

개정 TEPS 점수해석 및 사용에 대한 타당도 논증

전희성^{1†}

¹서울대학교

A Validity Argument for the Interpretation and Use of Revised TEPS Scores

Heesung Jun^{1†}

¹Seoul National University

ABSTRACT

Validity is an evaluative judgment of the appropriateness of evidence supporting test score interpretation and use (Messick, 1989). As a first step toward arguing for the validity of TEPS score interpretation and use, this paper presents an interpretation/use argument (IUA) for the revised TEPS, following Kane's argument-based approach to validation (Kane, 2013). The IUA for the revised TEPS outlines seven inferences—domain definition, evaluation, generalization, explanation, extrapolation, utilization, and consequence implication—with a warrant and one or more assumptions underlying each inference. Backing or evidence that is needed to support each assumption is also stated in the IUA. As a second step, a validity argument for the revised TEPS is made by describing the backing that has already been collected while pointing out the backing that is still needed. The paper concludes with further research that needs to be conducted to support the validity argument more fully.

Keywords: interpretation/use argument, validity argument, argument-based approach to validation, revised TEPS

1. 서 론

서울대학교 언어교육원에서 개발되어 1999년부터 시행되어온 TEPS(Test of English Proficiency developed by Seoul National University의 약자)는 그간의 사회적 변화에 따른 의사소통 방식의 변화, 영어교육 및 평가 분야에서의 연구 성과, 그리고 수험자의 요구에 따라 2012년부터 수행된 다양한 연구를 기반으로 개정되었고(본 특별호의 Jun et al., 2019 참조), 변경된 TEPS는 2018년 5월부터 시행되었다. 본 논문의 목적은 그동안의 개정 연구

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



Copyright © 2019 Language Education Institute, Seoul National University.

This is an Open Access article under CC BY-NC License (<http://creativecommons.org/licenses/by-nc/4.0>).

결과를 근거로 개정 TEPS의 점수 해석과 사용에 대한 타당도를 주장하는 것이다. 이를 위해 본 논문에서는 우선 개정 TEPS의 구인, 목적 및 대상에 대해 간략히 소개한 후 타당도의 정의와 타당도 검증 방법에 대한 이론적 배경을 설명하였다. 그 후에는 논증기반 접근법에 따른 개정 TEPS 점수의 해석과 사용을 위한 논증을 제시하였다. 이어 이미 수집된 타당도 근거에 대해 서술하는 방식으로 타당도 논증을 제시하였다. 마지막으로 타당도 논증을 더 잘 뒷받침하기 위해 필요한 근거를 모으기 위한 앞으로의 연구 계획을 소개하였다.

2. 개정 TEPS의 구인, 목적 및 대상

TEPS는 영어를 외국어로 사용하는 학습자의 영어 실력을 측정하기 위해 개발된 영어능력평가 시험이다. 비즈니스, 학술, 일상 생활 등 실제 생활에서 접할 수 있는 폭넓은 분야의 주제와 상황을 다루는 다양한 대화와 지문에서 문항이 출제되기 때문에 종합적인 영어숙달도를 평가한다. 평가영역은 청해, 어휘, 문법, 독해이다. 청해 영역에서는 대화의 흐름을 파악하는 능력, 대화나 담화의 주제를 파악하는 능력, 대화나 담화의 세부내용을 파악하는 능력, 대화나 담화의 내용을 바탕으로 추론하는 능력을 평가한다. 어휘 영역에서는 다양한 대화와 담화에서 쓰이는 어휘를 평가한다. 문법 영역에서는 다양한 대화와 담화에서 쓰이는 문법 요소를 평가한다. 마지막으로 독해 영역에서는 글의 주제를 파악하는 능력, 글의 세부내용을 파악하는 능력, 글의 내용을 바탕으로 추론하는 능력, 글의 흐름을 파악하는 능력을 평가한다(<http://www.teps.or.kr/info/teps>).

