

# 선제1) 시스템 자동화를 위한 청해 문항 분류 준거 개발

김정렬  
(한국교원대학교)

Kim, Jeong-ryeol. (2007). Developing classification criteria of listening comprehension test items for automating items selection. *Modern English Education*, 8(2), 138-163.

Ministry of Education announced that all the school exam items should be open to the public in 2008. Teachers feel they are not ready yet for this. Before they expose themselves with these high stake exams open to the public, they want a support system to assist them in making up the test items. This paper discusses the possibility of eliciting exam items automatically from the item pool. To construct such a system, selection criteria will be necessary with reference to input structure, content and inference types. The paper attempts to create such selection criteria from the literature survey and expert feedback as the first phase criteria and from the analyses of existing authentic test items as the second phase criteria to validate the first phase criteria. The result in input and inference types provides a working criteria to automate the elicitation process of exam items from a pool.

**Keywords:** [classification criteria/listening comprehension/automating items selection/선제시스템/청해문항분류]

## 1. 서론

교육인적자원부에서는 2008학년도부터 학교 성적의 객관성을 높이고 소위 성적 부풀리기를 방지하고 투명한 관리를 위해 인터넷 공개를 의무화 했다. 그러나 시험 문제 공개와 관련된 여론이 통일된 것은 아니며 교사 조직은 학교의 서열화와 시험 문항의 획일화로 인해서 교육력이 오히려 저해될 것으로 보는 반면 관리자인 교장,

---

1) 선제는 문항은행으로부터 출제 문항을 일정한 기준에 의거하여 선별하는 것을 말한다.

교감이나 학부모들은 대체로 정보 공개가 사회의 큰 추세인 점을 고려해서 투명한 학사 관리를 통해서 자녀들의 문제점을 파악하고 적절하게 대처할 수 있을 것이라고 주장하고 있다. 이와 같이 인터넷을 통한 시험문항 공개를 앞두고 교사들의 평가 권과 학부모들의 교육권 사이에 첨예한 대립각이 형성되고 있다.

사실 신뢰도와 타당도를 담보할 수 있는 좋은 시험 문항의 개발은 많은 지식과 훈련을 필요로 하는 기능이다. 교사들이 사범대학이나 교육대학에서 받는 강의 중심의 평가 관련 3학점으로는 문항 출제에 필요한 훈련을 받았다고 볼 수 없다. 예비 교사 시절에 받았던 부족한 문항 개발 훈련은 교사로 임용된 후에 시험출제를 부담스럽게 생각하고 시험 문항의 공개를 꺼리는 경향으로 나타나고 있다. 물론 시험문항 공개가 난이도의 차이로 학교를 서열화 할 것이라는 교사들의 목소리에도 귀를 기울여야 한다. 그러나 더 중요한 것은 평가와 교육을 따로 따로 볼 것이 아니라 평가를 좀 더 교육친화적인 것으로 만들 필요가 있다. 다시 말해 교육만 이루어지다가 대신 성적 산출에 필요한 중간고사와 기말고사와 같은 고부담 시험을 치를 것이 아니라 평소에 자주 교육과 평가가 함께 가는 교육친화적인 평가시스템이 필요하다. 교육을 좀 더 잘 하기 위해서 진단적 평가가 이루어지고 교육 후에 형성적 평가가 이루어져서 이러한 진단-교육-형성 평가의 사이클이 몇 개 합쳐져서 중간고사가 치루어지고 또 몇 개의 사이클이 더 합쳐져서 기말고사가 치루어지는 형태가 필요하다.

이러한 교육친화적인 평가를 고려한다면 교사들에게 진단평가와 형성평가를 위한 시험출제를 매 번 하도록 한다는 것은 부담되는 일이다. 이러한 부담을 경감하기 위한 방법으로 시험문항의 자동 출제를 고려해 볼 수 있다. 시험문항의 자동화는 문항제작 단계와 문항생성 단계와 문항채점 단계의 자동화로 각각 나누어 볼 수 있다. 문항제작 단계에서는 문항생성의 기본자료가 되는 교사들의 원시문항제작자료의 입력작업과 문항의 직접제작 작업을 컴퓨터를 통하여 수행하기 위한 출제제작 시스템의 자동화를 말하고 이와같은 문항제작 단계의 자동화는 거의 완성된 단계에 있다. 문항생성 단계에서는 제작된 문항 데이터베이스에서 적절한 문항을 발췌하여 단위문제를 편집하거나 지능형 평가에서 실시간 문항생성 업무를 컴퓨터를 통하여 수행하기 위한 문항생성시스템으로 나누어 볼 수 있다. 이에 대한 연구는 아직 미흡한 상황이다. 그리고 수험생이 시험을 본 결과를 컴퓨터로 자동화하여 수행내용을 분석하는 채점분석과정의 자동화 시스템이 있다. 본 논문에서는 출제제작 시스템의 자동화와 채점과정의 자동화는 이미 많은 연구가 되어 있고 기술적인 수준도 거의 완성된 단계에 이르렀는데 반해서 문항생성의 자동화에 대한 연구나 기술수준은 아직 많이 미흡하다는 문제의식을 가지고 출발했다. 문항생성의 자동화를 위한 첫 작

업으로서 자동 생성을 위한 준거가 될 문항의 색인화(item indexing) 과정에 대해서 알아보려고 한다.

문항생성의 자동화는 출제된 문항 데이터베이스에서 교사가 문항을 임의로 추출해서 출제하는 수동형 방식과 문항 데이터베이스로부터 자동으로 문항을 생성하여 자동으로 시험문항을 구성하는 지능형 문항 출제방식이 있다. 지능형 방식을 좀 더 세분화해서 살펴보면 사전 파일럿 평가에 의해서 얻어진 평가문항에 대한 모수정보가 있는 경우 문항분류준거에 모수정보를 고려하여 수험생의 문항반응정보에 따라서 모수 문항은행으로부터 다음문항을 자동 선택하여 평가하는 모수편성방식과 수험자의 문항반응정보에 따라서 컴퓨터에 의해 실시간으로 문항을 자동으로 생성하는 문항 편성방식으로 나누어 볼 수 있다. 지능형 문항생성 방식은 어떤 경우이든 문항의 생성에 개입되는 체계적인 문항분류준거가 있어야 한다. 좀 더 구체적인 본 논문의 목적은 청취 문항의 생성에 개입되는 체계적인 문항분류준거를 마련하는 것이다.

## 2. 이론적 배경

외국어의 발달 정도를 언어 기능별로 측정하는 외국어 평가 모형은 기본적으로 세상 지식을 외국어를 사용하여 이해하고 생산하는 세상 지식과 외국어 기능의 상호작용으로 볼 수 있다. 이러한 상호작용에 개입하는 요소들은 Bachman (1990)에서 주장한대로 수험자의 책략사용 능력, 심리-동작적 능력과 상황적 맥락에 기인한다. 그리고 상호작용에 개입하는 요소들에 평가방식의 측면도 수험생의 시험 결과에 영향을 미친다는 연구(Bachman & Palmer, 1996)를 통해서 평가방식측면 (Test Method Facets)을 추가하게 되었다. 평가방식측면에는 기본적으로 평가환경, 평가척도, 입력 자료, 요구되는 답변 자료, 입력과 반응의 상호작용 등으로 나누어서 제시하였다.

