

英語教育測定— 서울大學校 語學研究所作成 英語學力標準化検査의 結果報告⁽¹⁾

Standardized Test for Highschool Students of English

서울大學校 語學研究所

教育에는 恒常 測定이 따르기 마련이다. 測定이야 말로 教育의 成果를 判斷하고 教育內容의 長短과 虛實을 가름하는 것이기 때문이다. 이런 意味에서 自古로 語學教育에도 반드시 考查라는 測定方法이 뒤따르기 마련이었다.

그러면 이처럼 語學教育의 實質의 一部를 構成하고 있다고 볼 수 있는 考查乃至는 檢查가 果然 우리나라에서는 어느程度 진지하게 다루어져 왔던가? 出題의 容易함을 위하여正確한 測定이 不可能한 애매한 問題가 考查 또는 檢查로서 行세하여 오지 않았던가? 出題의 奇와 異 만을 쫓는 남어지 語學習得이라는 廣範한 分野가 갖는 知識을 網羅해야 할包括性이 餘地 없이 牺牲된 일은 없었던가? 宜當히 測定되어야 할 語學能力의 限界外의 것들이 出題技術上의 巧와 妙로 錯覺되어 온 일은 없었던가? 이러한 疑問에 대한 答은若干의 反省을 아끼지 않는다면 自明하다고 보는 것이 安當할 것 같다.

以上은 教育測定 自體가 갖는 元來의 特性이란 面에서의 反省이지만 여기에 또한가지 看過하지 못할 考慮點이 있으니 그것은 各級學校의 入學試驗이란 教育測定이 갖는 特性

(1) 本 檢查製作을 為한 委員會는 아래와 같이 構成되었다.

委員長	黃燦鏗	
顧問	李鍾洙	Fred Lukoff, Walter R. Owens
問項委員	金鍾云	
測定委員	金豪權	鄭元植
出題委員	黃燦鏗	安榮輝 李季順 金鍾云 李鍾贊 金鎮宇
研究補助員	安成道	金正雄 洪恩姬 李貞順
打字手	黃誠子	
謄寫整理手	李相圭	

이다. 即 우리는 入學試驗이 下級學校의 教育方法이나 內容에 미치는 影響의 重大함을 指摘하고 싶은 것이다. 考查나 檢查는 特히 이러한 面에서도 慎重히 다루어져야 한다고 우리는 늘 생각하여 왔다.

多幸히도 근자에 急激히 發達하게 된 言語學理論——特히 그 가운데에서도 言語의 比較分析教育法을 利用한다면 위에서 指摘한 바와 같은 盲點과 矛盾을 除去할 수 있다는 展望이 서오던차 이러한 試圖에 對한 各界의 贊同을 얻게되었으며 1964年3月에는 亞細亞財團의 財政支援을 얻기에 이르렀다. 于先 서울大學校의 入試英語考查의 바탕이 될 수 있도록 同校志望者가 比較的 많은 高等學校를 全國的으로 選定하여 學力検査를 兼한 標準化検査를 實施하자는 案이 具體화되기에 이른 것이다.

그後 戒嚴令宣布에 뒤이은 數個月의 空白을 包含하여 一年有餘 本研究所 教職員一同을 主軸으로 하는 標準化考查委員會와 學生指導研究所와 師範大學의 教育學科의 教授一部의 協力으로 檢查型의 作成과 全國的 檢查實施 그리고 이에 따르는 結果의 分析이 進行完了 되었다. 아래에 그 結果를 公開報告한다.

다만 이러한 種類의 檢查가 語學教育分野에서 그리 흔히 있는 일이 아님에 비추어 아래에 記述할 內容을 署純한 報告의 形式에 局限하지 않고 第一節에서 標準化検査에 대한一般的 概念 解說을 添加하였다.

第二節 以下是 通常의 報告의 形式을 쫓아 本研究所에서企劃한 標準化検査의 過程과 그 結果의 解釋과 本検査의 良好度(信賴度와妥當度)를 系數의으로 算出立證하였다. 本論에 들어가기에 앞서 이번의 標準化検査製作의 全過程을 通하여 Robert Lado博士의 力著 Language Testing⁽²⁾이 始終 重要한 指針書로 利用되었음을 밝혀 둔다.

第一節 標準化検査란?

