

## 外國語 能熟度의 評價方法 研究

文

龍

(英語 教育科)

### I. 序 論

이 研究는 영어를 중심으로 한 외국어의 평가의 방법에 관한 여러 문제점을 살피는 것을 목적으로 한다.

종래 특히 우리 나라에서는 외국어의 평가는 바로 「독해력」의 평가를 의미했고, 「독해력」의 평가는 주로 <번역>에 의지를 해왔었다. 그러다 얼마전 부터는 multiple choice型이 독해력 평가의 형식으로 널리 이용되고 있다.

그런데 외국어를 잘 한다는 것은 마땅히 듣기·말하기·읽기·쓰기의 네 技能을 잘 한다는 뜻이어서, 외국어의 평가가 독해력의 평가에만 머물 수는 없다. 그렇지 않아도 글을 통해서만 외국어의 효용을 찾으려던 시대가 지난 오늘날, 「듣기」, 「말하기」는 「읽기」 못지 않게 그 중요성을 인정받고 있다. 외국어의 듣기나 말하기에 능숙한 人力의 수요도 급격히 늘고 있고, 이에 따라 「듣기」나 「말하기」의 능숙도의 평가가 현실적으로 필요해지고 있다. 中·高校를 중심으로 한 외국어(영어) 교육의 현장에 눈을 돌린다면, 오늘날 만큼 「듣기」나 「말하기」가 강조되고 있는 때도 없다. 그런데 외국어의 활용 능력을 높이기 위한 이런 노력은 소기의 목적을 달성하기 위해서는 마땅히 「듣기」나 「말하기」의 능숙도의 평가를 수반하여야 한다. 다시 말해서 우리 나라에서도 외국어의 능숙도의 평가에서 「듣기」나 「말하기」의 평가는 절실하게 필요한 것이다. 그러나 이와 같은 필요성과는 별도로, 평가를 주관할 시험관이나 교사의 입장에서는 방법상 「듣기」나 「말하기」의 평가는 아직은 생소하기만한 것이 사실이다.

이런 사실을 감안해서 이 研究는 특히 「듣기」와 「말하기」의 능력의 평가에 관한 방법과 문제점을 살피는 데 중점을 두었으며, 그밖에 discrete point test와 integrative test에 관련된 논의 및 cloze test의 문제점 등을 통해서 외국어 능숙도의 평가에 관한 해외 학계의 최근 경향을 아울러 살폈다.

## Ⅱ. 項目別 評價(discrete point test)와 綜合 評價 (integrative test)에 關連된 문제

똑 같이 영어를 평가하는 문제라지만 다음 (A)와 (B)는 <무엇을 어떻게 평가하는가>란 문제에 관한 基本方向이 다르다.

## (A)

1. 밑줄 친 부분의 발음이 나머지 셋과 다른 것은?

- |                     |                   |
|---------------------|-------------------|
| (a) <u>ch</u> emist | (b) <u>ch</u> air |
| (c) <u>ch</u> oice  | (d) <u>ch</u> ild |

2. 다음 빈 칸에 알맞은 것은?

A: I'm leaving now.

B: You've turned off the lights, \_\_\_\_\_ ?

- |                |                 |
|----------------|-----------------|
| (a) didn't you | (b) haven't you |
| (c) did you    | (d) have you    |

3. 다음 밑줄 친 어구와 뜻이 같은 것은?

John was astounded to hear her answer.

- |                       |                      |
|-----------------------|----------------------|
| (a) greatly amused    | (b) greatly relieved |
| (c) greatly surprised | (d) greatly angered  |

4. 다음에 나오는 두 낱말의 발음을 자세히 듣고 두 낱말의 발음이 같으면 답안지의 해당란에 ○표를, 발음이 다르면 ×표를 하시오.

- |          |       |   |                      |
|----------|-------|---|----------------------|
| (1) they | day   | } | .....(녹음을 통해서 듣는 소리) |
| (2) sink | think |   |                      |
| (3) wood | wood  |   |                      |
| .        | .     |   |                      |

## (B)

1. 받아쓰기 (dictation)
2. 글 짓 기 (essay writing)
3. 面 接 (oral interview)
4. cloze test

외국어 평가에 관한 요저음의 文獻에서, 출제의 기본 방향이 다른 (A)와 (B)는 각각 discrete point test(項目別 評價)와 integrative test(綜合評價)로 불리우고 있다.

Discrete point test는 그 명칭이 나타내 보이듯이 외국어의 평가 문제는 문제 하나가 特定한 항목 하나만을 테스트할 수 있게 작성되어야 한다는 原則을 전제로 한다. 위에 예시한 (A)를 引用한다면, 1. 은 ch란 철자의 발음을, 2. 는 附加疑問文을, 3. 은 特定한 단어의 의미를, 그리고 4. 는 minimal pair를 이용한 音素의 差別을 테스트하는 discrete point test이다.

Discrete point test의 主唱者들은 하나의 문제가 하나의 항목만을 테스트하여야만 피시험자가 무엇을 알고 무엇을 모르는가를 精確하게 가려낼 수 있고, 외국어 수업의 現場에서는 어떠한 사항이 더 많은 학습을 필요로 하는가를 診斷할 수 있다고 주장한다. 이와 같은 주장의 밑바닥에는 외국어의 실력이라 결국 작은 항목 하나하나가 쌓이고 모여서 이루어진다는 推論이 깔려 있다.

다음은 이와 같은 推論을 圖表로 바꾼 예의 하나이다. (Harris: 1969 : 11)

(圖表 1)

Components	Language Skills			
	Listening	Speaking	Reading	Writing
Phonology/Orthography	a	b	c	d
Structure	e	f	g	h
Vocabulary	i	j	k	l
Rate and general fluency	m	n	o	p

이 16가지의 4각형은 다시 개별 항목으로 세분된다. 앞서 예를 든 (A)의 1. 은 4각형 (c)의 구성 요소의 하나를 이루는 항목일 것이고, 2. 는 (f)의, 3. 은 (i)나 (k)의, 그리고 4. 는 (a)의 구성 요소가 될 것이다. (a~p 표시는 필자)

이런 推論을 근거로 한다면 우리의 외국어 실력은 우리가 이 16가지의 4각형을 다시 細分한 구성 요소를 얼마나 master 했는가에 따라서 判定되는 것이다.

Integrative test는 어떠한가?

Discrete point test가 分析的이라면 integrative test는 글자 그대로 綜合的이다. discrete point test가 외국어의 능력이란 그것을 구성하는 작은 有限한 요소로 구분할 수 있다는 입장인데 반해서, integrative test는 의사 전달을 목적으로 언어를 사용하는 상황에서 우리가 드러내는 언어 능력(communicative competence)은 편리하게 구분할 수 있는 구성 요소의 總和 이상의 어떤 것이라고 주장한다.

간략한 예를 들어보자.

(1) I \_\_\_\_\_ he is honest.

우리가 문장 (1)을 듣고 빈 칸을 차지하는 소리를 가령 sink로 듣지 않고 think로 이해했을 때, 그것은 반드시 우리가 /s/와 /θ/란 두 音素의 식별을 할 수 있었기 때문만은 아니다. 상대방이 설사 sink라고 발음을 했을 망정 우리는(상대방의 발음을 어색하게 여기거나 자신의 귀를 의심해보기도 하겠지만) sink를 think로 고쳐서 들을 것이기 때문이다. 그렇게 생각해 본다면 (1)을 듣고 (그 뜻을 이해하게 하고) sink와 think를 식별하게 한 「듣기」의 能力은 <적어도> 다음과 같은 네 가지 항목을 익힌 能力이 <서로 얽혀서> 作用한 능력이다.

