

영어 듣기 평가 모형의 구인 타당도 연구

장재학

강원대학교

Chang, Jaehak. (2012). The construct validity of models of L2 listening ability. Modern English Education, 13(2), 207-229.

This study examines the adequacy of models of second language (L2) listening ability in order to better understand what its components are and the relationships among these components. The focus of the investigation was placed on two sets of the listening tests that are administered nationwide: The Nationwide Listening Comprehension Test (NLCT) and the College Scholarly Aptitude Test (CSAT). Data were collected from 97 high school students learning English in Korea. After a series of confirmatory factor analyses, the present study found a model of L2 listening ability that was supported by the data. According to the final model, L2 listening ability involves L2 learners' factual, inferential and global understandings of given information. The results also suggest that L2 listening ability is better examined when the listening construct is measured in relation to test items measuring ability to choose appropriate responses in communicative situations. Lastly, detailed findings, implications and suggestions for further research will be discussed.

[listening ability/construct validity/listening tests/confirmatory factor analysis
듣기능력/구인타당도/듣기시험/확인적 요인분석]

I. 서론

의사소통 활동에 학습자의 듣기 능력의 중요성이 강조되고 있다. 원어민의 경우, 일반적인 의사소통 중에서 듣기는 읽기나 말하기의 두 배, 그리고 쓰기의 다섯 배의 비중을 차지한다고 한다(Rivers, 1981). 이러한 의사소통 활동 중의 듣기의 중요성에 따라 영어 평가 분야에서는 듣기의 비중이 점차 증가되고 있는 추세이다. 예를 들어, 국내 최초로 전국적인 규모의 영어 듣기 능력 평가(이하 전국듣기평가)가 처음으로 실시된 것은 1982년이였다. 당시에는 중·고등학교 2학년만을 대상으로 실시 되었으나 대학 입시나 내신 등에 성적이 반영되지 않아 학습자의 관심을 끌지 못하였다고 한다(김일홍, 2006). 그 후

현재까지 매년 실시되고 있다. 2011년까지는 전국 15개 시·도 교육청이 문제 출제를 주관하였으나 2012년 올해부터는 대학 수학 능력 시험을 주관하는 한국교육과정평가원(KICE; 이하 평가원)에 문제 출제가 이관된다. 더 나아가 전국듣기평가 시행일에서는 한국교육방송공사(EBS)에서 듣기 시험이 라디오를 통해 송출되며 대부분의 학교에서 수행평가 성적에 반영되어 교사와 학습자들의 듣기평가에 대한 관심이 더욱 고조되고 있다.

듣기 평가나 듣기 교육의 위상을 한층 더 높인 것은 1994년에 치러진 대학 수학 능력 시험(이하 수능)이라 할 수 있다. 1994년에 처음으로 수능 시험에 듣기 평가(이하 수능듣기평가)가 실시된 이후 2011년까지 17회에 걸쳐 실시되고 있다. 1997년에는 순수 듣기 평가 문항 외에도 듣기와 연계한(소위 말하는 간접적) 말하기 평가 문항이 도입되기도 했다. 이는 듣기 기능이 말하기와 같은 표현 능력에의 전이 효과를 염두에 둔 것으로 보인다. 듣기 평가 문항 수도 점차로 확대되어 수능 시행 첫해에는 듣기 평가 문항이 8문항이었으나 2012년 현재 듣기 문항(간접 말하기 문항 포함)이 총 17문항이 출제되고 있다.

듣기 평가의 위상은 2012년 올해 시행될 국가 영어 능력 평가(NEAT; 이하 NEAT)에서 한 단계 더 높아질 전망이다. 수능 시험에서(간접 쓰기 문항을 포함한) 읽기 문항이 총 33문항이라는 점과 비교하면 수능듣기평가의 위상이 높아졌다고는 하나 읽기 평가에 비하여 절반 정도의 비율로서 여전히 상대적으로 낮은 수준이라고 할 수 있다. 반면에, 평가원의 김용명(2010)에 의하면 앞으로 치러질 NEAT에서는 듣기 평가와 읽기 평가 모두 같은 수의 문항으로 설계되어 있다. 2년 전의 자료라 시험의 실제 시행 시에는 달라질 수 있겠지만, 그가 제시한 NEAT의 구성은 2,3급 기준으로 듣기 평가 35문항과 읽기 평가 35문항으로 제시 되어 있다.

위에서 언급한 데로, 영어교육분야에서 의사소통 능력 중 듣기에 관한 위상은 점차 높아져 왔으나 듣기 관련 연구의 위상은 크게 나아지지 않은 듯이 보인다. 한 예로, 이명관, 박옥희(2010)는 *현대영어교육* 학술지 창간 후 10년(2000-2009년) 동안의 의사소통 능력 향상을 위한 듣기와 말하기 관련 논문을 분석하였는데, 총 241편의 논문 중에서 듣기와 말하기에 관한 논문은 24편으로 전체 논문에서 차지하는 비중이 10% 정도에 불과함을 밝힌다. 이들은 또한 듣기 논문이 전체 논문에서 차지하는 비중은 2.7%에 해당되며, 듣기와 말하기 모두에 해당되는 논문은 0.36%에 불과하다고 하였다. 따라서 듣기 관련 연구가 그 어느 때보다 더 절실히 요구 된다고 하겠다.

반면에, 올해 2012년은 영어 능력 평가 전반이나 듣기 능력 및 듣기 평가 관련 논문이 많이 쏟아질 것으로 기대된다. 첫째 이유는, 2012년은 평가원이 주관하여 전국듣기평가가 처음으로 실시되는 해이기 때문에 그 이전에 전국 15개 시·도 교육청 주관 시험과의 유사점과 차이점에 대한 논의가 있을 것으로 기대가 된다. 둘째로, 올해는 NEAT가 처음으로 일반 시행되는 해이기도 하다. 듣기

평가 관점에서는, 전국적으로 시행되는 듣기 평가의 비중이 읽기 평가와 동등하게 치러지는 원년이기도 하다. 듣기 문항 수가 늘어남에 따라 수능듣기평가와의 비교가 자연스럽게 이루어 질 것으로 보인다. 게다가 올해는 NEAT 시험에 대한 여론 수렴 결과에 따라 NEAT가 수능 시험의 대체 여부도 결정되는 해이고 보면(김용명, 2010), 수능 시험과 NEAT 시험의 각 영역에 대한 학문적인 논의는 더 활발해질 것으로 보인다. 이러한 시점에서, 본 연구는 수능듣기평가와 전국듣기평가에서 사용되는 듣기 평가 모형의 구인 타당도(construct validity)를 조사하여 평가원 주관으로 올해 처음으로 시행되는 전국듣기평가와 NEAT의 듣기 평가의 설계 및 문항 출제에 대한 이론적, 교육적, 평가론 적인 시사점을 알아보하고자 한다.

II. 문헌연구

1. 이론적 배경

수능듣기평가와 전국듣기평가는 고등학교 교육과정 내용의 성취 정도를 평가하는 성취도 검사의 성격과 영어능력을 전반적으로 평가하는 능숙도 검사의 성격을 모두 갖추고 있다. 따라서 두 듣기 평가 시험의 타당성을 검토하기 위해서는 먼저 현대 영어능력이론의 근간이 되고 있는 의사소통능력이론과 고등학교 영어과 교육 과정의 내용 구성을 함께 살펴볼 필요가 있다. 또한 듣기 능력은 영어능력의 하나의 주요 구성 요소 이므로 듣기 기능과 관련한 이론도 같이 살펴볼 필요가 있다.

