외국어능력 검정시험 개발 모델

최 인 철

경쟁이 치열해지고 있는 현대의 국제화 사회에서는 외국어능력의 평가에 있어서 집단 내에서의 개개인 능력의 상대적 차이뿐만 아니라 개인 능력의 절대적인 수준의 평가가 절실히 요구되고 있다. 현재 시행되고 있는 외국어 능력 검정시험들은 대부분 시험개발의 이론적 근거가 희박하여 엄격한 타당도 검증과정을 거치지 않고 활용되고 있으므로 시험의 對사회적인 책임을 다하지 못하고 있는 실정이다. 이런 점에서 본 논문은 외국어 테스팅의 최근 이론을 바탕으로 서울대학교 어학연구소에서 실시할 바람직한 외국어 능력 검정시험 개발을 위한 모델을 제시하고자 한다. 시험개발모델에 의거하여 개발되고 시행된 시험의 결과를 토대로 타당도 분석 및 검증에 관한 별도의 후속 연구가 진행될 것이다.

본 논문에서는, 우선 언어능력 요소와 평가의 이론적 개념들을 간략히 살펴 본 후, 본론적으로 외국어능력의 평가방법에 대한 고려사항 등을 언어영역별 내용, 구성 및 시험방식양상 등의 측면에서 체계적으로 고찰해 보고자 한다. 마지막으로 외국어능력 평가의 대표적인 유형들을 내용영역별로 예제와 함께 제시한다. 시험결과 분석의 후속연구에서 제시된 유형의 타당성이 체계적으로 검증될 것이다.

Ⅰ. 언어능력 요소, 시험내용 및 속성

1. 언어능력 요소

의사소통(communicative) 언어능력의 구성요소에 관한 이론들은 우수한 시험문제의 개발과 타당성 검증의 근본적인 이론적 바탕을 제공한다는 점에서 외국어시험 분야에서는 매우 중요한 역할을 한다. 여러 응용언어학자(Bachman and Palmer, 1981; Canale, 1983; Cummins, 1983)들의 의견을 종합해 보면 의사소통 언어능력은 (1) 언어적 능력, (2) 상황 판단적 능력, (3) 심리-생리적(psycho-physiological) 능력으로 대분되고 아래와 같이 세분된다(Bachman, 1990).

언어적 능력은 크게, (1) 구성(organizational) 능력(grammatical and textual)과 (2) 화용(pragmatic) 능력(illocutionary and sociolinguistic)으로 구분된다. 구성능력은 순수한 언어학적인 능력인 데 반해, 화용능력은 언어와 비언어의 상호작용의 원활한 통제능력을 말한다.

문장 구조와 의미의 규칙에 관한 문법적 능력은 다시 (1) 어휘, (2) 형태소, (3) 구문, (4) 음운으로 구성되며, 일관성 있는 언어구사에 관한 구문적 능력은 (1) 응집(cohesion), (2) 수사법적 구성(rhetorical organization/coherence)으로 구성된다.

언어기능에 관한 illocutionary competence는 (1) 정보-의사교환(ideational), (2) 상황조작-통제(manipulative), (3) 교육-인지(heuristic), (4) 창작-예술(imaginative) 등의 기능을 수행하는 언어능력으로 구성된다.

언어의 의미와 형태의 적합성(appropriateness)을 규정하는 사회문화적 규칙에 관한 사회 언어학적 능력은 (1) 방언(dialect and variety), (2) 전문용어(register), (3) 문화적 지식 및 수사법(cultural references and figures of speech), (4) 자연성(naturalness)으로 구성된다고 볼 수 있다.

위의 언어적 능력의 구성요소 모델에 근거하여 시험의 각 내용영역별로 측정할 대상의 세부언어능력 요소를 선정하여 시험 목적에 따라 각각의 구성 비중을 정한다. 과거 시험에서는 언어적 능력 중에서 특히 문법 및 구문을 대상으로 하는 구성능력에 치중했으나, 진정한 의사소통능력을 측정하려면 구성능력뿐만 아니라, 언어기능 및 사회언어학적 능력을 포괄한 화용적 능력까지도 평가영역에 포함시켜야 할 것이다. 의사소통능력의 평가를 위한 타당한 시험의 내용은 의사소통능력 요소 구성비중의 체계적인 고려를 통하여 의사소통 상황을 잘 반영하여 진정성(authenticity)을 극대화시키는 것이 가장 중요하다.

2. 시험내용 영역 및 구성

수험자들의 필요분석 결과와 아울러 시험시행상 실용성의 제약점 등을 고려해 볼 때 본 논문에서 다룰 시험 내용영역은 구두표현능력 및 작문능력을 제외한 청해, 독해, 어휘 및 문법 등 네 영역으로 한다.

2. 1. 문법

문법적인 능력은 국어 및 문어의 의사소통을 위해 실제 언어사용능력인 묵시적 문법능력(tacit grammatical competence)과 의사소통능력과는 직결되어 있지 않은 언어학적 지식인 명시적 문법지식(explicit grammatical knowledge)으로 대별된다. 이는 Krashen(1985)에 의하면 의식적 학습(learning)과

잠재적 습득(acquisition)과도 맥을 같이 하는 개념이다. 구어체 및 문어체의 문맥에 따라 적용되는 문법사항이 다소 차이가 있으므로 이를 분류하여 평가하는 것이 바람직하다.

의사소통능력 외국어시험에서는 구어 및 문어의 효율적인 의사소통을 위해 필수적인 내재화된 (잠재의식적) 문법능력이 평가의 대상이 되어야 한다.

2.2. 어휘

어휘능력은 파생어 (접사) 활용능력, 숙어력 등을 포함한 능력이다. 의미영역(semantic network) 및 사용 빈도수에 따라 난이도에 상당한 차이가 있다.

구어체 및 문어체의 문맥에 따라 어휘의 사용이 상당한 차이가 있으므로 이를 분류하여 평가하는 것이 중요하다. 문어체 어휘뿐만 아니라 구어체 의사소통 상황에서 활용빈도가 높은 어휘 및 二語동사(phrasal verbs) 등의 숙어까지 시험내용에 포함되는 것이 바람직하다.

2.3. 읽기

독해의 기능별로 보아 읽기 영역은 (1) 읽기 전 단계(saccade movement 전단계)부터 (2) 정독, (3) skimming(대의 파악), (4) scanning(원하는 정보 찾기) 능력 등을 일컫는다. 그리고 독해를 위한 여러 요소별로 보아 (1) alphabet 식별력, (2) 어휘(숙어)력, (3) 문법실력(구문력), (4) 세부내용 파악 능력, (5) 중심사상 인지 및 결론 유도, (6) 전후 내용의 논리적 유추, (7) 작가의 태도와 어조의 이해 및 추론, (8) 속독 능력, (9) 세상 지식 등이 있다.

외국어시험에서 어디까지 독해능력으로 평가해야 하는가 하는 점에서는 논란이 있을 수 있다. 그러나, 주로 외국어학습 초보자들이 활용하는 bottom-up 독서방법보다는 bottom-up 방법과 top-down 방법의 상호적(interactive) 독서과정을 활용할 때 유리하도록 사실적 세부내용 파악뿐만 아니라 전체적인 내용, 즉, 문맥 및 대의 파악과 논리적 추론을 중요시하는 독해능력 평가를 지향해야 한다(Goodman, 1967; Smith, 1982; Eskey & Grabe, 1986).

언어사용 상황에 직접적으로 관련되는 실생활에 사용되는(genuine) 지문내용을 선택하며 수험자 집단의 편파적인 이익 및 불이익을 초래할 수 있는 지엽적이며 전문적인 학술내용은 배제함으로써 내용 타당도를 제고한다.

2.4. 듣기

청해능력에는 기능별로 크게 micro-listening과 macro-listening으로 대별된

다(Rivers and Temperley, 1978). Micro-listening은 발음, 어휘, 구문의 식별 등 언어 요소의 분석을 통한 기초적 청해능력(perception)을 말하고, macro-listening은 언어 요소의 분석적 청해 과정을 넘어서, 연음되는 음성언어를 의사단락으로 크게 듣는 과정(chunking)을 통해, 사실적 내용파악뿐만 아니라 대의파악, 논리적 추론, 어조의 유추 등을 통한 광의의 의사소통을 위한 청해능력(comprehension)을 말한다. micro-listening은 bottom-up 독해과정에, macro-listening은 bottom-up과 top-down의 혼합적 독해과정에 비유된다.

청해능력을 위한 요소로서는 (1) 발음식별력(자모음, sandhi), (2) 억양, 휴지 및 강세 이해, (3) 구어의 특징(fillers, false starts, 주저함, 반복 및 redundancy 등) 이해, (4) 어휘/숙어력, (5) 문법실력(구문력), (6) 세부내용 파악능력, (7) 중심사상 파악능력, (8) 화자의 태도(얼굴 표정 및 몸 동작) 및 어조의 이해 및 추론, (9) 세상지식 등이 있다(Ur, 1984; Rost, 1990). 일반적인 의사소통 상황에서는 청해의 궁극적인 목적은 macro-listening의 성취에 있으므로, 청해능력의 평가에서도 청해의 기본이 되는 micro-listening 뿐만 아니라 전체적인 내용파악을 위한 macro-listening 능력에 초점을 두어야 한다.

실제 의사소통에서 나타나는 자연스러운 발음현상 및 속도를 청해시험에 반영하는 것은 중요하지만, C-R 평가속성상 수험자 집단의 청해력을 고려하여 문항별 난이도 수준에 따라서 발음속도를 통제해야 한다. 특히 초보 언어능력의 수험자 집단을 대상으로 하는 저곤란도의 문항 개발에 있어선 발음현상 및 속도를 통제하는 것이 난이도 및 변별력, 신뢰도, 내용타당도면에서 중요하다. 실제 의사소통 상황에서도 외국어 초보자에게는 모국어 사용자가 foreigner talk을 사용하는 사실을 볼 때도 본 시험의 수험자 집단의 다양한 언어능력을 신뢰성과 타당성 있게 평가하기 위해서는 발음속도의 통제는 중요한 고려사항이다.

3. 평가이론적 고려사항

3. 1. 시험종류

3. 1. 1. 이론적 근거

구어 및 문어의 종합적 의사소통능력을 측정하기 위해서는 언어분석 이론뿐만 아니라 의사소통을 위한 언어사용 이론에 바탕을 둔 전반적인 외국어능력의 시험(proficiency test)이다.

3. 1. 2. 시간제약

의사표현을 할 때 의식적인 지식체계가 감시장치(monitor: Krashen, 1985)로서 역할을 하려면 충분한 시간적 조건이 충족되어야 한다는 점을 고려하여, 시험시간을 적절히 배분하여 감시장치가 의식적으로 작동할 수 없도록 시간적 여유를 주지 않으면, (文語나 口語의) 의사소통능력의 바탕이 되는 잠재의식적인 지식체계를 좀 더 정확히 측정할 수 있다. 따라서 시험시간 제한 측면에서는, 의사소통능력 시험은 반드시 역량시험(power test)보다 속도시험(speed test)의 성격을 띠어야 한다.

시간 제약 시행에 대한 명확한 이론적 규정은 없으므로 수험자 집단의 실력수준을 고려하여 보통 시험시행의 경험상으로(rule of thumb) 결정한다. TOEFL같은 대부분의 대규모 시험에서는 시간 내에 총 수험자 집단의 80% 정도가 시험문제를 다 풀어볼 수 있도록 하며, 나머지 20% 정도의 수험자가 푼 문항의 수가 총 시험문제의 70% 정도에 달하도록 시험시간을 정한다. 일반적으로 표준화된 시험상황(standardized testing)에서는 본시험 실시 이전에 모형시험 실시를 통하여 얻어진 결과를 바탕으로 시험대상 수험자 집단능력에 알맞게 시험시간을 조정한다.

