

한국어능력시험(TOPIK) 읽기 텍스트의 난이도 연구 - TOPIKⅡ의 읽기 영역을 중심으로-

신 정 하

Abstract

Shin Jungha. 2016.03.31. Text difficulty of readings on the TOPIK: focusing on reading texts in TOPIKⅡ. *Bilingual Research* 62. 29-48. The goal of this paper is to measure the difficulty of reading texts on the TOPIK (Test Of Proficiency In Korean) in order to examine the distribution of text difficulty among the TOPIK's various test items. After the 35th TOPIK was administered, the test underwent reforms. It is necessary to examine text difficulty between test items on the reformed test. First, the difficulty texts appearing on the TOPIK are compared to Korean textbooks. The comparison targets consist of three Korean textbooks from level 3 to level 6. Text difficulty is measured by looking at the level of vocabulary, sentence length, and the difficulty of verbal endings. The programming language R was used for text analysis. The results show that the vocabulary found on the TOPIK is at a higher level of difficulty than that found in Korean textbooks. However, the TOPIK's sentence length and verbal endings difficulty are at a similar level when compared to the Korean textbooks. Second, four tests made available on the TOPIK website (the 35th, 36th, 37th, and 41st) are analyzed. This was done to determine the reliability of the test. The difficulty of text items from one test to another is very important affects the difficulty of the overall test itself. The results show no major difference between the four available tests. Therefore, the reformed TOPIK be considered to have a high level of reliability. (Korea University)

[Key words] 텍스트 난이도(text difficulty), 읽기 텍스트(reading text), 평가 신뢰도(test reliability), 읽기 평가(reading evaluation)

1. 서론

본 연구의 목적은 개편 후 한국어능력시험(Test Of Proficiency In Korean, TOPIK)의 읽기 텍스트 난이도를 자동적 방법으로 측정해 회차 내 난이도 분포와 회차 간 난이도 분포 양상을 확인하는 것이다.

TOPIK의 읽기 영역에 대한 연구는 주로 문항 유형, 평가 틀에 집중되어 있었으며(이상린, 2012; 董洋·이혁화, 2013) 읽기 지문 자체에 관한 연구는 매우 드물었다. 텍스트 난이도 관련 선행 연구 중 TOPIK의 읽기 텍스트 난이도 측정을 시도한 연구는 이지혜(2009), 김장식(2011), 장미경(2012), 구민지(2013), 구민지·양길석(2013) 등이 있었으나 이때 TOPIK의 텍스트는 난이도 측정법, 혹은 측정 결과의 타당성을 검증하는 비교 대상으로써 사용되었다. 이 연구들의 목적은 TOPIK 자체의 난이도를 확인하고 검증하는 것이 아니라 초·중·고급으로 나누어져 있었던 TOPIK을 기준으로 다른 측정 결과를 검증하는 것이었다. 그러나 현재 가장 공신력 있는 한국어평가로서, 한국어능력시험의 난이도의 검토는 평가 자체의 신뢰도를 위해 매우 필수적이라고 할 수 있다.

읽기 평가에서는 텍스트 자체의 난이도 외에도 주제와 문항 유형 등이 문항의 난이도에 영향을 미칠 수 있을 것이다. 그러나 한국어능력시험의 문항 유형과 주제는 회마다 크게 달라지지 않기 때문에 본고에서는 텍스트 자체의 난이도에 초점을 둔다. 그중에서도 자동적·기계적 측정이 가능한 문장 단위 이하의 난이도 요인을 중심으로 난이도를 측정할 것이다. 이러한 요인들은 직관적으로 난이도를 가늠하기 어려우며 계량적인 측정을 통해서만 그 정도를 파악할 수 있다.

본 연구의 측정 대상은 제35회 이후 개편된 TOPIKⅡ의 읽기 영역이다. 기존 초·중·고급으로 나누어 응시할 수 있었던 한국어능력시험은 개편 이후 초급이 TOPIK I으로 변경되었고 중급과 고급이 TOPIKⅡ로 통합되었다. 따라서 TOPIKⅡ는 중급과 고급의 수험자들을 모두 변별할 수

있어야 할 것이다. 비교 대상은 중급과 고급 수준의 한국어 통합 교재 읽기 텍스트의 난이도를 측정 한 자료이다. 텍스트 분석을 위해 TOPIK II의 텍스트로 코퍼스를 구축하였으며 참조 코퍼스로 한국어 교재 코퍼스(Korean Textbook Corpus, 이하 KTC)를 설정하였다. 현재 대부분의 한국어 교육 기관 및 교재, 그리고 한국어능력시험은 6등급 체제를 이루고 있는데, 분석 결과를 통해 한국어능력을 등분하는 기준의 차이도 간접적으로 확인할 수 있을 것이다.

또한 평가 자체 신뢰도 확인을 위해 총 4회의 텍스트들이 일관적인 난이도 수준을 유지하고 있는지도 검증하도록 하겠다. 4회의 시험 난이도를 요인별로 측정하여 그 양상이 일정하게 나타나는지를 비교할 것이다. 이상 본고의 연구 문제를 정리하면 다음과 같다.

