https://doi.org/10.23073/igll.2025..20.009

글로컬 시대 한국어능력시험 듣기 평가 문항의 구인타당도•

IELTS와의 비교 분석을 중심으로

Construct Validity of the TOPIK Listening Test in the Glocal Era: A Comparative Analysis with IELTS

김미경•• · 권태현•••

국문요약 본 연구는 글로컬 시대의 국제 기준에 부합하는 한국어 평가 체제를 마련하기 위한 방안으로 한국어능력시험(TOPIK)과 국제 공인 영어능력시험(IELTS)의 듣기 평가를 대상으로 구인타당도를 비교·분석하였다. 분석 결과, TOPIK은 짧은 담화에 소수의 문항을 배치하여 정보 확인과 의미 추론을 동시에 요구하였으나 실제로는 구성 지식 중심 문항이 주를 이루었으며, 기능적·사회언어학적 요소 반영이 미흡하였다. 반면, IELTS는 장문 담화 내 문항을 흐름에 따라 배열하여 실제적 수행 능력 측정에 효과적인 구조를 지녔으나, 구성 지식 중심 과제가 과도하게 반영되는 한계가 드러났다. 또한 두 시험 모두 이론적 설계와 실제 운영 간 간극이 확인되었다. 본 연구는 문항 내 요소를 다층적으로 분석함으로써 한국어 듣기 평가의 타당성을 국제 시험과의 비교 속에서 검토하였으며, 이를 통해 TOPIK의 개선 및 국제적 정합성 확보를 위한 기초 자료를 제공한다는 점에서 의의가 있다.

핵심어 한국어능력시험(TOPIK), 아이엘츠(IELTS), 듣기 평가, 구인타당도, 균등다중범주분석

차례

1. 서론
2. 이론적 배경 및 선행 연구
3. 연구 방법
4. 연구 결과
5. 결론 및 제언

1. 서론

최근 수년간 K-콘텐츠는 전 세계적으로 폭발적인 성장을 보이며 문화적·산업적 위상을 강화해 왔다. 드라마, 영화, 대중음악, 게임 등 다양한 분야에서 한국어가 주요 매개 언어로 활용됨에 따라, 한국어 학습에 대한 국제적 수요 또한 지속적으로 증가하고 있다. 이러한 수요는 단순히 문화 향유 차원에 그치지 않고, 고등교육 이수, 전문직 진출, 취업 및 이민 등 구체적이고 실질적인 목적과 결부되면서 한국어 교육의 세분화와 전문화를 요구하는 국면에 접어들고 있다.

이와 같은 한국어 교육의 확대와 맞물려 학습자의 언어 능력을 객관적이고 공정하게 측정할 수 있는 평가 체제 역시 점차 중요해지고 있다. 특히 한국어능력시험(Test of Proficiency in Korean, 이하 TOPIK)은 전 세계 한국어 학습자와 재외 동포를 대상으로 한국어 숙달도를 평가하는 대표적 시험으로 자리매김해 왔다. 그러나 한국어 교육의 수요가 양적으로 확대되고 학습자의 목적이 다양화되는 상황에서, 현행 TOPIK이 실제 의사소통 능력을 타당하게 반영하고 있는지에 대한 점검은 충분히 이루어지지 않고 있다.

특히 듣기 평가 영역은 언어 능력 전반의 토대를 이루

• 이 논문(연구실적물)은 2025학년도 충북대학교 연구년제 지원에 의하여 연구되었음(This work was conducted during the research year of Chungbuk National University in 2025).

•• 충북대학교 사범대학 국어교육과 대학원 석사과정(제1저자)
••• 충북대학교 사범대학 국어교육과 부교수(교신저자)

www.kci.go.kr

는 핵심 요소로, 실제 상황에서의 이해력과 의사소통 능력을 직접적으로 반영해야 한다. 그럼에도 불구하고 현행 TOPIK 듣기 문항이 학습자의 한국어 숙달도를 얼마나 정밀하게 측정하고 있는지에 대한 연구는 제한적으로 이루어져 왔다. 본 연구에서는 TOPIK 듣기 평가 문항이 실제 상황에서 요구되는 언어 능력을 구체적으로 반영하고 있는지를 구인타당도의 관점에서 점검해 보고자 한다. 이는 언어 능력 평가 체제의 고도화에 있어 필수적인 과제라 할 수 있다.

언어 평가의 타당성을 검증하기 위해서는 국제적으로 공신력을 인정받은 평가 체제와의 비교 분석이 효과적인 방법이 될 수 있다. 한국어 교육 환경이 단순한 국제화(globalization)를 넘어, 현지적 맥락과 국제적 표준이 동시에 작동하는 '글로컬(glocal)' 시대에 놓여 있다는 점에서 공신력 있는 국제 언어 능력 평가 체제와의 비교는 한국어능력시험의 질 제고에 유의미한 시사점을 제공할 수 있다. 이러한 관점에서 본 연구는 영어 능력 평가의 대표적 시험인 IELTS(International English Language Testing System)와의 비교를 통해 TOPIK 듣기 평가 문항의 구인타당도를 분석하고자 한다. IELTS는 오랜 기간 축적된 연구와 실증적 검증을 기반으로 전 세계적으로 언어 숙달도 평가의 표준적 지위를 확립해 온 시험이며, CEFR(Common European Framework of Reference for Languages) 수준별 기술 진술에 기반한 평가 방식을 채택하고 있어, 국제 기준에 따라 비교, 분석하기가 용이하다. 또한 IELTS는 Computer-delivered 시험 방식을 통해 디지털 기반 평가를 운영하고 있으며, 이는 최근 IBT 체제를 도입한 TOPIK과 구조적으로 유사하다고 할 수 있다.

지금까지의 선행 연구는 TOPIK과 외국어 시험을 비교 분석함에 있어 문항 수가 제한적이고, 타당도 분석에 있어서 단일 평가 요소만을 추출해 단순화된 분석에 머물렀다. 이에 본 연구에서는 타당도 분석에서 중복 집계를 허용함으로써 하나의 문항에 내재된 다양한 평가 요소를 동시에 반영하고, 보다 포괄적인 분석을 수행하고자 한다. 이를 통해 TOPIK 듣기 평가 문항이 실제 의사소통 능력 평가 도구로서 적합성을 지니고 있는지, 그리고 국제적 기준과의 비교를 통해 어떠한 개선의 여지가 있는지를 규명함으로써, 변화하는 한국어 학습 환경에 부합하고, 실제 언어 사용 상황을 충실히 반영하는 평가 체계 설계를 위한 실질적 개선안을 마련하는 데 기여하고자 한다.

2. 이론적 배경 및 선행 연구

2.1. 한국어 듣기 능력의 구성 요소

한국어 듣기 능력의 구성 요소를 고찰하기에 앞서 의사소통 능력의 구성 요소부터 살펴볼 필요가 있다. 의사소통 능력(communicative competence)이란 용어의 개념을 처음 도입한 Hymes의 연구 이후 의사소통 능력의 개념과 구성 요소는 Canale과 Swain, Canale, Bachman, Bachman과 Palmer 등에 의해 구체화되었다. 여러 연구자들이 구체화한 의사소통 능력의 구성 요소을 정리하면 〈표 1〉과 같다.

〈표 1〉 의사소통 능력의 구성 요소

구분	의사소통 능력	구성 요소
Hymes[1]	의사소통 능력	1) 언어지식 2) 언어사용
Canale과 Swain[2]	의사소통 능력	1) 문법적 능력 2) 사회언어학적 능력 3) 전략적 능력
Canale[3]	의사소통 능력	1) 문법적 능력 2) 사회언어학적 능력 3) 담화적 능력 4) 전략적 능력

1 Hymes, D, *On Communicative Competence*, In J. Price & J. Holmes (Eds.), Sociolinguistics, Harmonsworth, Penguin, 1972.

2 Canale, M. & M. Swain, "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing," *Applied Linguistics* 1, 1980, 1~47.