3. 1. 3. 참조근거

급수검정시험의 성격을 띨 본시험은 개인의 능력차이를 극대화하는 목적의 규준지향평가(N-R: norm-referenced)의 성격과 아울러 시험성적이 수험집단의 성적과 관계없이 수험자 개개인의 언어능력을 절대적으로 나타내주는 준거지향평가(C-R: criterion-referenced testing)의 속성을 갖추는 것이 수험결과의 타당하며 의미있는 해석을 위하여 중요하다(Cziko, 1983). 언어테스팅분야에서 많은 관심에도 불구하고 아직까지는 연구가 많지 못하기 때문에 C-R시험 개발을 위해선 구체적 모델이 없는 것이 현실이다(Brown, 1993). 따라서 C-R시험의 중요한 요소인 내용영역별 세분화 및 비중을 결정하기 위하여 Bachman의 언어능력 구성요소 모델을 이론적 근거로 한다.

앞으로, 모형시험의 시행과 타당도 검증과정을 통하여 수험자의 전반적 의사소통능력과 시험성적과의 상관관계를 분석하여, 타당한 능력별 급수체계를 지닌 N-R 및 C-R의 시험을 개발한다. 실력을 어느 정도 객관화시킬 수 있는 수험자집단을 대상으로 시험을 치러서 얻어진 결과에 대한 분석과 타당도 검증을 근거로 하여 C-R의 속성인 성적의 절대적인 기준을 마련한다.

등급 기준으로는, 文語 및 口語시험 각각, 측정대상의 언어를 모국어로 사

용하는 교육을 받은 집단만이 만점을 다 맞을 수 있도록 난이도를 조정하고, (예컨대, 100점 만점을 기준으로 할 때) (1) near-native 실력을 가진 최상위 집단(high-advanced)이 91~100점, (2) 일반적인 상위 집단(advanced)이 81~90점, (3) 중상위 집단(high-intermediate)이 71~80점, (4) 중위 집단(intermediate)이 51~70점, (5) 중하위 집단(low-intermediate)이 41~50점, (6) 하위 집단(elementary)이 31~40점, (7) 최하위 집단(lowest)이 30점 이하를 받도록 하여 실력별로 전체를 7등급 정도로 차등화하여 C-R의 시험의 속성을 최대한 반영한다.

3.2. 문항 지수

3.2.1. 난이도

지나치게 어렵거나 쉬운 시험은 실력의 우열을 가리기 어려우므로(floor or ceiling effect), 경쟁시험의 중요한 고려사항인 변별도면에서 상당히 문제가 있어 궁극적으로 시험의 타당도에 부정적인 영향을 미친다. 종합적인 의사소통능력시험(proficiency test)인 속도시험과 N-R 및 C-R의 시험속성을 균형있게 지니기 위해서는 정규분포를 따르도록 전체적 평균이 .5 정도를 유지하면서, 어려운 문항과 쉬운 문항, 그리고 중간 정도의 난이도를 가진 문항을 다양하고 균형있게 출제하여 변별력을 높이는 것이 바람직하다 (Gronlund, 1985).

問頭(stem) 및 독해/청해지문의 내용의 난이도(readability)를 좀 더 객관적으로 분석하기 위해서는 어휘빈도 사전(frequency dictionary)이나 구문 분석 프로그램을 활용하는 것이 효과적이다.

평가이론상 가장 타당한 시험은 수험자들의 실력수준에 맞는 난이도의 문항들로 구성된 시험이다. 결국 실력이 상이한 수험자마다 각기 내용이 상이한 시험을 치러야 하는 상황에서는, 紙筆式고사가 불가능하므로 새로운 평가이론인 문항반응이론(Item Response Theory: IRT)에 근거한 Computer Adaptive Test(CAT)를 활용해야 한다. CAT의 실용화가 아직 어려운 현실정에서는 고전적 평가이론(Classical Testing Theory)을 바탕으로 난이도는 쉬운 문항과 어려운 문항이 적절한 비율로 혼합되는 것이 바람직하다. 그러나, IRT를 활용하면 상이한 난이도의 문항에 대해 달리 나타나는 수험자의 문항반응에 따라 수험자의 실력을 보다 신뢰성 높게 추정할 수 있다.

3.2.2. 변별도

변별도는 문항점수와 총점과의 상관관계 계수인 양분상관계수(biserial)로

써 표시된다. EPT는 수험자 개인과 사회전반에 중요한 역할을 하는 경쟁 및 능력급수 측정시험이므로 실력의 우열 차이를 극대화함으로써 변별도를 가능한 한 높이는 것이 중요하여 .25 이상이 되는 것이 바람직하다.

문항개발의 개선책으로 상위 및 하위능력 집단별로 각 문항의 오답지의 선택률을 검토하는 오답지분석(Distractor Analysis)을 통하여 (multiple-choice 시험의 신뢰도 및 타당도에 심대한 영향을 미치는) 각 문항의 선택지 내용을 체계적으로 검토하여 시험내용을 개선한다.

변별도지수가 수험자들의 무분별한 추측(wild guess)의 영향을 받으므로 신뢰도에 많은 영향을 미치므로, IRT이론을 활용하여 추측 정도의 문제점을 체계적으로 분석함으로써 오답의 감점제도 실시 여부에 관하여 합리적인 결정을 할 수 있다.

3.3. 시험지수

3.3.1. 신뢰도

신뢰도는 타당도의 선행조건이므로, 대사회적인 책임이 큰 대규모의 시험 상황에서 .9 이상이 되는 것이 바람직하다. 현실적인 제약에서 신뢰도 및 타당도의 제고를 위해 전체 시험(test battery)의 문항수를 200~250문항 정도로 개발한다. 속도시험의 성격을 극대화하면 많은 문항수에도 불구하고 소요 수험시간은 2시간 내외가 걸리게 되므로 수험자들에게 심리적 부담을 크게 주지 않는다. 수험결과의 타당한 해석을 위해서 표준측정오차(SEM: Standard Error of Measurement)를 제시한다.

3.3.2. 타당도

3.3.2.1. 내용타당도 (Content Validity)

의사소통능력 요소의 모델(CLA & TMF; Bachman, 1990)과 사회언어학적 언어사용이론, 수험자의 필요 언어능력에 대한 설문조사를 바탕으로 시험의 내용을 균형있게 표본추출하여 내용타당도를 제고한다. 언어교육 전문가의 판단을 근거로 한 내용타당도의 평가는 다소 주관적이다. 따라서, 시험 내용의 이해난이도(易讀性, readability)를 분석하기 위하여 전산언어학(Computational Linguistics) 분야의 자연언어처리(Natural Langauge Processing) 구문분석(parsing) 프로그램을 활용하는 것도 내용타당도의 객관성을 보완하는 효과적인 방법이다.

시험의 내용만을 검증의 대상으로 하기 때문에, 수험자의 능력과 시험내용간에 상호작용은 고려하지 못하므로, 수험자의 실력을 사전에 파악하지

못했을 경우에는 좋은 내용타당도의 시험이라도 변별력이 낮을 수 있음을 유의해야 한다.

내용타당도는 시험이 측정하고자 하는 내용영역의 표본추출(sampling)의 정도를 의미하므로, (현실적인 제약하에서) 다양한 지문과 아울러 다양한 유형의 시험문항수가 많을수록 그만큼 내용타당도는 높다.

3.3.2.2. 기준참조(관련) 타당도 (Criterion-related Validity)

기준시험으로서의 근거가 주관적인 타 지필고사 수험결과와 상관관계를 분석하는 것은 별 의미가 없을 수 있다. 따라서, 구두면접이나 TSE (Test of Spoken English) 등을 실시하여 얻어진 의사소통능력의 평가결과와 현행 시험내용 영역별 및 시험방식별 수험결과의 관계를 비교하는 공인(共因, concurrent) 타당도와 현재 수험결과와 앞으로의 의사소통능력과의 관계를 비교하는 예언 유용도(predictive utility; Bachman, 1990)를 분석하는 상관관계 연구를 통해 가장 타당한 시험 내용영역 및 방식을 선택함으로써 전반적으로 시험을 개선한다.

3.3.2.3. 구인타당도 (Construct Validity)

구인타당도의 검증은 한 시험이 측정하고자 하는 바를 얼마나 타당하게 측정하는가를 조사하는 타당도 검증의 궁극적인 목표를 달성하고자 하는 것이기 때문에, 타당도의 개념상 가장 핵심적인 검증과정이다. 내용타당도와 기준참조타당도 검증은 모두 구인타당도의 검증을 위한 보조검증 과정이라 볼 수 있으므로, 위의 타당도 검증과 아울러 기술통계와 아울러 추론통계, 즉, ANOVA (Analysis of Variance), 상관관계 및 회귀분석 (Regression Analysis), MTMM (Multi-Trait Multi-Method) 기법, 요인분석 (Explanatory & Confirmatory Factor Analysis) 등의 제 통계기법을 활용하여 시험의 전반적인 타당도를 검증한다.

구인타당도의 검증을 위한 고려사항은 Bachman(1990)의 시험방식양상 모델이론에 근거를 둔다. 즉, (1) 시험 환경(testing environment), (2) 시험 구성 체제요소(test rubric), (3) 시험문제 내용 속성(nature of the input the test taker receives), (4) 요구되는 수험자의 응답 속성(nature of the expected responses to that input), (5) 시험문제 내용과 수험자 응답과의 관계(relationship between input and response)이다. 궁극적으로 좋은 시험의 개발은 위의 변수를 고려하여 타당성 검증과정을 거칠 때 가능하게 된다.

II. 외국어능력 평가 방법

1. 언어테스팅의 시험방식 양상

언어테스팅이 타 학문 분야의 능력 평가와 상이한 점은 평가 수단과 평가 대상 공히 언어라는 점이다. 언어테스팅의 내용과 형식의 이론적 바탕을 제공하는 시험방식양상(Test Method Facets: Bachman, 1990)이 얼마나 실제 의사소통상황의 언어기능과 내용에 근접하며 수험자들의 실력수준에 부합하는가에 따라서 시험의 진정성과 타당도가 좌우된다. 이런 점에서 외국어의 시험 출제와 타당성 검증과정에서는 시험방식양상의 세부적인 요소 하나하나를 심층적으로 고려하고 점검해야만 좋은 시험문제가 개발될 수 있고 체계적인 타당성 검증이 가능하다. 본 시험의 개발 및 타당성 검증의 제 과정은 위의 모델을 근거로 한다.

2. 영역별 시험 구성, 비중 및 방식

속도시험의 취지를 최대한도로 반영하면 수험 소요시간이 짧게 걸리므로 그만큼 문항수를 증가시킬 수 있다. 문항수를 극대화시키므로 신뢰도 및 타당도를 더욱 제고시키며, 수험시간을 단축하여 수험자의 심리적 부담 및 시험비용을 극소화시키도록 한다.

구성내용은 음성언어로 제시되는 (1) 청해시험과 문자언어로 제시되는 (2) 언어능력요소시험(문법, 어휘)과 (3) 독해시험으로 구성된다. 문항의 배열은 쉬운 문항부터 어려운 문항순으로, 짧은 지문부터 긴 지문순으로 배열한다.

청해시험은 100문항/50분 내외, 독해시험은 50문항/50분 내외, 문법은 50문항/10분 내외, 어휘는 50문항/10분 내외로 하여, 총 문항수는 250으로 하며, 총 소요시간은 2시간 내외로 한다. 각 내용영역별로 타당성이 입증된 유형을 다양하게 도입하여 시험방식에 따라서 수험자에게 주어지는 편파적인 이익/불이익을 최소화한다.

각 내용영역별로 시험방식 양상에 대한 유의점을 다음과 같이 알아본다.

