

영어 모의 TOEIC 시험의 구인 및 예언타당도 연구*

장재학
(강원대학교)

Chang, Jaehak. 2019. A study on the construct and predictive validity of mock TOEIC Tests. *Korean Journal of English Language and Linguistics* 19-3, 523-538. This study aimed to examine the construct and predictive validity of the reading section of a mock TOEIC test designed to measure first-year college students' general English proficiency. To this aim, the present study utilized a series of confirmatory factor analyses and examined a couple of reading models. The results suggested that L2 learners' reading ability measured by the mock TOEIC test consists of two constructs: language knowledge and reading skills. Language knowledge involves grammatical knowledge and lexical knowledge. Reading skills are concerned with factual understanding and inferential understanding of written English. The results also showed that the mock TOEIC test has some evidence for predictive validity. Some suggestions for future studies were also made.

Keywords: TOEIC, reading test, construct validity, predictive validity

I. 서론

TOEIC(이하 토익 시험)은 한국 수험생들이 취업, 승진, 졸업 등을 위하여 많이 보는 시험이다. 이들 수험생들은 또한 현재 자신의 영어 능력을 알아보기 위하여 토익 시험을 보기도 한다. 영어 능력이 현대 우리 사회에서 차지하는 중요성 때문에, 일부 대학에서는 신입생들의 영어 능력을 추정해보기 위해서 토익 시험을 실시하거나, 공인 토익 시험 점수를 요구하거나, 토익 시험과 유사하게 제작한 모의 토익 시험을 실시하기도 한다.

그러나 많은 학생들이 다양한 이유와 목적으로 가장 많이 보고 있는 공인 영어 시험 중 하나인 토익 시험이 진정 어떠한 영어 능력을 측정하는 지에 대한 연구는 거의 이루어지지 않았다. 그나마 토익 시험 전체에 대한 타당도 연구는 해외에서 일부 이루어졌으나, 토익 시험의 읽기 영역이나 듣기 영역과 같이 영역별로 어떠한 구인을 측정하는 지에 대해서는 조사된 바도 알려진 것도 거의 없다(Suzuki and Daza 2004). 이에 본 연구는 토익 시험이 측정하고자 하는 평가 구인이 무엇인지 알아보기 위하여, 토익 시험과 아주 유사하게 제작된 모의 토익 시험을 제작하였고, 이 중에서 특히 모의 토익 시험 읽기 영역의 평가 구인과

* 본 논문은 2017년도 강원대학교 대학회계 학술연구조성비로 연구하였음(관리번호-520170311).

예언 타당도를 조사하였다.

II. 선행연구 개관

본 장에서는 영어 교육 분야에서 시행된 토익 시험의 타당도 관련한 주요 연구에 대해서 간략히 살펴보고자 한다. 토익 시험의 타당도 연구는 2006년을 기점으로 그 전과 후를 구분할 필요가 있다. 현재의 토익 시험의 구성과 설계는 2006년 이후 변경되었다. 이전 토익 시험의 읽기 영역에서는 주어진 문장의 오류를 인식(error recognition)하는 유형의 문항이 포함되었으나 그 이후의 토익 시험의 읽기 영역에서는 배제되었다. 이러한 토익 시험의 구성과 설계의 변경은 토익 시험의 구인에도 변화가 있었음을 보여주는 것으로, 2006년 이전에 시행된 토익 시험의 타당도 연구나 그 결과는 현재 토익 시험의 구인타당도를 살펴보는 데 큰 의미가 없다고 할 수 있다. 따라서 본 연구에서는 2006년 이후 시행된 토익 시험의 타당도 연구 중에서 중요성을 갖는 일부 연구들을 중심으로 논의하겠다.

토익 시험을 주관하는 ETS에서 현재 토익 시험의 타당도에 대한 연구를 진행한 적이 있다. Powers(2010)는 한 토익 시험 연구 보고서에서 현 토익 시험의 타당도를 다양한 관점에서 조사하였으며 타당도에 대한 증거가 있다고 주장한다. 일례로, 이 연구는 토익 시험이 많은 토익 시험 전문가들이 토익 시험 문항이나 내용에 대해 철저한 검토 과정을 거치고 있기 때문에 타당도가 있다고 주장한다. 물론 전문가들의 철저한 검토 과정은 토익 시험의 타당도, 그 중에서도 특히, 내용타당도를 높이는 데 도움이 되는 것은 사실이다. 그러나 이러한 전문가의 판단이나 식견은 토익 시험의 구인타당도가 있다는 것의 증거로 사용될 수 없는 문제점이 있다.

그 외에도 Powers(2010)는 보고서에서 여러 종류의 타당도를 논의하지만 가장 주목할 점은 토익 시험이 예언타당도가 있다는 것이다. 그는 2007년 실시한 연구에서, 현 토익 시험을 치른 약 5천명의 학생들에게 토익 시험의 시험 과제들이 유사한 실제 직장 상황의 의사소통에 어느 정도의 어려움이 있는지 또는 도움이 되는 지에 대한 설문 조사를 실시하였다. 설문지는 수험생들에게 토익 시험의 읽기 과업과 유사한 실제의 여러 의사소통 상황들, 예를 들면, 사무실에서 단어나 문장으로 쓰인 업무 메모를 얼마나 잘 이해할 수 있는지, 회의 주요 내용에 대해 얼마나 잘 이해할 수 있는 지, 사업 제안서의 내용을 읽고 얼마나 잘 이해할 수 있는지 등을 물었다. 연구 결과는 토익 시험의 읽기 영역 성적이 높은 학생들이 낮은 학생들 보다 시험 이외의 실제 업무 상황에서 연관된 업무를 더 잘 수행할 수 있을 거라는 응답이 많았다. 읽기 영역 성적이 높으면 높을수록 긍정적인 응답 비율이 높았다. 이와 같은 결과에 기초하여, Powers(2010)는 현 토익 시험이 타당도가 있다고 주장하였다. 물론 이러한 경향은 현 토익 시험의 예언타당도가 어느 정도 있다는 것을 의미하지만, 현 토익 시험의 구인이 무엇인지 알려주지 못하는 한계가 있다.

