

NEW TEPS 와 TOEIC 간 점수 환산 연구

임의진 · 이규민 · 전희성 · 최지원 · 홍유민 · 박나갑*

Lim, Euijin, Lee, Guemin, Jun, Heesung, Choi, Jiwon, Hong, Youmin, & Park, Nagap. (2021). Score comparability of TEPS and TOEIC. *English Teaching*, 76(4), 143-163.

Scores from TEPS (Test of English Proficiency developed by Seoul National University) and TOEIC (Test of English for International Communication) are widely used in Korea for various purposes, including employment, admission, and certification. However, there is more than one score conversion relationship between these two tests, causing confusion among test takers. Moreover, the conversion relationships were established in 2004 and 2011 and have not been updated since then. With this background, a large-scale score comparability study was undertaken to develop a new score conversion table for TEPS and TOEIC. Data were collected from a total of 1,415 test takers using a single group design. Log-linear presmoothing and equipercentile methods were used for linking, with cross validation across subgroups. Results showed that the old conversions were disadvantageous to a certain test-taker group. A new conversion table between the two tests is then suggested.

Key words: concordance table, linking, TEPS, TOEIC/환산표, 연계화, 텡스, 토익

This work was funded by the TEPS Council, Seoul National University Foundation (TC202001).

*First Author: Euijin Lim, Professor, Department of Education, Chinju National University of Education

Corresponding Author: Guemin Lee, Professor, College of Educational Sciences, Yonsei University; 50 Yonsei-ro Seodaemun-gu, Seoul 03722, Korea; Email: guemin@yonsei.ac.kr

Co-Authors: Heesung Jun, Researcher, Test Research and Development Department, TEPS Center; Jiwon Choi, Researcher, Measurement and Statistical Research Department, TEPS Center; Youmin Hong, Graduate Student, College of Educational Sciences, Yonsei University; Nagap Park, Graduate Student, College of Educational Sciences, Yonsei University

Received 15 September 2021; Reviewed 19 October 2021; Accepted 21 December 2021



© 2021 The Korea Association of Teachers of English (KATE)

This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits anyone to copy, redistribute, remix, transmit and adapt the work, provided the original work and source is appropriately cited.

1. 서론

TEPS(Test of English Proficiency developed by Seoul National University)와 TOEIC(Test of English for International Communication)은 영어 능력을 평가하기 위해 한국에서 널리 사용되고 있는 표준화 검사로, 그 사용 목적은 대학교 입학, 편입학, 졸업 인증, 대학원 입학 등 학문적인 분야뿐만 아니라 기업 및 공공기관 채용, 인사 고과 평가, 해외 파견 근무자 선발 등 업무 상황까지 아우르고 있다. 당초에 두 시험은 검사 점수가 이렇듯 광범위하게 사용될 것을 고려하지 않은 채 각각 독립적으로 개발되었으나 현실적으로 오늘날 두 검사 모두 다양한 상황에서 사용되고 있다.

두 검사는 영어 능력을 평가한다는 광의의 목적은 같지만 검사 구성, 점수 체제 등이 서로 다르다. 따라서 한 기관에서 두 검사 점수를 함께 사용하기 위해서는 점수 환산 기준이 반드시 필요하다. 이렇듯 서로 같은 검사는 아니지만 그 구인과 사용 목적에 유사성이 있는 검사 점수 간 환산표(concordance table)를 산출하기 위해서는 연계화(linking) 연구가 수행되어야 한다. 그리고 산출된 환산표는 서로 다른 두 검사의 점수를 연계하여 사용하고자 하는 실용적인 목적을 위해 활용된다. 예를 들어 미국의 대학수학능력평가라고 할 수 있는 ACT Assessment와 SAT I Verbal-plus-Mathematics 검사는 서로 다른 검사 명세화에 따라 개발되고 측정하는 구인도 다르지만 대학에 진학하고자 하는 학생의 준비도를 측정한다는 공통의 목적을 위하여 개발되었고, 대학에서의 학생 선발, 장학금 지원 등의 의사결정을 위해 함께 사용된다. 따라서 두 검사 점수를 비교 가능하게 하려면 연계화를 실시할 필요가 있다 (Schneider & Dorans, 1999). 이를 위해 Dorans, Lyu, Pommerich와 Houston (1997)은 두 검사를 모두 치른 단일집단의 수험자들을 대상으로 동백분위 연계화(equipercentile linking)를 실시하여 두 검사 간의 환산 관계를 산출하였다.

TEPS와 TOEIC 또한 두 검사 점수를 서로 비교하며 사용할 수 있도록 해주는 환산 기준이 있다. 그러나 현재 두 검사 점수의 환산에는 두 가지 문제점이 있다. 첫번째는 두 종류의 환산 기준이 원칙 없이 혼용되고 있다는 것이고, 두번째는 환산 기준이 수립된 지 오랜 시간이 지났다는 것이다. 현재 TEPS와 TOEIC 간 점수 환산 기준은 두 종류가 있다. 인사혁신처 등 정부 부처와 그 산하 기관에서 사용하고 있는 TEPS-TOEIC 간 점수 환산 기준, 즉 TOEIC 700점이 TEPS 340점과 같다고 보는 기준은 2004년도 사법시험 및 군법무관 임용시험에서 도입된 이래로 그 타당성에 대한 검증 없이 지금까지 이어져 오고 있다. 또한 한국전력공사, 인천국제공항공사, 한국석유공사, 국민연금공단 등 일부 공공기관과 공기업에서는 2011년 서울대학교 통계연구소에서 수행한 환산 연구를 기반으로 TOEIC 700점이 TEPS 300점과 같다는 기준을 적용하고 있다. TEPS와 TOEIC 환산에 서로 다른 두 가지 기준이 함께 사용된다는 것은 동일한 TEPS 점수가 사용 기관에 따라 서로 다른 TOEIC 점수로 환산된다는 것이다. 이는 환산 기준의 신뢰성을 의심하게 하는 중차대한 문제라고 할 수 있다.

뿐만 아니라 현재 사용되는 두 종류의 환산 기준 모두 수립된 지 오랜 시간이 흘러

그 정확도에 의문이 제기된다. 정부 부처 및 그 산하 기관의 TEPS-TOEIC 간 환산 기준은 17년 전에 수립된 이후 현재까지 그대로 사용되고 있고, 2011년에 수행된 환산 연구도 10년이 경과되었다. 서로 다른 두 검사의 점수 환산 관계는 검사의 특성이 바뀌지 않더라도 검사 모집단의 변화에 따라 환산 관계가 바뀌기 쉽다. 5-10년의 기간이 지나면 모집단 변동(population shift)으로 인해 기존의 환산 기준에 의문이 제기된다(Kolen & Brennan, 2014). 만약 연계화하고자 하는 검사들 중 한 검사가 새롭게 개정되거나 척도 변환을 한 경우, 보다 정확한 환산 관계를 적용하기 위해 더 짧은 주기로 환산 기준을 변경하기도 한다. 예를 들어 SAT와 ACT의 경우 2016년에 SAT가 개정되면서 같은 해에 한 차례 SAT-ACT 환산표를 발표하였으나 2년 후인 2018년에 다시 한 번 변경된 SAT-ACT 환산표로 기존의 환산 기준을 대체하였다(College Board, 2018). TEPS는 2018년에 개정되며 (구)TEPS보다 문항 수가 줄어들고 시험 시간도 단축되었으며 점수 체제가 바뀌는 등 많은 변화가 있었다. TOEIC 또한 2016년에 새로운 문항 유형을 도입하고 파트별 문항 비율을 조정하는 등 개정을 거쳤다(Educational Testing Service, 2019). 그러나 현재 사용되고 있는 TEPS-TOEIC 환산 기준은 두 검사가 개정되기 이전에 구축되었다. 즉, 개정된 TEPS와 TOEIC 간의 환산 연구를 수행하지 않고 10년 이상 경과한 기존의 환산 기준에 (구)TEPS-TEPS 점수 환산을 적용한 2단계 과정을 거친 결과이다. 시간이 지나면서 발생한 두 검사 모집단의 변동과 검사 개정으로 인한 변화를 고려했을 때, (구)TEPS-TEPS 환산을 거친다 하더라도 (구)TEPS-TOEIC에 대한 환산 기준을 그대로 승계하여 사용하는 것은 근본적인 한계가 있다. 이러한 이유로 TEPS와 TOEIC 간 직접적인 점수 환산을 가능하게 하는 연구가 필요하다.