2. 1. 문법

2. 1. 1. 시간적 제약

Krashen의 주장대로 의사표현을 할 때 의식적인 지식체계(conscious learning)가 감시장치(Monitor)로서 역할을 하려면 충분한 시간이 있어야 한다는 점을 감안할 때, 감시장치가 작동할 수 없도록 시간적 여유를 주지 않으

면, 의사소통능력의 바탕이 된다는 잠재의식적인 지식체계(subconscious acquisition)를 좀 더 정확히 측정할 수 있다.

기존의 문법시험에서는 측정하는 문법지식이 무엇인지를 파악한 다음 암기한 문법지식을 한참 동안 생각해서 풀도록 하여 문법지식을 공식처럼 의미없이 암기하도록 조장함으로써 의사소통능력 배양의 교육현장에 매우 부정적인 파급효과(washback effect)를 끼치는 것이 사실이다. 실제 언어사용의 특성상 시간적 제약이 필연적이므로 역량(power) 시험보다는 시간에 제한을 두는 속도(speed) 시험의 양상을 취하는 것이 진정한 의사소통능력의 외국어시험을 위해서는 매우 중요하다.

2.1.2. 시험방식

시험방식으로는, 어구 빈칸채우기의 유형과 오류찾아내기 유형으로 구성된다. 어구 빈칸채우기의 유형은 주로 쉬운 문제로 간주되며, 작문의 기본적인 능력인 문장구성능력을 평가한다. TOEFL같은 기존 시험에서 많이 활용하는 오류찾아내기 유형에서는 부분적인 어구만을 제한적으로 밑줄로 그어서 틀린 부분을 찾게 함으로써 미시적인 문법능력을 평가하게 된다. 이런 문제점을 보완하기 위해서는 문어체 담화 및 구어체 대화 문장을 구성하는 의미단락을 사선(/ /)으로 분리하여, 문맥과 함께 사용되는 거시적인 문법능력인 진정한 구문판단능력을 측정한다.

2.1.3. 언어학적 문법지식 측정 지양

의사소통능력과는 관계가 먼 문법적인 용어사용으로 시험의 타당성에 문제가 있는 그런 방식의 출제는 배제되어야 한다. 언어학을 전공하지 않는 수험자들에게 불필요한 문법용어의 지식이 요구되거나 지엽적이며 단편적인 문법지식이 측정되는 것은 지양되어야 한다. TOEFL Section 2(Structure/Written Expression)나 UCLES (University of Cambridge Local Examination Syndicate) FCE/CPE Paper 3 (Usage) 영어구문이나 문법시험에서도 문법용어 지식이 요구되지 않는다는 사실을 유념할 필요가 있다.

2.2. 어휘

2.2.1. 시간적 제약

우리나라 학생들은 대개가 의사소통을 위한 목적이 아니라 주로 시험만을 위하여 영어공부를 하여서 문법뿐만 아니라 어휘력도 상황과 유리된 상태에서 기계적인 학습에 익숙해 있다. 따라서 사고 및 개념의 기본요소인

어휘를 통하여 인지과정(heuristic function)을 순간적으로 활성화시키지 못하는 어려움을 많이 겪는다. 즉, 필수적인 독해능력인 속독능력 없이 문법구조를 따지면서 단어의 뜻을 하나하나 풀이하면서 문자 그대로 「암호해독」(deciphering)을 하고 있기 때문이다. 어휘를 통한 순간적으로 느끼는 개념파악이 시간제약이 많은 실제(real-time) 의사소통에서는 매우 중요하므로, 어휘력시험도 잠재적인 어휘능력을 측정하기 위해 진정한 속도시험으로 시행한다.

2.2.2. 시험방식

일반적으로 빈칸완성형과 동의어선택형의 두 가지 유형이 가장 많이 사용된다. TOEFL의 전형적인 유형인 동의어선택형은 전후 문맥에 의존하지 않아도 밑줄 친 어구의 동의어를 고르게 함으로써 문맥과 유리된 어휘력을 측정하는 분리식 시험(Discrete-Point)의 양상을 띠는 문제점이 있다. 반면에 빈칸완성형은 정답을 고르기 위해선 문맥에 의존할 수밖에 없는 통합적(Integrative) 시험의 양상을 띠게 되므로 신뢰성 및 타당성이 제고된다(Oller, 1979).

2.2.3. 문두의 문맥

문두의 문맥은 자연스러운 의사소통상황을 반영하는 내용이어야 한다. 특히, 선다형 시험에서는 문두 내용의 진정성이 더욱 중요하다. TOEFL에서는 매우 전문적인 어휘나 잘 쓰이지 않는 단어들이 문맥에서 사용되는 경우가 적지 않은데 이는 타당성에 중요한 진정성에 위배되는 것이다.

TOEFL 및 TOEIC에서는 어휘력 시험의 문두는 한 문장으로 구성되어 있는데, 이는 구어 및 문어체 두 면을 고려하지 못한 데 기인한다. 진정한 의사소통능력의 측정을 하려면, 문어체 및 구어체의 두 가지 다른 차원의 어휘력을 측정해야 한다. 문어체 어휘력은 한 문장을 문두로 제시해도 되지만, 구어체 어휘력의 측정은 대화체인 두 문장으로 구성된 문두를 제시하는 것이 타당하다.

2.2.4. 시험내용 출처

의미영역 및 빈도수에 따라 난이도의 상당한 차이가 있으므로 어휘 빈도수에 관한 자료(frequency dictionary 등)를 활용하여 객관적으로 난이도의 수준을 통제해야 한다. 구어체 어휘력 평가를 위한 문항개발을 위하여는 실제 방송의 드라마나 단막희극(sitcom) 등의 대본을 concordance 프로그램 등으로 분석하여 빈도수에 근거하여 어휘를 선택하는 것도 효과적인 출제

방법이다.

2.3. 청해

2.3.1. 문항수 및 수험시간

청해시험은 (문항당 소요시간은 문항 내용에 따라서 다를 수 있겠으나) 자연스러운 언어사용 속도로 녹음하며 응답시간을 가능한 한 제한하여 속도시험의 속성을 최대한 살린다. 문법이나 어휘시험과 마찬가지로, 청해력 평가의 속도시험은 잠재적 종합적 언어능력을 측정하는 데 매우 효과적인 방식이다.

청해에서는 가장 중요한 변수가 시험내용을 읽는 발음속도이다. Rivers (1980)에 의하면, 보통 발음속도가 160~190 WPM(Words Per Minute)라고 하는데, 사실 여러 변수(상황, 話者, 격식 등)에 따라 속도는 얼마든지 달라진다(Choi, 1992). 청해속도면에서는 자연언어의 발음속도를 최대한도 반영하되 변별력과 신뢰도를 높이기 위해 우리나라 수험자들의 수준에 맞춰 난이도를 조정한다.

만일 TOEFL에서처럼 선택지를 문제지에 제시해야 할 경우에는 선택지를 읽는 시간도 매우 중요한 변수이다. 문제가 들려진 후 선택지를 읽고 답을 표시하는 데 걸리는 시간이 12초 정도, 즉 각 선택지당 약 2~3초 정도로 제한하여 속도시험을 개발하는 것이 바람직하다(참조: TOEFL: 약 15초; TOEIC: 약 10초).

2.3.2. 음성언어 對 문자언어

청해시험은 음성언어의 이해력을 측정하는 시험임에도 불구하고, 청해시험에서는 청해지문이나 질문을 듣고 글로 쓰여진 선택지를 보고 그 중에서 답을 고르는 형식이 가장 전형적인 시험방식인 것처럼 간주되고 있다. 그러나 괜찮아 보이는 이런 유형은 상당히 심각한 문제점들을 내포하고 있다.

우선, 청해력과 독해력이 혼합되어 측정하고자 하는 언어능력이 애매해지는 내용타당도 및 구인타당도의 문제점이 대두된다. 보통의 수험자들은 주어진 짧은 시간에 선택지를 읽고 이해하며 따라가기가 어렵게 느낀다. 이는 독해력의 변수가 수험과정에 많이 개입한다는 증거이며, 독해력 변수가 청해력 평가에 작용하여 내용타당도에 문제가 됨을 여실히 보여주는 예이다.

또한, 이런 시험유형에서는 수험자가 어떤 수험전략을 채택하는가에 따라 수험결과가 많이 차이가 나게 된다. 즉 주어진 선택지를 보고 난 후에 질문을 듣고 답을 하는 경우와, 선택지를 보면서 질문을 듣고 답을 하는 경우,

또한 질문을 듣고 난 후 선택지를 보고 답을 하는 경우의 세 가지 전략에 따라 수험 결과가 많은 영향을 받게 된다. 이런 시험 유형에서 통제할 수 없는 수험전략의 변수가 타당한 수험결과를 얻는 데 많은 장애요인이 된다.

나아가서, TOEFL같은 시험에서는 시험유형을 잘 아는 수험자들은 지시문이 들려지는 동안에 지시문을 듣지 않고 선택지의 내용을 한 번 죽 훑어봄으로써 (지시문대로 시험을 치를 때보다) 성적이 많이 향상이 된다. 이런 부정행위(cheating) 아닌 부정행위가 행해질 수 있는 시험방식은 필히 지양해야 한다.

이런 문제점 등을 해결하기 위해선 처음부터 끝까지 모든 청해지문, 질문과 선택지를 모두 음성만으로 제시하는 것이 타당하다. 이렇게 함으로써, 독해능력의 혼합을 배제하여 타당도를 제고하고, 선택지까지 들려주기 때문에 더 많은 시험내용(input)을 들려주게 되므로 시험의 가장 큰 제약점인 표본추출을 극대화하여 신뢰도를 제고한다. 다만 시·공의 제약이 많은 음성언어로 제시되기 때문에 수험자의 기억부담을 부당하게 요구하는 청해시험이 되지 않게 하기 위해 선택지는 짧게 제시하는 것이 바람직하다. 또한 독해시험이나 문법시험이 되지 않게 하기 위해서 선택지를 문법적으로나 의미적으로 틀리지 않게 제시한다.

2.3.3. 청취 횟수

정상속도로 녹음된 청해지문을 2번 들려준 후 문제를 풀도록 하는 방식은 몇 가지 중요한 목적을 지니고 있다. 이는 자연스러운 속도에 익숙하지 않은 수험자를 위한 인위적인 난이도 조정을 통하여 변별력을 향상시켜 신뢰도를 제고하기 위한 것이 일차적 목적이다. 더욱 중요한 이차적인 목적은 인위적인 시험유형을 자연스러운 의사소통방식에 근접시키기 위한 것이다. 즉 어떤 청해지문을 무턱대고 듣고 난 후 질문에 대한 대답을 찾아야 하는 수험자는 청해지문의 모든 내용을 기억하지 않는 한 매우 당황하기 마련이다.

더욱이, 청해시험의 전형적인 유형인 「Where is the conversation most probably taking place?」 등의 문제는 의사소통의 진정성에 매우 심각한 문제점을 내포하고 있다. 일상 의사소통상황에서 자신이 처해 있는 상황을 모르는 경우가 얼마나 될까? 눈이 가려진 상태에서 납치된 후 자신이 처해 있는 상황을 유추해야 되는 경우를 빼고는 그런 상황은 가정하기 매우 어렵다. 시험상황에서 청해지문을 두 번 들려줌으로써 수험자는 처음 들을 때 대략적인 의미파악(macro-listening)에 신경을 쓰고 질문을 듣고 나서 문제에서 요구하는 세부적인 정보를 찾기 위한 청해(micro-listening)를 함으로써 인위적인 점을 보완하는 청해시험 방식이다.