1980년대 이후 의사소통능력에 관한 이론 여러 이론들이 제안 되었고(Bachman, 1990; Canale, 1983; Canale & Swain, 1980) 여러 학자들에 의하여 수정 및 보완되었다. 그 중 대표적인 이론은 Bachman과 Palmer(1996)에 의해 제안된 언어능력이론이다. 이들에 따르면, 언어능력은 언어지식과 전략적 능력으로 이루어 졌으며, 언어지식은 조직적 지식과 화용론적 지식으로 나누어진다. 이들이 말하는 조직적 지식은 다시 문법적 지식과 담화적 지식으로 구분되며, 각각은 Canale과 Swain(1980)의 이론 중 문법적 지식과 담화적 지식의 내용과 아주 유사하다. 반면에, Bachman과 Palmer(1996)는 화용론적 지식을 사회언어학적 지식과 기능적 지식으로 분류하여 설명하고 있다. 사회언어학적 지식은 화자의 언어 사용력(register)에 대한 지식, 자연적인 표현, 관용 표현 및 비유적 표현의 의미에 관한 지식을 포함하고 있다. 여기서 사용력에 관한 지식은 화자들 간의 관계와 의사소통 상황에 따라 격식을 차리는 표현과 그렇지 않은 표현을 적절하게 이해하고 사용할 수 있는 지식을 말한다. 기능적인 지식은 언어의 다양한 의사소통 기능에 관한 포괄적인 지식을 말하며, 크게 언어 사용의 정보적 기능, 도구적 기능, 발견적 기능 및 상상적 기능에 대한 이해를 말한다. 끝으로, 이들

은 이러한 언어 지식과 함께 성공적인 의사소통을 위해서는 인지적 및 초인지적 전략을 효율적으로 사용할 수 있는 능력을 언어능력으로 보고 있다.

요약하면, 현대에서 말하는 의사소통 능력의 핵심은 언어에 관한 고른 지식이라고 할 수 있다. 언어 지식은 크게 언어의 형태에 관한 지식, 글자적 의미에 관한 지식, 사회적 사용에 관한 지식 및 의사소통 기능에 관한 지식을 고르게 갖추고 있는 것으로 정의할 수 있겠다. 언어의 형태와 의미에 관한 지식이 문법적인 지식과 관련이 많다면, 언어의 사용과 기능에 관한 지식은 화용론적인 지식과 연관이 많다고 할 수 있다. 이러한 의사소통능력에 대한 이해는 듣기 능력 평가가 언급된 두 가지 영역의 언어 지식을 모두 측정하는 문항으로 이루어져야 함을 시사한다고 할 수 있다.

의사소통능력의 일부인 듣기 능력은 의사소통과 관련된 언어 지식에 대한 이해와 연관되었을 뿐만 아니라 이해 기능의 일부로서 듣기 기능과 보다 밀접한 연관이 있다. Weir(1993)는 듣기 기능을 직접적, 추론적 및 공헌적 의미 이해로 구분하여 설명한다. 그가 말하는 공헌적 의미 이해의 주요 요소는 음운론적 자질, 문법적·어휘적 응집성 및 담화 표지어 등의 이해를 포괄하는 개념이다. 이는 듣기 능력이 언어 지식에 대한 이해와 연관이 있음을 보여주는 것이다. 직접적 의미 이해는 주제, 요지, 중요 정보에 대한 이해, 특정 정보 및 세부 정보에 대한 이해 등을 포괄하는 이해 기능을 말한다. 끝으로, 추론적 의미 이해는 어휘, 사회적 상황, 의사소통기능 등에 대한 추론과 일반적인 내용이해를 위한 추론을 포괄하는 기능을 말한다. Buck(2001) 또한 텍스트에 담겨 있는 주요 언어적 정보에 대한 이해 능력과 대화나 담화의 내용을 추론하는 능력이 듣기 능력의 가장 기본이 되는 이해 기능이라고 설명한다. 이들 학자들은 서로 표현은 약간 다르지만, 듣기 기능의 가장 기본적인 구성 요소는 대화나 담화 속의 사실적 정보를 이해하는 능력과 대화나 담화의 내용을 추론하는 능력으로 보고 있음을 알 수 있다.

듣기 기능은 사실적 이해 및 추론적 이해의 기본적 기능 이외에 대화나 담화의 분위기, 목적, 어조, 태도, 느낌 등을 파악하는 능력으로도 설명이 된다 (Brown, 2004; Buck, 2001; Mendelsohn, 1994; Weir, 1993). Mendelsohn(1994)은 대화나 담화의 상황 파악, 화자의 기분을 파악하는 것도 듣기의 주요 기능들 중 하나로 설명한다. Brown(2004)은 듣기 기능을 수십 가지의 하위 기능으로 나누어 설명하고 있는 데, 이들 중에 대화나 담화의 목적 및 화자의 태도를 파악하는 능력을 듣기의 주요 기능 중의 하나로 설명한다. 한국교육과정평가원(2012)은 더 나아가 대화나 담화의 분위기, 목적, 어조, 태도, 느낌 파악 능력을 듣기 능력의 주요 구성 요소로 보고 있으며, 이들 이해 기능을 종합적 이해라고 부르고 있다. 이상의 논의를 종합하면, 듣기 기능은 사실적 이해, 추론적 이해, 종합적 이해를 모두 포괄하는 개념으로 이해할 수 있겠다.

제 7차 고등학교 교육과정 해설(교육인적자원부, 2001)을 보면 학년 또는 단계별 듣기 성취 수준을 위에 언급한 세 가지 주요 듣기 능력의 세부

기능들로 기술하고 있다. 교육과정 해설서에 나타난 듣기 주요 기능들은 나열하면 다음과 같다: (1) 일반적인 주제에 관한 덜 친숙한 내용의 말을 듣고 대의, 주제, 요지 및 세부 사항을 이해한다; (2) 다양한 내용의 말을 듣고, 명시적, 함축적 의미를 이해하며, 대의 등을 추론하여 파악한다; (3) 다양한 내용의 말을 듣고, 화자나 글쓴이의 의도와 태도를 추론하고 그 근거를 찾는다; (4) 일반적인 주제에 관한 말을 듣고, 종합적으로 평가하며 이해한다; (5) 일반적인 주제에 관한 말을 듣고, 화자의 주관적 견해와 객관적 사실을 구별한다; (6) 일반적인 주제에 관한 덜 친숙한 내용의 말을 듣고, 전개방식, 원인과 결과, 결론과 근거 등을 이해한다.

한국교육과정평가원(2012)에 따르면, 수능 외국어 영역의 평가 목표는 영어로 의사소통을 하는 데 필요한 어휘, 문법성 판단력, 사실적 이해력, 추론적 이해력, 종합적 이해력 및 적용력을 측정하는 것이라고 되어있다. 이 중에서 특히 후자의 4가지 요소는 듣기 평가와 보다 직접적으로 관련이 되어있으며, 지금까지 논의된 학자들의 듣기 기능에 대한 이론과 교육과정에서 제시된 듣기 성취 수준과도 일치한다. 평가원에 의하면, 사실적 이해력은 대화나 담화에 진술된 핵심 내용을 신속 정확하게 파악하는 능력을 말하며, 추론적 이해력은 대화나 담화에 직접적으로 명시되지 않은 사항을 논리적으로 추론해 낼 수 있는 능력을 뜻한다. 구체적으로, 지칭 추론, 의도적으로 삭제한 표현의 추론, 요지, 제목, 주장 등의 추론 능력을 측정한다. 반면에, 종합적 이해력은 대화나 담화의 내용 등을 개별적 언어요소에 대한 분석이 아니라 종합적으로 파악하여 문제를 해결할 수 있는 능력을 말하는 데, 대화나 담화의 분위기, 목적, 어조, 태도, 느낌 파악하기 등의 이해력을 측정한다고 한다. 끝으로, 적용력은 대화나 담화의 내용을 이해하며 의사소통 상황에 알맞게 적용할 수 있는 능력을 말하는데, 특히 듣기 자료를 통해 이해한 바를 말하기 상황에 적용할 수 있는 언어 사용 능력을 간접적으로 측정한다고 되어있다. 이상에서 논의 된 내용을 정리하면, 듣기 능력이란 한마디로 문법적, 화용론적 지식과 같은 언어 지식에 기반하여 대화나 담화의 내용을 사실적, 추론적, 종합적으로 이해하는 능력이라 할 수 있으며, 이들 각각의 이해 능력은 또한 이들을 구성하고 있는 여러 가지 다양한 하위 구성요소로 이루어 졌다는 것도 알 수 있었다.

