

영어 학력 테스트의 내용과 방법

계 명 일*

- I. 서 론
- II. 학력 테스트의 내용
- III. 학력 테스트의 방법
- IV. 테스트 작성상의 문제점

I

현대에 있어서 국제교류의 양상에는 급속한 양의 증대와 더불어 심대한 질의 변화가 일어나고 있다. 교류의 주요한 매개체로서의 언어의 기능에도 질과 양의 변화가 발생하고 있는 것은 필연적인 현상이라고 할 수 있다. 언어학이나 외국어 교육이 과학으로서 또한 기술로서 괄목할 만한 발전을 이룩하고 있는 것은 언어 기능의 이러한 변화에 기인하기 때문일 것이다.

외국어 학습에 관해서도 학습자의 모국어와 학습하려는 외국어(Target language)의 언어구조와의 비교 또는 문화구조와의 비교를 통해서 학습의 문제점을 해명하고 거기에 근거하여 학습체계를 조직해야 할 것이며 학습체계에 있어서 테스트의 기능은 지극히 중요한 것이라고 말할 수 있다.

일반적으로 테스트에 있어서는 학습의 결과를 측정하는 기능이 가장 중요시 되며 입시에 있어서는 능력의 측정이나 선발의 목적이 가장 중요시된다. 그러나 테스트에는 학습의 결과를 측정하거나 정착시키는 기능 뿐만 아니라 다음 단계의 방향을 제시하는 기능도 있는 것이다.

테스트만큼 학습자의 심리에 강한 영향을 미치는 학습형태는 없다고 말할 수 있다. 따라서 방향제시의 기능에는 어느 면으로는 단순한 측정 이상으로 중요한 의미를 지니고 있다고 보아야 할 것이다. 필자는 테스트를 학습체계에 있어서 과거와 미래의 점점인 현재의 기능으로써 설정해야 한다고 생각한다. 대학입시도 중·고에서의 과거의 학습과 대학에서의 미래의 학습을 연결하는 현재의 점점으로써의 기능이 능력측정이나 선발에 못지 않게 중요하다 할 것이다.

언어 테스트는 언어학습에 있어서 가장 중요한 인자(Factor)이며 그 본질을 언어과학에 의해

* 文理科大學(서울) 英文學科 教授

서 해명하고 그 기법을 과학기술에 의해서 확립해야 할 시대가 도래한 것이다.

오늘날 대학입시가 고교의 교육을 빗나가게 하고 있다는 비난이 얼마간은 해소가 되었다고 하지만 아직도 학생들의 진정한 영어학력을 측정하고 있다고는 할 수 없다. 이것은 시험이 타당성과 신뢰성을 상실하고 단순히 「떨어뜨리기 위한」 수단에 지나지 않는다는 의미로 해석할 수도 있는 것이다. 따라서 필자는 이 논문을 입학시험에만 한정시키지 않고 일반적으로 공정한 영어 학력 테스트의 내용과 방법의 당위성을 규명해 보기로 한다.

II

오늘날은 테스트 교육 또는 테스트 인간의 시대라고 불리울 정도로 각양각색의 테스트가 거대한 힘으로 교육계를 지배하고 있다. 테스트란 과연 무엇인가에 대해서 근원적인 질문을 제기하지 않을 수 없게 된다. 교육 가운데서 평가의 역할, 평가 가운데서도 테스트가 차지하는 위치가 명백히 규정되어야 할 것이다. 간단히 말하자면 평가는 학습 지도의 방향 내용 및 방법의 진단에 의해서 학습과 교육활동을 추진시켜 나가는 것이며 테스트는 측정법의 유력한 하나의 방법으로써 객관적으로 수량화 할 수 있다는 점에서 평가활동에 큰 의미를 부여하고 있다.

영어 학력 테스트는 아래에 제시하는 작성과정에 나타나 있는 바와 같이 올바른 내용과 방법에 따라 실시되지 않으면 안된다.

