http://dx.doi.org/10.20880/kler.2021.56.231.

학문 목적 한국어 말하기 평가의 논증 기반 타당화 연구
— 외삽 추론을 중심으로

이성준 성균관대학교 문과대학 한국어교육과 강사(제1 저자)
홍은실 성균관대학교 학부대학 대우교수(공동저자)
오예림 서울대학교 국어교육과 박사과정(공동저자)
이승원 서울대학교 국어교육과 박사과정(공동저자)
조수진 서강대학교 한국어교육원 전임강사(공동저자)
민병곤 서울대학교 국어교육과 교수 / 국어교육연구소 겸무연구원(교신저자)

I. 서론
II. 이론적 배경
III. 연구 방법
IV. 연구 결과
V. 논의
VI. 결론

I. 서론

본 연구는 유학생의 학문 목적 말하기 능력을 평가하기 위하여 개발한 한국어 말하기 시험인 코스탭(Korean Speaking Test for Academic Purposes; KoSTAP)이 평가 목적과 관련하여 평가 상황 밖에 존재하는 실제적인 학문 목적 맥락에서도 타당성을 갖추고 있는지를 알아보는 것이 목적이다. 이를 위하여 본 연구에서는 Kane(1992, 2006, 2013)이 제안한 논증 기반 타당화 접근 체계 중 실제 수행과의 관련성을 기반으로 평가 결과를 해석하는 외삽 추론(extrapolation inference)의 논증 구조를 기반으로 코스탭의 타당도를 확인하고자 한다.

코스탭은 유학생이 대학 생활 및 학업 수행과 관련하여 요구되는 학문 목적 한국어 말하기 실력을 갖추고 있는지를 알아보기 위한 웹 기반 평가 도구이다. 2010년 이후 국내 대학의 유학생 규모가 증가한 가운데 범용 목적으로 개발한 한국어능력시험(TOPIK) 등급만으로 대학(원) 입학 여부를 판정하는 것에 대한 재고가 필요하며, 이와 관련하여 학문 목적 말하기 능력 평가 결과를 입학 판정의 근거로 활용하는 방안과 이를 위한 신뢰도 높은 평가

도구 개발에 대한 필요성이 제기되어 왔다(민병곤·이성준, 2016; 홍은실·박현정·조수진·민병곤·안현기·오예림 외, 2016). 학문 목적 한국어 말하기 평가에 관한 연구사 속에서는 평가 도구 구성(박광진, 2010; 한상미, 2009), 평가 과제 개발(권순용, 2015; 김상경·박동호, 2014; 박선윤, 2015), 요구 조사(홍은실 외, 2016), 평가 도구 개발(김상경, 2015; 민병곤·조수진·홍은실·박현정·강석한·이성준 외 2017; 서동선, 2015) 등의 다양한 접근이 이루어져 왔다. 이러한 연구 배경 속에서 코스탭은 유학생이 자신의 전공 공부와 대학 생활 및 학술 활동에 참여하는 데 필요한 한국어 말하기 능력을 측정하기 위하여 개발되었다(민병곤 외, 2017: 315).

개발된 학문 목적 한국어 말하기 평가 도구의 타당성을 확보하기 위해서는 대규모 시행을 통한 표준화 과정이 요구된다. 이는 해당 평가 도구가 고부담(high-stake) 평가로서의 사용 가능성을 확보하는 것을 의미한다. 교육 평가 분야에서는 이러한 일련의 과정을 타당화(validation)라고 부르는데, 이는 한 평가 도구가 그 목적에 맞게 개발 및 사용될 수 있는지 확인하기 위한 접근을 가리킨다. 언어 평가에서 타당도는 평가 도구의 질을 나타내는 지표 중 하나로서, 평가의 과정과 결과가 평가 목표를 중심으로 얼마나 긴밀하게 수행되었는지를 보여준다(민병곤·김호정·구본관·민현식·윤여탁·김종철 외, 2020: 108). 여기서 '긴밀한 수행'의 의미는 타당도의 개념을 확장하여 제시한 Messick(1989a: 13)이 검사의 해석과 사용 측면에서 "서로 다른 두 국면이 상호 연결된 통합된 타당도"라고 하면서 확장적으로 제시한 타당도 개념과 관련이 있다.

본 연구에서는 Messick이 제시한 "통합된 타당도" 개념을 평가 도구 개발과 사용 과정에 따라 체계화한 Kane(1992, 2006, 2013)의 논증 기반 타당화 접근법(argument-based approach)을 적용하여 말하기 평가 결과와 실제적인 과제 수행과의 관계를 살펴봄으로써 외삽(extrapolation) 추론에 대한 타당도를 검증하고자 한다. 연구 문제는 다음과 같다.

첫째, 유학생의 학문 목적 말하기 평가 결과를 타당화하기 위한 논증 기반 접근은 무엇이며, 실제 말하기 과제 수행으로부터 추론할 수 있는 타당도 논증은 무엇인가?

둘째, 학문 목적 한국어 말하기 평가와 실제 말하기 과제 수행의 결과는 어떠하며, 서로 어떤 관련성을 갖고 있는가?

셋째, 실제 말하기 과제 수행과의 비교를 통해 학문 목적 한국어 말하기 평가의 결과 활용과 관련하여 얻을 수 있는 시사점은 무엇인가?

II. 이론적 배경

1. 언어 평가의 타당화

전통적인 관점에서 언어 평가의 타당도는 '평가가 측정하고자 하는 것을 측정하는 정도(Lado, 1961: 321)'를 말하며, 이는 측정 결과인 평가 점수가 평가 도구의 고유한 속성을 포함하고 있는지를 확인하는 것을 가리킨다. 그런데 언어 평가에 수반되는 여러 물리적, 심리적, 인지적 조건들이 평가 과정에 영향을 미치며, 이에 따라 산출된 평가 결과의 타당도도 평가의 사용 목적과 관련된 다양한 해석을 고려하여 통합적인 접근이 필요하다는 주장이 제기되었다. 대표적인 사례로 Cronbach(1971: 445)는 검사 타당도의 증거 중심적인 접근의 한계를 지적하면서 "포괄적이고 통합적으로 검사에 대한 평가를 세분화할 필요가 있다"며 타당도 증거들의 통합과 상세화가 필요하다고 보았다. 또한 Messick(1989a: 6)은 타당도를 "검사 점수 또는 다른 평가 방식에 기반한 해석과 조치(interpretations and actions)의 적합성과 적절성을 실증적 증거와 이론적 원리가 뒷받침하는 정도에 대한 통합적이고 가치 평가적인 판단"이라고 정의하고, 주관적 판단에 취약한 단일 차원의 사

례적 접근이 아니라 사회적인 가치와 영향력을 고려하여 평가 전반에서 수집된 증거들에 기반한 통합적 접근이 필요하다는 관점을 제시하였다.

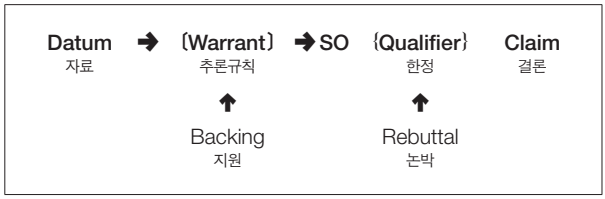
그런데 Cronbach나 Messick의 통합적 접근은 평가 점수의 해석과 사용의 측면에서 타당화를 위한 구체적인 지침을 제공하지 않는다는 점(Kane, 2006: 61), 그리고 여러 다양한 타당도 증거들의 우선순위와 그것이 평가 형태에 따라 어떻게 달라지는지에 대한 설명이 없다는 점(Cook & Beckman, 2006: 19) 등의 체계적인 접근이 부족하다는 한계가 있었다. 이에 Kane(1992, 2006, 2013)은 Cronbach(1988)가 제안한 "평가 논증에 근거한 타당도" 개념을 바탕으로 수험자의 수행에 대한 관찰과 그로부터 유도된 결론에 이르는 여러 단계의 추론을 타당도에 대한 논증으로 보고, 이러한 논증으로서의 타당도를 계획하고 평가하는 논증 기반 접근법(argument-based approach)를 제안하였다. 논증 기반 접근법은 타당화를 연쇄적인 논증으로 이루어지는 체계적인 과정으로 구성하는데, 평가 특성에 따라 맥락 및 사용 목적을 고려하여 유연하게 변형 및 적용이 가능하다는 장점이 있으며(Kane, 2012: 4), Mislevy et al.(2003), Chapelle et al.(2008), Bachman & Palmer(2010) 등의 연구에서 언어 평가 타당화의 틀로 이용되었다.