또한, 이런 방식은 음성의 시·공적인 제약을 보완해 주기도 한다. 청취라는 실제상황의 제약때문에 같은 수험장에서 옆의 수험자가 내는 소음(예: 재채기, 실수 등)때문에 시험내용의 청취가 순간 불가능할 경우가 얼마든지 있을 수 있으므로, 이에 대한 현실적으로 가장 타당한 대책이다. 사실, 실제 언어사용상황에서도 상대방의 말을 못 들었을 때(예: Pardon me? Sorry? 등의 표현으로) 상대방의 말을 다시 요청해서 들을 수 있음을 고려해 볼 때 이런 시험방식을 언어사용상황에 어긋나는 인위적인 방식이라고 볼 수는 없다. TOEFL보다 더 오랜 역사와 타당도를 지닌 UCLES FCE(First Certificate in English)나 CPE (Certificate of Proficiency in English) 시험에서도 이런 방식을 채택하고 있음을 참고할 필요가 있다(Bachman et al, in press).

2.3.4. 一지문 一문항 원칙

기존 시험에서 고난이도의 문제유형 중 흔히 볼 수 있는 것이 긴 지문을 주고 그에 따른 질문을 4~5개 정도 제시하는 유형이다. 독해시험에서도 많이 활용되는 이런 유형은 국부종속(local dependence)이라는 심각한 문제점을 지니고 있다. 국부종속이란 개념은 한 문항을 맞거나 틀리는 경우에 다음 문항을 맞거나 틀릴 확률에 영향을 미치는 정도를 의미한다. 즉 앞 문항을 맞으면 뒤의 문항도 쉽게 맞추거나, 앞 문항을 틀리면 뒤의 문항도 틀리기 쉬운 시험은 국부종속성이 높은 시험으로 바람직하지 않다. 기존의 한 지문당 여러 문제가 연결되어 있는 유형은 국부적으로 종속성이 높아서 신뢰성에 문제가 많다. 따라서 시험의 신뢰도 제고를 위해서는 한 지문당 한 문항으로 국한함으로써 국부독립성(Local Independence)의 원칙을 지켜야 한다.

더욱이, TOEFL의 청해시험 Part C에서는 청해지문이 너무 길어서 심지어 영어교육을 전공한 영어 모국어 사용자 교수들도 내용을 다 이해하여 문제를 풀기가 어렵다 (Choi, 1993). 즉 청해력시험이라기 보다는 기억력의 시험이 될 소지가 많다. 교육받은 모국인에게도 어려운 긴 내용의 시험이라면 당연히 외국인 수험자에게는 난이도 및 변별력에서 문제점이 많은 타당하지 못한 시험유형이다.

2.3.5. 청해지문 내용

난이도에 따라 차이는 있으나 시험내용에서는 자연스러운 속도의 언어에서 나타나는 구어의 특징을 최대한 반영한다. TOEFL의 청해시험 Part C에서처럼, 짧은 강의, 방송, 대화 등의 구두 지문을 '낭독'식으로 제시하는 것은 지양해야 한다. 자연적인 중복성이 부족하고, 오류 시작(false starts)이나, 머뭇거림 등의 자연스러운 구어의 특징이 전혀 없어서 매우 인위적인 시험

내용이 된다. 이러한 비자연적인 언어에만 익숙하다 보면, 실제상황에서 거의 항상 나타나는 구어의 특징들을 대처할 전략적 능력이 부족하게 되어 성공적인 듣기능력을 갖추기 어렵다. 예컨대, '…that (관계대명사) uh…'에서 말끄는 표현(filler) 'uh'를 관사 'a'로 착각하게 되는 오류 등이 계속되면 심각한 청해의 문제가 발생하게 된다. UCLES CPE/FCE에서처럼 자연적인 주저함, 말 바꾸기, 내용의 중복(redundancy) 등 자연언어의 특징을 최대한으로 반영하는 것이 내용타당도의 중요한 요소이다.

2.3.6. 청해지문 길이

EPT 청해시험에서 출제되는 간단한 문장단위의 paraphrasing 문제유형은 실제 언어사용상황을 반영하는 데 다소 미흡하나 하위 실력집단에 대한 변별력은 있다. 중상위 실력집단을 위해서는 실제 회화상황을 잘 반영하는 문맥을 제공하는 의미있는 단위(chunk)로써 청해지문의 길이를 조정하며 세부적인 청해(micro-listening) 능력을 측정하는 시험유형과 아울러, 거시적인 청해(macro-listening) 능력을 측정하는 진정한 종합적 언어능력인 청해능력을 측정한다. 변별도 및 신뢰도를 극대화하는 담화나 대화의 길이의 적절성에 대해서는 좀 더 많은 타당도 검증연구가 필요하다.

2.3.7. 다양한 시험방식

시험방식에 맞춘 수험자들의 수험전략에 따라 부당한 불/이익을 극소화하기 위해 다양한 유형의 활용이 바람직하다. 특히, 실생활에서 시청각 기관을 통하여 의사소통을 하는 점을 고려하여 청각뿐만 아니라 시각적 입력을 활용하는 시험유형도 포함한다(Ⅲ. 시험유형 참조). 길/방향 안내의 내용의 시험문제에선 선택지를 지도 등의 시각적 입력으로 제시한다. 물론 수험자들의 schemata 때문에 야기될 수 있는 그림의 해석상 문제를 극소화하기 위해 세심한 주의가 요구된다. 더 나아가서, 시험내용을 음성으로 제시할 뿐 아니라 효과음 또는 시각적 입력(그림 및 사진) 등도 제공하여 실제 의사소통상황에 좀 더 접근하여 상황을 풍부하게 제공하는 청해시험 방식도 고려해 볼 만하다.

2.4. 독해

2.4.1. 문항수 및 수험시간

현대사회에서 정보의 홍수에 능동적으로 대처하기 위한 외국어능력의 중

요한 요소인 속독능력을 측정한다는 차원에서도 속도시험 양상을 중요시해야 한다. 속도시험의 성격을 극대화할 수 있도록 수험 소요시간 및 문항수를 정하고, 난이도에 따라 문항당 어휘수를 정한다. 독해속도의 기준은 영·미 모국인들을 대상으로 한 연구와 우리나라 학습자들의 수준을 함께 고려하여 설정하는 것이 변별도 및 신뢰도 제고에 중요하다. 종합적 언어능력평가에선 통상 1분에 한 문항을 풀도록 하는 것이 바람직하다.

2.4.2. 독해지문 내용 및 一지문 一문항 원칙

청해에서와 마찬가지로, 국부독립성원칙(IRT의 필수 전제)이 신뢰도 높은 시험의 중요한 전제조건이다. 요인분석기법(Factor Analysis)에 의한 연구(Choi, 1991)에 의하면, 국부종속성이 높은 TOEFL의 독해시험은 특히 주제(topic) 요인에 따른 편파성이 심각한 문제를 야기한다. 즉 전문적인 특정 자연과학의 지문에 관련된 시험문제에서는 그 특정 분야의 배경지식이 부족한 수험자들이 한 독해지문에 관련된 몇 개의 문항을 연속적으로 거의 다 틀리게 되는 경우가 많이 발생한다. 반대로 인문사회과학 분야에서도 마찬가지 현상이 나타난다. 더욱이 한 지문에서 여러 문제를 출제하다 보니 앞 문항과 내용이 다소 중복되는 문제나 상식적인 내용의 문제 등 '문제를 위한 문제'를 출제하는 단점이 있다.

배경지식 없이 언어능력이 존재할 수 없다고는 하지만, 언어능력시험에서는 배경지식이란 변수가 수험결과에 미치는 영향이 가급적 최소화되어야만 타당도가 제고될 수 있다. 이런 점에서 시험내용의 주제선정은 매우 중요하다. 시험의 목적상 시험내용에는 다소 차이가 있겠으나, 전반적인 의사소통능력시험에서는 非전문적인 학문적인 글 외에도 실생활에 필연적으로 중요한 부분을 차지하는 독서물(신문/잡지 기사, 공익/상업 광고, 소설 등)을 포함한 지문 출처의 다양화를 꾀하여 진정한 의사소통능력의 배양을 유도하는 것이 바람직하다.

따라서, 이런 문제점을 해결하기 위해서는 한 독해지문에 한 문항을 제시하고, 지나치게 전문적인 내용의 지문은 지양해야 한다.

2.4.3. 시험유형별 문항제시

일반적으로 수험자들은 주로 문제를 먼저 보고 나서 시험에서 요구하는 것이 무엇인지를 기억하여 뇌의 기억용량을 불필요하게 낭비하며 독해지문을 읽고 나서(개인에 따라서는 다시 문제의 내용을 재확인하고) 문제를 푸는데, 이는 일반적인 독해과정과는 거리가 먼 인위적으로 강요된 수험상 인지과정이다. 그보다는 독해 목적이 우선 정해지고 나서 독해과정에 들어가

는 것이 일반적인 의사소통상황에 더 부합하므로, 문제의 세부내용을 하나하나 기억하는 부담을 수험자에게 주지 않는 시험방식이 바람직하다. 따라서 동일한 시험유형별로 문항을 연속해서 제시하는 방식은 수험자들이 연속되는 동일한 유형의 문제들을 풀 때 문제에 신경쓰지 않고 독해에만 충실할 수 있으므로 진정한 독해과정에 근사한 타당한 방식이다.

2.4.4. 시험방식

시험방식에 맞춘 수험자들의 수험전략에 따라 부당한 불/이익을 극소화하기 위해 타당성이 인정된 문제방식 중에서 다양한 유형을 채택하는 것이 바람직하다. 일반적인 시험유형으로, 선다형 독해지문, 핵심어구 빈칸채우기, 문장순서 재배열, 적절한 문장 삽입, 불필요 문장 삭제 등이 사용되고 있으나(Ⅲ. 시험 유형 참조), cloze 방식도 종합적 언어능력 평가라는 목적을 달성하는 통합적 시험방식이므로 고려해 볼 만한 가치가 충분히 있다. 채점의 실용성에 관한 문제를 해결하기 위해 다소 변형된 cloze 시험유형을 고려해 볼 수도 있다.

2.4.5. 핵심어구 빈칸채우기 유형

독해지문을 근거로 한 질문에 대한 선다형 선택지 유형의 독해시험 타당도에 가장 많은 영향을 미치는 요소 중 하나는 선택지 내용의 질적 문제이다. 출제자는 우선 정답을 작성하고 그에 따른 오답지를 작성하는데, 보는 관점에 따라서 정답의 명백성이 문제시될 때가 적지 않다. 이는 정답이 원문에 주어진 것이 아니라, 출제자의 주관적 관점에 따라서 인위적으로 작성되는 데서 기인하는 선다형의 내재적인 문제점이다.

이런 문제점을 보완하는 좋은 유형으로는 독해지문의 내용상 중요한 부분을 빈칸으로 제시하여 수험자가 원문에 있는 내용을 정답으로 고르도록 하는 빈칸채우기 유형이다. 잘 쓰여진 원문을 선택하여 내용파악을 잘해서 빈칸이 될 내용을 잘 선정하여 출제하면 정답이 이미 원문에 제시되어 있는 상태이므로 정답의 명백성에 대한 시비의 소지가 줄어든다. 대의파악, 추론 등의 종합적인 독해능력을 매우 효과적으로 측정할 수 있고 출제하기도 용이하므로 타당도 높은 시험유형이다.

2.5. 발음평가의 지필검사 방식

某 실용영어 검정시험에서 아직도 사용되는 지필고사식 발음평가 방식은 타당도면에서 문제가 있다. 음성언어의 표현력을 필기시험으로 평가한다는

데는 이론적으로나 실제적으로 많은 제약과 무리가 따른다. 실제 발음능력과는 관계없이 눈으로만 하는 암기를 통해서도 정답을 맞출 수 있기 때문에 정확한 발음의 교육 및 평가와는 유리된 타당하지 못한 유형이다. 현실적으로 학습현장에서 배우는 (틀린) 발음은 올바른 발음과 상당한 차이가 있을 수 있기 때문이다(예: '90년도 2학기 TV고교가정학습 영어특강 (90년 8월 16일 방송); answer의 w 발음, cost의 /ow/ 발음, develop의 3음절 강세, probably의 2음절 강세, recognize의 2음절 강세 등등). 청해시험에서 살아있는 발음으로 소리에 대한 인식력을 측정하기 때문에 지필시험 방식으로 평가하는 발음시험 문제는 그 존재 의미가 없다.