Bachman (1990)년 이전이나 이후에도 동일한 언어능력이라고 하더라도 입력 자료의 배경지식에 대한 수험자의 선형적 지식의 유무가 수험생의 평가 결과에 영향을 미친다는 연구는 많이 나와 있다(Finch, 2005; Alderson & Urquhart 1983, 1985; Bernhardt 1991; Carrell & Eisterhold 1983; Mohammed & Swales 1984; Salager-Meyer 1991; Steffensen & Joag-Dev 1984; Ulijn & Strother 1990). 입력의 내용뿐만 아니라 입력의 형태나 조직도 수험자의 평가 결과에 영향을 미친다는 연구들도 많이 나와 있다(Beck 1982; Davison & Kantor 1982; Duffy & Kabance

1982; Klare 1985; Reder & Anderson 1980).

Lado(1961), Carroll(1961)은 개인의 언어 능력을 언어 구성요소에 대한 지식과 기능으로 보아, 발음, 어휘, 문형, 문장 구조 등의 하위요소를 알고, 듣기·읽기·말하기·쓰기의 기능별로 나누어 처리하는 능력을 언어 능력이라 보았다. 그러나 기능이 단지 지식을 표출하는 채널인지 아니면 기능과 지식은 질적으로 다른지에 대한 분명한 설명이 없다. 이런 관점은 구조주의 언어학의 경우처럼 언어 사용의 맥락을 인식하지 못하는 단점이 있다.

## 2.1. 의사소통능력의 하위 구성요소

Bachman이 제시한 의사소통 능력의 세 가지 하위구성 요소 가운데 심리·생리적 기제는 언어 평가에서 실제로 측정한다는 것이 현재로서 불가능하고, 전략적 능력은 언어 평가에 있어서 수행평가의 한 측면으로 평가되는 것이 가능할 것이지만, 이 두 분야는 본 논문에서 추구하는 원시문항을 색인화해서 문항의 자동생성을 통한 문항의 표준화 작업과정에 직접적인 관련은 없다.

Bachman과 Palmer(1996)는 Bachman(1990) 모형의 기본을 유지하면서 약간 개선하여 변형한 형태의 의사소통능력 모형을 제시하고 있다. 언어 능력을 다시 하위 구성요소인 구성 지식과 실용 지식으로 다시 세분화하고 있는데, 구체적으로 살펴보면 다음과 같다.

표 1  
의사소통 능력과 하위 요소

구조적 능력		화용적 능력	
문법적 능력	담화적 능력	기능적 능력	사회언어적능력
어휘	응집성	생각 기능	방언식별
통사	수사	조작 기능	화역식별
음운		경험 기능	관용적 표현
철자		상상 기능	비유적 표현

Widdowson(2001)은 표 1의 모형에 대해서 최하위 13개의 구성요소들이 상호 관련성이 없다고 주장한다. 게다가, 의사소통능력은 근본적으로 이렇게 인위적으로 나눌 수 없으며, 이러한 인위적인 분리는 의사소통능력 평가가 극복하고자 했던 분리 평가를 오히려 더 지향하게 만든다는 것이다. 온전한 의사소통능력을 측정하기

위해서는 측정할 수 있는 의사소통능력의 온전한 모형을 만들어 내어야 하지만 그것이 가능하지 않기 때문에 의사소통능력을 온전하게 측정할 수는 없고, 다만 의사소통의 어떤 측면들을 평가할 수 있을 뿐이라고 주장한다. 따라서 의사소통능력의 하위요소들을 세세하게 구분할 것이 아니라, 하위 구성요소의 어떤 측면이 더 가치가 있는지를 정해서 가르치고 평가하는 것이 중요하다고 설명한다.

Widdowson의 이러한 지적은 진단과 형성평가의 경우에는 영어교육을 통해서 이루어져야 하고 이루어진 의사소통능력 구성요소를 분리하여 지도목표 및 내용과 연계할 필요가 있다는 것을 시사한다. 지금까지 평가의 측정 목표가 되는 의사소통능력의 개념 및 구성요소에 대한 여러 학자들의 이론을 살펴보았는데, 실제 상황에서 의미를 전달하고 이해할 수 있는 언어사용적인 측면을 강조하고 있다는 것이 공통점이다. 따라서 있을 법한 의사소통 상황 속에서 수험자가 교실 수업에서 학습하고 연습한 의사소통능력으로 평가 과제를 해결하는지 못하는지, 얼마나 성공적으로 해결하는지, 평가 과제를 해결하는 데 필요한 언어 지식은 소지하고 있으며 언어 지식을 의사소통과제를 해결하는데 사용하는 절차적 능력은 습득하였는지 등에 역점을 두고 평가를 시행하는 것이 합리적이라고 본다.

## 2.2. 듣기 시험 문항의 유형화

영어 듣기 시험의 문제를 유형화하는 데 있어서 그 기준은 학자마다 다양하다. 최인철(1991)의 청해 시험 유형, 이재희(2001)의 대학수학능력시험 듣기·말하기 능력 평가문항의 유형별 분류, 여효숙(2006)의 외고 입시 평가 시험 유형 분석을 살펴보고 수정 및 보완하여 차후에 최종 분석 기준을 재구성하기로 한다.

### 2.2.1 최인철(1991)의 청해시험 유형

최인철(1991)은 청해시험의 유형을 크게 그림/사진 제시형, 미시적 청해능력, 거시적 청해능력과 받아쓰기로 나누고 각각을 다시 다음과 같이 세분해서 설명하고 있다.

#### (1) 그림/사진(Picture Cues) 제시형

그림이나 사진의 제시와 함께 음성언어를 들려줌으로써 청해력을 평가하는 시험 양상으로서, 외국인에게 너무 문화적으로 생소하거나, 편파적인 것과 내용이 불분명한 것은 피하여야 한다. 특히, 선형지식구조(schema) 이론에 의하면, 같은 그림도 보는 사람에 따라 달리 보일 수 있기 때문에, 문화적인 차이를 항상 고려해야 한다.

글자를 모르는 초보자를 위해서 사용할 수 있을 뿐만 아니라, TOEIC이나 TSE 처럼 고 난이도수준의 시험문제에도 활용되는 유형이다.

① 단일 그림/사진의 적절한 묘사 선택 유형 : 문제지에 제시된 그림을 음성으로 묘사하는 선택지 중 정답을 고르는 유형이다.

② 한 상황묘사와 일치하는 그림/사진 선택 유형 : 문제지에 제시된 몇 가지 그림 중에서 음성묘사와 일치하는 것을 고르는 유형이다.

③ 일련의 상황묘사와 일치하는 그림/사진 순서 선택 유형 : 문제지에 제시된 몇 가지 그림을 음성묘사와 일치하도록 순서를 올바르게 배열하는 유형이다.

④ 문제지에 제시된 그림에 관한 질문의 응답 유형 : 문제지에 한 상황을 묘사하는 그림(사진)이 제시되고, 문제가 음성으로 제시되고 수험자는 알맞은 답을 보기에서 고르는 유형이다.

⑤ 길/방향 찾기: 일상생활에서 길/방향을 찾거나 알려주는 것도 매우 중요한 언어수행능력이므로, 하나의 독립적인 시험유형이 필요하다. 오해의 소지가 없는 명확한 삽화의 제시가 중요하다.