본 연구에서는 TEPS-TOEIC 환산표를 산출하기 위해 연계화를 수행하였다. 이를 위해 표본자료에 대한 기초분석을 실시하여 표본의 대표성을 확인하고, 하위 집단에 따라 연계화 결과를 교차 검증하여 연계화 방법을 타당화한 후 TEPS와 TOEIC을 모두 치른 단일집단 수험자의 성적 자료를 바탕으로 연계화를 실시하여 환산표를 산출하였다.

2. 문헌 연구

2.1. TEPS와 TOEIC

TEPS는 영어를 외국어로 사용하는 학습자의 실질적 영어 활용 능력을 평가하기 위해 서울대학교 언어교육원에서 개발한 영어 능력 평가시험이다. TEPS는 응시자의 영어 능력을 정확하게 측정하고, 올바른 외국어 학습 방향을 제시하는 것을 목적으로 개발되었다(Lee, Ban, Lee, & Jung, 2020; TEPS Council, 2020a). TEPS는 1999년 제1회 정기시험이 시행된 이후 20년간 다양한 영역에서 활용되었다. 그러나 시간이 지남에 따라 의사소통 방식 및 영어 사용 상황의 변화, 영어 시험에 대한 응시자들의 요구,

영어 교육 및 평가 분야의 연구 성과 등을 반영할 필요성이 꾸준히 제기되어 왔다(Lim, Jun, Yun, & Min, 2019). 이에 개정 연구를 거쳐 2018년 5월 제248회 정기시험 이후부터 개정된 TEPS가 시행되고 있다. 개정된 TEPS는 (구)TEPS와 비교했을 때 문항 수를 200문항에서 135문항으로 축소하였고, 시험 시간은 140분에서 105분으로 단축하였다. 일부 새로운 문항 유형을 도입하고 검사 구성을 변경하였으며 총점의 점수 척도 또한 기존 10-990점 척도에서 0-600점 척도로 변화하였다(TEPS Council, 2020b). TEPS는 청해, 어휘, 문법, 독해, 총 네 영역으로 구성되어 있으며, 영역별 점수의 합산인 총점을 기준으로 수험자의 능력을 10개 등급으로 구분하고 등급에 따른 설명을 제공한다(TEPS Council, 2020a).

TOEIC은 영어가 모국어가 아닌 사람들을 대상으로 의사소통 능력에 중점을 두고 일상생활 및 국제업무 등에 필요한 실용영어 능력을 평가하기 위해 미국 ETS(Educational Testing Service)에 의해 개발된 영어 능력 평가시험이다. 1979년 ETS에 의해 개발된 이래 TOEIC은 전 세계 160개 이상의 국가에서 해마다 약 700만 건 이상 시행되며, 14,000여 개의 기관 및 업체에서 선발과 승진 및 해외 파견 등의 목적으로 활용하고 있다. 우리나라에는 1982년에 도입된 이래 현재까지 범용 영어 평가 도구로 활용되고 있다(Lee et al., 2020; TOEIC, 2020a). 현행 TOEIC은 듣기와 읽기의 두 영역으로 이루어져 있으며, 영역별로 100문항씩 총 200문항으로 구성되어 있다. 검사 시간은 듣기 45분, 읽기 75분으로 총 120분이고, 정답의 수에 따른 각 영역별 점수와 그 합산인 총점을 제공한다. 총점은 5점 간격 10-990점 척도로 제공된다(TOEIC, 2020a). 또한 영역별 점수에 따라 듣기 영역에서는 3등급, 읽기 영역에서는 4등급으로 수험자를 구분하여 등급 설명을 제공한다(TOEIC, 2020b).

TEPS는 공인 영어 능력 평가로써 대학교 입학, 편입학, 졸업 인증, 대학원 입학, 기업 및 공공기관 채용, 인사 고과 평가, 해외 파견 근무자 선발 등의 기준 점수로 활용되고 있다. 또한 TEPS는 행정고시 등 주요 국가 고시와 각종 자격증 시험에서 영어 시험을 대체하는 자격시험으로 사용되고 있다(TEPS Council, 2020c). TOEIC은 기업, 공공/정부 기관의 채용에 활용되고 있으며, 일부 국가 자격시험에서는 TOEIC 검사 점수로 영어 및 외국어 과목을 대체할 수 있도록 하고 있다. 뿐만 아니라, 일부 대학과 대학원에서는 입학 및 편입학, 학점 인정, 장학금 지급, 해외연수/인턴/교환학생 선발, 졸업 인정 등에 활용하고 있다(TOEIC, 2020c). 국가 고시 및 자격증 시험, 기관 채용 요건 등에서 TEPS와 TOEIC을 최저 기준 점수로 활용하는 주요 사례는 다음의 표 1과 같다. 신뢰할 수 있는 단일 환산표의 부재로 인한 혼란은 표 1에도 나타난다. 예를 들어 7급 외무공무원 시험과 변리사 자격증 시험에서 요구하는 TEPS 점수는 385점으로 동일하지만, TOEIC 점수는 각각 790점과 775점으로 다르게 나타난다. 또한 5급 및 7급 공무원 시험, 법원행정고등고시, 입법고등고시 등은 TOEIC 700점과 같은 의미라고 인정되는 TEPS 점수가 340점이라고 공지했지만, 국민연금공단, 한국전력공사 등은 TEPS 300점을 제시했다. 더 심각한 예는 TOEIC 800점이 TEPS로 환산된 결과를 살펴보면 발견할 수 있다. TOEIC 800점과 같은 의미를 갖는 점수로 호텔경영사 자격증은 TEPS

404점을 제시했지만 인천국제공항공사는 TEPS 348점을 제시했고, 이는 관광통역안내사 자격증 시험에서 TOEIC 760점과 같다고 보는 TEPS 372점보다 오히려 더 낮은 점수이다.

TABLE 1

Examples of TEPS and TOEIC Score Uses

Examinations/Employment Qualifications	TEPS	TOEIC
Grade 5 diplomat candidate selection examination	452	870
Hotel general manager licensure examination	404	800
Incheon international airport corporation (employment qualification)	348	
Grade 7 foreign service consul selection examination		790
Certified patent attorney	385	775
Tourist guide licensure examination	372	760
Grade 5 and grade 7 civil servant selection examination		
Court administration examination		
Legislative civil service examination		
Certified appraiser licensure examination	340	
Certified public labor attorney licensure examination		700
Certified public accountant licensure examination		
Certified tax accountant licensure examination		
Actuary licensure examination		
National pension service (employment qualification)	300	
Korea electric power corporation (employment qualification)		

Source: TEPS Council, 2020c; TOEIC, 2020c

이러한 혼란은 현재 두 가지 TEPS-TOEIC 간 점수 환산 기준이 혼재되어 활용되고 있는 데 기인한다. 첫 번째 기준은 2004년 사법시험 및 군법무관 임용시험에서 최초로 영어 과목을 외부시험으로 대체하면서 TOEIC 700점이 (구)TEPS 625점과 같다고 보았고, 이후 행정고시, 외무고시 및 국가 자격시험에서 이를 준용하여 사용한 것에서 시작되었다(Ministry of Justice, 2004). 2018년에 TEPS가 개정된 후에는 (구)TEPS와 개정된 TEPS 점수를 연계하였고, 그 결과를 바탕으로 (구)TEPS 625점이 TEPS 340점으로 변환되어 TOEIC 700점이 TEPS 340점과 같은 의미로 사용되는 관행이 지금까지 이어져 내려왔다(Korea English Language Testing Association, 2018). 마찬가지로의 방법으로 TOEIC 790점이 TEPS 385점과 같고, TOEIC 870점이 TEPS 452점과 같다는 기준이 산출되어 지금까지 이어져 주요 국가고시 응시자격 등으로 활용되고 있다. 현재 활용되고 있는 두 번째 환산 기준은 2011년 서울대학교 통계연구소에서 수행한 (구)TEPS-TOEIC 환산 연구를 기초로 하고 있다(TEPS Council, 2020d). 2011년 연구에서는 (구)TEPS와 TOEIC을 모두 치른 수험자 집단의 성적 자료를 수집하고 동백분위 연계화 방법을 적용하여 환산표를 산출하였다. 다만 당시 산출된 환산표는 (구)TEPS → TOEIC 일방향으로 산출되어 TOEIC → (구)TEPS 환산은 단일한 점수로 제시되지 않는다. 이 연구 결과는 이후 (구)TEPS-TEPS 연계를 거쳐 인천국제공항공사, 국민연금공단, 한국전력공사 등에서 사용하는 환산 기준, 즉 TOEIC 700점이 TEPS 300점과 같고, TOEIC 800점이 TEPS 348점과 같다는 근거로

활용되었다. 이러한 혼란이 종식되고 어학시험 수험자들에게 점수 환산의 유효리가 발생하지 않기 위해서는 대규모 표본을 수집하여 신뢰할 수 있는 단일한 환산 기준을 산출할 필요가 있다.