(a) /s/와 /θ/의 식별

(b) sink와 think의 의미 차이

(c) I \_\_\_\_\_ he is honest.란 문장 구조

(d) 문장 (1)과 이 문장이 쓰인 言語外的인 狀況의 관련성

그런데 이 네 가지 능력이 <서로 얽혀서> 작용한 능력은 이 네 가지 능력의 단순한 총화는 아니다. 말을 바꾸어, 네 가지 능력으로 구분할 수 있는 능력은 아닌 것이다. 그런데다 한마디 덧붙인다면 문장 (1)을 이해하는 데 필요한 능력은 비단 위에서 언급한 네 가지 능력에 한하는 것이 아니다.

그러기 때문에 당초 구분하기도 어려운데다 구분할 수 있는 구성 요소의 총화 이상의 어떤 것인 언어 능력이란, 종합적으로 평가되어야지 구성 요소별로 평가될 수는 없는 것이다.

이런 관점에서 discrete point test를 자처한 앞서 예들 든 (A)의 2.를 吟味해 보자. (A)의 2.는 附加疑問文이란 항목에 대한 피시험자의 지식을 테스트하는 discrete point test이다. 그러나 과연 이 문제가 discrete point test일까? 이 문제의 正答 여부는 결코 부가의 문문에 대한 피시험자의 지식여부에만 달려 있는 것은 아니다. 이 빈 칸의 앞에 나오는 부분인 "A: I'm leaving now. B: You've turned off the lights,..."와 관련된 복합적인 지식 없이 正答을 맞출 수는 없는 것이다. 이 빈 칸 앞에 나온 부분을 잘 이해하지 못했기 때문에 誤答이 나올 수도 있다. 그렇게 본다면 discrete point test를 자처하고 있는 이 문제는 결코 discrete point test가 될 수 없다.

이와 같이 문장 (1)이나 (A)의 2. 처럼 짙막한 문장 하나를 이해하거나 生成하는 데 있어서도 복합적인 능력이 작용하고 있다면 discrete point test는 합리적인 평가의 요건의 하나인 validity를 실제로 부정하고 있는 셈이다.

比喩的인 例도 들어보자.

외국어를 배우고 가르치는 목적은 한마디로 communication이다. 이 communication은 흔히 「水泳」에 비유되어 왔다. 그래서 외국어 교수법에서는 문법 위주의 교수법은 수영을 물속에서 가르치지 않고 땅바닥에서 가르치는 교수법으로 비난을 받아오기도 한다. 수영이 물속에서 가르쳐져야 하는 것이 분명한 것처럼, 수영을 잘하는 사람이란 물속에서 더 빨리

그리고 더 멀리 가는 사람을 말하는 것도 분명하다. 다시 말해서 수영의 평가는 물속에서 피시험자가 얼마나 빨리, 그리고 멀리 갈 수 있는나를 기준으로 삼는 것이다. 그리고 이것이 수영을 communication으로 바꾸어 놓았을 때 integrative test가 志向하는 평가의 방향이다. 반면, 물과 팔다리의 운동을 분리시키고, 다시 팔다리의 운동을 세분해서 수영의 능력을 평가하려는 것이 discrete point test이다. Integrative test의 대표적인 형식으로는 앞서 지적한 것처럼 面接(oral interview)이나 글짓기(essay writing) 등이 있고, 「번역」도 integrative test에 포함시킬 수가 있다.

그런데 이런 integrative test의 난점으로는 역시 채점 방식의 주관성을 들 수 있다. (그리고 이와는 대조적으로 채점이 용이하고 적어도 겉으로는 客觀性이나 信賴性이 보장되어 있는 것처럼 보이는 것이 discrete point test이다.) Integrative test의 열렬한 主唱者로는 누구보다 John Oller를 손꼽을 수 있는데, Oller가 integrative test의 가장 바람직한 형식으로 dictation과 cloze test를 내세우고 있는 것은 이 두 형식이 채점 방법에 있어 크게 주관성을 배제할 수 있기 때문이기도 할 것이다.

하기는 Oller 처럼 discrete point test를 전적으로 배격하는 이도 있지만, 그 기본 방향이 판이한 이 두 형식은 과연 어떻게 대하는 것이 가장 바람직한 태도일까? 이 문제에 대해서 절충적 태도를 취하고 있는 이로는 Valette가 있다.

외국어의 평가 문제에 관한 定評 있는 概論書인 *Modern Language Testing*(1977)의 저자인 Valette는 第11章에서 Directions in Modern Language Testing을 다음과 같이 다루고 있다(pp.308~310). 즉, Valette는 당초 Bernard Spolsky가 발표했던 논문을 빌어 외국어 평가의 방향이 (1) 客觀性 또는 信賴性 등 평가의 기본 개념이 전연 없이, 교사의 주관적인 판단이 절대적이었던 단계에서, (2) 평가에 專門性이 도입되고, 정확하고 과학적인 평가 방식에의 志向이 (選多肢型) 객관식 문제 형식을 만들어내고, 그것이 構造言語學과 손을 맞잡아 discrete point test 優位論으로 발전한 단계를 거쳐, (3) 의사 전달(communication)의 가능성 여부에 초점을 맞추어 integrative test의 의의가 인정된 3단계로 발전되어 왔음을 示唆하고 있다.

원래 Valette의 이 *Modern Language Testing*은 1967년 版을 改訂한 것인데, 평가의 비교적 최근 경향을 말해주는 이런 論及은 1967년 版에는 물론 발견되지 않는다. 다만 1977년 版의 이 11章에는 1967년 版이 바로 위에서 언급한 첫 단계에 머물러 있는 어학 교수들에게 객관성을 강조한 두 번째 단계에 해당하는 평가의 기술을 더듬하게 하는 데 목적이 있었다는 언급이 있어, 1967년 版이 discrete point test 중심으로 집필되었음을 알 수 있다. 그렇다면 비교적 최근의 경향을 반영하고 있는 1977년 版은 전적으로 integrative test에 초점을 맞추고 있느냐 하면 반드시 그렇지는 않다. 여기에서 우리는 Valette의 절충적 태도를 엿볼 수 있는 것이다.

이 11장에서 Valette는 다음과 같은 뜻의 언급을 하고 있다.

「시험 문제를 작성하는 데 있어서 교사는 한편으로는 가능한 한 객관적 방법으로 수업 시간에 그가 가르친 어떤 특정한 사항을 학생들이 익혔는지를 확인하고 싶을 것이며 또 한편으로는 학생들이 의사 전달을 목표로 그동안 그들이 익힌 외국어를 얼마나 활용할 수 있는가를 알고 싶어할 것이다. 그러므로 저자는 평가 문제의 작성법을 다루는 데 있어 各章(의) 前半部는 discrete point test 중심으로 다루었고, 後半部는 integrative test에 초점을 두었다.……」

이 말은 「discrete point test는 achievement test에서 效用을 발견할 수 있고 integrative test는 proficiency test에서 有用하다」는 뜻으로 풀어 해석할 수 있을 것이다. 그렇지 않아도 평가의 목적이 단일하다면 몰라도, 목적이 단일하지 않다면, 목적에 따라 평가의 방법이나 형식은 마땅히 달라져야 한다. (모든 학과에 해당되는) 평가의 목적의 하나는 수업 현장에서 이루어지는 가르침과 배움에 대한 feedback의 제공이다. 교사는 그의 수업 방법의 효율성 내지 그가 가르친 특정한 학습 사항을 학생이 얼마나 익혔는가를 스스로 점검·확인할 필요가 있다. 이와 같은 목적으로 작성되는 것이 achievement test이다. 그러기 때문에 achievement test는 학생의 응답의 방향을 크게 통제할 필요가 있고, 이와 같은 통제를 필요로 하는 경우에는 integrative test보다 discrete point test가 알맞은 것이 분명하다.