## (2) 미시적 청해능력(Micro-listening) 평가 유형(단문)

간단한 문장을 제시하여 단편적인 능력요소인 자모음, 억양 및 강세, sandhi(빠른 발음 변이 현상)의 인식, 어휘 및 숙어, 구문력, 세부내용 파악능력 등을 평가하는 시험유형이다. Minimal pair는 음성언어 의사소통 능력의 가장 기본적인 발음식별 능력을 측정하는 미시적인 측정 유형이긴 하나, 문맥이 결여되어 있으므로 자연 음운론(natural phonology)의 내용 타당도 측면에서 의사소통을 위한 청해능력의 시험유형으로 미흡하다.

① 들려진 표현 선택 (단문) : 구어의 가장 기본적인 발음식별 및 청각이미지 기억력을 평가하는 초보적인 능력의 평가유형이다.

② 유사한 내용 선택(Paraphrasing: 단문) : 발음식별력과 아울러 기본적인 표현 및 억양 등을 들려지는 한 문장의 뜻과 가장 비슷한 것을 고르는 비교적 쉬운 난이도의 평가유형이다. 다양한 구어체 표현 및 구문력을 평가하는데 유용한 표현이다.

③ 적절한 응답 선택 (단문) : 대화자가 단문(single exchange) 말한 대화에서 질문에 대한 적절한 응답을 생각하게 함으로써, 종합적인 의사소통능력인 청해력을 평가하는 유형중 가장 기본적인 시험유형이다.

## (3) 거시적 청해능력(Macro-listening) 평가 유형 (대화/담화)

의미단위(meaningful chunk)의 대화 및 담화/강의 등을 구두로 제시한 다음, 수험

자가 문제지에서 적절한 선택을 고르게 하는 형태의 시험이다.

① 시간, 장소, 전/후 상황 추론 (짧은 대화) : 1-2 대화교환(exchange) 정도로 들려진 두 사람간의 대화에 관한 제 3자의 질문에 대한 적절한 대답을 고르는 유형으로서 간략한 의사소통상황을 반영하여 발음식별, 억양, 구어표현, 화자의 태도 유추 등의 청해력을 평가하는 시험이다.

② 화자·청자의 관계 유추 (짧은 대화) : 들려지는 두 사람간의 대화로부터 관계를 유추하는 유형이다.

③ 대의 및 세부정보 파악 (긴 대화 및 담화/강의) : 다소 긴 대화(5~7 exchanges) 내용이나 담화(10~15 문장)를 듣고 대의 및 세부정보를 파악하는 능력을 평가하는 시험 유형이다. 너무 긴 대화나 담화는 수험자에게 언어외적인 기억부담을 주어 신뢰도 및 타당도가 낮아진다. 질문들의 내용은 상호 국부 독립적(locally independent)이어야 한다.

④ 화자의 어조/태도 (대화) : 성공적인 의사소통을 위해서는 Gricean Maxim(quantity, quality, relevance, manner)을 지킬 뿐만 아니라, 대화하는 상대방의 감정 및 태도에도 적절히 반응하여 공손함(politeness)을 견지해야 되므로, 화자 태도의 추론에 대한 평가는 중요한 요소이다.

#### (4) 받아쓰기

적절한 시험 기법을 통한 받아쓰기 시험은 청해능력 뿐만 아니라 언어의 전반적인 능력을 가장 잘 대표한다고 밝혀졌다(Savignon, 1982; Oller & Streiff, 1975).

시험 진행 방식으로는, 일반적으로 지문을 세 번 읽어주게 된다. 첫 번째는 정상 속도로 읽히며 수험자는 듣기만 하고 대의를 파악하도록 해야 하며, 두 번째는 의미 단락으로 끊어서 읽히며, 각 의미 단락 간에 있는 휴지(pause) 기간에 수험자들은 답을 받아 적게 되며, 마지막 세 번째는 수험자들이 교정을 볼 수 있도록 지문을 다시 한 번 들려준다.

① 전체 받아쓰기(Full Dictation) : Cziko(1983)가 제작한 외국 유학생들에게 치러지는 영어시험, English Placement Test(EPT: University of Illinois at Urbana-Champaign)의 받아쓰기 시험의 단락제시(segmenting) 채점방식은 매우 독창적이며 대규모 시험의 효율적이니 타당성이 높은 채점방식으로서 시사하는 바가 매우 크다.

② 부분 받아쓰기(Partial/Spot Dictation; Oral Cloze) : 채점의 효율성을 감안하여, cloze와 전체 받아쓰기의 시험 형식을 절충한 형태의 시험 유형이다. 단답식의 채점방식을 도입한다면 대규모시험에 도입할 가치가 있는 시험 방식이다.

2.2.2 이재희(2001)의 대학수학능력시험 듣기·말하기 능력 평가문항의 유형별 분류  
 이재희(2001)는 표 2와 같이 듣기 능력 평가문항을 유형별로 분류하여 다음과 같이 분석하였다. 듣기영역에서는 특정 정보(물건의 치수, 수, 시각, 장소, 직업, 전공, 대화자의 관계, 목적, 의도, 원인, 심정 등) 파악하기 문항이 70% 정도를 차지하고, 그 다음으로 대화/담화 주제 파악하기(13.5%), 들은 내용에 맞는 그림 고르기(9.4%), 그림 상황에 맞는 대화 고르기(5.2%)의 순서로 사용하였으며, 들은 내용의 순서를 고르는 문항이 새롭게 추가되었다. 들은 내용에 맞는 그림 고르기 문항은 수험자의 심리적 부담을 덜어주기 위하여 항상 1번 문항으로 제시되었으며, 그림 상황에 맞는 그림 고르기는 1997학년도 시험 때부터 사용되었다. ‘대화/담화의 주제 파악하기’ 문항을 늘리고, ‘대화/담화의 내용 파악에 핵심적인 어휘의 이해’ 문항을 추가하면 좋을 것이다.

표 2  
 듣기 능력 평가문항의 유형별 분류



유형	학 년 도									합계 (비율)
	94년 1차	94년 2차	95년	96년	97년	98년	99년	00년	01년	
들은 내용에 맞는 그림 그 리기	1	1	1	1	1	1	1	1	1	9 (9.4)
특정 정보 과 악하기	6	6	5	7	9	8	8	10	9	68 (70.8)
대화/담화의 주제 파악하기	1	1	2	2	1	2	2	1	1	13 (13.5)
그림 상황에 맞는 대화 고 르기	.	.	.	.	1	1	1	1	1	5 (5.2)
들은 내용의 순서 정하기	.	.	.	.	.	.	.	.	1	1 (1.0)
합 계 (전체 문항 중 듣기 비율)	8 (16)	8 (16)	8 (16)	10 (20)	12 (21.8)	12 (21.8)	12 (21.8)	13 (23.6)	13 (26)	96 문항

이재희(2001)는 표 3과 같이 말하기 능력 평가문항을 유형별로 분류하여 다음과 같이 분석하였다. ‘대화 완성하기’ 문항이 63%를 차지한다. 그 다음으로 ‘담화를 대화로 바꾸기’ 문항(17.1%)이 많은데, 후반기에는 이 문항이 ‘담화 상황 이해 및 응답하기’로 전환하였다. 말하기 평가 문항의 비율이 점차 증가한 것은 고무적인 일이지만, 구어 의사소통능력에 중점을 두려면 문항수를 더 늘리면서 문항 유형도 다양화해야 할 것이다.