첫째, TOPIK II의 읽기 텍스트 난이도는 중급과 고급의 수험자를 모두 변별할 수 있는가?

둘째, TOPIK II의 읽기 텍스트 난이도는 일관성을 보이고 있는가?

2. 텍스트 난이도 요인 및 측정 방법

한국어교육용 텍스트의 난이도 측정 요인 및 방법은 선행 연구마다 조금씩 다르게 나타나고 있다. 본 연구의 측정 대상인 문장 단위 이하의 난이도 요인을 중심으로 선행 연구의 기준을 정리하면 <표 1>과 같다.

<표 1> 선행 연구의 난이도 요인 및 측정 방법

선행 연구	난이도 요인	측정 방법
이지혜 (2009)	어휘	어휘 목록 활용, 어려운 단어의 수 측정
	문장 수	전체 문장의 수 측정

김장식 (2011)	어휘	어휘 목록 활용, 어려운 단어의 수 측정
	문장 길이	문장별 평균 어절 수 측정
	문장 수	전체 문장의 수 측정
장미경 (2012) ¹⁾	어휘	어휘 목록 활용, 수준별 어휘의 분포 측정
	문장구조	문장 구조의 복잡성
	연결어미	어미 목록 활용, 수준별 어휘의 분포 측정
홍정하 외 (2011)	어휘	어휘의 평균 음절 수 측정
	문장 길이	문장별 평균 음절 수 측정
	문단길이	문단별 평균 음절 수 측정
구민지 (2013)	어휘	어휘 목록 활용, 수준별 어휘의 분포 측정
	문장 길이	문장별 평균 어절 수 측정
	문법	문형 목록 활용, 수준별 어휘의 분포 측정
	꾸밈표현	꾸밈표현 목록 활용

<표 1>에서 많은 연구들이 공통적으로 꼽고 있는 난이도 요인은 어휘, 문장 길이, 문법 요인으로 정리할 수 있겠다. 이 연구에서는 선행 연구의 난이도 측정 요인 및 방법을 종합적으로 검토하고 세부적인 난이도 측정 방법을 조정해 이를 자동적 방법²⁾으로 실현한 신정하(2016)의 기준과

- 1) 장미경(2012)에서는 <표 1>의 요인 외에도 주제 전개 구조, 텍스트 유형 등 담화 층위에서의 난이도 측정 평가 방법 또한 다루었는데, 본 연구에서는 직관적으로 평가하기 어려운 미시구조에 초점을 두었기 때문에 이를 별도로 제시하지 않았다. 텍스트 유형과 주제 전개 구조의 난이도를 평가하기 위해서는 질적인 접근이 필요할 것이다.
- 2) 컴퓨터를 활용한 자동적 난이도 측정은 다음과 같은 장점을 지닌다. 첫째, 일관적인 작업이 가능하다. 이는 컴퓨터를 활용한 언어 연구의 가장 큰 장점으로, 빈도를 세거나 연산 작업 시 수작업보다 정확하고 일관적인 작업이 가능하다(강범모, 2014). 두 번째 장점은 재현가능성(replicability)이다. 이는 첫 번째 장점과도 이어지는 것으로, 다른 연구자가 같은 도구를 사용한다면 같은 기준으로 쉽게 텍스트를 분석하거나 기존의 결과를 검증할 수 있는 것이다. 텍스트 처리에는 프로그래밍 언어 R(ver. 3.2.3)을 활용하였으며 측정에 사용된 스크립트는 신정하(2016)의 부록을 참고할 수 있다.

방법을 따르도록 할 것이다.

어휘, 문장 길이, 문법 요인 외에 구민지(2013)에서는 텍스트 난이도 요인으로 꾸밈 표현을 설정하였다. 꾸밈표현은 관용 표현, 수사적 표현, 간접 표현, 문화적 함축을 담은 표현 등을 가리키며 별도의 목록을 구성해 어휘처럼 수준별 분포를 통해 난이도를 측정하였다. 장미경(2012)에서도 어휘 난이도 측정 시 문화적 함축이 강한 관용어의 경우 맥락을 보고 그 등급을 조정하였다. 관용 표현 및 간접적인 표현이 한국어 텍스트의 이해를 방해한다는 점은 동의하지만 이러한 요인들은 독자 요인과 깊은 상관이 있으며 질적인 접근이 필요하기 때문에 측정 대상에서 제외하였다.³⁾

어휘 난이도 측정은 대부분의 선행 연구와 마찬가지로 수준별 어휘 목록을 활용하였다. 한 텍스트 내 수준별 어휘가 분포하고 있는 비율을 측정하였으며 수준별로 1점씩 가중치를 주어 점수를 측정하였다. 목록에 없는 어휘는 최고급으로 산정하였다.⁴⁾ 어휘 목록은 기존 어휘 목록 연구를 귀납적으로 정리하고 전문가들의 조정을 거친 강현화(2012, 2013, 2014a)를 활용하였다. 등장횟수를 세는 기준은 토큰(token)으로, 등장하는 모든 어형을 세었다.