3 Canale, M, *From Communicative Competence to Communicative Language Pedagogy* 1, London, Routledge, 1983.

www.kci.go.kr

Bachman[4]	의사소통적 언어 능력	1) 언어 능력 (1) 구성 능력: ①문법적 능력 ②문맥적 능력 (2) 실용 능력: ①발화수반 능력 ②사회언어학적 능력 2) 전략적 능력(strategic competence) (1) 판단 (2) 계획 (3) 실행 3) 생리심리학적 기제
Bachman과 Palmer[5]	언어 능력	1) 언어 지식 (1) 구성 지식: ①문법적 지식 ②문맥적 지식 (2) 실용 지식: ①기능적 지식 ②사회언어학적 지식 2) 전략적 능력 (1) 목표 설정 (2) 판단 (3) 계획

Canale과 Swain은 문법적 능력, 사회언어학적 능력, 전략적 능력으로 의사소통 능력을 구분하였으며, Canale은 여기에 담화적 능력을 추가하였다. 의사소통 능력에 대해 새로운 시각으로 접근한 Bachman은 의사소통적 언어 능력을 "언어 사용자와 담화 상황 간에 일어나는 역동적 상호작용을 포함하는 것"으로 보아 정보 전달 이상의 확장된 관점을 취했다. Bachman은 '언어 능력'을 의사소통 상황에서 실제로 언어를 활용할 수 있게 하는 구체적인 지식 체계로 정의하였다. 이와 함께 '전략적 능력'은 이러한 언어 지식을 효과적으로 사용할 수 있도록 지원하는 인지적 자원으로, 언어 사용을 조절하고 조율하는 정신적 기능을 의미한다. 또한 언어 수행의 실행 단계에서는 심리생리학적 기제가 작용하는데, 이는 언어 산출 및 이해와 관련된 신경학적·심리적 과정'으로 설명된다. 이후 Bachman과 Palmer는 Bachman을 보완, 수정하여 언어 능력(Language Ability) 모형을 제시하고 언어 능력 안에 언어 지식과 전략적 능력을 하위 구성 요소로 분류하였다. 언어 지식에서 구성 지식은 문장 단위의 규칙과 문장의 연결 관계를 지배하는 규칙을 모두 포함하여 우리가 사용할 언어 형식을 결정하는 모든 규칙을 말하고, 실용 지식은 언어의 기능적 측면과 관련되는 언표내적 능력과 공손함, 격식, 은유, 언어사용역, 언어 사용의 문화적 측면 등과 관련되는 사회언어학적 능력을 포함한다[6]

의사소통 능력에 대한 다양한 학자들의 연구를 종합해 보면, 의사소통 능력은 단순히 언어 지식을 보유하는 것에 그치지 않고, 특정 상황에서 그 언어 지식을 적절하고 효과적으로 활용하는 능력임을 알 수 있다. 이러한 관점에서 볼 때, 한국어 교육에서의 듣기 평가는 문장을 넘어 담화 단위로 확대되어야 하며, 사회언어학적 및 문화적 요소를 반영해야 한다. Bachman과 Palmer를 중심으로 한국어 듣기 능력을 규명하고자 할 때, 포함되어야 하는 요소를 살펴보면 다음과 같다.

우선 구성 지식은 문법 지식과 텍스트(담화, 이하 텍스트로 통일) 지식 두 가지로 구성되며, 문법 지식은 문장과 음운을 구별하고 이해하는 능력을 포함한다. 이는 언어 사용의 정확성을 요구하며, 어휘의 의미와 사용법, 문법 규칙, 문장 구조(통사적 요소), 발음 및 음운에 대한 지식을 포함하는 형태론적 능력을 의미한다. 즉, 언어의 규칙을 명확히 이해하고 이를 적절히 활용할 수 있는 능력이다. 한국어 듣기에서는 음운 식별, 단어 축약형 파악, 발음 규칙, 문법과 통사적 장치 이해 능력, 문장구조 이해, 어휘의 의미나 쓰임, 문맥 속의 어휘 의미 추측 등이 이에 포함된다. 이와 달리 문맥적 지식은 텍스트 지식으로 볼 수 있는데, 대화 내용을 효과적으로 이해하고, 들은 내용에 대해 논리적이고 정확하게 반응하며 체계적으로 정리할 수 있는 능력을 의미한다. 이는 단순히 문장 단위의 이해가 아닌 전체 담화를 포괄적으로 이해하는 것까지 요구한다. 한국어 듣기에서는 접속사 쓰임, 문맥의 의미, 인용문 이해, 생략이나 도치된 문장 이해, 담화의 상황이나 주제 기능에 대한 이해, 휴지에 대한 의미 이해, 화자의 중심 생각, 예시 등에 나타난 의미를 파악할 수 있는 언어능력을 포함한다.

다음으로 실용적 지식 중 기능적 지식은 의사소통의 목적과 언어 사용 상황에 따라 발화나 문장을 해석하는 능

4 Bachman, L. F., *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press, 1990.

5 Bachman, L. F. & A. S. Palmer, *Language Testing in Practice*, Oxford: Oxford University Press, 1996.

6 강승혜·강명순·이영식·이원경·장은아, 『한국어 평가론』, 태학사, 2006, 78쪽.

www.kci.go.kr

력을 의미한다. 이는 대화나 담화에서 다양한 언어적 기능을 이해하고 적절히 활용하는 능력으로, 자신의 생각과 감정을 표현하는 관념적 기능, 다른 사람의 행동을 유발하는 조작적 기능, 인지 교수학습 및 의도적 기억을 통해 독자적으로 언어를 사용하는 발견적 기능, 그리고 언어를 활용해 창작하거나 예술적 표현을 만들어내는 상상적 기능 등을 포함한다. 끝으로 사회언어학적 지식은 해당 언어권에서 사용되는 방언, 관용 표현, 그리고 더 넓게는 문화적 내용을 이해하는 능력을 포함한다. 이는 담화 참여자 간의 대화를 맥락과 상황에 맞게 해석하는 능력으로, 한국어 듣기에서는 높임말 이해, 종결어미의 의미, 그리고 화자의 의도 등을 이해하는 것이 중요한 요소로 포함된다.

본고에서는 TOPIK과 IELTS 두 시험의 타당도 검증을 위해 Bachman과 Palmer의 언어 능력 구성 요소 평가틀을 기본으로 사용하되, 한국적 듣기 능력 구성 요소를 보완적으로 추가하는 통합적 접근 방법을 시도하고자 한다.

2.2. 한국어와 외국어 평가의 타당도 비교

2000년대를 거치면서 국제적 평가 기준에 부합하는 기타 외국어 평가 시험들과의 비교 분석을 통해 한국어능력시험(TOPIK)의 방향성과 국제 표준과의 적합성을 고찰하여 보다 효과적이고 공정한 평가 방법을 모색하고 개선안을 마련하고자 하는 연구들이 수행되었다. 이 중에서도 듣기 평가와 관련된 대표적인 연구로는 유소정, 송영승 조성해의 연구를 들 수 있다.

유소정은 TOEFL iBT, 신 HSK, 신 JLPT와 한국어능력시험(TOPIK) 듣기 문항의 문항타당도를 비교 분석하였다.[7] 이 연구에서는 문항의 구인을 문법적, 기능적, 담화적, 사회언어학적 지식으로 분류하여 구인 타당도를 분석한 결과, 사회언어학적 지식을 측정하는 문제가 적게 출제되었음을 발견하였다. 또한 기능, 맥락, 내용, 텍스트 종류로 기준 설정하여 문항 내용 타당도 분석 결과 중고급 문항임에도 불구하고 학문 목적과 업무 수행 목적에서 충족을 이루지 못하여 내용 타당도 역시 떨어진다 분석하였다.