- ① 학력관의 확립
↓ (학력구조의 분석과 통합)
- ② 테스트 내용의 결정
↓ (학력을 반영하고 학력 전체로부터 넓은 분포에 걸친 테스트 내용을 sampling하여 결정한다)
- ③ 테스트 방법의 결정
↓ (테스트 내용을 어떠한 방법 어떠한 질문과 해답의 미디어를 써서 테스트할 것인가를 결정한다)
- ④ 타당화(Validitation)
↓ (문제항목 분석(Item analysis) 등에 의해서 타당화 한다)
- ⑤ 테스트의 구성

영어 학력 테스트의 문제항목은 영어 학력이라고 하는 모집단(Population)에서 따내온 표본(Sample)이다. 이 표본은 모집단을 훌륭하게 대표하는 것이 되지 않으면 안된다. 영어 학력의 내용을 테스트할 수 있는 성분(Testable components)으로 분석 기술함으로 해서 내용적으로는 타당한 학력 테스트가 가능하게 된다. 이러한 학력 테스트로써 우수한 성적을 올리는 것이 학력이 높다고 결론지을 수 있는 것이어야 한다. 여기서 내용적으로 타당성이 높은 테스트를 가능하게 하기 위해서는 영어 학력에 관해서 다음과 같은 방법에 의한 연구가 필요하다. 첫째

인자분석에 의한 방법, 둘째 언어 및 언어활동의 구조분석에 의한 방법, 셋째 언어학습의 문제점의 분석에 의한 방법, 넷째 언어능력의 이론적 모델에 의한 방법 등이다.

영어 학력을 네 가지 기능(4-skills)에 국한시킬 수는 없다. 네 가지 기능의 면에서 영어 학력에 아프로취하는 것은 가능한 일이지만 그것이 모두는 아니다. 네 가지 기능은 상호간에 어떠한 관계가 있으며 네 가지 기능을 양성하기 위해서는 어떠한 습득 과정을 밟아야 하는가 하는 문제는 앞으로 규명되어야 할 것이지만 현재로서는 Lado의 견해를 채택하는 것이 무리가 없을 것이다.

Lado는 외국어 학력을 반영하는 것으로서 언어 테스트의 변수(Variables)를 다음과 같이 정리하고 있다.¹⁾

(1) 언어요소(Elements)

- 1) 발음...sound segments, intonation and its border, stress and its sequences
- 2) 문법구조...morphology, syntax
- 3) 어휘(Lexicon)
- 4) 문화적 의미(Cultural meanings)

(2) 종합기능 테스트(Skills)...listening, speaking, reading, writing, translation

(3) 종합적 언어제어력 테스트(Over-all control of the language)

(4) 언어를 넘어서는 것(Beyond language)...cross-cultural understanding, higher values

이와 같은 변수를 될 수 있는 한 많이 도입하는 것이 필요하겠지만 여기서 문제가 되는 것은 elements와 skills의 관계이다. 이에 대해서 Lado는 다음과 같이 설명하고 있다.²⁾

기능과 요소의 선택의 결정은 테스트가 제작되는 목적과 테스트를 위해서 사용되는 시간, 게다가 테스트에 주어지는 조건에 크게 의존하고 있다. 만약 학생이 습득해야 할 문제점들을 집중적으로 조사하기 위해서 쓰이는 포괄적인 발음의 진단테스트를 하려고 한다면 명백히 음성체계의 전영역에 걸치는 테스트, 말하자면 언어의 그 요소에 대한 특별한 테스트를 선택하게 될 것이다...

만약 listening 만이든가 혹은 개개의 기능만이 아니면 일반적인 학력을 결정할 필요가 있으며 또한 포괄적인 테스트를 하는데 충분한 시간이 허용된다고 할 것 같으면 거기에다가 종합기능(Integrated skills)을 테스트하는 영역을 얼마간 설정하여 언어의 몇개의 요소를 망라하도록 선택하면 좋을 것이다.