TEPS 점수는 다양한 목적으로 사용되고 있다. 우선 외국어고, 과학고, 국제고 등 전국의 특수목적고와 자율사립고 등에서 학생들의 영어실력향상 및 평가를 위해 단체로 TEPS에 응시하고 있다. 또한 대학교 입학, 편입학, 졸업인증, 장학생 선발 등을 위한 기준으로 TEPS 점수가 사용되고 있으며 대학원 및 전문대학원 입학에도 TEPS 점수가 반영되고 있다. 기업에서는 신입, 경력, 인턴 채용 및 인사사고과, 해외파견근무자 선발 등 직원 영어능력평가를 위해 TEPS를 활용하고 있으며 중앙행정부, 지방자치단체, 공공기관에서도 직원 채용이나 해외 연수대상자 선발을 위해 TEPS 점수가 사용된다. 마지막으로 TEPS는 행정고시, 외무고시, 사법시험 등 주요 국가고시와 변호사, 공인회계사, 감정평가사 등 각종 자격증 시험에서 영어 시험을 대체하는 자격 시험으로 활용되고 있다(<http://www.teps.or.kr/use/school>).

3. 타당도의 정의 및 타당도 검증 방법

타당도의 정의는 교육평가 및 심리측정 분야의 발달에 따라 진화해왔다. 전통적으로 타당도는 검사도구가 측정하고자 하는 것을 얼마나 충실히 측정하였는가로 정의되었으며 타당도는 검사도구의 속성으로 생각되었다. 그리고 타당도를 내용타당도(content validity), 구인타당도(construct validity), 준거관련타당도(criterion-related validity) 세 종류로 구분하여 검

사도구의 타당도를 검증하였다(AERA, APA, & NCME, 1966). 하지만 타당도의 개념은 변화하여 이제는 타당도를 시험 점수의 해석과 사용의 적합성에 대하여 근거나 이론이 지지해주는 정도에 대한 판단으로 정의한다(Messick, 1989). 따라서 타당도는 시험 결과의 해석에 대한 근거의 적합성이지만 시험 자체의 속성이 아니다. 타당도의 정의가 변함에 따라 타당도의 종류 대신 근거리는 용어가 사용되기 시작했고(AERA, APA, & NCME, 1999), Messick(1989)은 내용과 관련된 근거, 구인과 관련된 근거, 준거와 관련된 근거에 결과(consequence)와 관련된 근거를 추가하면서 시험점수 사용을 둘러싼 사회적 가치와 윤리성의 문제까지도 제기하였다(Chapelle, 2012; Seong, 2002).

타당도의 정의가 변화하면서 타당도 검증 방법에 대한 이론도 변화하였다. 시험 점수 해석과 사용의 타당도를 주장하기 위해서는 적절성을 판단할만한 다양한 근거를 제시해야 하는데, 전통적으로는 타당도 근거의 종류 각각에 대한 검증을 통해 근거를 축적하는 접근법(accumulation-of-evidence approach)이 일반적이었다(Chapelle, 2008). Bachman and Palmer(1996)가 제시한 시험의 유용성(usefulness)을 평가하는 여섯 가지 기준인 신뢰도(reliability), 구인타당도(construct validity), 진정성(authenticity), 소통성(interactiveness), 영향력(impact), 실용성(practicality)도 이러한 접근법을 따른다고 볼 수 있다. 그러다가 교육평가 분야에서 Kane(1992, 2006, 2013)이 다양한 근거를 논리적이고 체계적으로 제시할 수 있는 방법으로 Toulmin(2003)의 논증구조(argument structure)를 기반으로 한 논증 기반 접근법(argument-based approach)을 제시하였고, Chapelle, Enright과 Jamieson(2008)이나 Bachman과 Palmer(2010) 등이 언어평가 분야에도 이를 적용하였다. 논증 기반 접근법에 서는 시험 점수 해석과 사용의 타당도를 판단할 만한 다양한 근거를 제시할 때 단계적인 논증을 통해 시험 점수의 의미와 의도된 점수 사용에 대한 여러 개의 연결된 주장을 하고 각 주장을 뒷받침하는 근거를 제시한다. 우선 시험 개발자가 점수 해석과 사용에 대한 논증(interpretation/use argument)을 전개하고, 이 논증을 뒷받침할 수 있는 다양한 근거를 이론과 연구를 바탕으로 모은 후, 이 근거를 해석·사용 논증과 종합한 타당도 논증(validity argument)을 작성하여 공개한다. 그리고 그 타당도 논증이 논리적이고 적절한지는 시험과 시험 점수를 사용하는 수험자, 교사, 강사, 고용주, 인사담당자, 입학심사관 등을 포함한 이해 당사자들(stakeholders)과 평가전문가들이 판단한다(Chapelle, 2008).