Proficiency test는 특히 외국어에 해당이 된다. Real-life situation에서 피시험자가 드러내는 communicative competence의 측정이 바로 proficiency test라면, proficiency test에서는 integrative test가 효과적이라는 결론을 내리는 것은 어려운 일이 아닐 것이다.

다만 이와 같은 折衷論을 펴는 데 있어서는 약간의 但書가 필요하다. Discrete point test와 integrative test의 對立은 절대적이라기 보다 상대적인 것으로 간주하여야 할 것이다. Discrete point test를 자처한 (A)의 2.만 해도 이미 살펴보았던 것처럼 이 문제는 순수한 discrete point test는 되지 못한다. 말하자면 discrete point test에 가까울 따름이다. Integrative test에 대해서도 비슷한 말을 할 수가 있다. Integrative test의 대표적인 형식인 oral interview나 essay writing을 예로 들 때, 그것이 印象的 評價에 그치지 않기 위해서는, 종합 평가의 전제로 評價의 기준이 될만한 여러 항목들을 설정하는 일은 흔히 필요한 절차에 속한다. Discrete point test적인 요소가 개입하는 것이다.

이와 같이 목적에 따라 평가의 방식이 달라질 수 있고, 이 두 형식의 대립이 절대적인 것이 아님을 확인할 때, 우리는 한편을 부당히 배제하는 일 없이 이 두 평가 방법의 장점을 살린 效用을 발견·활용할 수 있을 것이다.

### Ⅲ. 「말하기」의 評價——그 方法과 基準

Clark(1979)는 말하기의 평가를 直接的(direct) 방식, 間接的(indirect) 방식 및 折衷的

(semidirect) 방식으로 분류하고 있다.

直接的 방식이란 시험관과 피시험자가 얼굴을 맞대고 面接式으로 진행되는 방법을 말하며, 間接的 방식이란 이른바 paper and pencil test를 말한다. 간접적 방식이 말하기의 평가인 데도 말을 입밖내 내지 않는 평가의 방식이라면, 입밖으로 소리를 내고 그것이 평가의 대상이 되기는 하지만 직접적 방식에서처럼 시험관과 얼굴을 맞대고 면접식으로 진행되는 것이 아닌 방식이 절충적 방식이다. 즉 절충적 방식에서는 피시험자는 시험관이 직접 묻는 말에 대답하는 대신 미리 녹음되었거나 인쇄되어 있는 지시 또는 물음에 대하여 답을 하고 그 답은 tape에 녹음되었다가 추후에 채점되기도 하고, 또는 피시험자가 하는 말을 옆에서 주로 듣기만하는 시험관에 의하여 채점되기도 한다.

이 세 가지 방식은 각기 長點을 가지고 있다. 우선 간접적 방식은 그 시험과 채점이 용이하다. 그러나 말하기의 평가를 말 아닌 다른 방법으로 평가하여야 하는 致命的인 약점을 지니고 있다.

前章에서도 언급하였지만 말하기의 평가의 방법은 목적에 따라 달라진다. Clark는 Proficiency test에는 直接的 방식이 효율적이며, 어떤 특정한 요소(한 예로 /r/의 발음, 조동사의 용법, hold의 과거형, 특정한 반응, ...)만을 평가하고자 할 때, (즉 평가가 achievement test적 성격을 지닐 때는) 피시험자의 입으로부터 그런 요소를 유도하게끔 작성된 절충식 방식이 효율적이라고 말하고 있다. 한편 直接的 방식은 적어도 한·두 명의 시험관이 배석하여야 하고 한 명의 피시험자를 상대로 최소한 10분은 시간을 소비하여야 하는 절차상의 制約이 있다. 그러나 절충식 방식은 言語實習室을 이용한다면 많은 피시험자를 동시에 평가할 수 있다.

面接에 의한 「말하기」의 평가는 자칫하면 시험관은 묻기만 하고 피시험자는 대답하기만 해서, 一例로 질문, 요청, 부탁, 권유...등에 관련된 피시험자의 능력의 평가는 간과되기 쉽다. 이와 같은 결점을 지양하기 위해서 直接的 방식은 시험관과 피시험자가 어떤 role play를 演出한다든가, 여러 가지 그림을 제시하고 그림을 통해서 범위가 넓은 화제를 유도해 내는 방법을 쓰기도 한다.

그런데 여러 가지 문제점과 제약이 있으면서도 直接的 방식을 빼놓고 말하기의 評價를 운운할 수 없음을 분명하다.

말하기의 評價에서 또 하나 문제가 되는 것은 對象(피시험자)과 평가의 목적에 따라 달라질 능숙도의 기준에 관련된 문제이다. 단적인 예로 中·高校에서 6년간 영어를 익힌 학생의 경우 <어떤 주제에 관해서 얼마 만큼 어떻게> 말을 했을 때, 그는 <어느 정도> 말을 잘 한다는 평가를 받을 수 있는 것일까? 이와 같은 물음에 대한 명확한 답이 없어 그의 「말하기」의 능숙도를 제대로 평가할 수 없는 것은 말할 나위가 없다. 그러나 현재 이와 같은 물음에 대한 명확한 답이 마련되어 있다고는 보기 어렵다. (그리고 이것은 「듣기」도 마

작가지이다.) 하기는 이 물음에 대한 답은 이론적으로는 文敎部가 制定한 중학교·고등학교의 영어 교육과정에서 찾아야 할 것이다. 그런데 막상 교육과정을 펴보면 중학교나 고등학교나 「生活周邊과 일반적인 화제에 관하여 말할 수 있는 능력을 기른다」는 내용이 표현을 달리하며 반복되고 있는데 불과하다. 따라서 중학교 1학년부터 말하기를 가르친다면 중학교 2학년에 가서는 좀 더 많은 어휘와 좀 더 복잡한 구문을 이용한 회화가 가능할 것이요, 高II는 高I보다 상대적으로 더욱 수준 높은 회화가 가능하다는 막연한 전제가 근거가 되고 있을 뿐, 도달 목표는 명시되어 있지 않다. 앞으로 중·고등학교에서 외국어의 「말하기」와 「듣기」는 더욱 강조되고, 따라서 「말하기」와 「듣기」의 평가도 교사의 직관이나 경험에만 의지할 수는 없는 전망이다. 그렇게 볼 때 중·고등학교에서의 「말하기」와 「듣기」의 합리적인 평가를 위해서는 앞서 언급한 영어 교육과정의 상세화가 무엇보다 時急하다 하겠다.