표 3  
말하기 능력 평가문항의 유형별 분류

유 형	학 년 도									합계 (비율)
	94년 1차	94년 2차	95년	96년	97년	98년	99년	00년	01년	
대화 완성하기	2	1	2	1	4	3	3	3	3	22 (62.9)
답화를 대화로 바꾸기	2	2	1	1	·	·	·	·	·	6 (17.1)
대화/답화 내용 요약하기	·	·	·	·	1	1	1	1	·	3 (8.6)
답화 상황 이해 및 응답하기	·	·	·	·	·	1	1	1	1	4 (11.4)
합 계 (전체 문항 중 말하기 비율)	4 (12.9)	3 (9.7)	3 (9.7)	2 (6.5)	5 (16.1)	5 (16.1)	5 (16.1)	5 (16.1)	4 (8)	35 문항

2.2.3 여효숙(2006)의 외고 입시 평가 시험 유형 분류

여효숙(2006)은 외고 입시의 문항 유형을 크게 듣기, 듣고 말하기, 독해 듣기의 세 가지로 나누고 각 유형별로 문항유형들을 세분화해서 표 4와 같이 제시하였다.

표 4  
외고 입시 평가 시험 유형 분류

일반 듣기 유형	듣고 말하기 유형	독해 듣기 유형
지도·장소 파악	적절한 응답	글의 주제·요지
그림·인물 묘사	적절한 결론	글의 종류·제목 파악
순서배열	요약문 완성	세부정보·내용 일치
시각·기간 묻기	상황에 적절한 말	심경·태도·분위기
액수·요금·숫자 묻기	어울리지 않는 대화	글의 목적파악
전화목적 파악		제안·주장·의도·교훈
대화 장소·관계		빈칸 완성
그림상황에 맞는 대화 장면·상황 파악		요약·결론
직업 파악		지칭·함축 추론
분위기·심경·태도		전후문맥 문장삽입·배열
이유·의도 파악		무관한·적절한 문장 넣기
의견·주장·결론 도출		(표)계산 문제
구체적 정보 파악		통합교과형 문제
세부 내용 일치		
주제·요지 파악		
전후문맥·문맥상 의미 파악		
지칭추론		

### 3. 연구 방법

본 연구는 학습자들의 진단과 형성 평가를 위한 청해시험 문항 자동생성에 대한 준거개발이므로 문헌연구와 전문가와 교사의 피드백을 받아서 1차 준거를 개발한 다음 분석연구를 통해서 1차 준거에서 드러난 문제점을 수정보완하여 2차 준거를 개발하는 방법으로 접근한다. 우선 청해력을 측정할 때 게재될 수 있는 요인들을 문헌연구를 통해서 구성하고 이렇게 구성된 준거에 전문가 20인으로부터 피드백을 받아 1차 준거를 완성한다. 1차 준거를 기준으로 대입수능과 외고입시 등에 출제된 공신력 있는 청해력 문항들을 중심으로 분석준거를 적용하여 준거에 대한 적정화 과정을 거치고 문제가 되는 준거 항목을 수정보완하여 2차준거를 마련한다.

1차 준거를 적용하기 위한 분석대상 문항은 수학능력시험 2001학년도에서 2006

학년도까지 듣기 문항 (102)과 외고 입시문항은 가장 최근 것을 중심으로 온라인에서 구할 수 있는 문항들로서 대원외고 (120), 용인외고 (80), 대일 외고 (60), 중산외고(30), 고양외고(30), 명덕외고(30), 및 인천외고 (40) 총 492문항을 대상으로 하였다.

이론적 배경에서 논의된 평가방식측면에서 수험생의 평가 결과가 달라질 수 있다는 논의 내용을 토대로 분석의 틀을 마련한다. 평가 내용 및 방식과 관련된 요인들 중에서 문제를 출제할 때, 통제 가능한 요인들을 살펴보면 평가환경, 평가척도, 입력 자료, 요구되는 답변 자료, 입력과 반응의 상호작용 요인들 중에서 평가척도, 입력 자료, 답변 자료로 압축해 볼 수 있다. 평가척도와 입력과 반응의 상호작용 요인들은 평가문항의 제작 단계보다는 평가의 실시와 관계된 요인들이 많으므로 이는 제외하고 나머지 요인들을 중심으로 살펴보기로 한다.

입력 자료와 답변 자료를 중심으로 살펴보면, 자료의 구조와 내용에 대한 것을 나누어 볼 수 있다. 자료의 구조에는 지문의 형태, 언어적 복잡도 등을 준거로 제시할 수 있으며 내용적 요소에는 지문의 주제나 소재가 여기에 속한다. 그리고 요구되는 답변 자료는 추가적으로 문제의 유형을 결정하는 요인으로 들어갈 수 있다. 이를 중심으로 큰 범주를 나누어보면 다음과 같이 입력언어의 구조, 입력언어의 내용, 요구되는 학습자의 추론행동과 같은 세 가지 요인들로 나눌 수 있다.

먼저 입력언어의 구조는 지문의 형태와 언어적 복잡도로 나누어 볼 수 있다.

지문의 형태: 지문의 형태는 대화문인지 담화문인지를 구분하는 것이다. 대화문은 2인 이상이 영어로 상호작용하는 경우를 말하고 담화문은 1인의 독백이나 연설 또는 설명으로 이루어진 경우를 말한다.

언어적 복잡도: 언어적 복잡도에는 문법적 복잡도와 어휘적 복잡도로 나누어 볼 수 있다. 물론 담화적 복잡도도 있으나 이를 자동화할 수 있는 이론적 근거가 희박하기 때문에 여기서는 문법적 복잡도와 어휘적 복잡도를 준거로 활용한다.

입력 언어의 내용은 지문의 주제나 소재를 중심으로 나눌 수 있다.

지문의 주제나 소재: 지문의 주제나 소재는 언어의 내용적 요소로서 수험자의 평가 결과에 영향을 미친다. 수험자에게 익숙한 소재의 글이거나 선호하는 분야의 글이면 평가를 잘 받을 수 있는 확률이 그만큼 높아지기 때문이다. 언어의 내용적 요소 변인을 최소화하려면 주제나 소재를 유목화해서 나누고 문항의 자동 생성시에 가능한 다양한 내용적 요소들이 추출될 수 있도록 하는 것이 중요하다.

요구되는 학습자 추론 행동은 일반적으로 문제의 유형이라고 부르는 것으로 줄여서 문제의 추론 유형을 중심으로 나눈다.

문제의 추론 유형: 문제의 추론 유형은 문항에서 학습자들이 지문을 듣고 요구되

는 답변 자료의 양태에 대한 것으로 학습자들이 필요로 하는 듣기 방식이나 답을 찾아가는 기능과 관련된 것이다. 문제의 추론 유형은 문항의 외형적 모습과 상관관계가 높고 학습자들이 청취력을 키울 때에 요구되는 학습 요소이기도 하다. 이에 대한 연구는 비교적 잘 되어 있기 때문에 이론적 배경에서 나온 문항의 유형을 정리해서 살펴보도록 한다.