이 시기 또 다른 비교 분석 연구로 송영승의 연구를 들 수 있다. 송영승은 한국어능력시험의 초급 듣기영역 문항을 JLPT, 신 HSK, 캠브리지 ESOL(KET)과 비교 분석하여 TOPIK이 실질적 한국어 숙달도를 측정하는 데 얼마나 적합한지 평가하였다.[8] 외국어능력시험들의 구성요소와 평가 유형들의 특징을 살펴보았는데 구성요소로는 의사소통능력 측정을 위한 평가로 문법적, 화용적, 사실적, 추론적 능력으로 나눠 분석하였다. 그 결과 외국어능력시험은 구성요소와 문항 유형의 일치도가 한국어능력시험에 반해 높았으며 구성 요소에 적합한 문제가 좀 더 다양하게 출제되고 있음을 확인하였다. 그에 따른 한국어능력시험의 문항 유형 타당도를 높이기 위한 구체적인 개선안을 제안하였으나 문항 배열의 구성요소 비율 제시와 선택지 형식에 대한 재고 등의 제안에 있어서 주관성을 배제하기 어렵다는 한계점이 보인다.

조성해는 유학생의 한국어 듣기 평가를 위한 한국어능력시험의 내용 타당도를 유학생들 선발 목적을 위한 제 2 언어 평가로 안정적으로 사용되고 있는 TOEFL iBT와 비교 분석하였다.[9] 이 연구에서는 카이제곱 검정을 활용하여 두 시험의 듣기 영역에서 내용 타당도의 차이가 유의함을 우선 확인하고 각각의 문항 분석으로 내용 타당도의 차이를 구체적으로 규명하였다. 모든 지문들이 강의, 면담, 행정 담화들로 구성된 TOEFL과 달리 TOPIK은 목표 상황과 관련성이 떨어지고 면담 상황은 아예 고려되지 않았음이 분석되었고 문제 유형도 TOPIK은 특정 유형에 편중되어 고르게 분포되지 못한다는 결론 하에 현행 시험들은

7 유소정, 「외국어능력시험과의 비교를 통한 한국어능력시험 듣기 문항 분석 연구」, 경희대 교육대학원 석사학위논문, 2012.

8 송영승, 「한국어능력시험의 초급 듣기영역 문항 분석 연구」, 영남대 대학원 석사학위논문, 2013.

9 조성해, 「유학생의 한국어 듣기 평가를 위한 한국어능력시험의 내용 타당도 분석 연구」, 『외국어로서의 한국어교육』 64, 연세대 언어교육연구원, 2022.

www.kci.go.kr

유지하되 타당도를 높일 수 있는 개선 방안을 제안하였다.

이러한 선행 연구들은 외국어 능력 평가 시험들과의 비교를 통해 한국어능력시험(TOPIK)의타당성과 진정성을 검토하고, 문항 구성 및 평가 방식에서 개선점을 제시하는 데 중요한 역할을 해왔다. 그러나 이러한 연구들은 몇 가지 한계점을 보였다. 첫째, 많은 연구들이 특정 회차나 제한된 데이터를 분석 대상으로 하여 연구 결과의 일반화에 한계가 있다. 둘째, 문항 타당도와 진정성을 분석하였으나, 학문적 또는 실생활 상황에서의 실제적 활용성 평가는 미흡하였다. 셋째, 하나의 문항이 여러 평가 요소에 동시에 영향을 줄 수 있음에도 불구하고, 대부분의 기존 연구들은 단일 핵심 요소만을 중심으로 분석을 수행하여 문항이 지닌 복합적인 평가 기능을 충분히 반영하지 못하였다.

본 연구는 이러한 선행 연구들을 토대로, TOPIK과 IELTS의 듣기 평가 영역을 구체적으로 비교, 분석하고자 한다. 특히 평가 환경과 체제에 대한 고찰에 기반하여, 구인 타당도를 중심으로 두 시험의 문항 설계와 평가 방향을 분석할 것이다. 이를 통해 두 시험이 학문적, 업무적인 상황과 실생활에서 학습자의 실질적인 듣기 능력을 얼마나 유효하게 평가하는지를 심층적으로 검토하고자 한다. 나아가, 한국어 학습 환경에 부합하는 한국어 듣기 평가 방안의 개선점을 제안하고자 한다.

3. 연구 방법

3.1. 분석 대상

본 연구의 분석 대상은 TOPIK과 IELTS의 듣기 평가 문항이다. IELTS는 약 140개국에서 시행되고 있으며, 11,500개 이상의 교육 및 직업 기관에서 공신력 있는 평가로 인정받고 있어 학문적 및 실용적 목적 모두에서 활용도가 높다.[10] 또한 TOPIK과 IELTS 모두 유학, 이민, 취업 등 실용 목적의 비모국어 화자를 주요 응시 대상으로 하며, 학문 및 직무 상황에서의 언어 능력을 평가한다는 점에서 비교 분석에 타당성이 있다.[11]

TOPIK은 2023년부터 기존의 PBT(Paper-Based Test) 체제 외에 IBT(Internet-Based Test)가 도입되었으며, 일부 국가에서는 IBT 전용 시험이 운영되고 있다. IBT는 총 문항수와 시간은 축소되었으나 문항 구성 측면에서는 PBT와 실질적인 차이는 없다. 본 연구에서는 2014년부터 2023년까지의 TOPIK I 과 TOPIK Ⅱ 기출 문항을 대상으로 하였다.

IELTS도 Paper-based IELTS(PB IELTS)와 Computer-delivered IELTS(CD IELTS)가 있으며 시험 문제는 동일하다. 다만 IELTS는 보안성과 공정성을 유지하기 위한 조치로 공식적으로 과거 시험 문제 전체를 공개하지 않는다. 본 연구에서는 IELTS 출제 기관인 Cambridge Assessment English가 직접 편집, 출간한 가장 신뢰도 높은 실전형 기출 유사 자료를 연구 대상 자료로 채택하였다. 시험을 개발하는 기관에서 출판한 자료이므로 시험 체제와 형식이 실제와 동일하며 이는 연구 논문에서 신뢰 가능한 공식 준거 자료로 활용이 가능하다 판단하였다. 또한 대부분의 문제들이 과거 실제 시험에서 사용되었거나 그와 유사한 방식으로 재구성되어 있어 TOPIK과의 비교 연구에서 현실성 있는 문항 설계 비교가 가능하다.

이상에서 밝힌 본고의 분석 대상을 정리하면 〈표 2〉와 같다.

10 IELTS(2024). IELTS for immigration. Retrieved May 27, 2025, from https://ielts.org/organisations/ielts-for-organisations/ielts-for-your-sector/ielts-for-immigration

11 IDP IELTS(2024). Who takes IELTS and why. Retrieved May 27, 2025, from https://ielts.idp.com/about/news-and-articles/article-who-takes-ielts

www.kci.go.kr

〈표 2〉 분석 대상[12]

시험 종류	시험지	총 분석 문항 수
TOPIK	35회, 36회, 37회, 41회, 47회, 52회, 60회, 64회, 83회, 91회	TOPIK Ⅰ 총 300문항 TOPIK Ⅱ 총 500문항
IELTS	CAMBRIDGE IELTS 17 TEST 1-4회 CAMBRIDGE IELTS 18 TEST 1-4회 CAMBRIDGE IELTS 19 TEST 1-4회	총 480문항

3.2. 분석 기준

이 연구에서는 TOPIK과 IELTS 듣기 평가 문항의 구인 타당도 분석을 위해 언어 능력의 구성 요소를 다차원적으로 해석할 수 있는 통합적 접근 방식을 채택하였다. 기본적인 분석 틀로는 Bachman과 Palmer가 제시한 언어 능력 구성 요소 체계를 활용하였다. 이는 시험 문항이 요구하는 언어 능력을 구조적으로 파악하는 데에 객관적인 기준을 제공한다. 그러나 본 연구는 TOPIK 한국어 듣기 평가의 실제적 특성과 교육적 맥락을 개선안에 반영하기 위해 TOPIK 타당도 분석 시 Bachman과 Palmer의 틀에 더하여 국내 연구자들이 제시한 평가 요소를 보완적으로 통합하였다. 구체적으로는 강승혜 외[13]와 강현화[14]가 제안한 한국어 듣기 평가 목표와 세부 평가 기준을 참고하여, 한국어의 사회문화적 맥락 및 청해 전략이 반영된 분석 틀을 재구성하였다. 이를 통해 본 연구는 국제적 공신력을 확보하면서도 한국어 교육의 특수성을 반영하고자 시도하였다. 〈표 3〉은 본 연구에서 재구성한 듣기 평가 문항 분석 틀이다.