4. 개정 TEPS 점수 해석 및 사용에 대한 논증

Kane의 논증기반 접근법에 기반한 개정 TEPS 점수 해석 및 사용에 대한 논증의 구조는 미국 Educational Testing Service(ETS)에서 개발한 대단위 시험인 Test of English as a Foreign Language(TOEFL) 점수의 해석과 사용을 위한 타당도 논증을 담은 Chapelle et al.(2008)을 참고하여 작성되었다. 우선 논증에는 일곱 개의 추론(inference)이 포함되어 있는데, 영역 정의(domain definition), 평가(evaluation), 일반화(generalization), 설명(explanation), 확장(extrapolation), 활용(utilization), 그리고 결과 함의(consequence

implication)이다. 이 일곱 개의 추론이 연결되어 하나의 체인(chain)을 이루는데, 추론은 하나의 주장(claim)이 되며 그 주장은 다음 연결고리인 추론의 토대가 되는 자료(grounds)로 사용된다. 따라서 첫 추론인 영역 정의는 언어숙달도가 필요한 장소 혹은 상황을 지칭하는 목표 언어 사용 영역(target language use domain)과 수험자의 시험 수행의 관찰을 연결하고, 평가 추론은 수험자의 시험 수행의 관찰과 관찰 점수를 연결한다. 이 체인은 결국 시험 사용의 결과까지 이어지는데, 각각의 추론은 그것을 정당화하는 일반적인 원칙인 하나의 보증(warrant)으로 뒷받침되고, 각 보증은 하나 이상의 가정(assumption)으로 뒷받침된다. 점수 해석 및 사용에 대한 논증에는 각 가정에 대한 근거(backing)가 어떠한 연구방법을 통해 수집될 것인지도 제시된다.¹⁾ <표 1>은 개정 TEPS 점수 해석 및 사용에 대한 논증을 요약하고 있다.

표 1. 개정 TEPS 점수 해석 및 사용에 대한 논증

추론	보증	가정	근거(연구방법)
영역 정의	시험 수행의 관찰은 목표 언어 사용 영역(target language use domain)을 대표하는 관련 능력과 지식을 드러낸다.	1. 실제 생활에서 영어를 이해하기 위해 필수적인 영어 능력과 지식이 파악된다. 2. 실제 생활을 대표할만한 평가 과제들이 파악된다. 3. 필수적인 능력을 요구하고 도메인을 대표하는 평가 과제들의 제작이 가능하다.	1. 도메인 분석(요구 분석) 2. 전문가 자문 3. 문항 디자인 및 개발
평가	시험 수행의 관찰이 평가 되어 목표 언어 능력을 반영하는 관찰 점수를 제공한다.	1. 시험 수행 조건들이 목표 언어 능력의 증거를 나타내기에 적절하다. 2. 문항과 검사형의 통계적 특성이 표준지향 결정을 내리기에 적절하다.	1. 프로토타이핑, 파일럿 시험, 필드 시험 2. 문항 분석, 국부의존성 분석
일반화	관찰 점수는 수험자가 동등한 검사형과 시행 회차에서 받을 것으로 기대되는 기대 점수의 추정치이다.	1. 시험 평가틀(test specification)이 상세히 기술되어 있어 동형(parallel) 문항과 검사형을 제작할 수 있다. 2. 시험에 포함된 문항의 수가 수험자들의 수행에 대한 안정적인 추정치를 제공하기에 충분하다. 3. 시험 점수에 대해 적절한 척도와 동등화 절차가 사용된다.	1. 동형 문항과 검사형을 제작할 수 있는 시험 평가를 개발 2. 신뢰도 연구 3. 척도화와 동등화
설명	기대 점수는 일반적인 영어능력이라는 TEPS의 시험 구인에 기인한다.	1. 시험 과제를 성공적으로 완수하기 위해 필요한 영어 능력과 지식이 이론적 기대에 부응한다. 2. TEPS에서의 성취가 타 영어숙달도 시험들에서의 성취와 연관된다. 3. TEPS 점수의 내적 구조가 영어	1. TEPS 문항을 풀 때 사용되는 인지적 과정 조사(시험 후 설문조사, 자극회상, 시선추적 등) 2. TEPS 점수와 타 표준화 영어시험 점수와의 상관관계 3. 요인분석

1) 추론, 보증, 가정, 그리고 근거로 구성되는 기본단위는 Toulmin(2003)의 논증구조에 기반을 두고 있다.