In'nami와 Koizumi(2011)는 현 토익 시험의 구인타당도를 사실상 처음으로 조사하였다. 이들은 일본 소재 한 과학 기술대 소속의 학생들 569명의 토익 시험 성적 자료를 이용하여, 현 토익 시험 전체의 요인 구조를 확인적 요인 분석 방법을 활용하여 조사하였다. 이들은 이 연구에서 현 토익 시험이 듣기 기술(listening skills)과 읽기 기술(reading skills)의 두 가지 구인으로 이루어졌다고 가정하고 이 두 구인을 하나로 묶는 상위 개념인 이해 기능(receptive skills)이 존재한다고 가정한 위계 모형(higher-order model)과 읽기 기술과 듣기 기술은 서로 연관성이 있으나 다른 구분되는 능력이라고 가정한 상관 모형(correlated model) 등을 비교하였다. 연구 결과는 토익 시험 전체는 이해 기능이라는 상위 구인이 존재한 다기 보다는 읽기 기술과 듣기 기술이 서로 연관 되지만 서로 구분되는 다른 구인이라는 점을 밝혔다. 이 연구의 성과는 토익 시험 전체는 듣기와 읽기 영역으로 나누어 시험을 치르고 영역별 시험 성적을 부여해야하는 타당성의 근거를 제시한다는 점에서 매우 의의 있는 연구라 할 수 있다. 그러나 이 연구는 토익 시험 전체의 타당도에 초점이 맞추어져 있는 연구로서, 토익 시험의 읽기 영역 또는 듣기 영역은 몇 개의 구인으로 이루어졌는지에 대한 답변은 제시하지 못하는 아쉬움이 있다. 따라서 본 연구는 다음과 같은 연구 질문에 답을 얻고자 한다.

1. 모의 토익 시험을 통해 바라본 토익 시험의 읽기 영역의 구인은 무엇인가?
2. 모의 토익 시험의 읽기 영역의 예언타당도가 어느 정도인가?

Ⅲ. 연구 방법

1. 연구 참여자

본 연구에는 춘천 소재 K-대학의 신입생 3,814명이 참여하였다. K-대학의 신입생들은 학기 초에 의무적으로 대학에서 시행하는 모의 토익 시험을 보아야 한다. 이 시험은 신입생들의 어학 능력을 알아보는 것을 목적으로 한다.

2. 평가 도구

본 연구에 사용된 모의 토익 시험은 본 토익 시험의 구성과 설계와 아주 유사하게 제작하였으며, 듣기 100문항, 읽기 100문항 총 200문항으로 이루어졌다. 그러나 본 연구에서는 연구의 목적상 듣기 영역을 제외하고 읽기 영역만을 다룬다. 본 시험에서 사용된 모의 토익 시험의 읽기 영역의 구성 및 설계는 표 1에 제시되어 있다.

표 1. 모의 토익의 시험 설계

문항 유형 (test format)	문항 수	구인 (construct)	문항 수
Incomplete sentences	40	언어 문법 지식	33
Text completion	12	어휘 지식	21
Single passage	28	사실적 이해	39
Double passage	20	추론적 이해	7

3. 평가 구인

모의 토익 시험의 구인을 이해하기 위해서는 본 토익 시험의 구성과 설계에 대해서 먼저 알아 볼 필요가 있다. 본 연구는 토익 시험 중에서 읽기 영역만을 다루고 있으므로, 본 연구 주제와 연관 있는 토익 시험의 읽기 영역에 국한해서 논의하도록 하겠다. 토익 시험을 주관하는 미국 ETS에 따르면, 토익 시험의 읽기 영역은 표 1에 제시한 바와 같이 4개의 영역 총 100문항으로 이루어졌다. 문장 완성 형 40 문항, 담화 완성 형 12 문항과 하나의 읽기 지문을 제시하고 지문의 이해를 묻는 문항 유형 28 문항, 그리고 2개의 연관된 지문을 제시하고 지문들의 이해를 묻는 문항 유형 20문항이다.

이들 문항 유형에 대한 기술은 그러나 각 문항 유형이 무엇을 측정하려고 하는 지, 즉 평가 구인에 대해서는 알려주지 못하고 있다. 각 문항 유형이 측정하고자 하는 구인(construct)이 무엇인지를 알아보기 위해서는 문항 예제를 분석할 필요가 있다. 모의 토익 시험의 본 토익 시험을 최대한 본 따서 제작되었음을 감안하여 문항 분석을 해보면 각 문항 유형에서 무엇을 측정하는 지에 대해서, 즉 구인에 대해서 추론할 수 있을 것이다.

아래의 문장 완성형 문항을 살펴보면 토익 시험이 이 문항 유형에서 측정하고자 하는 구인이 문법 지식임을 알 수 있다.

문법 문항 예제

103. We consider long-term _____ in a delivery staff member to be more important than a willingness to work for long hours.

- (A) reliant
- (B) reliably
- (C) reliable
- (D) reliability

문항 103번의 선택지를 보면 모두 동일 어근을 가진 신뢰성에 관한 단어들을 알 수 있다. 따라서 이 문항은 유사한 뜻을 가진 단어들 중에서 주어진 문장을 완성하는 데 어떠한 단어 형태 또는 품사가 필요한 지를 묻는 것이다. 주어진 문장은 배달 직원이 오랜 시간 일하려는 마음보다 오랜 기간에 걸친 배달의 신뢰성을 더 중요하게 생각한다는 내용이지만, 문장의 전체적인 뜻을 다 이해하지 못하더라도 *consider*가 타동사이고 타동사는 명사나 대명사와 같은 명사 상당 어구의 목적어를 필요로 한다는 문법 지식만 있어도 문제를 푸는 데 지장이 없다. 또한 빈칸의 단어 앞에 *long-term*이라는 형용사가 있으므로 뒤에 오는 단어는 명사형인 (D) *reliability* 만이 정답이 된다는 것을 알 수 있다.

