

수준별 영어 시험 검사지 구성을 위한 외국어 영역 문항 유형의 외적 타당성 분석

김용명*

한국교육과정평가원

강문구

공주대학교

Kim, Yong-Myeong & Kang, Mun-Koo. (2012). The external analysis of the validation on item-types of foreign language (English) domain of CSAT. *Modern English Education*, 13(4), 239-270.

The purpose of this research is to design and compose the two level-differentiated test sets (A and B type) of the 2014 CSAT (College Scholastic Ability Test). To this end, this research performed a questionnaire survey on the test systems and the external validation on 33 item-types of foreign language (English) part of the current 2005 CSAT. Survey participants who were all involved in writing test items for the CSAT included 38 experts (professors and research fellows) and 76 high-school teachers. They were asked online to give their opinions on the current CSAT system. On the basis of results of the survey and the external validation, it constructs the test systems of the 2014 CSAT and comprises an *item-types bank* as an item-types pool from which item-types can be withdrawn to develop the two level-differentiated test sets (A and B type) of 2014 CSAT in terms of the *test set principles*, i.e., complementarity, integration, pivotality, and hierarchicality.

[college scholastic ability test/English/level-differentiated tests /item-types/
대학수학능력시험/외국어(영어)영역/수준별 시험/문항유형]

I. 서론

2012년대 한국 영어 평가에는 세 가지 주요한 변화가 진행 중에 있다. 이 변화의 중심에는 수능 외국어 영역의 듣기 문항 확대 방안, 2014 수능 체제의 수준별 시험, 국가영어능력평가시험(National English Ability Test, 이하 NEAT)의 수

* 제1저자: 김용명, 교신저자: 강문구

능 영어시험 대체(안)이 있다. 수능 영어 영역의 듣기 문항 확대 방안은 현행 외국어(영어) 영역 시험에서 듣기와 읽기 내용 영역 간의 불균형(34%: 66%)을 해소하고, 정부의 실용 영어 강화 정책에 부응하기 위해, 듣기 문항 수를 현행 17개에서 25개로 확대하는 것을 핵심 내용으로 한다(교육과학기술부, 2009; 김용명, 이완기, 김진석, 고현숙, 2010). 또한 2014 수준별 영어 시험은 현행 수리 영역과 같이, 영어도 두 수준, 즉, A형과 B형의 시험을 제공한다. A형은 현행 수능보다 낮은 수준으로 출제 범위는 줄이고, 보다 쉽게 출제하여 수험생의 부담을 최소화하며, B형은 현행 수능 수준과 동일한 출제 범위와 난이도를 유지한다. 이에 따라 수험생은 자신의 수준과 진로에 따라 A형 또는 B형을 선택하여 시험을 볼 수 있다(교육과학기술부, 2011). 마지막으로 NEAT의 수능 대체 여부는 2010년대 한국 영어 평가 변화의 최종 중착점이 될 것이다. 교육인적자원부(2006), 교육과학기술부(2008b)에 따르면, NEAT는 4개(듣기·읽기·말하기·쓰기)의 평가 영역으로 구성되며, 각 영역은 수준별(1급·2급·3급)로 구성된다. 1급은 성인용으로 대학 2, 3학년 수준으로 대학 졸업이나 취업 시에 토익, 토플 등을 대체하는 용도로, 2·3급은 고등학생용으로 대입 전형 자료로 활용될 것이다. 최근 교육과학기술부(2010)에 따르면, NEAT의 수능 외국어(영어) 시험 대체 여부는 여론 수렴 과정을 거쳐 2012년에 최종 결정하여, 2016학년도부터는 수능 영어시험을 대체하는 방안이 추진되고 있다.

현재 진행 중인 한국 영어 평가의 세 변화의 성패는, 김용명(2010a, 2010b)이 언급한 바와 같이, 어떤 문항 유형을 어떻게 조합하여 수준별 영어 시험 및 NEAT에 타당한 검사지를 구성할 수 있느냐에 달려있다. 김용명(2010a)은 Bachman과 Palmer(1996)의 시험 유용성 모델을 토대로 문항 유형 결정 원리를 제안하였으며, 또한 지난 6년간 수능 외국어 영역의 문항 분석 자료의 심층 분석을 토대로 검사지 구성 원칙을 제안하였다. 한편 김용명 외 3인(2010)은 현행 수능 외국어(영어) 영역의 개선 방향으로 고착화되어 있는 문항 유형을 개선하여 시험 요령이 통하지 않도록 신유형의 문항을 개발하거나 기존 유형을 재구성할 것을 제한했다. 이와 같은 선행 연구 결과에 따라 현행 수능 외국어(영어) 영역의 문항 유형에 대한 문제점을 진단하고, 이를 수준별 영어 시험, 더 나아가 NEAT 검사지 구성에 반영할 필요가 있다.

이 같은 연구 필요성에 부응하고, 또한 수준별 영어 시험 및 NEAT의 검사지 구성에 시사점을 제공하기 위해, 본 연구는 현행 수능 외국어(영어) 영역의 문항 유형에 대한 외적 타당성을 분석하고¹, 이를 토대로 2014 수준별 영어 시험

¹ 문항 유형의 타당성 평가는 외적 평가와 내적 평가로 대별할 수 있다. 외적 타당성 평가는 평가자의 경험과 직관에 따라 문항 유형을 총체적(holistic)으로 평가하는 것을 말한다. 내적 타당성 평가는 문항 유형 평가의 평가 요소(예, 연계성, 실제성, 상호 작용성, 환류 효과성, 출제 용이성 등)에 따라 각 문항 유형을 분석적(analytic)으로 평가하는 것을 말한다. 본 연구는 문항 유형에 대한 외적 평가에 초점을 두고 있다.

의 문항 유형 은행을 구성하고자 한다. 이 같은 연구 목적을 구체화하기 위해 본 연구는 다음과 같은 세 가지 연구 주제를 설정한다.

- (a) 현행 외국어(영어) 영역의 문항 유형에 대한 외적 타당성 분석
- (b) 2014학년도 수능 체제의 수준별 영어 영역의 시험 체제 구성
- (c) 수준별 영어 시험의 검사지 구성을 위한 문항 유형 은행 구성

이 같은 연구 주제에 대한 가능한 안을 탐색하기 위해, 제 2장에서는 시험의 질을 결정하는 문항 및 문항 유형을 분석할 수 있는 이론적 틀로서 **Bachman**과 **Palmer(2010)**가 제안한 문항 특성을 살펴볼 것이며, 또한 검사지 구성에 관한 이론적 틀로서 문항 유형 결정 원리(**김용명, 2010a**)와 검사지 구성 원칙(**김용명, 2010b**)에 관해 설명할 것이다. 제 3장에서는 최근 5년간 수능 출제 참여자를 대상으로 현행 외국어(영어) 영역의 약 33개 문항 유형에 대한 외적 타당성 평가(설문조사)를 하고, 그 결과에 관해 논의할 것이다. 제 4장에서는 현행 외국어(영어) 영역의 문항 유형에 대한 외적 타당성 평가를 토대로 2014학년도 수능 체제의 수준별 영어 영역의 시험 체제를 구성할 것이며, 이어서 수준별 영어 시험(A·B형)의 검사지 구성을 위한 문항 은행을 제시할 것이다. 마지막 제 5장에서는 결론 및 시사점에 관해 논의할 것이다.

II. 이론적 배경

1. 문항 유형 특성 체계

시험의 질은 문항을 질을 능가할 수 없다. 이는 문항의 특성이 시험의 타당성을 결정하기 때문이다. **Bachman**과 **Palmer(2010)**는 문항 특성과 시험 자질과의 관계를 분석할 수 있는 시험 과업 특성 틀을 제안했다². 시험 과업 특성 틀은, 표 1에서 보듯이, 5개의 구성 요소로 구성된다.

먼저 시험 환경 특성은 시험 환경과 관계된 특성을 말하는데, 시험장의 환경(위치, 소음, 조명 등), 시험 참가자(수험생, 감독관) 등이 이 특성에 속한다. 둘째, 시험 체제 특성은 시험 절차와 시험 구성과 관련된 특성을 말하는데, 지시문, 시험지의 구조, 시험 시간, 점수화 방법 등이 이 특성에 속한다. 셋째, 입력 특성에는 시험의 방법과 관련된 입력 양식과 시험 내용과 관련된 입력 언어가 있다. 입력 양식에는 채널, 형식, 언어, 길이/시간, 수단, 속도, 유형 등이 있다.

² **Bachman**과 **Palmer(2010)**는 과업을 시험 과업과 언어 사용에서의 목표 언어 과업 나누어 기술하고 있으나, 본 연구에서는 논의의 편의를 위해 시험 과업만 논의하고자 한다.

채널은 시험 내용을 제시하는 방법을 말하는데, 시험 내용을 구두로, 시각으로, 또는 두 방식 모두로 제시할 수도 있다. 형식은 시험 내용을 제시하는 형식을 말하는데, 이를 언어로, 비언어적으로, 또는 둘 다로 제시할 수 있다. 언어는 시험 내용을 언어로 제시할 경우, 모국어나 목표어로 또는 둘 다로 제시할 수 있다. 수단은 입력을 전달하는 수단을 말하는데, 목소리로, 녹음으로, 또는 둘 다로 전달할 수도 있다. 길이/시간은 문항 자료의 길이 또는 이를 처리하는데 걸리는 시간을 말한다. 유형은 시험 문제의 형식을 말하는데, 여기엔 문항과 서술형이 있다. 문항은 선택지에서 정답을 선택하는 선택형과 단어나 문장으로 답하는 단답형으로 구분된다. 서술형은 답화 수준의 확장된 응답을 요하는 것을 말한다. 입력 언어는 언어 특성과 화제적 특성으로 구분할 수 있다. 언어 특성은 채고자 하는 언어 능력을 말하는데, 문법과 답화 구조와 관련된 조직적 특성과 언어의 기능과 사용과 관련된 화용적 특성으로 대별된다. 화제적 특성은 시험이 담고 있는 주제 또는 화제와 관련된다. 넷째, 예상된 응답의 특성은 답지의 구성 방법과 답지의 내용과 관련된 것으로 입력 특성과 동일한 내용을 담고 있어 논의를 생략하고자 한다. 마지막, 입력과 응답의 관계는 입력과 예상된 응답이 어떻게 상호 관련되는가를 말하는데, 상호작용, 범위, 직접성에 따라 그 관계를 파악할 수 있다. 상호 작용은 입력과 응답이 후속 입력과 응답에 직접적으로 영향을 주는 정도로 정의할 수 있는데, 영향의 정도에 따라 상호작용 과업, 적응형 과업, 비상호작용 과업으로 구분할 수 있다. 상호작용 과업은 구두 면접에서와 같이, 면접관의 입력에 대해 응답을 하면, 이 응답에 대해 피드백이 주어지고, 후속 응답이 이에 영향을 받게 되는 경우를 말한다. 적응형 과업은 입력(문항)이 주어지면, 수험생이 이를 해결하면, 더 높은 난도의 문항을 제시하고, 그렇지 못하면, 더 낮은 난도의 문항을 제시하는 경우를 말한다. 비상호작용 과업은, 에세이와 같이, 입력과 응답 사이에 피드백과 상호작용도 없는 경우를 말한다. 범위는 응답을 찾기 위해, 수험생이 처리해야 할 정보의 양과 범위를 말하는데, 넓은 범위와 좁은 범위로 대별할 수 있다. 넓은 범위는 요지 추론 문항과 같이, 응답을 찾기 위해 주어진 입력을 모두 처리해야 하는 경우를 말하며, 좁은 범위는 세부 사항 찾기 문항과 같이, 응답을 하기 위해, 입력의 특정 부분만 처리하는 경우를 말한다. 마지막으로 직접성은 응답을 하기 위해 주어진 입력에 의존하는 정도를 말하는데, 직접 관계와 간접 관계로 대별할 수 있다. 직접 관계는 지문 안에서 응답을 찾을 수 있는 경우를 말하며, 간접 관계는 지문 밖의 정보(예, 배경지식)를 이용해야 답을 찾을 수 있는 경우를 말한다.