〈표 3〉 듣기 평가 문항 분석 틀

상위 구성 요인	하위 평가 요인	하위 능력 요인
문법 지식	어휘 지식	– 어휘의 의미와 사용법 이해 – 문맥 속 어휘 의미 추측
	통사 지식	– 문장 구조(통사적 요소) 이해 – 문법 규칙의 이해와 활용 – 통사적 장치 이해
	음운적 지식	– 음운 식별 능력 – 발음 규칙 이해 – 단어 축약형 파악
텍스트 지식	응집성 지식	– 접속사 쓰임 이해 – 생략, 도치 등 구조적 연결 장치 이해 – 휴지(pause)의 의미 이해
	수사적 구성 지식	– 담화의 주제, 기능, 전개 방식, 사실 정보 이해 – 화자의 중심 생각 파악 – 구체적 수치 및 비교 정보 이해 – 인용문 이해 – 예시 등을 통한 의미 파악
기능적 지식	자신의 생각 표현 기능	– 자신의 생각, 느낌 표현
	조작적 기능	– 상대의 행동 유발을 위한 발화 이해 – 지시, 요청 기능 이해
	독자적 학습 기능	– 학습을 통한 독자적 언어 기능 수행
	상상적 기능	– 창작적/예술적 표현에 필요한 상상력 발휘
사회언어학적 지식	방언/전문용어/은어 지식	– 지역 방언, 전문용어, 은어에 대한 이해
	관용적 표현 지식	– 관용어, 속담 등 관용적 표현의 의미 파악
	문화적 지식	– 한국어 담화의 문화적 맥락 이해 – 높임 표현 및 종결어미의 문화적 의미 이해 – 겸손/겸양 표현의 사회적 맥락 이해

3.3. 분석 방법

본 연구에서는 분석자 간 일치도 검증을 통해 분석의 타당성과 신뢰도를 확보하였다. 본 연구자(한국어 교원 자격증 2급 소지자)를 비롯하여 현 한국어 교육 기관에서 활동 중인 경력 5년 이상의 연구자 두 명이 제 3의 분석자로 참여하여 동일한 기준으로 문항에 대한 분석을 수행하였으며, 범주 적용 결과의 일치도를 확인하였다. 이를 통해 분석 기준의 적용이 임의적으로 변하지 않도록 검증하였고, 분석자 간 판단 차이를 최소화하였다.

또한 본 연구에서는 문항의 구인타당도 분석을 위해

12 TOPIK 듣기 평가는 한 회당 Ⅰ은 30문항, Ⅱ는 50문항으로 출제된다. 그러므로 TOPIK Ⅰ 듣기 평가 10회차, 총 300 문항과 TOPIK Ⅱ 듣기 평가 10회차, 총 500문항을 분석하였다. 또한 IELTS의 경우 CAMBRIDGE IELTS 17, 18, 19권의 TEST 1~4 총 12회분, TEST 당 40 문항으로 총 480문항을 대상으로 분석을 실시하였다.

13 강승혜 · 강명순 · 이영식 · 이원경 · 장은아, 『한국어 평가론』, 태학사, 2006, 103~107쪽.

14 강현화 · 김선정 · 김은애 · 김정숙 · 이미혜 · 최은규, 『외국어로서의 한국어 교육학』, 한국방송통신대 출판부, 2011, 199쪽.

www.kci.go.kr

기존과 다른 접근 방식을 취하였다. 기존의 선행 연구들에서는 주로 각 문항에서 하나의 핵심 평가 범주만을 지정하여 통계를 산출하는 우선 순위 방식을 채택해 왔다. 이 방식은 각 문항의 대표적인 특성을 간단히 요약할 수 있고, 분석 결과의 해석이 비교적 용이하며, 제한된 시간과 자료 환경에서도 효율적으로 적용할 수 있다는 장점이 있다. 그러나 각 문항은 실제로 문법적, 담화적, 기능적, 사회문화적 요소가 복합적으로 작용하여 정답 도출에 기여하는 경우가 많기 때문에, 단일 범주만을 선택하는 방식은 문항의 복합적 특성을 충분히 반영하지 못하는 한계가 있다. 더불어 핵심 범주 선정 과정에서 분석자의 주관이 개입될 가능성도 존재하여, 분석 결과의 객관성과 재현성을 저해할 우려가 있다.

이에 본 연구는 이러한 한계를 극복하고, 문항의 구성적 타당성을 보다 정확히 파악하기 위해 균등 다중 범주 분석 방식을 채택하였다. 이 방식은 하나의 문항이 두 개 이상의 평가 범주에 기여할 경우, 모든 해당 범주를 식별하여 포함시키고, 각 범주에 대해 동일한 비중으로 빈도를 산출하는 방식이다. 즉, 특정 요소에 가중치를 부여하거나 배제하지 않고, 문항이 포함하고 있는 언어 능력 구성 요소를 고르게 반영하여 전체 시험의 구인 구성을 파악하고자 하였다.

특히 TOPIK과 IELTS처럼 다양한 상황과 담화를 포괄하는 공인 언어 능력 평가에서는 하나의 문항이 문법 규칙의 이해, 담화의 논리적 전개 파악, 지시 수행과 같은 기능적 목적, 사회문화적 맥락 해석 등 다차원적 언어 처리 능력을 동시에 요구한다. 이러한 특성은 하나의 평가 범주만으로 설명이 어려우며 복수 요소가 상호작용하는 실제 의사소통 상황을 반영하기 위해서는 각 평가 범주의 동등한 반영이 필수적이다. 균등 다중 범주 방식은 이러한 복합성과 실제성을 분석 과정에 충실히 반영함으로써, 분석자의 주관적 판단 개입을 줄이고, 분석 결과의 객관성 및 신뢰성을 높일 수 있다는 점에서 타당성이 있다고 판단하였다.

4. 연구 결과

4.1. 평가 범주별 분석 과정

TOPIK은 짧은 담화에 1~2개의 문항이 연계된 구조로 단일 발화의 핵심 정보와 암시적 의미, 맥락적 단서를 함께 고려해야 정답에 도달이 가능하도록 설계되어 있다. 따라서 복합적 언어 지식이 작동해야 하며 이는 타당도 분석 시 복수의 하위 평가 범주가 동시에 반영되어야 하는 근거가 된다. 반면, IELTS는 하나의 장문 담화에 연속적으로 제시된 10개 문항이 특정 구간에 정밀하게 대응되도록 구성되어 있으며, 대부분 담화 전개 흐름 속에서 문항별로 특정 요소 하나가 정답 결정의 근거가 되도록 설계되어 있다.

본 연구에서는 타당도 검증의 분석 과정에서 모호성을 줄이고 평가 요소 판정의 신뢰도를 높이며 실제적 교육적 맥락을 반영하기 위해 세 가지 절차를 사전 설정하였다. 첫째, 두 시험의 비교 분석에는 일관성과 객관성을 확보하기 위해 〈표 3〉과 같이 Bachman과 Palmer의 언어 능력 구성 요소를 기본 분석 틀로 적용하되, 개선안 제시를 위해 강승혜 외와 강현화의 한국어 청해 평가 기준을 보완적으로 통합한 통합적 접근 방식을 채택하였다. 둘째, 문항당 하위 평가 요인의 균등 다중 범주 방식을 채택하였으나, 동일 문항에서 같은 하위 평가 요인의 중복을 피하고 각 평가 요인 당 하위 능력은 하나씩만 반영되도록 설정하였다. 이는 하위 평가 요인들의 균등한 범주 분포를 확보하여 분석의 초점을 명확히 하고 동시에 제한된 시간과 제약된 자료 환경 속에서 효율적 적용을 위한 방법론적 선택이었다. 셋째, 사전 정의된 분석 기준을 확립하여 문항 내에 명시적으로 제시되었거나 정답 유도에 실질적으로 기여한 하위 평가 범주만 분석에 포함하였다. 예로 문법 지식(어휘, 통사, 음운적 지식)은 모든 문항에서 정답 도출을 위해 기본적으로 작동하지만, TOPIK II에서는 정답 결정의 핵심 기제가 아니므로 분석에서는 모두 제외하였다. 또한 판

www.kci.go.kr

단의 일관성을 유지하기 위해 임의 해석 개입을 최소화하였다. 이를 위해 담화 지문 및 음성 정보뿐만 아니라 선택지의 언어적 구성과 기능도 판단의 근거로 활용하여 분석을 진행하였다. 구체적으로 선택지 분석 시 발화만으로는 정답 도출이 불가능하고 선택지 간 비교 추론이 요구되는 경우는 기능적 지식의 독자적 학습 기능에 포함하였으며, 반대로 지문 내 명시적 발화로 정답이 확정되는 경우는 제외하였다.