2.2. 연계화

검사는 개인의 지능이나 적성, 학업성취도 등 인간의 잠재 능력(latent trait)을 측정하기 위하여 개발되는 측정 도구이다(Cronbach, 1990; Ebel & Frisbie, 1991; Kim, Lee, Ban, & Cha, 2004; Linn & Gronlund, 2000). 문항 보안 등의 문제로 하나의 검사를 반복하여 사용할 수 없는 경우, 검사 제작 기관은 같은 측정 목적을 가진 검사이지만, 서로 다른 문항으로 구성된 여러 개의 검사형(test form)을 개발하여 사용할 수 있다. 비록 검사형이 동일한 내용 및 통계적 명세화에 따라 제작된다고 하더라도 각기 다른 문항을 포함하고 있기 때문에 검사형 간 난이도에 차이가 있을 수 있고, 이를 보정하여 상호 호환하여 사용할 수 있도록 통계적 과정인 동등화(equating)를 수행한다(Kolen & Brennan, 2014).

동등화의 대상이 되는 검사형들은 기본적으로 같은 구인을 측정해야 한다. 만일 두 검사가 서로 다른 구인을 측정한다면 동등화를 적용할 수 없고, 대신 연계화를 통해 검사 결과의 비교 가능성을 확보할 필요가 있다(Kim et al., 2004). 연계화란 한 검사 결과를 다른 검사 결과와 비교할 수 있도록 만들어주는 과정으로, 유사한 구인을 측정하지만 서로 다른 검사에서 나온 결과를 비교하고자 하는 다양한 요구에 부응하기 위해 발전된 측정학적 접근이다(Kim et al., 2004; Linn, 1993; Mislevy, 1992).

Mislevy(1992)와 Linn(1993)의 분류에 따르면 연계화의 유형에는 동등화, 추정(calibration), 통계적 조정(statistical moderation), 예측(prediction), 사회적 조정(social moderation)이 있다. 즉, 서로 같은 검사 명세화를 기반으로 제작되어 동일한 구인을 측정하는 검사형 간의 점수를 통계적으로 조정하는 동등화는 연계화의 가장 엄격한 형태라고 할 수 있다. 추정이란 동일한 구인을 측정하지만 검사 길이가 서로 다르거나, 혹은 서로 수준이 다른 집단에서 검사를 시행하는 경우에 검사 점수를 상호 비교 가능하게 하는 연계 방법이다. 통계적 조정은 다양한 검사로부터 얻은 검사 결과를 기준 검사를 사용하여 비교 가능한 점수 체제로 조정하는 방법으로, 두 검사가 동일한 구인을 측정한다는 사실을 입증하기 어려운 경우에 사용한다. 예측은 측정하고자 하는 잠재적 특성, 혹은 검사 점수 간의 상관관계를 기초로 한 검사 점수를 이용하여 다른 검사 점수를 예측하는 것으로, 이때 점수 예측은 양방향성이 아닌 일방향으로 이루어진다. 마지막으로 사회적 조정은 전문가들의 합의(consensus moderation)를 기준으로 서로 다른 검사 점수를 비교하는 방법이다(Kolen & Brennan, 2014). TEPS와 TOEIC은 영어를 외국어로 사용하는 학습자의 영어 능력을 평가한다는 점에서 유사한 목적을 갖지만 세부적인 구인, 검사 구성 및 문항 유형 등이 다르기 때문에 동일한 구인을 측정한다고 보기는 어렵다. 또한 양방향으로 점수 환산이 가능해야 하므로 통계적 조정을 통해 연계화 관계를

산출할 수 있다.

통계적 조정으로 연계화를 수행하기 위해서는 엄밀한 자료 수집 설계를 통해 검사로 인해 발생하는 차이와 수험자 특성으로 인해 발생하는 차이를 분리할 수 있어야 한다. 자료 수집 설계는 크게 단일집단 설계(single group design), 무선집단 설계(random groups design), 공통문항 비동등집단 설계(common-item nonequivalent groups design)로 구분할 수 있다. 단일집단 설계는 동일한 수험자 집단이 두 검사를 모두 치르게 하는 것으로, 수험자 집단이 동일하므로 발생한 점수 차이가 곧 검사 특성으로 인한 차이라고 파악할 수 있다. 무선집단 설계는 수험자들을 각 검사에 무선할당하여 동등한 집단을 구성하는 방법이다. 각 검사를 치른 두 집단은 서로 다른 수험자로 구성되어 있지만 무선할당 절차를 통해 집단 간 동등성을 확보했으므로 발생한 점수 차이는 검사 간 차이로 인한 것이라고 볼 수 있다. 공통문항 비동등집단 설계는 두 검사에 공통문항을 포함하여 각 검사를 치른 두 비동등집단이 얼마나 능력 차이가 있는지 파악하고, 이를 반영하여 검사로 인한 차이를 규명하는 방법이다. TEPS와 TOEIC은 서로 문항 유형이 다르므로 공통문항을 포함하기 어렵고, 무선집단 설계는 더 많은 수험자 표본을 요구하므로 본 연구에서는 단일집단 설계를 적용하였다.

통계적 조정을 위해 사용될 수 있는 연계화 방법은 전통적 방법과 문항반응이론(item response theory)을 이용한 방법으로 구분할 수 있다. 많은 경우 연계화 상황에서 서로 다른 두 검사의 문항 반응을 확보하기 어려우므로 문항반응이론을 이용한 방법보다는 선형 연계화(linear linking), 동백분위 연계화 등 검사 총점을 기반으로 하는 전통적 방법을 많이 사용한다. 선형 연계화 방법은 연계화 후 두 검사 점수가 같은 평균과 표준편차를 갖도록 한 검사 점수를 선형 변환하는 방법이다. 동백분위 연계화 방법은 연계화 후 두 검사 점수의 분포가 같은 형태가 되도록 동백분위 연계화 함수를 사용하여 점수를 비선형 변환하는 방법이다. 단일집단 설계에서의 동백분위 연계화 함수는 다음과 같다.

$$\begin{aligned} e_Y(x) &= \frac{\frac{P(x)}{100} - G(y_U^* - 1)}{G(y_U^*) - G(y_U^* - 1)} + G(y_U^* - 0.5), & 0 \leq P(x) < 100, \\ &= K_Y + 0.5, & P(x) = 100. \end{aligned}$$

위의 식에서 검사 X 와 Y 의 검사 점수는 각각 x 와 y 로 나타낸다. $e_Y(x)$ 는 검사 X 의 점수를 검사 Y 척도로 연계하는 연계화 함수이다. $P(x)$ 와 $G(y)$ 는 각각 검사 X 와 Y 점수의 백분위를 의미하며, y_U^* 는 특정 백분위 P^* 보다 큰 누적 백분위 $[100 G(y)]$ 를 만족시키는 최소 정수를 가리킨다. K_Y 는 검사 Y 의 총점을 의미한다. 동백분위 연계화를 시행하면 두 검사에서 같은 백분위에 위치했을 때 점수가 호환될 수 있다고 연계한다. 즉, 두 검사 점수의 분포가 같아지게 된다(Kolen, 1988; Kolen & Brennan, 2014).