여기서 결론은 요소와 기능을 각각 따로 테스트할 필요가 있으며 테스트의 목적과 조건에 따라 전자 또는 후자 혹은 그것들을 모두 합친 것을 선택하게 되리라는 것이다.

그동안 수많은 외국어 학력 테스트가 표준화되어 출판되었지만 4기능을 모두 망라하고 있는 테스트는 (1) MLA Foreign Language Proficiency Tests for Teachers and Advanced Stu-

1) Robert Lado, *Language Testing* (Longmans, 1961) Part II—IV.

2) *Ibid.*, pp. 27—28.

dents³⁾ (2) MLA Cooperative Foreign Language Tests⁴⁾ (3) Pimsleur Modern Foreign-Language Proficiency Tests⁵⁾ 등 세 가지 정도를 들 수 있다.

그런데 이 가운데서 첫번째만은 4기능 이외에 응용언어학(40분), 문명과 문화(30분), 전문준비(Professional preparation)(40분)도 포함되어 있지만 목적에 따라서 상기한 일곱 가지 전부의 테스트, 4기능 만의 테스트, 응용언어학 문명과 문화 전문준비의 테스트 등으로 나누어져 이용할 수도 있다. 여기서는 (2)와 (3)의 테스트 내용을 간략하게 정리하여 보기로 한다.

(2) MLA Test

- ① Listening (25분) comprehension (pictorial & written cues)
- ② Speaking (10분)
 - a) Mimicry
 - b) Reading aloud a passage
 - c) Picture questions
- ③ Reading (35분)
 - a) Reading
 - b) Vocabulary
 - c) Grammar
- ④ Writing (35분)
 - a) Item of fill-in-type (articles, pronouns, etc.)
 - b) To receive sentences
 - c) To write a short paragraph or dialogue test

(3) Pimsleur Test

- ① Listening (20분)
 - a) Perception (sounds, stress, intonation)
 - b) Comprehension
- ② Speaking (20분)
 - a) Vocabulary
 - b) Pronunciation
 - c) Fluency
- ③ Reading (35분)
 - a) To grasp the literal meaning of a passage
 - b) To draw inferences on the basis of information presented in the passage
- ④ Writing (35분)
 - a) To complete a paragraph or dialogue
 - b) To supply the correct form of a given verb

3) *MLA Foreign Language Proficiency Tests for Teachers and Advanced Students*, (Princeton, New Jersey: Educational Testing Service, 1962)

4) *MLA Cooperative Foreign Language Tests*, (Princeton, New Jersey: Educational Testing Service, 1964)

5) *Pimsleur Modern Foreign-Language Proficiency Test*, (New York: Harcourt, Brace & World, 1967)

- c) Sentence transformation
- d) A free-writing exercise

이것은 영어학력 테스트 내용의 하나의 패턴에 지나지 않지만 어떠한 영역의 어떠한 테스트 활동을 어떠한 비율로 테스트해야 할 것인가를 아르켜 주고 있다고 할 것이다.

또한 David. P. Harris의 연구결과에 의하면 테스트 항목수와 시간과의 비율을 다음과 같이 보여주고 있다.⁶⁾

Part	Number of Items	Minutes (items/m)
1. Listening Multiple-choice	40	20 (2)
2. Structure Short-answer	70	45 (1.5)
3. Vocabulary Short-answer	40	45 (1)
(Total)	150	110
Administration		10
		120

(예를 들면 short-answer 형식의 structure를 테스트하는 경우 우수한 학생은 평균 1분에 1½ 항목의 대답을 할 수 있다. 따라서 45분 동안에 70 항목에 대한 해답을 할 수 있게된다. Administration은 테스트 전후에 테스트의 배포, 회수, 일반적 주의 등에 필요한 시간을 의미한다)

4- skills을 테스트하는데 있어서는 이들 네 가지 기능 가운데 각 기능의 내용 이를테면 hearing ability는 어떠한 인자구조를 갖고 있는가를 보다 구체적으로 분명히 할 필요가 있다. 그러나 현단계로서는 각각의 기능에 관한 하위 테스트를 다음과 같이 생각할 수가 있다.