人間의 여러가지 能力이나 學業成績 또는 性格特性을 測定하려고 할 경우에 우리는 흔히 考查 또는 檢査를 만들어 使用한다. 그런데 이런 考查 또는 檢査는 그 作成의 目的, 機能作成者の 性質에 따라서 보통 非公式的 檢査(informal test)와 標準化検査(standardized test)의 두가지 種類로 나눌 수 있다. 非公式的 檢査의 대표적인 例는 흔히 學校에서 教師

(2) Robert Lado, Language Testing, Longmans, London, 1962.

들에 의해서 만들어지고 使用되는 教師作成検査(teacher-made test)이다. 한 單元의 學習이 끝난 뒤에 그 學習成果(achievement)를 評價하기 위한 考査, 學期末에 시행되는 學期末考査 등이 모두 이런 非公式的 教師作成検査에 포함된다. 그런데 이런 非公式的 教師作成検査는 보통 그 用途가 极히 制限되어 있어서 서로 다른 教育 課程計劃, 서로 다른 教科書, 서로 다른 教師 밑에서 배운 여러 學生들의 成績을 公平하게 評價할 수가 없다. 예컨대 A學校에서 만들어진 考査는 A學校의 學生들에게는 알맞는 考査가 될지 몰라도, B學校의 學生들에게는 적합한 考査가 아닐 수도 있다. 따라서 이와 같은 非公式的 檢査의 特殊性, 局地性을 克服하여 모든 學生들의 成績을 公平하고 보편적인 方法으로 测定하기 위해서, 檢査의 目的을 一般化하고 檢査의 節次를 標準化한 것을 標準化検査라고 한다. 이런 標準化検査의 구체적인 特徵을 살펴보면 다음과 같다.

첫째, 檢査의 目的이 一般的이다.

어떤 特定 學校의 學生들에게만 알맞는 考査가 아니라, 全國의 어떤 學校의 學生들의 成績도 비교적 公平하게 测定할 수 있게 되어 있다. 몇몇 教師들의 主觀的인 判断에 의해서 出題될 것이 아니고, 教科 專門家들에 의하여 보편적인 基礎에서 각 學科領域의 重要的 문제들이 選定된다. 또 單元考査나 學期末考査와 같이 短期間의 學習成果가 아니라, 한 學科의 全領域에 걸친 一般的의 學習 成就度(achievement)를 测定하게 되는 것이다.

둘째, 测定理論上으로 洗練된 문제들로서 檢査가 構成된다.

非公式的인 教師作成検査에서는 각 문제가 测定理論에 입각한 엄격한 檢討를 거치지 않고 考査에 수록되는 것이 보통이다. 따라서 이런 考査의 문제들 중에는 测定理論에 비추어 不良한 문제들이 섞일 可能성이 있고 따라서 學生들의 成績을 엄밀 정확하게 测定하기 어렵다. 標準化検査에서는 檢査의豫備實施(tryout)에 의하여 각 문제의 良好度를 일일히 檢討하여 测定理論上的 基準을 만족시키는 문제들만을 수록하기 때문에 一般的으로 檢査의 信賴度, 妥當度가 极히 높다.

셋째, 檢査의 實施節次가 一定化되어 있다.

非公式的 檢査에서는 檢査 實施時의 指示, 說明 등이 明確하게 規定되어 있지 않기 때문에, 檢査 實施者마다 相異한 指示, 說明을 할 可能성이 있다. 더구나 檢査時間, 採點方法 조차 仔細히 規定되어 있지 않아서, 여러 다른 檢査實施者 밑에서 受檢하는 學生들

의 成績을 直接 비교할 수 없는 경우가 있을 수 있다. 標準化検査에서는 이런 檢查實施條件, 採點條件를 一定化 또는 固定化하여 얻은 成績들을 더욱 엄밀한 條件下에서 서로 비교할 수 있도록 해준다.

네째, 檢查結果를 서로 廣範하게 比較할 수 있다.

非公式的 檢查의 結果는 앞에서 말한 여러가지 理由 때문에 그 可比較性이 매우 희박하다. 그러나 標準化検査에서는 檢查의 目的 自體가 보편的일 뿐만 아니라 흔히 넓은 地域 또는 全國에 걸쳐 事前에 實施되어 檢查點數의 分布가 規準(norm)으로 提示되어 있기 때문에 한 學生의 成績을 한 學級, 한 學校內에서뿐만 아니라 全國的인 見地에서 比較할 수 있고 解釋할 수 있다.