3. 선택지 개발시 유의사항

선다형 시험유형의 타당성을 좌우하는 가장 큰 변수는 선택지, 즉 오답지(distractors)와 정답(key)의 내용타당도이다. 수험자 입장에서 볼 때 내용의 좋고 나쁨이 시험문항의 난이도, 변별도, 신뢰도 및 타당도 등 모든 변수에 심각한 영향을 끼친다. 선다형시험의 선택지 내용이 잘 출제만 된다면 선다형의 객관식 시험도 얼마든지 통합식 시험의 바람직한 양상을 띠게 되어 높은 신뢰도 및 타당도를 얻을 수 있다. 선택지를 개발하는 데 필요한 기본적인 고려사항은 다음과 같다.

3. 1. 세상지식 對 언어능력

언어의 구성요소 중 세상지식이 중요한 부분을 차지하곤 있지만, 문제풀이 과정에서 수험자들의 세상지식 자체가 직접적인 영향을 미치지 않도록 선택지를 개발하여야 한다. 따라서 주어진 지문의 이해를 위한 충분한 언어능력 없이 상식만으로는 풀 수 없는 시험문제를 개발하거나 지문에 의거하지 않고 상식만으로 풀면 정답을 구할 수 없는 문제를 출제한다.

3. 2. 오답지 내용 및 정답 배열

정답 및 오답지 모두 어법상 틀린 곳이 없도록 해야 하며, 오답지의 역할을 잘 해내기 위해선 내용면에서도 지문의 내용과 상식적으로 볼 때 정답같이 느껴지는 매력도가 있어야 한다.

정답의 길이는 너무 길거나 짧지 않게 함으로써 수험자의 수험재능(test-wiseness)의 영향을 최소화해야 한다. 더욱이 오답지의 내용이 정답을 고르는 데 실마리가 되지 않도록 각별히 신경써야 한다. 정확한 정답을 만드는 데만 신경쓰다 보면, 정답의 어휘가 타 오답선택지보다 어려워지는 경향이

있으므로 이점 유의해야 한다.

또한, 정답을 임의 순서로 배열함으로써 수험자들의 무분별한 추측(wild guessing)을 막아야 한다.

3.3. 독해/청해 지문의 난이도와의 관계

시험의 난이도는 시험목적에 따라 얼마든지 달라질 수 있으나, 각 문항의 난이도 조정을 위해서는 선택지와 지문내용의 두 가지 변수를 동시에 고려해야 한다. 선택지 내용의 난이도는 독해/청해 지문 내용의 난이도와 보완관계가 있다. 즉 난이도의 균형을 맞추기 위해선 독해/청해 지문이 어려우면 선택지를 쉽게, 지문이 쉬우면 선택지를 어렵게 만든다.

3.4. 정답의 명확성

오답지의 정답 가능성을 수험자 입장에서 가능한 모든 경우를 검토해 보아야 한다. 일차적으로 선택지를 만든 후 높은 수준의 교육을 받은 native speaker에게 실제 시험을 쳐보게 하여 정답의 시비를 완전히 배제함으로써 내용타당도를 완벽하게 검증하는 검토과정이 필수적이다.

각 문항의 난이도, 변별력과 각 선택지의 선택률 정도를 조사하는 문항분석과정을 통해 시험출제시에 발견하지 못했던 문제점들을 도출하는 과정이 절대적으로 중요하다.

3.5. 지시문 및 선택지의 언어

지시문 및 선택지를 평가대상언어와 모국어 중 어느 언어를 선택하여 제시할 것인가 아니면 혼용하여 제시할 것인가 하는 것도 중요한 시험방식 요소이다. 종합적인 언어사용능력시험의 성격에 비추어 볼 때, 측정대상언어로 지시문을 제시하는 것이 특히 액면타당도(face validity)가 있어 보인다. 그러나, 실제로 언어능력이 미흡한 수험자를 고려해 볼 때는 시험방식에 대한 지시문의 이해가 부족하여 자신의 언어능력을 십분 발휘하지 못하는 일은 구인타당도면에서 문제가 있으므로, 음성(청해) 및 문자언어시험(어휘, 문법, 독해) 공히 측정대상언어와 함께 모국어도 함께 제시하는 것이 바람직하다. TOEFL이나 Cambridge proficiency test batteries 등의 국제적인 시험인 경우에는 상이한 모국어 배경의 수험자를 대상으로 하므로 어쩔 수 없이 영어로만 지시문을 제시하고 있다.

선택지에 관해서는, 독해시험인 경우에는 독해지문의 양과 함께 선택지의 양도 수험자의 독해처리량에 많은 영향을 주므로, 이를 잘 고려하여 선택지

의 양을 결정하여 적절한 속독능력이 요구되도록 한다. 다만 독해지문 본문 내용의 이해력을 평가하는 데 시험의 주 목적이 있으므로, 선택지의 내용이 너무 어려워 본문내용 파악능력의 정확한 평가를 왜곡시키지 말아야 한다. 따라서 문제의 난이도 수준에 지대한 영향을 미치는 선택지는 문항의 난이도에 걸맞는 수준의 내용으로 측정대상언어로 제시하는 것이 바람직하다.

3.6. 선택지 수 및 오답 채점

IRT를 근거로 수험결과의 오답분석을 통하여 무분별한 추측이 심각한 문제로 분석된다면, GRE(Graduate Record Exam)에서처럼 오답에 대한 감점 제도를 도입하여 선다형 문제유형과 속도시험의 속성상 단점을 보완하도록 한다.

선택지의 수에 대해선, 답지의 수를 넷에서 다섯으로 조정하는 것도 추측에 의한 정답률을 낮춤으로써 부분별한 추측을 다소 방지하는 방법이 될 수 있다. 그러나 실제로 다섯의 선택지를 개발하려면 오답지를 위한 오답지를 만들게 되는 문제점이 있다. 또한 청해력 시험에서는 구두 선택지를 들려주는 상황에서는 수험자들의 기억부담을 고려하여 현행대로 사지선다 유형이 바람직하다.

4. 출제 4차원 요소

문제시 고려해야 할 가장 중요한 4가지 요소를 각 영역별로 보면 다음와 같다. 청해시험은 (1) 세부 능력 요소, (2) 시험 유형, (3) 내용(의사소통상황, 기능 및 의미), (4) 난이도로 구성되며, 어휘시험은 (1) 문맥(context), (2) 시험 유형, (3) 격식(formality), (4) 난이도(frequency, semantic network)이며, 문법시험은 (1) 문맥, (2) 시험 유형, (3) 격식, (4) (문법사항의 내재적) 난이도이며, 독해시험은 (1) 세부 능력 요소, (2) 시험 유형, (3) 내용(주제), (4) 난이도(readability) 등으로 구성된다.

실제 외국어능력시험을 출제할 때에는 위의 각 영역별로 4가지의 차원을 동시에 고려할 때만이 타당한 시험을 출제할 수 있는 필수조건을 갖추게 된다.

5. IRT를 활용한 능력급수 결정

문항별 난이도에 상관없이 총 정답 개수만을 수험자의 실력으로 간주하는 고전적 평가이론과는 달리, 확률적 모델인 IRT이론에서는 문항별 난이도(b), 변별력(a), 추측도(c)에 따라 달라지는 수험자집단의 반응양상(item response pattern)을 근거로 각 수험자의 실력지수(θ)를 계산함으로써, 수험

자의 실력을 보다 정확히 평가할 수 있다(Choi, 1993). 다차원적인 언어테스팅에도 model-data fit의 큰 제약없이 IRT가 적용될 수 있으므로, 수험자능력의 급수를 추정하는 데 가장 효과적인 평가기법이다(Choi, 1991).

6. 종합능력 평가 및 진단 평가

신속한 수험결과를 시험내용 영역에 따라 어휘, 구문(문법), 독해, 청해능력 등의 세부언어능력을 체계적으로 통보한다. 따라서 수험자 자신의 언어능력을 종합적인 평가와 아울러 분석적으로 평가함으로써 수험자 개개인의 강점과 약점을 파악할 수 있다.

좀 더 많은 연구를 토대로 하여 독해 및 청해능력의 고차적인 인지능력의 세부적인 내용영역별로도 분석이 가능할 것이다. 유의할 점은 이런 체계적인 분석은 Lado(1965)의 기능별요소 격자표(Skill-by-Component Matrix)를 근거로 한 단편적인 평가방식인 분리식시험(Discrete-Point)과는 그 성격이 다르므로 혼동이 없어야 한다.

Ⅲ. 외국어능력 시험 유형

위에서 언급한 언어능력의 구성적 기본 요소들은 의사소통을 위한 언어기능에 복합적으로 얽혀 있다. 비록, 언어의 구성요소들이 언어사용능력을 직접 보여 주지는 않지만, 언어 학습자의 文語 및 口語의 의사소통능력의 기본 바탕이 된다는 점과 분석적으로 진단하는 데 도움이 된다는 점에서 이런 언어 요소 능력들을 독립적으로 측정할 필요도 있다. 진정한 화용론적인 능력의 측정시험(implicature test)은 이론의 미정립으로 타당성이 증명된 시험의 개발이 현실적으로 아직 어려운 단계이다.

시험 출제단계에서 반드시 기억해야 할 사항은 무슨 종류의 언어 시험이건 간에, 좋은 문맥의 제공은 매우 중요하다는 사실이다. 선다형 시험의 시험개발은 먼저 측정 언어능력의 세부요소를 선정하고 나서, 그 언어능력 측정에 적합한 유형을 시험유형 및 난이도를 결정한 후, 그에 적합한 내용문맥을 선별하고 난 후, 적절한 오답지를 제작하는 순서로 진행된다(Madsen, 1983).

아래 제시된 문항은 대규모 수험집단을 대상으로 표준화된 외국어능력 검정시험의 객관식 시험유형들을 제시하기 위한 예시이므로 여기에 제시된 시험유형 이외에도 다양한 형태의 시험유형이 있을 수 있다. 또한 시험유형의 예시이므로 예시문항의 난이도는 매우 다양한 수준으로서 절대적인 척도가 될 수 없다.

1. 문법 시험

문법 시험은 어미변화로부터 구문까지의 정확성에 관한 능력을 측정하는 시험이다. 종합적 의사소통능력에 가장 기본적인 요소인 문법능력을 속도시험방식을 통해 측정함으로써 의사소통능력의 유창성까지 간접적으로나마 진단할 수 있다.

1. 1. 완성형

〈예〉 The old man is_____his dinner.

[1] eating ② ate ③ eats ④ eaten

1. 2. 세부적 오류지적형

〈예〉 TOEFL의 Structure 시험의 유형으로서, 오류를 지적하는 시험이다.

One of the girls lost her wallet and don't have any money for the game.

① [2] ③ ④

1. 3. 포괄적 오류지적형

2. 2와 같은 오류지적형의 유형이지만, 선택지에 밑줄을 긋는 대신에 의미단락을 사선(/)으로 구분하는 유형이다. 밑줄 쳐진 오류의 가능성이 있는 부분만을 미시적으로 보게 하는 대신 좀 더 거시적으로 문장의 오류를 찾아낼 수 있도록 유도하는 시험 유형이다. 2. 2의 세부적 오류지적형보다 좀 더 어려울 수 있으나 단편적인 문법지식보다는 잠재적인 문법판단력을 측정하는 데 바람직한 유형이다. 문어체 및 구어체 두 가지를 모두 고려하는 것이 바람직하다.

1. 3. 1. 문어체

〈예〉 For as long as there has been rich countries and poor,/

[1]

the poor have tried to catch up/ and the rich have tried

② ③

to stay ahead./ Unfortunately, this tendency shows no

④

signs of abating in the near future.

