다. 첫 번째 시기의 주요 내부 연구 과제는 <표 1>에 요약되어 있다.

표 1. TEPS를 개발하기까지의 연구

연구 과제	연구 내용	결과
Park et al. 1992	794명을 대상으로 한 설문조사의 결과, 직장에서 쓰이는 실용적인 영어를 평가 할 영어능력시험의 필요성이 대두됨	실용적이고 일반적인 영어능력시험의 개발이 시작됨.
Choi, 1993; Choi, 1994;	두 개의 파일럿 시험이 개발되어 각각 738명과 314명의 한국 대학생을 대상으로 시행됨.	첫 번째 파일럿 시험 후 변경사항을 반영한 두 번째 파일럿 시험을 제작함. 문제은행 가동의 시행 가능성을 탐색하기 위한 후속 연구가 진행됨.
Choi, 1997	408명의 한국 대학생을 대상으로 파일럿 시험을 시행하여 시험의 대규모 시행을 위한 문제은행 가동의 가능성을 탐색함.	문항 및 검사 통계를 바탕으로 시험을 수정함.

시험 개발의 초기 단계에서 가장 먼저 출판된 논문은 Choi(1993)와 Choi(1994)로, TEPS의 전신인 SNUCREPT 시험의 내용과 구인 타당도(construct validity)를 논의하였다. 시험 구조, 시험 내용, 시험 방식에 대한 설명과 함께 파일럿 시험의 결과가 제시되었는데, 4개 영역의 알파 계수가 모두 0.80 이상으로 만족스러운 수준의 신뢰도를 나타내었다.⁴⁾ 뒤이어 SNUCREPT에 대한 두 편의 외부 논평이 출판되었다. Oller(1995)와 Bachman(1996)은 SNUCREPT에 대해 대체적으로 긍정적인 평가를 하였다. 특히 Oller는 시험이 가지고 있는

4) 시험의 점수 신뢰도 수치가 어느 정도가 만족스러운 수준이냐의 문제에는 학자마다 다소 다른 기준을 제시할 수 있다. 하지만 알파 신뢰도 계수가 대략 0.70-0.80 사이이면 용인될 수 있는 수준으로 보고 있고 0.80-0.90 아주 좋은 수준으로, 0.90이상이면 아주 탁월한 수준이라 할 수 있다.

속도성(speededness) 특성을 지지하였는데, 언어 처리는 보통 시간 제약이 높은 상황에서 이루어 진다는 사실을 반영하기 때문이라고 하였다. Bachman은 SNUCREPT가 잘 설계되었고 상세한 관련 연구 결과에 근거해 개발된 영어능력 시험이고 응용언어학과 심리측정학 이론에 온전히 근거한 시험이라고 평가하였다. 마지막으로 Choi(1997)는 SNUCREPT의 시험 방식 요소들에 대한 수험자의 인식을 조사하였다. 시험 방식의 특징으로는 (1) 어휘와 문법 영역에서 말하기와 쓰기에서 쓰이는 어휘와 문법 사항 구분, (2) 청해 영역에서 시험 내용을 음성으로만 제시, (3) 청해 영역 일부 파트에서 지문과 질문을 두 번씩 제시하여 거시적 듣기(macro-listening)와 미시적 듣기(micro-listening)라는 2단계 과정을 가능케 함, (4) 청해와 독해 영역에서 1지문 1문항 원칙을 유지하여 국부의존성(local item dependence) 효과를 최소화 함, 그리고 (5) 제한된 시간 내에 많은 수의 문항을 제시하는 속도성(speededness)의 특성이 있다. SNUCREPT의 파일럿 시험을 치른 314명의 수험자는 이러한 시험 방식 요소들이 대체로 타당하고 바람직하다고 인식했다. 이러한 수험자 인식에 관한 설문조사의 결과는 위에 언급된 시험 특성들을 왜 TEPS에 반영하고 유지해야 하는가의 실질적 근거를 제공하였다.