둘째, 문장 길이는 한 문장의 평균 음절 수를 측정한다. 문장 길이는 음절 수 외에도 어절 수, 형태소 수 등 여러 가지를 기준으로 하여 측정이 가능하다. 신정하(2016)에서 형태 분석 말뭉치로 다중회귀분석을 실시한 결과, 문장 길이를 측정하는 여러 기준 중 한 문장의 평균 음절 수가 난이도에 통계적으로 유의한 영향을 미치는 것으로 나타났기 때문에

3) 미국 공통 학습 기준(CCSS, 2010)에서는 ‘언어의 축어적 의미와 비유적 의미’, ‘언어의 명료성과 모호성’, ‘언어의 친숙성과 고어체 정도’를 질적 요인으로 간주하고 있다. 또한 구민지·양길석(2013)에서도 꾸밈표현은 텍스트 수준에 가장 낮은 영향력을 미치는 요인으로 밝혀졌다.

4) 초급은 1점, 중급은 2점, 고급은 3점, 최고급은 4점으로 상정하였다.

본고에서도 이를 기준으로 문장 길이를 측정하도록 한다.

셋째, 문법 요인은 구문 분석 말뭉치가 아닌 형태 분석 말뭉치를 활용하는 한계로 문장에 나타난 어미의 난이도 분포를 통해 확인하도록 하겠다. 어미 목록은 김중섭(2010)과 학습자의 습득 양상 연구를 기반으로 김중섭(2010)의 연결어미 목록을 조정한 장미경(2012)의 목록을 활용한다. 지금까지 한국어의 통사적 난이도를 측정 한 시도들은 모두 내포문이 많을수록 어려운 문장으로 판단하였기 때문에(이혜진, 2007; 김의수·정은주, 2010; 정은주, 2010; 장미경, 2012), 어미의 난이도를 살펴보는 것이 이러한 관점을 포괄할 수 있으리라 본다. 어미는 연결어미, 전성어미, 종결어미로 한정하여 측정한다. 선어말어미는 문장을 구조적으로 복잡하게 만들지 않기 때문에 제외하였다. 어미 난이도 역시 어휘 난이도처럼 수준별로 1점씩 가중치를 두어 점수를 측정할 것이다. 한 문장 당 평균 어미 난이도 점수를 구하는 것이므로 가중치를 곱한 후 전체 문장 수로 나누게 된다. 본 연구의 난이도 측정 방법을 정리하면 <표 2>와 같다.

<표 2> 신정하(2016)의 텍스트 난이도 측정 방법

요인	측정 방법
어휘	강현화(2012, 2013, 2014a)의 목록을 활용하여 한 텍스트의 초·중·고급 어휘의 분포를 토른 기준으로 측정. 수준별로 어휘에 가중치 부여. 점수 합계 = $\frac{\text{초급어휘수}}{\text{전체어휘수}} \times 1 + \frac{\text{중급어휘수}}{\text{전체어휘수}} \times 2 + \frac{\text{고급어휘수}}{\text{전체어휘수}} \times 3 + \frac{\text{최고급어휘수}}{\text{전체어휘수}} \times 4$
문장길이	문장별 평균 음절 수 측정.

어미	<p>장미경(2012)의 연결어미, 김중섭(2010)의 종결어미와 전성어미 목록을 활용하여 한 문장의 초·중·고급 어미 분포를 측정. 수준 별로 어미에 가중치 부여.</p> <p>점수 합계 = $\frac{\text{초급어미수} \times 1 + \text{중급어미수} \times 2 + \text{고급어미수} \times 3 + \text{최고급어미수} \times 4}{\text{문장수}}$</p>
----	--

3. 연구 대상 및 방법

본 연구의 분석 대상은 한국어능력시험 공식 홈페이지(www.topik.go.kr)에 공개되어 있는 TOPIKⅡ의 읽기 영역 텍스트이다. 현재 문제 은행 구축을 위해 기출문제가 선택적으로 공개되어 있기 때문에, 공개된 모든 시험(제35, 36, 37, 41회, 총 4회)을 본고의 분석 대상으로 삼고 시험의 텍스트로 형태 분석 코퍼스를 구축하였다⁵⁾.

참조 코퍼사인 KTC는 여러 대학 기관에서 제작한 일반 목적 통합 교재를 형태 분석 코퍼스로 구성한 것이다. 본고에서 활용한 KTC는 신정하(2016)에서 사용된 수정판으로, 개편되어 사용되지 않은 교재들을 제외한 후 읽기 활동의 텍스트만을 가려낸 코퍼스이다⁶⁾.

그중에서도 텍스트 유형 요인을 통제하기 위하여 제보적 기능만을 추출하여 구성한 텍스트를 사용하였다. 제보적 기능이란 독자에게 텍스트 생산자가 갖고 있는 지식, 정보, 의견을 전달하는 기능이다(한국텍스트 언어학회, 2004). 그런데 신정하(2016)과 본 연구에서는 텍스트의 의사소

5) 형태 분석에는 울산대학교 한국어처리연구실에서 개발한 Utagger를 사용하였다. 분석 표지는 21세기 세종계획의 형태 분석 표지를 따르고 있다.