표 1

과업 특성 틀(Bachman & Palmer, 2010, pp. 66-68)

<p>I 시험 환경 특성(Characteristics of the setting)</p> <p>A 시험장 물리적 환경(Physical characteristics)</p> <p>B 시험 참가자(Participants)</p> <p>C 시험 시간(Time of task)</p>	
--	--

<p>II 시험 체제 특성(Characteristics of the rubric)</p> <p>A 지시문(Instructions)</p> <ol style="list-style-type: none"> 1. 언어(Language (native, target, both)) 2. 채널(Channel (aural, visual)) 3. 시험 및 점수화 절차(Specification of structures, procedures to be followed by test-takers, and procedures for producing assessment records) <p>B 시험지 구조(Structure)</p> <ol style="list-style-type: none"> 1. 부분/과업 수(Number of parts/tasks) 2. 부분/과업 구분(Salience of parts/tasks) 3. 부분/과업 순서(Sequence of parts/tasks) 4. 부분/과업 상대적 중요도(Relative importance of parts/tasks) 5. 부분/과업 당 문항 수(Number of tasks/items per parts) <p>C 시험 시간(Time allotment)</p> <p>D 점수화 방법(Recording method)</p> <ol style="list-style-type: none"> 1. 점수화 형식(Types of assessment record (score, description)) 2. 정오의 기준(Criteria for correctness) 3. 점수화 절차(Procedures for producing an assessment record) 4. 점수화 방법(Recorders (scorers/raters, describers)) 	
--	--

<p>III 입력 특성(Characteristics of the input)</p> <p>A 입력 양식(Format)</p> <ol style="list-style-type: none"> 1. 채널(Channel (aural, visual, both)) 2. 형식(Form (language, non-language, both)) 3. 언어(Language (native, target, both)) 4. 길이/시간(Length/time) 5. 수단(Vehicle (live, reproduced, both)) 6. 속도(Degree of speededness) 7. 유형(Type (item, prompt, input for interpretation)) <p>B 입력 언어(Language of input)</p> <ol style="list-style-type: none"> 1. 언어 특성(Language characteristics) <ol style="list-style-type: none"> a) 조직적 특성(Organizational characteristics) <ol style="list-style-type: none"> (1) 문법적 특성(Grammatical (vocabulary, syntax, phonology/graphology)) (2) 담화적 특성(Textual (cohesion, organization(rhetorical/conversational)) b) 화용적 특성(Pragmatic characteristics) <ol style="list-style-type: none"> (1) 기능적 특성(Functional (ideational, manipulative, heuristics, imaginative)) (2) 사회언어적 특성(Sociolinguistic (genre, dialect/variety, register, naturalness, cultural references, and figures of speech)) 2. 화제적 특성(Topical characteristics) 	
---	--

 IV 예상된 응답의 특성(Characteristics of the expected response)

A 양식(Format)

1. 채널(Channel (aural, visual))
2. 형식(Form (language, non-language, both))
3. 언어(Language (native, target, both))
4. 길이/시간(Length/time)
5. 문제 유형(Type (selected, limited production, extended production))
6. 속도(Degree of speededness)

B 언어 응답(Language of expected response)

1. 언어 특성(Language characteristics)
 - a) 조직적 특성(Organizational characteristics)
 - (1) 문법적 특성(Grammatical (vocabulary, syntax, phonology/graphology))
 - (2) 담화적 특성(Textual (cohesion, organization(rhetorical/conversational))
 - b) 화용적 특성(Pragmatic characteristics)
 - (1) 기능적 특성(Functional (ideational, manipulative, heuristics, imaginative))
 - (2) 사회언어적 특성(Sociolinguistic (genre, dialect/variety, register, naturalness, cultural references, and figures of speech))
 2. 화제적 특성(Topical characteristics)
-

V 입력과 응답의 관계(Relationship between input and response)

A 상호작용(Type of external interactiveness: reciprocal, non-reciprocal, adaptive)

B 범위(Scope of relationship (broad, narrow))

C 직접성(Directness of relationship (direct, indirect))

이상에서 논의한 문항 특성 틀이 수준별 영어 시험의 문항 선별 및 검사지 구성에 줄 수 있는 시사점에 대해 살펴보자. 먼저 문항 유형 특성 틀은 수준별 영어 시험에 포함될 문항 유형의 특성과 실제 언어 사용 상황에서의 실제 과업의 특성이 상호 일치하는 정도, 즉 실제성의 정도를 파악할 수 있는 수단을 제공해준다. 따라서 수준별 영어 시험의 검사지를 보다 실제성이 높은 문항 유형으로 구성할 수 있을 것이며, 이는 수준별 영어 시험에서의 수행의 정도(시험 결과)와 실제 사항에서의 언어 수행력의 정도와 일치할 가능성을 높여 줄 것이다. 또한 문항 특성 틀을 일종의 검사표로 활용하여, 현행 외국어(영어) 영역의 문항 유형의 타당성을 평가하여, 비 시험 상황에서의 언어 사용을 목표로 하는 TLU 과업과의 일치의 정도가 낮은 문항 유형(실제성이 낮은 유형)과 수험생의 언어 능력과 관계하는 정도가 낮은 유형(상호작용성이 낮은 유형)을 배제할 수 있을 것이다³. 마지막으로 문항 유형 특성 틀에 따라 국내·외 표준화 시험(NEAT, TEPS, TOEIC 등)의 문항 유형을 분석하여, 수준별 영어 시험에 부합하는 유형을 선별하거나 문항 유형 특성의 구성 요소 중, 한 두 개를 변형하여, 수준별 영어 시험의 적합한 새로운 유형을 개발할 수 있을 것이다.

³ Bachman과 Palmer(2010, p. 60)는 수험생이 시험 상황 밖에서 언어 사용 과업을 수행해야 할 특정 상황을 목표 언어 사용 영역(Target Language Use(TLU) Domain)으로 정의했으며, 이 특정 TLU 영역에서의 언어 사용 과업을 TLU 과업으로 정의했다. TLU 과업은 시험 수행 결과(점수)를 시험 상황을 넘어 일반화하는 것과 관계된다.

2. 문항 유형 결정 원리와 검사지 구성

김용명(2010a)은 NEAT 구성에 관한 밑그림을 제시하기 위해, “어떤 원리에 따라 NEAT에 적합한 문항 유형을 개발하고 선별할 것인가?”라는 연구 질문에 대한 가능한 답으로 Bachman과 Palmer(1996)의 시험 유용성 모델을 토대로 ‘문항 유형 결정 원리’를 제안했다(pp. 377-385). 문항 유형 결정 원리는 연계성(connection), 실제성(authenticity), 상호작용성(interactiveness), 환류효과성(washback effectiveness)으로 구성된다.

연계성은 수준별 영어 시험의 평가 목표, 평가 내용, 수준 및 평가 요소는 개정 영어과 교육과정(교육과학기술부, 2008a)의 교육 목표, 교육 내용, 성취 기준과 연계되어야 한다는 원리를 말한다. 따라서 연계성 원리는 수준별 영어의 수준별(A·B형), 기능별(듣기·읽기) 시험의 출제 체제 구성에 논리적 타당성을 제공하며, 특히 각 수준별 시험의 출제 범위, 평가 내용, 평가 기준 설정에 논리적 근거를 제시해 준다. 실제성은 시험 과업의 특성과 실제 언어 사용 상황에서의 TLU 과업의 특성은 상호 일치해야 한다는 원리로 문항 유형 개발 및 선별의 범위와 한계를 설정하는 역할을 한다. 따라서 실제성 원리는 수준별 영어 시험의 평가 목표 분류표 상의 내용 영역 및 행동 영역 체계의 주요 구성소를 설정하고, 이의 타당성을 검증하는 수단이 된다. 상호작용성은 시험 과업과 학습자의 언어능력이 상호작용하는 정도를 말하는데, 상호작용성이 높은 문항일수록 언어능력과 관계하는 정도가 높으므로 상호작용성 원리는 문항의 질적 통제 역할을 한다. 마지막으로 환류 효과성은 시험의 시행 결과는 교육과정, 교수·학습 활동, 평가에 긍정적 영향을 주어야 한다는 원리로 교육과정, 교수·학습, 교육 평가 간의 피리를 줄이고, 일체화에 기여한다.

3. 검사지 구성 원칙

김용명(2010a)은 NEAT 구성에 관한 설계도를 제시하기 위해, “어떤 원칙에 따라 어떤 문항 유형으로 NEAT에 타당한 검사지를 구성할 것인가?”라는 연구 질문을 제기하고, 이에 대한 가능한 답으로서 지난 6년간 수능 외국어 영역의 문항 분석 자료의 심층 분석을 토대로 검사지 구성 원칙을 제안하였다. 검사지 구성 원칙은 상보성(complementarity), 통합성(integration), 주축성(pivotality), 위계성(hierarchicality) 등으로 구성된다.

상보성은 문항 또는 문항 유형이 언어능력, 내용 영역 및 행동 영역의 각 구성소가 서로 상보적 분포를 이룰 수 있도록 검사지를 구성해야 한다는 원칙이다. 다시 말해, 문항 a, b, c가 평가하고자 하는 언어능력의 구성소, 재고자 하는 내용 영역의 구성소, 측정하고자 하는 행동 영역의 구성소 각각에 대하여 상보적 관계에 있을 때 상보성 원칙을 만족시킨다고 할 수 있다. 예를 들며, 내용 일치, 빈칸 추론, 목적 추론 문항은 사실적 이해력, 추론적 이해력, 종합적 이해

력을 각각 측정하고, 또한 각각 상향식, 상호작용식, 하향식 읽기 이해 능력을 평가한다는 점에서 상보적 분포를 이루고 있으므로 상보성 원칙을 만족시킨다. 반면, 주제, 제목, 요지 추론 문항은 모두 추론적 이해력을 측정하고, 또한 모두 하향식 읽기 능력을 평가한다는 점에서 동일한 능력을 측정하고 있으므로 상보성 원칙에 위배된다. 이런 점에서 상보성은 ‘필수 불가결한 문항을 필요 최소한’으로 검사지를 구성해야 한다는 원칙으로 이해할 수 있으며, 따라서 각 수준별, 기능별 검사지에 포함될 적정 문항 유형의 수를 결정하는데 논리적 타당성을 제공할 수 있다. 또한 상보성 원칙은 동형 문항에 대해서는 상호 배타성을 가지므로 시험의 동등화와 관련하여 동형 검사형을 구성하는 역할을 한다.

통합성은 각 기능별 시험(듣기, 읽기, 말하기, 쓰기)에서 해당 기능의 능력을 측정하는 기능 독립형 문항으로 검사지를 구성하되, 기능 통합형(예, 듣기와 말하기 통합) 또는 기능 연계형 문항(읽기와 쓰기를 연계)도 필요 최소한으로 검사지에 포함되어야 한다는 원칙을 말한다. 이에 대한 논리적 근거는 최근 교수이론(Bachman & Palmer, 1996; Brown, 2007; Ellis, 2003)과 영어과 교육과정(교육과학기술부, 2008a)에서 강조하고 있는 4기능의 통합 교수 원리에서 찾을 수 있다.