1. 3. 2. 구어체

〈예〉

① A: John's new color notebook computer is really fast!
② B: Actually, it's not new. He's had it for a while.
③ A: Really? I haven't seen the same model here before.
4 B: Well, John said he had bought it three months ago in the U. S.

1. 4. 선택형

주어진 答肢에서 지시문에 따라 문법적으로 올바른/틀린 문장을 고르는 시험이다. 구어체 및 문어체의 격식(formality)이 상이하므로 수험자의 혼동을 일으킬 우려가 있기 때문에 구어 및 문어체의 문장을 따로 분리하여 출제하는 것이 바람직하다.

1. 4. 1. 〈예〉 다음에서 어법상 올바른 문장을 고르시오.

① I have been to the United States two years ago.
② If were it not for your help, I could not have finished it.
3 I would appreciate your responding to my letter ASAP.
④ He did recommend that she took charge of the group.

1. 4. 2. 〈예〉 다음에서 어법상 틀린 문장을 고르시오.

① How do you think I got there?
② This is too good a chance to lose.
③ Please call me when she comes back tomorrow.
4 I've been to the United States about two years ago.

2. 어휘력 시험 (단어, 숙어, 접사)

어휘력 시험에서 문장이나 문맥이 없이 단어만이 독립적으로 제시된 상태에서 어휘력을 측정하는 일은 지양한다. 어휘력 측정 목적은 구어 및 문어체 의사소통에서 사용되는 중요한 어휘의 정확한 개념파악과 고난이도의 시험유형으로는 UCLES의 FCE/CPE에서처럼 어감상 차이(뉘앙스, 함축적 의미)에 대한 이해력까지도 평가한다.

2. 1. 완성형

앞뒤 문맥을 보고 빈칸에 알맞은 어구를 채우는 유형이다.
〈예〉 한 문장의 빈칸에 맞는 단어를 고르는 유형이다.
She quickly_____her medicine.
① drank ② ate [3] took ④ had

2. 2. 동의어 선택형

TOEFL 어휘력 시험의 전형적인 유형으로 밑줄 친 단어와 뜻이 가장 비슷한 어휘나 구를 보기에서 고르는 유형이다. 전후 문맥없이 밑줄친 단어만을 보고 풀 수 있으므로 지엽적인 어휘력만을 평가할 우려가 있는 분리식 시험 유형이다.

〈예〉 He was irate when he heard about the new plans.
① interested ② surprised [3] angry ④ sad

3. 청해 시험

청해 시험의 유형을 대별하면, (1) 그림/사진 제시형, (2) 미시적 평가유형, (3) 거시적 평가유형 등이 있으며, 각 유형은 아래와 같이 세분화된다.
TOEFL처럼 유형을 몇 가지만으로 고정시킬 것이 아니라, 실제 언어사용 상황을 최대한으로 반영하는 다양한 시험유형을 개발함으로써 시험유형에 따라 달라지는 수험자의 수험전략의 영향을 최소화하도록 출제해야 한다. UCLES FCE/CPE의 청해시험에서는 구어체언어의 특성을 담은 대화를 내용으로 일반적인 시험방식 틀 안에서 다소 융통성 있는 유형(다소 긴 대화, 방송언어, 전화대화 등)으로 구성되어, 세 가지 시험유형으로 틀에 박힌 TOEFL 청해 시험보다 더 좋은 내용타당도를 지니고 있다.

주: 모든 문장은 음성언어로 제시되며, 가능한 한 선택지의 길이는 짧게 제시한다.

3. 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, 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 TSE처럼 고난이도 수준의 시험문제에도 활용되는 유형이다.

3. 1. 1. 단일 그림/사진의 적절한 묘사 선택 유형

문제지에 제시된 그림을 음성으로 묘사하는 선택지 중 정답을 고르는 유형이다.

〈예〉 문제지의 그림의 적합한 묘사를 들려주는 보기에서 고르시오.
[수업시간에 선생님이 칠판에 판서하는 장면의 그림/사진]
① The teacher is lighting something near the blackboard.
② The teacher is riding a bike to school.
③ The teacher is right in front of the students.
[4] The teacher is writing something on the blackboard.

3. 1. 2. 한 상황묘사와 일치하는 그림/사진 선택 유형

문제지에 제시된 몇 가지 그림 중에서 음성묘사와 일치하는 것을 고르는 유형이다.

〈예〉 다음의 말을 듣고 말의 내용에 맞는 그림을 고르라. 간단한 한 문장의 상황묘사 이상으로 두 명의 화자간의 대화를 통해 대화상황을 제시할 수도 있다.
(Several people are standing in line to board a bus.)
① 몇 사람이 버스안에 타고 있는 모습
② 몇 사람이 버스에서 내리고 있는 모습
[3] 몇 사람이 버스를 타기 위해 줄 서있는 모습
④ 몇 사람이 줄을 잡고 있는 모습

3. 1. 3. 일련의 상황묘사와 일치하는 그림/사진 순서 선택 유형

문제지에 제시된 몇 가지 그림을 음성묘사와 일치하도록 순서를 올바르게 배열하는 유형이다.

〈예〉 다음 말에서 나오는 이야기의 차례대로 나열한 그림순서를 선택하라.
(Paul attended a job interview yesterday. This morning, he received a telephone call informing him that he had not been selected for the ap-

pointment. This means that Paul should keep reading newspapers for job openings and writing application letters.)

A: 한 남자가 전화를 받는 모습

B: 한 남자가 소파에서 신문을 보는 모습

C: 한 남자가 다른 남자와 인터뷰하는 모습

D: 한 남자가 책상에서 무엇을 쓰는 모습

① ABCD

② CBDA

[3] CABD

④ BADC

3. 1. 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.

〈예〉 [다리에서 일어난 자동차 사고 장면의 사진/그림]

(Q: How could this traffic accident have been avoided?)

① No, it wasn't on the street.

② Several people may have been hurt.

[3] Probably with more careful driving.

④ Sure, it took place on the bridge.

3. 1. 5. 길/방향 찾기: 그림 제시

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

〈예〉 다음 대화를 듣고 여자가 찾아가고자 하는 곳을 지도에서 고르라.

W: Excuse me. Could you please tell me how to get to the nearest post office?

M: Yes. Well … Walk this way to Orchard Drive and turn right.

W: Uh-huh. Orchard Drive and turn right.

M: Yes. Then, go two more blocks to Main Street.

W: Okay, Main Street.

M: Then, turn left on Main Street and look for the post office on your

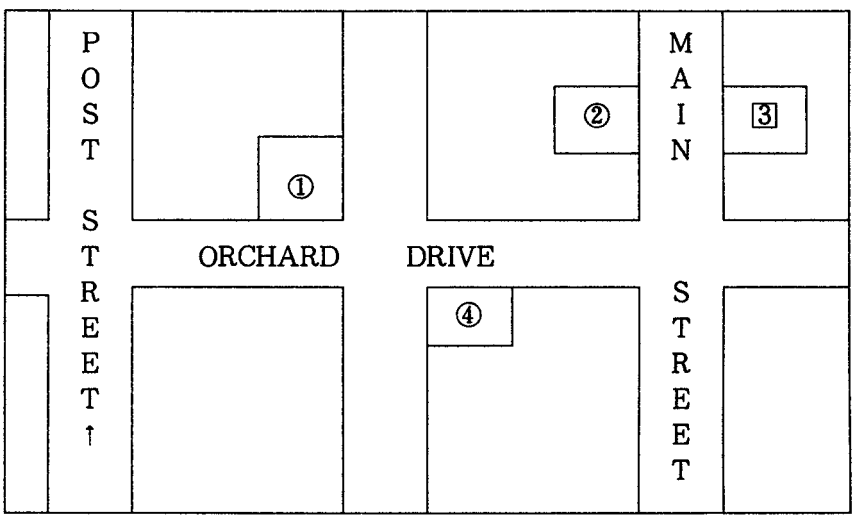
right.

Have you got that?

W: Yeah. Turn left on Main Street and the post office is on my right.

M: Exactly.

W: Thanks. I appreciate it.



3. 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 의사소통능력의 가장 기본적인 발음식별능력을 측정하는 미시적인 측정 유형이긴 하나, 문맥이 결여되어 있으므로 natural phonology의 내용타당도 측면에서 의사소통을 위한 청해능력의 시험유형으로는 미흡하다.

3. 2. 1. 들려진 표현 선택 (단문)

구어의 가장 기본적인 발음식별 및 청각이미지 기억력을 평가하는 초보적인 능력의 평가유형이다.

〈예〉 다음에 들려진 말과 똑같은 표현을 고르시오.

(I've gotta pass the test this year.)

① I got a pass of dish test.

② I got to pass the test of dish.

③ I've got a past test of dish ear.

[4] I've got to pass the test this year.

3. 2. 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 평가하는 데 유용한 표현이다.

〈예〉 (I'll keep my fingers crossed.)

[1] I'll wish you luck.
② I have hurt my fingers.
③ I'll have my fingers treated.
④ I'll keep my promise.

3. 2. 3. 적절한 응답 선택 (단문)

대화자가 단문(single exchange)을 말한 대화에서 질문에 대한 적절한 응답을 생각하게 함으로써, 종합적인 의사소통능력인 청해력을 평가하는 유형 중 가장 기본적인 시험유형이다.

〈예〉 (Are they coming soon?)

[1] No, I don't think so.
② No, they haven't.
③ No, I am not coming soon.
④ Yes, they do.

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

의미단위(meaningful chunk)의 대화 및 담화/강의 등을 구두로 제시한 다음, 수험자가 문제지에서 적절한 선택을 고르게 하는 형태의 시험이다.

3. 3. 1. 시간, 장소, 전/후 상황 추론 (짧은 대화)

1-2 대화교환(exchange) 정도로 들려진 두 사람간의 대화에 관한 제3자의 질문에 대한 적절한 대답을 고르는 유형으로서 간략한 의사소통상황을 반영하여 발음식별, 억양, 구어표현, 화자의 태도 유추 등의 청해력을 평가하는 시험이다.

〈예〉 (M: Should I fill it?
W: Yes. And would you also check the oil?
M: Okay. I'll see if the tires are up, too.
Q: Where are these people?)
① At a bank.
[2] At a gas station.
③ At a police station.
④ At a department store.

3.3.2. 화자-청자의 관계 유추 (짧은 대화)

들려지는 두 사람간의 대화로부터 관계를 유추하는 유형이다.

〈예〉 (W: What seems to be the problem?
M: I'm not sure, but I'm afraid I have broken my finger.
W: I'm afraid the doctor is not available right now. He's out to lunch. He told me he'd be back in about 15 minutes. Can you wait?
M: Well, I think I've got no other choice.
Q: What is the relationship between the two speakers?
① doctor-nurse
② friend-friend
[3] nurse-patient
④ doctor-waitress

3.3.3. 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 국부독립적(locally independent)이어야 한다.

〈예〉 (A: Check in here?
B: Sure. Can I see your ticket, please?
A: Here it is. I'm going to L. A.
B: OK. Smoking or nonsmoking?
A: Nonsmoking, please. By the way, can I get an aisle seat?
B: Well, all right. That's 8-D. Put your bag on the scale, please.

A: Sure.
B: I'll check it straight through. Here's your boarding pass. That'll be Gate A-16. They'll be boarding in about an hour.
A: Oh, really? I mean, is it late?
B: About thirty minutes.
A: And, uh, is it a breakfast flight?
B: Yes, it is. Have a good day. Next, please.)
(Q: What time of day is it?)
[1] morning
② afternoon
③ evening
④ night
(Q: What is the customer doing?)
① Buying an airplane ticket.
② Checking on someone's flight.
[3] Checking in at an airline terminal.
④ Seeing a friend off at an airport.

3. 3. 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.