이론적 배경을 토대로 나온 준거들을 기초로 5년 이상 교직경력이 있는 현장교사 10명, 수능출제경험이 있는 교수 및 교사 7명, 외교에서 현직교사로 있으면서 외교입시 출제경험이 있는 교사 3명으로 총 20명의 전문가로부터 피드백을 받아서 수정보완 하여 1차 준거를 마련한다. 그리고 2차 준거는 1차 준거를 기출문항으로 구성된 문항은행에 적용해서 적정화 과정을 거치도록 하였다. 그 적정화 과정에서 드러난 문제점들에 대한 수정보완 과정을 거쳐서 2차 준거를 마련하였다.

## 4. 연구 결과

### 4.1 문헌 및 전문가 피드백에 의한 결과 (1차 준거)

1차 준거에서는 입력 언어에 해당하는 항목인 지문유형과 문장의 복잡도 부분에 대한 내용으로 교사 및 전문가 집단으로부터 피드백을 받기 전에는 문헌조사를 바탕으로 지문유형은 대화와 담화로 나누고 문장복잡도는 어휘 복잡도와 문장내의 단순 어휘빈도수를 중심으로 설정하였다. 그러나 교사 및 전문가의 피드백을 받아서 최종적인 1차 준거는 조금 다른 모습으로 나타났다. 구체적인 피드백에 의해서 수정된 내용은 아래에 언급하였다.

#### 4.1.1 입력언어의 구조

입력언어의 구조에는 이론적 배경을 바탕으로 구성된 1차 초안에는 지문유형으로 대화와 담화의 두 가지 유형과 입력언어복잡도로 어휘 복잡도만 참고로 했다. 그러나 교사 및 전문가들은 외교 입시의 경우 담화로 단순 분류하기 어려운 상당히 긴 독해형 지문이 출제되므로 대화와 담화 외에 다른 범주의 지문유형이 필요하다는 피드백을 하였다. 그리고 두 분의 전문가가 입력언어의 복잡도를 단순히 어휘의 복잡도로만 따지는 것에 문제를 제기하였다. 다시 말해 문장의 복잡도가 지문의 복

잡도를 결정짓는 중요한 요소라는 주장이었다. 따라서 지문유형에는 독해형 지문이 추가되고 입력언어 복잡도에는 문장 복잡도가 추가되었다.

(1) 지문 유형

① 대화 ② 담화 ③ 독해

외고입시 출제 경험이 있는 교사들의 피드백 중에 독해형 지문이 있으므로 지문의 종류에 독해형 지문을 추가하였다.

(2) 입력언어 복잡도

어휘 분석은 어휘 자동 분석 프로그램에 의해 난이도를 분석하였다. 어휘 자동 분석 프로그램은 Paul Nation이 구성한 Range로 General Service List를 어휘 목록으로 활용하여 어휘 복잡도의 구간을 자동적으로 표시해주는 프로그램인데, 여기에 우리 교육과정에 나온 초등용 어휘 500단어로 구성된 목록을 한 단계 더 추가하여 수정 변형한 프로그램을 사용하였다. 자동화 프로그램은 전체 어휘의 80% 이하가 빈도 어휘수 500단어 범위내에 있으면 1단계로 분류하고, 빈도 어휘수 500-1000 단어 사이에 20% 이상의 어휘가 속하며 3단계가 아니면 2단계에 속한다. 그리고 빈도 어휘수 1000-2000 사이에 지문 어휘의 20%이상이 있으면서 4단계가 아니면 3단계로 분류하고 지문 어휘의 20%이상이 빈도 어휘수2000-3000 사이에 있고 5단계가 아니면 4단계로 분류하였다. 마지막으로 지문 어휘의 20%이상이 빈도 어휘수 3000 단어를 넘어서면 5단계로 분류하였다.

교사 및 전문가 피드백을 통해서 입력언어 복잡도를 단순히 어휘의 빈도범위로 분류하기에는 무리가 있다는 주장들이 많았다. 따라서 기술적으로 가능한 범위에서 가장 단순한 형태로 문장 복잡도를 반영하기로 하였다. 우선 문장단위의 복잡도는 교육과정의 내용을 참고해서 문장당 어휘수를 계산해서 이들 문장의 20%이상이 최상위 단계이면 해당 단계에 속하는 것으로 간주하였다. 각 단계는 문장당 어휘수가 7개 미만인 문장이 전체문장의 80%이상이면 1단계, 7개에서 10개 미만인 문장이 20%이상이고 3단계에 속하지 않으면 2단계, 10개에서 20개 미만인 문장이 20%이상이고 4단계에 속하지 않으면 3단계, 문장당 어휘수가 20개에서 30개미만이고 5단계에서 속하지 않으면 4단계, 30개 이상인 것이 20%를 넘으면 5단계로 계산하였다.

4.1.2 입력언어의 내용

입력언어의 내용에 해당하는 지문의 소재는 크게 일상생활, 자연, 사회, 인문으로 나누고 각 범주별로 아래에 열거한 내용으로 다시 소분류하였다.

- ① 일상생활 : 건강, 운동, 오락, 여행, 쇼핑, 친교, 학교, 가정, 관계, 신소재
- ② 자연 : 동물, 식물, 수학, 물리, 화학, 환경, 컴퓨터, 통신, 지구, 우주, 해양, 신소재
- ③ 사회 : 도덕, 정치, 경제, 역사, 지리, 상업, 직업, 신소재
- ④ 인문 : 음악, 미술, 문학, 문화, 언어, 신소재

#### 4.1.3 문항의 추론유형

문항별 학습자로부터 기대되는 답변 행동에는 다음과 같은 것들로 구성했다.

① 그림, 도표: 사진, 광고문, 각종 안내문, 비행기시간표, 식당 메뉴 등과 같은 시각자료를 보면서 대화를 이해하는 사실적 이해능력을 요구한다. 시각 자료의 내용이 대화의 대의와 세부내용 중 어느 것과 관련된 것이냐에 따라, 대화의 내용과 전개가 달라진다. 대의와 관련된 경우에는 대화의 내용이 시각 자료의 어느 한 항목에 집중되며, 세부내용과 관련된 경우에는 대화의 내용이 시각 자료의 여러 항목과 관련되므로 하나씩 체크하면서 도표에 비추어서 사실성 여부를 체크할 수 있어야 한다.

② 심경, 심정: 수험자의 대화 내용을 듣고 화자의 심리 상태를 파악하는 종합적 추론 능력을 요구한다. 사람의 감정 상태를 기술하는 어구나 표현이 대화 가운데서 찾아서 답할 수 있어야 한다.

③ 장소: 추론적 이해를 체크하고자 하는 문항으로 대화를 듣고 대화의 장소를 추론하는 능력을 요구한다. 대화가 어디에서 일어나고 있는지 암시하는 표현이 대화 가운데 적절한 곳에 적절하게 나와 있고 이를 근거로 답변을 할 수 있어야 한다.

④ 이유, 원인: 지문 내용을 근거로 인과관계에 대한 추론적 능력을 체크하는 문항이다. 대화가 어떤 이유나 원인으로 말미암아 이루어졌는지를 전체적으로 이해하고 답해야 한다.

⑤ 관계, 직업: 대화 내용을 근거로 하여 화자의 상대적 관계나 직업을 추론하는 능력을 측정한다. 대화가 어떤 목적에서, 어떤 배경에서 이루어지는지 전체적으로 이해하면, 배경지식에 의해 화자가 어떤 사람인지 추론할 수 있다.