전술한 바와 같은 분석 기준 설정에 기반하여 문항 분석의 실제를 제시하면 다음과 같다. 다음 문항은 기능적 지식 중 조작적 기능이 나타나는 문항의 사례이다.

※ [9~12] 다음 대화를 잘 듣고 여자가 이어서 할 행동으로 알맞은 것을 고르십시오. (각 2점)

9. 남자: 여보, 바쁘지 않으면 넥타이 매는 것 좀 도와줘요. 이게 잘 안 매지네요.
여자: 알겠어요. 어? 근데 셔츠에 단추가 떨어졌네요. 금방 달아 줄 테니까 줘 봐요.
남자: 아, 여기. 이걸 못 봤네요. 아까 입을 때는 괜찮았는데…….
여자: 지금 바로 돼요. 잠깐만 기다려 봐요.

① 단추를 단다. ② 셔츠를 준다.
③ 셔츠를 입는다. ④ 넥타이를 맨다.

〈그림 1〉 TOPIK II 41회 9번 문항 지문과 문제[15]

위 문항은 대화 속 여자가 이어서 할 행동을 추론하는 문제로 정답을 도출하기 위해 수험자는 발화 흐름 전체를 해석하고 여자가 언급한 "금방 달아 줄 테니까 줘 봐요"로 행동 예고를 듣고 반영하여야 한다. 이는 지시, 요청 발화에 대한 반응에 대한 이해가 가능한지를 평가하는 문항으로 기능적 지식의 조작적 기능이 요구된다. 정답은 1번이며, 또한 선택지 간 비교 추론이 가능해야 정답 판단이 가능하므로 독자적 학습 기능도 관여한다. 균등 다중 방식에 따라 이 문항의 하위 평가 범주를 분석하면 〈표 4〉와 같다.

〈표 4〉 기능적 지식-조작적 기능 균등 다중 방식 타당도 하위 평가 범주 빈도수 및 비율 통계 예시

하위 평가 범주	빈도수	비율
수사적 구성 지식	1	33.3%
조작적 기능	1	33.3%
독자적 학습 기능	1	33.3%

다음으로 〈그림 2〉는 기능적 지식 구성 요인을 보여주는 IELTS 문항이다. 이 문항은 CAMBRIDGE IELTS 19 TEST 3 Part 3 듣기 26번 문항에서는 실험에 사용할 생쥐의 수와 조건을 묻고 있다. 정답은 C. age로 수험자는 대화 흐름 속에서 제시된 실험 설계 조건으로 발화 안에서 "all young ones, not a mixture of old and young." 즉, 모두 어린 쥐여야 한다는 지시 사항을 정확히 이해하고 반영해야 한다. 21번부터 25번까지 대화 전반에서 수사적 구성 지식을 요구하는 문항들이 제시된 이후, 본 문항 26번에서는 단일한 언어 능력 요소인 기능적 지식, 그 중 조작적 기능을 근거로 정답이 도출된다.

CLARE: So can I talk through a possible procedure for the experiment where mice are given a sugar supplement?
JAKE: Sure. I did a similar experiment in college actually.
CLARE: Great. So how many mice would I need?
JAKE: I'd say about 12. And all young ones, not a mixture of old and young. Q26

Complete the flowchart below.

*Choose **FIVE** answers from the box and write the correct letter, **A–H**, next to Questions 26–30.*

A size
B escape
C age
D water
E cereal
F calculations
G changes
H colour

Choose mice which are all the same **26** ..... .

↓

자료: CAMBRIDGE IELTS 19

〈그림 2〉 CAMBRIDGE IELTS 19 TEST 3 Part 3 26번 문항 해당 지문과 문제[16]

15 TOPIK(https://www.studytopik.go.kr)

16 Cambridge University Press, *Cambridge IELTS 19 Academic: Student's Book with Answers*, Cambridge: Cambridge University Press, 2022.

www.kci.go.kr

〈표 5〉 기능적 지식 포함한 IELTS타당도 하위 평가 범주 빈도수 및 비율 통계 예시

하위 평가 범주	빈도수	비율
조작적 기능	1	100.00%

앞서 언급한 사례에서 확인한 바와 같이 IELTS 문항은 하나의 장문 담화에 연속적으로 10개의 문항이 제시되며, 각각 특정 담화 구간에 정밀하게 대응되도록 설계된다. 따라서 각 문항은 대체로 특정 시점의 정보를 기반으로 하나의 기능을 수행하여 정답이 결정된다. 이는 TOPIK 문항이 다중 범주적 요소를 동시에 반영하는 특성과 뚜렷한 대조를 이룬다.

4.2. TOPIK 시험의 듣기 평가 구인 분석

본 연구는 TOPIK I 듣기 평가 10회차, 총 300문항과 TOPIK II 듣기 평가 10회차, 총 500문항에 대해 타당도 분석을 위해 하위 평가 범주별 빈도 분석을 실시하였다. 전체 문항에 대한 상위 구성 요소 비율은 〈표 6〉와 같다.

〈표 6〉 TOPIK 상위 구성 요소 비율

상위 구성 요소	하위 평가 범주	TOPIK I	TOPIK II
문법 지식	어휘 지식, 통사 지식, 음운 지식	27.78	0.00
텍스트 지식	응집성 지식, 수사적 구성 지식	41.67	62.35
기능적 지식	자신의 생각 표현 기능, 조작적 기능, 독자적 학습 기능, 상상적 기능	26.90	35.90
사회언어학적 지식	방언/전문용어/은어 지식, 관용적 표현 지식, 문화적 지식	3.64	1.74
비율 합계		100.00	100.00

분석 결과, TOPIK I 에서는 문법 지식(27.78%)과 텍스트 지식(41.67%)을 포함한 구성 지식이 전체의 69.45%를 차지한 반면, 기능적 지식(26.90%)과 사회언어학적 지식(3.64%)으로 구성된 실용 지식의 반영 비율은 상대적으로 낮게 나타났다. 이는 Bachman과 Palmer의 언어 능력 구성 이론을 기준으로 볼 때, 전체 검사지가 구성 지식에 과도하게 편중되어 출제되고 있음을 시사한다.