완곡화(smoothing)는 동백분위 방법을 적용했을 때 표집 오차를 줄이기 위해

적용하는 수리적인 기법이다. 연계화에서 발생할 수 있는 오차는 무선 오차와 체계적 오차로 구분할 수 있다. 이 중 무선 오차는 모집단 전체가 아닌 표본을 사용함으로써 발생하는 표집 오차로, 완곡화는 수리적 모형을 적용함으로써 모집단에서의 연계화 관계를 보다 정밀하게 추정하고자 하는 절차이다(Cope & Kolen, 1990; Kolen, 1984). 완곡화는 적용 시점에 따라 연계화 이전에 두 검사의 점수 분포를 완곡화하는 사전완곡화(presmoothing)와 연계화 후 두 검사 점수의 관계를 완곡화하는 사후완곡화(postsmoothing)로 구분할 수 있다. 사전완곡화는 표본의 점수 분포에서 관찰되는 고르지 못한 양상이 표집 오차에서 기인한다고 보고 로그 선형(log-linear) 함수나 4모수 베타 이중분포 모형(four-parameter beta compound binomial model)을 가정하여 이를 모집단에서 기대되는 완만한 곡선으로 모형화하는 것이고, 사후완곡화는 연계화 후 두 검사 점수 간의 관계가 고르지 못한 양상을 보일 때 이를 3차 스플라인(cubic spline) 함수 등을 이용해 완만한 곡선으로 모형화하는 것이다(Fairbank, 1987; Hanson, Zeng, & Colton, 1994; Kolen & Brennan, 2014).

본 연구에서 적용하고자 하는 로그 선형 사전완곡화 방법은 다음과 같은 다항 함수로 나타낼 수 있다.

$$\log[N_x f(x)] = \omega_0 + \omega_1 x + \omega_2 x^2 + \dots + \omega_c x^c$$

이 때 N_x 는 표본 크기를 나타낸다. 즉, 각 점수 x 의 빈도 $N_x f(x)$ 에 로그를 취한 값을 다항식으로 표현하는 것이다. 이때 우변의 $\omega_0, \omega_1, \omega_2, \dots, \omega_c$ 는 최대우도추정방법(maximum likelihood estimation)으로 추정된다. C 는 다항식의 차수를 나타내는데, 로그 선형 함수로 모형화한 분포는 원래의 표본 분포와 C 번째까지의 적률이 일치한다는 특성이 있다(Kolen & Brennan, 2014). 즉, C 가 4인 경우 두 분포가 1차 적률부터 4차 적률까지 일치하므로 같은 평균, 표준편차, 왜도, 첨도를 갖게 된다. 따라서 차수가 높아질수록 로그 선형 완곡화를 수행한 결과가 원래의 표본 분포에 가까워진다. 다항식의 차수가 너무 높으면 완곡화 실시 후에도 무선 오차를 거의 줄일 수 없고, 차수가 너무 낮으면 체계적 오차를 과도하게 도입하게 되므로 적절한 완곡화의 수준을 결정하는 것이 중요하다. 모형의 적합도를 비교하기 위해 카이제곱 검정, 카이제곱 차이 검정, AIC(Aikake information criterion; Akaike, 1981) 통계치 비교 등의 방법을 적용할 수 있다. 본 연구에서 사용한 카이제곱 차이 검정은 차수가 인접한 모형 간 카이제곱의 우도비를 비교하여 그 차이가 통계적으로 유의한지 검정하는 방법이다. 만약 그 차이가 유의하다면 한 차수를 높이는 것이 적합도를 의미 있는 수준으로 개선한다고 보고 높은 차수의 모형을 선택한다. 만약 차이가 유의하지 않다면 더 간결한 낮은 차수의 모형을 택하게 된다.

3. 연구 방법

3.1. 연구 대상

이 연구에서는 단일집단 설계 모형을 적용하여 TEPS와 TOEIC을 모두 치른 응시자 표본자료를 수집하였다. TEPS와 TOEIC은 동등화를 통해 검사형 간 난이도 차이를 보정한다고 밝히고 있으므로 서로 다른 일자에 시행된 검사 점수를 호환하여 사용하였다(Educational Testing Service, 2021; TEPS Council, 2021). 4차례에 걸쳐 총 1,415명의 자료를 수집하였고, 회차별 자료 수집 조건 및 인원은 다음 표 2와 같다.

TABLE 2
Data Collection

Group	Conditions	N
1st	Examinee who has valid TEPS and TOEIC score reports between January and June, 2020 - Compensation: Free TEPS voucher	501
2nd	Examinee who has a valid TOEIC score report between May and July, 2020 and has not taken TEPS before - Compensation: 30,000 won	190
3rd	Examinee who has valid TEPS and TOEIC score reports between July and September, 2020 - Compensation: Free TEPS voucher	393
4th	Examinee who has valid TEPS and TOEIC score reports between January and October, 2020 - Compensation: Free TEPS voucher	331
	Total	1,415

1-4차 집단 중 1, 3, 4차 집단의 경우 TEPS SNS 계정 및 홈페이지 공지를 통해 TEPS 성적 보유자를 대상으로 TOEIC 성적표를 수집하였다. 이때 TOEIC 성적을 둘 이상 보유한 수험자는 가장 최근 성적을 제출하도록 하였고, 응시자가 직접 스캔한 TOEIC 성적표를 업로드하면 그 진위 여부를 확인하여 검증 절차를 거쳤다. 그리고 수집된 TOEIC 성적의 응시 날짜와 가장 가까운 날짜에 응시한 TEPS 성적을 추출하는 방식으로 자료를 구축하였다. 반면에 2차 집단은 TOEIC 고사장에서 안내문을 배부하고 TOEIC 응시자가 많을 것으로 예상되는 취업 준비 사이트에 공지하여 기존에 TEPS를 치른 적 없는 TOEIC 응시 집단을 대상으로 TOEIC 성적표를 수집한 후 TEPS 정기시험을 치르게 하였다. 이는 연계화 표본을 수집함에 있어서 TEPS와 TOEIC 응시자 모집단의 특성이 다를 수 있다는 점을 고려하여 대표성 있는 표본을 구성하기 위한 것으로, 1, 3, 4차 집단과 2차 집단의 특성이 다소 상이하다고 할 수 있다. 많은 기관에서 자격기준으로 사용하는 영어 능력 검사의 특성상 검사를 반복해서 응시하는 응시생이 많으나 본 연구에서는 자료 수집 대상을 이전에 자료 수집에 참여하지 않은 사람으로 한정하여 동일한 응시자의 성적 자료를 2회 이상

반복 수집하지 않도록 하였다. 이러한 절차를 통해 응시자 집단의 성별, 연령, 소속, 학력, 전공을 배경 변수로 수집하였고, TEPS와 TOEIC 총점, 영역별 점수, 응시 일자를 확보하였다.

3.2. 분석 방법

본 연구에서는 수집된 자료의 특성을 확인하기 위해 응시자 배경 변수의 분포를 확인하고, TEPS와 TOEIC 총점 및 영역별 점수에 대해 집단별 기초분석을 실시하였다. 또한 단일집단 설계 적용의 적절성을 확인하기 위해 수험자들의 TEPS와 TOEIC 응시 일자 차이를 분석하였다. 동일 수험자라 하더라도 두 시험을 치른 응시 일자 차이가 크면 그 사이에 영어 능력 변화가 있을 수 있어 TEPS-TOEIC 간 점수 환산에 오차 요인으로 작용할 수 있기 때문이다. 이러한 기초분석에 활용한 프로그램은 SPSS 24.0(IBM Corp., 2016)이다.

다음으로는 집단에 따른 환산 결과의 일관성을 알아보기 위해 전체 집단을 두 집단으로 무선할당하여 연계화 결과를 교차 검증하였다. 마지막으로 전체 집단을 대상으로 TEPS-TOEIC 연계화를 수행하여 환산표를 산출하였다. 연계화를 위해서는 로그 선형 사전완곡화 후 동백분위 연계화 방법을 적용하였고, 이를 위해 *Equating Recipes*(Brennan, Wang, Kim, & Seol, 2009)가 사용되었다.

4. 연구 결과

4.1. 기초분석

본 연구에서는 표본의 특성을 확인하기 위하여 전체 1,415명의 표본 응시자 배경 변수에 대한 빈도분석을 실시하였다. 더불어 표본 집단이 가장 많이 응시한 TEPS, TOEIC 회차를 기준으로 실제 두 시험의 응시 집단 분포와 표본 집단 분포를 비교하여 표본의 대표성을 확인하였다(TEPS Council, 2020e; TOEIC, 2020e). 자료 수집 기간인 2020년 1월부터 10월 사이에 표본 집단이 가장 많이 응시한 TEPS 회차는 제291회(2020년 9월 5일 시행) 시험이고, 표본 집단이 가장 많이 응시한 TOEIC 회차는 제405회(2020년 5월 31일 시행) 시험이었다. 먼저 표본 응시자들의 배경 변수 빈도분석 결과는 다음의 표 3과 같다.