2. 논증 기반 타당화

논증 기반 타당화란 논리적으로 구성한 검사에 대한 해석 논증(interpretive argument)을 타당도 논증(validity argument)으로 평가하는 접근법을 말한다(Kane, 2006: 63). 평가의 한 구성 성분에 의존하는 기존의 준거나 내용, 구인 중심의 타당화와 달리, 다른 평가 점수나 이론적인 검토 등의 증거를 근거로 평가 결과를 해석 및 사용하기 위한 논증 체계를 설계하고, 수립한 논증의 타당성을 평가하는 것에 초점을 둔다(이성준, 2018: 89). 논증 기반 접근에서 평가 수행에 부여한 점수에 대한 해석이 타당한지를 입증하기 위하여 구성하는 해석 논증은 수험자의 수행을 관찰한 것으로부터 그에 대

한 논리적인 결론을 내리는 추론을 포함하는데, 이때 추론을 지지하는 합리적인 증거인 타당도 논증이 뒷받침되어야 그 결론을 사실로 받아들일 수 있다. 따라서 논증 기반 타당화가 성립되기 위해서는 구성한 논증을 분석하여 평가 점수를 해석하는 데 필요한 이론이나 증거들이 결론을 지지하는지를 확인해야 한다.

이와 관련하여 Kane(1992, 2006)은 Toulmin(2003)이 제시한 논증 분석을 위한 도식을 이용하여 추론의 타당성을 입증하는 모형을 제안하였다.



〈그림 1〉 Toulmin의 논증 도식(Toulmin, 2003; Kane, 2006: 71 재인용)

〈그림 1〉의 Toulmin의 논증 도식에서는 소전제에 해당하는 자료(Datum)와 대전제인 추론 규칙(Warrant)[1]으로부터 주장하고자 하는 바인 결론(Claim)을 도출함으로써 논증적 추론이 성립하는데, 이 과정에서 추론 규칙을 확정하기 위한 실제적인 지원(Backing)을 필요로 하며, 판정 결과와 관련하여 예외 조건이 있을 경우 이에 대한 논박(Rebuttal)에 대처할 수 있도록 불확실성을 한정(Qualifier)해야 함을 보여 준다.

이를 언어 평가 상황에 적용하면 평가 결과로부터 추론된 주장은 수험자들의 언어 과제 수행으로부터 수집된 자료를 기반으로 하며, 이러한 주장이 성립되기 위해서는 타당한 근거(추론 규칙)를 필요로 하는데, 이는 타당화하려는 구체적인 맥락에서 수집된 특정한 증거(지원)에 기반한다.

---

1) Toulmin의 'Warrant'는 정당한 이유, 보증, 혹은 추론 규칙으로 번역되어 왔으며, 본 연구에서는 타당화가 평가 결과에 대한 '추론의 타당성을 입증하는 일'로 보고, 이에 대해 평가 도구가 갖추어야 할 형식적인 질서로서 '추론 규칙'으로 명명하였다.

〈그림 2〉 논증 기반 타당화 체계

논증 기반 타당화는 평가 결과로부터 타당한 해석을 제공하기 위하여 요구되는 맥락에 따라 여러 유형의 추론으로 이루어진다.[2] 〈그림 2〉의 ① ~ ⑤는 평가 타당화를 위한 각각의 추론 유형을 제시한 것이다. 먼저 수험자의 과제 수행으로부터 관찰 점수를 이끌어내는 채점 추론(①)과 관찰 점수로부터 비슷한 수준의 수험자가 받을 것으로 예상되는 점수를 주장하는 일반화 추론(②), 예상 점수로부터 평가에서 측정하는 이론적인 구인이나 과제 특성을 도출하는 이론 기반 추론(③), 평가에서 측정하는 구인과 같은 특성으로부터 관련 맥락에서의 수행에서 획득할 것으로 기대되는 목표 점수를 유도하는 외삽 추론(④), 끝으로 목표 점수로부터 평가 결과 사용의 타당성을 입증하는 유용화 추론(⑤)으로 구성된다(Chapelle et al., 2008: 15).[3] 이와 같은 체계를 따라 이루어지는 타당화 과정에서는 각각의 논증을 구성하는 자료와 그에 대

2) 추론의 유형은 평가 도구의 목적이나 특성에 따라 달라질 수 있으며, 본 연구에서는 논거 기반 타당화 접근에서 가장 보편적으로 사용되는 추론의 유형을 제시하였다.

3) Chapelle et al.(2008)에서는 Kane의 논증 기반 접근법에 기반하여 토플 iBT의 타당화를 위해 6개의 추론이 포함된 논증 체계를 제시하였다. 각각의 추론은 그것을 정당화하는 일반적인 원칙인 규칙으로 뒷받침되고 또한 하나 이상의 지원으로 뒷받침된다. 예를 들어 토플 iBT의 타당성 주장에 대한 추론은 '토플이 평가하는 학문적 언어 능력의 구인이 고등 교육기관의 언어 수행을 반영한다'는 규칙에 근거하며, 추론 규칙의 지원은 '토플에서의 수행이 학문적 맥락에서 필요한 언어 능력의 다른 준거들과 관련이 있다'가 된다(Chapelle et al., 2008: 21).

한 추론의 결론을 이끌어내도록 하는 추론 규칙, 그리고 추론 규칙을 뒷받침하는 지원의 구조를 바탕으로 추론의 타당성이 입증되었는지를 확인한다.

3. 코스탭의 타당화와 외삽 추론

1) 코스탭의 타당화

코스탭은 대학 및 대학원 입학을 목적으로 하는 한국어 학습자의 학문 목적 말하기 능력을 알아보기 위하여 개발된 웹 기반 준직접식 말하기 시험이다. 코스탭은 개발 단계를 세분하여 그 절차 및 결과를 분석함으로써 타당화 증거들을 지속적으로 수집해 왔다. 개발 초기에 학문 목적 한국어 말하기 능력 평가에 대한 잠재적 수요자와 이해관계자들의 인식과 요구를 파악하고(홍은실 외, 2016), 전문가 델파이 조사를 통해 평가 구인을 선정하는 과정을 구체적으로 기술하였다(이승원 · 조수진 · 이성준 · 홍은실 · 안현기 · 강석한 외, 2020). 이 결과를 전제로 평가 틀 및 문항을 개발하고, 예비 검사를 시행한 결과를 분석하였다(민병곤 외, 2017), 대규모 본 검사는 예비 검사에서 나타난 문항 특성에 대한 고전검사이론 분석을 토대로 피평가자 집단 특성에 따른 차이를 최소화하기 위하여 개선한 평가 문항을 바탕으로 이루어졌는데, 문항반응이론을 활용하여 문항 양호도에 대한 객관화된 분석 결과를 제공하였다(박현정 · 박민호 · 이성준 · 강석한 외, 2018). 또한 점수가 의미하는 구체적인 수준 정보를 제공하기 위해 분할점수를 도출하는 기준 설정(standard setting) 절차를 수행하였다(이성준 · 홍은실 · 민병곤 · 박민호 · 이승원 · 오예림 외, 2020). 그밖에도 채점 타당도 확보를 위하여 채점자 워크숍 설계(홍은실 · 오예림 · 이성준 · 조수진 · 안현기 · 이승원 외, 2020) 및 채점자 변인 분석(강석한 · 박현정 · 이성준 · 민병곤 · 홍은실 · 안현기, 2020)을 실시하였고, 평가 과정을 통해 수집된 피평가자의 응답 특성에 대한 분석(강석한 · 안현기 · 홍은실 · 민병곤 · 조수진 · 이성준 외, 2017; 이승원 · 조수진 · 홍은실 · 이성준 · 오예림 · 민병곤, 2019)을 실시하여 검사 결과의 타당도를 다각도로 증명해 왔다.

코스탭의 타당화를 위한 지금까지의 접근을 논증 기반 타당화의 관점에서 보면, 평가 영역 검토와 예비 검사를 통한 평가 도구 개발에서부터 문항별, 검사 유형, 과제 유형 등에 따른 점수 분석(채점 및 일반화 추론), 평가 구인 및 과제 선정 분석(이론 기반 추론)의 접근이 이루어져 왔으며, 이는 주로 코스탭을 구성하는 내부적인 증거에 기반한 접근이라고 말할 수 있다. 현재의 과제는 코스탭이 평가 도구 외적으로 존재하는 실제적인 목표 영역인 실제 학문 목적 맥락에서 이루어지는 말하기 과제 수행과 어떤 관련성을 갖고 있는지를 살펴보는 것이며, 이는 논증 기반 타당화 체계의 외삽 추론과 관련이 된다.