주축성은 주축 문항을 시험의 향상성과 동등성을 유지할 만큼 필요 최대한으로, 시험의 다양성을 해치지 않을 만큼 필요 최소한으로 검사지 구성에 포함해야 한다는 원칙으로 정의할 수 있다. 수능 외국어(영어)의 문항 분석 자료⁴에 따르면, 문항 유형을 주축 문항(pivot item)과 주변 문항(peripheral item)으로 구분할 수 있다. 주축 문항은 평균 정답률을 기준으로 정답률의 편차(variation)가 상대적으로 작으며, 문항의 복잡도(어휘적, 언어적, 개념적, 인지적 복잡도)가 높아지면, 이에 따라 난이도도 올라가는 경향을 갖는다(문항의 복잡도와 난이도는 정비례 관계). 이런 경향성에서 주축 문항은 학습자의 언어능력에 지배를 받는다는 것을 추론할 수 있으며, 시험의 향상성 유지와 동등화에 기여할 수 있다는 점을 예상할 수 있다. 반면 주변 문항은 정답률의 편차가 상대적으로 크고, 문항의 복잡도가 높아지면, 일정 수준까지 난이도가 올라가지만, 그 수준 이상에서 더 이상 난이도의 변화가 없는 시험 고원(testing plateau) 현상이 생겨나는 경향을 보인다. 이런 점에서 주변 문항은 언어 능력보다는 시험 요령, 학습자의 정의적 특성, 배경 지식 등의 지배를 받는다는 것을 추론할 수 있으며, 시험의 다양성에 기여할 수 있다는 점을 예상할 수 있다.

위계성은 Y. M. Kim(2006, 2007a, 2007b)의 시험가능성(testability)⁵ 개념과 수능

⁴ 한국교육과정평가원의 2005년부터 2010까지의 수능문항 분석 비공개 자료집에 근거하였다.

⁵ 시험 가능성(Testability) 가설은 Piennemann(1985, 1998)의 교수 가능성(teachability) 가설을 근거한 것으로 시험 과업은 학습자(수험생)의 현재의 언어 발달 단계보다 한 단계 다음 단계의 규칙과 구조와 관련될 때, 그 과업을 수행 가능하다고 할 수 있다는 것을 말한다(Y. M. Kim, 2007b, p. 51).

외국어(영어) 영역의 문항 분석 결과를 토대로 구성된 개념으로 검사지를 구성하고 있는 각 문항의 복잡도 또는 난이도는 각 수험생(또는 각 집단)의 수행가능 단계와 일치할 수 있도록 위계화 되어야 한다는 원칙을 말한다. 외국어(영어) 영역의 문항 분석에 따르면, 각 문항 유형에 대한 수험생(학습자)의 문항 반응 곡선은 해당 문항 유형의 특성에 따라 L형, M형, H형으로 정형화할 수 있다⁶. L형 문항 반응 곡선은 하위 등급(예, 9, 8, 7등급)에서는 등급이 올라감에 따라 정답률도 올라가지만(등급과 정답률이 정비례 관계에 있지만), 일정 등급(예, 7등급) 이상을 넘어서면 정답률(예, 80%대)이 고정되어 시험 고원을 형성하는 문항 유형을 말한다. L형 반응 곡선의 특성을 보이는 문항 유형에는 a, b, c 등의 유형이 있으며, 이 유형들은 하위 등급 간의 난이도 및 변별도 조정에 결정적 역할을 할 것이다⁷. 한편 M형의 문항 반응 곡선은 하위 등급(예, 7등급 이하)에서 정답률이 정체(예, 20%대)되어 시험고원을 형성하지만, 일정 등급(예, 7등급)에 도달하면 등급이 올라감에 따라 정답률이 올라간 후, 다시 일정 등급(예, 3등급)을 넘어서면 정답률이 정체(예, 80%대)되어 또 하나의 시험 고원을 형성하는 유형을 말한다. M형 곡선의 특성을 보이는 문항 유형에는 o, p, q, r 등의 유형이 있으며, 이런 문항 유형들은 중위 등급 간의 난이도 및 변별도 조정에 결정적 역할을 할 것이다. 반면, H형의 문항 반응 곡선은 일정 등급(예, 3등급)까지는 정답률이 정체(예, 30%대)되어 시험 고원을 형성하지만, 일정 등급(예, 3등급)을 넘어서면 등급과 정답률이 정비례 관계에 있는 유형을 말한다. H형의 문항 반응 곡선의 특성을 보이는 문항 유형에는 x, y, z 등의 유형이 있으며, 이들 문항 유형은 상위 등급 간의 난이도 및 변별도 조정에 결정적 역할을 할 것이다. 따라서 위계성은 한 검사지에서 L형, M형, H형의 구성 비율을 적절히 조정함으로써 검사지 총체적 난이도와 변별도를 적정 수준으로 통제할 수 있고, 더 나아가 시험의 안정성과 향상성을 유지할 수 있고, 시험의 동등화에도 기여할 수 있다. 예를 들어, L형에 속하는 문항 유형(a, b, c)의 비율을 늘리면, 검사지의 총체적 난이도는 내려갈 것이며, 그 결과 하위 학습자(수험생)에 대한 변별력이 높아지는 경향을 보일 것이다. 역으로 H형에 속하는 문항 유형(x, y, z)의 비율을 늘리면, 검사지의 총체적 난이도는 올라갈 것이며, 그 결과 상위 학습자(수험생)에 대한 변별력이 높아지는 경향을 보일 것이다.

⁶ 문항 반응 곡선 L형, M형, H형은 각각 low, mean, high의 첫 글자를 딴 것으로 하위 학습자, 중위 학습자, 상위 학습자를 보다 잘 변별한다는 의미를 담고 있다.

⁷ 수능 외국어(영어) 영역 시험의 문항 분석을 통해, L형, M형, H형에 속하는 문항 유형을 식별해낼 수 있었지만, 수능 자료가 보안인 관계로 이를 구체적으로 명시하지 않고, a, b, c 등으로 표시하였다.

4. 모듈형(Modular) 평가 목표 분류표

고정형 평가목표 이원목적분류표는 현행 수능 체제와 같이 단일 수준의 단일 시행 체제에서는 시험의 항상성과 동일성을 유지하기가 비교적 용이하고, 출제 과정도 비교적 안정적으로 운영할 수 있는 장점이 있다. 그러나 수준별 영어 시험과 NEAT와 같이 수준별(A·B형 또는 2·3급), 기능별(듣기·읽기·말하기·쓰기) 복수 시행 체제에서는 매 시행되는 수준별, 기능별 시험 간에 등거리성을 확보하기가 어렵고, 각 문항 간에 상보성을 파악하기가 용이하지 않아, 각 시험의 평가 요소 간의 간접 효과가 있을 수 있다(김용명, 2010b, pp. 171-178).

고정형 평가목표 이원분류표의 이 같은 한계점을 완화하기 위해, 2014 수능 체제의 수준별 영어는 모듈형 평가목표 분류표에 따른다. 모듈형 평가목표 분류표는 실제성 원리에 따라 평가목표 분류표의 각 영역과 그 하위 영역의 구성소를 가능한 한 세분하여 구성하지만, 상보성 원칙에 따라 각 영역과 그 하위 영역 간에 서로 상보적 분포를 이루도록 구성한 것을 말한다. 모듈형 평가목표 분류표는 현행 수능 및 교육 현장에서 활용되고 있는 고정형 평가목표 이원목적분류표의 내용 및 행동 영역에 더하여 문항 유형 특성 체계라는 별도의 독립된 체계를 두고 있다(김용명, 2010b). 문항 유형 특성 체계의 구성 요소는 검사지 구성 원칙, 즉, 상보성, 통합성, 주축성, 위계성으로 구성된다(전술한 검사지 구성 원칙 참조).

이 같이 구성된 모듈형 평가목표 분류표는 고정형 이원 분류표의 한계점을 극복할 수 있을 뿐만 아니라 수준별(A·B형, 3·2급), 기능별(듣기·읽기) 시험 간에 상보성 파악이 용이하고, 수준별(A·B형), 기능별(듣기·읽기) 시험 간 등거리성을 확보할 수 있음으로써 평가 요소 간의 상호 간접 효과를 최소화할 수 있으며, 또한 문항 유형 은행식 또는 문항 공모식 출제 체제 하에서 문제은행 시스템 또는 문항 개발의 매트릭스 역할을 할 수 있다.

III. 현행 외국어 영역의 문항 유형에 대한 외적 타당성 분석

1. 현행 외국어 영역 시험의 문항 유형

현행 수능 외국어(영어)영역의 문항 수는 총 50문항이며, 이 중 듣기 평가 문항 수는 총 17문항(간접 말하기 5문항)이며, 읽기 평가 문항 수는 총 33문항(간접 쓰기 5문항)으로 구성되어 있다.

듣기 평가는, 표 2에서 보는 바와 같이, 총 13개 유형으로 구성되어 있어 다양한 유형으로 듣기 평가가 이루어지고 있음을 알 수 있다. 행동 영역별로 살펴보면, 사실적 이해가 가장 많은 평균 7문항이며, 보다 복잡한 사고력을 요구하는 적용력, 종합적 이해와 추론적 이해는 각각 5문항, 2문항, 3문항이다. 이를

통해 듣기 평가의 난도가 별로 높지 않음을 예측할 수 있다.

또한 표 3에서 보는 바와 같이, 읽기 평가는 총 20개 유형으로 구성되어 있으며, 매우 다양한 유형으로 읽기 평가가 이루어지고 있음을 알 수 있다. 행동 영역별로 살펴보면, 논리적 복합적 사고력을 요하는 추론적 이해와 적용력을 묻는 문항이 각각 14문항, 5문항이며, 사실적 이해를 묻는 문항은 단 4문항에 지나지 않는다. 이를 통해 읽기 평가가 듣기 평가보다 난이도가 높음을 유추할 수 있다.