⑥ 수치: 대화를 듣고, 날짜, 시간, 요금, 가격 등의 숫자 정보를 파악하는 특정 정보에 대한 사실적 이해 능력을 요구한다. 여러 숫자 정보가 혼합되어 있으면서도 대화가 자연스럽게 이어질 수 있는 대화를 이해하고 파악해서 필요한 정보를 찾아

내는 능력이 요구된다.

⑦ 내용 일치: 대화나 담화에 나오는 특정 정보에 대한 사실적 이해를 측정한다. 대화나 담화 가운데 묻는 특정 정보로서 흔히 묻는 형태의 특정 정보는 어떤 일을 하는 방법, 어떤 일에 대한 이유, 하려고 하는 일 등에 관한 것이다.

⑧ 요지, 주장, 목적, 속담: 담화나 대화를 듣고, 그 목적이나 주장이 무엇인지 파악하는 종합적 추론 능력을 요구한다. 담화나 대화의 목적이 내용의 전반적인 흐름에 자연스럽게 포함되어 전체 내용을 파악하면서 답변할 수 있어야 한다.

⑨ 주제, 소재: 대화나 담화의 전체적인 내용에 비추어 지문이 무엇에 관한 것인지 이해하는 능력을 측정한다. 세부적인 내용을 정확하게 듣지 못하였다 하더라도, 반복적으로 나오는 표현이나 유사 어구를 듣고 배경지식과의 상호작용에 의해 지문의 주제를 파악하여 답해야 한다.

⑩ 말하기 (응답): 대화 내용을 근거로 대화의 마지막 말에 적절한 응답을 고르는 문제이다. 대화의 전체적인 내용 이해와 마지막 부분의 의사소통 기능의 이해가 함께 돼야 문제를 풀 수 있도록 출제해야 한다. 마지막 부분만을 듣고 답할 수 있도록 해서는 안 된다.

⑪ 대화문 찾기: 말을 한 번씩 주고받는 짧은 대화들을 듣고 주어진 그림이나 묘사된 상황에 가장 적절한 대화를 고르는 능력을 체크한다. 그림의 핵심적인 내용을 특징적으로 묘사하는 대화를 찾을 수 있어야 한다.

여기에 덧붙여서 교사 및 전문가들의 피드백에서 듣기와 읽기 통합유형의 문제로 대화문 찾기와 빈칸, 순서 유형을 추가할 것을 자문하였다.

⑫ 빈칸, 순서 (듣기, 읽기 통합): 수능에서는 읽기에 나오는 문항인데 외고입시에서 듣기와 읽기 기능의 통합문제로 듣고 빈칸에 들어갈 표현을 고르거나 듣고 순서에 맞도록 문단을 배열하는 능력을 가려내는 문항이다.

#### 4.2 문항 분석을 통한 수정보완 (2차 준거)

492문항 중에서 지문유형으로 나누어보니 수능 102문항은 주로 담화와 대화로 나눌 수 있으나 외고입시 문항 390문항은 다양한 듣기 문항이 출제되었다. 독해형 지문으로 듣기와 읽기를 결합한 것도 있고, 실제로 과학이나 수학과 같은 내용 교과 의 지식을 묻는 영어 논술식 듣기 지문도 있었다. 따라서 독해형 지문을 독해지문과 내용중심의 통합형 지문으로 나누어 분류할 필요가 있었다.



표 5  
지문유형별 분류  
N=492

지문유형	빈도	비율
대화	57	11.6
담화	103	20.9
독해	176	35.8
통합	156	31.7

#### 4.2.1 언어 복잡도

어휘 분석 자동화 프로그램을 돌려서 지문을 분석해 본 결과 그 분포는 다음과 같이 나왔다. 1000에서 2000 단어 사이에 속하는 3단계 지문이 가장 많았고 2단계와 4단계는 거의 비슷한 비율로 뒤를 이었다. 대화문은 주로 1-2단계 어휘로 이루어진 지문이었고, 담화는 1-3단계 사이에 분포했으며 독해형 지문과 통합형 지문이 비교적 높은 단계의 어휘 복잡도를 가진 지문들이 쓰였다.

표 6  
어휘의 단계별 분류  
N=492

어휘의 단계	빈도	퍼센트
1단계 (500이하)	37	7.5
2단계 (500-1000)	125	25.4
3단계 (1000-2000)	167	33.9
4단계 (2000-3000)	124	25.2
5단계 (3000이상)	39	7.9

문장의 복잡도는 1차 준거에서는 단순히 문장당 평균 어휘수를 중심으로 1단계에서 4단계까지 분류했다. 그 결과는 다음과 같이 나타났다.

표 7  
문장 복잡도

N=492

문장당 평균 어휘수	빈도	퍼센트
1단계 (7개 미만)	117	23.8
2단계 (7-10개 미만)	131	26.6
3단계 (10-20개 미만)	142	28.9
4단계 (20-30개 미만)	96	19.5
5단계 (30개 이상)	6	1.2

문장당 어휘의 평균만으로 문장의 복잡도를 보는 것에는 예측했던 대로 일부 문제점이 보였다. 가장 대표적인 문제점은 and나 or같은 등위 접속사로 문장의 일부 구성성분이 연결되어서 계속 나타나는 경우에 아래에 제시된 예문과 같이 문장의 구조 자체가 그렇게 복잡한 것은 아니면서 어휘 수는 많아지는 경우가 자주 나타났다.

1) With Richards and Jagger out on bail and their friends shortly gathered and discussed... [25개 어휘]

2) Ts'ao Kung, Li Ch'uan, Meng Shih, Tu Yu, Tu Mu and Mei Yao-ch'en have notes to the effect that... [32개 어휘]

이 문제를 해결하기 위해서 t-unit의 개념을 도입할 필요성이 있었다. t-unit을 보면 1)의 경우는 2개이고 2)의 경우도 2개가 된다. 문장당 t-unit의 평균수를 보면 모문과 삽입절의 구조를 짐작할 수 있고 문장의 복잡도를 계산할 수 있다. 여기에 덧붙여 문장당 어휘의 수를 t-unit의 가중치로 곱한 값으로 처리하면 문장의 상대적 복잡도를 계산할 수 있다. 예를 들어 다음의 한 문장 지문이 있다고 가정하고 이들 지문에 대해서 복잡도를 계산하면 다음과 같다.

1) Simple sentences(단문)- 주어+동사가 하나만 있는 문장.

a. She loves me. (1 t-unit \* 3 = 3)

2)Compound sentences(중문)-주어 하나와 두 개 이상의 동사가 갖추어진 2개의 문장이 등위 접속사 and, but, for, or, so등으로 연결된 것.

- a. She loves me and everybody around me. (1 t-unit \* 7 = 7)
- 3) Complex sentences(복문)- 주어+동사가 갖추어진 2개의 문장이 등위 접속사가 아닌 종속 접속사(that, if, when, as, though 등)로 연결된 문장.
- a. I think that she loves me. [2 t-unit \* 6 = 12]
- 4) Compound-complex sentences(혼문)-한 문장안에 단문, 중문, 복문이 같이 섞여 있는 것.
- a. I think that she loves me, but she doesn't really love me. [3 t-unit \* 9 = 27]

#### 4.2.2 소재

소재의 경우는 1차 준거에서 예상했던 것들이 지문 속에 많이 등장했으나 새로운 장르의 소재도 실제로 문항 분석 과정에 많이 나타났다. 실제로 분석과정에서 나타난 이러한 신소재들도 추가로 목록에 넣었다. 그 구체적인 내용을 일상생활, 자연, 사회, 인문으로 나누어서 표로 정리해서 살펴보면 다음과 같다.