(1) Auditory test

Hearing ability는 discrimination과 comprehension으로 나누어진다. 전자는 청각력(Aural perception)이며 후자는 청해력(Aural comprehension)이다. G.E. Perren의 연구결과에 의하면 청각력은 학습이 진전되어 감에 따라 spoken language의 total command에의 관련도는 낮아진다.⁷⁾ 따라서 청해력 테스트가 중심이 되는 셈이다.

청해력 테스트에 있어서 주요한 문제점은 다음 두 가지가 있다.

① 기억 부담(Memory burden)이 지나치게 커져서 기억력의 테스트가 되어 버려서는 안된다. 보다 짧고 보다 많은 회화적 발화(Utterance)를 문제항목으로 삼는 편이 좋다.

6) David, P Harris, *Testing English as a Second Language*, (New York:McGraw-Hill, 1969) p. 99.

7) George E. Perren, "Testing Spoken Language:some unsolved problems," *Language Testing Symposium*, ed. A. Davies, (New York: Oxford University Press, 1968) p. 85.

② 문자를 매개체로 하여 청해력을 측정하려는 기법은 청해력에 부가해서 독해력을 필요로 하므로 순수한 청해력 테스트로써는 바람직하지 않다.

(2) Oral production test

Speaking에 관한 working definition이 부족하다는 것과 실시용이성(administrability)이 낮으므로 구두발표력 테스트는 그리 실시되지 않고 있다. 그러나 이 문제에 대한 연구결과가 조금씩 나오고 있으며 현단계로서는 다음 몇 가지 점을 고려해서 학력 테스트의 한 분야에 부가해서 실시하여 grading에 가미하는 것이 바람직하다.

① 암창(Recitation)이나 음독(Reading aloud)은 speaking 활동의 일면에서의 아프로취로서는 가능하지만 전달능력(Ability to communicate)의 테스트나 평가는 되지 않는다. T.B. Kolivoda도 지적하고 있는 바와 같이 학습자 자신의 아이디어를 표현하는 speaking behavior가 중심이 되지 않으면 안된다.⁸⁾

② 발음을 페이퍼 테스트로 측정하는 방법은 시간적 절약, 집단실시가 가능하다는 잇점은 있으나 이것을 speaking test로서 대용할 수는 없다. Speaking은 다른 기능 이상으로 social-productive skill이기 때문이다.

③ 채점에 있어서는 scorer reliability를 높이기 위하여 해답례와 허용범위를 정한 scoring key를 만들어 1시(혹은 1인) 1항목 주의로 채점하지 않으면 안된다.

(3) Reading comprehension test

Passage reading에 있어서는 100~250 단어의 passage에 4~6의 문제 항목수가 바람직하다⁹⁾. 그러나 장문독해로서 흔히 실시되는 500 단어 정도의 passage에 15자 정도의 질문 항목수 보다는 250 단어 정도의 passage 두 개를 선정하여 거기에 각각 7개 정도의 질문 항목을 설정하는 편이 낫다. Translation technique는 잘 쓰이고는 있지만 어디까지나 translation의 능력을 측정하기 위한 것 이외에는 사용해서는 안될 것이다.

(4) Writing proficiency test

여기서는 주로 objective type나 subjective type나가 문제가 되지만 이 문제는 나중에 언급하기로 한다. Dictation에 관해서는 Lado (1961)나 H.S. Otta (1968)에서 보는 바와 같이 테스트로서는 한계가 있다. Dictation은 오히려 학습을 위한 테스트이며 하나의 teaching device로 간주하는 편이 옳은 것이다. 설사 이것이 학력 테스트로써 사용된다 치더라도 reading speed나 grouping 같은 데 세심한 주의가 필요하다.