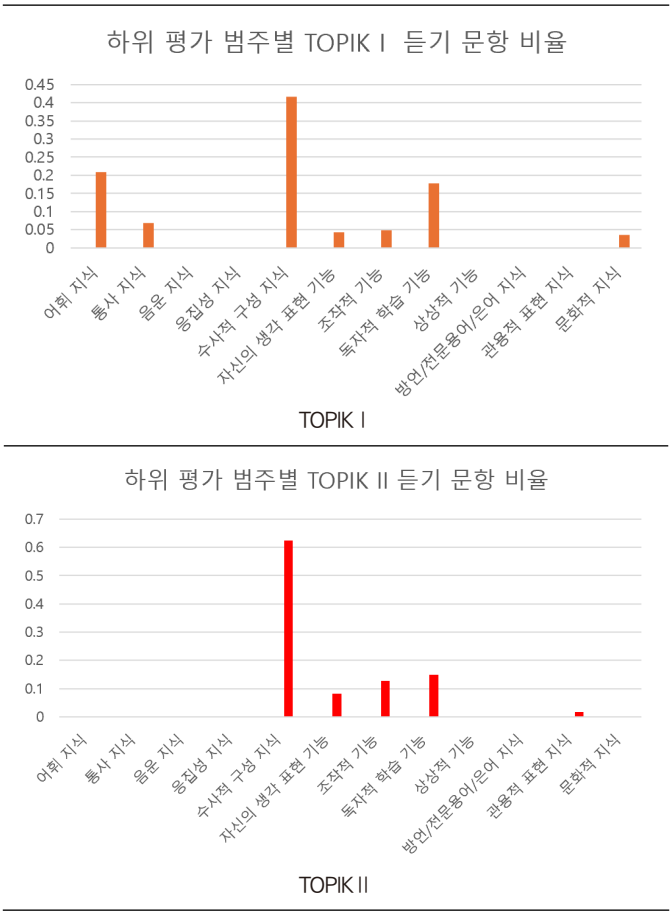
TOPIK II에서는 텍스트 지식이 전체의 62.35%로 가장 높은 비중을 차지하였으며, 문법 지식은 0%, 기능적 지식은 35.90%, 사회언어학적 지식은 1.74%로 나타났다. TOPIK I 에 비해 문법 지식의 비중이 크게 줄고 텍스트 지식의 비중이 크게 늘어났다. 그러나 이러한 구인 비율은 언어 능력의 핵심 축인 구성 지식(문법 지식, 텍스트 지식)과 실용 지식(기능적 지식, 사회언어학적 지식) 중 일부 요소만이 평가에 과도하게 반영되고 있으며 구성 지식에 62.35%로 편중되어 문항이 출제되고 있음을 보여준다. 또한 구성 지식 중에서도 텍스트 지식에 집중되어 문법 지식은 정답 도출에 실질적으로 작용하지 않았으며, 실용 지식 중 사회언어학적 판단을 요구하는 문항 또한 매우 적었다.

이러한 비대칭적 평가 구성은 TOPIK I과 TOPIK II 모두에서 나타났으며 이는 TOPIK 시험이 언어 능력의 다면성을 충실히 반영하지 못한다는 점에서 구인타당도의 결함을 나타낸다. 즉, TOPIK 시험은 듣기 문항을 통해 다양한 의사소통 상황에서의 언어 능력을 평가하고자 하나, 실제 문항 구성은 특정 능력에만 편중되어 있어 의사소통 수행 능력을 포괄적으로 측정하는 데 한계가 있음을 보여준다.

하위 평가 범주별 문항 비율을 살펴보면 〈그림 3〉과 같다. 분석 결과, TOPIK I은 초급 대상 평가임에도 불구하고 상대적으로 다양한 언어 능력 요소가 반영된 것으로 나타났다. 이는 문항 난이도가 상승하면서 지문의 길이가 길어지나, 정답 도출에필요한 정보의 복잡성과 위치가 언어 능력을 사용하는데 더 큰 영향을 미친다는 점을 보여준다. 예로 짧은 지문이라도 정답 도달을 위해 선택지 간 비교나 추론이 필요한 경우 기능적 유추 능력과 텍스트의 구조적 이해 능력이 동시에 작동하여 복수의 평가 범주가 요구되었다. 반면, 발화 내에서 정답이 명시적으로 제시된 경우

www.kci.go.kr

언어 능력 사용이 제한적으로 나타나 평가 범주가 단일화되는 현상이 관찰되었다. 실제로 TOPIK I은 어휘 지식과 통사 지식이 일정 비율 반영되고 동시에 자신의 생각 표현 기능, 조작적 기능, 독자적 학습 기능 등의 기능적 지식도 일정 부분 반영되어 초급임에도 상대적으로 폭넓은 능력이 필요하도록 설계된 경향을 보였다.



〈그림 3〉 **하위 평가 범주별 TOPIK 듣기 문항 비율**

반면, TOPIK II는 하위 평가 범주가 상대적으로 제한적이었다. 수사적 구성 지식이 절대적으로 높은 비율을 차지하였으며, 그 외 자신의 생각 표현 기능, 조작적 기능, 독자적 학습 기능이 일부 보완적으로 작동하는 수준에 그쳤다. 그러나 어휘 지식과 통사 지식은 사실상 반영되지 않았는데, 이는 중고급 단계에서 개별 어휘 의미나 단순 문장 구조보다 담화 전체의 흐름과 정보 관계를 추적하는 능력이 더 중시됨을 시사한다. 또한 상당수 문항은 지문 속에서 제시되는 정보를 직접 확인하는 수준에 머물러, 특정 하위 평가 범주(수사적 구성 지식)에 과도하게 편중되는 양상을 보였다. 특히 일부 과제에서는 전문적 스키마 기반의 단일 이해만으로 정답 선택이 가능하도록 설계되어 있었다.

종합하면, TOPIK 두 급 모두에서 언어 능력의 다면성을 균형 있게 반영하지 못하고 특정 범주에 집중된 출제 경향을 보였다. 비교적 TOPIK I은 어휘, 통사 지식에서 기능적 지식까지 상대적으로 더 넓은 범주가 작동한 반면, TOPIK II는 난이도가 상승되었음에도 불구하고 수사적 구성 지식 중심으로 과도하게 편중되는 양상이 더욱 두드러졌다.

4.3. IELTS 시험의 듣기 평가 구인 분석

이어 본 연구에서는 CAMBRIDGE IELTS 17, 18, 19권의 TEST 1~4 총 12회분(총 480문항)을 대상으로, 타당도 검토를 목적으로 하위 평가 범주별 빈도 분포를 분석하였다. IELTS 듣기 평가 문항의 상위 구성 요소 비율은 〈표 5〉와 같다.

〈표 5〉 **IELTS 듣기 문항의 상위 구성 요소 비율**

상위 구성 요소	하위 평가 범주	비율(%)
문법 지식	어휘 지식, 통사 지식, 음운 지식	2.29
텍스트 지식	응집성 지식, 수사적 구성 지식	69.79
기능적 지식	자신의 생각 표현 기능, 조작적 기능, 독자적 학습 기능, 상상적 기능	27.71
사회언어학적 지식	방언/전문용어/은어 지식, 관용적 표현 지식, 문화적 지식	0.21
비율 합계		100.00

하위 평가 범주별 빈도 분석을 실시한 결과, IELTS에서는 텍스트 지식(69.79%)과 기능적 지식(27.71%)의 비율이 매우 높게 나타났으며, 문법 지식(2.29%), 사회언어학적

www.kci.go.kr

지식(0.21%)의 비중은 극히 미미한 것으로 확인되었다. 이와 같은 구성 비율은 IELTS 듣기 평가의 문항 설계 구조와 평가 방식의 특수성을 반영한다.

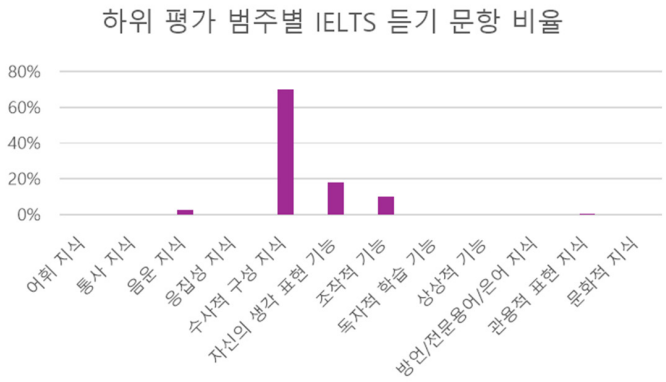
IELTS 듣기 평가에서는 하나의 장문 담화를 기반으로 연속된 10개의 문항이 출제되며, 각 문항은 해당 담화 내 특정 구간에 대응되어 설계된다. 수험자는 발화의 흐름을 따라가며 지시된 시점의 정보만을 정확히 청취하여 정답을 도출하면 되며, 대부분의 문항이 단일한 핵심 언어 능력 요소(특정 정보 확인, 발화 목적 이해 등)에 근거하여 정답이 결정되도록 구성되어 있다. 이러한 구조적 특성은 각 문항이 요구하는 능력을 명확히 분리하고, 수험자의 수행을 보다 객관적으로 측정할 수 있다는 장점이 있다.