6) KTC는 경희대, 고려대, 서강대, 연세대, 이화여대, 서울대 총 다섯 종의 일반 목적 한국어 교재로 구성되어 있다. 그런데 경희대, 이화여대, 서울대의 경우 개편 전의 교재였기 때문에 분석대상에서 제외하였다.

통 목적을 함께 고려하여 분석 대상 텍스트의 기능을 판정하였다. 예를 들어 텍스트가 접속 기능의 소개글이라 하더라도 이어진 과제가 ‘일치하는 내용 고르기’라면 이를 제보적 기능의 텍스트로 분류하였다. 따라서 본 연구에서는 TOPIKⅡ의 11~24번, 28~50번 문항의 텍스트가 분석 텍스트가 된다.⁷⁾

사용된 통계 기법은 두 가지로, 교재와 한국어능력시험의 난이도 점수의 비교에는 단일 표본 t 검정을 실시하였다. t 검정은 두 집단 사이의 평균을 비교할 수 있는 통계 기법이다. 회차 간 난이도 비교에는 분산 분석을 실시하였다. 분산 분석이란 세 개 이상 집단 간의 평균을 비교할 수 있는 통계 기법이다. 이때의 통계 분석 역시 텍스트 난이도의 요인별로 각각 실시할 것이다. 통계에는 프로그래밍 언어 R(ver. 3.2.3)을 활용하였다.

4. 난이도 측정 결과

4.1 교재 텍스트 난이도 비교

4.1.1 어휘 난이도

TOPIKⅡ 텍스트와 KTC의 어휘 난이도 측정 결과를 요약하면 <표 3>과 같다. TOPIKⅡ의 어휘 난이도 평균은 2.205점이었으며 누적 1분위 수부터 3분위 수, 즉 난이도 점수가 가장 밀집된 범위는 2.000~2.436점이었다.

7) 제35회의 14번 문항, 제36회의 13번 문항, 41회의 14번 문항의 텍스트 기능이 본 연구의 대상에 부합하지 않아 분석 대상에서 제외되었다. 따라서 총 113개의 텍스트가 분석 대상이 된다.

<표 3> 어휘 난이도 측정 값 요약 및 비교

	최솟값	1분위수	중앙값	평균	3분위수	최댓값
TOPIK II	1.432	2.000	2.217	2.205	2.436	2.821
KTC	1.125	1.464	1.709	1.681	1.855	2.291

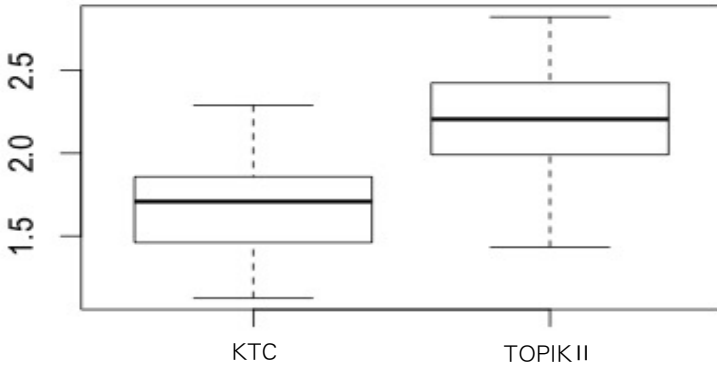
TOPIK II 텍스트의 어휘 난이도는 KTC보다 높게 나타나고 있었다. 이런 차이는 두 집단의 난이도 점수 분포를 상자그림⁸⁾으로 나타낸 <그림 1>을 보면 더욱 뚜렷하다. KTC에서 높은 어휘 난이도 수치를 보인 텍스트는 TOPIK II에서는 일반적인 수준에 불과하다. 즉, TOPIK II 텍스트의 일반적인 어휘 수준은 교재와 비교했을 때 상당히 높은 편으로 드러났다. 이러한 차이는 한국어 교재에 사용된 어휘를 양적으로 분석한 강현화(2014b)에서도 드러나 있다. 강현화(2014b)에서는 한국어 교재와 한국어능력시험에서 중복되어 사용된 어휘의 수를 분석하였는데, 그 결과 한국어 교재보다 한국어능력시험이 더 다양한 어휘를 사용하고 있는 것으로 나타났다.

이러한 양상은 하나의 평가 도구로 학문 목적과 일반 목적 수험자를 모두 포괄해야 하기 때문에 나타난 것으로 보인다. 어휘의 난이도는 텍스트의 주제에 많은 영향을 받는다. 그런데 학문 목적 읽기 능력을 평가하기 위한 텍스트는 일상적인 주제보다 전문적인 주제가 적합하기 때문에 문항의 어휘 수준이 올라가는 것이다. 일반 목적의 한국어 교재에서도 고급 수준에서는 일상적인 주제 이상을 다루지만, 여전히 의사소통적 읽

8) 상자그림이란 최댓값, 최솟값, 중앙값, 사분편차를 사용해 자료들의 분포를 한눈에 보여주는 그래프이다. 상자 가운데의 가장 짙은 선은 중앙값을 의미하고, 상자의 위쪽과 아래쪽의 가장자리는 각각 3분위 수와 1분위 수를 나타낸다. 3분위 수와 1분위 수는 각각 측정치 분포에서 누적된 75%, 25%에 해당하는 값을 의미하므로 이 사이의 값에는 측정치들이 가장 몰려서 분포하고 있다. 상자에 연결된 선의 양끝은 최댓값과 최솟값을, 상자 위아래의 작은 동그라미는 전체 데이터의 1% 미만에 속하는 특잇값(outlier)을 의미한다.