8) T. B. Kalivoda, "Oral Testing in Secondary School, *MLJ* LIV (May, 1970) 55.

9) David .p. Harris, *op.cit.*, p. 64.

III

영어의 학력은 지극히 많은 인자(Factor)에 의한 복잡한 중층구조를 이루고 있다고 생각할 수 있다. 따라서 테스트로서 측정하려고 하는 내용과 동시에 측정하는 방법에 따라 서로 상이한 결과가 발생한다. 우리들은 매우 간단하게 학생을 「잘 한다」, 「못한다」로 구별해 놓고 있다. 하나의 지도법 하나의 평가법에 의해서 말이다. 다른 지도법이나 평가법을 쓰면 「못한다」는 쪽에 설 학생이 「잘한다」는 쪽의 학생으로 뒤바뀔 가능성은 충분히 있는 것이다. 이를 보면 필자가 조사한 바에 의하면 중학교 3학년의 K라는 학생이 학교에서 정기적으로 실시하는 폐이퍼 테스트로는 100점 만점에서 불과 7점밖에 못받아 클래스에서 제일 뒤떨어진 학생이었는데 auditory test(청각력 테스트와 청해력 테스트)를 해 보니 67점이나 딸 수 있었다. 그러므로 「못한다」는 딱지가 붙어 교사의 기억에서 잊혀져 버린 학생에게는 음성면에서 자신과 의욕을 관기시켜 주어야 할 것이다.

테스트의 방법, 예를 들면 receptive type와 productive type, 문제수와 배점 등에 따라 측정된 결과가 달라지기 마련이다. 영어 학력 테스트의 방법의 차이는 동일한 학력 내용을 상이한 방법으로 측정하는 것이라기 보다는 오히려 여러가지 다양한 종류의 학력 정보를 제공해주는 것이기 때문에 변화있는 방법으로 테스트하는 것이 중요하다.

테스트 방법에 있어서 객관적 테스트나 주관적 테스트가 자주 논의가 되지만 잘못 이해되는 경우가 많다. 객관적 테스트는 그 테스트의 결과가 객관적이라고는 말할 수 없다. 채점자의 신뢰성(Scorer reliability)이 객관적 테스트가 높다고 할 수 있다. 그러나 객관적 테스트, 주관적 테스트나 그 자체로 볼 때는 어느 쪽이 신뢰성과 타당성이 높은가는 일률적으로 말할 수 없다. E.G. Pilliner도 지적하고 있듯이 어떠한 테스트라도 주관적으로 작성되어 해답이 된다. 주로 채점자가 주관적이 되느냐 객관적이 되느냐가 두 테스트를 구별하는 차이점이다.¹⁰⁾ 필자는 또 하나의 차이점으로서 문제항목의 형식에 있어서 발표적(Productive)인가 수용적(Receptive)인가 즉 문제항목이 요구하는 productivity의 비율의 차이를 들고 싶다. 보다 수용적인 문제 항목으로 구성되는 객관적 테스트는 필연적으로 학력을 수용적 피동적으로 측정하기 쉽게 되고 보다 발표적인 주관적 테스트는 발표적, 능동적으로 측정하기 쉽게 된다. 여기서 당연히 문제되는 것은 수용력은 어느 정도 발표력을 예측할 수 있는가 다시 말하면 수용력과 발표력과의 관계이다. 이 문제는 아직 결론은 나와있지 않지만 J. B. Carroll에 의하면 모국어의 경우보다도 외국어의 경우에 있어서 양자의 상관관계는 낮으므로 수용력과 발표력은 별개로 테스트하지 않

10) E.G. Pilliner, "Subjective and Objective Testings," *Language Testing symposium*, (New York: Oxford University Press, 1968), p. 21.