특히 전체 문항의 약 70%에 해당하는 문항이 텍스트 지식, 그 중 수사적 구성 지식을 기반으로 출제되었으며, 이는 장문 담화의 전개 구조, 세부 정보의 위치, 논리적 흐름 이해 등 담화 구성 능력을 주요 평가 대상으로 삼고 있음을 보여준다. 기능적 지식은 전체의 약 27.71%로 나타났으며, 이는 주로 발화자의 의도, 견해, 계획 등을 유추하는 과제를 통해 실용적 의사소통 능력을 측정하고자 한 구성으로 해석된다.

반면, 문법 지식(2.29%)이나 사회언어학적 지식(0.21%)은 거의 반영되지 않았으며, 이는 IELTS의 평가 목적이 구조적 언어 지식이나 화용적 사회 판단보다는, 특정 상황 내 정보 이해와 기능적 수행 능력에 보다 초점을 맞추고 있음을 시사한다.

이와 함께 IELTS의 각 하위 평가 범주별 듣기 문항의 비율을 그래프로 나타내면 〈그림 4〉와 같다. 주목할 점은 IELTS에서 음운 지식을 직접 평가하는 특수 유형이 존재한다는 것이다. 이는 TOPIK에는 존재하지 않는 문항 유형으로 특정 장소명이나 고유명사를 발화자가 철자화(spelling)하여 제시하고, 이를 정확히 식별하도록 요구한다. 이는 수험자가 단순히 의미를 이해하는 것이 아닌 발화자가 철자 하나하나를 음소 단위로 변별하고 기억해야 정답 도출이 가능하다. 음운 지식을 직접적으로 평가하기 위한 것으로 소리 기반 언어 단위의 인식 능력을 측정한다.

또한 해당 그래프에서 알 수 있듯이 IELTS 문항 전반은 장문 담화의 논리적 흐름을 따라 특정 시점의 정보를 찾아내는 문제 풀이 방식이 반복되며 평가 범주가 다양하게 분산되기 보다 특정 하위 범주, 수사적 구성 지식 중심에 지나치게 편중되는 경향이 뚜렷하다. 이는 TOPIK과 유사하게 담화 구조 이해와 사실 확인 능력에 과도하게 집중되어 있다는 공통점이 발견됨을 보여준다.



〈그림 4〉 **IELTS 하위 평가 범주별 듣기 문항 비율**

4.4. TOPIK과 IELTS 듣기 평가 구인의 비교 분석

앞선 분석을 통해, TOPIK과 IELTS는 문항 구성 체제와 평가 방식에서 구조적으로 상이한 특징을 지니며, 이는 각 시험의 언어 능력 요소 반영 방식과 구인 타당도의 해석에 영향을 준다는 점을 확인할 수 있었다.

TOPIK은 하나의 짧은 실용 담화에 대해 1~2개의 문항이 출제되는 구조로, 단일 발화를 통해 핵심 정보뿐 아니라 암시적 의미나 맥락적 단서를 함께 고려해야 정답 도출이 가능하도록 설계되어 있다. 이러한 문항 특성은 문법 지식이나 텍스트 지식 외에도 기능적, 사회언어학적 요소까지 복합적으로 작동할 수 있는 여지를 가지지만, 실제 분석 결과에서는 구성 지식에 편중된 문항이 대부분을 차지하였으며, 실용적 언어 능력을 반영하는 기능 중심 과업

www.kci.go.kr

의 비율은 상대적으로 낮게 나타났다. 특히 장문 담화의 경우에도 발화의 흐름이나 맥락적 전개보다는 명시된 정보 확인에 초점을 맞춘 문항이 주를 이루며, 문항 유형의 다양화 및 기능적 요소 반영의 필요성이 제기된다. 더불어 사회언어학적 지식, 문화 간 의사소통 능력, 맥락에 따른 언어 조정 능력의 반영 또한 매우 제한적으로 나타나, 실제적 의사소통 능력 평가라는 목적에 부합하지 못하는 구조적 한계가 보인다.

반면, IELTS는 하나의 장문 담화에 대해 10개의 문항이 연속적으로 출제되는 구조로, 각 문항은 해당 담화 내 특정 구간에 정밀하게 대응되도록 설계된다. 수험자는 담화의 전개 흐름을 따라가며 시점별로 필요한 정보를 청취하고 이를 정확히 식별하여 정답을 도출해야 하며, 각 문항은 대부분 단일한 언어 능력 요소에 기반해 정답이 결정되도록 구성되어 있다. 이러한 체제는 구성 요소별 언어 능력 평가의 명확성을 높이는 한편, 기능 중심 평가 설계의 일관성을 유지함으로써 실제적 수행 능력 측정에 효과적인 구조로 볼 수 있다. 그러나 본 연구의 문항별 정밀 분석 결과에 따르면, 실제 IELTS 듣기 문항 역시 구성 지식 중심의 정보 이해 및 사실 확인에 과도하게 집중되어 있는 경향이 확인되었다. 이는 객관적인 채점 체계 유지와 대규모 시험 운영의 효율성 확보 등 제도적 현실을 고려한 결과로 해석될 수 있으며, 시험 설계 이론과 실제 문항 구현 간에는 간극이 존재함을 보여준다.

그럼에도 불구하고 IELTS는 CEFR 기준과 Bachman과 Palmer의 언어 능력 구성 요소 이론을 기반으로 고안된 국제 공인 시험으로, 다양한 언어 능력 요소가 균형 있게 반영되도록 문항이 설계되어 있으며, 이에 따라 기능성과 실제성을 중시하는 공신력 있는 평가 체제로 널리 인정받고 있다. TOPIK의 경우, 아직 CEFR 기반 외국어 교육의 설계 원칙이 충분히 반영되지 못한 국가 공인 시험으로서 발전 단계에 있으며, 특히 듣기 문항에서는 기능적 지식과 사회언어학적 지식의 반영이 매우 미흡한 반면, 구성 지식 중심의 문항이 과도하게 출제되고 있음이 분석을 통해 확인되었다. 이는 TOPIK이 학습자의 실제 의사소통 능력을 평가하기보다는, 정보 처리 및 정답 추출 중심의 평가에 초점을 맞추고 있음을 시사하며, 구성 타당도 측면에서 구조적 한계를 내포하고 있음을 보여준다.

본 연구는 TOPIK과 마찬가지로 IELTS 듣기 문항 역시 구성 지식 중심의 정보 이해 및 사실 확인에 과도하게 집중되어 있는 경향을 파악하였으나 그럼에도 불구하고 TOPIK이 국제적 기준에 부합하는 방향으로 체제적 전환을 모색할 수 있는 근거를 마련하고, 평가 목적과 언어 능력 구성 요소 간의 정합성을 확보해 나가는 개선 방향을 제시할 수 있다는 점에서 학술적, 실천적 의의를 가진다. 특히 구인 타당도 분석을 통해 도출된 기능 중심 문항 유형 설계는 TOPIK의 평가 체계가 국제적 정당성을 갖춘 방향으로 나아가는 데 시사점을 제공할 수 있을 것이다.

5. 결론 및 제언

본 연구에서는 글로컬 시대 국제 기준에 부합하는 한국어 평가 체제를 정립하기 위해 한국어능력시험(TOPIK)과 국제 공인 영어능력시험(IELTS)의 듣기 평가 시험의 구인타당도를 분석하였다. 구인 타당도 분석 결과, 두 시험은 문항 체제와 평가 방식에서 상이한 특성을 보였다. TOPIK은 하나의 짧은 담화에 대한 소수 문항 구성으로 정보 처리와 의미 추론을 함께 요구하지만, 구성 지식 중심의 문항이 주를 이루고 있다. 기능적 지식이나 사회언어학적 요소는 미흡하며, 특히 장문 문항에서도 정보 확인 중심의 접근이 주를 이루었다. 반면, IELTS는 장문 담화 내에서 다수의 문항이 흐름을 따라 배치되어 각 문항은 특정 언어 능력 요소를 중심으로 평가 일관성과 실제적 수행 능력 측정에 효과적인 평가 구조이다. 다만, IELTS도 실제 문항 구현에서는 구성 지식 중심의 정보 이해가 과도하게

www.kci.go.kr

반영되어, 이론적 설계와 시험의 제도적 현실 운영 간 간극이 존재함이 확인되었다.