2. 선행연구

듣기 능력에 관한 평가 모형의 타당도에 관한 연구나 듣기 능력 모형의 타당도에 관한 연구는 거의 이루어 지지 않았다. 드물게 행해진 연구 중에서, Buck(1992)의 듣기 구인에 관한 연구는 주목할 만하다. 그는 일본 오사카 소재 대학생 353명을 대상으로 하여, L2 언어 학습자들의 영어능력 중에서 이해 기능인 듣기와 읽기 기능이 과연 서로 구별될 수 있는 독립적인 구인인지를 확인적 요인 분석으로 조사였다. 그는 듣기 기능과 읽기 기능은 둘 다 이해

기능이기는 하나 서로 구분될 수 있는 언어 기능임을 밝힌바 있다. 그러나 그의 연구는 듣기 능력 구인이 어떤 구성 요소로 이루어졌는지를 알아보는 것이 그 주된 목적은 아니었으므로 듣기 능력 구인의 구성요소에 관한 후속 연구의 필요성을 제기한 바 있다.

반면에, 최근 국내에서 이루어진 주목할 만한 연구 3개를 소개하면 다음과 같다. 배태일(2009)은 총 285명의 대학생을 대상으로 이들의 영어 듣기 청취 모형의 적합성을 조사하였다. 구체적으로, 그는 영어 듣기 청취에 영향을 미치는 여러 요소들의 직접적, 간접적 관계를 구조방정식 모형을 이용하여 조사하였다. 그는 의사소통능력의 일부인 문법이나 어휘 등의 언어지식이 학습자의 영어 청취력에 통계적으로 유의미한 긍정적인 영향을 미친다는 것을 밝히고 있다. 그는 또한 영어 불안감은 학습자의 영어 청취력에 부정적인 영향을 미치지만, 듣기 전략은 학습자의 영어 청취력에 직접적인 영향은 미미하지만 영어불안감을 통해 간접적인 영향을 미치고 있음을 밝히고 있다. 이 연구는 방법론적 우수성을 바탕으로 듣기 청취 모형의 타당도 조사를 시도하였다는 데 의의가 크다고 할 수 있다. 그러나, 이 연구는 듣기 능력 모형 자체의 구인 타당도를 조사하지 않은 아쉬움이 있다.

S. K. Shin(2006)은 총8명의 학습자를 대상으로 특정한 듣기 평가 문항 유형의 타당도를 조사하였다. 그는 일종의 표준화 시험인 MELAB(Michigan English Language Assessment Battery)의 듣기 영역에서 사용되는 강조 문항 20개를 연구에 사용하였다. 이 문항들은 학습자들의 운율적 청취능력(prosodic competence)을 측정하고자 고안된 것으로 사고과정 분석법(verbal protocol analysis)을 이용하여 문항의 타당성을 분석하였다. 분석 결과, MELAB의 듣기 강조 문항들은 어느 정도 타당성을 갖추었다는 것을 밝히고 있다. 이 연구는 듣기 문항 관련 타당성 연구라는 점에서 의의가 있으나, 듣기 능력의 작은 일부에만 초점을 맞추고 있어 전반적인 듣기 능력 모형의 적합성을 알 수 없는 아쉬움이 있다.

정행(2010)은 수능듣기평가의 내용 타당도를 조사한 의미 있는 연구이다. 그는 1994년부터 2010년까지의 총 17회의 수능시험의 듣기/말하기 평가 문항을 연구 대상으로 하였다. 구체적으로, 그는 수능 음성 언어 평가 영역의 내용을 언어기능, 의사소통기능, 소재, 언어자료유형, 평가 유형 별로 분류하여, 문항 내용과 교육과정과의 연관성 정도를 조사하였다. 조사 결과, 그는 수능 음성언어(듣기/말하기) 문항의 평가 내용이 고등학교 교육과정과의 상관성이 높으며, 일부 수정 제안 사항은 있으나 전반적으로 타당성이 있다고 보고하고 있다. 그는 나아가서 수능 시험에 관한 구인 타당도 연구의 필요성도 주장하고 있다.

이상에서 살펴보았듯이, 국가적으로 시행되는 고부담 시험인 수능듣기평가와 같은 시험의 구인 타당도를 조사한 연구는 거의 이루어 지지 않고 있다. 더군다나 수능시험에 비해 상대적으로 덜 주목 받고 있는 전국듣기평가의 타당도에 관한 연구 역시 거의 전무한 실정이다. 따라서 본 연구는 수능듣기평가 2회와 전국듣기 평가 2회를 이용하여 이들 영어듣기 평가 모형의 구인 타당도를 조사

하고자 한다. 본 연구의 연구질문은 다음과 같다:

1. 수능듣기평가와 전국듣기평가를 통해 알아본 영어듣기 평가 모형의 구인 타당도는 어떠한가?
2. 수능듣기평가와 전국듣기평가를 통해 알아본 영어듣기 능력의 주요 구성 요소는 무엇인가?

III. 연구방법

1. 연구참여자

본 연구는 강원도 원주 소재 한 고등학교 3학년 97명의 학생을 대상으로 이루어졌다. 이 지역은 고교 비평준화지역으로 학교에 따라 학습자의 영어 수준에 차이가 있다. 본 연구에 참여한 학습자들은 지역 기준으로는 상위권에 해당하며, 수능 영어 내신 성적 기준으로는 전국의 중상위권에 해당한다.

2. 연구도구

본 연구는 듣기 평가 모형의 구인 타당도를 알아보기 위하여 전국 시·도 교육청 주관 영어듣기평가 2회분과 대학수학능력시험 영어듣기평가 2회분을 사용하였다. 전국듣기평가는 2009년 2회와 2010년 2회를 사용하였고, 수능듣기평가는 2010학년도와 2011학년도 문항을 사용하였다. 전국듣기평가는 수능듣기평가보다 3문항이 더 많은 총 20문항으로 이루어졌으나, 측정능력이나 구성요소 등이 수능과 아주 유사하게 설계되었다고 보여지므로 수능듣기평가의 설계에 맞추어 문항을 분류하여 보았다. 본 연구에 사용된 듣기 능력 평가의 이론적 설계와 문항유형은 표 1에 제시되어 있다. 표에서 보듯이, 수능듣기평가와 전국듣기평가의 시험 설계는 전반적으로 유사함을 알 수 있다. 다만, 수능듣기평가가 상당히 정형화되어 있는 반면, 전국듣기평가는 출제연도에 따라 평가지 문항 구성이 조금씩 달라지는 특성이 있다. 끝으로 표에 제시된 각 측정 영역 별 문항 예제는 부록에 제시되어 있다.

표 1
듣기 평가의 구성

측정 능력	문항유형	문항수	평가지 별 문항번호				
			전국1	전국2	수능1	수능2	
사실적 이해	묘사대상	4	1	1	1	1	
	일치/불일치	7	6, 13, 14	7, 15	12	12	
	할일/부탁한일	9	7, 11	8	4, 7, 10	4, 7, 10	
듣기	특정정보	5	15	9, 14	11	11	
	관계추론	4	2	3	9	9	
	추론적 이해	주제/요지/화제	7	8, 12	10, 12, 13	3	3
		장소추론	3	-	2	8	8
종합적 이해	정보추론	4	4	5	5	5	
	심경 파악	4	3	4	2	2	
	목적 파악	5	5, 10	6	6	6	
적용	그림에 적절한 대화	4	9	11	13	13	
	대화응답	14	16-19	16-19	14-16	14-16	
	담화응답	4	20	20	17	17	

전국1=2010학년도 전국듣기평가 2회; 전국2=2009학년도 전국듣기평가 2회;
수능1=2010학년도 수능듣기평가; 수능2=2011학년도 수능듣기평가

3. 자료 수집 및 절차

수능듣기평가와 전국영어듣기평가는 4차례에 걸쳐 시청각 시설이 잘 갖추어진 교실에서 약 20분간 동일 조건하에 실시하였다. 4차례의 시험 중 한 차례라도 시험에 응시하지 못한 학습자는 연구 대상에서 제외되었다. 시험 실시 후, 평가 문항은 정오표에 의하여 채점하였으며, 연구 목적 상 맞으면 1점 틀리면 0점을 부과하였다.

4. 자료분석

채점이 끝난 자료는 일련의 통계 분석을 실시하였다. 듣기 평가 영역별의 평균과 표준 편차를 알아 보기 위하여 기술 통계를 실시하였다. 시험의 신뢰도를 알아보기 위하여 신뢰도 분석도 실시하였으며, Cronbach의 알파를 사용하였다. 변인들의 관계를 알아보기 위하여 상관관계 분석도 실시하였다.