동일한 문장 완성 형 유형이지만 아래에 제시된 문항 131은 문법 지식이 아닌 다른 언어 지식을 측정하고 있다.

어휘 문항 예제

131. The government has earmarked eight percent of gas tax _____ for the development of alternative-energy vehicles.

- (A) reversions
- (B) reminders
- (C) remainders
- (D) revenues

문항 131은 *gas tax* 뒤에 이어질 말을 고르는 문제로 *tax* 뒤에는 *tax*의 수식을 받을 수 있는 명사형이 와야 하는 데 제시된 단어 모두 명사형이라는 형태적인 공통점이 있다. 따라서 이 문항은 문법과 관련된 형태적인 지식을 묻는 것이 아니며 동일 형태 중에서 의미에 알맞은 단어를 고르는 문항, 즉 어휘 지식을 묻는 문항으로 봐야 할 것이다. 주어진 문장은 정부가 유류세 수입의 8 퍼센트를 대체 에너지 차량 개발에 책정했다는 내용이 돼야 하므로 세수에 해당하는 (D) *revenues*가 정답으로 선택되어야 한다. 선택지 (A) *reversions*은 전환, 상속권을 뜻하며, (B) *reminders*는 무엇인가 생각나게 하는 사람(사물) 등을 뜻하며, (C) *remainders*는 나머지, 잔여라는 뜻이므로 모두 의미상 문맥에 사용될 적절한 단어들일 수 없다.

본 논문에서는 지문형 문항의 평가 구인에 대해서도 알아보았다. 문항 165-168은 구인 광고에 대한 주어진 지문을 읽고 질문에 답하는 문제 유형이다.

사실적 이해/추론적 이해 문항 예제

Questions 165-168 refer to the following classified advertisement.

Job #16 Data Entry Assistant

Temporary weekend position lasting approximately 5–7 weeks. Basic computer skills, good writing skills strongly preferred.

Job #17 Guest Service Representative

Full-time. We need an experienced manager who has supervised a staff of at least 20 people. Candidates will possess outstanding communications and customer service skills. Some familiarity with Spanish or Korean and ability to work under pressure a plus.

Job #18 Accounts Receivable Clerk

Part-time entry-level position. Duties include photocopying, filing, and handling incoming phone calls. Individuals must enjoy working with people. Will receive between 20 and 25 weekday hours.

Job #19 Room Attendant

Seeking energetic individual. Must be able to work all shifts and occasionally work overtime. Previous housekeeping experience preferred, but we are willing to train.

Please reply by fax for managerial positions. All other positions may apply for an on-site interview in person. Please send your resume to 3010 S. Broadway Street, Palos Verdes, CA, or fax to Jean Pearson (801) 623–7272.

165. Where are these positions being offered?

- (A) At a software company
- (B) At a department store
- (C) At a travel agency
- (D) At a hotel

167. For how long will the temporary position last?

- (A) 5 to 7 weekdays
- (B) 16–19 days
- (C) About 2 months
- (D) 20 to 25 hours

문항 165는 어디에서 제시된 일자리를 제공하고 있는가를 묻고 있다. 그러나 주어진 지문 어디에도 이들 광고주에 대한 명시적인 언급은 없다. 다만 일자리 16에서 19까지의 내용을 읽고 광고주나 지원자들이 어디에서 일을 하게 되는 지에 대한 추론을 할 수는 있다. 문항 165의 선택지를 보면 소프트웨어 회사, 백화점, 여행사, 그리고 호텔로 되어 있다. 일자리 16번의 내용을 본다면, 혹자는 컴퓨터에 자료 입력 등을 보조해야 하므로 소프트웨어 회사와의 연관성이 있는 것이 아닌가라는 생각을 할 수도 있을 것이다. 그러나 일자리 17번 내용을 보면 지원자들은 의사소통을 잘해야 하고 손님 응대에 대한 기술을 소유해야 한다는 내용이 나오므로 소프트웨어 회사라기보다는 많은 고객들을 상대해야 하는 백화점, 여행사, 그리고 호텔이 더 적절해 보인다.

무엇보다도 일자리 19번의 제목과 내용을 보면 정답이 호텔이어야 한다는 것을 알 수 있다. 일자리 19번의 제목과 내용에는 *Room Attendant*와 *housekeeping*이라는 단어가 제시되어 있다. *Room Attendant*는 우리말로 객실 청소부라고 해석할 수 있고 문맥상 *housekeeping*은 객실 정리 정돈하는 일임을 알 수 있다. 객실을 청소하고 정리 정돈하는 일을 하는 곳은 백화점, 여행사, 호텔 중에서 호텔이 가장 가까우므로 호텔에서 낸 구인 광고로 유추하는 것이 가장 합리적인 판단일 것이다. 따라서 주어진 지문 그 어디에도 어느 호텔에서 사람을 구한다는 문구가 명시적으로 제시되지 않았으므로 문항 165는 명시적으로 언급되지 않은 내용을 문맥을 통해서 추론하는 능력, 즉 추론적 이해력을 측정하는 문항이라고 할 수 있다.