5.2. TEPS의 안정화와 유지를 위한 연구

TEPS 연구의 두 번째 시기는 TEPS가 첫 시행된 1999년 이후부터 2010년경까지이다. 이 시기에는 TEPS를 안정화시키고 계속 발전시키고 유지하기 위한 목적의 연구가 수행되었다. 두 번째 시기의 전반에는 문항 출제(Kim & Lee, 2004; Kim et al., 2001, Lee, 1999), 대화와 지문 내용의 민감성(Ahn, 2002; Ahn, 2004), 문항 난이도 예측(Lee, Kim, & Park, 2003; Park, 2004), 문항 분석(Choi, 2004), 그리고 다양한 측정 관련 문제들(Park, 2000; Choi & Park, 2001; Choi & Lee, 2002; Hwang, 2001)에 관한 연구가 이루어졌다. 두 번째 시기의 후반에는 기존 문항 출제 절차의 개선(Cho et al., 2010), 시험 검토 과정 개선(Lee, 2009; Lee et al., 2010; Kim & Kim, 2009), 그리고 문항 데이터베이스 관리(Sohn et al., 2010) 등에 관한 연구를 통해 더욱 효율적으로 TEPS를 제작할 수 있는 방안을 마련하였다. 두 번째 시기의 주요 연구 과제는 <표 2>에 요약되어 있다.

표 2. TEPS의 안정화와 유지를 위한 연구

연구 과제	연구 내용	제안점/결과
Lee, 1999	TEPS 첫 6개 검사형에 포함된 문법과 어휘 문항을 문법적, 어휘적 특성에 따라 분석하여 비율을 조사함.	구어적이고 관용적 표현이 수험자의 TEPS 난이도에 대한 인식에 영향을 미치므로 개수를 조절할 필요가 있음.
Kim et al., 2001	TEPS 문항 출제를 위한 종합적인 안내서를 작성함.	출제자들이 안내서에 기반하여 문항 출제를 함.
Park, 2000; Choi & Park, 2001; Choi & Lee, 2001b, 2002; Hwang, 2001	채점과 관련된 심리측정학적 문제들이 연구됨.	연구 결과에 따라 채점 절차를 지속적으로 개선함. 시험 시간이 유지됨.

Lee, Kim & Park, 2003	독해 문항에 대한 연구원의 예상 난이도와 실제 난이도 간의 상관관계가 이독성(readability) 지수와 실제 난이도 간의 상관관계보다 높음.	이독성 지수가 독해 문항의 난이도를 예측할 수 있는 유일한 변수가 될 수는 없다. 연구원의 경험에 의거한 주관적 판단도 중요한 변수가 될 수 있음.
Ahn, 2002; Ahn, 2004	대화와 지문에서 피해야 하는 민감한 주제에 대한 지침이 작성됨.	지침이 TEPS 출제자 가이드와 검토자 가이드에 추가됨.
Choi, 2004	문항 통계에 의해 문제의 가능성이 있는 문항들이 분석됨.	문항의 난이도를 예측할 때 선택지에 사용된 어휘와 표현의 난이도를 고려하도록 문항 검토자들을 훈련함.
Kim & Lee, 2004	7개의 TEPS 검사형에 포함된 문법 문항들을 분석하여 어떠한 문법적 특성들이 서로 그룹화 되어야 할지 제안함.	문항 출제 가이드를 수정하여 문법 문항에서 문법적 특성을 분류하라는 제안을 반영함.
Park, 2004	200개의 어휘 문항의 예측 난이도와 실제 난이도를 비교·분석하였다. 정답 이외의 선택지들의 난이도가 어휘 문항의 난이도에 영향을 미치는 것으로 나타남.	어휘 문항의 난이도를 예측할 때 선택지 간의 관련성과 언어(collocations)를 고려하도록 문항 검토자들을 훈련함.
Choi et al., 2010	문항 출제 절차를 분석하여 효율성을 높일 수 있는 방안을 제시함.	새로 출제되는 문항의 관리가 더욱 용이해짐.
Kim & Kim, 2009	메뉴와 기능을 업데이트하여 온라인 시험 검토 시스템을 개선함.	TEPS 센터에서의 문항 검토 절차가 더욱 용이해짐.
Lee, 2009; Lee et al., 2010;	시험 검토 절차를 분석하여 효율성을 높일 수 있는 방안을 제시함.	문항 출제 가이드와 시험 검토 가이드가 개정 및 업데이트되어 시험 검토 절차가 더욱 용이해짐.
Sohn et al., 2010	문항 데이터베이스 프로그램을 업데이트하였고 시험 검사형 제작 절차를 문항 출제, 문항 검토, 시험 검사형 제작 세 단계로 간소화함.	TEPS 센터에서의 시험 검사형 제작이 더욱 용이해짐.