끝으로, 듣기 평가 모형과 이론 모형의 구인 타당도를 확인하기 위해서 확인적 요인분석(confirmatory factor analysis)을 실시하였다. 기술 통계, 상관 관계 분석, 신뢰도 분석은 IBM SPSS Statistics 20.0을 이용하였으며, 구조방정식 모형(structural equation modeling)의 일종인 확인적 요인 분석을 위해서는 EQS 6.0을 사용하였다.

IV. 결과 및 논의

1. 기술 통계

표 2는 듣기 평가의 기술 통계, 문항 유형간의 상관 관계 계수를 나타낸다.

표 2
듣기 평가 기술 통계 및 문항 유형 간의 상관 관계

	사실적 이해			추론적 이해				종합적 이해		적용			
	1	2	3	4	5	6	7	8	9	10	11	12	13
상관관계													
1	1												
2	.10	1											
3	.16	.35	1										
4	.04	.29	.32	1									
5	.12	.10	.22	.11	1								
6	-.09	.37	.27	.32	.13	1							
7	-.02	.15	.22	.07	.29	.18	1						
8	.05	.34	.37	.17	.11	.28	.35	1					
9	.04	.26	.41	-.06	-.11	.23	.16	.19	1				
10	.11	.42	.43	.17	.27	.18	.24	.40	.20	1			
11	.19	.03	.17	.15	-.08	.13	.09	.11	.09	-.05	1		
12	.16	.43	.52	.24	.27	.41	.16	.31	.11	.39	.16	1	
13	.03	.31	.30	.11	.22	.31	.12	.19	.29	.27	.15	.42	1
기술통계													
k=	4	7	9	5	4	7	3	4	4	5	4	14	4
M	3.58	4.58	7.15	2.82	2.8	5.12	2.23	2.18	3.76	3.52	3.75	9.76	2.4
SD	.57	1.41	1.44	1.22	.74	1.25	.60	1.08	.57	1.15	.46	2.48	1.04
%	89.5	65.4	79.0	56.4	70.0	73.0	74.0	54.5	94.0	70.4	93.8	69.7	60.0

1. 묘사대상 2. 일치/불일치 3. 할일/부탁한 일 4. 특정정보 5. 관계추론 6. 주제/요지/화제 7. 장소추론
8. 정보추론 9. 심경파악 10. 목적파악 11. 그림에 적절한 대화 12. 대화응답 13. 담화 응답
k= 문항 수; M= 평균; SD= 표준편차; %=100점 환산 점수

표에서 보듯이, 듣기 평가의 전체 평균은 74점 만점에 53.66이었고 표준편차는 8.06이었다. 최저점은 33점이며 최고점은 72점이었다. 100점 만점으로 환산한 평균은 72.5점이며, 대부분의 학습자의 점수는 51점에서 94점 사이의 분포를 보였다. 전체적으로, 4번의 듣기평가의 평균적인 난이도는 본 연구 참여자들에게는 약간 쉬웠다고 할 수 있다. 학습자들이 가장 쉽게 느낀 문항은 심경 파악이었다. 이들은 4점 만점에 3.76점을 받았고 100점 만점으로 환산하면 94점이 된다. 그 다음으로 쉬운 문항은 4점 만점에 3.75점을 받은 그림 상황에 알맞은 대화 고르기 유형이었다. 100점 만점에 93.8점에 해당한다. 그 다음으로 쉬운 문항은 4점 만점에 3.58점을 받은 그림 보고 묘사 대상 고르기 유형이다. 100점 만점에 89.5점에 해당한다. 반면에 학습자들이 가장 어렵게 느낀 문항은 4점 만점에 2.18점을 받은 정보추론(가격, 요금 등) 유형이다. 100점 만점에 54.5점에 해당한다. 그 다음으로 어렵게 느낀 문항은 5점 만점에 2.82점을 받은 특정정보 찾기 유형이다. 100점 만점에 56.4점에 해당한다. 학습자들은 일치/불일치 문항에 어려움을 느꼈다. 학습자들은 7점 만점에 4.58점을 받았으며 100점 만점 환산 점수는 65.4점에 해당한다. 끝으로, 표에 제시된 문항 유형간의 상관관계를 보면 상관관계계수가 $-.09$ 에서 $.52$ 까지 다양하게 분포 되어있다. 일부 문항 유형간의 상관관계가 음의 수치거나 0에 가까운 것으로 보아 이들 일부 문항은 듣기 구인의 척도로서 적합하지 못하거나 신뢰하기 어려움을 시사한다.

2. 평가의 신뢰도

듣기 평가의 전체 신뢰도는 $.829$ 이었다. 이는 듣기 평가 결과는 전반적으로 신뢰할 만한 수준이라고 할 수 있지만 공신력 있는 국가 기관이 주관하는 전국 규모의 시험이라는 점과 학습자의 진로나 내신에 큰 영향을 미치는 고부담 시험(high-stake exams)이라는 것을 고려하면 그리 높은 신뢰도라고 보긴 어렵다. 상대적으로 낮은 신뢰도를 보고 판단한다면, 평가 결과는 전체적으로는 신뢰할 만한 수준이나 평가의 일부 문항 유형은 개선이 필요함을 보여준다.

3. 듣기 평가 모형의 타당도 분석

위에서 언급한 듣기 평가의 신뢰도를 바탕으로 듣기 능력 모형의 구인 타당도를 조사하였다. 문헌 연구에서 언급한 듣기 능력의 구성 요소와 평가원이 제시한 수능 듣기 평가의 목표를 통해 추정할 수 있는 듣기 능력의 이론적 개념을 그림으로 도식화하여 표현하면 그림 1과 같다.

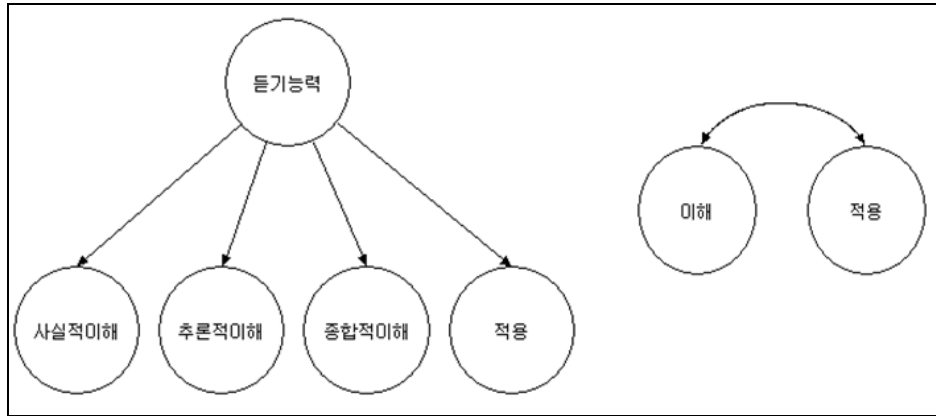


그림 1 듣기 능력의 이론적 모형

왼쪽의 도식은 듣기 능력이 사실적 이해력, 추론적 이해력, 종합적 이해력 및 적용력의 4가지 구성 요소로 이루어졌음을 시각적으로 보여주고 있다. 각각의 구성 요소가 원으로 표시되었는데, 여기서 원은 이론적 개념 또는 구인을 상징적으로 나타낸다. 반면, 오른쪽 도식은 왼쪽 도식에 담긴 듣기 능력에 대한 개념을 보다 단순하게 나타낸 것이다. 즉, 듣기 능력은 이해력과 적용력의 최소 두 가지의 연관된 구인으로 설명할 수 있다고 보는 것이다. 여기서 물론 이해력은 사실적, 추론적, 종합적 이해력을 모두 포함하는 개념이다. 본 연구는 바로 그림 1에서 제시된 듣기 능력의 이론적 모형의 타당성을 알아보는 것이 그 주된 목적이라 할 수 있다. 또 다른 목적은 이들 듣기 능력 구인들을 효과적으로 측정하고 있는 문항 유형과 그렇지 못한 문항 유형을 알아봄으로써 NEAT와 같은 국가 규모의 고부담 시험의 평가 문항 유형 개발 및 문항 출제에 대한 시사점도 함께 알아보는 것이다.