위에서 알 수 있는 바와 같이 標準化検査를 作成하는데 있어서는 教科專門家와 測定 專門家의 協調가 必須의 要素이다. 教科專門家들의 教科에 對한 지나친 獨占性과 主觀 고집 때문에 入學試驗에 想像과 能力 以上的 문제들이 나와 受驗生들을 당황하게 하는例를 흔히 볼 수 있다.勿論 入學試驗이란 選拔을前提로 한 一種의 競爭시험에서 얼핏 생각하기에는 標準化検査와는 無關한 것으로 보일지도 모른다. 그러나 標準化検査의 一般的인 特性으로 볼 때, 入學試驗이야말로, 中學校 全教科課程, 高等學校 全教科課程을 總網羅하여 그 成就度를 測定하는 것으로서 가장 標準化되어야 할 性質의 試驗이다. 더우기 入學競爭 試驗은 경쟁자의 選拔로 끝나는 것이 아니고 上級學校에로의 教育의 延長이라는 것을 생각할 때 教科專門家와 測定專門家 間의 相互 긴밀한 協調로써 各級學校의 入學試驗의 標準화가 時急히 要請되며 또한 좋지 못한 入試出題 傾向으로 因하여 下級學校의 教育에 미치는 惡影響을 除去해야 할 것이다.

第二節 語學研究所 英語學力標準検査의 製作過程

1.豫備検査(Tryout)

豫備検査란 最初 出題된 많은 問項中 가장 適切한 것만을 選擇하여 하나의 完成된 最終型에 到達하기까지 거쳐야 하는 몇 차례의 試験을 말한다. 完成型을 얻기까지에는 于先 英語教科專門家들에 의해 檢查測定될 問題範圍(problem area)의 討議 決定이 있어야 한다. 卽 이 檢查에서 무엇을 測定할 것인가 하는 出題의 範圍를決定하는 것이다. 여기에 있어서 標準化検査의 特性인 一般性 客觀性 包括性에 留意해야 함은勿論이다.

다음 段階로는 위에 定한 領域을 基礎로 하여 教科專門家들이 出題하고, 出題된 問項들을 比較 檢討하고 取捨選擇하여 몇 개의豫備考查型을 만드는 일이다. 이와 같이 하여 이루어진豫備考查型을 몇 學校를 選擇하여 實施하게 된다.豫備檢查란 이 實施段階만을 일컫는 경우가 많다. 實施對象學校數는 한 學級當 學生數를 60名으로 잡는다면 七個校로서 應試 總人員數가 400餘名 程度면 적당하다. 이때의 學校 選定은 教科專門家의 힘만으로는 不可能하고, 測定專門家의 協調를 얻어 各集團을 고루고루 代表할 수 있도록 해야 한다.豫備檢查는 一次로써 끝낼 수도 있겠으나 더 좋은 問項을 얻기 爲하여 出題問項數를 많이하고 그 回數도 三次로 늘이는 것이 보통이다.

一次와 二次에서는 時間에 別로 拘碍됨이 없이 좋은 問項選擇에 더 무게를 두고, 거의 完成에 가까운 型을 만들어 가지고 三次에서 時間配定 關係를 測定하게 된다.

以上 말한 것이 대개 本檢查實施前 完成型을 얻기까지의 過程인데, 다음에는 本研究所에서 擇한 過程을 段階的으로 說明 記述하려 한다.

1. Problem-area 와 Sub-area 의 選定

本研究所(以下 本所로 省略)에서는 教科專門家로 構成되는 委員會를 數次 열어 問題領域을 發音 文法 語彙 讀解 作文 등 五個部로 定하고 각部는 여러 細部項目을 包括하도록 하였다. 그것을 다음에 提示해 보인다.

Problem-areas and Sub-areas to Be Tested

Area	Weight	Sub-areas
I . Phonology	10	1. Consonants 2. Vowels (stressed and unstressed) 3. Allophonic variation 4. Word stress 5. Phrase stress 6. Intonation 7. Pronunciation of endings 8. Assimilation 9. Blending 10. Thought group
II . Grammatical structure		1. Word order 2. Number 3. Case 4. Agreement 5. Kinds of verb 6. Tense 7. Sequence of tense 8. Gerund 9. Participle 10 Infinitive 11. Voice 12. Mood 13. Complement 14. Comparison 15. Article 16. Preposition 17. Relatives 18. Punctuation
III . Vocabulary		1. Lexical meaning 2. Form 3. Compounds 4. Idioms 5. Connotations 6. Figure of speech 7. Derivatives 8. Inflexions

- | | |
|-------------------|--|
| IV. Comprehension | 1. Contextual understanding 2. Cultural understanding |
| V. Composition | 1. Sentence building: simple, compound and complex. |