5.3. TEPS의 개정을 위한 연구

2012년부터 현재에까지 이르는 세 번째 시기에는 TEPS를 개정하기 위한 노력의 일환으로 여러 개의 연구 과제가 수행되었다. 연구 과제로는 전문가 회의(Kim et al., 2012; Jun et al., 2014), 요구 분석 및 프로토타이핑 연구(Lee et al., 2015), 예비 가능성(feasibility) 연구(Lee et al., 2016), 파일럿 시험 및 필드 시험(Kwon et al., 2018), 그리고 안정화 및 타당도 연구(Lee et al., 2019)가 있었다. 기존의 TEPS를 개정하거나 혹은 완전히 새로운 시험을 개발하기 위한 여러 기초 연구를 수행한 결과, 기존의 TEPS는 고능력 영어학습자를 잘 변별해내는 장점에도 불구하고, 많은 수험자가 기존 TEPS를 다소 길고 어렵게 느끼는 것을 확인하였다. 기존 독해 영역에 포함된 40 문항은 각각 1지문 1문항의 형태였기 때문에 수험자는 총 40개의 지문을 읽어야 했고, 청해 영역의 60 문항도 마찬가지였다. 또한 고능력 수험자를 변별할 목적으로 포함되는 일부 어휘의 높은 난이도에 대한 수험자의 비판도 있었다.

언어교육원 TEPS 센터는 이러한 의견을 적극적으로 수용하여, 총 문항 수를 총 200 문항에서 135 문항으로 줄이고, 각 영역의 문항 수도 축소하였다. 좀 더 상세히 살펴보면, 청해 영역은 60 문항에서 40 문항으로, 어휘와 문법 영역은 종전의 각각 50 문항에서 각각 30 문항으로, 독해 영역은 40 문항에서 35 문항으로 축소되었다. 뿐만 아니라 청해 및 독해 영역에 ‘1지문 2문항’ 형식을 도입하고, 독해 영역에 더욱 다양해진 의사소통 상황을 반영하여 실제 현실과 보다 밀접하게 연계된 새로운 지문 형식(이메일, 인터넷 신문 기사, 휴대폰 메시지 등)의 활용 가능성을 모색하였다. 수 차례의 파일럿 시험 및 필드 시험을 통해 개정 가능성을 확인한 후 개정 TEPS에 이러한 새로운 특성들을 반영하기로 결정하였고, 개정된 TEPS는 2018년 5월부터 시행되었다. 세 번째 시기의 주요 내부 연구 과제는 <표 3>에 요약되어 있다.

표 3. TEPS의 개정을 위한 연구

연구 과제	연구 내용	제안점
Kim et al., 2012; Jun et al., 2014	전문가들은 최근 영어교육 및 평가 이론의 변화와 영어사용 환경의 변화를 더 잘 반영하도록 TEPS를 개정할 필요가 있다고 의견을 모음.	구제적인 변경 사항을 제안하고 변경 가능성을 탐색해야 함.
Lee et al., 2015	요구 분석에 의하면 한국 영어학습자들은 TEPS를 어렵게 느끼지만 청해와 독해 영역에 새로운 문항 유형을 추가하는 것에 대해서는 긍정적인 반응을 보임.	개정된 TEPS 시행의 가능성을 탐색해 보아야 함.
Lee et al., 2016	개정된 TEPS 시행의 가능성을 조사함.	개정 TEPS 검사형을 개발하여 파일럿 시험을 거쳐야 함.
Kwon et al., 2018	세 차례의 대규모 파일럿 시험과 한 차례의 대규모 필드 시험을 통해 개정 TEPS의 문항 및 검사 통계의 적절성을 보여 주었다. 수험자들은 개정 TEPS에 포함된 변경 사항에 대해 긍정적인 반응을 보임.	TEPS 센터는 개정 TEPS를 안정화시키고 타당도 근거를 마련하기 위해 적극적인 연구를 지속해야 함.
Lee et al., 2019	개정 TEPS 청해 영역 및 독해 영역에 대한 수준설정 연구를 수행하였고, 성적표에 제시되는 세부능력 점수의 안정성에 대한 연구를 수행함.	TEPS 센터는 어휘 영역과 문법 영역에 대한 추가적인 수준설정 연구와 타 표준화 영어능력시험과의 비교 연구를 수행하여 점수 해석과 사용 가능성을 높여야 함.