듣기 능력의 이론적 모형의 타당성을 알아보기 위해선 측정 모형(measurement model)과 측정변인(measured variables)의 신뢰성이 우선 전제 되어야 한다. 예를 들어, 이론적 모형의 구인 타당도 조사 결과가 좋지 않게 나오는 경우, 복합적인 원인으로 인해 해석상의 어려움이 발생한다. 이론적 구인을 측정하는 데 사용된 평가도구의 신뢰성이 좋지 못하면 측정 오류로 인하여 이론적 모형의 타당도는 나쁠 수 밖에 없게 된다. 이 경우에는 듣기 능력을 개념적으로 옳게 가정했다고 하더라도 타당도의 결과는 좋지 않게 나오게 된다. 다시 말하면, 타당도 조사 결과는 수치상으로 좋지 않지만 개념적으로 듣기 능력의 구인들과 그들의 관계를 올바르게 가정하고 있을 수 있다는 뜻이다. 이론적 모형의 타당도가 좋지 못한 또 하나의 이유는 듣기 능력에 대한 불완전한 이해에 기인할 수 있다. 이론적으로 듣기 구인과 이들의 관계를 잘 못 이해한 경우, 잘못된 가정에 기초한 분석이므로 평가도구의 신뢰성이 확보된 경우라고 하더라도 구인 타당도 결과는 좋지 않게 나오게 된다. 이때는 이론을 수정 보완할 필

요가 있다. 끝으로, 이론적 모형의 타당도가 좋지 못한 이유는 평가도구의 신뢰성도 좋지 않고 이론적 가정도 잘 못한 경우이다. 이 경우도 첫 번째 경우와 같이 평가도구의 신뢰성 자체가 확보되지 못한 경우이므로 이론적 모형의 적절성을 확인할 길이 없어진다. 즉, 듣기 능력의 이론적 모형의 타당성을 확인할 수 있는 경우는 이들 구인들을 측정하기 위해 사용한 측정변인들의 신뢰성을 확인할 수 있는 경우 밖에 없다는 것이다. 쉽게 말하면, 무엇을 측정하든, 정확하게 측정해야 의미 있는 해석을 할 수 있다는 것이다.

위에 언급된 이유로, 본 연구는 듣기 능력 평가의 측정 모형 타당도 조사를 먼저 하였고, 그 결과는 그림 2에 도식과 수치로 제시되어 있다. 그림에서 보면, 원으로 표시된 것은 측정하고자 한 듣기 능력의 이론적 구인을 뜻하며 네모로 표시된 것은 이들 구인을 측정하기 위한 평가 문항 유형 또는 측정변인을 뜻한다. 이들 측정변인을 구인들과 혼동하지 말아야 한다. 이들 평가 문항 유형들은 측정변인이므로 측정 오류에서 자유롭지 못하다. 예를 들어, 세부 정보 파악 능력을 측정하기 위한 문항들이 실제로는 의도하지 않은 추론적 이해 능력을 측정할 수도 있다. 의도 하지 않은 경우이므로 측정 오류라고 할 수 있다. 또한 문항의 난이도가 너무 쉬운 나머지 거의 모든 학습자가 맞춘 경우도 문항의 변별력이 아주 낮거나 심지어는 역변별력을 보이는 경우도 있다. 이러한 측정 오류도 평가 모형의 전체적 신뢰도를 낮추며 이러한 문항 유형들이 많이 있으면 측정 모형의 타당도 역시 낮아지게 된다. 그림 2를 자세히 보면 듣기 평가에 사용된 모든 평가 문항 유형들이 측정변인으로 포함되고 있음도 알 수 있다. 이들 모든 측정변인이 포함된 모형을 편의상 듣기 평가 기준 모형이라고 부르겠다.

그림에서 보듯이 듣기 기준 모형의 구인 타당도를 나타내는 지수는 기준치에 상당부분 미치지 못하고 있다. Hu와 Bentler(1999)에 의하면 평가모형의 적합성 여부를 알려주는 권고 수치는 .95이다. CFI 지수가 .95라는 뜻은 검증 대상인 평가 모형이 그 모형을 구성하고 있는 변인들의 관계를 약 95%정도 설명을 하고 있다는 것을 의미한다. 반면에 모형의 부적합 정도를 나타내는 지수인 RMSEA 수치는 .05이하로 낮아야 한다. 본 연구에서는 듣기 평가 모형의 적합도를 나타내는 수치인 CFI = .876이었으며 NNFI = .849이었다. 반면에 모형의 부적합 정도를 나타내는 지수인 RMSEA = .064이었다.

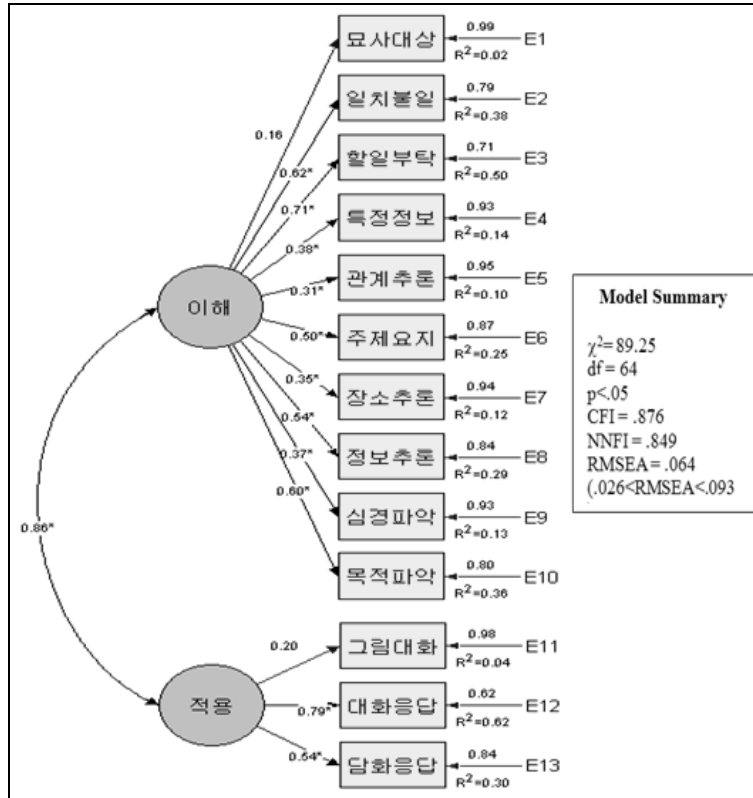


그림 2 듣기 기준 측정 모형의 구인타당도

듣기 평가 측정 모형의 구인타당도가 좋지 못한 원인을 알아보기 위하여 개별 지수의 적합성을 살펴보았다. 그림 2를 보면, 구인들과 측정변인들 간에 화살표로 이어져 있는 것을 볼 수 있다. 이들 각각의 화살표 옆에는 수치들이 표시되어 있는데 이것을 요인 적재량이라고 부른다. 이들 수치는 듣기 능력을 설명하는 각 측정변인들의 상대적 중요성을 나타낸다. 좀 더 구체적으로, 위의 기준 모형에서 요인 적재량의 수치가 가장 큰 몇 가지 측정변인은 활일/부탁한 일 파악하기(.71), 일치/불일치(.62), 목적 파악(.60) 및 정보추론(.54) 유형의 문항이다. 이들 문항은 듣기 능력의 구인을 효과적으로 측정하고 있다고 해석할 수 있다. 반면에 요인 적재량이 .3이하인 유형을 보면 그림 보고 묘사 대상 고르기(.16) 유형과 그림 상황에 맞는 대화 고르기(.20) 유형이 있다. 이들 두 유형의 평가 문항은 이들이 측정하고자 의도한 듣기 구인을 효과적으로 측정하고 있지 못하고 있기 때문에 듣기 구인의 척도로서의 기능을 잘 하고 있지 못하고 있음을 시사한다. 이들이 듣기 구인을 나타내는 훌륭한 척도라면 이들 측정변인들이 평가 모형에서 제외되면 모형의 적합도 수치는 훨씬 낮아지게 된다. 반대로 훌륭한 척도라면 이들을 모형에서 제거하면 모형의

적합도 수치는 올라가게 된다. 이러한 원리를 이용하여 주어진 데이터 안에서 측정 오류가 최소가 되는 듣기 평가 측정 모형을 찾기 위해서 일종의 신뢰도 분석 과정과 아주 유사한 일련의 분석을 행하였으며 그 결과는 그림 3에 제시되어 있다.