**표 8**  
**입력언어의 소재**  
 N=492

소재		빈도수	백분율	
일상생활	건강	19	3.9	
	운동	20	4.1	
	오락	17	3.5	
	여행	21	4.3	
	쇼핑	14	2.8	
	친교	22	4.5	
	학교	29	5.9	
	가정	17	3.5	
	관계	16	3.3	
	신소재	교통/사고	15	3.0
		지시	9	1.8
		식당/음식	15	3.0
		날씨	10	2.0
		심정	7	1.4
전화		13	2.6	
의복		12	2.4	
화재예방	11	2.2		
소계		267	54.3	

<표 계속>

자연	동물	4	0.8	
	식물	3	0.6	
	수학	2	0.4	
	물리	2	0.4	
	화학	3	0.6	
	환경	4	0.8	
	컴퓨터	5	1.0	
	통신	6	1.2	
	지구	3	0.6	
	우주	2	0.4	
	해양	2	0.4	
	신소재	발명	2	0.4
		에너지	2	0.4
의학/질병		3	0.6	
기계		3	0.6	
<b>소계</b>		<b>46</b>	<b>9.3</b>	
사회	도덕	12	2.4	
	정치	5	1.0	
	경제	9	1.8	
	역사	5	1.0	
	지리	7	1.4	
	상업	9	1.8	
	직업	15	3.0	
	신소재	사회복지	6	1.2
		교육	7	1.4
		인구	3	0.6
		인권	3	0.6
		범죄예방	2	0.4
	<b>소계</b>		<b>83</b>	<b>16.9</b>
인문	음악	9	1.8	
	미술	11	2.2	
	문학	21	4.3	
	문화	17	3.5	
	언어	21	4.3	
	신소재	철학	7	1.4
		고고학	3	0.6
		미신	2	0.4
		인물	5	1.0
<b>소계</b>		<b>96</b>	<b>19.5</b>	

표 8에서 보여준 대로 2차 준거 마련을 위한 문항 분석을 통해서 조사한 바에 따르면 신소재의 유형들은 각 대범주마다 골고루 나타나고 있는데 이들의 예를 대범주별로 다소 특이한 것들만 살펴보면 다음과 같다.

① 일상생활 신소재 유형: 교통 / 사고

대화를 듣고, 여자의 교통법규 위반행위에 포함되지 않는 것을 고르시오.  
 M: May I see your driver's license, please?  
 W: Did I do anything wrong, officer?  
 M: You were speeding. Let me see your license and car registration, please.  
 W: Sure. It's here somewhere in the glove compartment. Hold on a minute. Here you are.  
 M: Ma'am, do you know how fast you were going in the school zone?  
 W: What? I'm not sure, but I think I was driving about 40 miles.  
 M: You were going 60 miles in a 20-mile zone. And you failed to come to a complete stop at the intersection back there.  
 W: Didn't I stop there?  
 M: Absolutely not... And you're not buckled up... Do you know both of your break lights are out?  
 W: Really? I didn't know that.  
 M: Wait a minute. This is not your car, and your driver's license expired ten months ago.

② 일상생활 신소재 유형: 식당/음식

다음을 듣고, Cindy가 좋아하는 음식으로 가장 적절한 것을 고르시오.

W: Yesterday was Cindy's birthday. Cindy and Danny went to a Japanese restaurant for dinner. There were so many kinds of food that they didn't know what to order. Cindy preferred fish but Danny didn't like fish. So they decided to order fish for her and pork for him. The food was great!

- ① fish                      ② kimchi                      ③ salad  
④ pork                      ⑤ cake

### ③자연

다음을 듣고, 손목시계의 기능으로 언급되지 않은 것을 고르시오.

[Script]

M: Every busy modern person needs a watch. Not only do watches tell the time, but the latest versions can do a great many other things. For example, many watches show the exact month and date. But that is not all. You can find watches that tell the temperature and tell the time of other time zones. Some watches have a calculator. A great addition is the watch that has an alarm. Business people set the alarm to ring when they need to make an important call, while traveling.

다음을 듣고, 손목시계의 기능으로 언급되지 않은 것을 고르시오. (1.5점)

- ① showing month and date  
② telling the temperature  
③ telling the time of other time zones  
④ measuring heartbeats  
⑤ ringing an alarm

## ④사회-사회복지

다음을 듣고, 남자가 주장하는 바로 가장 적절한 것을 고르시오.

M : As I have shown, low-income senior citizens make up approximately 30 percent of the elderly population. Senior citizens are among the weakest members of society because they depend so heavily on government programs for food, shelter, and medical needs. They will suffer most severely if the government cuts back on its social welfare programs. Therefore, more social benefits should be provided for the elderly.

다음을 듣고, 남자가 주장하는 바로 가장 적절한 것을 고르시오.

- ① 빈부 격차를 줄여야 한다.
- ② 정부의 예산을 삭감해야 한다.
- ③ 실업 수당을 대폭 늘려야 한다.
- ④ 노인들에게 일자리를 주어야 한다.
- ⑤ 노인들을 위한 복지 제도를 확대해야 한다.

## 4.2.3 문항 유형

문항 유형은 소재에 비해서는 일반적으로 예측했던 1차 준거에서 크게 벗어난 것은 없었다. 다만 어휘와 관련된 1차 준거에서 예측하지 못했던 유형이 다수 있었기 때문에 추가해야 할 필요성이 있었다. 2차 준거 마련을 위한 분석결과를 살펴보면 다음과 같았다.



표 9  
문항유형분류  
N=492

문항 유형	빈도수	백분율
그림 정보파악	21	4.3
장소 추론	27	5.5
심정/성격파악	28	5.7
한일(할일)파악, 부탁	19	3.9
목적 파악	19	3.9
이유/원인 파악	22	4.5
관계/직업추론	23	4.7
수치파악	21	4.3
특정 정보 파악	19	3.9
일치/불일치	81	16.5
요지파악	36	7.3
주제 파악	23	4.7
도표 및 시각 자료 활용	23	4.7
그림 상황 적절한 대화 찾기	28	5.7
적절한 응답 찾기	35	7.1
상황에 적절한 말 찾기	26	5.3
어휘정의	41	8.3

2차 준거에서 새로이 추가된 어휘정의 관련 문항은 예상외로 8.3%를 구성할 정도로 많이 출제되는 경향이 있었고 이러한 추세는 2005년도 이후에 좀 더 뚜렷해지는 추세를 보이고 있다. 구체적인 예시 문항을 살펴보면 다음과 같다.