표 2
현행 외국어(영어)영역의 듣기 문항 유형

내용 영역	유형	문항수	행동 영역			적용
			사실적	종합적	추론적	
듣기	그림 설명	1	o			
	화자의 심정	1		o		
	대화 지시대상	1	o			
	한/할/부탁한 일	3	o			
	담화 내용 일치	1	o			
	담화 목적	1		o		
	숫자 관련 정보	1	o			
	대화자 관계	1			o	
	대화 장소	1			o	
	시각 자료 활용	1			o	
말하기	그림에 적절한 대화	1				o
	마지막 말에 대한 응답	3				o
	상황에 적절한 응답	1				o
계	13유형	17	7	2	3	5

표 3
현행 외국어(영어)영역의 읽기 문항 유형

내용 영역	유형	문항 수	행동 영역				적용
			문법	어휘	사실적	추론적	
읽기	지칭 추론	1				o	
	어법(밑줄형/선택형)	2	o				
	어휘(밑줄형/선택형)	2		o			
	빈칸추론(단어/구/절)	6				o	
	연결사 추론	1				o	
	도표의 내용 일치	1			o		
	내용일치/불일치	2			o		
	주제	2				o	
	제목	2				o	
	요지	2				o	
	글의 목적	1					o

	필자의 주장	1			0		
	분위기/심경/어조	1			0		
쓰기	무관한 문장 찾기	1			0		
	주어진 문장 삽입	1			0		
	글의 이어질 순서	1			0		
	문단 요약	1			0		
읽기/ 쓰기	1지문 2문항	1		0	0		
	1지문 3문항	1		0	0	0	
계	20유형	33	2	2	4	14	5 5

2. 현행 외국어 영역 문항 유형의 외적 타당성에 관한 설문 조사

1) 현행 외국어 영역 문항 유형의 타당성에 관한 설문 대상과 방법

현행 외국어(영어) 영역 문항 유형의 타당성에 관한 설문 대상자는 과거 5년에 걸쳐 수능 또는 문제은행 출제 참여자로 한정하였으며, 지역별, 연령별로 임의로 표집 하였고, 온라인 방식(구조화된 설문지에 자기 기입 방식)으로 설문 조사를 하였다. 설문 대상자는, 표 4에서 보는 바와 같이, 전문가(교수) 29명, 교사 76명, 연구원 및 장학사 9명으로 총 114명이었다. 성별은 남성이 74.6%, 여성이 25.4%이었다. 연령별로는 40대가 가장 많은 53.5%였다. 경력은 15년 이상이 50.9%로 과반 이상을 차지하였다. 지역별로는 서울시가 28.4%, 광역시 35.8%, 중소도시 27.5%, 읍·면 지역은 8.3%였다.

설문 결과는 SPSS Win 12 프로그램을 이용하여 빈도 분석을 통해 전반적인 응답 내용의 분포를 관찰하고, 교차 분석, 평균 분석, 집단별 분석을 하였다.

표 4
설문대상자의 인적 정보

	전 체	114	100.0
대상별	전문가(교수)	29	25.4
	교사	76	66.7
	연구원 및 장학사	9	7.9
성별	남자	85	74.6
	여자	29	25.4
연령	30대	23	20.2
	40대	61	53.5
	50대	30	26.3
학력	대졸	16	14.5
	석사	56	49.1
	박사	42	36.8

경력	5년 이상-10 미만	13	11.4
	10년 이상-15년 미만	42	36.8
	15년 이상	58	50.9
거주 지역	특별시	32	28.4
	광역시	41	35.8
	중소 도시지역	31	27.5
	읍면 및 기타	10	8.3

2) 현행 외국어 영역 문항 유형의 타당성에 관한 설문 내용

현행 외국어(영어) 영역 문항 유형의 타당성 분석을 위한 설문 내용은 다음과 같다. 먼저 영어 교육 전문가(교수), 교사, 연구원 등의 배경 변인을 포함하였으며, 통상 설문 조사에서와 마찬가지로 성별, 연령, 지역, 학력 등에 대해 설문하였다. 둘째, 듣기 문항이 17문항에서 25문항으로 확대되고, 읽기 문항이 33문항에서 25문항으로 축소됨에 따라 현행 외국어(영어) 영역의 문항 유형 33개(듣기 13개 유형, 읽기 20개 유형)에 대한 외적 타당성을 설문하였다. 셋째, 듣기 및 읽기 영역에서 새로운 유형의 도입할 것인지, 또 도입한다면 몇 문항이 적절한지 등을 조사하였다. 넷째, 듣기 영역에서도 현행 읽기 영역과 같이 1대화·담화문 2문항을 도입할 것인지, 도입 한다면 몇 문항이 타당한지를 조사하였다. 다섯째, 현행 듣기 영역에서 간접 말하기 문항의 확대 여부를 설문하였다. 마지막으로 현행 듣기 문항의 녹음 속도가 적절한지, 듣기 문항을 몇 번 들려주는 것이 타당한지 등을 설문하였다(부록 참조).

3. 현행 외국어 영역 문항 유형의 외적 타당성에 관한 설문 결과 분석

듣기 문항 확대 방안에 따른 듣기 및 읽기 영역 33개 문항 유형에 대한 설문 결과 및 1대화·담화문 2문항 도입 여부, 간접 말하기 문항의 확대 여부, 듣기의 녹음 속도 및 듣기 횟수 등에 대한 설문 결과를 제시하면 다음과 같다.

첫째, 듣기 영역 13개 유형 중, 삭제할 필요가 있다면 어느 유형을 수준별 영어 검사지 구성에서 삭제해야 한다고 생각하는가라는 질문에 대한 응답 결과는, 그림 1에서 보는 바와 같이, ‘화자의 심정 파악(43.9%)’, ‘그림의 상황에 맞는 대화(41.2%)’, ‘대화 장소 추론(32.5%)’, ‘대화로 설명된 대상 찾기(19.3%)’, ‘대화자 관계 추론 (18.4%)’ 등의 순이었다. 삭제의 이유로는 타당성과 실제성이 낮아서, 환류 효과성이 없어서, 출제의 용이성이 낮아서 등을 들었다.

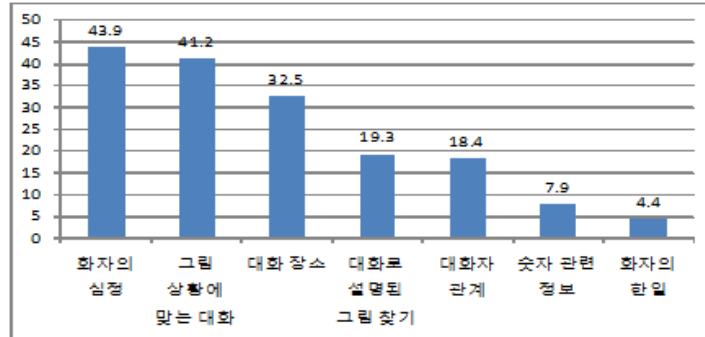


그림 1 듣기 문항 유형 중 삭제할 필요가 있는 유형

둘째, 읽기 영역 20개 유형 중, 삭제할 필요가 있다면 어느 유형을 검사지 구성에서 삭제해야 한다고 생각하는가라는 질문에 대한 응답 결과는, 그림 2에서 보는 바와 같이, ‘필자의 심경(45.6%)’, ‘글의 제목(31.6%)’, ‘글의 요지(24.6%)’, ‘주제 파악(21.9%)’, ‘지칭 추론 (14.9%)’, ‘도표의 세부 내용(14.0%)’, ‘글의 목적(13.2)’ 등의 순이었다. 삭제의 이유로는 실제성과 환류 효과성이 낮아서, 두 문항씩 존재하므로 등을 제시했다.

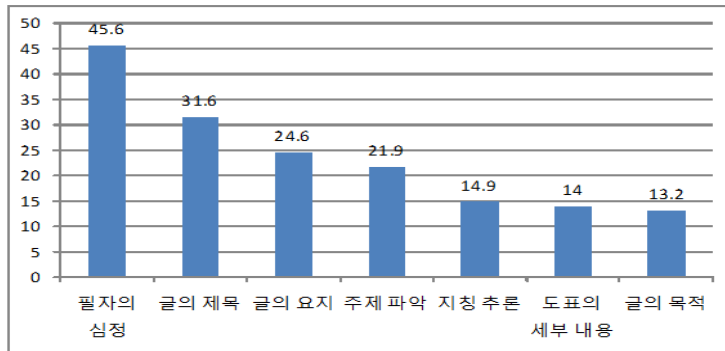


그림 2 읽기 문항 유형 중 삭제할 필요가 있는 유형

셋째, 1대화·담화문 2문항 도입에 대해서는, 그림 3에서 보는 바와 같이, 응답자의 88.6%가 1대화·담화문 2문항의 유형을 도입하는 것에 대해 찬성하였으며, 적당한 도입 세트는 평균 2.2개로 조사되었다. 대화·담화문 2문항 도입에 찬성하는 이유로는 ‘변별력 높은 문항 제작이 가능해서(15.8%)’, ‘실제 언어 사용 상황을 반영할 수 있어서(7.9%)’, ‘심층적인 듣기 능력 테스트에 도움이 되어서(6.9%)’ 등의 순으로 나타난 반면, 도입에 반대하는 이유로는 ‘기억력 테스트가 될 가능성 때문에(25.0%)’, ‘도·농 간의 격차가 때문에(8.3%)’, ‘난이도가 상승하므로(8.3%)’, 사교육을 야기할 것이므로(8.3%) 등의 순이었다.

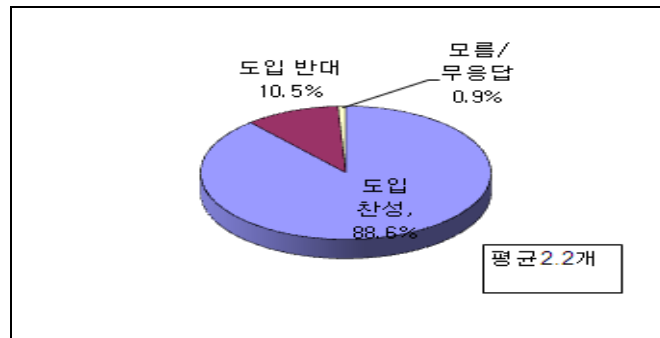


그림 3 1대화·답화 2문항 도입에 대한 설문 결과

넷째, 1대화·답화문 2문항 들려주기 횟수에 대해서는, 그림 4에서 보는 바와 같이, 2번 들려주는 것에 찬성하는 의견이 1번 들려주는 것에 찬성(43.0%)하는 것보다 9.6% 높은 52.6%였다. 두 번 들려주는 것에 찬성하는 이유는 ‘학생들의 심리적 부담을 덜어줄 수 있어서(18.3%)’이고, 한 번 들려주는 것에 찬성하는 이유는 ‘실제 대화는 한 번만 이루어지기 때문(20.4%)’으로 응답하였다.

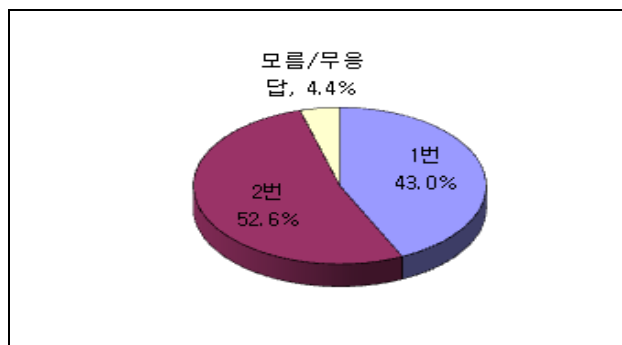


그림 4 1대화·답화문 2문항을 들려주는 횟수에 대한 설문 결과

다섯째, 현행 듣기 영역에서 (간접)말하기 5개 문항보다 늘리는 것에 찬성한다는 의견은, 그림 5에서 보는 바와 같이, 현행 유지(0.9%)보다 90.6배나 높고, 줄여야 한다는 의견보다는 6.6배나 높은 81.6%로 나타났다. 현행보다 문항을 ‘늘여야 한다’고 응답한 경우는 평균 4개를 늘이는 것이 적당하다고 생각하고, 현행보다 문항을 ‘줄여야 한다’고 응답한 경우는 평균 3개를 줄이는 것이 적당한 것으로 응답했다. 또한 늘여야 할 이유로는 ‘듣기 문항과 비례해서(11.8%)’, ‘의사소통 능력을 측정해야 하므로(9.7%)’, 말하기 측정 문항이 적어서(9.7%)’ 순으로 나타났다. 한편 줄여야 할 이유로는 ‘실질적 말하기 시험이 아니어서(21.4%)’, ‘상황 구성이 인위적이므로(14.3%)’가 상대적으로 높게 나타났다.