문항 167은 이상에서 언급한 문항 165와는 다른 유형의 이해 능력을 묻는 문항이다. 문항 167은 주어진 구인 광고에서 제시된 임시직이 얼마 동안 지속 될 것인가를 묻고 있다. 이 문항의 선택지를 보면 (A) 5 일에서 7 일, (B) 16 일에서 19 일, (C) 약 2 개월, (D) 20 시간부터 25 시간까지로 되어있다. 이 문제를 올바르게 풀기 위해서는 근무 시간, 근무 일수 등에 대한 정보를 지문에서 찾아야 할 뿐만 아니라 이들 숫자 정보가 될 의미하는 지도 문맥에서 정확하게 파악해야 한다. 선택지 (A)에 표시된 5와 7의 숫자는 일자리 16번의 5에서 7 주에서 온 것으로 보이며, 선택지 (B)의 16과 19의 숫자는 일자리 번호인 16에서 19에서 온 것으로 보인다. 또한 선택지 (D)의 20과 25의 숫자는 일자리 18번 지문 내용을 보면 20에서 25시간이라는 내용이 명시적으로 나오는 데 여기서 따온 것으로 보인다. 즉, 이들 선택지들은 지문 내용에 명시적으로 제시된 숫자 정보이기는 하나 정확한 내용 이해 없이 정답을 찍는 학습자들을 위한 매력적인 오답의 역할을 하고 있다고 할 수 있다. 끝으로 선택지 (C)는 일자리 16번의 내용을 잘 살펴보면 (C) 번이 왜 정답인지를 알 수 있다. 일자리 16번의 내용을 잘 살펴보면 임시직으로 주말에 약 5에서 7주간의 일한다고 명시되어 있다. 선택지에는 5 주에서 7 주의 기간이 없으므로 이를 풀어 쓴 선택지 (C)번 약 2 개월의 기간이 가장 본문 내용과 일치한다는 것을 알 수 있다. 이상에서 논의한 데로, 문항 167은 지문 내용에 명시적으로 언급된 내용을 정확하게 이해하고 있는지, 즉 사실적 이해 능력을 측정하는 문항이라고 할 수 있다.

본 연구는 위와 같은 문항 분석을 통해서 모의 토익 시험의 읽기 영역에 제시된 모든 문항을 살펴보았으며, 그 결과 평가 구인을 추론할 수 있었다. 모의 토익 시험 읽기 영역의 평가 구인은 넓게 보면 언어 지식(language knowledge)과 이해 기능(comprehension skills)로 구분할 수 있다. 첫 번째 평가 구인인 언어 지식은 문법 지식과 어휘 지식의 두 구성 요소로 이루어졌으며, 두 번째 평가 구인인 이해 기능은 사실적 이해력과 추론적 이해력의 두 구성 요소로 이루어졌다. 이들 평가 구인은 표 1에 제시되어 있다.

4. 연구 절차

모의 토익 시험의 읽기 영역은 총 75분간 실시되었다. 학습자들의 응답은 맞으면 1점, 틀리면 0점으로 채점되었으며, 이 후 다양한 통계 분석에 이용되었다. 기술통계, 상관관계, 신뢰도 분석은 SPSS 24.0 for Windows를 이용하여 실시하였다. 본 연구의 주요 목적인 토익 시험 읽기 영역의 구인타당도 검증을 위한 일련의 분석에는 EQS 6.1 for Windows를 이용하였다.

IV. 자료 분석 및 결과

1. 시험의 신뢰도

모의 토익 시험의 읽기 영역 전체의 신뢰도를 알아보기 위하여 신뢰도 분석을 실시하였다. 신뢰도 분석에는 내적 신뢰도 지수인 크론바크 알파 계수(Cronbach's alpha)를 이용하였다. 본 연구의 전체 참여자인 3,814명과 읽기 영역 총 100문항을 모두 포함하여 전체 신뢰도를 구하였으며 그 결과 읽기 영역의 신뢰도 지수는 .877로 나타났다. 따라서 읽기 영역 전체의 신뢰도는 상당히 높은 편으로 본 연구 결과는 모두 신뢰할 수 있다고 하겠다.

2. 기술 통계와 상관 계수

표 2와 표 3은 모의 토익 시험의 읽기 영역에 관한 기술 통계와 타당도 검증에 필요한 변인 간의 상관 계수를 제시한다. 각 영역 별 문항 수가 서로 다른 관계로 영역 별 시험 점수 비교를 위하여 100점 만점으로 환산한 점수를 표 맨 아래 환산 점수 란에 제시하였다. 환산 점수 기준으로, 1지문 독해(single passage) 영역 평균이 43.9로 가장 높았고, 문장 완성 영역 평균이 42.8로 그 다음으로 높았다. 담화 완성 영역 평균은 40.9이었으며, 2지문 독해(double passage) 영역의 평균 점수는 33.6으로 다른 영역에 비해 상대적으로 낮았다. 변인간의 상관계수는 모두 양의 상관을 보였고 아주 높지도 아주 낮지도 않았다. 상관 계수

는 낮게는 .329에서 높게는 .630까지의 다양하였다.

표 2. 읽기 기술통계 및 상관계수 (N=3,814)

변인	1	2	3	4
Incomplete sentences	1.000			
Text completion	.563**	1.000		
Single passage	.630**	.497**	1.000	
Double passage	.423**	.329**	.489**	1.000
평균(M)	17.1432	4.9183	12.3059	7.3339
표준편차(SD)	5.7710	2.0378	4.9830	3.3383
환산 점수	42.8580	40.9858	43.9496	36.6695

1: Incomplete sentences, 2: Text completion, 3: Single passage, 4: Double passage

** $p < .01$

표 3. 읽기 기술통계 및 상관계수 (N=3,814)

변인	1	2	3	4
문법 지식	1.000			
어휘 지식	.632**	1.000		
사실적 이해	.576**	.616**	1.000	
추론적 이해	.389**	.417**	.576**	1.000
평균(M)	12.9422	9.8291	15.8866	3.0433
표준편차(SD)	4.4310	3.6421	5.9601	1.6084
환산 점수	39.2188	46.8051	40.7348	43.4757