2) 코스탭의 외삽 추론

코스탭의 목표인 유학생의 학문 목적 한국어 말하기 능력 수준에 대한 판정과 관련하여 외삽 추론에서는 코스탭 평가 결과가 실제 맥락에서 갖는 설명력을 확인하며, 이를 보장해 주는 명제와 그에 대한 증거를 '실제적인 맥락에서 이루어진 말하기 과제 수행'을 통해 살펴볼 수 있다(Flockton & Crooks, 2002).

코스탭의 외삽 추론을 구성하는 논증 구조는 〈그림 2〉에서 제시한 바와 같이 코스탭 각 문항을 통해 측정한 학문 목적 말하기 능력의 특성이 실제 학문 목적 맥락에서 이루어지는 말하기 수행에 대한 설명력을 갖추었다는 결론을 기본으로 한다. 구체적인 코스탭의 외삽 추론 논증은 툴민의 논증 요소를 활용하여 아래와 같이 설정할 수 있다.

〈표 1〉의 외삽 추론 논증 구조에서 추론 규칙(W)은 자료(D)로부터 추론 과정을 통해 타당한 결론(C)을 이끌어내기 위해 요구되는 것이며, 추론 규칙과 자료는 각각의 특성에 따라 이를 입증하는 지원(B)을 요구한다. 여기서 추론 규칙(W)의 지원(B)은 코스탭과 유사한 실제 상황에의 말하기 과제 수행 간에 관련성이 있다면, 해당 상황에 대한 설명력이 있다는 것을 입증하는 이론적 증거들이다. 자료(D)의 지원(B)은 코스탭 결과와 실제 상황에서 수행한 결과가 유사하다는 명제를 입증하는 증거들이며, 두 평가 결과의 상

관 분석 결과에 기초한다.

〈표 1〉 코스탭의 외삽 추론 논증

〈코스탭의 외삽 추론 논증〉

추론 규칙(W): 어떤 평가 도구(구인 및 과제)로부터 부여된 점수가 관련된 실제 상황(구인 및 과제)에서의 수행 점수와 유사하다면 그 평가 도구는 실제 상황에 대한 설명력을 갖고 있다고 볼 수 있다.

⇦ 지원(B1-1): 언어 평가 타당도에 대한 이론이 이러한 추론 규칙을 지지한다.

⇦ 지원(B1-2): Cronbach(1971)나 Messick(1989)의 준거 타당도 이론이나 Kane(2006)의 논증 기반 타당화 이론에서는 "준거 관련 타당도의 증거가 검사 점수와 타당한 준거 측정치의 직접적인 결속을 확립"한다고 하였다.

자료(D): 코스탭의 평가 맥락(구인, 과제)과 이와 유사한 실제 상황의 수행(구인, 과제)에서 코스탭과 유사한 결과를 얻었다.

⇦ 지원(B2-1): 코스탭과 실제 상황에서의 수행에서 유의미한 상관 분석 결과가 도출되었다.

⇦ 지원(B2-2): 코스탭의 수준 판정 결과와 실제 상황에서의 수행에 대한 수학 가능성 평정 결과에서 유의미한 상관이 도출되었다.

결론(C): 그러므로 코스탭은 실제 학문적 상황의 말하기 과제 수행에 대한 설명력을 갖고 있는 평가 도구이다.

본 연구에서는 이상에서 제시한 코스탭의 외삽 논증 구조에서 추론의 타당성을 뒷받침하는 지원(B)을 확인하기 위하여 코스탭 결과와 실제 목표 영역에서의 수행 결과를 비교하기 위한 실험을 설계하였다.

III. 연구 방법

1. 연구 참여자

본 연구에 참여한 이들은 대학 부설 한국어교육 기관에서 5급 이상에 재학 중인 30명의 유학생들로서, 모두 대학 및 대학원에 진학하여 유학을 희

망하거나, 현재 유학 중인 이들이다.

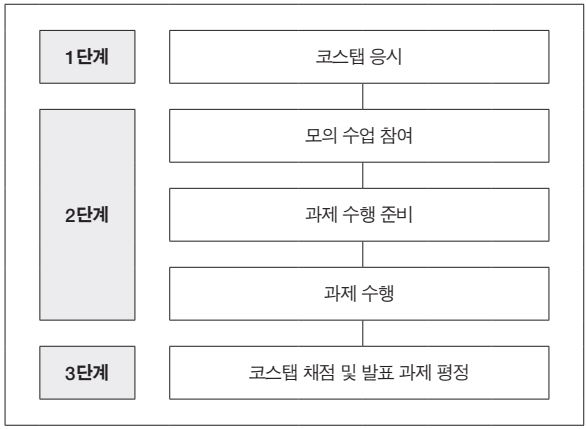
〈표 2〉 연구 참여자 특성

특성	빈도(명)	비율(%)	특성	빈도(명)	비율(%)
연령			**TOPIK 등급**		
10대	1	3.3	2급	1	3.3
20대	22	73,3	4급	11	36.7
30대	4	13.3	5급	9	30.0
			6급	6	20.0
			무응답	3	10.0
성별			**학습 기간**		
남	6	20.0	1년 미만	7	23.3
여	24	80.0	1~2년	14	46.7
			2년 이상	8	26.7
			무응답	1	3.3
한국어수준			**희망 전공**		
5급 이상	30	100.0	인문·사회	13	43.3
			경제·경영	5	16.7
			예체능	4	13.3
			과학·공학	4	13.3
			무응답	4	13.3
전체	30	100.0	전체	30	100.0

〈표 2〉의 연구 참여자 특성을 보면, 주로 20대, 여성의 비율이 높았으며, 모두 한국어 교육 기관에서 5급 이상에 재학 중으로 일상적인 한국어 의사소통 능력을 갖고 있는 이들이었다. 연구 참여자 가운데에는 과거에 한국어능력시험(TOPIK)에 응시하여 4~6급을 받은 이들이 전체 중 86.7%를 차지하고 있었으며, 한국어 학습 기간이 1년 이상인 경우가 73.4%였다. 아울러 진학 중이거나 진학을 희망하는 전공을 조사한 결과 인문·사회 계열이 43.3%로 다수를 차지하였다.

2. 연구 절차

코스탭의 타당도 논증에서 외삽 추론의 타당성을 확인하기 위하여 수험자가 코스탭에 응시한 결과와 실제 대학 강의 상황에서 말하기 과제를 수행한 결과를 수집하여 분석할 필요가 있다. 이를 위하여 본 연구 참여자들에게 코스탭 검사에 응시하도록 하였으며, 실제적인 학문 목적 말하기 능력을 확인하기 위하여 모의 수업 상황에서의 말하기 과제 수행 능력을 확인하기 위한 실험을 〈그림 3〉과 같이 설계하였다. 코스탭 검사와 말하기 과제 수행은 학습 효과 및 기타 간섭 요인을 최소화하기 위하여 같은 날 오전과 오후 시간에 진행되도록 하였다.



〈그림 3〉 연구 절차

먼저 연구 참여자들의 학문 목적 한국어 말하기 능력을 측정하기 위하여 코스탭에 응시하게 하여 평가 결과를 확보하였다. 코스탭 응시 과정에서 수험자들은 평가 웹 사이트에 접속하여 응답 녹음 방법을 연습하였으며, 이후 〈표 3〉과 같이 6개의 학문 목적 말하기 능력 평가 문항으로 구성된 준직접식 말하기 시험에 임하였다.