기에 초점을 두고 있기 때문에 그 어휘 수준이 높지 않다.

그런데 <그림 1>에서 측정치 분포 양상을 보여주는 두 집단의 상자그림을 보면 분포의 양상과 폭은 상당히 유사하게 나타나고 있다. 자료의 값이 평균으로부터 얼마나 가까이 분포하는지를 보여주는 표준 편차 값도 KTC가 0.252, TOPIK II가 0.3으로 큰 차이를 보이지 않았다.



<그림 1> 어휘 난이도 측정 결과 비교

유의수준 0.01에서 두 집단에 t검정을 실시한 결과가 <표 4>로, p값은 $2.2e-16^9$ 보다 작은 것으로 나타났다. <그림 1>의 그래프와 마찬가지로 두 집단의 어휘 난이도 점수 차이는 통계적으로 유의하다.

9) $2.2e-16$ 의 e는 부동소수점으로, 선행하는 숫자 밑이 10인 지수를 곱한 수치를 나타낸다. e+는 소수점을 오른쪽으로, e-는 소수점을 왼쪽으로 이동시킨다. 즉, $2.2e-16$ 은 2.2×10^{-16} 라는 값을 의미한다.

<표 4> 어휘 난이도 t 검정 결과

t	-15.163
df	219.97
p 값	< 2.2e-16 ***

*p < 0.05 **p < 0.01 ***p < 0.001

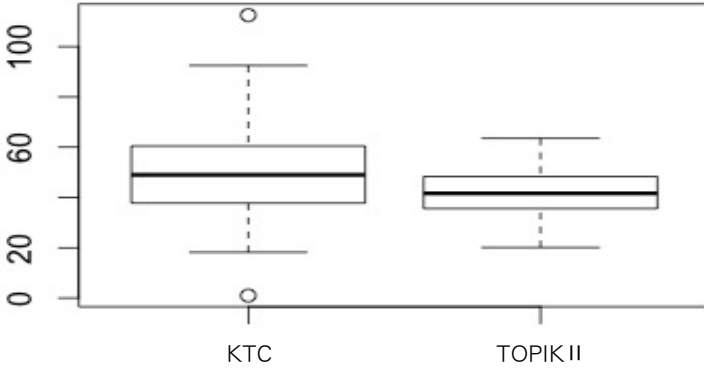
4.1.2 문장 길이

TOPIKⅡ 텍스트의 문장 길이 측정 결과 및 비교 집단인 교재 텍스트의 측정 결과를 요약한 것이 <표 5>이다. TOPIKⅡ의 문장 길이 평균은 42.47점이었으며 문장 길이 점수가 가장 밀집된 범위는 35.75~48.40점이었다.

<표 5> 문장 길이 측정 값 요약 및 비교

	최솟값	1분위수	중앙값	평균	3분위수	최댓값
TOPIKⅡ	20.13	35.75	41.80	42.47	48.40	63.50
KTC	20.00	37.78	49.00	50.18	60.46	97.50

<표 5>에서 두 집단의 중앙값 및 평균은 서로 비슷한 모습을 보이고 있음을 확인할 수 있다. 누적 1분위 수와 3분위 수는 KTC의 측정치가 T OPIKⅡ보다 높게 나타나고 있었다. 특히 최댓값을 보면 교재 텍스트에서 가장 긴 문장은 TOPIKⅡ의 평균 문장 길이의 2배 이상이었다. 측정치를 상자그림으로 나타내면 <그림 2>와 같다.



<그림 2> 문장 길이 측정 결과 비교

두 집단 사이 단일 표본 t 검정을 실시한 결과는 <표 6>과 같다. p 값은 2.466e-07로, 두 집단의 차이는 통계적으로 유의하다고 해석할 수 있겠다. 그렇지만 <그림 2>를 보면 KTC는 TOPIK II보다 더 넓은 범위를 가진 것으로 나타나고 있다. 두 집단의 통계적 차이는 어느 한 집단보다 다른 집단이 아주 높거나 아주 낮은 점수를 보이기 때문이 아니다. KTC가 TOPIK II보다 더 넓은 범위를 포함하고 있어 그 분포의 범위가 다르게 나타났기 때문에 통계적으로 차이를 보인 것이다.