1: 문법 지식, 2: 어휘 지식, 3: 사실적 이해, 4: 추론적 이해

** $p < .01$

표 3은 모의 토익 시험의 읽기 영역에 관한 기술 통계와 타당도 검증에 필요한 변인 간의 상관 계수를 제시한다. 표 2와 표 3이 서로 다른 점은 학습자들의 읽기 능력을 문제 유형별 또는 영역별이 아니라 평가 구인 중심으로 살펴보는 것이다. 구인 중심으로 살펴보면, 학습자들이 어느 지식 또는 능력이 상대적으로 더 부족한지를 보다 명확하게 알 수 있는 장점이 있다. 본 표에서 보듯이 100점 만점으로 환산하였을 때, 어휘 지식이 평균 46.8로 점수가 가장 높았다. 추론적 이해력 평균 43.4로 그 다음으로 높았으며, 사실적 이해의 평균은 40.7이었다. 문법 지식은 평균 39.2로 가장 낮았다. 이상의 결과에서, 문법 문제와 사실적 이해 문제의 비중이 상대적으로 높은 토익 시험을 학습자들이 보다 더 잘 보기 위해서는 이들 두 영역에 보다 더 집중하여 공부할 필요가 있음을 보여준다.

또한 표 3을 보면, 문법 지식과 어휘 지식의 상관관계가 .632로 가장 높으며, 어휘 지식과 사실적 이해와의 상관관계가 .616으로 그 다음으로 높음을 알 수 있다. 사실적 이해와 추론적 이해의 상관관계는 .576이었으며, 추론적 이해와 문법 지식과의 상관관계는 .389로 가장 낮았다. 이러한 상관관계 분석 결과는 변인들 간에 수렴타당도(convergent validity)와 구분타당도(divergent validity)가 어느 정도 있음을 시사한다. 수렴타당도는 이론적으로 유

사하다고 생각되는 변인들 간의 상관계수가 이론적으로 서로 연관성은 있으나 다르다고 여겨지는 변인들의 상관계수 보다 더 높을 것은 말한다. 구분 타당도는 수렴타당도의 반대 개념이다. 본 연구에서는 언어 지식들 간의 상관관계가 언어 지식과 읽기 기술의 상관관계보다 모두 다 높게 나왔다. 예를 들어, 문법 지식과 어휘 지식의 상관관계는 .632이었고, 문법 지식과 추론적 이해의 상관관계는 .389이었으며, 어휘지식과 추론적 이해와 상관관계는 .417이었다. 이상에서 변인들 간에는 적당한 상관관계 및 구분타당도와 수렴타당도에 기초하여 토익 시험의 읽기 능력 모형의 구인타당도를 살펴보았다.

3. 모형의 타당도 분석

1) 평가 유형을 중심으로 한 읽기 모형의 구인타당도 분석

지금까지 논의한 신뢰도, 상관관계 분석 결과를 바탕으로, 본 연구에서는 모의 토익 시험의 읽기 능력 모형의 구인타당도를 조사하였다. 본 연구는 기존의 타당도 연구에서 많이 사용되는 3가지 기준(예, Brown and Cudek 1993, Chang 2017, Hu and Bentler 1999)에 근거해서 모형의 타당도를 조사하였다. 본 연구에 사용된 세 가지 기준은 모형의 적합도 지수, 모형의 부적합 지수, 그리고 개별지수의 적합성이다. Hu와 Bentler(1999)는 모형의 상대적 적합도 지수(Comparative Fit Index; CFI)가 .95이상이어야 한다고 주장한다. 참고적으로, 모형의 상대적 적합도 지수의 최대치는 1.0이다. 다시 말하면, 모형이 변인들의 관계를 95%이상 잘 설명할 때에만 모형의 타당도가 있다는 것이다. Brown과 Cudek(1993)은 모형의 부적합도 지수인 RMSEA의 수치가 .05 이하이어야 한다고 주장한다. 이론적으로 모형의 부적합성이 전혀 없다면 RMSEA의 수치는 0이 된다. 부적합 지수는 일종의 오류 값으로 오류 값이 100분에 5 이하이어야만 모형의 타당도가 있다는 것이다.

끝으로 개별지수의 적합성 여부이다. 개별지수는 이론적 허용 범위 안에 존재해야 한다. 이론적으로 변인들 간의 상관계수는 1.0을 넘을 수 없는 데, 타당도가 결여된 모형에서는 두 변인 간의 상관계수가 종종 1.0이거나 1.0을 넘는 경우가 있다. 또한 변인들의 상관 계수가 0이거나 0에 가깝거나 음의 수치를 보이는 경우도 있다. 이들은 개별지수가 이론적인 허용치의 범위를 벗어난 경우로 해당 변인들이 담고 있는 이론적 개념에 문제가 있음을 보여준다. 당연히 개별지수에 문제가 있는 모형의 타당도에는 문제가 있다고 할 수 있다.

그림 1은 평가 유형을 중심으로 한 읽기 모형의 타당도 조사 결과를 보여준다. 그림에서 원(circle)으로 표시된 것은 구인(construct)을 뜻한다. 구인은 직접적으로 관측은 할 수 없지만 우리가 학습자들의 머릿속에 존재한다고 강하게 믿고 있는 이론적인 개념인 동시에 시험에서 궁극적으로 측정하고자하는 변인이다. 모의 토익 시험의 읽기 영역은 학습자들의 읽기 능력을 측정하고자 하는 시험이므로, 이 시험의 평가 구인은 읽기 능력(reading ability)이라고 할 수 있다. 네모(rectangle)로 표시된 것은 측정 변인을 뜻하며, 본 연구에서는 모

의 토익 시험의 읽기 영역의 4가지 평가 유형들로부터 얻은 학습자들의 점수를 나타낸다. 화살표는 변인들 간의 관계를 나타낸다. 그림에서는 읽기 능력 구인이 이를 구성하는 네 가지 평가 유형 또는 구성 요소들과 모두 화살표로 연결이 되어있다. 즉 이것은 읽기 능력이 미완성의 문장을 완성하는 능력, 담화를 완성하는 능력, 지문 내용을 이해하고 질문에 답하는 능력 등과 관련이 있다는 것을 보여준다. 한마디로, 미완성의 문장이나 담화를 완성하고 주어진 일련의 지문의 내용을 이해하는 능력을 모두 합쳐 하나의 읽기 능력이라고 가정한 것이다.