TABLE 3
Frequency Distribution of the Background Variables

Variables	Categories	N	Percentage
Gender	Male	555	39.2
	Female	860	60.8
Status	Middle school	1	0.1
	High school	5	0.4
	Community college	9	0.6
	University	607	42.9
	Graduate school	74	5.2
	General	452	31.9
	No answer	267	18.9
Education level	Elementary school	3	0.2
	Middle school	4	0.3
	High school	505	35.7
	Community college	14	1.0
	University	473	33.4
	Graduate school	100	7.1
	Others	17	1.2
No answer	299	21.1	
Major	Business/Economics	137	9.7
	Engineering	282	19.9
	Education	58	4.1
	Law	42	3.0
	Social science	109	7.7
	Art/Music/Sports	21	1.5
	Medicine/Pharmacy	96	6.8
	Liberal arts	124	8.8
	Natural science	204	14.4
	Others	48	3.4
No answer	294	20.8	
Age	10-19	9	0.6
	20-29	1,051	74.3
	30-39	280	19.8
	40-49	65	4.6
	50-59	8	0.6
	60-69	1	0.1
70-79	1	0.1	
Total		1,415	100.0

표 3에 따르면 표본 중 여성이 60.8%로 39.2%의 남성에 비해 많았다. 제291회 TEPS 응시자 분포에 따르면, 해당 회차 응시자 중 남성은 51.4%, 여성은 48.6%로 나타나 본 연구의 표본 집단에 비해 남성 비율이 다소 높은 것으로 나타났다(TEPS Council, 2020e). 제405회 TOEIC 응시자 분포에 따르면 전체 TOEIC 응시자 중 남성은 41%, 여성은 59%로 나타나 본 연구의 표본 집단은 실제 TOEIC 응시자 분포와 유사한 성별 분포를 보였다. 전체 1,415명의 표본 수험자 중 대학생이 전체의 42.9%로 가장 많았고, 일반 수험자가 31.9%로 두 번째로 많았다. 제291회 TEPS 응시자 분포에 따르면 대학생이 44.3%로 가장 많고 일반 응시자가 37.9%로 뒤를 이어, 본 연구의 표본은

실제 TEPS 수험자 집단의 소속과 대체로 유사한 분포를 따르고 있음을 확인하였다. 또한 표본 수험자의 학력 분포를 살펴봤을 때 고등학교 졸업자가 35.7%로 가장 많았고, 그 다음은 대학교 졸업자가 33.4%를 차지했다. 수험자 표본의 전공 분포를 살펴보면 공학 전공자가 응답자의 19.9%로 가장 많았고 그 다음은 자연과학 전공 14.4%, 경제/경영 전공 9.7%, 인문학 전공 8.8% 순이었다. 표본 응시자의 연령은 10대부터 70대까지 넓게 분포했고, 20대가 74.3%, 30대가 19.8%로 가장 많았다. 제291회 TEPS 응시자 분포에 따르면, 20대가 68.4%로 가장 많았으며 30대가 14.8%로 그 뒤를 이었다. TOEIC의 경우 응시자 연령 분포를 연도별로 제공하고 있는데, 가장 최근에 제공된 2019년 응시자 연령 분포에 따르면 2019년 전체 TOEIC 응시자 중 20대가 77%로 가장 많았고, 두 번째는 30대로 13%를 차지했다(TOEIC, 2020d). 이를 통해 본 연구에서 수집한 표본의 연령 분포가 실제 TEPS 및 TOEIC 수험자 집단의 연령 분포와 유사함을 알 수 있다.

표본 자료의 점수 분포 특성을 파악하기 위해 표본 응시자의 TEPS와 TOEIC 점수에 대한 기술통계 분석을 실시한 결과는 다음의 표 4와 같다. 표 4에 따르면 표본에서 TEPS 총점은 76에서 598점의 범위를 가지고 있으며, 평균은 348.96, 표준편차는 90.66으로 나타났다. TOEIC 총점은 190에서 990점의 범위를 가지고 있으며 평균은 825.78, 표준편차는 141.11로 나타났다.

TABLE 4
Descriptive Statistics for TEPS and TOEIC Scores

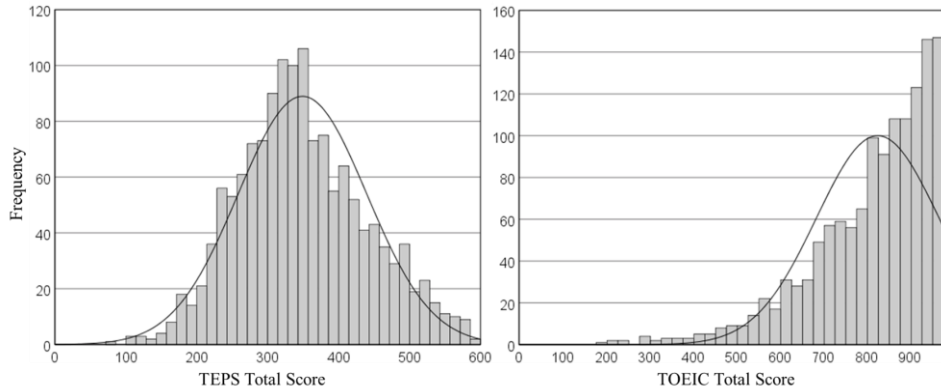
Test	Min	Max	<i>M</i>	<i>SD</i>
Total	76	598	348.96	90.66
TEPS				
Listening comprehension	0	240	135.93	37.35
Vocabulary	0	60	37.60	9.38
Grammar	0	60	38.66	10.01
Reading comprehension	9	240	136.77	42.61
TOEIC				
Total	190	990	825.78	141.11
Listening comprehension	105	495	431.36	65.47
Reading comprehension	80	495	394.42	82.04

TEPS와 TOEIC 총점에 대한 빈도분석 결과는 그림 1에 히스토그램으로 제시되었다. 그림 1에 나타난 바와 같이 동일한 표본 집단에서 TEPS 총점은 정규분포에 가까운 형태로 분포하는 반면, TOEIC은 표본의 수험자의 절반 이상이 TOEIC 800점 이상 범위에 분포하여 부적 편포의 형태를 보였다. 즉, TEPS와 TOEIC의 검사 난이도에 차이가 있다는 것을 유추할 수 있다. 제291회 TEPS 응시자 분포에 따르면, 해당 회차 응시자 집단의 TEPS 총점은 271-330점 구간에 가장 많이 분포했으며 이 구간을 중심으로 대체로 정규 분포에 가까운 분포 형태를

보인다(TEPS Council, 2020e). 이를 통해 본 연구의 표본 집단은 실제 TEPS 응시자 집단과 유사한 점수 분포를 보인다는 것을 알 수 있다. 제405회 TOEIC 응시자 분포에 따르면, 해당 회차 TOEIC 응시자 집단의 TOEIC 총점은 705-750점이 전체 10.77%로 가장 많았고 755-800점은 전체 10.42%, 805-850점은 전체 10.32%, 655-700점은 전체 10.26%로 나타났으며 다소 부적 편포된 점수 분포를 보였다(TOEIC, 2020e). 이와 비교했을 때 본 연구의 표본 집단은 실제 TOEIC 응시자 집단에 비해 비교적 높은 TOEIC 점수 분포를 보이지만 분포의 형태는 실제 TOEIC 응시자 집단의 점수 분포 형태와 유사하게 나타났다.

FIGURE 1

TEPS and TOEIC Total Score Distributions



다음으로 본 연구에서는 자료 수집 기간 및 방법에 따라 구분되는 1-4차 집단의 TEPS, TOEIC 총점에 대한 기초통계량을 분석하였다. 표 5에 나타나는 것과 같이 1-4차 집단의 TOEIC 총점 평균은 집단에 따라 큰 차이를 보이지 않았고, TEPS 총점 평균은 2차 집단에서 다른 집단에 비해 약간 낮게 나타났다.