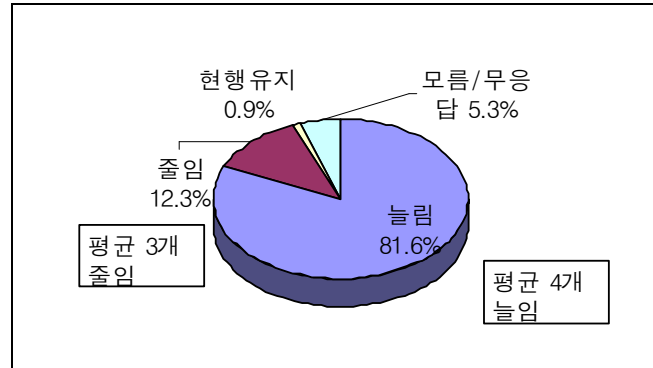


그림 5 간접 말하기 문항을 늘려야 하는 것에 대한 설문 결과

여섯째, 현행 듣기 평가의 녹음 속도에 대해서는, 그림 6에서 보는 바와 같이, 응답자의 57.1%가 느리다고 생각했으며, 또한 그 이유로 ‘문항별 속도 차이가 필요해서(4.6%)’, ‘자연스런 속도 제시 요망(4.6%)’ 등으로 나타났다.

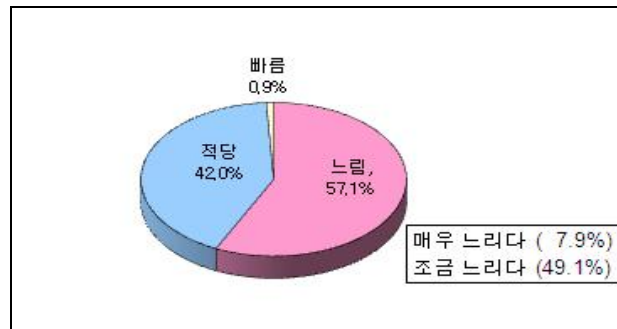


그림 6 듣기 평가의 녹음 속도에 대한 설문 결과

일곱째, 듣기 문항 확대 방안에 따라 새로운 문항 유형을 추가로 도입하는 것에 대해, 그림 7에서 보듯이, 응답자 대부분인 93.0%가 찬성하였으며, 새로운 유형 도입한다면, 적당한 개수는 평균 3개로 조사되었다. 새로운 문항 유형을 추가로 도입하는 것에 찬성하는 이유는 ‘현재 문항이 고정화되어 있어서 (28.3%)’, ‘의사소통 능력을 신장 위한 수업을 확대할 필요(3.8%)’, ‘변별력을 높이기 위해(3.8%)’ 등의 순으로 나타났다.

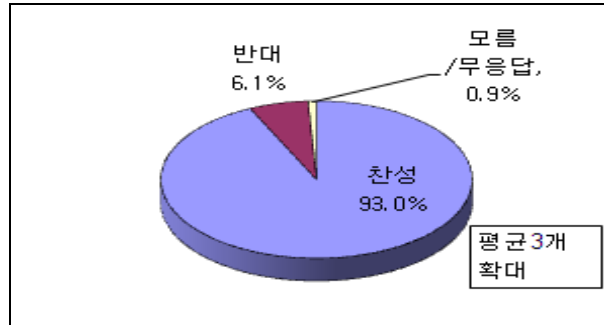


그림 7 신유형 도입에 대한 설문 결과

여덟째, 듣기 영역에서 신유형에는 ‘짧은 대화의 응답’, ‘대화의 내용과 그림의 일치/불일치’, ‘5W1H 세부 사항’, ‘대화/담화에 언급/불언급된 것’, ‘의견 묻기’ 등이 제안되었다. 읽기 영역에서 신유형에는 실용문(안내문, 광고문) 등을 활용한 내용일치/불일치’ 등이 제안되었다.

IV. 수준별 영어 영역의 시험 체제 및 문항 유형 은행의 구성

1. 수준별 영어 영역의 시험 체제의 구성

설문 조사 결과 및 2014학년도 수능 체제 개편 방안(교육과학기술부, 2011)⁸ 토대로 전문가 협의회, 세미나 및 공청회를 거쳐 수준별 영어 영역의 시험 체제는 표 5와 같이 구성되었다.

먼저 총 문항 수는, 2014학년도 수능 체제 개편 방안의 수험생의 시험 부담 완화를 위해, 현행 수능 체제보다 5문항 줄어든 45문항(듣기 22문항, 읽기 23문항)으로 구성하기로 하였다. 시험 시간은 현행 70분 그대로 유지하기로 하되, 듣기 문항 증가에 따라 듣기 시간을 현행보다 10분 늘려 총 30분으로 하였으며, 읽기 영역은 10분 줄어든 총 40분으로 정하였다. 또한 적정 난이도 및 변별력 확보를 위해 5지 선택형은 현행 체제 그대로 유지하기로 했으며, 총 배점도 현행과 같이 100점 만점으로 하기로 했다. 문항의 난이도에 따라 차등 배점도 현행과 같이 유지하지만, 배점 방식을 달리하여 차등 배점 1점은 없애고, 2점과 3점만 두기로 결정하였다⁹. 문두(지시문, 질문 등) 제시 언어는 A형과 B형 모두

⁸ 교육과학기술부에서 발표한 2014학년도 수능 개편 방안의 기초는 ‘과도한 시험 준비 부담이 없는 수능’, ‘학교 수업을 통해 준비할 수 있는 수능’, ‘고교교육 정상화에 기여하는 수능’이었다. 이에 따라(수험생의 시험 부담 완화를 위해) 문항 수를 현행 50개에서 5개 줄였다.

⁹ 현행 차등 배점 방식을 유지할 경우, 1점 배점이 5문항, 2점 배점이 25문항, 3점 배점

수험생의 시험 부담 완화(특히 하위 학습자)를 위해 한국어로 제시하기로 하였다. 듣기 회수는 설문 조사 결과에 따라 1대화·담화문 1문항의 경우, 현행 수능과 같이 1회 들려주지만, 1대화·담화 2문항의 경우, 2회 들려주기로 하였다. 듣기 속도도 설문 조사 결과를 반영하여, A형은 현행 수능과 동일한 속도로 들려주지만, B형은 현행 수능 보다 빠르게 들려주기로 하였다.

또한 설문 조사 결과에 따라 현행 읽기 영역에서와 같이 듣기 영역에서도 1대화·담화문 2문항 유형을 1세트 도입하기로 결정되었다¹⁰. 간접 말하기 문항도 현행 5문항에서 3문항 늘려 총 8문항으로 검사지를 구성하기로 하였다. 새로운 유형으로 짧은 대화의 응답, 그림과 대화의 일치 여부, 의견 묻기, 5W1H 등의 유형을 도입하였다.

표 5
2014 수능체제의 수준별 영어 영역의 시험 체제

시험 체제 구성 요소	A형		B형	
	듣기 영역	읽기 영역	듣기 영역	읽기 영역
총 문항 수	22문항	23문항	22문항	23문항
시험 시간	30분	40분	30분	40분
선택지의 수	5			
총점	100점			
배점(차등배점)	2점(35문항), 3점(10문항)*			
듣기 회수	1회 (1지문1문항), 2회 (1지문2문항)			
듣기 속도	현행 수능과 같은 속도		현행 수능 보다 빠르게	
문두 언어 (질문, 선택지 등)	한국어			

2. 수준별 영어 영역의 문항 유형 은행(Item-types Bank)의 구성

현행 수능 체제에서는 고정형 평가 목표 이원분류표에 따라 매 시험마다 문

이 15문항이 최적 조합으로 나타났다. 이 조합은 수험생에게 상당한 혼란을 줄 우려가 있어 1점 배점을 삭제하고, 대신 2점과 3점 배점을 늘려 각각 35문항, 10문항으로 단순화하였다.

¹⁰ 설문 결과에 의하면, 2세트 도입이 적정한 것으로 나타났지만, 도입 초기이고, 지역(도·농) 및 계층(상·하) 간의 영어 격차를 우려하여 최종 1세트를 도입하기로 하였다.

항 유형과 형식이 고정되어 있는 고정형 검사지를 구성하고 있다. 이와 같이 매 시험마다 동일한 문항 유형이 동일한 형식으로 배열되어 있는 고정형 검사지는 언어 능력의 지배를 받기 보다는 시험 요령의 지배를 받는다는 비판을 받았다(김용명 외 3인, 2010). 이 같은 문제점을 최소화하기 위해, 2014 수능 체제에서는 모듈형 평가 목표 이원분류표로부터 수준별(A·B형) 기능별(듣기·읽기) 시험마다 상보적으로 구성되는 상보형 평가 목표 이원분류표를 구성하고, 이에 따라 매 시험마다 문항 유형과 형식을 융통성 있게 구성할 수 있는 상보형 검사지를 구성하기로 결정하였다. 이와 같이 고정형 검사지 대신 상보형 검사지를 구성함으로써 시험 간 평가 요소 간의 간섭효과를 최소화할 수 있을 뿐만 아니라 시험 요령의 지배도 최소화할 수 있다. 상보형 검사지를 구성하기 위해서는 먼저 문항 유형 은행을 구성해야 한다. 문항 유형 은행은 수준별 영어시험에 포함될 수 있는 가능한 모든 문항을 저장하고 있는 일종의 문항 유형 은행을 말한다.

이 같은 문항 유형 은행을 구성하기 위하여, 3장에서 논의한 문항 유형 외적 타당성(설문조사) 분석에 더하여 국·내외 표준화 시험과 인접 국가인 중국과 일본의 국가 주도 영어 시험(한국의 수능, 중국의 까오카오, 일본의 입시센터)의 문항 유형을 분석하였다¹¹. 현행 문항 유형에 대한 외적 타당성 평가 및 국·내외 표준화 시험과 동양 3국의 국가 주도 영어 시험의 문항 유형 분석을 토대로 수준별 영어 영역의 문항 유형 은행을 다음 표 6, 표 7과 같이 제안하였다.

문항 유형 은행은 문항 유형의 속성에 따라, 표 6과 표 7에서 보는 바와 같이, 몇 개의 모듈로 구성할 수 있다. 듣기/읽기의 이해 방식에 따라 하향 이해 방식이 작동하는 ‘대의 파악 모듈’, 상향 이해 방식이 작동하는 ‘세부 사항 파악 모듈’, 상호 작용 이해 방식이 작동하는 ‘상호작용 모듈’, 이 세 가지 이해 방식이 복합적으로 작동하는 ‘복합 모듈’로 각각 범주화할 수 있다. 또한 재고자 하는 능력에 따라 말하기 능력을 재는 ‘말하기 모듈’, 쓰기 능력을 재는 ‘쓰기 모듈’, 문법·어휘 능력을 재는 ‘문법·어휘 모듈’로 범주화할 수 있다.