		<p>능력은 서로 밀접한 연관이 있는 구성요소들로 이루어졌다는 영어능력에 대한 이론적 관점과 일치한다.</p> <p>4. TEPS에서의 성취가 영어 학습 경험의 양과 질에 따라 달라진다.</p>	4. TEPS 점수와 영어 학습 내용 및 기간과의 상관관계 검토
확장	TEPS가 평가하는 일반적인 영어능력이라는 구인이 목표 언어 사용 영역에서의 영어 수행의 질을 설명한다.	1. TEPS에서의 성취는 다른 일반적인 영어능력의 기준들과 연관이 있다.	1. TEPS 점수와 타 표준화 영어시험 점수, 수험자의 자기평가, 수험자의 영어 능력에 대한 교사, 강사 및 고용주의 평가와의 상관관계
활용	목표 언어 사용 영역에서의 수행의 질의 추정치는 수험자들에 대한 판단을 내리기에 유용하다.	1. TEPS 점수는 수험자의 일반적인 영어능력에 대해 유용하고 의미 있는 정보를 제공하고, TEPS 점수의 의미를 수험자와 잠재적 고용주를 포함한 모든 이해당사자가 명확히 해석할 수 있다.	1. 성적표 개발, 수준설정, 이해당사자 인터뷰와 설문조사
결과 함의	TEPS 사용의 결과와 내려진 결정들이 수험자와 다른 이해당사자들에게 유익하다.	1. TEPS가 일반적인 영어의 학습과 교육 방식에 긍정적인 영향을 끼친다. 즉, 긍정적인 환류 효과를 촉진시킨다.	1. 환류효과 연구(교사, 강사 및 전문가 인터뷰, 수험자 인터뷰 및 설문조사)

5. 개정 TEPS 타당도 근거

이 장에서는 개정 TEPS 점수 해석 및 사용에 대한 논증에서 각 가정을 뒷받침하기 위해 필요한 근거 중 이미 수집된 것을 소개하고 아직 마련되지 않은 근거가 무엇인지도 명시한다. 우선 개정 TEPS 개발과 초기 시행 및 안정화 과정에서 수집된 다양한 타당도 근거를 요약하면 <그림 1>과 같다.

TEPS 개정을 위한 준비	개정 TEPS 개발	개정 TEPS 초기 시행 및 안정화
<ul style="list-style-type: none"> 개정이 필요한 부분과 개정의 방향 제시(Kim et al., 2012) 수요자 분석과 개정안 제시(Jun et al., 2014) 요구 분석, 새로운 문항 유형 개발(prototyping)(Lee et al., 2015) 새로운 시험 개발과 시행의 가능성(feasibility) 탐색(Lee et al., 2016) 	<ul style="list-style-type: none"> 문항 수에 따른 신뢰도 시물레이션(Kim, 2016) 1, 2차 파일럿 시험의 국부 의존성 분석(Kim & Jun, 2017) 파일럿 시험 3회와 필드 시험 1회의 문항 분석, 신뢰도 분석, 요인분석, 수험자 설문 분석(Kwon et al., 2018) 개정 TEPS 평가를 작성 및 척도화·동등화 절차 수립(TEPS Center, 2019a, 2019b, 2019c) 	<ul style="list-style-type: none"> 초기 시행 회차의 문항 분석, 신뢰도 분석, 요인분석; 성적표에 제공되는 세부능력 수준 정보의 안정성 연구(Lee et al., 2019) TEPS 청해·독해 영역 점수와 CEFR 연계 수준설정(Jun, Lim, & Lee, 2019; Lim, Jun, & Lee, 2019) TEPS-GEPT 비교연구(Lim et al., 2018)

그림 1. 개정 TEPS 개발과 초기 시행 및 안정화 과정에서 수집된 타당도 근거

가장 먼저 영역 정의 추론을 뒷받침하기 위해서는 목표 언어 사용 영역을 대표하는 필수적인 언어 능력과 지식을 파악하고 그것을 평가할 평가 과제를 제작해야 한다. 일단 TEPS의 개정은 기존 TEPS의 시행이 시작된 이후 지속적으로 수집된 요구 분석 자료와 그동안의 사회적 변화를 통해 새롭게 요구되는 의사소통 방식(이메일, 인터넷 신문기사, 메신저 대화 등)을 반영하여 이루어졌다. 또한 수 차례의 프로토타이핑, 세 차례의 파일럿 시험, 그리고 한 차례의 필드 시험을 통해 적절한 평가 과제 제작이 가능하다는 것을 보여주었다(Jun et al., 2014; Kim et al., 2012; Kwon et al., 2018; Lee et al., 2015; Lee et al., 2016)(본 특별호의 Jun et al., 2019 참조).