TABLE 5

Descriptive Statistics for TEPS and TOEIC Total Scores by Group

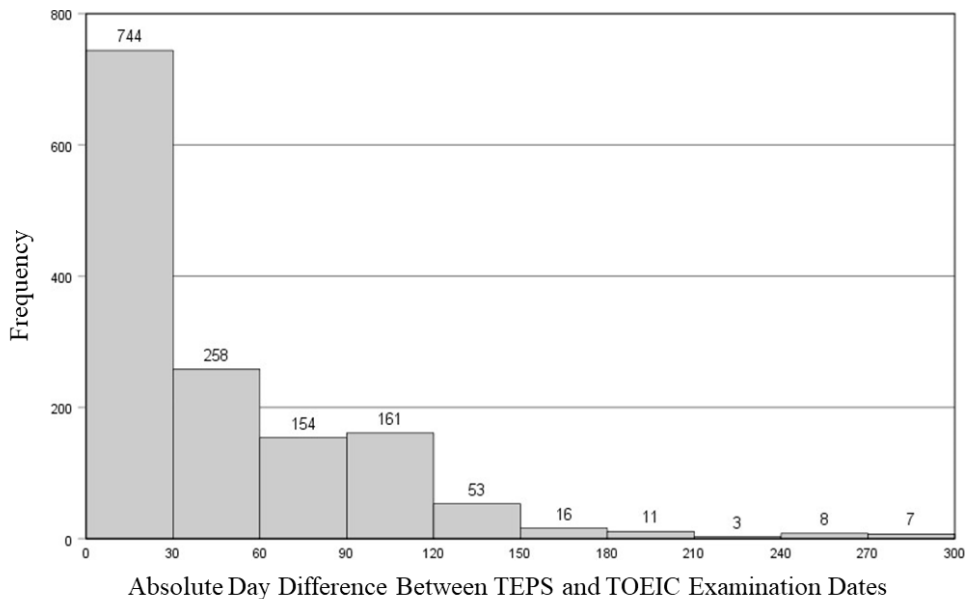
Group	N	Test	Min	Max	M	SD
1st	501	TEPS	105	598	346.51	91.55
		TOEIC	235	990	832.93	142.96
2nd	190	TEPS	114	574	321.72	78.76
		TOEIC	540	990	828.24	114.99
3rd	393	TEPS	162	588	359.87	90.87
		TOEIC	190	990	816.73	146.48
4th	331	TEPS	76	585	355.34	92.40
		TOEIC	210	990	824.27	145.31

본 연구의 표본자료가 단일집단 설계에 충실하게 수집되었는지 확인하기 위해

전체 수험자 1,415명이 TEPS와 TOEIC에 응시한 일자 차이에 대해 기술통계량을 산출하였다. 수집된 1,415명의 수험자 중 TEPS 응시일이 더 빨랐던 수험자는 569명이었고, TOEIC 응시일이 더 빨랐던 수험자는 846명이었다. 수험자들의 TEPS와 TOEIC 응시 일자 차이에 절댓값을 취해 분석한 결과, TEPS-TOEIC 응시 일자 차이의 절댓값은 1일에서 286일의 범위를 가지고 있으며, 평균은 46.08, 표준편차는 48.58로 나타났다. 그림 2는 응시 일자 차이의 절댓값에 대한 히스토그램이다. 그림 2에 나타난 바와 같이 TEPS와 TOEIC 응시 일자 차이가 30일 미만인 응시자가 744명으로 가장 많았고, TEPS와 TOEIC 응시 일자 차이가 30일 이상, 60일 미만인 응시자는 258명으로 나타났다. 반면 TEPS와 TOEIC 응시 일자 차이가 150일 이상인 응시자는 45명으로 상대적으로 적었다. 수집된 자료에 나타난 응시 일자 차이에 대한 분석 결과, 1) 수험자들의 평균 응시 일자 차이가 약 46일로 짧다는 점, 2) 수험자들의 절반 이상이 30일 이내에 두 시험에 모두 응시했다는 점, 3) 응시 일자 차이가 150일 이상인 수험자가 표본에서 차지하는 비율이 매우 낮다는 점 등을 고려했을 때 수집된 자료로 환산 연구를 수행하는 것이 비교적 적절하고 타당하다고 할 수 있다.

FIGURE 2

Frequency Distribution of Absolute Day Difference Between Examination Dates



4.2. 연계화

이 연구에서는 완곡화가 있는 동백분위 연계화 방법을 이용하여 TEPS-TOEIC 연계화를 실시하였다. 이를 위하여 먼저 로그 선형 사전완곡화의 완곡화 정도를 조정하는 모수를 결정하였고, 완곡화가 있는 동백분위 연계화 방법의 타당성을 확인하기 위해 집단에 따른 교차 검증을 실시하였으며, 최종적으로 전체 표본 집단을 대상으로 연계화를 실시하여 환산표를 산출하였다.

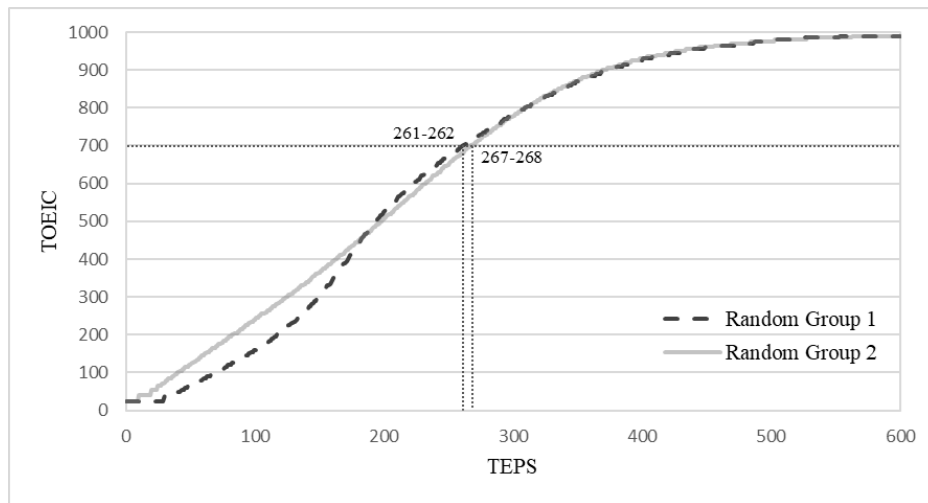
로그 선형 사전완곡화를 실시하기 위해서는 최적의 완곡화 모수 C 를 결정해야 한다. 이를 위하여 전체 표본 집단의 두 검사 총점을 대상으로 C 값에 따른 카이제곱 우도비 및 우도비 차이를 산출하였다. 표 6에 나타난 통계량을 검토했을 때 TEPS의 경우 $C=5$ 인 경우와 $C=6$ 인 경우 카이제곱의 우도비 차이가 4.260으로 자유도가 1인 카이제곱 분포에서 유의수준 .05 임계치에 해당하는 3.84보다 크게 나타나 통계적으로 유의하였다. 즉, $C=5$ 에서 $C=6$ 으로 완곡화 정도를 조정할 경우에 통계적으로 유의미한 수준의 개선이 있었다. 그러나 $C=7$ 이상부터는 완곡화 모수를 더하는 것이 더이상 유의미한 변화를 낳지 않았다. TOEIC의 경우 $C=4$ 인 경우와 $C=5$ 인 경우 카이제곱의 우도비 차이는 4.142로 자유도가 1인 카이제곱 분포에서 유의수준 .05 임계치에 해당하는 3.84보다 크게 나타나 통계적으로 유의하였다. 즉, $C=4$ 에서 $C=5$ 로 완곡화 모수를 증가시키는 것이 통계적으로 유의미한 수준의 개선으로 나타났다. 그러나 $C=6$ 이상부터는 더이상 유의미한 변화를 보이지 않았다. 이러한 카이제곱의 우도비 차이 검정 결과에 따라 TEPS의 완곡화 모수는 $C=6$ 으로, TOEIC의 완곡화 모수는 $C=5$ 로 결정하였다.