한편 요지음은 특히 官公置나 企業體에서 해외 파견 등 여러 가지 이유로 직원의 「말하기」를 평가하는 일이 늘고 있다. 그렇게 실시되는 「말하기」의 평가는 흔히 面接方式을 취하는데, 면접에서는 아직까지는 피시험자의 身上 문제를 중심으로 한 질문과 응답이 오가고 <몇 마디 물어보면 얼마나 말을 잘 하는지 알 수 있다>는 시험관 나름의 경험과 직관에서 울어 난 판단에서 평가가 이루어지고 있는 것이 通例이다. 이런 평가 방법은 영어 解得人口의 「말하기」의 능숙도가 평균적으로 낮고, 면접을 필요로 하는 피시험자의 수도 한정되어 있는 데 구실이 없는 것은 아니다. 그러나 영어 해득 인구의 「말하기」의 능숙도도 점차 높아지는 한편 각 개인에 따라 능숙도의 차이가 크게 벌어지는 추세를 감안한다면 문제는 달라진다. 뭇풀이가 되지만 첫머리에서 지적한 능숙도의 수준과 등급(및 채점 기준의 객관적 설정)이 결실한 문제로 提起되는 것이다.

이렇게 문제를 提起해볼 때 무엇보다 미국의 FSI(Foreign Service Institute)에서의 面接의 평가 방법은 참고가 될만하다.

다음은 FSI의 oral interview(面接官: 2명, 소요 시간: 30분)의 평가 항목, 능숙도의 등급 및 배정의 분류표이다. (圖表 2)

(圖表 2)

항 목		능숙도					
		1	2	3	4	5	6
발	음	0	1	2	2	3	4
문	법	6	12	18	24	30	36
어	휘	4	8	12	16	20	24
유	창	2	4	6	8	10	12
듣	고 이해 하기	4	8	12	15	19	23

「말하기」를 평가하는 데 있어 발음, 문법, 어휘, 유창도 및 듣고 이해하기란 다섯 가지의 항목을 설정하되 문법의 比重이 가장 크고 발음의 비중이 가장 낮음을 알 수 있다. 「말



하기]의 평가에 「듣고 이해하기」란 항목을 두고 있는 것도 주목할 만하다. 능숙도의 정도는 여섯으로 區分했는데 숫자가 높아질수록 능숙도도 높은 셈이다. 이 능숙도의 평가 기준은 다음과 같다.

### 發 音

- (1) 대부분의 경우 무슨 말을 하는지 이해하기 어렵다.
- (2) 잘못된 발음 탓으로 말의 뜻을 이해하기 어렵고 여러 번 반복시킨 다음에야 그 뜻을 이해할 수 있다.
- (3) 발음이 어색해서 귀를 기울여 들어야 한다. 잘못된 발음 때문에 그 뜻을 오해하기 쉽다.
- (4) 발음이 native speaker와는 判異하다. 가끔은 잘못된 발음을 하기도 한다. 그러나 무슨 뜻인지 알아 들을 수는 있다.
- (5) 그다지 잘못은 없으나 native speaker의 발음은 아니다.
- (6) native speaker와 다를 바가 없다.

### 文 法

- (1) 상투적으로 쓰는 어구를 제외하고는 말이 전연 문법에 맞지 않는다.
- (2) 主要文型이나 문법 사항을 모르고 있는 증거가 되는 잘못을 빈번히 저지른다. 따라서 의사 전달이 제대로 되지 않는다.
- (3) 主要文型이나 문법 사항을 완전히 익히지 않는 데서 생기는 잘못을 저지른다. 따라서 상대방이 말뜻을 오해하는 경우가 생긴다.
- (4) 主要文型 또는 문법 사항을 익히고 있지만 그밖의 문형 또는 문법 사항을 완전히 익히지 않아서 생기는 잘못을 저지른다. 그러나 상대방이 그 말뜻을 오해하는 정도는 아니다.
- (5) 간혹 잘못을 저지른다.
- (6) 面接이 끝날 때까지 한 가지 이상의 잘못을 저지르지 않는다.

### 語 彙

- (1) 지극히 간단한 회화에 필요한 어휘를 제대로 쓰지 못한다.
- (2) 사용하는 어휘가 개인의 身上이나 일상 생활에 필수적인 基本的 어휘에 국한되어 있다.
- (3) 어휘 선택이 부정확한 경우가 있고 어휘가 한정되어 있어 직업상 공통적인 회제나 사회 일반에 관한 회제를 제대로 다루지 못한다.
- (4) 특정한 관심사를 토의하는 데 적절한 전문적 어휘를 알고 있다. 가끔은 적절한 어휘를 몰라 말을 풀어 쓰기도 하고 다른 말로 돌려 말하기도 하나, 非專門的 회제도 다룰 수 있을 만큼 일반적인 어휘를 알고 있다.
- (5) 전문적인 어휘를 광범위하게 그리고 정확하게 쓴다. 복잡한 實社會的인 문제와 여러 생활에 대처할 수 있을 정도로 일반적인 어휘를 익히고 있다.
- (6) 교육을 받은 native speaker와 다를 없이 어휘 사용이 적절하다.

### 流暢度

- (1) 말이 자주 중도에서 끊기고 단편적이기 때문에 對話가 불가능하다.
- (2) 짧고 상투적인 어구를 제외하고는 말의 속도가 너무 느리고 고르지 못하다.
- (3) 더듬거리고 고르지 못한 경우가 자주 생긴다. 도중에서 말을 중단하기도 한다.
- (4) 가끔 더듬거린다. 말을 다시 바꾸어 표현하거나 적절한 어휘를 찾느라고 말이 도중에서 끊기기 때문에 고르지 못한 경우가 가끔 생긴다.

- (5) 말이 입밖으로 험들이지 않고 나오며 거침이 없다. 그러나 그 속도가 native speaker로 보기는 어렵다.
- (6) 모든 화제를 native speaker와 다름 없이 유창하게 다룰 수 있다.

### 듣고 이해하기

- (1) 간단한 대화도 거의 이해하지 못한다.
- (2) 일상적인 화제를 극히 간단한 構文을 사용하여 천천히 이야기하면 알아듣는다. 번번히 되풀이해서 들려주어야 하고 말을 다시 풀어서 설명해 주어야 하기도 한다.
- (3) 어느 정도 간단한 構文을 뚜렷한 발음으로 들려주면 알아 듣는다. 되풀이해서 들려주거나 말을 풀어서 설명하여야 하는 경우가 생긴다.
- (4) 교육을 받은 사람이 주고 받는 말을 正常的인 속도로 들려주면 알아 듣는다. 가끔 되풀이해서 들려주거나 말을 풀어서 설명해 주어야 한다.
- (5) 지극히 俗語的인 표현이나 드물게 사용되는 어휘 또는 지나치게 빠르거나 명확치 못한 말을 제외하고는 교육을 받은 사람이 정상적으로 주고 받는 말을 알아 듣는다.
- (6) Native speaker와 다름 없이 格式的인 표현이나 俗語的인 표현을 다 듣고 이해한다.

이와 같은 기준에 의해서 채점이 되면 그 결과에 따라 다음과 같은 等級(FSI level)이 매겨진다.

得點	等級(FSI level)
16-25	0
26-32	1
33-42	1+
43-52	2
53-62	2+
63-72	3
73-82	3+
83-92	4
93-99	4+

위의 等級은 의사 전달이 전연 불가능한 level 0와 native speaker와 조금도 다름이 없는 level 5를 일단 제외하고 볼 때 다음과 같은 수준과 對應한다.

- Level 1—여행을 하거나 社交上的 예의를 최소한 지키는 데 흔히 필요한 정도의 말을 할 수 있다.
- Level 2—일상적인 社會生活과 한정된 업무 수행에 필요한 정도의 말을 할 수 있다.
- Level 3—일상적이거나 사회적인, 또는 직업과 관련된 話題에 관한 格式的이거나 허물이 없는 대화에 능동적으로 참여할 수 있고 상당히 정확한 문장과 어휘를 구사할 줄 안다.
- Level 4—업무 수행상 아무 지장이 없을 정도로 모든 분야에 걸쳐 유창하고 정확한 말을 구사할 줄 안다.