이러한 분석을 토대로, TOPIK 듣기 평가의 구인타당도를 제고하고 학습자의 실제 듣기 능력을 정밀하게 평가하기 위한 개선 방안을 제시하면 다음과 같다. 첫째, 기능적 언어 능력 측정을 강화할 수 있는 문항 유형의 확대가 요구된다. 현재 듣기 평가 문항은 정보를 단순히 인식하고 확인하는 수준에 머무는 경우가 많아, 발화자의 의도나 청자 역할 수행, 사회적 상호작용 기능을 평가하는 데에는 부족하다. 따라서, 발화의 맥락을 해석하고 목적에 맞게 의미를 추론하거나, 다양한 의사소통 기능을 수행하는 문항 유형을 개발하여 기능적 지식의 반영 비율을 확대할 필요가 있다.

둘째, 사회언어학적 지식이 반영된 문항 구성의 확대가 필요하다. 한국어는 사회적 맥락에 따라 높임법, 어휘 선택, 표현 방식이 달라지는 언어적 특성을 가지므로, 사회적 관계나 담화 목적에 따른 언어 사용의 적절성을 판단하는 문항이 포함되어야 한다. 이를 통해 언어의 사회적 맥락 내 활용 능력을 타당하게 측정할 수 있으며, 실제적인 언어 능력 평가로서의 공신력을 제고할 수 있다.

셋째, 고급 문항의 비중 확대와 평가 영역의 세분화가 요구된다. 현재 상위 등급 문항의 수가 제한적이어서 학문적, 직업적 환경에서의 고급 언어 수행 능력을 충분히 측정하기 어려우며, 이에 따라 평가 결과가 학습자의 실제 수준을 정밀하게 반영하지 못하는 한계가 발생한다. 고급 문항 내에서도 논리적 추론, 비판적 이해, 상위 개념 도출 등 고차 사고 기반 과업을 다양하게 포함시켜야 하며, 이를 세부 능력 단위로 나누어 구조화된 평가 체계를 마련할 필요가 있다.

끝으로 평가의 실제성과 경제성을 높이기 위해 자연스러운 구어체 표현, 특히 축약 표현을 반영한 문항 구성의 확대가 필요하다. 일상적인 한국어 담화에서는 의미를 해치지 않는 범위에서 축약된 표현이 자주 사용되며, 이를 반영한 문항은 학습자의 실제 청해 능력을 보다 효과적으로 측정할 수 있다.

본 연구는 문항 내 포함된 구성 요소를 다층적으로 식별하되, 각 요소의 기여도를 균등하게 반영하는 다중 범주 분석 방식을 채택하였다. 이는 실제 청해 상황에서 요구되는 복합적 언어 능력을 포괄적으로 반영하기 위해 선택된 방식이나, 정답 도출에 실질적으로 기여하는 중심 요소의 중요도가 상대적으로 희석되었다는 점에서 핵심 범주와 보조 범주 간 가중치 조정을 포함한 혼합 분석 방식의 필요성이 제기될 수 있다. 연구 자료의 한계도 있다. 특히 IELTS는 공식 기출 문제의 공개가 제한되어 있어, 본 연구는 Cambridge에서 제공하는 모의 평가 자료를 바탕으로 분석을 진행하였다. 해당 자료는 실제 시험과 유사성을 가지나 완전한 대표성을 확보했다고 보기에는 제한적이다.

이러한 한계점에도 불구하고, 본 연구는 TOPIK의 듣기 평가를 공신력 있는 국제 시험과 비교함으로써 국제적 기준에 부합하는 방향으로 평가 체제를 전환하기 위한 구체적 근거를 제시하고, 구인 타당도의 개선 방안을 마련하고자 한다는 점에서 실천적인 가치가 있다. 향후 연구에서는 보다 다양한 평가 체제를 포함한 비교 분석과 함께, 가중치 반영 방식의 정교화, 분석 기준의 객관화, 자료의 대표성 확보 등을 통해 연구의 한계를 보완해 나갈 필요가 있다.

www.kci.go.kr

참고문헌

1. 단행본 및 논문

강승혜 · 강명순 · 이영식 · 이원경 · 장은아, 『한국어 평가론』, 태학사, 2006.

강현화 · 김선정 · 김은애 · 김정숙 · 이미혜 · 최은규, 『외국어로서의 한국어 교육학』, 한국방송통신대학교 출판부, 2011.

송영승, 『한국어능력시험의 초급 듣기영역 문항 분석 연구』, 영남대학교 대학원 석사학위논문, 2013.

유소정, 『외국어능력시험과의 비교를 통한 한국어능력시험 듣기 문항 분석 연구』, 경희대학교 교육대학원 석사학위논문, 2012.

조성해, 「유학생의 한국어 듣기 평가를 위한 한국어능력시험의 내용 타당도 분석 연구」, 『외국어로서의 한국어교육』 64, 연세대학교 언어교육연구원, 2022.

Bachman, L. F., *Fundamental Considerations in Language Testing*, Oxford: Oxford University Press, 1990.

Bachman, L. F. & A. S. Palmer, *Language Testing in Practice*, Oxford: Oxford University Press, 1996.

Canale, M., *From Communicative Competence to Communicative Language Pedagogy* 1, London, Routledge, 1983.

Canale, M. & M. Swain, "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing", *Applied Linguistics* 1, 1980.

Cambridge University Press, *Cambridge IELTS 17 Academic: Student's Book with Answers*, Cambridge: Cambridge University Press, 2022.

_____, *Cambridge IELTS 18 Academic: Student's Book with Answers*, Cambridge: Cambridge University Press, 2022.

_____, *Cambridge IELTS 19 Academic: Student's Book with Answers*, Cambridge: Cambridge University Press, 2022.

Hymes, D. *On Communicative Competence*, In J. Price & J. Holmes (Eds.), Sociolinguistics, Harmonsworth, Penguin, 1972.

2. 기타 자료

Council of Europe(2020), Common European framework of reference for languages: Companion volume with new descriptors, Retrieved May 29, 2025. https://www.coe.int/en/web/common-european-framework-reference-languages/cefr-companion-volume-and-its-language-versions

IDP IELTS(2024), Who takes IELTS and why, Retrieved May 27, 2025. https://ielts.idp.com/about/news-and-articles/article-who-takes-ielts

IELTS(2024), IELTS for immigration, Retrieved May 27, 2025. https://ielts.org/organisations/ielts-for-organisations/ielts-for-your-sector/ielts-for-immigration

TOPIK 한국어능력시험. https://www.topik.go.kr.

Abstract

Construct Validity of the TOPIK Listening Test in the Glocal Era

A Comparative Analysis with IELTS

Kim, Mi-Kyung | Chungbuk National University

Kwon, Tae-Hyun | Chungbuk National University

This study conducted a construct validity analysis of the Test of Proficiency in Korean (TOPIK) and the International English Language Testing System (IELTS) listening assessments to establish a Korean language testing framework aligned with international standards. The analysis revealed that TOPIK employs short passages with a limited number of items requiring both information retrieval and meaning inference. However, most items primarily assessed compositional knowledge, with insufficient incorporation of functional and sociolinguistic elements. By contrast, IELTS structures multiple items along the flow of extended discourse, enabling more effective measurement of authentic listening performance, yet it was also found to overemphasize compositional knowledge. Both tests exhibited gaps between their theoretical design and operational practice. By conducting a multi-layered analysis of test items, this study evaluated the validity of Korean listening assessment in comparison with an international benchmark. The findings provide a foundation for improving TOPIK and enhancing its alignment with global testing standards.

Keywords TOPIK, IELTS, listening assessment, construct validity, Multiple Correspondence Analysis

이 논문은 2025년 9월 20일에 투고 완료되어
2025년 9월 30일부터 10월 15일까지 심사위원이 심사하고
2025년 10월 20일에 심사위원 및 편집위원 회의에서 게재가 결정된 논문임

www.kci.go.kr