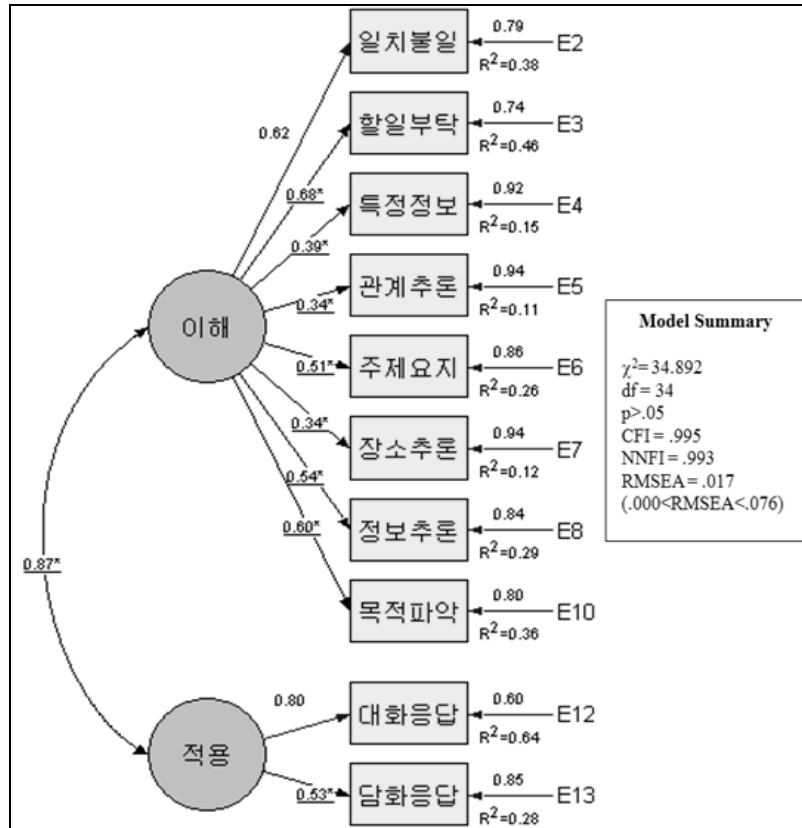


그림 3 듣기 최종 측정 모형의 구인타당도

그림에서 보듯이, 듣기 평가 최종 측정 모형의 적합성 여부를 알려주는 지수인 CFI = .995이었으며, NNFI = .993으로 모두 1.0에 가까웠다. 또한 모형의 부적합 정도를 나타내는 지수인 RMSEA는 .05보다 훨씬 낮은 .017을 나타내었다. 개별 지수의 적합성도 살펴 보았으며 모두 허용치 범위 안에 있었다. 다시 말하면, 듣기 평가의 최종 측정 모형에서는 측정 오류는 최소화될 뿐만 아니라 측정변인들간의 관측된 상관관계를 가장 잘 설명하고 있다라고 할 수 있다. 한마디로, 듣기 평가의 최종 측정 모형 속에 포함된 문항 유형들은 모두 신뢰할 수 있는 측정변인들이라고 할 수 있다.

최종 측정 모형은 기준 측정 모형과는 몇 가지 면에서 차이가 있다. 최종

모형에서는 기본 모형에 있던 ‘묘사대상’의 측정변인이 삭제되어 있다. 또 하나의 차이는 최종 모형에는 기본 모형에 있던 ‘그림대화’로 표기된 측정변인 역시 삭제되어 있다. 끝으로, 한 가지 더 주목할 점은 기본 모형에 있던 ‘심경파악’ 측정변인이 최종 모형에서 역시 제외된 점이다. 이들 세 측정변인들의 공통점은 다른 영역의 문항들에 비하여 문항의 난이도가 상당히 낮은 문항이라는 점이다. 쉽게 말하면, 아주 쉬운 문항들이다. 예를 들어, ‘심경파악’ 유형은 100점 만점 기준 평균이 94점이었고 ‘그림대화’ 영역 역시 평균이 94점에 가까웠다. 아마도 이들 영역 문항의 난이도가 너무 쉬워서 잘하는 학습자와 그렇지 못한 학습자의 영어능력을 변별하는 데 어려움이 있었던 것으로 보인다. 또한 거의 모든 학습자들이 맞춘 문항으로 이들의 듣기 능력을 추론하는 데 있어서 특별히 도움이 되는 정보를 주고 있지 못하다고도 볼 수 있겠다. ‘심경파악’ 영역은 또한 이론적으로 종합적 이해력을 측정하는 평가 유형으로 보았는데, 종합적 이해라 함은 대화나 담화의 내용의 사실적 이해력과 추론적 이해력 모두 요구하는 문항으로도 설명할 수 있겠다. 다시 말하면, 사실적 이해력을 묻는 문항들과 추론적 이해력을 묻는 문항들만으로도 학습자의 종합적 능력의 일부를 충분히 추론할 수 있다면 심경파악 유형의 문항은 듣기 능력을 추론하는 데 있어 중추적인 문항 유형이라기 보다는 일종의 잉여적 또는 중복의 성격을 가진 변인이 아닌가라는 추정을 해 볼 수 있겠다.

듣기 최종 측정 모형은 이론적으로 여러 가지를 시사한다. 첫째는 듣기 능력은 추론적 이해력, 사실적 이해력, 그리고 종합적 이해력을 모두 포괄하는 개념이라는 것이다. 그림 3에서 보면 8가지의 듣기 능력 지표 중에서 종합적 이해로 분류된 ‘목적 파악’을 제외한 7가지의 지표가 사실적 이해와 추론적 이해를 측정하는 문항 유형이라는 것을 알 수 있다. 또한 요인 적재량을 통해서 알 수 있듯이, 정보추론 능력과 주제/요지 추론 능력이 장소나 관계를 추론하는 능력에 비하여 추론적 이해력을 보다 잘 나타내는 주요한 지표라는 점을 알 수 있다. 반면에 할일/부탁한 일과 일치/불일치 문항 유형은 특정 정보 파악 유형 보다 사실적 이해력을 나타내는 보다 더 중추적인 지표라는 것도 알 수 있다. 끝으로, 학습자의 입장에서 보면, 수능 및 전국듣기평가 시험은 이해 능력을 측정하는 하나의 평가 영역과 적용력을 평가하는 또 다른 평가 영역이 서로 다르게 구분되어 학습자들에게 인식되고 있음을 유추할 수 있다.

이러한 이론적 시사점은 듣기 교육이나 평가에도 그대로 적용될 수 있다. 예를 들어, 듣기 지도 시에는 화자의 심정이나 태도, 분위기 파악 등에 시간을 많이 할애 하기 보다는 주제나 요지 파악, 세부 정보 파악 등에 보다 더 많은 시간을 할애하는 것이 학습자의 듣기 능력 향상에 더 도움이 될 수 있을 것이다. 듣기 평가 시에도 일치/불일치 유형이나 정보 추론, 주제, 요지 파악 및 목적 추론 등의 문항 유형은 듣기 시험에 반드시 포함 시켜야 듣기 평가의 유용성을 높일 수 있을 것이다. 이들은 듣기 능력의 주요 하위 기능들이므로

듣기 활동 시에도 이들 기능들을 중심으로 이루어지는 것이 바람직 할 것으로 보인다. 듣기 최종 측정 모형의 또 다른 이론적 시사점은 듣기 평가를 대화 내용에 적절한 응답 고르기와 같은 적용력을 측정하는 문항과의 연계를 통한 평가가 순수 이해 기능만을 평가하는 것보다 학습자의 전반적인 영어능력 또는 의사소통능력을 평가하는 데 도움이 된다는 것이다. 적용력 또한 사실적, 추론적, 종합적 이해력과 같이 모두 하나의 듣기 구인을 측정한다는 가정하에 모형의 적합도를 조사해보았으나 이해 능력과 적용력의 두 가지 구인을 측정한다고 가정한 최종 모형의 적합도에 비하여 결과가 좋지 못하였다는 점이 이런 사실을 잘 말해준다.