그런데 得點에 따라 이런 등급이 정해진다고는 하지만 실제 面接의 과정에서는 이 등급과 앞에서 언급한 채점 기준은 相互補完的인 성격을 갖는다. 즉 피시험자를 대했을 때 시험관은 우선 level 1에 해당하는 정도의 대화를 시작하며 상대방의 반응에 따라 그 등급을

조금씩 올라가면서 채점이 이루어지는 것이다.

주지하는 바와 같이 이 FSI의 방식은 美國務省 傘下의 외교관 또는 이에 준하는 人力의 외국어 실력을 평가하기 위해서 작성한 것이다. 그러한 것을 이렇게 길게 引用한 것은 「말하기」의 목적이나 대상을 달리하는 경우도 이 기준은 하나의 point of reference로 삼을 만하고 이를 근거로 修正案을 마련해 볼 수 있기 때문이다.

실제로 외국의 여러 대학이나 연구 기관에서 채택하고 있거나 평가에 대한 概論書 등에 제시되어 있는 「말하기」 能力的 等級設定의 基準은 이 FSI의 방식을 참고로 하고 있고 그것을 적절히 수정 보완한 것들이다. 一例로 다음은 英國의 대학에서 수학하려는 외국 학생의 말하기의 能力을 6等級으로 나눈 기준으로 Dr. Frank Chaplen이 작성한 것이다.

Rating 6 : 극히 우수하다. 모든 면에서 native speaker와 다름이 없다.

- 5 : 우수하다. Native speaker로 간주되지는 않지만 본인의 의사는 충분히 전달할 수 있고 상대방의 말을 듣고 이해하는 데 지장이 없다.
- 4 : Native speaker와 대화를 나누는 데 지장은 없으나 가끔 문법·어휘·발음 등에서 잘못을 저지른다.
- 3 : Native speaker가 그의 말을 이해하기 어려운 경우가 생긴다. 자꾸 말을 반복하기도 하고 말을 시작했다가 다른 말로 다시 바꾸어 표현하기도 한다.
- 2 : 외국인들이 사용하는 영어에 익숙하지 않는 native speaker는 그의 말을 알아듣기가 힘들다. 日常적인 것에 대한 의사 전달은 가능하나 발음, 문법, 어휘 등에 걸쳐 많은 잘못을 저지른다.
- 1 : 의사 소통이 불가능하다.

FSI나 FSI를 참고한 위의 기준등을 참고했을 때 가령 우리나라 고등학교 3년 정도의 「말하기」의 능력은 어떠한 등급과 기준으로 평가할 수 있을 것인가? 이미 언급한 바와같이 이런 등급의 설정과 기준은 고등학교 영어 교육과정의 詳細化를 전제로 하여야 하지만, 잠정적으로는 다음과 같은 試案을 마련해 볼 수도 있을 것이다. 等級은 A,B,C,D,F로 한다.

等級 F : 극히 단편적인 말 몇 마디를 제외하고는 전연 의사 소통이 불가능하다.

等級 D : 간단한 인사를 포함하는 상투적인 말을 할 수 있고 들어서 이해한다. 간단한 일상 단어를 입밖에 낼 줄 안다.

等級 C : (1) 개인의 신상에 대해서 단편적인 이야기를 할 수 있다.

(2) 다음과 같은 내용에 대해서도 단편적인 말을 할 수는 있다. 그러나 입밖에 나오는 말이 단편적이어서 대화를 이어나가지 못한다.

- a. 간단한 인사
- b. 간단한 요구
- c. 긍정·부정·의문 표시

等級 B : (1) 다음과 같은 사항에 대하여 한정된 정보를 제공할 수 있다.

- a. 개인의 신상(본인과 가족)
- b. 학교 생활과 일상 생활
- c. 지역 사회



- (2) 다음과 같은 사항에 대해서도 본인의 의사를 표현할 수 있으나 그 표현에 밀도가 없다.
- 여러 가지 인사(만남과 헤어짐, 간단한 소개, 축하, 사의, 사과)
  - 위와 같은 인사를 받았을 때의 응답
  - 긍정·부정·의문
  - 동의·반대·의문과 확신·찬성·불찬성...

等級 A : (1) 다음과 같은 사항에 대해서 서로 대화를 이어갈 수 있을 정도로 말을 할 수 있다.

- 개인의 신상(본인과 가족)
  - 학교 생활과 일상 생활
  - 지역 사회
- (2) 다음과 같은 사항에 대해서 밀도 있게 본인의 의사를 표현할 수 있고 상대방에 따라 서로 조금은 style을 달리 할 줄 안다.
- 여러 가지 인사
  - 위와 같은 인사를 받았을 때의 응답
  - 긍정·부정·의문
  - 동의·반대·의문과 확신·찬성·불찬성...

여기에서 「말하기」의 평가를 또 한번 「水泳」의 평가에 비유해보자. 이미 言及하였듯이 한 개인이 수영을 얼마나 잘 하는가 하는 것은 원칙적으로 그가 <얼마나 빨리 그리고 얼마나 멀리 갈 수 있느냐>에 따라 평가된다. 고등학교 3학년 학생을 대상으로 한 「말하기」의 평가에서 그 등급을 다섯으로 나누고 각기 등급에 따라 기준을 설정해 본 위에서의 시도는 바로 <얼마나 빨리 그리고 얼마나 멀리 갈 수 있느냐>에 따라 수영의 능력을 평가한 원칙을 적용한 것이라고 말할 수 있다.

다만 「水泳」의 경우는 <얼마나 빨리 그리고 멀리 갈 수 있느냐>의 판단이 지극히 객관적으로 가능한데 반해서, 「말하기」의 평가는 위의 기준만으로는 자칫하면 印象의 평가로 흐를 수도 있다. 이와 같은 난점을 補完하기 위해서는 역시 앞서 例示한 圖表 2와 같은 항목과 배점등을 명시한 또 하나의 보조적 基準表가 필요해진다.

다음은 고등학교 3학년을 대상으로 한 「말하기」 평가에서의 항목·등급·배점의 기준 試案이다.

(圖表 3)

항 목		등 급			
		D	C	B	A
발	음	4	8	12	16
문	법	10	20	30	40
어	휘	7	14	21	28
유	창	4	8	12	16

위의 試案에서 FSI의 방식과 달리 「듣고 이해하기」를 빼 놓은 것은 「듣기」의 평가는 독립시킨다는 전제하에 것이고, 항목별 배점에서는 역시 문법에 가장 큰 비중을 두었다. 또한 등급 F는 제외시켰다.

다음은 네 가지 항목의 채점 기준이다.

**발 음**

- D. 간단한 단어 또는 상투적인 표현 이외는 무슨 뜻인지 이해하기 어렵거나, 말을 입밖내 내지 않는다.
- C. 대부분의 경우 영어의 음소를 한국어의 음소로 대응한다.
- B. 아직도 영어의 음소를 한국어의 음소로 대응하는 경향이 있으나 대체적으로는 무슨 뜻인지 이해할 수 있다.
- A. 한국어에 없는 음소도 비교적 잘 발음하고 stress도 상당히 정확하다.