으면 안된다.¹¹⁾

P. Robinson은 주관적 테스트와 객관적 테스트로 나누지 않고 participation이라는 용어를 써서 분류하고 있다. 독해력에 관한 한가지 예를 들면 다음과 같다.¹²⁾

(1) Zero participation:

The students reads a text. He then reads a number of questions and selects what he believes to be the most appropriate of a number of possible answers.

(2) Limited participation:

The student reads a text. He then reads a number of questions and is asked to limit his answer to a certain number of words.

(3) Extended participation:

The student reads a text and is asked to make a summary.

(4) Complete participation:

The student reads a text and is asked to comment on the text.

필자가 중학교 2, 3년생을 대상으로 writing test의 문제형식에 관해서 조사한 바에 의하면 spelling, vocabulary, grammatical items의 각각에 대해서 수용형식(Receptive type)과 발표형식(Productive type)간의 상관관계는 .50~.60정도였다. Godshalk에 의하면 60분간의 writing에 관한 객관적 테스트와 작문 테스트와의 상관관계는 대체로 .70이었다.¹³⁾

이것들을 종합하여 판단해 보면 다음과 같은 사실을 말할 수 있을 것이다.

(1) 주관적 테스트, 객관적 테스트 등 여러가지 테스트 형식에 의해서 테스트하는 것이 바람직하다. 왜냐하면 주관적 테스트나 객관적 테스트 등으로 측정하고 있는 것이 동일하다고는 할 수 없기 때문이다.

(2) 객관적 테스트는 그 하위 테스트형식(수용형식 부분발표형식 등)의 종류와 문제항목수를 늘림으로써 학력의 바람직한 테스트에 접근할 수 있다.

테스트는 다음 두 가지 용도에 따라 사용됨이 바람직하다.

(1) 평정(評定)을 위한 테스트...주로 진단적 측정 기능

(2) 학습을 위한 테스트...주로 연습적 기능

테스트 그 자체에 학습(연습)기능이 있는 셈이지만 평가를 주목적으로 하느냐 평가는 생각하지 않고 언어활동의 하나로서 테스트 형식을 사용할 것인가를 일단 구별하는 편이 좋다. 이를테면 free composition이나 dictation 등의 hybrid tests는 후자의 학습을 위한 테스트로써 보다 큰 효과를 발휘할 수 있을 것이다.

11) J.B. Carroll, "The Psychology of Language Testings," *Language Testing Symposium*, (New York: Oxford University Press, 1968), pp. 52-53.

12) Peter Robinson, "Oral Expression Tests: *ELT*, XXV, 2, (February, 1971), 98.

13) F.I. Godshalk, ed al., *Studies of the Question Types in the CEES English Composition Test as Predictors of an Essay Criterion*, (Princeton, New Jersey: Educational Testing Service, 1965) p. 22.

IV

테스트 작성상의 여러가지 문제 가운데서 첫째로 문맥설정 (Contextualization)을 들 수 있다. 일반적으로 문맥을 길게 잡으면 잡을수록 테스트의 신뢰성과 타당성은 높아지지만 진단성은 낮아진다. 하나의 관점에 짧은 하나의 문제항목(1시 1항목주의)의 경우에는 그 문제항목의 正否가 그대로 진단에 이어지지만 문맥이 길어지면 어디서 학생이 실수를 했는지 어디에 학생의 장단점이 있는지 진단하기가 어려워지게 된다. 그러나 이 원칙은 발음, 어휘, 문법구조 등의 언어 요소를 테스트하는 경우에 보다 크게 들어 맞는다. 예를 들면 테스트에서 선도부 (Lead)의 문맥은 그 어휘에 대해서 최소한의 정의를 내릴수 있는 것이어야 한다. Lado가 들고 있는 예로 말해 보면 다음과 같다.¹⁴⁾

선도부 : He wears glasses.