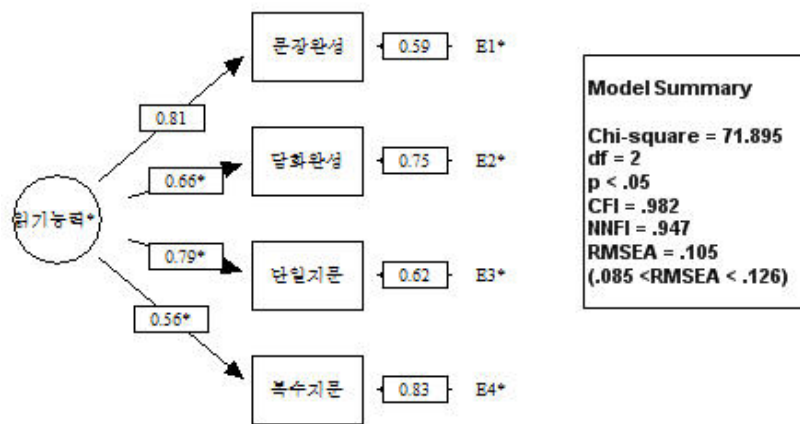


그림 1. 유형 중심 읽기 모형의 타당도

그림 1은 평가 유형 중심 읽기 모형의 적합성에 문제가 있음을 보여준다. 전체 모형의 상대적 적합도를 나타내는 CFI 지수는 .982였다. 이 수치 하나만 본다면 모형의 적합도 준거인 .95보다 높았으므로 제시된 모형의 구인타당도가 있다고 볼 수 있겠다. 또한 모형에 제시된 개별지수들 역시 그 수치가 0이거나 1이거나 음의 값을 갖지 않았고 모두 이론적인 허용 범위 내에 있었다. 이상에서 보면 제시된 읽기 모형은 얼핏 구인타당도가 있어 보인다. 그러나 문제는 모형의 부적합도 지수에 있다. 모형의 부적합도 지수인 RMSEA의 값을 보면 .105로 기준 값인 .05보다 큼을 알 수 있다. 부적합도 지수가 높다는 것은 그림 1에 제시된 변인들의 관계를 보다 더 잘 설명하는 모형이 여럿 존재할 수 있음을 시사한다. 따라서 본 모형은 구인타당도 기준 3가지를 모두 충족하지 못하였으므로, 모형의 구인타당도는 낮다고 판단할 수 있다.

2) 평가 구인을 중심으로 한 읽기 모형의 구인타당도 분석

이상의 결과에 기초해서, 평가 구인 중심의 읽기 모형의 타당도를 검증하였으며 그 결과는 그림 2에 제시되어 있다.



그림 2. 구인 중심 읽기 모형의 타당도

그림을 보면, 읽기 능력은 크게 두 가지 구인으로 이루어졌다고 설명한다. 두 구인 중 하나는 언어지식이고 다른 하나는 읽기 기술이다. 읽기 능력을 언어 지식과 읽기 기술로 설명하려는 것은 새로운 생각이 아니며, 영어 교육 분야의 많은 학자들이 공통적인 견해이다(예, Alderson 2000, Choi and Cheon 2006). 그림 2는 또한 읽기 능력의 구인 중 하나인 언어 지식은 문법지식과 어휘 지식을 모두 포함하는 개념이며, 읽기 기술은 사실적 이해력과 추론적 이해력 모두를 포함하는 개념이라는 것을 도식적으로 보여준다. 또한 이들 두 구인은, 쌍방향 화살표로 이어진 두 원으로 표시하였듯이, 연관성은 있으나 서로 다른 능력이라는 것도 나타내고 있다.

그림 2는 구인 중심 읽기 모형의 적합성이 아주 뛰어남을 보여준다. 그림에서 보듯이, 전체 모형의 상대적 적합도를 나타내는 CFI 지수는 1.000이었다. 모형의 부적합 정도를 나타내는 RMSEA 지수는 .000이었다. 또한 모형에 제시된 개별지수들 역시 그 수치가 0이거나 1이거나 음의 값을 갖지 않았고 모두 이론적인 허용 범위 내에 있었다. 따라서 본 모형은 구인타당도 기준 3가지를 모두 충족하였을 뿐만 아니라 제시된 수치 모두가 모형의 구인타당도가 아주 높다는 것을 보여주었다. 이상에서 논의된 모든 결과를 종합하면, 본 토익 시험을 모방하여 본 연구에서 사용된 모의 토익 시험은 신뢰도와 타당도를 모두 갖춘 교육적으로 유용한 평가도구라고 할 수 있다.

위의 연구 결과에 기반을 두어 다음과 같은 이론적인 해석을 할 수 있다. 첫째, 모의 토익 시험을 통해서 유추해본 본 토익 시험의 평가 구인도 모의 토익 시험의 그것과 유사할

것이다. 그림 2에서 보듯이, 토익 시험의 평가 구인도 크기는 언어지식과 읽기 기술로 이루어졌다고 보아야 할 것이다. 그림을 보면 또한 구인들과 이들의 구성요소인 변인들과의 관계가 화살표로 표시되어 있고 화살표 옆에는 수치가 제시되어 있다. 이들 수치를 요인적재량이라고 부르는데, 요인적재량은 구인의 구성요소로서 변인들의 상대적 중요성을 나타낸다고 할 수 있다. 언어지식의 구성요소로서의 어휘지식의 요인적재량은 .82로 문법지식의 요인적재량인 .77보다 높다. 이것은 모의 토익 시험이 본 토익 시험의 구인을 잘 모방하여 설계되었다는 전제 하에서는, 어휘지식이 문법지식 보다 더 중요한 읽기 능력의 일부이며, 언어지식의 일부라는 것을 시사한다. 또한 사실적 이해와 추론적 이해의 요인적재량을 보면 각각 .92와 .62이다. 적어도 모의 토익 시험이나 본 토익 시험에 관한, 사실적 이해력이 추론적 이해력보다 더 중요한 읽기 능력의 일부이며 또한 읽기 기술의 일부라고 할 수 있다.