<표 6> 문장 길이 t 검정 결과

t	5.3065
df	249.65
p 값	2.466e-07 ***

*p < 0.05 **p < 0.01 ***p < 0.001

두 집단의 문장 길이 분포 범위 차이를 통해 TOPIK II의 문장 길이는 중급 수준의 교재 텍스트와 유사한 양상을 보인다는 것을 추측할 수 있다. 교재 텍스트의 문장 길이, 문단 길이를 측정한 홍정하 외(2011)에서

문장 길이는 중급 수준에서만 유의한 요인이었다. 신정하(2016)에서도 중급에서는 문장 길이를 짧게 통제하였지만 고급 수준에 이르면 문장 길이는 매우 다양하게 나타나 난이도 변별에 큰 영향이 없다는 점을 밝혀졌다. 다시 말하면 문장 길이는 중급 학습자에게는 여전히 유의미한 난이도 요인이며, 이는 중급 학습자를 변별할 때 문장 길이를 활용할 수 있다는 점을 시사한다.

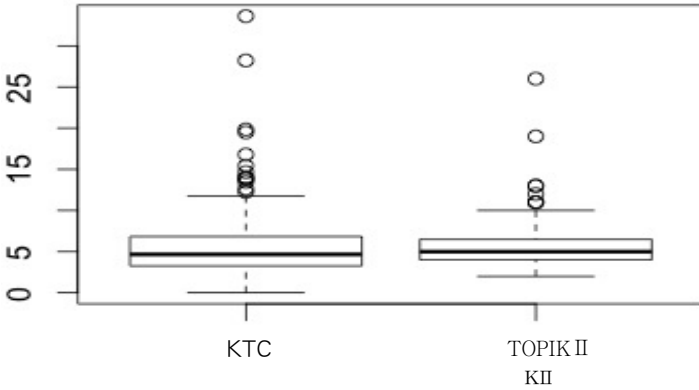
4.1.3 어미 난이도

TOPIK II 텍스트의 어미 난이도 측정 결과 및 비교 집단인 교재 텍스트의 측정 결과를 요약하면 <표 7>과 같다. TOPIK II의 어미 난이도 평균은 4.873점이었으며 측정치들이 가장 몰려있는 범위는 4.00~5.714점이었다.

<표 7> 어미 난이도 측정 값 요약 및 비교

	최솟값	1분위수	중앙값	평균	3분위수	최댓값
TOPIK II	2.250	4.000	4.600	4.873	5.714	10.380
KTC	0	3.261	4.667	5.719	6.809	33.6

어휘 난이도 및 문장 길이와 달리 어미 난이도는 KTC와 유사한 모습을 보이고 있었다. 두 집단의 측정치를 상자그림으로 나타낸 <그림 3>을 보면, 두 집단의 상자는 거의 비슷한 모습이지만 KTC의 값들이 조금 더 넓게 분포하고 있으며 많은 특잇값을 가지고 있다.



<그림 3> 어미 난이도 측정 결과 비교

두 집단이 통계적으로도 유사한 모습을 보이는지 확인하기 위해 유의수준 0.01에서 단일 표본 t 검정을 실시하였다. 분석 결과를 정리하면 <표 8>과 같다. p 값은 0.7954로 유의수준인 0.01보다 크게 나타났다. 두 집단의 차이는 통계적으로 유의미하지 않았다. 따라서 TOPIK II와 한국어 교재의 어미 난이도 수준은 유사한 양상을 보인다고 해석할 수 있다.

<표 8> 어미 난이도 t 검정 결과

t	-0.25957
df	283.79
p-value	0.7954

교재와 TOPIK II가 유사한 어미 난이도 양상을 보인 이유는 고급으로 올라갈수록 문법에 대한 학습량과 그 범위가 제한적이기 때문이다. 어휘의 경우 학습자의 관심사와 목적에 따라 고급 수준에서 확장하고자 하면 별도의 자료를 통해 무한으로 확장할 수도 있겠으나, 어미(문법)의 경우는 그렇지 않다는 것이다.

4.2 회차간 난이도 비교

이번 절에서는 총 4회의 텍스트들이 서로 유사한 수준을 보이는지를 검토하도록 하겠다. 일정한 수준의 난이도를 유지하는 것은 평가의 신뢰도 및 액면 타당도와 관련되어 매우 중요한 문제이다.

4회의 평가의 어휘 난이도 점수에 대하여 분산 분석을 실시한 결과는 <표 9>와 같다. 이때 유의수준은 0.01로 설정하였다.

<표 9> 분산 분석 결과

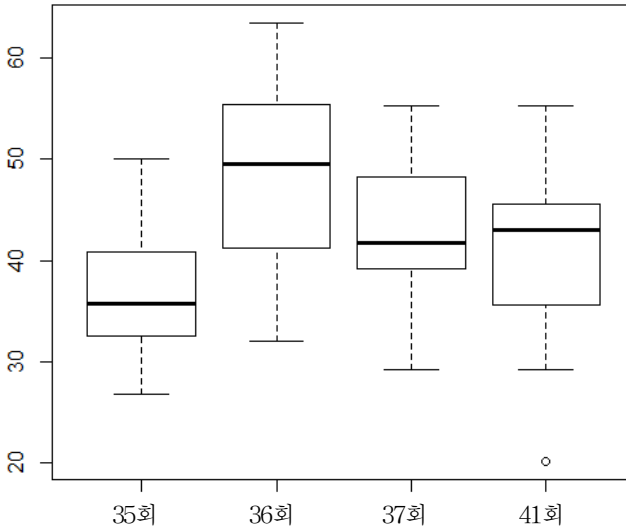
어휘 난이도	F	3.5907
	num df	3
	denom df	112
	p-value	0.01596
문장 길이	F	13.219
	num df	3
	denom df	112
	p-value	1.891e-07 ***
어미 난이도	F	3.3137
	num df	3
	denom df	58.694
	p-value	0.02602

p < 0.01 *p < 0.001

분석 결과, 통계적으로 유의한 차이를 보인 것은 문장 길이 요인뿐이었으며 회차 간 어휘 난이도와 어미 난이도 점수는 유사한 모습을 보이고 있다는 것이 통계적으로 검증되었다.