둘째로 평가 추론을 뒷받침하기 위해 시험 수행 조건과 문항 및 검사형의 통계적 특성이 적절하다는 것을 보여야 한다. 개정 TEPS 파일럿 시험과 필드 시험의 성적결과와 수험자 설문 결과를 통해 시험 조건에 해당하는 시험 시간 및 시험 방식이 적절하다는 것을 파악하였고(Kwon et al., 2018), 개정 TEPS 시행 이후에는 시험 환경에 대한 수험자의 만족도를 답안지 뒷면의 설문조사를 통해 TEPS관리위원회에서 지속적으로 관리하고 있다. 그리고 파일럿 및 필드 시험의 문항 분석과 국부의존성 분석을 통해 개정 TEPS 문항의 적절성을 검토하였으며(Kim & Jun, 2017; Kwon et al., 2018), 개정 TEPS가 시행된 이후에도 계속해서 문항 및 검사 통계가 안정적으로 산출된다는 사실을 확인하였다(Lee et al., 2019)(본 특별호의 Lim, 2019 참조).

셋째로 일반화 추론을 뒷받침하기 위해서는 충분한 수의 문항이 포함된 동형(parallel) 검사형들이 제작되어야 하고 척도화와 동등화를 통해 검사형이나 회차에 상관없이 일반화 가능한 점수를 수험자에게 제공해야 한다. 일단 TEPS의 개정을 위해 문항 수를 줄이는 과정에서 시뮬레이션을 통해 적정 신뢰도인 영역별 신뢰도 0.80-0.85 내외와 총점 신뢰도 0.9 이상을 유지하는 선에서 문항 수를 축소하였고(Kim, 2016; Kwon et al., 2018), 모든 문항의 출제와 검토 및 검사형 제작은 상세한 시험 평가틀(test specification)에 따라 이루어지고 있다(TEPS Center, 2019a, 2019b). 또한 매 회차 측정평가 전문가가 표준화된 척도화 및 동등화 절차에 따라 산출한 척도 점수를 수험자에게 제공하고 있다(TEPS Center, 2019c).

넷째로 설명 추론을 뒷받침하기 위해 개정 TEPS 점수가 이론적인 구인에 기인한다는 것을 보여야 한다. 우선 개정 TEPS와 타 표준화 영어시험간의 비교 연구를 통해 개정 TEPS 점수와 대만 Language Training and Testing Center(LTTC)가 개발·시행하는 General English Proficiency Test(GEPT) 점수와의 상관관계가 0.791(총점 기준)로 상당히 높다는 것을 알 수 있었다(Lim et al., 2018). 이전 연구에서도 기존 TEPS 점수와 영국문화원, 호주 IDP, Cambridge Assessment English에서 개발한 International English Language Testing System(IELTS) 점수와의 상관관계가 0.815(총점 기준)로 상당히 높음을 보인 바가 있다(Choi, 2009). 또한 확인적 요인분석을 통해 개정 TEPS의 요인구조가 기존 TEPS의 요인구조와 동일하며 모형 적합성이 적절하다는 사실을 확인하였다(Kwon et al., 2018)(본 특별호의 Lim et al., 2019 참조). 아직 확보되지 않은 근거는 시험 중 수험자의 인지적 과정(cognitive processing)에 대한 연구인데, 시험 후 설문조사, 자극회상(stimulated recall) 인터뷰, 시선추적(eye-tracking) 등과 같은 방법으로 수집할 수 있다. TEPS 점수와

영어 학습 내용 및 기간과의 상관관계도 살펴보아야 하는데, 영어 학습 경험에 대한 정보는 앞으로 수험자 설문조사를 통해 수집할 수 있을 것이다.