듣기 평가의 문항 유형 은행은, 표 7에서 보는 바와 같이, 총 21개의 유형으로 구성된다. 이 중, 순수 듣기 유형은 17개 유형이며, 간접 말하기 유형은 ‘대화/담화에 대한 응답 유형’ 등을 포함하여 4개 유형이다. 듣기 평가에 새로 도입된 신유형은 짧은 대화에 응답, 1대화·담화문 2문항, 5W1H 유형, 대화·담화에서 언급된 것/언급되지 않은 것, 의견 묻기, 그림과 대화의 일치/불일치, 대화·담화의 요지 등 10개 유형이다¹². 또한 현행 유형 중, 문항 유형 타당성 평가에

¹¹ 문항 유형에 대한 외적 타당성 평가가 현행 외국어(영어) 영역에 대한 분석인 반면, 국·내외 표준화 시험(TEPS, TOEFL, TOEIC, Eiken, CET 등)과 동양 3국(한, 중, 일) 국가 주도 영어 시험의 문항 유형 분석은 수준별 영어 영역에 새로이 도입할 신유형에 대한 시사점을 얻고자 수행한 연구였다. 지면 관계상 본 연구에서는 상세한 논의를 할 수 없다(김용명, 권오량, 이완기, 김진석, 고현숙(2011) 참조).

¹² 짧은 대화에 응답은 TEPS의 유형을 변형한 것이며, 1대화·담화문 2문항, 5W1H 유형,

서 상대적으로 낮은 점수를 받았고, 2014 수준별 영어의 출제의 용이성을 높이기 위해, 그림 설명 유형과 그림에 적절한 대화 찾기 유형은 삭제하기로 결정하였다.

모듈에 따라 듣기 영역의 문항 유형 은행을 분류하면 크게 4모듈로 나눌 수 있다. 먼저 하향 이해 방식에 따라 대화문과 담화문의 대의를 파악하는 능력을 측정하는 ‘대의 파악’ 모듈이 있으며, 목적, 주장, 주제, 제목, 요지 추론 유형 등이 이 모듈에 속한다. 상향 이해 방식에 따라 대화문이나 담화문의 구체적 세부 내용을 파악하는 능력을 측정하는 ‘세부 사항 파악’ 모듈이 있으며, 내용 일치·불일치, 언급·불언급, 한·할·부탁한 일 등이 이 모듈에 속한다¹³. 또한 상향 또는 하향 이해 방식을 작동하여 복합 대화문이나 담화문의 대의나 구체적 세부 사항을 파악하는 능력을 측정하는 ‘복합’ 모듈이 있으며, 1대화·담화문 2문항 또는 3문항이 이 모듈에 속한다. 마지막으로 듣기 자료를 통해 이해한 바를 가상의 말하기에 적용할 수 있는 능력을 측정하는 ‘말하기’ 모듈이 있으며, 짧은 대화에 응답, 마지막 말에 대한 응답, 담화에 응답, 그림에 맞는 대화 등이 이 유형에 속한다.

표 6
수준별 영어 시험의 듣기 영역의 문항 유형 은행

내용 영역	모듈	문항 유형	외적 타당성분석결과 (%)*
듣기 유형	대의 파악 (하향이해)	담화·대화 목적	3.5
		담화·대화 주장	신유형
		담화·대화자 의견	신유형
		담화·대화의 주제	0.9
		담화·대화 요지	신유형
		담화·대화 제목	신유형
		화자의 심정	43.9
		대화 장소	32.5
		대화자 관계	18.4
	세부 사항 파악 (상향 이해)	할·한·부탁한·일	3.8
		숫자파악	7.9
		5W1H 세부 사항	신유형
		도표 내용일치	0.9
		그림 대화일치	신유형

대화·담화에서 언급된 것/언급되지 않은 것, 의견 묻기, 그림과 대화의 일치/불일치 등은 NEAT와 연계를 위해 NEAT 유형을 수능에 맞게 변형한 것이다.

¹³ 스키마 이론(schema theory)에 따르면(김용명, 1991; Nuttal, 1996), 읽거나 듣기의 이해 방식에 따라 하향식(top-down), 상향식(bottom-up), 상호작용식(interactive)으로 구분할 수 있으며, 본 연구에서는 이를 토대로 읽기 유형의 모듈을 구성하였다.

		대화·담화 내용 일치	2.6
		대화·담화 언급 일치	신유형
	복합	1대화·담화문 2문항	신유형
		그림에 맞는 대화	신유형
간접 말하기	말 하 기	짧은 대화 응답	신유형
		대화에 응답	4.4
		담화에 응답	4.4

* 현행 듣기 유형에 대해 검사지 구성에서 배제해야 할 유형에 대한 응답률임

읽기 영역의 문항 유형 은행은, 표 7에서 보는 바와 같이, 현행 외국어 영역의 유형, 신유형 등을 포함하여 총 20개의 유형으로 구성된다. 이 중에서 문법·어휘 유형이 3유형, 순수 읽기 유형은 ‘글의 목적’, ‘빈칸 추론’ 등 총 11개 유형이며, 쓰기 유형은 ‘주어진 문장 삽입’, ‘글의 순서’ 등 4유형이며, 복합(세트) 유형은 1지문 2·3문항 등 2유형이다. 신유형에는 안내문, 광고문, 초정장 등과 같은 실용 지문을 활용하여, 내용일치/불일치를 측정하는 유형이 도입되었다.

표 7
수준별 영어 시험의 읽기 영역의 문항 유형 은행

내용 영역	모듈	문항 유형	외적 타당성 분석 결과 (%)	행동 영역				문항 유형 특성 체계		
				사 실 적	적 용 력	추 문 적	종합 적	통합성 단일형	통합성 통합형	주축성 주축형
문법 어휘	문법 어휘	어법(밑줄형)	7.0							
		어법(박스형)	5.3							
		어휘(박스형)	8.8							
		어휘(밑줄형)	1.8							
		지칭추론(지시어)	14.9							
읽기	대의 과약 (하향 이해)	글의 목적	13.2							
		필자의 주장	21.9							
		글의 주제	21.9							
		글의 요지	24.6							
		글의 제목	31.6							
		글의 분위기·심경	45.6							
		상호 작용 (추론)	빈칸추론 (단어·구·절·문장) 연결사 추론	0.0 7.0						
세부	실용자료 내용 일치	신유형								

	사항 (상황 이해)	도표 내용 일치	14.0
		내용일치·불일치	7.9
복합		1지문 2문항	-
		1지문 3문항	-
간접 쓰기	쓰기	주어진 문장 삽입	1.8
		글의 순서 찾기	5.3
		흐름에 무관한 문장	2.6
		문단 요약	0.9

모듈에 따라 읽기 영역의 문항 유형은행을 분류하면 크게 6개의 모듈로 나눌 수 있다. 먼저 주어진 글의 전체적인 의미를 파악하는 능력을 측정하는 ‘대의 파악’ 모듈에는 목적, 주장, 주제, 제목, 요지 추론 등이 있다. ‘세부 사항 파악’ 모듈에는 내용 일치/불일치, 도표 자료의 이해, 실용 자료의 이해 등이 있다. 상향 및 하향 이해 방식을 통합적으로 작동하여, 주어진 지문의 내용을 토대로 명시되지 않는 정보를 논리적으로 추론할 수 있는 능력을 측정하는 ‘상호 작용’ 모듈이 있으며, 빈칸 추론(단어, 구, 절, 문장)이 이 모듈에 속한다. 또한 상향 또는 하향 이해 방식을 작동하여 복합 지문의 대의나 구체적 세부 사항을 파악하는 능력을 측정하는 ‘복합’ 모듈이 있으며, 1지문 2문항 또는 3문항이 이 모듈에 속한다. 또한 읽기 자료를 통해 이해한 바를 가상의 쓰기에 적용할 수 있는 능력을 측정하는 ‘쓰기’ 모듈이 있으며, 흐름과 관계없는 문장 찾기, 주어진 문장 삽입, 글의 순서 찾기, 요약 등이 이 모듈에 속한다. 마지막으로 문법적 능력과 어휘적 능력을 측정하는 ‘문법·어휘’ 모듈이 있다.

이상에서 논의한 바와 같이, 문항 유형 은행의 각 모듈은 동일 능력을 측정하는 유사한 문항 유형을 묶어 놓은 소위 ‘메타 문항’으로 구성되어 있다. 따라서 수준별(A·B형), 기능별(듣기·읽기) 시험의 상보형(개별) 검사지를 구성할 때, A형과 B형의 적정 수준 및 난이도를 조정하기 위해, 어떤 모듈에서 몇 문항 인출할 것인지, 어떤 모듈에서 어느 문항을 어떤 문항으로 교체할 것인지, 또 어떤 문항을 삭제할 것인지에 대한 논리적 답을 제공해 줄 수 있다. 따라서 매 시험마다 문항 유형에 따른 검사지의 구성은 달라질 수 있지만, 동일 모듈 내에서 문항이 교체되므로 시험의 수준 및 난이도는 항상성과 동일성을 유지할 수 있다.

3. 수준별 영어 영역의 상보형(Complementary) 검사지의 구성

이제 이와 같이 구성된 문항 유형 은행으로부터 문항 유형 특성 체계(통합성, 주축성, 위계성)의 각 구성소의 적정 구성 비율에 따라 해당 문항 유형을 문항 유형 은행으로부터 인출하여 수준별(A·B형), 기능별(듣기·읽기) 시험의 상보형

검사지를 구성할 수 있다.

먼저 통합성 원칙에 따라 수준별 영어 시험의 단일형과 통합형 문항의 적정 구성 비율이 결정되겠지만, 2014 수능 체제에서 검사지 구성에 급격한 변화를 주게 되면, 수험생과 사회에 혼란을 초래할 가능성이 있다. 이점을 고려하면, 2014 수능 체제의 수준별 영어시험의 단일형과 통합형 문항의 적정 구성 비율이 '4:1'이 타당할 것이며¹⁴, 이에 해당 하는 비율만큼 문항 유형 은행으로부터 해당 특성을 지닌 유형을 인출하여 수준별 영어 A형과 B형의 검사지를 구성할 수 있다.

또한 주축성 원칙에 따라 주축성 문항과 주변성 문항의 적정 비율이 결정되겠지만, 2014학년도 수능 체제 개편 방안(교육과학기술부, 2011)에 따르면, 수준별 영어시험 A형은 현행 수능 보다 쉽게 출제하고, B형은 현행 수능 수준으로 출제한다고 발표하였다. 이에 따라 A형의 경우, 현행 수능보다 쉽게 출제하기 위해서는 주축성 문항의 비율을 낮추는 반면, B형의 경우, 현행 수준으로 출제하기 위해서는 주축성 문항의 비율을 현행대로 유지하는 것이 바람직할 것이다. 이점을 고려하면, 주축성 문항과 주변성 문항의 적정 비율은, A형의 경우, '3:1', B형의 경우, '4:1'이 타당할 것이다¹⁵. 전자의 경우, 시험에 변화성 및 다양성을 확보하기가 상대적으로 쉬울 것이며, 후자의 경우, 향상성과 안정성을 확보하기가 상대적으로 용이할 것이기 때문이다. 마지막으로 위계성 원칙에 따른 H형, M형, L형의 적정 구성 비율은 A형의 경우, '2:6:2', B형의 경우, '3:4:3'이 타당한 것으로 생각된다. 이에 대한 논리적 근거는 A형은 총체적 난도는 B형에 비해 상대적으로 낮아야 하며, 중하위 학습자(수험생) 간에 변별력이 있어야 한다는 데서 찾을 수 있으며, B형은 총체적 시험의 난도는 A형에 비해 상대적으로 높아야 하며, 상위 학습자(수험생) 간에 변별력이 있어야 한다는 데서 찾을 수 있다.