회차별로 고른 점수를 보이지 않은 문장 길이 요인을 별도로 살펴보기 위해 점수 분포를 <그림 4>와 같이 상자그림으로 나타내었다. <그림 4>를 확인한 결과, 개편 후 처음 실시되었던 제35회와 36회가 결과에 가

장 큰 영향을 미친 것으로 나타났다.



<그림 4> 4회 시험의 문장 길이 측정 결과 비교

제37회와 제41회의 상자그림은 앞의 두 시험(제35회, 제36회)보다 비슷한 양상을 보이는데, 이 두 집단의 값이 통계적으로도 유사한지를 살펴보기 위해 t 검정을 실시하였다. 분석 결과는 <표 10>과 같다. p 값은 0.5493으로 유의수준인 0.01보다 크게 나타났으므로 제37회와 제41회의 읽기 지문은 같은 수준을 유지하고 있다고 해석할 수 있다.

<표 10> 제 37회, 41회 문장 길이 t 검정 결과

t	0.60248
df	54.715
p-value	0.5493

결국 문장 길이가 일정하지 않았다는 <표 9>의 결과는 개편 직후의 두 시험이 원인으로 보인다. 한국어능력시험은 더 이상 기출문제를 매회 공개하지 않기 때문에 확인이 쉽지는 않겠지만 회를 거듭할수록 유사한 모습을 보이며 안정적으로 난이도가 유지될 것으로 예상된다.

5. 결론

본 연구는 지금까지 공개된 4회의 TOPIK II의 텍스트를 바탕으로 개편된 한국어능력시험이 중급과 고급의 학습자를 모두 아우르며, 회차 간 동일한 텍스트 난이도 수준을 유지하고 있는지를 검토하고자 하였다.

이를 위해 두 가지의 연구문제가 제기되었는데 첫 번째 연구 문제는 중급과 고급을 통합한 TOPIK II의 변별력에 관한 것이었다. 비교 집단으로 한국어 교재의 읽기 텍스트를 설정하였고, 두 집단의 측정치를 요인 별로 비교한 결과는 다음과 같다.

첫째, 텍스트 난이도 요인 중 TOPIK II와 교재 텍스트가 가장 유사한 모습을 보인 것은 어휘 난이도였다. 이는 어휘와 달리 문법의 학습량과 범위가 고급으로 갈수록 크게 확장되지 않기 때문으로 보인다. 둘째, 어휘 난이도와 문장 길이는 교재의 텍스트와 다른 모습을 보였다. 먼저 어휘 난이도는 교재 텍스트의 어휘보다 더 어려운 수준으로 나타났다. TOPIK에서 중급 수준에 기대하는 어휘 수준은 교재의 고급 텍스트의 어휘 수준이었다. 문장 길이는 그 분포가 교재의 분포보다 더 좁았다. TOPIK II의 문장 길이는 교재보다 더 다양하게 나타나지 않았으며 중급 수준의 교재 텍스트와 마찬가지로 문장 길이가 통제되고 있었다.

이상의 분석 결과를 통해, TOPIK II의 텍스트 수준이 중급과 고급 학습자들을 모두 효과적으로 가려낼 수 있도록 텍스트 개발에 대해 두 가지를 제언할 수 있을 것이다.

첫째, 전체적인 어휘 수준을 조금 하향할 필요가 있다. 현재 한국어능

력시험 텍스트의 높은 어휘 수준은 학문 목적 한국어능력시험의 부재로 인하여 일반 목적과 학문 목적의 학습자 모두를 하나의 도구로 평가해야 하는 현실과도 관련이 있을 것이다. 학문 목적 한국어능력시험이 개발된다면, 일반 목적 한국어능력시험의 어휘 난이도를 낮추는 것이 더 용이해질 것이다.

둘째로 지문 제작 시 문장 길이를 난이도 조절에 활용할 필요가 있을 것이다. 문장 길이는 고급 수준의 학습자에게는 크게 변별력 있는 요인은 아니지만, 중급 수준에서는 여전히 영향력이 있는 난이도 요인이다. 또한 실제적인 한국어 텍스트의 문장 길이 역시 매우 다양하게 나타나기 때문에 다양한 문장 길이의 텍스트를 제공하는 것은 실제적인 의사소통 상황에서의 읽기 능력을 평가하는 것과도 이어질 것이다.