〈예〉 (W: I wonder if you could stop smoking in here.
M: I think I have the right to smoke.
W: Can't you see the No-Smoking sign? (to be pronounced sarcastically)
(Q: What is the woman's attitude toward the man?)
① satisfied
② grateful
③ apologetic
[4] accusatory

3. 4. 받아쓰기 (Dictation)

적절한 시험 기법을 통한 받아쓰기 시험은 청해능력뿐만 아니라 언어의

전반적인 능력을 가장 잘 대표한다고 밝혀졌다(Savignon, 1982; Oller and Streiff, 1975).

시험 진행 방식으로는, 일반적으로 지문을 세 번 읽어주게 된다. 첫번째는 정상속도로 읽혀지며 수험자는 듣기만 하고 대의를 파악하도록 해야 하며, 두번째는 의미단락으로 끊어서 읽혀지며, 각 의미단락간에 있는 휴지(pause) 기간에 수험자들은 답을 받아 적게 되며, 마지막 세번째는 수험자들이 교정을 볼 수 있도록 지문을 다시 한 번 들려 준다.

3. 4. 1. 전체 받아쓰기 (Full Dictation)

Cziko(1983)와 Bachman이 제작한 외국 유학생들에게 치뤄지는 영어시험, English Placement Test(EPT: University of Illinois at Urbana-Champaign)의 받아쓰기 시험의 단락제시(segmenting) 채점 방식은 매우 독창적이며 대규모 시험의 효율적인 타당성이 높은 채점 방식으로서 시사하는 바가 매우 크다.

(가) 2단어로 가장 짧은 부분으로부터 21단어로 제일 긴 부분의 단락을 들려줌으로써 기억의 한계에 도전을 줌으로써, 기억력의 측정이 아닌, 진정한 언어능력의 개인별 차이를 극대화할 수 있는 N-R시험의 장점을 살렸다.

(나) 이렇게 상이한 단락이 타당한 Guttman scale이 될 수 있으므로, 개인의 총점이 낱낱의 단락의 답에 대한 좋은 예측치가 될 수 있다는 점이다. 즉 개인별 점수가 그 자체로서 의미가 있다는 점에서 C-R 시험의 장점이 있다는 것이다.

(다) 정확철자(exact-spelling), 단락 채점 방식의 채택으로 오자 허용(acceptable spelling), 단어별 채점(word-by-word) 방식보다 3~4배 시간이 절약되었다.

(라) 전자의 방식과 후자의 방식으로 채점한 결과의 상관 관계는 .92로 매우 높았다.

〈예〉 (UCLA의 외국인 유학생 청해시험. /에서 休止기간을 둔다. 물론 休止 사이의 간격은 수험자의 수준에 따라 다르다.)

A taste for music,/ a taste for anything,/ is an ability to consume it/ with pleasure./ Taste in music/ is preferential consumption/; a greater liking/ for certain kinds of it/ than for others./ A broad taste in music/ involves the ability to consume/ with pleasure/ many kinds of it./ Vast numbers of persons,/ many of them highly intelligent,/ derive no pleasure at all/ from organized sound./ An even larger number/ can take it

or leave it./ They find it agreeable/ for the most part,/ stimulating to the sentiments,/ and occasionally interesting to the mind.

3. 4. 2. 부분 받아쓰기 (Partial/Spot Dictation; Oral Cloze)

채점의 효율성을 감안하여, cloze와 전체 받아쓰기의 시험 형식을 절충한 형태의 시험 유형이다. 단답식의 채점 방식을 도입한다면 대규모 시험에 도입할 가치가 있는 시험 방식이다.

〈예〉 (괄호 부분이 받아쓰기 부분: /에서 休止기간을 둔다.)

A taste for music,/ a taste for anything,/ is (an ability to consume it/ with pleasure)./ Taste in music/ is preferential consumption/; a greater liking/ for certain kinds of it/ than for others./ A broad taste in music/ (involves the ability to consume)/ with pleasure/ many kinds of it./ Vast numbers of persons,/ many of them highly intelligent,/ (derive no pleasure at all)/ from organized sound./ An even larger number/ can take it or leave it./ (They find it agreeable)/ for the most part,/ stimulating to the sentiments,/ and occasionally interesting to the mind.

4. 독해 시험

기존 시험에서 주로 평가하는 세부내용 파악능력이나 주제 및 대의 파악능력 등의 사실적 이해력뿐만 아니라, 담화/문맥(discourse)의 논리적 추론능력 및 필자의 어조 파악능력 등의 고차적인 독해력까지 망라하여 전반적인 문어 의사소통능력을 측정하는 다양한 독해 시험 유형은 다음과 같다.

4. 1. 지문 (reading passage) 이해 유형

주제(topic) 및 대의(main idea) 파악이나, 세부내용 파악능력, 논리적 유추능력, 작가의 태도/어조 파악능력 등의 다양한 능력을 측정하는 독해력의 전형적인 시험 유형이다.

4. 1. 1. 주제파악

〈예〉 Choose the best topic of the following passage.

"Man since the beginning of time has sought peace," declared General Douglas MacArthur. Indeed there has been noteworthy efforts to attain that objective. But another famous general, Charles de Gaulle, observed

"The sword is the axis of the world." What a paradox that the statements of both are accurate!

Peace, however men may define it, has been sought after. But the sword has more usually been used to resolve differences between nations.

① statements of two generals
② the history of peaceful war
③ accurate definition of peace
[4] paradoxical means of achieving peace

4. 1. 2. 대의 파악

〈예〉 다음 글의 대의로 가장 적절한 것은?

A look at history clearly shows that even enduring empires do pass away.

Egypt became the greatest empire at the dawn of human history. It fell. The mighty Assyrian empire and its capital of Nineveh were utterly destroyed.

Sometimes nations seem to fall with stunning rapidity after just having climbed to the height of power. Babylon is one example. The awesome Roman empire—the most invincible empire the world has known—had its ups and downs for centuries.

[1] No empire is forever.
② History repeats itself.
③ History witnessed many empires.
④ Roman empire was the most invincible.

4. 1. 3. 세부내용 파악

〈예〉 다음 글의 내용과 일치하는 것은?

Social security systems varies from state to state. Medical care system, for instance, is very different in England and in the United States. In England medical care is national, which means the government pays for all medical care. On the other hand, the United States has a private system, in which people have to pay their own medical care bills. English doctors, dentists and hospitals are free. In the U. S., however, they are quite expensive. Everyone in England gets medical care,

ever the poor people. That is not always the case in the U. S. Only Americans with a lot of money can afford to get good medical care. Poor Americans sometimes manage to get financial help from the government.

[1] Medical care is free of charge in England.
② Even the poorest Americans can afford good medical care.
③ American medical care is supported mainly by the government.
④ Only the rich English people can afford expensive medical care.

4. 1. 4. 논리적 유추

〈예〉 윗글이 나올 가능성이 가장 큰 곳은?

A short drive brought Mary to the electronic highway. She entered her destination into the computer. The car's computer automatically sent this information to the highway control computer. Because traffic was not heavy, the computer immediately put the car into automatic to enter the Superway.

① historical document
[2] science fiction
③ maintenance manual
④ computer catalogue

4. 1. 5. 전후 문맥 유추

〈예〉 다음 글의 바로 뒤에는 다음 중 어떤 내용이 있겠는가?

Who could have dreamed a decade ago that a reform program called perestroika could have caused such a chain reaction? Now that political reforms are in place, what challenges face the peoples of the shattered socialist realm?

And where will the political and religious changes and the economic troubles of this largely Slavic region most likely lead?

Choose the one that will most probably follow the above passage.

① capitalism's economic laws
② reasons for Communism's collapse
③ advantages and disadvantages of socialism
[4] predictions of former socialistic world's future

4. 1. 6. 작가 태도/어조 유추

독해도 필자와 독자간의 글을 통한 의사소통상황임을 고려해 볼 때 작가의 태도/어조 유추는 중요한 독해능력 요소이다.

〈예〉 What is the author's attitude toward Karl Marx?

Your article "Communism–What went wrong?" was so frank and in depth.

Most leaders of workers' (trade) unions in the Third World countries—especially Africa—have advocated communism as a paradise for the impoverished working class and peasants; but now I and many other readers know better.

So long as man will be the leader, there is no hope in communism for equitable distribution of wealth and resources. It will take only God to implement the lofty ideology of Karl Marx.

① optimistic
② supportive
[3] accusatory
④ complimentary

4. 2. 중요 어구 완성형

독해지문의 대의 파악에 핵심적인 중요 어구를 찾아내는 빈칸 넣기 유형의 선다형 시험으로서, 대의파악, 글의 응집성(coherence: 내용; cohesion: 표현), 논리적 사고력 등을 평가하는 데 타당하며, 각 지문당 하나의 문항을 출제하므로 상대적으로 많은 글의 독해를 유도함으로써 속독능력을 요구하는 일석이조의 시험유형이다. 자연스러운 독해과정을 유도하기 위해 빈칸은 지문의 앞 부분을 피하고 가능한 지문 후반에 제시한다. 괄호의 전후에 주어진 문맥이나 맞춤법 등을 통해 정답을 유추하지 못하도록 각별히 유의하여 출제해야 한다.

4. 2. 1. 중요 어구 빈칸채우기 (Coherence)

선다형 지문이해 시험유형의 타당도는 오답지의 내용의 질에 따라 절대적인 영향을 받는 데 반해, 본 시험유형은 간단한 어구로 구성된 오답지에 따른 영향이 다소 감소되어 선다형 시험유형 중에 타당도의 문제점이 가장 적은 시험 유형이다. 선다형 지문이해 유형과 병행하여 사용될 수 있는 좋

은 시험유형이다.

〈예〉 What is the most suitable for the blank?

After a naval battle, it was customary for the defeated ship to lower its flag to signal ( ) and make room for the victor's flag above it. After this tradition faded, a lowered flag continued to be used as a universal sign of despise.

① compromise
[2] surrender
③ alliance
④ victory

4. 2. 2. 중요 접속사/연결사 채우기 (Cohesion)

글의 내용을 논리적으로 연결하기 위해 필요한 접속사/연결사(cohesive marker)를 채우는 방식으로서 언어의 중요한 요소인 논리적 사고력의 연결을 위한 표현의 능력을 평가하는 유형이다.

〈예〉 What is the most suitable for the blank? (다음 글의 내용의 빈 칸에 들어갈 알맞은 것을 고르시오.)

Entertainment doesn't have to be a time-waster or filler. It can be a profitable way to enrich and improve life. One can become a better person for it. _____being frivolous or unnecessary, certain entertaining activities can help maintain balance and health in life. One never need feel guilty about enjoying good entertainment. We all need a regular measure of enjoyable relaxation and an uplifting change of pace in our lives.

① As for
[2] Far from
③ In addition to
④ Outside of

4. 2. 3. 전후문맥 유추 빈칸 채우기 (Coherence/Inference)

주어진 글의 처음이나 끝에 주어진 빈칸에 들어갈 가장 적합한 말을 고르는 방식으로서, 대의 파악을 한 후 전후 문맥을 유추할 수 있는 논리적 능력을 평가하는 유형이다.

〈예〉 What is the most suitable for the blank? (다음 글의 내용의 빈칸에 들어갈 알맞은 것을 고르시오.)

It is harder to learn some things when you get older. Your brain does not work as quickly. That is why we sometimes say,"_____"

① A watched pot never boils.
② What is done cannot be undone.
[3] You can't teach an old dogs new tricks.
④ You can't have your cake and eat it too.

4. 3. 문장 배열/제거/삽입

응집에 대한 이해력은 독해력뿐만 아니라 작문력과도 관련된 다차원적인 언어능력의 중요한 요소이므로, 작문능력의 간접적 평가의 성격도 지니고 있다.