〈표 3〉 코스탭 문항

문항	담화 유형	과제	자료
1번	서사	경험 이야기하기	-
2번	설명	두 대상을 비교하여 설명하기	매체 비교 사진
3번	설명	자료의 내용을 분석하여 설명하기	인구 변화 그래프
4번	설득	글을 읽고 문제 해결 방안 설명하기	사회 문제에 대한 글
5번	설득	대화를 듣고 주장과 근거 제시하기	캠퍼스 문제에 대한 대화
6번	설명	강의를 듣고 요약하여 설명하기	강의 음성 자료

코스탭 응답에 대한 채점은 본 검사에서 채점자 간 신뢰도가 우수하고, 문항반응이론 분석에서 채점자 영향이 가장 적었던 두 명의 채점자에 의해 이루어졌다. 다음으로 연구 참여자의 실제적인 학문 목적 말하기 능력을 확인하기 위한 실험에서는 유학생들이 기초 교양 수업을 수강하는 상황을 가정하였다. 연구 참여자들이 제공된 강의를 시청한 후에 소집단 토의에 참여하여 학습 내용에 대한 이해를 확인하고 심화시킨 후, 이를 바탕으로 강의에 관한 학문 목적 말하기 과제를 청중 앞에서 발표하는 수업 상황을 설계하였다. 그리고 수행한 발표 과제를 녹음하여 학문 목적 한국어 교육을 담당하고 있으며 대학 및 대학원 강의 경력이 있는 2명의 교수에게 제공하고, 1~5점의 총체적 척도로 평가하게 하였다. 또한 과제 수행에 대한 평가 결과를 다면적으로 확인하기 위해 학문 목적 수학 가능성에 따라 Pass/Fail로 양분하여 표시하도록 하였다.

3. 모의 수업

연구 참여자들이 코스탭 응시 이후에 참여한 모의 수업은 대학에서 배우는 기초 교양 과목들 중 경제학의 기본 개념을 다루는 '경제와 생활' 수업에서 '기회비용'에 대한 강좌로 구성하였다. 모의 수업에서는 실제성 확보와 교수 방식에 따른 편차를 통제하기 위하여 연구 참여자들이 함께 온라인 강의를 시청하

도록 하였으며, 전체적으로는 학문 목적 한국어교육 전문가가 진행하였다.[4)]

연구 참여자들은 동영상 강의를 시청한 후에 조별로 강의 내용에 관한 소집단 토의에 참여하였다. 이는 활동 기반으로 진행되는 대학 강의 상황에서 내용에 대한 학습 후에 자신의 이해를 점검하고, 다른 동료 학습자와 관련 정보 및 경험들을 공유하면서 학습 내용에 대한 비계를 형성하는 실제성을 고려한 것이며, 동시에 연구 참여자들이 동영상 강의 내용 중에 잘 이해하지 못한 부분에 대해 보완할 수 있도록 기회를 제공한 것이다. 소집단 활동에서는 유학생들이 대학에서 강의를 듣는 실제 상황 요소를 고려하여 모든 조마다 한국인 대학생을 한 명씩 배치하였으며, 참여한 한국인 대학생들이 연구 참여자의 내용 이해에 미칠 수 있는 직접적인 영향을 통제하기 위하여 토의 과정에서 진행하는 역할을 맡도록 하였다. 토의 후에 연구 참여자들은 무작위로 선정된 발표 순서를 따라 과제 수행을 위해 대기실로 이동한 후에 강의실에서 발표 과제를 수행하였다. 모의 수업의 구체적인 진행 과정 및 내용은 〈표 4〉와 같다.

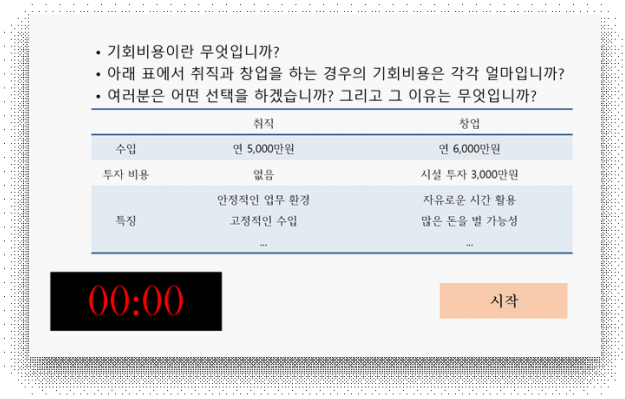
〈표 4〉 모의 수업 진행 과정 및 내용

활동 단계	교사 발화	학생 활동
인사 및 활동 소개	안녕하세요? 오늘 강의의 진행을 맡은 ○○○입니다. 만나서 반갑습니다. 오늘은 '기회비용'에 대한 동영상 강의를 들을 것입니다. 이 동영상 강의는 〈경제와 생활〉이라는 교양 수업의 일부입니다.	강의 형식 및 주제 확인하기
동영상 강의 시청	동영상 강의를 보면서 기회비용이 무엇인지 그리고 기회비용을 어떻게 계산하는 것인지 잘 들어보세요. 나눠준 종이에 메모하면서 들으면 됩니다.	동영상 강의 시청 및 주요 내용 필기
조별 활동	기회비용에 대한 동영상 강의를 잘 들으셨어요? 그럼 이제 스크린에 제시된 질문들에 대해 조별로 이야기해 봅시다. 시간은 15분입니다. 1) 기회비용이란 무엇입니까? 2) 자신이 경험한 '기회비용의 예'를 이야기해 보세요.	기회비용 개념, 사례, 경험 등을 함께 이야기하기
과제 수행 준비	기회비용에 대한 이야기를 이제 마무리하십시오. 이제 개별 과제 순서를 정하도록 하겠습니다. 연구원이 옆을 지나가면 순서표를 뽑으세요. 1번을 뽑은 학생은 연구원을 따라 이동해 주세요. 그리고 나머지 학생들은 지금 나눠주는 연구 참여 확인서를 작성하면서 기다려 주세요.	순서를 따라 대기실로 이동하여 발표 준비

4) 본 연구에서 이용한 강의 영상의 출처는 '매경 전국민 지식나눔 미시경제편'에 등록된 동영상(https://tv.naver.com/v/108811)이다. 강의 상황에서 교수자의 차이 및 상호작용으로 인하여 발생할 수 있는 변수를 통제하기 위하여 동영상 강의를 시청하게 하였다.

4. 학문 목적 말하기 과제 수행

모의 수업 참여 과정에서 연구 참여자들은 강의 영상을 통해 '기회비용'이 무엇인지에 관해 개념과 사례 등을 학습하였다. 그리고 소그룹 토의를 통해 자신이 이해한 '기회비용'이란 무엇이며, 자신의 경험 속에서 이를 적용할 수 있는 사례에 대해 이야기해 보았다. 모의 수업을 바탕으로 연구 참여자들에게 제시한 말하기 과제는 〈그림 4〉와 같다.



〈그림 4〉 말하기 과제

연구 참여자들은 말하기 과제 수행 전에 대기실에서 과제를 안내 받고, 필요한 단어나 내용을 검색하면서 발표 내용을 정리하는 준비 시간을 가졌다. 준비 시간이 끝나면 연구 참여자들은 발표장으로 이동하여 기회비용에 대한 말하기 과제를 수행하였다. 말하기 과제는 강의 내용에 관한 개인 발표 과제였으며, 이때 토의에 참여했던 한국인 대학생과 한국어교육 전문가가 청중으로 참여하여 학문 담화 맥락의 실제 상황을 구현하고자 하였다. 구체적인 말하기 과제 수행의 진행 과정과 내용은 〈표 5〉와 같다.

〈표 5〉 말하기 과제 수행 과정과 내용

활동 단계	교사 발화	학생 활동
문제 확인	기회비용에 대한 동영상 강의도 보고, 기회비용의 예에 대한 조별 토의를 통해 기회비용을 더 잘 이해하게 되었지요? 이제 문제지를 보고 발표를 준비하세요.	과제 파악하기 발표 내용 정리
과제 수행	기회비용에 대한 동영상 강의 시청과 조별 토의를 바탕으로 발표를 해 주십시오. 기회비용의 개념은 무엇인지, 표에 제시된 취직과 창업의 경우, 기회비용은 각각 얼마인지, 여러분은 어떤 선택을 할 것인지, 그리고 그 이유는 무엇인지 말해 주세요. 발표 시간은 3분이며, 화면에 시계를 통해 시간을 확인할 수 있습니다.	교사 및 한국인 학생 앞에서 개인 발표 수행

5. 분석

연구 참여자들의 코스탭 검사 결과와 모의 수업 과제 수행 결과의 관계를 확인하기 위하여 채점을 실시하였다. 코스탭의 채점은 다수의 코스탭 문항 채점 경험을 갖고 있는 한국어교육 전공자 2인이 복수 채점을 하였다. 말하기 과제 수행에 대한 채점은 대학에서 학문 목적 한국어 강의를 맡고 있으며, 한국어교육을 전공한 2인의 강사가 복수로 채점하였다. 코스탭의 결과는 6개 문항의 응답에 대하여 내용, 조직, 어휘와 문법, 발음과 유창성의 4가지 측면으로 구성된 채점 척도를 바탕으로 문항별로 1~5점을 부여하는 총체적 채점을 통해 산출하였다.