**문 법**

- D. 기계적으로 암기하고 있는 상투적 문장 이외는 문법에 맞는 표현을 하지 못한다.
- C. 기본적 문법 사항에 관해서도 잘못을 흔히 저지른다.
- B. 주요 문형을 알고 있다.
- A. 주요 문형을 구사할 줄 알고, 기능어의 사용도 상당히 정확하다.

**어 휘**

- D. 극히 한정된 어휘만을 안다.
- C. 어휘가 한정되어 있다.
- B. 일상적인 어휘를 비교적 많이 알고 있다.
- A. 일상적인 어휘를 상당히 활용할 줄 안다.

**유창도**

- D. 한 단어나 두 단어를 입밖내 낼 수 있을 정도이다.
- C. 거의 기계적으로 암기하고 있는 문장을 제외하고는 일췌 더듬거린다.  
그 다음 말을 이을 때까지의 pause가 길고 도중에서 말을 끊어 버리기도 한다.
- B. 더듬거리고 속도도 느리지만 당초 의도했던 뜻은 결국 전달한다.
- A. 대부분의 경우 듣는 이를 답답하게 만들지 않는다.

IV. 「듣기」 評價의 問項作成에 관련된 문제

외국어의 「듣기」 문제의 대표적인 예로는 TOEFL의 일부를 이루는 listening comprehension test를 들 수 있다. 그런데 대상과 목적에 따라 여러 가지로 다를 수 있는 「듣기」평가의 문항을 작성하는 데 있어서 前提要件이 되는 것은——비단 「듣기」에 관한 것은 아니지만 특히 「듣기」의 경우——「듣기」란 언어 활동의 본질에 대한 이해이다.

「듣기」, 좀더 정확하게 말해서 「듣고 이해하는(listen and comprehend)」 능력은 단적으로 그 언어를 모르는 제三者가 들으면 그저 소음에 불과하는 「소리」를 통해서 「의미」를 알아듣는 능력을 말한다. 소리를 통해서 의미를 알아듣기 위해서는 첫째 소리의 기본 단위가 되는 音素(와 異音素)를 식별할 줄 알아야 한다. 그렇지 않으면 가령 sink와 think의 뜻을 구별못할 것이요, rising intonation과 falling intonation으로 발음했을 때의 He went home.의

차이를 알지 못할 것이다. 음소의 식별 여부를 테스트하는 여러 가지 방법은 테스트를 다룬 웬만한 概論書에 빠짐 없이 예가 나와 있으므로 굳이 여기에서는 언급하지 않는다. 그런데 이와 같은 음소의 식별은 실상 소리의 연속 속에서 의미를 찾아내는 능력에서 큰 비중을 차지하는 것은 아니다.

Sink와 think는 다르다. 그러나 실생활에서는 문맥이 cue가 되어서 가령 상대방이 I sink he is honest.란 말을 했더라도 sink를 think로 고쳐듣는 것이 「듣고 이해하는」 능력이기 때문이다.

구체적인 예를 또 하나 들어보자. 영어 교실의 「받아쓰기」 연습에서 학생이 [bækənfiθ]를 듣고 back an forth나 back on forth라고 받아썼다든가, [gud rizÁlt]란 녹음의 한 구절을 good without로 들었다면 그 학생은 어디에 결함이 있는 것일까? 그의 청각 기능에 고장이 생겼기 때문일까? 그렇지 않은 것이다. Native speaker라면 [gud rizÁlt]의 경우 실제로 녹음은 [gudwiðÁut]에 가까웠다 하더라도 good result로 들었을 것이요, [bæk]와 [fiθ] 중간의 발음이 아무리 모호했을 망정 back and forth라고 들을 것이다. 그것은 그가 good 다음에 result가 오면 하나의 완전한 의미가 표현 전달되지만 전치사가 뒤따르면 의미가 통하는 어구가 될수 없다는 것, back와 forth는 접속사로 연결되어야만 의미가 통한다는 것을 알고 있기 때문이다. 그러기 때문에 native speaker는—drugstore—place—have lunch.란 문장을 듣고 (-)부분의 발음이 모호한데도 A drugstore is the place to have lunch.란 소리를 빼놓지 않고 들었다고 「착각」하는 것이다.

우리 말의 예를 들어보자.

우리가 가령 「꼬치 피고…」라든가 「부를 떼고 나무를 심꼬」란 소리를 듣고 그 의미를 알아 들었을 때, 우리는 이와 같은 「物理的 音聲」을 그저 「物理的 音聲」으로 듣지는 않는다. 우리는 「꼬치 피고」를 「꽃이 피고」로 바꾸어서 듣고 「부를 떼고」를 「불을 떼고」로 들을 것이다. 그리고 「불을 떼고」의 「고」와 「나무를 심꼬」의 「꼬」는 물리적 음성은 분명히 다르지만 다르다는 의식을 갖지 않는 것이다.

영어의 경우 native speaker가 [lemigetəm]을 그저 물리적 음성으로 듣지 않고 Let/me/get/them으로 바꾸어 듣는 것도 똑같은 현상으로 파악할 수 있다. 「듣기」의 능력이란 결코 귀에 들려오는 물리적 음성을 물리적 음성으로 정확하게 듣는 능력은 아니다. 이렇게 생각해 볼 때 「듣기」의 평가에서 큰 비중을 차지하여야 하는 것은 「소리」보다는 「날말」이고 「날말」보다는 「문장」이다. 이런 사실을 다시 한번 다짐하면서, 다음에 「듣기」 문항 작성의 유의점을 예를 들어가며 따져 보기로 한다.

우선 다음 (1~2)와 같은 「듣기」 문항은 어떠한가?

1. 은 주어가 3인칭 단수인 경우 술어동사에 -s를 붙이는 문법 규칙을 알고 있는가를 테스트하는 문제인데, 문장 자체의 이해와 전연 관계가 없는 이런 문제는 listening test로는

결코 바람직하지 못하다. 한편 2.의 (1)은 When does Christmas come?이란 문장을 듣고 그 뜻을 이해했는가를 테스트하는 문제가 못된다. 이 문제는 <Christmas는 언제냐?>란 뜻은 피시험자가 이해하고 있는 것으로 전제가 되고 있는 셈으로, 다만 달의 뜻을 이해하고 있는가를 테스트할 수 있을 뿐이다. 그러나 달의 이름을 테스트하는 방법은 이밖에도 다른 방법이 있을 것이다.

1. 다음 녹음을 듣고 문법에 맞지 않는 것을 고르시오.
  - A. I get up early in the morning.
  - B. You get up early in the morning.
  - C. He get up early in the morning.
  - D. Tom gets up early in the morning.
2. 다음 녹음을 듣고 물음에 대한 올바른 답을 고르시오.
  - (1) When does Christmas come?
    - A. It comes in May.
    - B. It comes in June.
    - C. It comes in November.
    - D. It comes in December.
  - (2) Why are you going home?
    - A. Six o'clock.
    - B. Yes, I am.
    - C. To help my mother.
    - D. By Bus.

1.~2.에서 「듣기」 평가의 문항으로는 2.의 (2)가 내용상 가장 적절한데, 2.의 (1)을 (2)처럼 고쳐보면 다음과 같다.

3. When does Christmas come?
  - A. It comes in December.
  - B. December is the last month.
  - C. It is very cold in December.
  - D. We have snow.