선택 : Something 1. for the rain
 2. to see with
 3. for ice
 4. to drink with

이 선도부 (Lead)의 문맥은 응답하기에 충분한 문맥이지만 He wears glasses to read인 경우엔 key word의 glasses는 몰라도 올바른 해답을 할 수 있기 때문에 길이가 지나친 문맥이 된다. 한편, 선도부를 glasses 만으로 한다면 2. 나 4.가 모두 정답이 되어 지나치게 짧은 문맥이 된다.

종합기능(Integrated skills)의 경우 이룰때면 hearing에 있어서는 기억부담(Memory burden)의 면에서도 짧은 회화적 발화가 가장 좋고 reading의 경우에는 보다 서술적 발화가 좋다.

두번째 문제는 문제항목의 난이도(Item difficulty)이다.

테스트 전체의 난이도(정답율)는 50~60%정도가 되는 테스트가 바람직하다. 지나치게 쉽거나 어려우면 그 테스트로 실력을 판별할 수 없기 때문이다. 각 문제 항목에 관해서는 이룰때면 다선지문제항목(multiple-choice item)의 경우에는 너무 쉬운 문제항목(92% 이상의 정답율)과 너무 어려운 문제항목(30%이하의 정답율)은 타당성을 해치기 때문에 채택해서는 안된다.¹⁵⁾

세번째 문제는 무엇이 신뢰성을 해치는 요인이 되는가 하는 것이다.

테스트는 타당성과 함께 높은 신뢰성(안정성)을 겸비하고 있지 않으면 안된다. 신뢰성은 테스트의 신뢰성과 채점자의 신뢰성(Scorer reliability)으로 나누어 진다. 전자인 테스트의 신뢰

14) Robert Lado, op. cit., p. 192.

15) David, P. Harris, op. cit., p. 105.

성은 문제항목수와 각 문제항목에 관한 선택수를 늘림으로 해서 높아진다. 신뢰성을 상실하는 요인은 이것 이외에 A. Paterno가 지적한 속에 현저하게 나타나 있다. 결국 다음과 같은 요인에 의해서 테스트의 신뢰성이 낮아지게 된다. 특히 ③과 ④는 가장 신뢰성이 낮다.¹⁶⁾

- ① Ambiguous questions,
- ② Questions set in such a way as to form lucky guesses or to allow the pupil to proceed to the correct result by a process of elimination,
- ③ Questions which cover only a small sample of the skills or knowledge involved,
- ④ Questions which invite responses of different kinds and place too much stress on the qualitative judgement of the examiner.

네번째 문제는 잘못된 형(Incorrect form)의 출제이다.

교실에서의 테스트(Classroom testing)는 학습과정을 추진해나가야 하는 것이기 때문에 문제항목에 여하한 incorrect form도 도입해서는 안된다. 이러한 의미에서나 또한 짐작(Guessing)이란 면에서나 이자택일의 테스트형과 다를 바 없는 true-false test는 바람직하지 않다. 또한 평상시의 지도에 있어서도 테스트로는 불가능했던 말하자면 attainments 보다도 deficiencies에 눈을 돌리지 않도록 하는 것이 중요하다.

다섯번째는 본 테스트에 앞서 예비 테스트가 바람직한 경우이다.

타당성과 신뢰성은 일반적으로는 말할 수 없으므로 각 학력의 영역 그 테스트의 목적, 학습자의 연령과 학습경력 등에 의해서 결정되지 않으면 안된다. 표준 학력 테스트의 경우에는 사전에 그 테스트를 타당화(Validation)하고 표준화하기 위해서 예비 테스트를 한다. 교사작성 테스트(teacher-made test)의 경우에도 될 수 있는 한 동일한 예비 테스트의 단계가 설정되어야 한다. 그러기 위해서는 평상시의 테스트(Small test)를 본 테스트의 예비 테스트로써 타당화시켜 놓는 것이 바람직하다.