6. 맺는 말

본 논문에서는 지난 20여년 간의 TEPS의 연구개발, 성장, 개정 과정을 서울대학교 언어교육원 50여년 외국어 평가 역사의 연장선 상에서 조망하고 그 주요 성과들을 검토해 보았다. 아울러

TEPS 시험 개발, 개선, 개정 및 타당도 검증 연구들을 검토하고 요약함으로써 TEPS의 지속적인 성장을 위해 바람직한 발전방향과 주요 연구과제들을 정리해 보고자 하였다.

이러한 정리 과정을 통해 확인한 몇 가지 중요한 과제들을 언급하면 다음과 같다. 우선, TEPS 시험은 여러 긍정적인 시험의 특성과 장점에도 불구하고 다소 시험의 길이가 길고 시험의 난도가 높다는 의견들이 있었다. 이런 의견들을 반영하여 TEPS의 개정작업을 최근 수년간 진행하였고 개정된 TEPS 시험을 2018년 5월부터 시행하고 있다. 그러나 여전히 개정 TEPS 보다도 더 쉬워진, 중급 영어학습자들을 주요 대상으로 하는, 새로운 영어 숙달도 시험 개발에 대한 요구와 필요성에 대한 논의가 계속되고 있는 상황이다. 따라서 개정 TEPS는 개정 TEPS대로 기존 주요 수험자집단을 대상으로 한 범용 영어 숙달도 시험(*general-purpose English proficiency test*)으로서의 역할을 계속 해 나가도록 하게 하고, 동시에 초·중등학생, 취업준비생, 직장인 등의 영어능력의 척도로 활용할 수 있는 초·중급 학습자를 위한 영어 숙달도 시험들을 단계적으로 개발하는 시도를 해 볼 필요가 있다.

앞으로도 국내에서 자체적으로 이미 개발된 혹은 앞으로 개발될 다양한 영어능력 시험이 안정적으로 정착하려면 여러 가지 조건이 만족되어야 한다. 우선 영어시험의 개발연구, 설계, 문항 출제 및 검토 관련 전문성 배양, 문항 관리 데이터베이스의 구축, 시험의 인프라 관리, 문항 분석, 채점, 척도 설정 및 시험 동등화, 그리고 점수 환산표 제작 등과 관련된 심리측정학적 전문성의 축적을 위한 지속적이고 체계적인 노력이 필요하다. 아울러 이러한 분야에 꾸준한 연구 및 투자가 이루어져야 하고, 새로운 시험이나 시험 관련 서비스 개발을 위해 사업 수익의 일정 부분은 이러한 분야의 개발연구에 우선적으로 재투자 되어야 할 것이다.

둘째, 정책적인 측면에서 볼 때, 정부가 국내에서 개발된 영어 시험들이, 현재 국내 시장을 압도하고 있는 국제적인 영어 시험들과 함께, 발전해 나갈 수 있도록 공정한 생태계를 조성하기 위해 더 적극적인 역할을 하는 것이 필요하다. 그동안 정부에서는 ‘자격기본법’이나 ‘자격 기본법 시행령’ 같은 법규정에 근거해서 한국직업능력개발원 및 관련부처들을 통해 외국어 시험을 포함한 각종 민간자격 시험들을 관리·감독해 왔다. 하지만 이러한 규정과 지침은 국내에서 개발된 시험에는 비교적 엄격하게 적용되고 있지만, 압도적인 자본과 축적된 경험을 가지고 국내에 들어와 영어 평가 시장을 선점하고, 각종 정부 기관들에서 채택되어 사용되고 있는 해외 시험들에게는 제대로 적용되고 있지 않은 것으로 보인다. 현재 국내 기관에서 개발되어 시행 중인 여러 영어시험들은 정부나 각종 공공 기관의 영어능력 척도로 인정받기 위해 국가공인 민간 자격의 지위를 취득하고 유지하기 위해 온갖 노력을 다하고 있지만, TOEFL 혹은 TOEIC 같은 국제적인 시험은 그러한 공인 자격의 취득 없이도 국내 시험들보다 더 많은 정부 및 공공 기관에서 인정받고 사용되고 있는 실정이다. 물론 이러한 국제적인 영어 시험들도 국내 심의 절차를 부과 받게 되면 다 무난하게 통과할 수 있는 역량을 갖추었을 수도 있겠지만, 원칙적으로 볼 때 공정한 정책이나 관행이라고 보기는 어렵다. 아울러 정부는 국내에서 개발되어 시행 중인 시험들에게 국가공인 시험 지위의 획득과 유지를 위한 책임과 의무만을 부여할 것이 아니라, 이러한 자격의 획득이 해당 평가기관에 실질적인 혜택이 돌아가게 하고 또 그러한 조치로 인한 긍정적인 효과가 사회적으로 파급 및 확산되도록 유도하는 정책을 입안하고 실행할 필요가 있다.