발표 평가에 대한 채점은 발표의 수준을 고려하여 '매우 미흡'에 해당하는 1점부터 '매우 우수'에 해당하는 5점까지의 점수를 부여하는 것으로 이루어졌다. 아울러 발표 평가 결과의 의미를 해석하기 위하여 해당 참여자의 학문 목적 말하기 능력 수준이 대학에서 수학 가능한 수준인지에 대한 양분 평정(Pass/Fail)을 함께 요청하고, 연구 참여자별로 'Pass'를 받은 수의 평균값을 산출하였다. 수집한 평가 자료의 분석은 각 측정값에 대한 기본적인 신뢰도 확인과 함께 각 평가 간의 수준 차이 및 관계를 파악하기 위하여 기술 통계 및 상관 분석을 중심으로 진행하였다.

IV. 연구 결과

1. 기술 통계

1) 코스탭

연구 참여자들의 코스탭 응시 결과, 〈표 6〉와 같이 전체 6개 문항 유형 중에서 1번 문항이 5점 만점에서 평균 3.22(SD=0.63)점으로 평균 점수가 가장 높았고, 2번 문항은 평균 2.43(SD=1.11)점으로 가장 낮은 것으로 나타났으며, 1, 2번 문항을 제외한 나머지 문항들은 평균값에서 대체로 비슷한 수준인 것으로 나타났다.

〈표 6〉 코스탭 결과(타당화 연구 참여자)

문항	사례 수	최소값	최대값	평균	표준편차
1번	30	1.00	5.00	3.22	0.63
2번	30	0.00	4.00	2.43	1.11
3번	30	1.00	4.00	2.68	0.70
4번	30	1.00	4.00	2.87	0.71
5번	30	1.00	5.00	2.80	0.65
6번	30	0.00	4.00	2.77	0.96

코스탭 기술 통계치에 대한 해석을 위해 2017년에 실시한 코스탭 본 검사[5]에서 5급 이상으로 한국어 수준을 표시한 응시자들의 검사 결과를 살펴

5) 코스탭 본 검사는 2017년 6월부터 8월까지 전국 11개 대학에서 396명의 유학생을 대상으로 진행되었다. 본 검사의 시행 과정과 종합적인 결과는 박현정 외(2018)를 통해 보고되었다.

보았다.

〈표 7〉 코스탭 본 검사 결과(한국어 수준 5급 이상)

문항	사례 수	최소값	최대값	평균	표준편차
1번	90	0.00	5.00	2.87	1.22
2번	90	0.00	5.00	2.23	1.48
3번	90	0.00	5.00	2.60	1.35
4번	90	0.00	5.00	2.19	1.44
5번	90	0.00	5.00	2.60	1.39
6번	90	0.00	5.00	2.47	1.51

〈표 7〉에서 5급 이상 한국어 학습자의 코스탭 본 검사 결과를 보면, 1번 문항의 평균 점수가 5점 만점에서 2.87로 가장 높았으며, 4번 문항이 2.19로 가장 낮은 것으로 나타났다. 이를 〈표 4〉의 결과와 비교해 보면, 본 연구의 참여자들이 본 검사 응시자들에 비해 모든 문항 유형에서 평균 점수가 높으며, 상대적으로 낮은 곤란도를 나타내었다.

2) 말하기 과제 수행 평가

연구 참여자들이 수행한 발표에 대하여 2인의 전문가들이 1~5점으로 평정한 결과, 〈표 8〉과 같이 평균 2.42(SD=1.02)점으로 나타났다.

〈표 8〉 말하기 과제 수행 평가 결과(점수 평정)

사례 수	최소값	최대값	평균	표준편차
30	1.00	5.00	2.42	1.02

아울러 대학 수학 가능 여부에 대한 평정에서는 〈표 9〉와 같이 전체 30

명의 연구 참여자에 대하여 1인당 평균 1.3명의 교수자가 수학 가능하다는 판정을 내렸으며, 채점자 중 1인이라도 수학 가능하다고 판단(Pass)한 경우가 전체 30명 가운데 26명이었고, 2인 모두 가능하다고 판단한 경우는 13명, 2인 모두 불가능하다고 판단한 경우가 4명으로 나타났다.

〈표 9〉 말하기 과제 수행 평가 결과(P/F 평정)

사례 수	전체 평균 (Pass 빈도)	평정 사례별 빈도		
		P×2	채점자 1인 이상 Pass	채점자 0인 Pass
30	1.3회	13명	26명	4명

2. 신뢰도 분석

코스탭과 발표 평가에 대한 채점이 신뢰할 만한 수준으로 이루어졌는지를 확인하기 위하여 급내 상관 계수(Intra-class correlation coefficient)를 확인하였다.

〈표 10〉 채점자 간 신뢰도 분석 결과표

측정 도구	급내 상관	95% 신뢰구간
코스탭	.820***	.758 ~ .866
말하기 과제 수행(점수)	.824***	.630 ~ .916
말하기 과제 수행(P/F)	.485***	-.082 ~ .755

***: $p < .001$

코스탭의 채점에 대한 신뢰도 검정 결과, 급내 상관이 .824($p < .001$)로 양호한 수준의 채점자 간 신뢰도가 확보된 것으로 나타났다.

다음으로 말하기 과제(발표) 수행에 대하여 전문가 2인이 점수로 평정

한 결과와 Pass/Fail 평정의 결과에 대한 급내 상관을 확인하였다. 그 결과, 점수 평정에서는 급내 상관이 .824(평균 측도, 유의확률: .000)로 우수한 것으로 나타났으나, Pass/Fail 평정에서는 .485(평균 측도, 유의확률: .000)로 낮은 급내 상관이 나타났다.

3. 상관 분석

코스탭이 실제 학문 목적 맥락에서의 수행과 관련하여 어느 정도의 관련성을 갖고 있는지를 확인하기 위하여 연구 참여자들이 코스탭에 응시한 결과와 발표 수행 결과와의 상관 분석을 실시하였다.

〈표 11〉 상관 분석 결과

	점수 평정	P/F 평정
상관 계수	.503**	.514**
유의확률	.005**	.004
사례 수	30	30

**: $p < .01$

〈표 11〉과 같이, 연구 참여자들의 말하기 과제 수행과 코스탭 결과의 상관을 Pearson 적률상관법을 적용하여 산출한 결과, 점수 평정 결과와는 .503, 수학 가능성 평정과는 .514로 보통 수준의 상관이 있는 것으로 나타났다($p<.01$).[6)]

발표 과제 수행에 대한 코스탭의 문항 유형별 영향을 구체적으로 확인하기 위하여 연구 참여자들이 코스탭의 6개 문항 유형에서 얻은 점수와 과

6) 상관 계수는 +1과 −1 사이의 위치하는 상관 관계 크기의 값을 의미하며, 일반적으로 0.7 이상은 강한 상관 관계, 0.3~0.6은 보통 상관 관계, 0.3 이하는 약한 상관 관계로 해석할 수 있다. 상관 관계에 대한 해석은 유의 확률값을 기준으로 하며, 대립 가설이 채택될 경우에 선형적인 관계를 가정할 수 있다.

제 수행의 상관을 재산출하였다.

〈표 12〉 코스탭 문항별 상관 분석

말하기 과제 수행 평정	코스탭 문항	1번	2번	3번	4번	5번	6번
점수 평정	상관 계수	.369	.346	.466	.334	.217	.479
	유의확률	.045*	.061	.010*	.071	.249	.007**
	사례 수	30	30	30	30	30	30
P/F 평정	상관 계수	.234	.287	.506	.458	.355	.460
	유의확률	.213	.124	.004**	.011*	.054	.011*
	사례 수	30	30	30	30	30	30

**: $p < .01$, *: $p < .05$

〈표 12〉에서 코스탭 문항별 말하기 과제 수행 평정의 상관을 보면, 보통 수준의 상관을 나타낸 경우가 점수 평정에서는 3, 6번, Pass/Fail 평정에서는 3, 4, 6번으로 나타났다.