3) 모의 토익 시험의 예언타당도 분석

이상에서 논의된 모의 토익 시험의 구인타당도에 기초하여, 모의 토익 시험의 예언타당도(predictive validity)를 조사하였다. 모의 토익 시험의 예언타당도는 모의 토익 시험을 치른 신입생들 중에서 졸업 자격을 갖추기 위해 영어 공인 성적으로 졸업 전에 본 토익 시험 성적을 제출하여 졸업한 82명의 학생을 대상으로 하였다. 본 토익 시험 성적을 제출한 학습자 수가 상대적으로 적은 이유는 다음과 같다. K-대학의 학생들은 졸업 자격을 맞추기 위해 공인 토익 시험 성적만을 제출해야 되는 것은 아니다. TEPS나 TOEFL 등 다른 영어 시험 성적을 제출하여도 무방하다. 재학생들은 시험 비용이 보다 저렴한 TEPS 성적을 선호하는 경향이 있으며, 일어나 독어, 불어 등 다른 외국어 공인 성적을 제출해도 된다. 또한 전산 관련 자격증을 따거나 과목을 수강하여도 졸업 요건을 갖추 수 있는 등 다양한 방법으로 졸업 요건을 충족할 수 있기 때문이다.

모의 토익 시험의 예언타당도 결과를 분석하기 전에 주목할 점이 있다. 첫째, 본 연구에서는 모의 토익 시험의 읽기 영역 점수만을 다루고 있으나 본 토익 시험 점수는 듣기와 읽기 두 영역의 점수를 합산한 점수라는 점이다. 동일한 읽기 영역 성적의 관계를 분석하였을 때와 동일한 해석을 하기에는 어려운 점이 있다. 둘째, 재학생들의 단과 대학별로 졸업요건에서 정한 토익 시험 점수는 서로 달랐던 점도 결과 해석에 주의해야 할 필요가 있어 보인다. 예를 들어 사범대 영어 교육과 학생들은 본 토익 시험 점수 기준으로 최소 800점 이상을 제시하여야 하지만 다른 단과 대학 다른 학과 학생들은 그 보다 낮은 점수를 제출하여도 졸업에 지장이 없다. 셋째, 이러한 최소 기준이 있다는 것은 학생들이 본교 재학 중에 본 토익 시험을 봐서 일정 점수 이상 넘으면 굳이 자신의 최고 점수를 제출할 필요가 없다는 것을 의미한다. 따라서 어떤 학생은 1학년 2학기에 또 어떤 학생은 4학년 1학기에 본 토익 시험 점수를 제출할 수도 있으며, 자신의 최고점이 아닌 최소 기준 이상의 점수를 제

출한 학생이 다수라고 판단할 수 있다.

표 4는 모의 토익 시험의 읽기 영역과 본 토익 시험의 예언 타당도를 나타낸다.

표 4. 모의 토익 시험과 본 토익 시험의 예측 타당도 (N=82)

변인	1	2	3
본 토익 시험 점수	1.000		
모의 토익 시험 언어지식 점수	.601**	1.000	
모의 토익 시험 읽기기술 점수	.497**	.661**	1.000

1: 본 토익 점수, 2: 모의 토익 언어지식 점수, 3: 모의 토익 읽기기술 점수

** $p < .01$

구인 타당도 분석 결과 모의 토익 시험에서 측정한 읽기 영역의 평가 구인은 언어 지식과 읽기 기술이었다. 표 4는 이들 두 변인들과 본 토익 시험 총점과의 상관관계를 보여준다. 표에서 보듯이, 모의 토익 시험의 언어지식 및 읽기 기술 영역 점수와 본 토익 시험의 총점과의 상관계수는 음의 수치가 없었고, 모두 양의 상관관계를 보였으며 상관계수가 1이거나 1에 가까운 것도 없었다. 이들 변인들 간의 상관계수는 낮게는 .497에서 높게는 .661까지 분포하였다. 모의 토익 시험의 언어지식과 본 토익 시험의 총점과의 상관계수는 .661이었으며, 모의 토익 시험의 읽기 기술과 본 토익 시험 총점과의 상관 계수는 .497이었다. 이들 수치를 보고 판단하면, 모의 토익 시험의 예측타당도가 아주 높지도 않지만 그렇다고 낮지도 않은 수준의 예언타당도를 갖추었다고 볼 수 있다. 모의 토익 시험의 예측타당도가 아주 높지 않다는 것은 모의 토익 시험을 아주 잘 본 학생들 모두가 반듯이 본 토익 시험 또한 아주 잘 볼 것이라고 강하게 예측하기에는 무리가 있지만, 대체로 모의 토익 시험 우수자의 상당수가 본 토익 시험 또한 대체로 잘 보는 경향을 보인다는 것을 시사한다. 예언타당도 결과는 또한 모의 토익 시험 점수로 본 토익 시험 점수를 예측하려면 수험생의 모의 토익 시험 언어지식 영역의 점수가 읽기기술 점수 보다 더 예언타당도가 높다는 것도 시사한다.