셋째, 정부 정책과 관련해서 추가적으로 꼭 언급해야 할 필요가 있는 분야로는 정부가 사용하고

있는 국내의 영어시험 간의 점수 환산표의 공정성 문제이다. 해외에서 개발되어 국내에서 널리 사용되고 있는 시험들은 압도적인 자본과 축적된 경험을 소유하고 있을 뿐만 아니라 국내 국가 민간자격공인 제도의 제외법권 지역에 위치해 있다. 여기에다 정부에서 사용하고 있는 시험 간 점수 환산표에서도 유리한 입지를 누리고 있다는 의견들이 수험자들과 언어평가 전문가들 사이에서 오랫동안 제기되어 왔다. 국내외의 여러 기관에 의해 제작되어 시행되고 있는 주요한 시험들 간의 점수 환산표 제작을 위한 연구는 어느 특정 평가기관이 주도하기 보다는 제 3자적 입장의 감독관리자의 위치에 있는 해당 정부기관에서 충분한 예산을 배정해서 발주하고 주도하는 연구프로젝트 형식을 통해서 수행하는 것이 공정성 시비를 없애는 길이다. 해당 정부 기관이 직접 연구를 발주하거나 주도하기 어려운 경우에는, 해당 국내의 평가기관들이 공동으로 참여하여 연구 프로젝트를 수행하도록 주선을 할 수도 있을 것이다. 앞으로 우리 나라의 '외국어 평가의 주권'을 지키고 관련 분야를 계속 발전시켜 나가기 위해서는 국내 평가기관들의 자체적인 노력뿐만 아니라 정부의 지속적인 관심과 노력과 지원이 절실히 필요한 시점이다.

넷째, 언어교육원 TEPS 센터에서 개발하여 시행하고 있는 기존의 지필 TEPS 시험뿐만 아니라 컴퓨터 기반 시험에 대한 연구와 개선 작업도 계속 할 필요가 있다. 언어교육원에서 개발하여 시행하고 있는 컴퓨터 기반 시험으로는 영어 말하기 및 쓰기 시험인 TEPS-Speaking과 TEPS-Writing과 i-TEPS 시험이 있다. 의사소통능력(communicative competence)이나 상호작용능력(interactive competence)이 강조되는 오늘날의 언어교육이나 평가 분야의 이론적 경향과 실제적인 추세를 볼 때 말하기 평가와 쓰기 평가는 앞으로도 계속 관심을 갖고 발전시켜 나가야 할 분야이다. 더욱이 모든 것이 자동화되고 연결되는 초연결 사회가 다가 오고 있기 때문에 종래의 지필시험 체제의 발전과 개선을 위한 연구뿐만 아니라 컴퓨터, 인터넷을 포함한 멀티미디어 기반 영어평가 시스템에 대한 연구와 투자도 계속해 나가는 것이 미래의 지속적 성장과 발전을 담보하는 길이라 생각된다.

마지막으로, 이제 20돌을 맞은 TEPS는 시야를 넓혀 조금 더 멀리 내다볼 필요가 있어 보인다. 현재 우리 사회는 제 4차 산업혁명의 새로운 시기로 진입하고 있고, 이와 함께 인공지능, 사물 인터넷, 자연언어처리, 자동통·번역, 스마트 스피커, 인간형 로봇 등과 같은 4차 산업혁명의 핵심 첨단 기술의 물결이 밀려오고 있는 것을 목도하고 있다. 이러한 새로운 물결은 앞으로 외국어의 교육 및 평가에 대한 수요와 발전에도 지대한 영향을 미칠 것으로 예상된다. 특히 심층학습(deep learning)이나 심층신경망(deep neural network) 등과 같은 기계학습 및 인공지능 관련 기술들은 이미 언어평가 분야에도 영향을 미치고 있다. 언어평가 분야에서는 일찍부터 자동 말하기 및 쓰기의 평가 시스템이 개발되어 실용화되고 있었는데 여기에 인공지능이 기법이 최근 결합되어 채점시스템의 정확성과 신뢰도를 향상시켜 나가는 연구가 이루어지고 있는 상황이다. 사실 우리는 이전에 경험해보지 못했던 새로운 도전과 기회에 직면해 있다. TEPS가 앞으로 펼쳐질 새로운 변화와 도전의 시기에도 계속 영어학습자의 영어능력을 평가하는데 유용한 도구로 발전하고 진화해 나가며, 국제적인 시험으로도 인정받을 수 있도록 노력해야 할 것이다. 이를 위해서는 미래의 여러 기술적, 사회적 변화가 TEPS와 그 자매 시험들에 어떤 영향을 주게 될지 열린 마음으로 바라보고 능동적이고 진취적으로 대응해 나가는 것이 필요하다.