V. 논의

본 연구에서는 유학생의 학문 목적 한국어 말하기 평가로서 개발된 코스탭이 실제 강의 상황에서의 학문 목적 말하기 과제 수행과 관련하여 타당한 해석을 제공할 수 있는지를 알아보고자 연구를 수행하였다. 이 장에서는 II장에서 제시한 코스탭 외삽 추론의 논증구조를 구성하는 추론 규칙(W)과 자료(D), 논박(R)을 중심으로 이를 뒷받침하는 증거인 지원(B1-1, 1-2), (B2-1, 2-2)의 의미를 해석하고 관련 요소들을 살펴보고자 한다.

1. 추론 규칙(W)

> 어떤 평가 도구로부터 부여된 점수가 관련된 실제 상황에서의 수행 점수와 유사하다면 그 평가 도구는 실제 상황에 대한 설명력을 갖고 있다고 볼 수 있다.

코스탭의 개발 목적은 유학생이 국내 대학에서 유학을 하는 데 필요한 한국어 말하기 실력을 어느 정도로 갖추었는지를 알아보기 위한 것이며(민병곤 외, 2017), 본 연구에서는 대학 현장, 즉 실제 강의 맥락에서 요구되는 한국어 말하기 능력을 코스탭 결과를 통해 설명할 수 있을 것인지를 알아보고자 하였다. 이러한 명제의 성립을 위하여 뒷받침되어야 할 지원으로는 첫째, 언어 평가 이론이 이러한 추론 규칙을 지원해야 하며(B1-1), 둘째, 직접적으로 준거 타당도 이론이나 논증 기반 타당도 이론에서 이러한 접근을 포함하고 있어야 한다(B1-2). 본 연구의 평가 도구인 코스탭은 학문 목적 말하기 평가로서 이를 실제 맥락에 적용할 수 있으려면 평가 외부에 존재하는 준거, 즉 준거 관련 타당도(criterion-related validity)의 지원을 수반해야 한다.

이와 관련하여 Weir(2005: 36)는 준거 관련 타당도 가운데 평가 도구가 목표로 하는 관련된 실제 상황 맥락에서의 수행에 대한 예측력을 확인하는 예측 타당도(predictive validity)를 고려할 수 있으나, 측정상의 어려움으로 서로 다른 영향을 고려해야 하는 두 검사로부터 충분한 결론을 도출하기에 어려움이 따른다고 하였다. 이에 본 연구에서는 실제 맥락에서의 수행과 코스탭 검사 사이에 간섭할 수 있는 요인을 최소화하기 위하여, 기간이나 학습 등의 변인의 영향을 통제할 수 있는 〈그림 3〉의 연구 설계를 마련하여 측정의 타당성을 확보하고자 하였다.

아울러 본 연구는 타당화 연구 설계 과정에서 Cronbach(1971), Kane(1992, 2006, 2013), Messick(1989b)이 제안한 평가 타당화에 대한 통합적인 접근의 가치에 주목하였는데, 목표 영역과의 관계를 확인하는 일을 통해 개별 평가 결과를 일반화된 전집 점수(universe score)로 추론 가능하게 한

다는 점(Kane, 2006), 이러한 접근이 다양한 이해 관계자가 평가 결과를 이해하는 데 도움이 된다는 점(Messick, 1989b), 그리고 직접적으로 수험자가 과제를 어떻게 수행하는지에 대한 정보를 제공할 수 있다(Cronbach, 1971)고 언급된 바 있다. 이러한 이론적 검토의 결과를 바탕으로 본 연구에서 설정한 외삽 추론 논증 구조에서 추론 규칙(W)의 정당성을 확인할 수 있다.

2. 자료(D)

> 코스탭의 평가 맥락과 유사한 실제 상황의 수행에서 코스탭과 유사한 결과를 얻었다.

먼저 자료에 대한 첫 번째 지원(B2-1)과 관련하여 코스탭의 평가 결과가 실제 상황에서의 말하기 과제 수행에 대한 평가 결과와 유사한지를 둘 사이의 선형적 관계를 보여줄 수 있는 상관 계수를 통해 확인할 수 있다. 이와 관련하여 〈표 11〉에서 제시한 상관 분석 결과에서 코스탭과 말하기 과제 수행 평정의 상관은 점수 평정과 Pass/Fail 평정에서 보통 수준으로 나타난 바 있다. 이러한 결과와 관련하여 다른 대규모 검사의 타당화 연구에서 평가 도구에서 측정한 값과 실제 현장의 교사나 다른 이해 관계자가 평정한 값의 상관 혹은 다른 평가 결과와의 상관을 비교하는 준거 타당도 조사 결과를 참조할 수 있다.

관련하여 박현정 외(2018)에서는 준거 기반 타당도를 확인하기 위하여 코스탭과 수험자의 한국어 능력 수준과의 상관을 조사하였으며, 그 결과 약한 상관(.4)의 관계에 있음을 확인하였다. 국외 평가 가운데 TOEFL의 타당화 연구(Chapelle et al., 2008: 343)에서는 외삽 추론과 관련하여 조사한 TOEFL 점수와 교수자 평정과의 상관 분석을 하였는데, 낮은 수준에서 보통 수준의 상관(.27~.51/.42~.49)이 나타났다. 이와 같은 실제 학문 목적 맥락과 관련하여 수행한 준거 타당도 조사 결과와 비교하였을 때, 본 연구에서 나타난 코스탭과

실제 상황에서의 말하기 과제 수행의 상관(.503~.516)은 수용 가능한 수준으로 볼 수 있다. 이는 코스탭이 대학 상황에서의 말하기 과제 수행에 대해 수용 가능한 설명력을 갖고 있음을 보여주는 것이라고 할 수 있다.

다음으로 자료에 대한 두 번째 지원(B2-2)과 관련하여 연구 참여자들이 코스탭에서 받은 점수를 이성준 외(2020)에서 제시한 코스탭 수준별 분할 점수(〈표 13〉)를 기준으로 구분해 보면, 전체 30명의 연구 참여자 가운데 코스탭 문항별 점수의 총점이 14.86점 이상으로 대학 수학 가능 수준에 해당하는 경우는 21명이며, 9명은 입학 불가에 해당하는 수준인 것으로 나타났다.

〈표 13〉 코스탭 문항유형별 분할 점수

문항 유형	평균 (표준편차)		
	독립적 수학 가능	보완적 수학 가능	제한적 수학 가능
1번	4.65 (.129)	3.92 (.243)	2.78 (.217)
2번	4.58 (.134)	3.75 (.263)	2.60 (.294)
3번	4.43 (.172)	3.68 (.235)	2.51 (.335)
4번	4.31 (.231)	3.39 (.244)	2.29 (.258)
5번	4.51 (.126)	3.74 (.202)	2.50 (.253)
6번	4.43 (.166)	3.35 (.296)	2.18 (.374)
총점	26.91	21.83	14.86

본 연구 참여자 가운데 코스탭의 수준별 분할 점수를 기준으로 '제한적 수학 가능'보다 낮은 수준, 즉 대학 입학이 어려울 것으로 판단되는 수준의 학생 수 9명은 앞서 〈표 9〉의 Pass/Fail 평정 결과에서 채점자 2인 모두 수학 불가로 판정한 사례(P×0, 4명)보다는 많고, 1인만 수학 불가로 판정한 사례(P×1, 13명)보다는 적은 결과이다. 한 개인에게 있어 어떤 일의 가능성을 예측한다는 것은 절대적으로 큰 부담이 따르는 일이므로, 이 때문에 교육 평가

영역에서도 교육의 책무성(accountability)을 강조하며, 평가에 참여하는 이들이 이해 가능한 평가가 이루어질 수 있도록 할 것을 강조하고 있다. 이러한 관점에서 코스탭의 평가 결과를 바탕으로 수험자에게 대학에서의 수학이 어렵다는 판정을 하기 이전에 이들이 어떤 점에서 보완이 필요한지에 대한 정보를 제공할 수 있어야 할 것이다. 특히 코스탭 평가 결과, 입학 불가로 판정된 9명 중 실제 말하기 과제 수행에서 채점자 1인이 Pass를 부여한 경우는 판정 결과가 양분되어 판단이 어려운 사례에 해당한다. 따라서 이를 합리적이고 온건한 해석으로 귀결시킬 수 있도록 하려면 코스탭 측정값에 대한 상세한 검토와 측정 오차를 가져온 다른 영향은 없는지, 그리고 해당 수험자의 다른 과제 수행에서의 결과는 어떤지 등을 종합적으로 검토할 필요가 있다. 또한 코스탭 평가 내적으로는 수험자를 비롯한 이해관계자들이 기대하는 학문 목적 말하기 능력의 특성이 응답에 포함될 수 있도록 과제에서 제시하는 맥락이나 채점 기준 등을 보완할 필요가 있다.