(2~3)에서 A~D의 答肢는 녹음으로 들려주는 방법도 있고 글로 제시할 수도 있다. 이 (2~3)은 TOEFL 등에서도 흔히 대할 수 있는 형식인데, 答肢가 녹음으로 제시되는 경우는 答肢까지도 「듣기」의 평가의 일부를 구성하는 장점을 갖는 반면 그 답지가 길면 길수록 피시험자의 기억의 부담이 늘어나 「듣기」의 능력을 제대로 평가하는 데 부정적인 영향을 미치는 단점이 있다. 그렇게 본다면 녹음으로 제시되는 答肢는 가능한 짧은 것이 바람직하다. 또한 다음과 같은 형식도 고려할 수 있다. 「듣기」의 評價에서까지 四選肢의 형식이 으

례히 (기계적으로) 적용될 필요는 없을 것이다.

4. 다음에 나오는 질문을 잘 들으십시오. 그 다음 여러분은 다른 목소리로 이 물음에 대한 답을 듣게 됩니다. 이 답이 먼저 들은 물음에 알맞으면 주어진 칸에 ○표를 하고, 알맞지 않으면 ×표를 하시오.

(1) Q: How's the weather today?

A: 9:30 sharp.

(2) Q: May I see you this afternoon?

A: I hope you can come again soon.

(3) Q: Are you free tomorrow?

A: Yes, it's Sunday.

앞서 언급한 것처럼 drugstore—place—have lunch.를 A drugstore is the place to have lunch.로 「착각」하는 것이 바로 「듣기」의 능력의 일부를 이루듯이, 「듣기」의 능력의 일부를 이루는 요소로는 「豫測」의 능력도 있다. 모국어의 경우 일상 언어 생활에서 우리는 예사로 상대방의 말을 어느 정도 들었을 때 상대방이 말을 끝마치기 전에 상대방의 말을 앞질러 말하기도 하고, 상대방이 한 말을 정정해 주기도 한다. 「듣기」문항 작성에는 당연히 이와 같은 「듣기」의 能力도 고려되어야 할 것이다.

다음과 같은 일종의 完成問題는 이런 豫測의 능력을 평가할 수 있는 예가 되겠다.

5. 여러분이 듣는 문장은 마지막 낱말을 보충하여야만 완전한 문장이 됩니다. 녹음을 자세히 들은 다음 신호가 울리면 마지막 말을 답지에서 고르시오.

(1) I didn't go to school. I wasn't feeling well, so my mother made me stay (신호).

A. school

B. home

C. out

(2) Korean children are taught not to talk while eating. But the Western custom is that everybody talks and a lively conversation always goes on at the (신호).

A. food

B. room

C. table

주지하는 바와 같이 「듣기」의 평가에서는 「그림」을 이용하는 방법도 많이 보급되어 있다. 그림을 이용하는 경우에 유의할 사항은 특수한 경우(즉 낱말의 테스트 등 한정된 경우)를 제외하고는 녹음으로 들려주는 문장이나 대화 또는 문단 전체의 이해와 그림의 내용이 직결되도록 그림을 꾸미는 일이다. 예를 들어 A puppy is following a lady.란 문장을 들려주고, 다음과 같은 세개의 그림을 제시해서 맞는 그림을 고르게 하는 문제를 생각해 보자.



- A. 강아지를 안고 가는 부인
- B. 부인을 뒤따라 가는 강아지
- C. 고양이에게 밥을 주고 있는 부인

그런데 위의 문제에서 정답 B는 녹음으로 들려주는 follow라는 낱말 하나만 들어도 맞출 수가 있다. 차라리 그림 C를 「강아지를 뒤따라 가는 부인」의 그림으로 바꾸면 이 문제는 A puppy is following a lady.란 문장 전체의 이해 여부를 테스트할 수 있을 것이다.

### V. Cloze test에 관한 論議

오늘날 외국어의 능숙도의 평가를 운운할 때 논의의 대상에서 빼놓을 수 없는 것은 Cloze test이다. Cloze test는 주어진 문장의 빈 칸을 메꾸는 이른바 完成問題와 비슷해 보이지만, 주어진 text의 n번째에 해당하는 단어를 기계적으로 빈 칸으로 바꾸는 것을 원칙으로 삼는 특징을 지니고 있다. n번째라고 하지만 대체적으로는 5-7번째의 단어를 빈 칸으로 바꾸는 방식이 가장 흔히 쓰인다. 그러므로 Cloze test는 잘막한 單一文章만 가지고는 만들 수가 없다. Heaton(1975)에 의하면 빈 칸이 많으면 많을수록 Cloze test의 신뢰도는 높아지는데 40~50 정도의 빈 칸을 갖는 문제가 바람직하다. 따라서 다섯이나 여섯번째의 단어를 빈 칸으로 바꾼 Cloze test는 250 내외의 단어로 구성된 문단이 text로 적합한 셈이다.

우리는 가령 신문을 읽을 때 활자의 오식을 쉽게 찾을 수 있고, 지워져 잘 보이지 않는 활자도 앞뒤의 문맥을 근거로 무슨 뜻인지를 짐작할 수 있다. 이와 같이 앞뒤의 문맥을 통해서 어떤 빈 곳의 의미를 찾아내는 능력은 단어와 문법 전반에 걸친 지식을 포함하는 종합적인 언어 능력의 소산이다. Cloze test는 바로 이런 종합적인 언어 능력을 테스트해 주는 이점이 있다. 즉, Cloze test는 integrative test인 것이다. 그런데다 「번역」이나 「作文」 등의 integrative test는 채점의 절차상 <주관성의 개입>이란 단점을 가지고 있는데 반해서, Cloze test는 그 채점이 용이하고 채점의 객관성도 보장될 수가 있다. —Oller를 위시한 Cloze test의 주창자들은 이렇게 주창한다.

하기는 Cloze test가 긍정적으로만 받아들여지고 있는 것은 아니다. Cloze test에는 다음과 같은 변수가 작용하고 있음을 흥미로운 실험을 통해서 보여주고 있는 것은 Alderson (1979)이다.

#### (1) text 자체의 난이도

—즉, text 자체가 쉽건 어렵건 Cloze test의 결과는 동일한 것인가의 문제

#### (2) 채점 방식

—즉, 특정한 단어 하나만을 정답으로 삼느냐, 의미가 비슷한 단어는 어느 것이나 정답으로 삼느냐에 따라 채점 결과나 테스트의 타당도가 어떻게 달라지는가의 문제

## (3) 단어 삭제의 빈도

—즉, 단어를 빈 칸으로 바꿀때 그 빈도가 채점결과나 테스트의 타당도에 미치는 영향 Cloze test에서 이런 변수가 미치는 영향을 살펴보기 위해서 Alderson이 실시한 실험의 절차는 다음과 같다.

(1) 난이도가 다른 세 가지 text를 선정하였다.

(2) 이 세 가지 text를 이용해서 6, 8, 10, 12번째의 단어를 각각 빈 칸으로 메꾼 12종의 테스트를 작성하였다.

(3) 하나의 답안지는 다음 다섯 방식으로 채점되었다.

- A. 원본에 나왔던 단어 하나만을 정답으로 한다.
- B. 의미가 유사한 낱말은 전부 정답으로 한다.
- C. 품사가 같으면 정답으로 한다.
- D. 기능이 같으면 정답으로 한다.
- E. 문법상 맞는 것은 전부 정답으로 한다.

(4) 채점 결과를 피시험자가 별도로 본 ELBA test(영국의 여러 대학에서 입학을 원하는 외국어 학생에게 과하는 영어 시험)의 결과와 비교·분석해서 상관성을 알아보았다.

Alderson은 이 실험의 결과를 다음과 같이 요약하고 있다.