2. 出題와豫備検査型作成

앞에 말한 五個部分에 基礎를 두고 各 教科專門家들에 의하여 作成된 千餘個의 많은 問項中에서 檢査初案 作成委員會는 高校 英語教育目標 및 本 檢査의 目的에 留意하여 三百 問題를 精選하였다. 그래도 問項數가 너무 많아서 모든 問項을 同一人에게 實施하는 것이 不可能하였으므로, 精選된 300 문항을 A,B,C, 3型에 각각 100 문항씩 分類하여 수록하였다. 여기서 委員會가 特히 腹心한 點은 各型 問의 平均, 標準偏差를 같게 하고 또 各型마다 測定할 問題領域과 細部項目을 다 均等하게 包含하도록 하는 것이었다.

3. 三次豫備検査와 問項選定

前述한 바와 같이 一, 二次豫備検査에서는 時間에 制約을 주지 않고 問項選定에만 重點을 두었다. 一次検査 實施 學校는 서울市內 高等學校中에서 各 集團을 代表하는 7個校의 7個學級(學生數 約 400名)을 擇하였다. 二次에서는 一次検査에 參加한 學校를 避하고 다른 學校中에서 7學級을 擇하였다. 이는 同一人이 同一한 問題를 重複하여 풀게 될 때 必然的으로 發生하게 되는 問項困難度(item difficulty)의 變動을 없애고, 集團의 範圍를 擴大시키자는 데서였다. 第三次豫備検査에서는 一, 二次 檢査의 結果 좋은 問項으로選定된 것들 중에서 110 問項을 뽑아 最終型으로 만들어, 時間關係에 重點을 두고 實施하여 보았다. 時間 測定 方法으로는 受檢者의 約 80%에 해당하는 數의 學生들이 檢査를 다 끝내고 난 후의 時間을 채서, 이 時間들의 平均을 擇하였다. 一, 二次 檢査에서는 엄격한 時間制約은 없었지만 110分 程度로서, 受檢者들이 모든 問項을 할 수 있도록 하였던 것인데 三次 檢査의 結果 110 問項을 푸는데 所要되는 時間을 110分으로 決定하였다.

다음에는 參考로 一, 二次 檢査에서 問項分析에 의하여 좋은 問項을 選定하는 方法을 說明하기로 한다. 一般的으로 말해서 한 問項의 좋고 나쁨은 檢査 結果 算出되는 問項의 困難度(正答率)와 問項의 辨別力(discriminating power)로써 決定된다.豫備検査에서 얻은 全 資料中에서 時間 未達項(non-reach)과 除外項(omitt)이 적은 答案紙 370枚를 추려서 上位 27%에 해당하는 100枚와 下位 27%에 속하는 100枚를 擇하여 積率相關係數를 推測하여 얻은 數値가 discriminating power다. 問項選定의 標準으로는 正答率이 25~75%,

辨別力이 20 이상의 것을 採擇하였음을 밝히는 바이다.

2. 本 檢 查

1. 對象校의 選定과 實施

前述한 三次의豫備検査에서 完成型을 만든은 다음 全國 高等學校中 31個校를 擇하여 本所 職員들을 直接 派遣하여, 全國的으로 實施하였다. 校校選擇은, 本標準化検査의 目的이 서울大學校 入試의 標準을 定하기 為한 것이기 때문에, 서울大에 入學한 學生이 있는 學校를 擇하였다.

2. 採點 其他 統計的 計算

全國의으로 採集된 答案紙 中에서 未達項과 除外項이 적은 答案紙 1655枚를 擇하여 總點의 平均과 標準偏差를 計算하고 各部別 平均과 標準偏差도 알아 보았다. 또 檢查의 良好度를 알기 위하여 信賴度, 妥當度를 計算하였다. 以上의 統計的 計算은 「結果의 解釋」에서 모두 仔細히 論하기로 하고 여기서는 다만 檢查의 한 過程으로 言及에 그칠 뿐이다.

参考로 이 檢查의 總點 平均과 標準偏差, 各部別 平均과 標準偏差를 表示하면 다음과 같다.