3. 논박(R)

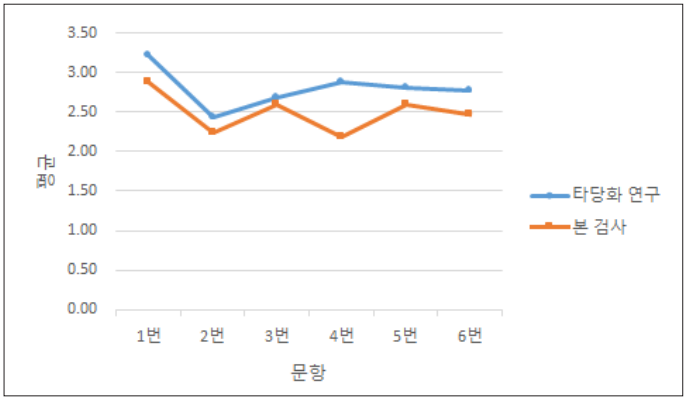
> 코스탭 평가 결과와 실제 학문적 상황의 말하기 과제 수행의 관련성을 부정하는 증거가 있다.

본 연구에서 수행한 외삽 추론 논증의 타당성을 저해하는 요인과 관련하여 측정 과정에서 발생할 수 있는 오차 요인과 구인 무관 변인 등의 영향을 생각해 볼 수 있으며, 연구 참여자의 개인적 특성과 코스탭 검사 사이의 상호작용으로 인한 영향이 나타날 가능성이 있다.

이와 관련하여 〈표 10〉의 채점자 간 신뢰도 분석 결과에서 신뢰도가 낮게 나타난 말하기 과제 수행 평가에서의 Pass/Fail 평정의 경우, 평정을 맡은 교수자 2인이 주관적인 경험만을 바탕으로 양분된 척도로만 평정하도록 하였고, 대상 피험자 수도 30명으로 적은 상황이었다는 점 등의 영향을 받은

것으로 예상된다.

아울러 연구 참여자의 코스탭 결과가 일반화된 전집을 모든 문항에서 대표하는지와 관련하여 〈표 12〉에서는 약한 상관이 나타난 문항이 있음을 확인한 바 있다. 이를 〈표 6〉와 〈표 7〉을 통해 일반화 전집에 해당하는 유사한 한국어 수준의 코스탭 본 검사의 문항별 평균값을 비교해 보았을 때, 곤란도가 가장 높은 것으로 나타난 문항 유형은 2번 '비교하여 말하기'였으며, 가장 낮은 것으로 나타난 문항은 1번 '경험 말하기'였다. 이러한 결과는 4번 '자료 읽고 의견 제시하기' 문항의 곤란도가 가장 높았던 코스탭 본 검사 결과와는 다른 경향이었다.



〈그림 5〉 코스탭 문항별 평균 비교(타당화 연구-본 검사)

이와 관련하여 두 응시자 집단의 코스탭 문항별 평균 차이를 명시적으로 확인하기 위하여 〈그림 5〉를 제시하였다. 앞서 언급한 바와 같이, 본 타당화 연구 참여자들의 코스탭 점수는 평균적으로 전체 문항에서 본 검사 응시자에 비해 높은 것으로 나타났으나 대체로 각 문항별로 두 집단의 상승과 하강의 흐름이 유사하게 나타났다. 그러나 4번 문항의 경우에는 두 집단의 격차가 심화되어 나타나고 있었는데, 타당화 연구 참여자 집단은 해당 문항의 평균 점수가 3번이나 5번 문항보다 높은 반면에 본 검사 참여자 집단의 경우

에는 3번이나 5번 문항보다 낮았다. 4번 문항의 경우, 사회적인 문제에 대한 짧은 글을 읽고, 자신의 의견을 제시하는 문제였는데, 이는 학문 목적 맥락에서 일상적으로 수행하게 되는 과제라는 점에서 타당화 연구 참여자 집단이 유사한 과제 유형에 대한 경험을 많이 갖고 있었거나, 또는 5급 이상이라는 공통 조건 가운데 한국어 교육기관 수료 등급인 6급이나 최고급 과정인 7급 학생들의 비율이 상대적으로 더 높았을 때 발생할 수 있다.

아울러 2번 문항의 경우 뉴스를 접하는 두 매체의 차이를 비교하는 문제가 제시되었는데, 제시한 자료와 응답 요건이 응답을 충분한 정도로 유도하고 있지 못하였다는 것을 알 수 있다. 이러한 현상의 원인을 파악하기 위하여 해당 문항의 발문과 제시한 사진 자료, 채점 기준 기술문 등을 연계하여 다시 살펴볼 필요가 있을 것이며, 아울러 친숙도가 떨어지는 문제에 대한 개선을 위해 안면 타당도 조사도 고려할 수 있다.

끝으로 외삽 추론의 타당도 증거로서의 타당성을 높이기 위해서는 향후 코스탭의 측정 및 분석 설계를 보완하는 방법 이외에도 대규모 본 검사의 지속적인 실행과 그에 따른 데이터베이스의 확장을 기반으로 타당도 논증의 기반이 되는 전집 점수를 보완해 나갈 필요가 있다.

VI. 결론

본 연구에서는 학문 목적 한국어 말하기 평가인 코스탭이 실제적으로 유학생의 학문 목적 말하기 과제 수행에 관한 정보를 제공할 수 있는지를 알아보기 위하여 검사 타당화에 대한 논증 기반 접근법을 바탕으로 외삽 추론 논증 구조를 설계하고, 이를 검증하기 위한 타당화 연구를 설계 및 실행하였다. 연구 참여자들은 코스탭에 응시한 후에 모의 수업 상황에 참여하여 강의를 듣고, 그에 관한 말하기 과제를 수행하였으며, 이 과정에서 수집된 응답에

대하여 채점을 실시하고, 그 결과를 바탕으로 외삽 추론의 타당성을 확인하였다.

연구 결과에서 연구 참여자들의 코스탭 응시 결과와 발표 과제 수행에 대한 교수자 평정 결과를 통해 코스탭이 대학에서 요구하는 실제적인 학문 목적 말하기 능력을 평가하는 말하기 시험으로서 기능할 수 있으며, 특히 강의 맥락과 관련된 말하기 과제 수행과 관련성이 높은 문항들을 갖추고 있음이 나타났다. 아울러 코스탭과 말하기 과제 수행의 채점 신뢰도가 확보되었다는 점, 그리고 코스탭 본 검사 결과와의 유사성 등의 증거는 향후 코스탭이 유학생의 학문 목적 한국어 말하기 능력을 사전에 확인할 수 있는 유용한 도구로서 사용될 수 있음을 보여주고 있었다.

본 연구는 연구 과정에서 이용한 학문 목적 말하기 과제의 유형이 개념 학습에 대한 발표 과제에 한정되어 있다는 점과 말하기 과제 수행 평정 과정에서 평정자의 주관성 개입이 나타난 점 등의 한계를 갖고 있다. 관련하여 본 연구의 결과를 바탕으로 향후 코스탭뿐만 아니라 학문 목적 말하기 능력을 평가할 수 있는 다양하고 통합적인 말하기 과제 유형에 대한 검토가 이루어져야 할 것이며, 이들의 담화적 속성에 대한 언어적 분석과 내용의 질적인 분석 등이 이루어진다면 학문 목적 말하기 평가 문항의 측정 기능 향상에 기여할 수 있을 것이다. 또한 본 연구에서 실시한 외삽 추론을 넘어 실질적으로 수험자들의 학문 목적 말하기 평가 결과 활용에 대한 여러 이해 관계자들의 요구를 반영할 수 있는 평가 결과 사용에 대한 유용화 추론에 관한 연구들도 이루어질 것을 기대한다.

* 본 논문은 2021.2.7. 투고되었으며, 2021.2.18. 심사가 시작되어 2021.3.16. 심사가 종료되었음.

참고문헌

강석한·박현정·이성준·민병곤·홍은실·안현기(2020),「온라인 방식 한국어 학문 목적 말하기 평가에서 변인 연구」,『한국언어문화학』17(1), 35-62.

강석한·안현기·홍은실·민병곤·조수진·이성준·박현정(2017),「외국어로서의 한국어 발음과 유창성 연구」,『이중언어학』67, 1-29.