위의 결과에 기초하여, 듣기 측정 모형의 신뢰성과 측정 오류를 최소화하는 측정변인들을 사용하여 듣기 능력의 이론적 모형의 타당성을 조사하였다. 이를 위하여, 그림 3에 제시된 10개의 측정변인들을 이용하여 이들 변인들이 측정한다고 이론적으로 가정한 각각의 구인들을 나타내는 통합 변인(composite variables)을 만들었다. 구체적으로는, 일치/불일치, 할 일/부탁한 일과 특정정보 파악 문항 유형을 한데 묶어 이들 변인들이 측정한다고 이론적으로 가정한 ‘사실적 이해’라는 통합변인을 만들었으며, 관계추론, 주제/요지 파악, 장소추론 및 정보추론 문항 유형을 하나로 묶어 이들 변인들이 측정한다고 가정한 ‘추론적 이해’라는 통합변인을 만들었다. 목적파악은 ‘종합적 이해’의 통합변인으로, 대화 응답 및 담화 응답 유형 문항들은 하나로 묶어 ‘적용’이라는 통합변인을 만들었다. 이들 통합변인들 간의 상관 관계를 이용하여 듣기 이론적 모형의 타당성을 유추해 보았다. 이들 변인들의 상관 관계는 표 3에 제시되어 있으며, 이론적 모형의 타당도 조사 결과는 그림 4에 제시되어 있다.

표 3
듣기 능력 구인들간의 상관 관계

	사실적 이해	추론적 이해	종합적 이해	적용
사실적 이해	1			
추론적 이해	.516**	1		
종합적 이해	.467**	.413**	1	
적용	.556**	.500**	.405**	1

** .p<0.05

표에서 보듯이 듣기 능력 구인들간의 상관관계계수는 .405에서 .556까지 분포하고 있으며, 모두 통계적으로 유의미하였다. 이들 상관 계수를 통해서, 듣기 능력 구인들간에는 그리 강하지도 약하지도 않은 관계가 존재하고 있음을 유추할 수 있다. 표 2에 제시되었던, 문항 유형들간의 상관계수와 비교하여도, 관측된 상관계수가 확연히 차이가 남을 알 수 있다. 표 2에 제시된 문항 유형들간의 관계는 문항에 따라 상당히 낮거나 0에 가까운 상관 계수들이 일부

관측된 반면에, 표 3에 제시된 구인들간의 상관관계를 보면 서로 유사하며 일정한 수치를 유지하고 있음을 알 수 있다. 이는 문항 유형 수준에서는 측정 오류에 따라 서로의 관계가 약하게 보이거나 왜곡되어 보일 수 있다는 것을 시사한다고 할 수 있으며, 표 3에 제시된 구인들 간의 상관 관계는 측정 오류가 최소화된 상황에서 구인들간의 관계 정도를 수치로 나타내어 주므로 보다 신뢰할 수 있다고 하겠다.

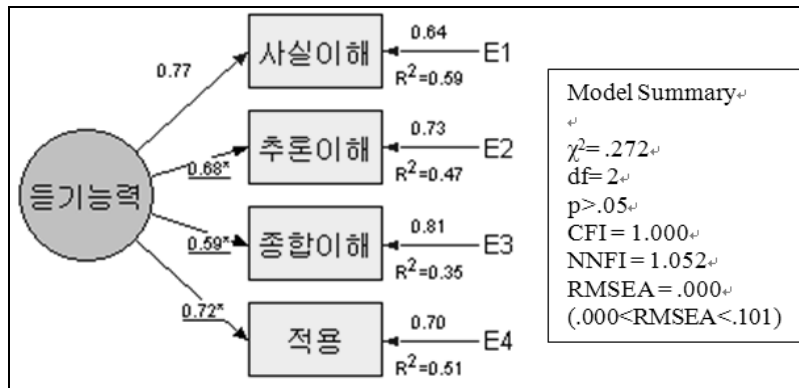


그림 4 듣기 이론 모형의 구인타당도

사실이해 = 사실적 이해; 추론이해 = 추론적 이해; 종합이해 = 종합적 이해

그림 4는 수능 및 전국듣기평가를 통해 추정해 본 듣기 이론적 모형의 타당성 정도를 나타낸다. 그림에서 보듯이, 듣기 평가 이론 모형의 적합성 여부를 알려주는 지수인 CFI = 1.000이었으며, NNFI = 1.052으로 이론적으로 가정한 듣기 모형의 타당성이 상당히 설득력이 있음을 보여준다. 또한 모형의 부적합 정도를 나타내는 지수인 RMSEA는 .05보다 훨씬 낮은 .000을 나타내고 있다. 개별 지수의 적합성도 모두 허용치 범위 안에 있었다. 한마디로, 수능 및 전국듣기평가에 사용되는 영어 듣기 능력 모형은 간결하면서도 상당히 설득력이 있는 이론적 모형이라고 할 수 있겠다.

V. 결론 및 제언

본 연구는 총 4회의 전국듣기평가와 수능듣기평가 문항을 이용하여 전국적으로 시행되는 영어듣기 평가 모형의 구인 타당도를 조사하였다. 총 13개의 듣기 평가 문항 유형을 모두 사용하여 듣기 평가 측정 모형의 적합도를 알아보았으나 기준치에 미치지 못하였다. 일련의 확인적 요인 분석을 통하여 주어진 측정변인들의 관계를 가장 잘 설명하는 듣기 최종 측정모형을 찾았고, 모형의 적합도에 가장 부정적인 영향을 미치고 있는 3가지 문항 유형도 확인

할 수 있었다. 이들 3개의 문항 유형은 그림 보고 묘사 대상 고르기, 그림 상황에 맞는 대화 고르기와 심경파악 유형이었다. 이들 3가지 문항 유형은 문항의 난이도가 지나치게 낮을 뿐만 아니라 듣기 능력의 좋은 척도가 되지 못함을 알 수 있었다. 본 연구에서는 또한 이들 3개의 문항 유형을 교체한다면, 듣기 능력 평가 모형의 구인 타당도는 대단히 높아진다는 것도 알 수 있었다. 또한 본 연구 결과는 듣기 평가를 대화 내용에 적절한 응답 고르기와 같은 적용력을 측정하는 문항과의 연계를 통한 평가가 순수 이해 기능만을 평가하는 것보다 학습자의 전반적인 영어능력 또는 의사소통능력을 평가하는 데 도움이 된다는 것도 시사한다. 이러한 연구 결과는 앞으로 치러질 평가원 주관 전국듣기평가나 NEAT 시험의 듣기 평가의 평가지 설계나 구성 및 문항 제작에 시사하는 바가 크다고 할 수 있다.

본 연구는 또한 이론적 및 교육적으로 시사하는 바가 크다. 이론적으로 보면, 듣기 능력은 추론적 이해력, 사실적 이해력, 그리고 종합적 이해력을 모두 포괄하는 개념이라는 것이다. 하지만 이들 세가지 영역 중에서 듣기 능력의 중추적 구성 요소는 사실적 이해력과 추론적 이해력이라는 점이다. 예를 들어, 본 연구 결과는 대화나 담화의 주요 내용에 대한 일치/불일치를 파악하는 능력, 대화의 내용을 듣고 화자가 부탁한 일이나 대화 후의 할일 파악하기 등과 같은 듣기의 하위 기능들은 사실적 이해력의 중추적 구성요소이며, 정보 추론과 같은 하위 기능은 듣기의 추론적 이해력을 나타내 주는 중추적인 구성요소라는 사실도 알려준다. 이와 같은 연구 결과는 듣기 지도 시에도 동일 시간을 투자하여 학습자의 듣기 능력을 향상시키고자 할 때 듣기의 어느 영역에 얼마나 더 많은 비중을 두어야 하는 지에 대한 구체적인 교육적인 정보도 제공하고 있다.