다섯째로 확장 추론을 뒷받침하기 위해서는 개정 TEPS 점수가 목표 언어 사용 영역에서의 다른 일반적인 영어능력의 기준들과 연관이 있다는 것을 보여야 한다. 앞서 설명 추론에서 거론된 개정 TEPS와 타 표준화 영어시험간의 비교 연구 결과가 확장 추론의 근거도 된다. 앞으로 수험자의 자기평가, 그리고 수험자의 영어 능력에 대한 교사, 강사 및 고용주의 평가 정보를 수집하여 개정 TEPS 점수와의 상관관계를 구하면 확장 추론을 더 잘 뒷받침할 수 있을 것이다.

여섯째로 활용 추론의 근거로 TEPS 점수가 수험자의 영어숙달도에 대한 판단을 내리기에 유용하며 이해당사자들이 점수의 의미를 쉽게 해석할 수 있다는 사실을 밝혀야 한다. 개정 TEPS 성적표를 개발할 때 수험자에게 더 유용한 정보를 제공하기 위하여 세부능력에 대한 수험자의 수준을 Basic, Intermediate, Advanced 세 단계로 분류하여 추가하였고 세부능력 수준 정보의 안정성에 대한 후속 연구를 수행하였다(Lee et al., 2019). 그리고 기존 TEPS와 개정 TEPS 점수를 상호 호환하여 사용할 수 있도록 점수환산표가 개발되었다(KELTA, 2018). 또한 청해 영역과 독해 영역 점수를 유럽 언어 공통기준(Common European Framework of Reference for Languages; CEFR)과 연계하는 수준설정 연구를 수행하여 점수 사용자가 TEPS 점수의 의미를 더 잘 이해할 수 있도록 하였다(Jun, Lim, & Lee, 2019; Lee et al., 2019; Lim, Jun, & Lee, 2019). 아직 수집되지 않은 정보는 TEPS 점수의 유용성과 해석 용이성에 대한 TEPS 수험자와 이해당사자들의 의견이며, 이는 앞으로 설문조사와 인터뷰를 통해 수집할 수 있을 것이다.

마지막으로 일곱째 결과 합의 추론의 근거로 TEPS 사용의 긍정적인 환류 효과를 밝혀야 한다. TEPS의 환류 효과는 아직 공식적으로 연구된 바가 없으므로 교사, 강사 및 전문가 인터뷰, 그리고 수험자 인터뷰 및 설문조사를 통해 근거를 적극적으로 수집할 필요가 있다.

6. 맺는 말

개정 TEPS가 2018년 5월부터 시행되기까지 수행된 다년간의 연구를 통해 개정 TEPS 점수 해석과 사용의 타당도를 뒷받침할 다양한 근거가 마련되었다. 특히 점수 해석과 사용에 대한 논증 중 점수 해석 부분에 해당하는 영역 정의, 평가, 일반화, 설명 추론에 대한 근거가 상당히 많이 마련되었다. 반면 점수 사용 부분에 해당하는 활용 및 결과 합의 추론에 대한 증거는 상대적으로 적어 추가적으로 수집할 필요가 있다.

개정 TEPS 점수 해석과 사용에 대한 타당도 논증을 더 잘 뒷받침하기 위해 필요한 근거를 모으기 위한 연구계획은 다음과 같다. 우선 설명과 확장 추론의 근거가 될 개정 TEPS와 타 표준화 영어시험과의 비교 연구를 추가로 계획하고 있다. 특히 아시아 지역의 영어시험인 일본 EIKEN과 TEAP, 중국 CET 등의 시험과 비교 연구를 추진하고 있다. 그리고 활용 추론을 더 잘 뒷받침하기 위해서 개정 TEPS의 어휘 영역과 문법 영역을 유럽 언어 공통기준과

연계하기 위한 연구가 계획되어 있다. 또한 앞으로 수험자를 포함한 이해당사자 인터뷰와 설문조사를 통해 수험자의 영어 학습에 대한 정보, 수험자의 영어 능력에 대한 평가 정보, TEPS 점수의 유용성과 해석 용이성에 대한 의견을 수집하고 TEPS의 환류 효과를 연구하여 설명, 확장, 활용, 결과 합의 추론을 뒷받침할 근거를 계속해서 마련하고자 한다. 타당도 논증은 한 번 작성한 것으로 끝이 아니라 계속적인 타당도 근거 수집을 통해 보완할 수 있는 것이다(Chapelle, 2008). 따라서 본 논문에 제시된 TEPS 점수 해석과 사용에 대한 타당도 논증을 앞으로 지속적으로 보완하여 더욱 완성도를 높여 나갈 것이다.