권순용(2015),「학문 목적 한국어 학습자의 말하기 평가 도구 개발 연구」, 부경대학교 석사학위논문.

김상경·박동호(2014),「한국어교육학: 학문 목적 한국어 말하기 평가 과제 유형 개발 연구: 다국면 라쉬 모형과 일반화가능도 이론 적용을 중심으로」,『새국어교육』100, 115-141.

김상경(2015),「학문 목적 한국어 학습자의 말하기 능력 평가 방안 연구」, 경희대학교 박사학위논문.

민병곤·김호정·구본관·민현식·윤여탁·김종철·고정희·윤대석·강경민·김가람·김영미·김지혜·김혜진·신현단·유민애·이성준·이수미·이슬비·홍은실(2020),『한국어교육학개론』, 서울: 태학사.

민병곤·이성준(2016),「학문 목적 한국어 말하기 평가 도구 개발의 쟁점과 과제」,『국어교육연구』38, 67-110.

민병곤·조수진·홍은실·박현정·강석한·이성준·오예림·이승원·안현기(2017),「학문 목적 한국어 말하기 평가 도구 개발 연구」,『국어교육』157, 309-340.

박광진(2010),「학문 목적 한국어 학습자 대상 교실 말하기 평가 방안 연구」, 배재대학교 석사학위논문.

박선윤(2015),「학문 목적 한국어 말하기 평가 도구 개발 방안」, 중앙대학교 석사학위논문.

박현정·박민호·이성준·강석한(2018),「문항반응이론을 활용한 학문 목적 한국어 말하기 평가 문항 및 채점자 특성 분석」,『이중언어학』73, 177-203.

서동선(2015),「외국인 유학생을 위한 대학 수학 목적 한국어능력 평가 방안 연구」, 경희사이버대학교 석사학위논문.

이성준·홍은실·민병곤·박민호·이승원·오예림·조수진·박현정(2020),「학문 목적 한국어 말하기 평가의 기준 설정 연구」,『한국어교육』31(4), 217-249.

이성준(2018),「한국어 말하기 평가 타당화를 위한 논거 기반 접근법의 이해와 적용」,『화법연구』41, 85-116.

이승원·조수진·이성준·홍은실·안현기·강석한·박현정·민병곤(2020),「학문 목적 한국어 말하기 평가의 구인 선정 연구: 일반 목적 평가와의 비교를 중심으로」,『새국어교육』123, 315-350.

이승원·조수진·홍은실·이성준·오예림·민병곤(2019),「학문 목적 말하기 평가에서 읽기 통합형 과제 응답의 담화적 특성: 읽기 자료 사용을 중심으로」,『화법연구』44, 81-119.

한상미(2009),「학문 목적 한국어 말하기 평가 연구: 대학 입학 전 과정을 중심으로」,『한국어교

육』 20(1), 207-238.
홍은실 · 박현정 · 조수진 · 민병곤 · 안현기 · 오예림 · 강석한(2016), 「학문 목적 한국어 말하기 능력 평가 도구 개발을 위한 요구 분석」, 『한국어교육』 27(4), 243-268.
홍은실 · 오예림 · 이성준 · 조수진 · 안현기 · 이승원 · 민병곤(2020), 「한국어 말하기 평가의 채점 타당도 연구: 워크숍 기반의 채점자 훈련을 중심으로」, 『국제어문』 84, 513-544.
Bachman, L. F. & Palmer, A. S. (2010), *Language assessment in practice: Developing language assessments and justifying their use in the real world*, Oxford: Oxford University Press.
Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008), *Building a validity argument for the test of English as a foreign language* New York, NY: Routledge.
Chapelle, C. A., Jamieson, J., & Hegelheimer, V. (2003), "Validation of a web-based ESL test", *Language Testing 20*(4), 409-439.
Cook, D. A. & Beckman, T. J. (2006), "Current concepts in validity and reliability for psychometric instruments: theory and application", *The American Journal of Medicine 119*(2), 166.e7-e16.
Cronbach, L. J. (1971), Test validation, In R. L. Thorndike (Ed.), *Educational measurement*(2nd ed.), Washington, DC: American Council on Education.
Cronbach, L. J. (1988), "Five perspectives on validity argument", In H. Wainer & H. Braun(Eds.), *Test Validity*, Hillsdale, NJ: Lawrence Erlbaum.
Flockton, L. & Crooks, T. (2002), *Social studies assessment results 2001*, Dunedin: Educational Assessment Research Unit, University of Otago.
Kane, M. (1992), "An argument-based approach to validity", *Psychological Bulletin 112*(3), 527-535.
Kane, M. (2006), Validity, In R. Brennan (Ed.), *Educational Measurement*(4th ed.), Westport, CT: American Council on Education and Praeger.
Kane, M. (2012), "Validating score interpretations and uses", *Language Testing 29*(1), 3-17.
Kane, M. (2013), "The argument-based approach to validation", *School Psychology Review 42*(4), 448-457.
Lado, R. (1961), *Language testing: The construction and use of foreign language tests*, New York, NY: McGraw Hill.
Messick, S. (1989a), Validity, In R. L. Linn(Ed.), *Educational Measurement* (3rd ed.), New York: American Council on Education and Macmillan.
Messick, S. (1989b), "Meaning and values in test validation: The science and ethics of assessment", *Educational Researcher 18*(2), 5-11.
Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003), "On the structure of educational assessments", *Measurement: Interdisciplinary Research and Perspectives 1*(1),

3-62.
Toulmin, S. (2003), *The uses of argument* (update ed.), Cambridge: Cambridge University Press.
Weir, C. J. (2005), Language testing and validation: An evidence-based approach, New York, NY: Palgrave Macmillan

초록

학문 목적 한국어 말하기 평가의 논증 기반 타당화 연구
— 외삽 추론을 중심으로

이성준 · 홍은실 · 오예림 · 이승원 · 조수진 · 민병곤

본 연구는 학문 목적 한국어 말하기 평가(코스탭)가 실제적인 맥락에서 유학생의 학문 목적 말하기 과제 수행에 대한 정보를 제공하는 것과 관련하여 외삽 추론의 타당성을 갖추고 있는지를 알아보는 것을 목적으로 한다. 이를 위하여 논증 기반 타당화 체계 가운데 외삽 추론의 논증 구조를 수립하고, 툴민의 논증 요소를 기반으로 이를 검증하기 위한 연구를 설계하였다. 외삽 추론의 타당도 증거는 이론적인 검토와 코스탭 결과와 실제 목표 영역에서의 학습자의 수행 결과의 비교를 통해 수집되었다. 연구 참여자인 유학생은 모의 수업 상황에서 강의를 듣고 그에 관한 말하기 과제를 수행하였다. 분석 결과 코스탭과 실제 말하기 과제 수행 간에 보통 수준의 상관이 나타났으며, 이는 기존의 교육 측정 및 언어 평가 이론의 지지와 함께 대학 강의 맥락과 관련하여 실제적인 학문 목적 말하기 능력을 평가하는 도구로서 코스탭이 기능할 수 있음을 보여준 것이다. 또한 타당화 과정에서 확인한 코스탭 본 검사 결과, 기준 설정 결과에 따른 수학 가능성 평정과의 유사성 증거 등은 향후 코스탭의 유용성과 활용 가능성을 보여준다.

핵심어 학문 목적 말하기 평가, 타당도, 논증 기반 타당화, 외삽 추론, 실험 연구

ABSTRACT

Argument-based Validation for the Korean Speaking Test for Academic Purposes
— Examining the Extrapolation Inference

Lee Sunjun · Hong Eunsil · Oh Yelim ·
Lee Seungwon · Cho Sujin · Min Byeonggon

This article investigates the extrapolation inference for the Korean Speaking Test for Academic Purposes(KoSTAP) in relation to providing information on the performance of international students in academic speaking tasks in practical contexts. The Argument schemes of extrapolation inference and validation study were designed based on the argument-based approach. Evidence of extrapolation inference was collected by comparing the KoSTAP scores with learners' performance in the real-world target domain. International students who participated in this study, listened to a lecture in a simulated class and performed related speaking tasks. The extrapolation inference was verified by analyzing the correlation between the result of the speaking task evaluation and the KoSTAP score. This study shows that KoSTAP functions as a tool to evaluate the practical academic speaking skills required by universities in relation to the lecture context. It also presents the future usefulness of KoSTAP.

KEYWORDS Speaking Test for Academic Purposes, Validity, Argument-based Validation, Extrapolation Inference, Experimental Research