여러가지 형식의 객관적 테스트를 작성하는 경우는 최종적으로 필요한 문제항목수보다 30~40%정도 여유있게 준비해 두어야 한다. 난이도, 판별력, 신뢰성계수(Reliability coefficient) 등의 점에서 적절하지 않는 문제항목을 버리기 위해서이다. 학기말이나 학년말의 보다 정식 테스트에서는 그때까지 실시한 테스트의 결과에 입각한 문제항목 분석(Item analysis)에 의해서 테스트를 작성해야 한다.¹⁷⁾

영어 학력·테스트의 효력을 규정하는 것은 통계학적 조작도 중요하겠지만 어떠한 테스트나 문제항목이 높은 타당성과 신뢰성을 지니고 있는가에 대한 종합적 견해이다. 바람직한 테스

16) Adelaida Paterno, "Foreign Language Testing," *Teaching English as a Second Language*, ed. H.B. Allen, (New York: McGraw-Hill, 1965) p. 379.

17) Robert Lado, op. cit., pp. 342-360.

트란 무엇인가에 대한 시점과 그것을 살리고 테스트에 반영하여 개인 개인에 의한 타당한 평가를 하는 노력이 요구되는 것이다.

참 고 문 헌

1. Carroll, John B., "The Psychology of Language Testing," *Language Testing Symposium*, ed. by A. Davies (London: Oxford University Press, 1968)
2. Harris, David P., *Testing English as a Second Language*, (New York: McGraw-Hill, 1969)
3. Godshalk, F.I., et al., *Studies of the Question Types in the CEGB English Composition Test as Predictors of Essay Criterion*, (Princeton, New Jersey: Educational Testing Service, 1965)
4. Kalivoda, T.B., "Oral Testing in Secondary Schools," (*MLJ*, LIV, 1970)
5. Lado, Robert, *Language Testing*, London: Longmans, 1961)
6. _____, "Testing Control of the Structure of a Foreign Language," (*LL*, 1951)
7. Paterno, Adelaida, "Foreign Language Testing," *Teaching English as a Second Language*, ed. by H.B. Allen, (New York: McGraw-Hill, 1965)
8. Perren, George, "Testing Spoken Language: some unsolved problems", *Language Testing Symposium*, (London: Oxford University Press, 1968)
9. Pilliner, A.E.G., "Subjective and Objective Testing," *Language Testing Symposiums*, (London: Oxford University Press, 1968)
10. Pimsleur, Paul, "Testing Foreign Language Learning," *Trends in Language Teaching*, ed. by A. Valdman, (New York: McGraw-Hill, 1966)

Language Testing

—Its Content and Method—

Kay, Myung-il

Summary

1. The function of testing is not only to evaluate the result of learning but also to give direction to the next stage.

2. Language testing is one of the most important factors in language learning. The essence of testing must be made clear by linguistic science, and its technique is to be established by scientific description of language.

3. The achievement test in English is to be made according to the following process: (1) establishment of the level of intended scholastic ability (2) determination of the content and method of test (3) validation (4) organization of test.

4. The items in a language test are a sample which comes from population called linguistic ability. The sample must represent the population. The validity of test requires the research using the method (1) by factor analysis (2) by structure analysis of language and language activity (3) by problem analysis of language learning (4) by theoretical models of linguistic ability.

5. The result of a test may vary according to the method of testing, for example, whether it is a receptive type or productive type, the number of items, and score distribution. By adopting different kinds tests, various kinds of information about scholastic ability can be obtained.

6. In organizing test, the following must be considered:

- 1) Contextualization (Broad context brings high validity and reliability but low diagnosis.)
- 2) Item difficulty (The desirable percentage of correct answer is 50–60%.)
- 3) Reliability (Test becomes reliable by increasing number of items and their choices.)
- 4) Undesirable introduction of incorrect forms (true-false test is not desirable.)