TABLE 6

Likelihood Ratio Chi-Square and Difference Statistics				
Test	C	χ^2	Degree of Freedom	$\chi^2_C - \chi^2_{C+1}$
TEPS	2	582.256	598	34.457
	3	547.800	597	0.109
	4	547.691	596	5.761
	5	541.929	595	4.260
	6	537.669	594	0.041
	7	537.628	593	0.185
	8	537.443	592	0.370
	9	537.074	591	
	TOEIC	2	138.854	194
3		138.834	193	0.158
4		138.676	192	4.142
5		134.533	191	0.707
6		133.826	190	1.809
7		132.016	189	0.012
8		132.005	188	0.622
9		131.383	187	

연계화 방법의 타당성을 교차 검증하기 위해 전체 표본 집단을 두 집단으로 무선할당한 후 사전완곡화가 있는 동백분위 연계화를 적용하여 결과를 비교하였다. 다음의 그림 3은 무선집단 1과 무선집단 2의 TEPS-TOEIC 점수 환산 결과에 어떤 차이가 있는지 그래프로 비교한 것이다. 임의로 나눈 두 집단에서의 환산 결과를 비교하면 낮은 점수대에서는 표본의 크기가 작아 연계화 결과에 차이를 보이지만 대다수 응시자가 포함되어 있는 점수 구간에서는 유사한 연계화 결과를 보인다. 무선집단 1의 경우 TEPS 261-262점이 TOEIC 700점으로 환산되고, 무선집단 2의 경우 TEPS 267-268점이 TOEIC 700점으로 환산됨을 알 수 있다. 즉, 집단에 따라 동일한 TOEIC 700점이 TEPS 점수에서 최대 7점의 차이를 보였다. 또한 동일한 TEPS 261점이 무선집단 1에서는 TOEIC 700점으로 환산되나 무선집단 2에서는 TOEIC 680점으로 환산되어 TOEIC 척도 상에서는 20점의 차이가 발생하였다. 즉, 임의로 나눈 두 집단 사이에 환산 결과가 큰 차이를 보이지 않아 사전완곡화가 있는 동백분위 연계화 방법이 안정적인 환산 결과를 산출한다는 것을 확인하였다.

FIGURE 3
TEPS-TOEIC Linking Relationships in Random Groups 1 and 2



교차 검증 결과를 바탕으로, 모든 응시자를 포함하는 전체 자료에 대해 사전완곡화가 있는 동백분위 연계화를 적용하여 최종 연계화 결과를 산출하였다. TOEIC → TEPS 환산 결과는 TEPS → TOEIC 환산표 산출 후 동일한 TOEIC 점수와 연계되는 TEPS 점수의 중앙값으로 산출하였고, 점수가 없는 지점에 대해서는 선형보간법(linear interpolation)을 적용하였다. 주요 점수 포인트에서의 TEPS-TOEIC 환산 결과는 표 7과 같다. 표의 왼쪽은 TEPS에서 TOEIC으로의 환산, 표의 오른쪽은 TOEIC에서 TEPS로의 환산을 나타낸다. 이는 TEPS와 TOEIC의 점수 척도가 다르기 때문에 한 검사의 여러 점수 포인트가 다른 검사의 한 점수 포인트로 환산될 수 있고,

이로 인한 점수 사용자의 혼란을 막기 위한 구분이다. 표에 포함되지 않은 다른 점수 포인트에 대한 환산 결과는 TEPS Council (2020d)에서 확인할 수 있다. 표 7의 왼쪽에 제시된 TEPS → TOEIC 환산 기준은 특정 TEPS 점수에 대해 환산 가능한 TOEIC 점수를 제공한다. 예를 들어 응시자의 TEPS 점수가 348점인 경우 이와 환산 가능한 TOEIC 점수는 870점임을 알 수 있다. 반대로 표 7의 오른쪽에 제시된 TOEIC → TEPS 환산 기준에 따라 특정 TOEIC 점수에 대해 환산 가능한 TEPS 점수를 살펴볼 수 있다. 예를 들어 응시자의 TOEIC 점수가 865점인 경우 TEPS 346점으로 환산된다.

TABLE 7
TEPS-TOEIC Conversion Table

TEPS → TOEIC		TOEIC → TEPS	
TEPS	TOEIC	TOEIC	TEPS
0	10	10	0
1~5	15	15	3
6	20	20	6
7~11	25	25	9
12~16	30	30	14
⋮	⋮	⋮	⋮
264~265	700	700	265
⋮	⋮	⋮	⋮
304~305	790	790	305
⋮	⋮	⋮	⋮
339~341	855	855	340
342~344	860	860	343
345~347	865	865	346
348~351	870	870	350
⋮	⋮	⋮	⋮
504~525	980	980	515
526~557	985	985	542
558~600	990	990	600

본 연구에서 산출된 환산 결과를 표 1에 제시된 기존 환산 기준과 비교하여 기존 환산 기준을 사용했을 때 수험자에게 미치는 영향을 파악할 수 있다. 예를 들어 TOEIC 700점은 본 연구에 따르면 TEPS 265점으로 환산된다. 그러나 2004년 환산 기준을 따르는 5급 및 7급 공무원 시험, 법원행정고등고시, 입법고등고시 등은 TEPS 340점과 TOEIC 700점을 같은 의미로 사용하여 TEPS를 치르는 수험자에게 본 연구 결과보다 75점이나 더 높은 점수를 요구한다. TEPS 340점은 본 연구 결과에 따르면 TOEIC 855점에 해당한다. 즉, TEPS 340점과 TOEIC 700점이 같다고 보는 2004년 환산 기준을 유지할 경우 TEPS 수험자들은 TOEIC 855점을 받을 수 있는 영어 실력이 있어야만 TEPS 340점 이상을 받고 요구 조건을 충족시킬 수 있다. 또한 7급 외무공무원 시험에서 TOEIC 790점과 같은 의미로 요구하는 TEPS 점수는 385점인데, 이는 표 7에 따른 TEPS 305점보다 80점이나 더 높다. 마찬가지로 5급 외교관후보자 선발시험에서는 TOEIC 870점 혹은 TEPS 452점을 요구하는데, 본 연구 결과에 따르면 TOEIC 870점은 TEPS 350점으로 환산되어 무려 102점의 차이가 있다. 즉,

정부 부처 및 주요 산하기관에서 2004년부터 사용하고 있는 환산 기준은 본 연구에서 산출한 결과에 비해 TEPS 수험자들에게 75-102점가량 더 높은 점수를 요구한다. 2011년 환산 기준 역시 정도의 차이가 있을 뿐 동일한 TOEIC 점수 기준에 대해 본 연구의 환산 결과보다 더 높은 TEPS 점수를 요구한다.

5. 결론 및 시사점

TEPS와 TOEIC은 영어 능력을 평가하는 표준화 검사로 다양한 목적을 위해 사용되고 있다. 두 검사가 유사한 목적으로 함께 사용되는 상황이 많은 현실을 고려할 때 두 검사 점수의 환산 기준을 명확하게 하는 것이 매우 중요하다. 그러나 현재 사용되고 있는 환산 기준은 도입 이후 짧게는 10년, 길게는 17년이 경과하여 그 사이에 발생한 모집단 변동을 반영하지 못하고 있고, 또한 2018년에 TEPS가 개정되면서 점수 체제가 변화하였음에도 환산 기준은 (구)TEPS-TOEIC 환산 결과를 기반으로 하고 있어 타당성 측면에서 의문이 제기된다. 이에 본 연구에서는 TEPS와 TOEIC의 점수 환산이 신뢰롭고 타당하게 이루어질 수 있도록 대규모 표본 자료를 이용하여 두 검사 점수 간 직접적인 연계화를 실시하였다.

연구 결과, TEPS와 TOEIC 간의 직접적인 연계화를 통해 산출된 환산 결과는 현재 정부 부처와 그 산하기관 등에서 사용하고 있는 2004년 환산 기준이나 일부 공공기관 및 공기업 등에서 사용하고 있는 2011년 환산 기준과 상당한 차이가 있음이 나타났다. 현행 환산 기준을 적용할 경우 TOEIC에 응시한 수험자에 비해 TEPS에 응시한 수험자들이 더 높은 영어 실력을 갖춰야만 같은 요구 기준을 통과할 수 있는 것이다. 이는 같은 영어 실력을 갖고 있더라도 어떤 검사를 치르느냐에 따라 유불리가 발생한다는 것을 의미한다. 채용, 선발, 승진 등 중요한 인사 결정에 반영되는 영어 능력 검사 점수의 특성을 고려했을 때, 특정 검사의 응시자들이 이러한 불이익을 받는다는 것은 추후 법적 문제로 이어질 가능성이 있다. 따라서 본 연구를 통해 산출된 TEPS-TOEIC 간 직접적인 환산 결과를 바탕으로 새로운 환산 기준을 적용하는 것이 시급할 것으로 보인다.