V. 결론 및 제언

본 연구는 본 토익 시험을 모방하여 특별히 제작한 모의 토익 시험의 구인타당도와 예언 타당도를 조사하였다. 본 연구는 평가 유형 중심의 모의 토익 시험의 읽기 능력 모형의 구인타당도와 구인 중심의 모의 토익 시험의 읽기 능력 모형의 구인타당도를 구조방정식 모형의 일종인 확인적 요인분석을 통하여 비교, 분석하였다. 구인타당도 분석 결과, 구인 중심의 읽기 능력 모형의 구인타당도가 유형 중심의 읽기 능력 모형의 그것 보다 훨씬 더 높다는 점이 밝혀졌다. 이론적인 측면에서, 본 연구는 토익 시험의 읽기 영역의 구인에 대한 이해를 높였다. 구체적으로, 모의 토익 시험의 읽기 영역은 언어 지식과 읽기 기술의 두 가지

구인으로 이루어 졌으며, 언어 지식은 문법 지식과 어휘 지식으로 이루어졌고, 읽기 기술은 사실적 이해력과 추론적 이해력의 두 구성요소로 이루어졌다는 것도 확인하였다. 본 토익 시험의 구인도 이와 유사하리라는 추정이 가능함도 보여주었다. 본 연구는 모의 토익 시험의 읽기 영역의 구인타당도에 기초하여 이 시험과 본 토익 시험과의 예언 타당도도 조사하였다. 우리는 모의 토익 시험의 읽기 영역이 본 토익 시험 점수와 연관해서 아주 높지도 낮지도 않은 중간 수준의 예언타당도를 갖추었다는 것도 알았다.

이러한 결과는 본 연구의 방법론적인 그리고 교육적인 시사점도 제시한다. 방법론적으로 본 연구는 교육적으로 유용한 시험의 저비용, 고효율적인 제작 방법을 제시한다. 일반 연구자들은 자신의 마음대로 본 토익 시험의 구성이나 설계를 바꿀 수 없다. 하지만 본 연구에서 사용한 방법대로, 본 토익 시험의 문항 분석을 통하여 본 토익 시험과 아주 유사한 모의 토익 시험을 상대적으로 저렴한 비용으로 제작할 수 있을 뿐만 아니라, 일련의 신뢰도, 타당도 분석을 통하여 교육적으로 유용한 시험으로 사용할 수 있음을 보여주었다. 본 연구는 또한 모의 토익 시험의 타당도 연구를 통하여 더 나아가서 본 토익 시험의 타당도 개선 방안 또한 파악할 수 있다는 것을 시사한다. 본 연구에서는 모의 토익 시험의 타당도를 거시적으로 살펴보는 데 그 초점이 있었으므로, 사실적 이해력이 어떠한 구성요소로 이루어졌는지에 대한 조사도 논의도 하지 않았다. 그러나 본 연구에서 사용한 연구 방법론을 이용한다면, 본 토익 시험의 사실적 이해력의 핵심적인 구성요소가 무엇인지 상대적으로 덜 중요한 구성요소가 무엇인지 등을 조사할 수도 있으며 이러한 결과가 누적된다면 본 토익 시험의 개선으로 이어질 수 있는 가능성을 제시하였다. 앞으로 읽기 능력의 구인을 보다 더 세부적으로 살펴보거나 본 연구에서는 다루지 않았던 듣기 능력 관련한 후속 연구가 필요하다고 하겠다.

끝으로, 본 연구는 교육적인 시사점도 제시한다. 모의 토익 시험의 읽기 영역을 통해서 바라 본 K-대학 신입생들은 문법지식보다 어휘지식이 상대적으로 높았다. 따라서 본 연구 결과는 이들 학생들이 공인영어시험인 본 토익 시험을 잘 준비하기 위해서는 토익 시험에서 요구하는 문법영역에 좀 더 학습을 집중할 필요가 있음을 보여주었다. 또한 본교 신입생들은 추론적 이해력보다 사실적 이해력이 상대적으로 낮았기 때문에, 이들이 본 토익 시험에서 보다 더 높은 점수를 얻으려면 다양한 지문에 명시적으로 제시된 사실을 보다 정확하게 파악하는 데 보다 더 많은 노력이 필요함을 시사한다. 교양 영어 과목 등 학생들에게 일반 영어를 가르치는 교수나 강사들은 학습자들의 문법, 어휘, 사실적, 추론적 이해 능력이 균형 있게 발전할 수 있도록 교과 내용을 구성할 필요도 있어 보인다.

참고문헌

Alderson, C. 2000. *Assessing Reading*. Cambridge: Cambridge University Press.

- Chang, J. 2017. A multigroup analysis on the validity of models of reading processes. *Secondary English Education* 10-3, 47-67.
- Browne, M. and R. Cudek. 1993. Alternative ways of assessing model fit. In K. Bollen and J. Long, eds., *Testing Structured Equation Models*, 136-162. Newbury Park, CA: Sage.
- Choi, E. and E. Cheon. 2006. *Teaching of Reading in English as a Foreign Language: Principles and Classroom Applications*. Seoul: Hankookmunhwasa.
- Hu, L. and P. Bentler. 1999. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6, 1-55.
- In'Nami, Y. and R. Koizumi. 2011. Factor structure of the revised TOEIC test: A multi-sample analysis. *Language Testing* 29-1, 131-152.
- Powers, E. 2010. *Validity: What Does It Mean for The TOEIC Tests?* Princeton, NJ: Educational Testing Service.
- Suzuki, M. and C. Dana. 2004. A review of the reading section of the TOEIC. *TESL Canada Journal* 22-1, 16-24.

예시 언어(Examples in): 한국어(Korean)
 적용가능 언어(Applicable Languages): 한국어(Korean)
 적용가능 수준(Applicable Level): 대학생, 성인(Tertiary)

장재학(Chang, Jaehak)
 교수(Professor), 영어교육과(Department of English Education)
 강원대학교(Kangwon National University)
 강원도 춘천시 강원대학길 1(1 Kangwon Daehak-gil, Chuncheon-si, Kangwon-do)
 TEL: 033-250-6634
 Email: jaehakchang@kangwon.ac.kr

논문 접수(Received): 2019년 8월 10일 (August 10, 2019)
 논문 수정(Revised): 2019년 9월 10일 (September 10, 2019)
 게재 확정(Accepted): 2019년 9월 19일 (September 19, 2019)