(1) 삭제되는 단어의 빈도에 따라 채점 결과는 여러 가지로 달라진다.

(2) 상대적으로 내용이 어려운 test를 사용한 Cloze test의 채점 결과는 ELBA test의 문법·어휘 영역의 채점 결과와 상관도가 높았으며, 쉬운 text를 사용한 Cloze test는 ELBA test의 「듣기」 평가와 상관도가 높았다.

(3) 채점의 방식으로는 의미가 유사한 낱말은 전부 정답으로 처리하는 방식이 가장 타당도가 높아 보인다.

다시 말해서 Alderson은 Cloze test는 본질적으로 文章依存的(sentence-bound)이며 문법·어휘 등 기본적인 어휘 능력을 평가할 수는 있지만 고도의(종합적인) 언어 능력을 테스트 하는 힘은 약하다는 결론을 짓고 있다. 이 말은 “they (cloze tests) require utilization of discourse level constraints as well as structural constraints within sentences”라고 말한 Oller (1979: 347)와는 어긋나는 주장인 것이다.

이와 같은 상반된 견해는 어떻게 받아들여져야 할 것인가?

Cloze test는 당초 종합적인 언어 능력의 평가를 표방한 테스트의 한 형식이다. 그런데 종합적인 언어 능력은 sentence-bound test보다는 discourse-centered test가 효율적인 것이 분명하다. 만약 Cloze test가 문장 의존적이라면 단순한 완성 문제와 다를 바가 없다. 그리고 어휘나 문법 사항은 단순한 완성 문제나 선택 문제만으로도 테스트할 수 있는 것이 사실이다. 그러나 cohesion이나 coherence가 discourse의 기본 요건임을 감안할 때, 250 내의

의 단어로 된 discourse를 text로 삼는 Cloze test의 모든 빈 칸이 전부 문장 의존적일 수는 없지 않을까? 다만 단어의 삭제가 기계적이고 보면, 문장 의존적인 요소의 삭제도 불가피 하기는 하다. 이렇게 본다면  $n$ 번째의 단어를 기계적으로 빈 칸으로 바꾸는 원칙을 수정하고, 문장보다는 전체문맥과 관련이 깊은 단어를 삭제토록 하는 방법을 가미함으로써 문장 의존적인 요소는 크게 조정될 수 있을 것이다.

Cloze test는 여러 가지 문제를 안고 있음이 분명하다. 이 글에서 Alderson의 실험을 길게 소개한 것은 Cloze test가 안고 있는 여러 문제의 일단을 드러내려는 데도 목적이 있었다. 하지만 외국어의 능숙도를 평가하는 형식이나 문제로 현재 통용되고 있는 것치고 문제를 안고 있지 않는 것이 과연 있을까?

Cloze test는 여러 가지 문제를 안고 있고 이와 같은 문제를 해결하는 데에 우리는 중지를 모아야 하겠지만, 특히 독해력과 관계되는 종합적 능력을 평가하는 데 있어 Cloze test가 효율적으로 이용될 가능성을 지니고 있음은 분명하다.

## VI. 結 語

이상으로 「듣기」와 「말하기」의 평가에 관련된 문제점과 海外學界의 動向 등을 살펴 보았다. 結論은 우리나라에서 「듣기」와 「말하기」의 평가의 문제가 가장 절실한 중·고등학교의 외국어 교육의 현장에 초점을 맞추기로 한다.

序論에서도 언급했듯이 우리 나라에서도 「듣기」나 「말하기」의 능숙도의 평가는 현실적으로 필요해지고 있다. 그런데 중·고등학교의 경우 「듣기」나 「말하기」의 올바른 평가는 학년별로 <어떤 주제에 관해서 얼마 만큼 어떻게 말을 하고 말을 알아들었을 때 어느 정도로 말을 잘하고 말을 잘 알아듣는다>고 할 수 있을 것인가란 基準의 設定이 전제가 된다. 그리고 이와 같은 기준의 설정은 외국어 교육 과정의 상세화를 필요로 한다. 아직까지는 중·고등학교의 수업의 현장에서 외국어의 평가는 역시 문법과 독해력이 중심이 되고 있어서 「듣기」나 「말하기」의 평가는 그 형식이나 내용보다 시행 자체에 의의를 찾고 있는 것이 실정이다. 그러나 앞으로 중·고등학교의 외국어 교육에서 「듣기」나 「말하기」는 「읽기」 못지 않게 강조되어야 할 것이고, 그런 수업이 효율을 거두기 위해서는 「듣기」와 「말하기」의 올바른 평가가 수반되어야 한다. 그런데 올바른 평가를 위해서는 무엇보다 외국어 교육 과정의 상세화가 시급한 것이다. 또한 평가를 전제로 한 교육과정의 상세화는 교육 과정의 의미를 가능하게 하고 교육과정 자체의 補完에도 기여할 것이다.

## 參 考 文 獻

- Allen, J.P.B., and Alan Davies. 1977. *Testing and experimental methods*. London: Oxford University Press.
- Alderson, J. Charles. 1979. 'The Cloze Procedure and Proficiency in English as a Foreign Language', *TESOL Quarterly* 13, no. 2, pp.219-27.
- Briere, Eugene and Frances Butler Hinofotis. (eds.) 1979. *Concepts in language testing: Some recent studies*, Washington, D.C.: TESOL
- Clark, John L.D. 1972. *Foreign language testing: theory and practice*. Philadelphia: Center for Curriculum Development
- Clark, John L.D. 1979. 'Direct vs. semi-direct tests of speaking ability.' In Briere and Hinofotis, 35-49.
- Davies, Alan. 1975. 'Two tests of speeded reading.' In Jones and Spolsky, 119-130.
- Harris, David P. 1969. *Testing English as a second language*. New York: McGraw Hill
- Heaton, J.B. 1975. *Writing English language tests*. London: Longman
- Heaton, J.B. (ed.) 1982. *Language Testing*. Modern English Publications, Ltd.
- Johnson, Robert K. 1982. 'Questioning some assumptions about cloze testing'. In Heaton (1982), 59-72.
- Jones, R.L. and Bernard Spolsky (eds.) 1975. *Testing language proficiency*. Arlington, Virginia: Center for Applied Linguistics
- Lado, Robert. 1961. *Language testing*. New York: McGraw Hill
- Oller, J.W 1979. *Language testing at school*. London.: Longman
- Palmer, Leslie and Bernard Spolsky (eds.) 1975. *Papers on language testing: 1967-74*. Washington, D.C.: TESOL
- Spolsky, Bernard. 1976. 'Language testing: art or science'. Paper read at the Fourth International Congress of Applied Linguistics. In *Proceedings of the fourth international congress of applied linguistics*. Stuttgart, Germany: Sonderdruck, 9-28.
- Upshur, John A. 1971. 'Objective evaluation of oral proficiency in the ESOL classroom'. *TESOL Quarterly* 5, 47-60.
- Valette, Rebecca M. 1977. *Modern language testing*. Second edition. New York: Harcourt, Brace and Jovanovich

Some Theoretical and Practical Considerations in Testing  
Foreign Language Proficiency

MOON Yong

**Abstract**

One of the most important considerations in designing a test of foreign language proficiency is the purpose for which the test is to be used. Under this assumption, the respective merits and demerits of discrete point test and integrative test are discussed.

In Korea, interest in testing listening and speaking abilities has grown remarkably in recent years. In keeping up with this increasing interest, this paper also proposes some criteria that hopefully increase objectivity in evaluating oral proficiency especially in high school English classes.