본 연구는 국내의 수많은 영어 능력 시험 수험자들과 그 점수 사용 기관에 신뢰롭고 타당한 점수 환산의 근거를 제공할 것으로 기대된다. 또한 본 연구에서 적용한 환산표 산출 방법은 TEPS와 TOEIC 외 다른 영어 능력 시험 점수 간 환산 관계를 산출하고 점검하기 위해 적용될 수 있으리라 생각된다. 특히 두 검사와 비슷한 목적으로 사용되고 있는 FLEX, G-TELP, TOEFL 등의 시험에 대해 타당한 환산 기준을 산출하는 것이 당면 과제라 할 수 있다. 또한 영어 말하기 능력에 대한 관심과 요구가 증대되고 있는 상황을 고려할 때 IELTS, OPIc, TEPS Speaking, TOEFL, TOEIC Speaking 등 영어 말하기 능력을 측정하는 검사들 간의 환산 기준을 재점검하고 그 타당성을 확인하여 검사에 따른 수험자 유불리를 방지할 필요가 있다.

이러한 함의에도 불구하고 본 연구가 가지고 있는 제한점 및 후속 연구를 위한

제언은 다음과 같다. 첫째, 대규모 표본 자료를 구축함에 있어서 TEPS 수험자 모집단과 TOEIC 수험자 모집단을 모두 포함하여 표본의 대표성을 확보하고자 하였으나 그 구성 비율에는 차이가 있었다. 특히 1, 3, 4차 집단의 경우 수험자 개인이 점수를 활용할 목적을 갖고 자발적으로 TEPS와 TOEIC에 모두 응시하여 성적표를 보유하였지만, 2차 집단의 경우 그렇지 않았다. 2차 집단 수험자들은 TOEIC에 대해서는 개인의 필요에 의해 자발적으로 응시하여 성적표를 보유하였지만 TEPS에 대해서는 문항 형식이나 검사 시간 등에 익숙하지 않은 상태에서 참가자 보상을 받기 위해 시험을 치렀다고 할 수 있다. 이는 TEPS는 치러본 경험이 없고 TOEIC에 익숙한 수험자가 많은 현실적인 상황을 반영한 부분이나, 이로 인해 발생한 오차가 환산 결과에 반영되었을 수 있다. 둘째, 표본에 포함된 수험자들의 점수 분포를 살펴보면 상대적으로 영어 능력이 우수한 수험자들이 많았다. 즉, 연계화의 조건부 오차가 낮은 점수대에서 상대적으로 더 높을 것으로 예상된다. 이러한 제한점을 고려할 때 본 연구에서 산출한 환산 결과의 타당성을 확인하기 위해 지속적인 자료 수집 및 후속 연구가 필요하다. 인접한 응시일자에 두 검사를 모두 치른 수험자의 성적 자료를 꾸준히 수집하여 대규모 표본을 구축하고 주기적으로 연계화 연구를 수행하여 환산 기준을 점검할 필요가 있다.

Applicable level: Tertiary

REFERENCES

- Akaike, H. (1981). Likelihood of a model and information criteria. *Journal of Econometrics*, 16(1), 3-14.
- Brennan, R. L., Wang, T., Kim, S., & Seol, J. (2009). *Equating recipes* (Version 1.0) [Computer software]. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, University of Iowa.
- College Board. (2018). *Guide to the 2018 ACT/SAT Concordance*. Retrieved on February 19, 2021, from <https://collegereadiness.collegeboard.org/pdf/guide-2018-act-sat-concordance.pdf>
- Cope, R. T., & Kolen, M. J. (1990). *A study of methods for estimating distributions of test score* (ACT Research Report 90-5). Iowa City, IA: American College Testing.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.

- Dorans, N. J., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT Assessment and recentered SAT I sum scores. *College and University*, 73(2), 24-35.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Hoboken, NJ: Prentice Hall.
- Educational Testing Service. (2019). *TOEIC® listening & reading test user guide - updated version*. Retrieved on November 8, 2021, from <https://www.ets.org/s/toEIC/pdf/toEIC-listening-reading-test-user-guide.pdf>
- Educational Testing Service. (2021). *Frequently asked questions about the TOEIC® listening and reading test*. Retrieved on November 8, 2021, from <https://www.ets.org/toEIC/organizations/listening-reading/faq>
- Fairbank, B. A. (1987). The use of presmoothing and postsmoothing to increase the precision of equipercentile equating. *Applied Psychological Measurement*, 11(3), 245-262.
- Hanson, B. A., Zeng, L., & Colton, D. (1994). *A comparison of presmoothing and postsmoothing methods in equipercentile equating* (ACT Research Report 94-4). Iowa City, IA: ACT.
- IBM Corp. (2016). *IBM SPSS Statistics for Windows, Version 24.0* [Computer Software]. Armonk, NY: IBM Corp.
- Kim, A., Lee, G., Ban, J.-C., & Cha, J.-E. (2004). Linking old and new vocational aptitude batteries for adults. *Journal of Educational Evaluation*, 17(2), 107-129.
- Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, 9(1), 25-44.
- Kolen, M. J. (1988). Traditional equating methodology. *Educational Measurement: Issues and Practice*, 7(4), 29-37.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking*. New York: Springer.
- Korea English Language Testing Association. (2018). *Comparability and score conversion between revised TEPS and authorized English tests*. Seoul, Korea: Author.
- Lee, G., Ban, J.-C., Lee, M., & Jung, J. (2020). *Scale development for the TOPIK speaking test*. Seongnam, Korea: National Institute for International Education.
- Lim, E., Jun, H., Yun, J., & Min, S. (2019). Evidence of construct validity for new TEPS. *Language Research*, 55(s), 37-50.
- Linn, R. L. (1993). Linking results of distinct assessment. *Applied Measurement in Education*, 6(1), 83-102.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Hoboken, NJ: Prentice Hall.

- Ministry of Justice. (2004). *2004 bar examination and military judge advocate examination plan announcement*. Retrieved on December 20, 2020, from <https://viewer.moj.go.kr/skin/doc.html?rs=/result/bbs/146&fn=1545276375198101>
- Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: Educational Testing Service.
- Schneider, D., & Dorans, N. (1999). *Concordance between SAT® I and ACT™ Scores for Individual Students*. Research Notes. RN-07. College Entrance Examination Board.
- TEPS Council. (2020a). *Examination guide*. Retrieved on December 20, 2020, from <https://www.teps.or.kr/Info/Committee>
- TEPS Council. (2020b). *2018 New TEPS: Revision*. Retrieved on December 20, 2020, from <https://www.teps.or.kr/html/NEWTEPS/html/change.html>
- TEPS Council. (2020c). *Score uses*. Retrieved on December 20, 2020, from <https://www.teps.or.kr/Use/School>
- TEPS Council. (2020d). *Conversion table*. Retrieved on December 20, 2020, from <https://www.teps.or.kr/InfoBoard/ConversionTable>
- TEPS Council. (2020e). *TEPS analysis: 296th administration*. Retrieved on December 20, 2020, from <https://www.teps.or.kr/InfoBoard/TepsAnalysis>
- TEPS Council. (2021). *Scoring and level descriptions*. Retrieved on November 8, 2021, from <https://www.teps.or.kr/Info/Teps#>
- TOEIC. (2020a). *Introduction*. Retrieved on December 20, 2020, from <https://exam.toeic.co.kr/common/template/viewContents.php?contentsCode=19>
- TOEIC. (2020b). *Proficiency description*. Retrieved on December 20, 2020, from <https://exam.toeic.co.kr/common/template/viewContents.php?contentsCode=20>
- TOEIC. (2020c). *Score uses*. Retrieved on December 20, 2020, from <https://exam.toeic.co.kr/common/template/viewContents.php?contentsCode=21>
- TOEIC. (2020d). *Score analysis – by year – 2019 TOEIC and TOEIC Speaking score analysis*. Retrieved on December 20, 2020, from <https://exam.toeic.co.kr/result/statisToeic.php>
- TOEIC. (2020e). *Score analysis – by administration – 405th administration*. Retrieved on December 20, 2020, from <https://exam.toeic.co.kr/result/statisToeic.php>