<p>다음 문장을 듣고, "shortly"의 문맥상 뜻으로 가장 적절한 것을 고르시오.  W: <i>The two boys set off on the river, and landed down the river shortly after dark.</i></p> <p>다음 문장을 듣고, shortly의 문맥상 뜻으로 가장 적절한 것을 고르시오.</p> <p>① on time                      ② easily                      ③ suddenly  ④ at last                      ⑤ soon</p>
---

## 5. 결론

시험 문항의 자동생성을 위해서는 원시문항의 데이터베이스가 구성되어야 하는데, 이 과정의 자동화는 거의 완성단계에 있다고 볼 수 있다. 일선 학교에서 교사들이 출제한 시험문항을 공개하도록 하면 문항출제의 표준화가 가속화 될 것이고 이러한 원시문항들을 교육청의 교수학습센터나 교육과정평가원과 같은 기관에서 수집하여 문항데이터베이스를 구성해서 다른 교사들이 이용할 수 있도록 서비스하면 교사들의 입장에서는 자신의 시험문항을 제공하고 전체 데이터베이스의 서비스를 받는 상호 호혜적인 관계를 설정할 수 있다. 교육청이나 교육과정 평가원에 마스터 서버를 두고 네트워크를 통해 마스터 서버와 연동되는 다수의 서브 서버 및 서로 다른 시험 문제 출제 시스템으로 구성할 수 있다. 마스터 서버는 각 시험 문항에 포함된 각각의 선택지에 문항과 선택지 번호를 포함하는 식별코드가 지정된 베이스 출제지 파일을 입력받은 후, 상기 베이스 출제지 파일을 XML로 변환하여 문제 문항수, 문항별 선택지 수, 선택지 식별코드의 정상 부여 여부를 검수하는 과정을 거친다. 생성된 전체 개별 시험지의 문항별 입력구조, 입력어 내용 및 요구되는 추론 행동 등에 관한 식별표를 활용하여 시험지 전체의 혼합도를 추출하고 문항의 재변경이 가능한 가변 문항 정보까지 산출해 낼 수 있다. 혼합율에 따라 문항 정보 세트를 개별 시험지 파일 수에 맞게 생성한 후, 생성된 문항 배열 정보 세트를 서브 서버의 수에 따라 그룹핑하여 선택지 배열 정보 세트 그룹 각각을 선택지 변환 정보 파일로 시험 문항을 최종적으로 생성하여 다양한 시험문제를 비교적 쉽게 생성할 수 있도록 자동화한다.

이러한 문항출제시스템은 교사들이 시험문항자동생성 프로그램을 활용해서 시험 문항을 추출해서 수업을 위한 진단과 형성평가와 같은 저부담 시험에 사용할 수 있도록 해준다. 교사의 입장에서 진단과 형성평가의 필요성을 누구나 인정하면서도 단위별 진단과 형성 평가 문항을 구안하여 매번 시험을 치른다는 것은 현실적으로 힘든 일이다. 따라서 문항 출제의 부담을 줄이면서 이미 단위 내용을 대상으로 구성된 검증을 거친 문항들로 이루어진 데이터베이스가 있고 자동문항추출시스템이 있다면 교사들이 이를 사용하여 좀 더 교수학습과 연계되고 친화적인 시험으로 학생들의 수업내용 이해도에 대해서 좀 더 과학적으로 진단과 형성 과정을 살펴볼 수 있다.

문항의 자동생성 과정에서 필수적인 요소가 각 문항에 대한 자동화 준거인데, 이를 위해서 기존의 연구자들이 분류해 놓은 시험의 문항유형에 덧붙여서 듣기 지문

의 형태와 소재를 덧붙이고 입력문의 복잡도를 자동으로 측정할 수 있도록 해서 이러한 요소들을 복합적으로 고려한 문항 추출이 이루어질 수 있는 기본적인 문항정보의 구성에 대한 논의를 하였다. 문항의 자동추출을 위해서 원시문항 출제자인 교사들이 제공하게 될 관련 단원, 문항의 목표와 같은 이원분류표에 있는 기본적인 내용 외에 입력언어의 구성과 내용, 문항의 유형과 같은 정보들을 분류준거로 사용하여 자동적으로 추출하게 된다. 기본적인 이원분류표 정보는 해당 정보를 찾아서 1차 문항풀로 활용하고 모수정보에 해당하는 입력언어와 문항 유형에 대한 것은 1차 문항풀 가운데서 입력언어의 경우에는 구조적으로 해당 단계의 학생들 수준에 맞고, 내용은 지문의 소재별 비율을 고려한 다양한 형태의 지문이 제공될 필요가 있다. 그리고 문항의 유형에 관한 정보도 활용하여 가능한 학습자의 수행목표와 관련된 문항들 가운데서 다양한 문항유형이 추출될 수 있도록 문항유형 준거가 사용되어야 할 것이다.

### 참고문헌

- 여효숙. (2006). *외국어 고등학교 입학시험 내용과 시험방식 분석*. 미출간 석사학위논문. 성신여자대학교, 서울.
- 이재희. (2001). 대학수학능력시험 영어 영역의 총체적 개선 방안. *영어교육*, 56(2), 333-364.
- 최인철. (1991). 언어(영어) 테스트의 이론과 실제. 이홍수 (편), *영어평가 및 멀티미디어 교육론* (pp. 33-104). 서울: 한국문화사.
- Alderson, J. C., & Urquhart, A. H. (1983). *Reading in a foreign language*. New York: Longman.
- Alderson, J. C., & Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing*, 2, 192-204.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language test*. Oxford: Oxford University Press.
- Beck, J. (1982). Texture segmentation. In J. Beck (Ed.), *Organization and representation in perception* (pp. 285 - 317). Hillsdale, N.J.: Erlbaum.
- Bernhardt, E. B. (1991). *Reading development in a second-language*. Norwood, N.J.: Ablex.
- Carrell, P. & Eisterhold, J. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17, 553-573.
- Carroll, B. J. (1961). *Fundamental considerations in testing for English language*

- proficiency of foreign students*. Washington, D.C.: Center for Applied Linguistics.
- Cziko, G. (1983). Psychometric and edumetric approaches to language testing. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 274-297). Rowley, MA: Newbury House.
- Davison, A. & Kantor, R. N. (1982). On the failure of readability to define readable texts: A case study from adaptations. *Reading Research Quarterly*, 17, 187-209.
- Duffy, T. M. & Kabance, P. (1982). Testing a readable writing approach to text revision. *Journal of Educational Psychology*, 74, 733-748.
- Finch, A. (2005). Students as test-makers. *Modern English Education*, 6(2), 3-23.
- Klare, G. R. (1985). *How to write readable English*. London: Hutchinson.
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Mohammed, M. A. H. & Swales, J. M. (1984). Factors affecting the successful reading of technical instructions. *Reading in a Foreign Language*, 2, 206-217.
- Oller, J. W. & Streiff, .V. (1975). Dictation: A test of grammar based on expectancies. *English Language Teaching*, 30, 25-36.
- Reder, L. M., & Anderson, J. R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology*, 12, 447-472.
- Salager-Meyer, F. (1991). Reading expository prose at the post-secondary level: The influence of textual variables on L2 reading comprehension (a genre-based approach). *Reading in a Foreign Language*, 8, 645-662.
- Savignon, S. J. (1982). Dictation as a measure of communicative competence in French as a second language. *Language Learning*, 32, 33-51.
- Ulijn, J. M. and Strother, J. B. (1990). The effect of syntactic simplification on reading EST texts as L1 and L2. *Journal of Research in Reading*, 13, 38-54.
- Widdowson, H. G. (2001). Communicative language testing: The art of the possible. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara & K. O'Loughlin (Eds.). *Experimenting with uncertainty: Essays in honour of Alan Davies*. Cambridge: Cambridge University Press.

김정렬  
 한국교원대학교 초등교육과  
 충북 청원군 강내면 다락리  
 043)230-3537  
 jrkim@knue.ac.kr

**166**

김정렬

논문접수일자: 2007. 6. 10

게재확정일자: 2007. 7. 20