References

- Ahn, J. (2002). *Seeking a balance in TEPS passages: A study in sensitivity issues [1]* (Research Report No. 35). Seoul: SNU Language Education Institute.
- Ahn, J. (2004). *Seeking a balance in TEPS listening conversations: A study in sensitivity issues* (Research Report No. 38). Seoul: SNU Language Education Institute.
- Cho, C., Byun, C.-G., Yoo, M., & Ryu, D. (2010). *Plans for improving the item-writing system of TEPS* (Research Report No. 61). Seoul: SNU Language Education Institute.
- Choi, I.-C. (1993a). Construct validation study on SNUCREPT (Seoul National University Criterion-Referenced English Proficiency Test). *Language Research*, 29(2), 243-275.
- Choi, I.-C. (1993b). *The development of a language proficiency test* (Research Report No. 2). Seoul: SNU Language Education Institute.
- Choi, I.-C. (1994a). Content and construct validation of a criterion-referenced English proficiency test. *English Teaching*, 48, 311-348.
- Choi, I.-C. (1994b). *Development of Seoul National University criterion-referenced English proficiency test* (Research Report No. 3). Seoul: SNU Language Education Institute.
- Choi, I.-C. (1995). A comparability study on SNUCREPT and TOEIC. *Language Research*, 31(2), 357-386.
- Choi, I.-C. (1997a). *Construct validation on SNUCALTS (Seoul National University Computer-Assisted Language Testing Service)* (Research Report No. 6). Seoul: SNU Language Education Institute.
- Choi, I.-C. (1997b). Essential test method facets of a general English proficiency test and their validity as perceived by test-takers. *Language Research*, 33(4), 773-799.
- Choi, I.-C. (1999). Test fairness and validity of the TEPS. *Language Research*, 35(4), 571-603.
- Choi, I.-C. (2004). *Content validation of TEPS* (Research Report No. 39). Seoul: SNU Language Education Institute.
- Choi, I.-C. (2019, January 1). Written Interview.
- Choi, I.-C., & Lee, H.-O. (2001a). A trend analysis of individual test takers' performance of TEPS. *Language Research*, 37(3), 675-710.
- Choi, I.-C., & Lee, H.-O. (2001b). *A trend analysis of test takers' performance of TEPS* (Research Report No. 30). Seoul: SNU Language Education Institute.
- Choi, I.-C., & Lee, H.-O. (2002). *The effects of speededness of the TEPS on test performance* (Research Report No. 33). Seoul: SNU Language Education Institute.
- Choi, I.-C., & Park, J. (2001). *Validation of ability parameter estimation methods for TEPS scoring schemes* (Research Report No. 29). Seoul: SNU Language Education Institute.
- Hwang, I. (2001). *A study on test set generation using simulated annealing in TEPS system* (Research Report No. 28). Seoul: SNU Language Education Institute.
- Jun, Y., Ryu, D., Park, Y. J., Shin, S., Jun, H.,...Byun, J. (2015). *Research for the development*