두 번째 연구 문제는 난이도의 일관성에 관한 것이었다. 4회의 난이도 측정 결과에 대해 분산 분석을 실시한 결과, 문장 길이를 제외한 나머지 요인들은 통계적으로 유의한 차이를 보이지 않았다. 유일하게 차이가 있었던 문장 길이 요인은 개편 직후의 두 시험이 원인으로 나타났다. 두 회의 시험이 개편 초기에 출제되었음을 감안한다면 TOPIK의 읽기 영역은 일관적으로 텍스트의 난이도를 유지하고 있으며 신뢰도를 지닌 평가라고 볼 수 있을 것이다.

이 연구는 개편된 TOPIK이 중급과 고급 학습자를 모두 효과적으로 가려낼 수 있는지, 일관적인 수준의 읽기 지문을 제공하고 있는지를 객관적인 방법을 통해 검토하였다는 의의가 있다. 단 양적인 요인에 치중하여 문항의 출제 의도를 바탕으로 과제 유형, 담화 유형까지를 종합적으로 고려하지 못했다는 한계를 지닌다. 또한 본 연구의 결과를 바탕으로 문항반응 분석을 통해 자동적으로 측정된 결과가 실제 학습자가 체감하는 난이도와 일치하는지를 검증하는 후속 연구 역시 이루어질 필요가 있다.

<참고 문헌>

- 강범모(2014). 언어, 컴퓨터, 코퍼스 언어학, 고려대학교 출판부.
- 강현화(2012). <한국어 교육 어휘 내용 개발 1단계>. 국립국어원 연구보고.
- 강현화(2013). <한국어 교육 어휘 내용 개발 2단계>. 국립국어원 연구보고.
- 강현화(2014a). <한국어 교육 어휘 내용 개발 3단계>. 국립국어원 연구보고.
- 강현화(2014b). 국내 한국어교육기관 교재 어휘 분석 연구1 -양적 분석을 중심으로-, <외국어로서의 한국어교육> 제41권, 연세대학교 언어연구교육원 한국어학당. 37쪽~60쪽
- 구민지(2013). 한국어 읽기 교육을 위한 텍스트 난이도 측정 방법 연구, 박사학위논문, 카톨릭대학교.
- 구민지·양길석(2013). 한국어 읽기 텍스트 난이도 판별분석, <이중언어학> 제52권, 이중언어학회. 1쪽~18쪽.
- 김의수·정은주(2009). TOPIK 읽기 영역 지문의 난이도와 균질성에 관한 통사론적 접근, <한국언어문학> 제71권, 한국언어학회. 189쪽~213쪽.
- 김장식(2011). 한국어능력시험 읽기 영역 텍스트의 이독성 분석 연구, <한국어문화교육> 제6권 제1호, 한국어문화교육학회. 1쪽~24쪽.
- 김중섭(2010). <국제 통용 한국어교육 표준 모형 개발>. 국립국어원 연구보고.
- 신정하(2016). 한국어교육용 텍스트의 난이도 연구 - 제보적 기능을 중심으로-, 석사학위논문, 고려대학교.
- 서혁(2011). 읽기(독서) 교육 체계화를 위한 텍스트 복잡도(Text Complexity) 상세화 연구 (1), <國語教育學研究> 제42권, 국어교육학회. 433쪽~460쪽
- 이상린(2012). 한국어능력시험 읽기 영역 평가 틀에 대한 비판적 검토, <우리어문연구> 제43권, 우리어문학회. 255쪽~279쪽.
- 이혜진(2007). 문장의 단어 수와 구조를 통한 글 형식 난도 연구, 석사학위논문, 경인교육대학교.
- 장미경(2012). 한국어 읽기 교육을 위한 텍스트 난이도 평가 방안 연구, 박사학위논문, 고려대학교.
- 정은주(2010). 한국어능력시험(TOPIK) 읽기 영역 지문의 통사론적 난이도 연구, 석사학위논문, 한국외국어대학교.
- 홍정하·최재웅·유석훈(2011). 텍스트 수준과 가독성: 한국어 학습 교재를 이용한 검증과 응용, <언어정보> 제12권, 고려대학교 언어정보 연구소. 111쪽~148쪽.
- 한국텍스트언어학회(2004). 텍스트언어학의 이해, 박이정.
- 董洋·이혁화(2013). 한국어능력시험(TOPIK) 중급과 고급 읽기 영역의 문항 유형과 구성, <人文研究> 제69권, 영남대학교 인문과학연구소. 111쪽~144쪽.
- Council of Chief State School Officers and the National Governors Association(2010).

Common Core State Standards for English Language Arts.

Homan, S., Hewitt, M., & Linder, J. (1994), The development and validation of a formula for measuring single-sentence test item readability. *Journal of Educational Measurement*, 31(4), pp. 349-358.

R Core Team. (2015), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

신정하(Shin Jungha)

고려대학교 국어국문학과

서울시 성북구 안암로 145

전화번호: 02-3290-1960

전자우편: shinjungha@korea.ac.kr

접수일자: 2016년 2월 4일

심사(수정)일자: 2016년 3월 7일

게재확정: 2016년 3월 10일