그럼에도 불구하고, 본 연구 결과를 절대적으로 해석하는 데는 주의가 필요하다. 본 연구의 참여자들은 수능영어평가 기준으로 전국 중위권에 해당하는 학생들이다. 본 연구 참여자 보다 상대적으로 듣기 능력이 낮은 학습자들이 연구에 많이 참여 하였다면 위에서 언급한 3가지 문항 유형의 적합성이 나쁘지 않은 결과가 나왔을 수도 있다. 또한 본 연구 참여자들은 모두 특정한 지역에 속하였으므로 서울, 경기와 같이 다양한 지역의 학습자가 모두 포함 되었다면 본 연구 결과가 달라졌을 수도 있다. 요컨대, 영어 능력의 구인 타당도는 어느 특정한 한 연구 결과에 따라 좌우 되는 것이라기 보다는, **Bachman**과 **Palmer(1996)**이 옳게 지적하였듯이, 끊임 없이 진행되어야 하는 일련의 과정인 것이다. 따라서, 듣기 능력 평가 모형의 타당도 연구는 좀 더 다양한 듣기 평가지와 좀 더 다양한 특성을 가진 학습자를 대상으로, 본 연구에서 제시된 듣기 이론 모형의 일반화 가능성에 대한 후속 연구가 이어져야 할 것이다.

참고문헌

- 교육인적자원부. (2001). *고등학교 교육과정 해설: 외국어(영어)*. 서울: 대한 교과서.
- 김용명. (2010). 국가영어능력평가지험(NEAT)의 문항 유형의 개발과 선별 원리 및 검사지 구성의 원칙. *영어교육*, 65(4), 313-342.
- 김일홍. (2006). 전국단위 영어듣기 평가의 시행효과 및 발전과제. *언어학*, 14(1), 183-202.
- 배태일. (2009). 영어 듣기 설명 모형 연구. *영어교육*, 64(4), 399-418.
- 이명관, 박옥희. (2010). 의사소통 능력 향상을 위한 듣기와 말하기 논문 고찰: *현대영어교육* 학술지를 중심으로. *현대영어교육*, 11(2), 40-59.
- 정행. (2010). 대학수학능력시험 영어 음성언어 평가문항 분석. *Studies in English Education*, 15(2), 177-197.
- 한국교육과정평가원. (2012). *2012 학년도 대학수학능력시험 대비 학습 방법 안내*. 서울: 한국교육과정평가원.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Brown, H. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson.
- Buck, G. (1992). Listening comprehension: Construct validity and trait characteristics. *Language Learning*, 42(3), 313-357.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Canale, M. (1983). On some dimensions of language proficiency. In J. Oller (Ed.), *Issues in language testing research* (pp. 333-342). Rowley, MA: Newbury House.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Mendelsohn, D. (1994). *Learning to listen: A strategy-based approach for the second language learner*. San Diego, CA: Dominie Press.
- Rivers, W. (1981). *Teaching foreign language skills*. Chicago, MI: The University of Chicago Press.
- Shin, Sang-Keun. (2006). Construct validity of listening test items: A verbal protocol study. *English Teaching*, 61(3), 293-305.
- Weir, C. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

부록
평가 영역별 예제 문항

2009학년도 전국듣기평가 2회

사실적 이해 - 일치/불일치

[script]

7번입니다. 대화를 듣고, 내용과 일치하지 않는 것을 고르세요.

M: What brings you in, Sumi?

W: I feel tired and weak.

M: Let me check your medical chart. Everything seems to be normal. Have you experienced any drastic changes recently?

W: I don't seem to have any appetite and I've lost a lot of weight.

M: Is there any reason why?

W: I came from Korea about 6 months ago. I miss my family and friends very much. I dream about them every night.

M: How often do you call your family and friends?

W: Just once in a while.

M: Hmm, you must be homesick. Take some vitamins, eat nutritious food and exercise regularly. You'll feel energized in a few days.

W: Thank you, Dr. Brown.

7. 대화를 듣고, 내용과 일치하지 않는 것을 고르시오.

- ① Sumi는 식욕이 없고 몸무게가 많이 줄었다.
- ② Sumi는 자신의 꿈을 이루기 위해 한국에 왔다.
- ③ Sumi는 가족과 친구들에게 가끔 연락을 한다.
- ④ Dr. Brown은 Sumi가 향수병에 걸렸다고 말한다.
- ⑤ Dr. Brown은 Sumi에게 비타민과 운동을 권한다.

2010학년도 전국듣기평가 2회

추론적 이해 - 주제/요지/화제

[script]

12번입니다. 대화를 듣고, 무엇에 관하여 이야기하고 있는지 고르세요.

M: Thanks for joining us today, Kristina. Are there any fabulous events running right now?

- W: Yes, Andy. If you feel bored and have nothing special to do this weekend, go to Sun Beach and enjoy the Sand Festival.
- M: That sounds interesting. Tell me more about it.
- W: There are a variety of events, ranging from artistic to extreme, all based on an eco-friendly sand theme. Visitors to the festival can enjoy masterpieces by world-renowned sand sculptors, Beach Ball Games, live music and even participate in a sand sculpture competition themselves. You can refer to the guidebook provided at the beach for more information.
- M: How long does the Sand Festival last?
- W: The festival is held from this Friday, June 4th to Monday, June 7th. During the festival period, over 90 businesses near Sun Beach, including lodgings, restaurants, and hot-spring facilities, will offer 15% discounts to visitors.
- M: Wow, it's very attractive. Stay tuned for more. We have lots to talk about with Kristina.

12. 대화를 듣고, 무엇에 관하여 이야기하고 있는지 고르시오.

- ① Book Café
- ② Global Music
- ③ Current Issue
- ④ Sports Highlight
- ⑤ Weekend Outings

2010학년도 수능듣기평가

종합적 이해 – 심경파악

[script]

2. 대화를 듣고, 여자의 심정으로 가장 적절한 것을 고르시오.

[phone rings.]

M: Hello?

W: John, why didn't you pick up the phone? I kept calling you!

M: Sorry, I missed your calls. What's up?

W: I wasn't able to complete my part of our project.

M: What? It's due at 5 p.m. There's only half an hour left.

W: I know, but I couldn't open the file you sent. The password didn't work.

M: What's the password you used?

W: It's 013456, isn't it? That's what you wrote on the memo.

M: Oh, no! It's not 01. It's o1, o13456.

W: What? I've been trying the wrong password until now. I totally lost my patience.

M: I'm sorry. I should've told you it was a combination of letters and numbers.

W: It's too late now. I won't be able to finish by the deadline.

2. 대화를 듣고, 여자의 심정으로 가장 적절한 것을 고르시오.
 ① frustrated ② jealous ③ indifferent ④ satisfied ⑤ grateful

2011학년도 수능듣기평가

적용 - 대화응답

[script]

14. 대화를 듣고, 여자의 마지막 말에 대한 남자의 응답으로 가장 적절한 것을 고르시오.

W: Alex, where have you been? I've been looking all over for you.

M: I'm sorry. I was in a meeting. What's up?

W: You know I have to organize the company fashion show this Friday.

M: Yeah. The boss is expecting an amazing show.

W: That's why I need your help. I have a big problem.

M: I thought everything was going planned. What happened?

W: One of my models came down with the flu. Could you help me find a replacement?

M: I can call one of my friends who works at a modeling agency.

W: Great. But the final rehearsal is in two days. Can you find out for me as soon as possible?

M: _____

14. 대화를 듣고, 여자의 마지막 말에 대한 남자의 응답으로 가장 적절한 것을 고르시오.

Man: _____

- ① I'm here to buy a suit for my concert.
 ② I'll tell you who's available by tomorrow.
 ③ I hope you're feeling better from the flu.
 ④ I think we should put on the show right now.
 ⑤ I want to thank you for finding a replacement.

예시언어(Examples in): English
적용가능 언어(Applicable Languages): English
적용가능 수준(Applicable Levels): Secondary

장재학
강원대학교 영어교육과
200-701 강원도 춘천시 강원대학길 1
Tel: (033) 250-6634/ H.P.: 010-2220-6304
Email: jaehakchang@kangwon.ac.kr

Received 18 March 2012

Revised 7 May 2012

Accepted 19 May 2012