내용의 응집성(coherence)을 측정하는 시험 유형에는 재배열, 제거, 삽입 등 세 가지로 대분된다. 지문의 내용을 암시하는 첫 한두 문장은 선택지에서 반드시 제외하여 문제를 푸는 실마리 글(clue statement)로서 제시되어야 한다.

4. 3. 1. 문장 순서 배열

담화의 중요한 요소인 응집(coherence) 및 논리성의 이해력 평가하기 위한 시험 유형으로서 논리적 일관성이 존재하는 글을 구성하는 문장으로 분리하여 순서를 임의로 섞어 놓은 후 다시 논리적으로 명백한 순서로 재배열하도록 하는 시험유형이다. 전체의 글을 지나치게 많이 분류하여 순서를 섞으면 논리적 응집성의 재배열이 매우 어려우므로 전체 글을 3~4 단락 정도로 분류하여 조합한 후 선택지를 만든다.

〈예〉 주어진 문장으로 시작하는 다음의 글을 논리적으로 가장 알맞게 배열한 순서를 고르라.

In the United States, 49,000 people were killed on the nation's roads in 1988.

(A) That is some 1,600 more than the number of Americans killed in battle during the entire Vietnam conflict. But U. S. roads are by no means the most dangerous.

(B) By comparison, the U. S. rate for the same year was 19.2 and in Chile 6.5.

(C) According to Pacific Islands Monthly, the traffic accident death rate in Tahiti for 1985 was 55.

① A–B–C
② B–C–A
③ C–A–B
④ A–C–B

4. 3. 2. 불필요 문장 제거

글의 응집에 대한 이해력을 평가하는 또 다른 방법으로서 전체 글의 흐름과 무관한 문장을 제거하는 유형이다.

〈예〉 다음 글 속에서 전체의 흐름과 무관한 문장 하나를 고르라.

When I was walking down the street yesterday, I happened to notice a small wallet lying on the sidewalk.

① Many people liked to carry a small wallet.
② I picked it up.
③ I opened it to see if I could find out the owner's name.
④ There was nothing inside it except some change and an old photograph.

4. 3. 3. 필요 문장 삽입

글의 내용의 논리적 일관성을 유지하기 위해 필요한 문장을 적절한 곳에 삽입하는 시험 유형이다.

〈예〉 다음 글의 내용으로 볼 때 아래 박스 안에 있는 문장이 들어갈 수 있는 가장 적절한 곳은?

American and British people both speak English of course. But sometimes it does not seem like the same language. In fact, there are some significant differences between British and American English. 〈가〉 First of all, they sound very different. Often, Americans say several words together. They may say "Whaddya say?" instead of "What did you say?" However, the British are more careful in their speech. 〈나〉 This may have to do with their cultural differences. 〈다〉 Words sometimes have different meanings too. Some American words are never used in England. 〈라〉 The same thing is true of some British words in

America. For example, the back of a car is called trunk in America, whereas in England it is a boot. All these differences can be confusing if you are learning English. But most languages are like this.

Sound is not the only difference between British English and American English.

① 가 ② 나 [3] 다 ④ 라

4. 4. Cloze 관련 유형

타당도에 대한 논란이 많은 cloze시험은 채점의 비실용성때문에 대규모 시험에서 활용이 불가하겠지만, 채점의 실용성을 제고한 Cloze의 변형된 선다형 유형은 대규모 시험에서 활용가능하다.

4. 4. 1. Cloze

1) 출제 방식
   (가) 고정비율(fixed ratio): 매 n (통상 5-10)번째 단어를 생략한다.
   (나) 변동비율(rational/variable ratio): 지문 내용과 시험 목적을 고려하여 적절한 단어를 생략한다.
2) 진행 방식
   수험자들에게 전체 지문을 한 번 읽게 한 후 빈칸을 메꾸게 한다.
3) 채점 방식
   (가) 적확 어휘(exact word): 원문으로부터 얻어진 답
   (나) 적합 어휘(acceptable word): 원어민에 의해 적절하다고 인정된 답수험자들의 심리적인 부담 외에는 통계적으로 별 차이 없다고 한다 (Stubbs & Tucker, 1974).

4. 4. 2. 선다형 Cloze

수험자들의 수준과 지문의 난이도를 고려하여 정답과 distractor를 보기로 주어서 선다형의 형태를 취할 수도 있다(Heaton, 1988). 객관성 및 실용성을 고려해 볼 때 대규모의 시험상황에서 충분히 가용한 시험유형이다.

〈예〉 Is your home safe? Do 1) it [2] you 3) them 4) each protect your family? Are you safe 1) of 2) for 3) with [4] from fire or falling? Do you guard against 1) watching 2) knowing [3] eating 4) sending unsafe food or drinking poison? It 1) perhaps [2] is 3) always 4) so

important to keep your home safe [1] and 2) since 3) because 4) for develop good safety habits.

4. 4. 3. C-Test

Cloze 시험의 변형으로서 Klein-Braley & Raatz(1984)에 의해서 1981년에 개발되어 C-test라고 명명되었다. C-Test란 [2의 규칙(the rule of 2)]을 바탕으로 5~6개의 다소 짧은 지문으로 구성된 시험. 두번째 문장부터, 매 두째번 단어의 후반부가 생략된 시험 형태이다.

장점으로는, (1) exact word scoring만이 가능하여 채점이 매우 효율적이다.

(2) 몇 개의 다른 지문을 사용하여 적어도 100개의 많은 문항을 얻게 됨으로써 높은 신뢰도를 얻게 된다.

(3) Cloze보다 수험자들에게 심리적으로 부담감을 덜 준다.

단답식의 채점방식이기는 하나, 대규모 시험에서 답안지가 기계채점(machine scoring)을 가능하도록 제작되어야 하는 문제점이 있다.

〈예〉 There are usually five men in the crew of a fire engine. One o__ them dri__the eng__. The lea__ sits bes__ the dri__. The ot__firemen s__ inside t__cab o__ the f__ engine. T__ leader h__ usually be__ in t__Fire Ser__ for ma__ years. H__ will kn__ how t__ fight diff__sorts o__s fires. S__, when t__ firemen arr__ at a fire, it is always the leader who decides how to fight a fire. He tells each fireman what to do.

4. 4. 4. C-Test 변형

어휘의 후반부를 빈칸으로 두 단어마다 빈칸을 제시하는 유형인 C-test의 변형으로서 빈칸의 첫글자를 제시하여 채점의 실용성을 높이며 cloze보다는 쉽고 C-test보다는 어렵게 난이도를 조정함으로써 cloze와 C-test의 약점을 보완한 유형이다. 빈칸의 빈도는 rational ratio로 하는 것이 바람직하다.

〈예〉 American and British people both speak English of course. But sometimes it does not seem like the s_____ language. In fact, there are some significant d____ between British and American English.…

참고문헌

Bachman, Lyle F. (1990) *Fundamental Considerations in Language Testing,*

Oxford: Oxford University Press.

Bachman, Lyle F. and Palmer Adrian (1981) 'The Construct Validation of the FSI Oral Interview,' *Language Learning* 31.1, 67–86.

Bachman, Lyle F., Fred Davidson, Kathy Ryan, and Inn-Chull Choi (in press) *An Investigation into the Comparability of Two Tests of English as a Foreign Language: The Cambridge-TOEFL Comparability Study*, UCLES.

Brown, James Dean (1993) 'A Comprehensive Criterion-referenced Language Testing Project,' In Carol Chapelle and D. Douglas (eds.), *A New Decade of Language Testing Research*, Alexandria: TESOL, 163–184.

Canale, Michael (1983) 'On Some Dimensions of Language Proficiency,' In John W. Oller, Jr. (ed.), *Issues in Language Testing Research*, Rowley, MA: Newbury House, 333–42.

Choi, Inn-Chull (1989) 'Past, Present, and Future of Language Testing,' *English Teaching* 38, 95–135, The College English Teachers Association of Korea.

Choi, Inn-Chull (1991) *Theoretical Studies in Second Language Acquisition: Application of Item Response Theory to Language Testing*, New York: Peter Lang Publishing Inc.

Choi, Inn-Chull (1992) 'Interplay of Natural Phonology and Sociolinguistic Variable of Formality,' *English Teaching* 44, 65–88, The College English Teachers Association of Korea.

Choi, Inn-Chull (1993) 'Construction Validation Study on SNUCREPT (Seoul National University Criterion-Referenced English Proficiency Test),' *Language Research* 29.2, Seoul National University.

Cummins, J. P. (1983) 'Language Proficiency and Academic Achievement,' In John W. Oller, Jr. (ed.), *Issues in Language Testing Research*, Rowley, MA: Newbury House, 108–26.

Cziko, Gary (1983) 'Psychometric and Edumetric Approaches to Language Testing,' In John W. Oller, Jr. (ed.), *Issues in Language Testing Research*, Rowley, MA: Newbury House, 289–308.

Dubin, F, Eskey D. E., and Grabe W. (1986) *Teaching Second Language Reading for Academic Purposes*, Addison-Wesley Publishing.

Goodman, Kenneth S. (1967) 'Reading: A Psycholinguistic Guessing Game,' *Journal of the Reading Specialist* 4, 126–135.

Gronlund, Norman (1985) *Measurement and Evaluation in Teaching*, (5th ed.) NY: Macmillan Publishing Company.

Heaton, J. B. (1988) *Writing English Language Tests*, New York: Longman Inc.

Klein-Braley, Christine & Raatz, Ulrich (1984) 'A Survey on Research on the C-test,' *Language Testing* 1, 2, 134-146.

Krashen, Stephen D. (1985) *The Input Hypothesis*, London: Longman Inc.

Lado, Robert (1961) *Language Testing*, NY: McGraw-Hill.

Madsen, Harold S. (1983) *Techniques in Testing*, NY: Oxford Univ. Press.

Oller, John W. Jr. (1979) *Languages Tests at School: A Pragmatic Approach*, London: Longman Inc.

Oller, John W. Jr. & Virginia Streiff (1975) 'Dictation: A Test of Grammar Based on Expectancies,' *English Language Teaching* 30, 25-36.

Rivers Wilga (1980) *Teaching Foreign Language Skills*, MA: Harvard Univ. Press.

Rivers, Wilga and Mary Temperley (1978) *A Practical Guide to the Teaching of English*, NY: Oxford Univ. Press.

Rost, Michael (1990) *Listening in Language*, New York: Longman Press.

Savignon, Sandra J. (1982) 'Dictation as a Measure of Communicative Competence in French as a Second Language,' *Language Learning* 32, 33-51.

Smith, Frank (1982) *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read* (2nd ed.), NY: Holdt, Rinehart and Winston.

Stubbs, Joseph b. & Tucker, G. Richard (1974) 'The Cloze Test as a Measure of English Proficiency,' *The Modern Language Journal* 58.5, 239-241.

Ur, Penny (1984) *Teaching Listening Comprehension, Cambridge*: Cambridge Univ. Press.

ABSTRACT

A Model of Foreign Language Proficiency Test Development

Inn-Chull Choi

In globally competitive modern societies where a good command of foreign language is considered a prerequisite, it is increasingly important to develop a valid tool by which to measure language proficiency from the perspective of criterion-reference as well as norm-reference. It should be pointed out, however, that the commercially available language test batteries have been lacking in sound framework of language test development. Hence, they have failed not only to fulfil the rigorous criteria of language testing theories but also to meet the social expectation of valid measurement process.

The present study presents a desirable model intended to develop valid and reliable measures to be employed by Seoul National University Language Research Institute. The pilot tests developed on the basis of this framework will go through systematic process of content, criterion-related and construct validation in further studies.

This paper deals with the basic concepts of communicative language ability components and language testing. Thereafter it elaborates on some key issues of the fundamental considerations of Test Method Facets (TMF; Bachman, 1990) in terms of test contents, and tasks. Finally, it provides the example question items pertaining to each task category.

151-742
서울시 관악구 신림동
서울대학교 어학연구소