마지막으로 서울대학교 언어교육원에서 개발하고 시행하는 시험으로는 TEPS 외에도 영어 말하기 및 쓰기 시험인 TEPS-Speaking & Writing과 영어능력의 4기능(듣기, 읽기, 말하기, 쓰기)을 통합하여 평가하는 컴퓨터 기반 시험인 i-TEPS, 그리고 중국어, 일본어, 프랑스어, 독일어, 스페인어, 러시아어 시험인 Seoul National University Language Test(SNULT)가 있다. 가까운 시일 내에 이 시험들의 점수 해석과 사용에 대한 타당도 논증도 작성하여 공개한다면 서울대학교 언어교육원이 개발하는 모든 시험들을 더욱 널리 알리고 공신력을 높이는 계기가 될 것으로 기대한다.

References

- AERA, APA, & NCME. (1966). *Standard for educational and psychological testing*. Washington DC: American Psychological Association.
- AERA, APA, & NCME. (1999). *Standard for educational and psychological testing*. Washington DC: American Psychological Association.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson. (Eds), *Building a validity argument for the Test of English as a Foreign Language™* (pp. 319-352). New York: Routledge.
- Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 21-33). New York: Routledge.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language™*. New York: Routledge.
- Choi, I.-C. (2009). *Comparability study on TEPS & IELTS* (Research Report No. 50). Seoul: SNU Language Education Institute.
- Jun, H., Lim, E., & Lee, Y.-W. (2019, March). Enhancing the interpretability and usefulness of TEPS section scores through alignment with the CEFR. Paper presented at the 41st Language Testing Research Colloquium. Atlanta, GA.
- Jun, Y. C., Ryu, D.-S., Park, Y. J., Lee, Y., Shin, S.-H., Jun, H. . . . Byun, J. (2014). *Research for the development of TEPS 2.0* (Research Report No. 77). Seoul: SNU Language Education Institute.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 17-64). Westport, CT: Greenwood Publishing.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- KELTA. (2018). *A study on the compatibility and score conversion between the revised TEPS and other certified language proficiency tests*. Seoul: KELTA.
- Kim, J. Y. (2016). *Reliability and test length* (internal document). Seoul: TEPS Center.
- Kim, J. Y., & Jun, H. (2017). An investigation of local item dependence in testlets and its causes in a large-scale English proficiency test. *Journal of Educational Evaluation*, 30(4), 837-858.
- Kim, M.-W., Jeong, S., Kim, J.-W., Lee, Y.-W., Lee, Y., Shin, S.-H. . . . Lee, G. (2012).

- Basic research for a complete revision of the regular TEPS* (Research Report No. 73). 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. J., So, Y.-S., Lee, Y., & Jun, H. (2016). *Verification of preparation for New TEPS* (Research Report No. 79). Seoul: SNU Language Education Institute.
- Lee, B., Park, Y. J., So, Y.-S., Lee, Y.-J., Kim, C., Jun, H., . . . Yeom, S. (2015). *The development of New TEPS* (Research Report No. 78). Seoul: SNU Language Education Institute.
- Lee, Y.-W., Kwon, H., Lee, Y., Lim, E., Jun, H., Song, M., . . . Min, S. (2019). *Research for the stabilization and validation of the New TEPS* (Research Report No. 81). Seoul: SNU Language Education Institute.
- Lim, E., Jun, H., & Lee, Y.-W. (2019). Standard setting to relate an English reading comprehension test to the CEFR: A comparison of modified Angoff and Bookmark methods. *Journal of Educational Evaluation*, 32(3), 523-548.
- Lim, E., Jun, H., Kim, J., Kwon, H., Wu, R. Y. F., Liao, C., . . . Pan, L. (2018, October). Alignment of the standards-based tests in Asia: TEPS and GEPT comparison. Paper presented at the 5th Asian Association of Language Assessment Conference. Shanghai, China.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). Washington, DC: American Council on Education/Macmillan.
- Seong, T. (2002). *Modern educational evaluation*. Seoul: Hakjisa.
- TEPS Center. (2019a). *TEPS item writer's guide* (internal document). Seoul: TEPS Center.
- TEPS Center. (2019b). *TEPS reviewer's guide* (internal document). Seoul: TEPS Center.
- TEPS Center. (2019c). *TEPS technical report: 2018 administration* (internal document). Seoul: TEPS Center.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge: Cambridge University Press.