- of *TEPS 2.0* (Research Report No. 77). Seoul: SNU Language Education Institute.
- Kim, M., Jung, S., Kim, J., Lee, Y.-W., Lee, Y., Shin, S.,...Lee, G. (2012). *A basic study for the overall amendment of regular TEPS* (Research Report No. 73). Seoul: SNU Language Education Institute.
- Kim, S., & Kim, Y. (2009). *Research on improving the intranet program for item reviews* (Research Report No. 56). Seoul: SNU Language Education Institute.
- Kim, S., & Lee, K. (2004). Developing TEPS Grammar Items—Focusing on the basic sentence structure items (Research Report No. 41). Seoul: SNU Language Education Institute.
- Kim, M. & Moon, Y. (2018, December 13). Personal Interview.
- Kim, S., Park, Y., Lee, K., & Leaper, D. (2001). Making a comprehensive manual for test writing (Research Report No. 32). Seoul: SNU Language Education Institute.
- Kwon, H., Lee, Y.-W. Lee, Y., Park, Y. J. Kim, J., Jun, H.,...Park, H. (2018). *Development and validation of a pilot test form for the revised TEPS* (Research Report No. 80). Seoul: SNU Language Education Institute.
- Lee, B., Kim, C., Park, Y., So, Y., Lee, Y., Jun, H.,...Yeom, S. (2016). *Preliminary feasibility verification project for implementation of New TEPS in 2016* (Research Report No. 79). Seoul: SNU Language Education Institute.
- Lee, B., Park, Y., Lee, Y., So, Y., Lee, Y., Kim, C.,...Yeom, S. (2015). *The development of New TEPS* (Research Report No. 78). Seoul: SNU Language Education Institute.
- Lee, K. (2000). *A study on the content composition of TEPS grammar and vocabulary* (Research Report No. 26). Seoul: SNU Language Education Institute.
- Lee, K., Kim, S., & Park, Y. J. (2003). *Assessing the difficulty levels of TEPS reading comprehension items* (Research Report No. 36). Seoul: SNU Language Education Institute.
- Lee, S. (2004). *State Examination System of Korea*. Seoul: Korean Studies Information.
- Lee, Y. (2008). Fifty years of development in English language testing and the future research direction. *Korean Journal of Applied Linguistics*, 24(3), 87-103.
- Lee, Y. (2009). *Evaluation and proposal for reforms on the current TEPS reviewing process* (Research Report No. 60). Seoul: SNU Language Education Institute.
- Lee, Y., Anderson, P., & Ahn, J. (2010). *Revising the review/writing manuals and improving the training process of researchers* (Research Report No. 62). Seoul: SNU Language Education Institute.
- Lee, Y.-W., Kwon, H., Lee, Y., Lim, E., Jun, H., Song, M.-J.,...Yoon, M. (2019). *Research for the stabilization and validation of the New TEPS* (Research Report No. 81). Seoul: SNU Language Education Institute.
- Lee, Y.-W., Jun, H., & Kim, J. (in press). TEPS and its family of tests. In L. I.-W. Su, C. J. Weir, and J. R. W. Wu (Eds.), *English language proficiency testing in Asia: A new paradigm bridging global and local contexts*. New York: Routledge (Taylor & Francis Group).

- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. New York: Wiley.
- Park, N. (2018, November, 28). Personal interview
- Park, N., Lee, S.-U., Hong, J., Shin, J., & Choi, I.-C. (1992). *A report of survey analysis results for the reformation of foreign language examinations* (Research Report No. 1). Seoul: SNU Language Education Institute.
- Park, S. (2000). *A study to reduce IRT scoring error for TEPS* (Research Report No. 27). Seoul: SNU Language Education Institute.
- Park, S., & Shin, D. (2015). A study on the social dimension of second language testing for citizenship: Focus on KINAT. *Multicultural Education Studies*, 8(1), 49-75.
- Park, Y. J. (2004). *A comparative study on difficulty levels between predictions and student responses* (Research Report 43). Seoul: SNU Language Education Institute.
- Sohn, C., Hwang J., & Cho, H. (2010). *Plans for improving TEPS item-writing and item-review processes* (Research Report 68). Seoul: SNU Language Education Institute.
- Spolsky, B. (1977). Language testing: Art or science. In G. Nickel (Ed.), *Proceedings of the Fourth International Congress of Applied Linguistics* (Vol. 3, pp.7-28). Stuttgart, Germany: Hochschulverlag.
- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.
- Spolsky, B. (2008). Language assessment in historical and future perspective. In E. Shohamy and N. H. Hornberger (Eds.), *Encyclopedia of Language and Education (2nd ed.)*, Volume 7: *Language Testing and Assessment* (pp.445-454). New York: Springer.
- Weir, C. (1990). *Communicative language testing*. Hemel, Hemstead, UK: Prentice Hall.