V. 결론 및 시사점

서론에서 언급한 바와 같이, 현재 한국 영어 평가에는 세 가지 변화, 즉, 듣기 문항 확대 방안, 2014 수능 체제의 수준별 영어 시험, NEAT의 수능 대체 여부가 진행 중에 있다. 2010년대 이 세 가지 주요한 변화의 성패는 어떤 문항 유

¹⁴ 현행 외국어(영어) 영역 문항 분석에 따르면, 독립형과 통합형의 비율이 약 4:1이다. 이에 근거하여 수준별 영어시험의 독립형 문항과 통합형 문항의 구성 비율을 4:1로 제안하였다.

¹⁵ 현행 외국어(영어) 영역 문항 분석에 따르면, 주축성 문항과 주변성 문항의 비율이 약 4:1이다. 이에 근거하여 수준별 영어시험 B형의 주축성 문항과 주변성 문항의 구성 비율을 4:1로 제안하였다.

형을 어떻게 배합하여 수준별 영어 시험 및 NEAT에 타당한 검사지를 구성할 수 있느냐에 달려있다. 이 같은 영어 평가의 변화에 대처하기 위해, 본 연구에서는 현행 외국어(영어) 영역의 시험 체제 및 문항 유형에 대한 외적 타당성을 분석하여, 이를 수준별 영어 시험 및 NEAT의 검사지 구성에 반영하고자 하였다.

이 같은 연구 목적을 실현하기 위해, 본 연구에서는 수능 및 문제 은행 출제 경험을 갖고 있는 교수, 교사 및 연구원 114명을 대상으로 현행 수능 외국어(영어) 영역의 시험 체제 및 문항 유형 타당성에 관한 설문 조사를 수행하였다. 설문 결과를 토대로 현행 외국어 영역의 문항 유형에 대한 외적 타당성 평가를 분석한 결과는 다음과 같았다. 첫째, 듣기 영역 13개 유형 중, 삭제할 필요가 있는 유형에는 ‘화자의 심정 파악(43.9%)’, ‘그림의 상황에 맞는 대화(41.2%)’, ‘대화 장소 추론(32.5%)’, ‘대화로 설명된 대상 찾기(19.3%)’, ‘대화자 관계 추론(18.4%)’ 등의 순으로 나타났다. 한편, 읽기 영역 20개 유형 중, 삭제할 필요가 있는 유형에는 ‘필자의 심정(45.6%)’, ‘글의 제목(31.6%)’, ‘글의 요지(24.6%)’, ‘주제 파악(21.9%)’, ‘지칭 추론(14.9%)’, ‘도표의 세부 내용(14.0%)’, ‘글의 목적(13.2)’ 등의 순으로 나타났다. 둘째, 응답자의 88.6%가 현행 듣기 영역에 1대화문 2문항 또는 1담화문 2문항의 유형을 도입하는 것에 대해 찬성했으며, 적당한 도입 세트는 평균 2.2개로 조사되었다. 셋째, 1대화문 2문항 또는 1담화문 2문항 들려주는 횟수에 대해서는 ‘2번’ 들려주는 것에 찬성하는 의견이 ‘1번’ 들려주는 것에 찬성(43.0%)하는 것보다 9.6% 높은 52.6%였다. 넷째, 현행 듣기 영역에서 (간접)말하기 5개 문항보다 늘리는 것에 찬성한다는 의견은, 81.6%였으며, 평균 4개를 늘리는 것이 적당한 것으로 나타났다. 마지막으로 듣기 영역에서 신유형에는 ‘짧은 대화의 응답’, ‘대화의 내용과 그림의 일치/불일치’, ‘SWIH 세부 사항’, ‘대화/담화에 언급/불언급된 것’, ‘의견 묻기’ 등이 제안되었다. 읽기 영역에서 신유형에는 실용문(안내문, 광고문) 등을 활용한 내용일치/불일치’ 등이 제안되었다.

이와 같은 현행 수능 외국어(영어) 영역의 시험 체제에 관한 설문 조사 결과를 토대로 2014학년도 수준별 영어 영역의 시험 체제를 다음과 같이 구성하였다. 총 문항 수는 현행보다 5문항 줄어든 45문항(듣기 22, 읽기 23)으로 구성하였다. 문항의 난이도에 따라 차등 배점 방식도 현행과 같이 유지하지만, 차등 배점 방식을 달리하여 차등 배점 1점은 없애고, 2점과 3점만 두기로 결정하였다. 듣기 회수는 설문 조사 결과에 따라 1대화·담화문 1문항의 경우, 현행 수능과 같이 1회 들려주지만, 1대화·담화 2문항의 경우, 2회 들려주기로 하였다. 듣기 속도도 설문 조사 결과를 반영하여, A형은 현행 수능과 동일한 속도로 들려주지만, B형은 현행 수능 보다 빠르게 들려주기로 하였다.

또한 현행 외국어(영어) 영역의 문항 유형에 대한 외적 타당성에 대한 분석 결과를 토대로 문항 유형 은행을 구성하였다. 먼저 듣기 평가의 문항 유형 은행은 총 21개의 유형으로 구성되며, 이 중, 순수 듣기 유형은 17개 유형이며,

간접 말하기 유형은 ‘대화/담화에 대한 응답 유형’ 등을 포함하여 4개 유형이다. 듣기 평가에 새로 도입된 신유형은 짧은 대화에 응답, 1대화·담화문 2문항, 5W 1H 유형, 대화·담화에서 언급/불언급, 의견 묻기, 그림과 대화의 일치/불일치, 대화·담화의 요지 등 10개 유형이다. 읽기 영역의 문항 유형 은행은 총 20개의 유형으로 구성된다. 이 중에서 문법·어휘 유형이 3유형, 순수 읽기 유형은 ‘글의 목적’, ‘빈칸 추론’ 등 총 11개 유형이며, 쓰기 유형은 ‘주어진 문장 삽입’, ‘글의 순서’ 등 4유형이며, 복합(세트) 유형은 1지문 2·3문항 등 2유형이다. 신유형에는 안내문, 광고문, 초정장 등과 같은 실용 지문을 활용하여, 내용일치/불일치를 측정하는 유형이 도입되었다.

이와 같이 구성한 문항 유형 은행으로부터 문항 유형 특성 체계(통합성, 주축성, 위계성)의 각 구성소의 적정 구성 비율에 따라 해당 문항 유형을 문항 유형 은행으로부터 인출하여 수준별(A·B형), 기능별(듣기/말하기·읽기/쓰기) 시험에 타당한 상보형 검사지를 구성할 수 있음을 보여주었다. 결론적으로, 문항 유형에 대한 타당성 분석 결과를 토대로 문항 유형 은행을 구성하고, 이 문항 유형 은행으로부터 매 시험마다 문항 유형 특성 체계의 각 구성소의 정해진 적정 비율에 따라 해당 특성을 지닌 문항 유형을 인출하여 상보형 검사지를 구성함으로써 수준별, 기능별 시험 간 평가 요소 간의 간섭 효과를 최소화할 수 있을 뿐만 아니라 시험 요령의 지배도 최소화할 수 있다.

이상에서 논의한 본 연구의 결과는 NEAT, 국가수준학업성취도평가 및 교실 수업에서 수준별 평가에 적용가능성이 있다. 현행 NEAT, 국가수준학업성취도평가 또는 수준별 교실 평가의 기존 문항 유형 및 새로이 도입하고자 하는 유형에 대해 해당 시험 출제 전문가(또는 교사)에게 외적 타당성 평가를 하게 하고, 그 결과를 토대로 문항 유형 은행을 구성할 수 있을 것이다. 이와 같이 구성한 문항 유형 은행으로부터 전문가(또는 교사)들이 설정한 문항 유형 특성 체계의 각 구성소의 적정 구성 비율에 따라 NEAT의 등급별 시험(2·3급), 국가수준 학업성취도평가의 수준별 시험(초·중·고), 수준별 교실 평가의 수준(상·중·하)에 부합하는 상보형 검사지를 구성할 수 있을 것이다.

본 연구에서 수행한 문항 유형에 대한 외적 타당성 분석은 일종의 총체적 척도에 의한 평가이다. 이와 같이 총체적 척도에 따른 문항 유형에 대한 평가는 평가자의 경험과 직관에 의존할 수 밖에 없고, 해당 문항 유형이 어떤 평가 영역에서 강점과 단점을 지니고 있는지 파악할 수 없다는 한계점을 지니고 있다. 이 같은 한계점을 보완하기 위해, 총체적 척도에 의한 평가와 상호 보완적 관계에 있는 분석적 척도에 따라 문항 유형 타당성 평가를 할 필요가 있다. 구체적으로 말하면, 가령, 문항 유형 결정 원리(김용명, 2010a)를 토대로 문항 유형의 질적 타당도를 평가하기 위한 평가 영역(예, 연계성, 실제성, 상호작용성, 환류 효과성)을 설정하고 그에 따른 하위 영역(기준) 및 평가 요소를 설정한다. 이 같은 평가 요소에 따라 문항 평가자에게 문항 유형에 대한 타당성을 Likert n 점 척도로 평가하게 할 수 있을 것이다. 이제 이 같은 외적 타당성 평가의 문

제점을 보완하기 위해 분석적 척도에 따른 문항 유형 내적 타당성 평가를 수행해야 할 때이다.

참고문헌

- 교육인적자원부. (2006). 인터넷 기반 국가영어능력인증시험 시행 계획 발표. *보도자료*(2006.11.3).
- 교육인적자원부. (2007). 국가영어능력평가시험 도입 기본 계획 수립 발표. *보도자료*(2007.07.30).
- 교육과학기술부. (2008a). *외국어과 교육과정(I)*: 교육과학기술부 고시 제 2008-160호 [별책 14]. 서울: 교육과학기술부.
- 교육과학기술부. (2008b). 국가영어능력평가시험 개발 계획 발표. *보도자료* (2008.12.18).
- 교육과학기술부. (2009). 대통령 업무 보고. *보도자료*(2009.12.23).
- 교육과학기술부. (2010). 국가영어능력평가시험 대입 수시에 반영 발표. *보도자료*(2010.1.7).
- 교육과학기술부. (2011). 2009 개정 교육과정에 따른 2014학년도 수능시험 개편 방안. *보도자료*(2011.2.26).
- 김용명. (1991). 상호작용 읽기 모델의 관점에서 *Good/Poor Reader*의 읽기 전략 비교 연구. 미출간 석사학위 논문, 서울대학교, 서울.
- 김용명. (2010a). 국가영어능력평가시험(NEAT)의 문항 유형 개발과 선별 원리 및 검사지 구성 원칙. *영어교육*, 65(4), 369-398.
- 김용명. (2010b). 국가영어능력평가시험(NEAT)의 검사지 구성의 원칙과 절차: 문항 유형 확정 모델. *영어어문교육*, 16(4), 153-184.
- 김용명, 이완기, 김진석, 고현숙. (2010). 수능 *외국어(영어)* 영역 개선 연구. 한국교육과정평가원, 연구보고 CAT 2010-11.
- 김용명, 권오량, 이완기, 김진석, 고현숙. (2011). 수능 *외국어(영어)* 영역 평가 방법 개선 후속 연구. 한국교육과정평가원, 연구보고 CAT 2011-2.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy*. New York: Longman.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Kim, Yong-Myeong. (2006). A common metric scale (CMS) on the parallel developmental sequence model. *English Teaching*, 61(4), 77-107.
- Kim, Yong-Myeong. (2007a). Diagnosis and remedy system (DRS) for teaching English

on the Common Metric Scale (CMS) model. *English Teaching*, 62(2), 47-77.

Kim, Yong-Myeong. (2007b). Validation of the common metric scale (CMS). *English Language & Literature Teaching*, 14(1), 21-44.

Nuttal, C. (1996). *Teaching reading skills in foreign language* (2nd. Ed.) Oxford, UK: Heinemann.

Pienemann, M. (1985). Learnability and syllabus construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp.23-112). Clevedon, Avon: Multilingual Matters.

Pienemann, M. (1998). *Language processing and second language development*. Amsterdam: John Benjamins.

부록
설문지

동계법 13조(비밀의 보호)에 의거 본 조사에서 개인의 비밀에 속하는 사항은 엄격히 보호됩니다. (전문가용)

수능 외국어(영어) 영역 듣기 및 읽기 영역 문항 유형 선정에 관한 의견 조사

안녕하십니까? 저희 한국교육과정평가원에서는 교육과학기술부의 위탁을 받아 실용영어교육 강화 방안의 일환으로 대학수학능력시험 외국어(영어) 영역 평가 방법 개선 후속 연구들 수행해 오고 있습니다. 위 연구의 핵심 내용은 수능 시험의 듣기 문항수들 17개→25개로 확대하는 방안과, 읽기 문항수들 33개→25개로 축소하는 방안에 관한 것입니다. 이에 대해 선생님의 의견을 청하고자 아래 설문을 마련하였습니다. 응답해 주시는 설문지 결과는 대학수학능력시험 외국어(영어) 영역 듣기 및 읽기 문항 유형 확정에 소중한 자료로 활용될 것입니다. 선생님의 응답 내용은 동계법 13조에 의해 원저하게 비밀이 보장되며 오직 통계분석 용도로만 사용됨을 약속드리며, 설문지 응해 주시는 성의와 노고에 감사드립니다.

2011년 1월
 ◎ 한국교육과정평가원 수능 외국어(영어) 개선 방안 연구팀
 김용명 (02-3704-3533)/고현숙 (02-3704-3922)

- 아래 ()에 √로 표시하거나 기타에 쓰시오.**
- ☐ 성 별: () 1. 남자 () 2. 여자
 - ☐ 나 이: () 1. 20대 () 2. 30대 () 3. 40대 () 4. 50대 () 5. 60대
 - ☐ 학 령: () 1. 대학교 졸업 () 2. 석사 () 3. 박사
 - ☐ 교육 경 령: () 5년 미만 () 5년 이상 10년 미만 () 10년 이상 15년 미만 () 15년 이상
 - ☐ 직 종: () 1. 교수 () 2. 교사 () 3. 장학사 () 4. 교육 행정 5. 기타
-
- ☐ 소속 시 도: () 1. 서울 () 2. 부산 () 3. 대구 () 4. 인천 () 5. 광주 () 6. 대전 () 7. 울산 () 8. 경기 () 9. 강원 () 10. 충북 () 11. 충남 () 12. 전북 () 13. 전남 () 14. 경북 () 15. 경남 () 16. 제주
 - ☐ 근무 지역 단위: () 1. 특별시 및 광역시 () 2. 도시(중소)지역 () 3. 읍면지역 () 4. 도서벽지

☐ 자녀가 있으십니까? (하나만 √ 표시함) () ① 네, 있습니다. ↓	() ② 아니오, 없습니다. ↓
☐ 자녀(들)은 어느 학교급에 속합니까? (자녀들의 해당하는 학교급에 모두 √ 표시함) () 초등학교 () 중학교 () 고등학교 () 대학생 () 기타 () ↓	↓

■ [듣기 문항 확대에 대하여]

※ 수능 외국어(영어) 듣기 문항 확대 방안에 따라 듣기 문항은 현행 17개에서 25개로 확대됩니다. 현행 수능시험의 듣기 문항 17개를 문항 유형별로 분석해 보면 총 13개의 유형으로 구분할 수 있습니다. 먼저 다음과 같은 현행 13개의 유형의 적절성 및 효과성에 대해 선생님의 의견을 듣고자 합니다.

문1. 다음의 13개 유형 중, 삭제할 필요가 있다면 어느 유형을 수능 영어에서 삭제해야 한다고 생각하십니까? (가장 우선적인 2개 항목에 √ 표시하시고, 그 이유를 간단히 적어주십시오.)

- ☞ 문항 유형의 적절성을 평가하는 기준으로는 다음과 같은 내용도 고려하실 수 있습니다.
 - 타당성: 문항 유형이 측정해야 할 듣기 또는 (간접) 말하기 능력을 측정하고 있는지?
 - 실제성: 문항 유형이 실제 언어 사용 상황에서 활용 가능성이 높은지?
 - 상호작용 : 문항 유형이 학습의 내재된 언어능력과 관계되는 정도가 높은지?
 - 환류효과성: 문항 유형이 교실 수업과 평가에 긍정적 영향을 주는지?
 - 출제의 용이성: 문항 유형을 제작하기가 용이한지?

☞ 아래 표의 문항 번호는 2011학년도 문항 번호를 기준으로 한 것입니다.

☞ √ 표시를 복사해서 삭제할 유형에 표시해주시면 됩니다.

문항 번호	내용 영역	유형(13개)	수능 영어에서 삭제해야 할 유형	
			삭제할 유형에 √ 로 표시	이유
1	듣기	대화로 설명된 그림 찾기		
2		화자의 심정 파악		
3		주제 파악		
4		화자의 할일		
5		숫자 관련 정보		
6		글의 목적 추론		
7		화자가 부탁한 일		
8		대화 장소 추론		
9		대화자 관계 추론		
10		화자의 할일		
11		도표 자료 활용		
12		담화의 내용일치 불일치		
13	간접 말하기	그림의 상황에 맞는 대화		
14-16		대화에 대한 적절한 응답		
17		담화에 대한 적절한 응답		

※ 듣기 문항을 현행 17개→25개로 확대하면, 현행 수능보다 8문항을 추가하는 것입니다. 이에 대한 선생님의 의견을 듣고자 합니다.

문2. 현행 듣기 영역에서는 (간접) 말하기 문항이 총 5문항입니다. 선생님께서는 간접 말하기 문항을 현행보다 더 늘려야 한다고 생각하십니까 아니면 줄여야 한다고 생각하십니까?

아래 () 안에 ✓ 로 표시하시고, 그 이유를 간단히 적어 주십시오.

늘리는데 찬성 (), (몇 문항?) _____ 더 늘려야 한다.

줄이는데 찬성 (), (몇 문항?) _____ 더 줄여야 한다.

이유: _____

문3. 현행 읽기 영역에서 출제되는 1지문 2문항 또는 1지문 3문항을 듣기 영역에 도입하여, 1대화 2문항 또는 1담화 2문항을 출제하는 안에 대해 어떻게 생각하십니까? ()안에 ✓ 로 표시해 주십시오.

▪ 도입에 찬성 (), 도입에 반대 ()

▪ 찬성 또는 반대 이유: _____

▪ 도입한다면 1대화 2문항 또는 1담화 2문항이 몇 세트가 적당하다고 생각하십니까?
_____ 세트

▪ 현행 듣기 영역에서는 대화나 담화를 한 번 들려주고 있습니다. 1대화 2문항 또는 1담화 2문항의 경우에 한 번 들려주는 것과 TEPS/TOEIC처럼 두 번 들려주는 것에 대해 어떻게 생각하십니까? ()안에 ✓ 로 표시해 주십시오.

· 한 번 들려주는 것에 찬성 (), 두 번 들려주는 것에 찬성 ()

· 이유: _____

문4. 현행 듣기 영역의 녹음 속도에 관해 어떻게 생각하십니까? ()안에 ✓ 로 표시해 주십시오.

()	()	()	()	()
매우 느리다	조금 느리다	적당하다	조금 빠르다	매우 빠르다

기타 의견: _____

문5. 듣기 문항 확대 방안에서는 새로운 문항 유형을 추가로 도입하고자 합니다. 이에 대해 선생님께서는 어떻게 생각하십니까? ()안에 ✓ 로 표시해 주십시오.

▪ 도입에 찬성 (), 도입에 반대 ()

▪ 찬성 또는 반대 이유: _____

▪ 도입 한다면 몇 가지의 문항 유형을 추가로 도입하는 것이 적당하다고 생각하십니까? _____ 가지

문6. 선생님께서 수능 영어 듣기 문항에 도입되었으면 하는 신유형의 문항이 있으시면 그 개요를 아래에 적어 주시면 연구에 적극 참고 하겠습니다(필요하시다면 PDF 또는 다른 파일 첨부 가능합니다).

■ [읽기 문항 축소에 대하여]

※ 수능 외국어(영어) 듣기 문항 확대 방안에 따라 읽기 문항은 현행 33개→25개로 축소됩니다. 이에 아래 제시된 현행 20개의 문항 유형에 대해 선생님의 의견을 듣고자 합니다.

문7. 아래 20개 문항 유형 중, 어느 유형을 수능 영어에서 삭제해야 한다고 생각하십니까? (가장 우선적인 2개 항목에 √ 로 표시하시고 그 이유들 간단히 적어주십시오. 단, 동일 유형의 2개 문항 중, 1개를 삭제하는 선택도 가능합니다.)

☞ 문항 유형을 평가하는 기준으로는 다음과 같은 내용도 고려하실 수 있습니다.

- 타당성: 문항 유형이 측정해야 할 읽기 또는 (간접) 쓰기 능력을 측정하고 있는지?
- 실제성: 문항 유형이 실제 언어 사용 상황에서 활용 가능성이 높은지?
- 상호작용: 문항 유형이 학습의 내재된 언어 능력과 관계되는 정도가 높은지?
- 환류효과성: 문항 유형이 교실 수업과 평가에 긍정적 영향을 주는지?
- 출제의 용이성: 문항 유형을 제작하기가 용이한지?

☞ 아래 표의 문항 번호는 2011학년도 문항 번호를 기준으로 한 것입니다.

☞ √ 표시를 복사해서 삭제할 유형에 표시해주시면 됩니다.
(단, 동일 유형의 2개 문항 중, 1개를 삭제하는 선택도 가능합니다.)

문항 번호	유형(20개)	수능 영어에서 삭제해야 할 유형	
		삭제할 유형에 √로 표시	이유
18	글의 목적		
19	지침 내용 주문		
20	문맥상 판단(민중형)		
21	문맥상 판단(박스형)		
22	글의 전체 흐름		
28	필자의 주장		
80	필자의 성격		
81	문맥에 적절한 어휘(박스형)		
82	문맥에 적절한 어휘(민중형)		
88	영문사 주문		
84	주제 파악		
85	주제 파악		
88	내용 일치		
87	내용 불일치		
88	도표의 세부 내용		
88	글의 묘지		
40	글의 묘지		
41	글의 제목		
42	글의 제목		
48	순서 배열하기		
44	문장 삽입		
45	문단 요약		

문8. 선생님께서 수능 영어 읽기 문항에 도입되었으면 하는 신유형의 문항이 있으시면 그 개요를 아래에 적어 주시면 연구에 적극 참고 하겠습니다.
(필요하시다면 PDF 또는 다른 파일 첨부 가능 합니다.)

<< 설문에 끝까지 답변해 주셔서 감사합니다. >>

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

김용명
한국교육과정평가원
서울특별시 중구 정동길 21-15 한국교육과정평가원 321호
Tel: (02) 3704-3533
E-mail: mencius@kice,re.kr

강문구
공주대학교 사범대학 영어교육과
충남 공주시 신관동 182 공주대학교 사범대학 영어교육과
Tel: (041) 850-8190
E-mail: kangmunkoo@hanmail.net

Received 25 June 2012
Revised 14 August 2012
Accepted 20 August 2012