	平均	標準偏差
總 點 (100점)	41.7	16.98
第一部 (10점)	4.7	1.04
第二部 (20점)	9.3	3.58
第三部 (35점)	15.3	6.15
第四部 (20점)	8.1	4.86
第五部 (15점)	5.8	3.08

3. 規準表(Norm) 作成

前述한 諸 統計的 計算에서 얻은 資料를 가지고 規準表을 作成하였다. 本所 作成 規準表에는 크게 나누어 個人規準表(individual norm)와 學級規準表(class norm)가 있다. 또 個人規準表는 檢查 總點規準表와 各部別 規準表가 있어서, 總點은 勿論 各部分 部分에서도 個人の 位置를 全國의인 見地에서 알 수 있도록 하였다. 規準表에서의 尺度로는 原點數

에 해당하는 百分位點數와 T點數를 記入하여 各尺度間의 缺點을 相補하여 解釋에 正確을 期하도록 하였다. 이 規準表에 對하여는 다음에 論한 「結果의 解釋」에서 다시 한 번 仔細히 說明하고자 한다.

I. 個人 規準表 (Individual-norm table)

A. 總點 規準表

100點을 滿點으로 하였다.

原點數 (raw score)	百分位點數 (percentile Scores)	T點數 (T-Scores)	原點數 (raw score)	百分位點數 (percentile Scores)	T點數 (T-Scores)
100	—	—	83	98.7	74.3
99.5	—	—	82.5	98.7	74.0
99	—	—	82	98.5	73.7
98.5	—	—	81.5	98.4	73.4
98	—	—	81	98.2	73.1
97.5	—	—	80.5	97.9	72.8
97	—	—	80	97.7	72.5
96.5	—	—	79.5	97.3	72.2
96	—	—	79	97.1	71.9
95.5	99.9	81.8	78.5	96.9	71.6
95	99.9	81.5	78	96.6	71.4
94.5	99.9	81.2	77.5	96.4	71.1
94	99.9	80.8	77	96.2	70.8
93.5	99.9	80.5	76.5	95.8	70.5
93	99.9	80.2	76	95.5	70.2
92.5	99.9	79.9	75.5	95.2	69.9
92	99.9	79.6	75	94.9	69.6
91.5	99.9	79.3	74.5	94.6	69.3
91	99.7	79.0	74	94.3	69.0
90.5	99.7	78.7	73.5	94.1	68.7
90	99.7	78.4	73	93.9	68.4
89.5	99.6	78.1	72.5	93.7	68.1
89	99.6	77.8	72	93.5	67.8
88.5	99.6	77.5	71.5	93.0	67.5
88	99.6	77.2	71	92.7	67.2
87.5	99.5	76.9	70.5	92.3	66.9
87	99.4	76.6	70	91.9	66.6
86.5	99.3	76.4	69.5	91.6	66.4
86	99.3	76.1	69	91.4	66.1
85.5	99.1	75.8	68.5	91.2	65.8
85	98.9	75.5	68	90.8	65.5
84.5	98.9	75.2	67.5	90.4	65.2
84	98.8	74.9	67	90.1	64.9
83.5	98.7	74.6	66.5	89.7	64.6

66	88.9	64.3	43.5	60.3	51.1
65.5	88.5	64	43	59.2	50.8
65	88.1	63.7	42.5	58.4	50.5
64.5	87.6	63.4	42	57.4	50.2
64	87.3	63.1	41.5	56.3	49.9
63.5	86.7	62.8	41	54.7	49.6
63	86.2	62.5	40.5	53.8	49.3
62.5	85.8	62.2	40	52.9	49.0
62	85.2	61.9	39.5	52.0	48.7
61.5	84.5	61.6	39	51.2	48.4
61	83.8	61.4	38.5	50.2	48.1
60.5	83.4	61.1	38	49.4	47.8
60	83.0	60.8	37.5	48.1	47.5
59.5	82.2	60.5	37	47.6	47.2
59	81.7	60.2	36.5	46.6	46.9
58.5	81.2	59.9	36	45.5	46.6
58	80.7	59.6	35.5	44.5	46.4
57.5	79.9	59.3	35	43.6	46.1
57	79.5	59.0	34.5	42.5	45.8
56.5	78.7	58.7	34	41.3	45.5
56	78.1	58.4	33.5	40.1	45.2
55.5	77.5	58.1	33	38.8	44.9
55	76.9	57.8	32.5	37.7	44.6
54.5	76.3	57.5	32	36.3	44.3
54	75.9	57.2	31.5	35.1	44.0
53.5	75.2	56.9	31	33.8	43.7
53	74.6	56.6	30.5	31.7	43.4
52.5	74.1	56.4	30	30.1	43.1
52	73.3	56.1	29.5	28.2	42.8
51.5	73.0	55.8	29	26.3	42.5
51	72.3	55.5	28.5	24.8	42.2
50.5	72.0	55.2	28	23.4	41.9
50	71.1	54.9	27.5	21.7	41.6
49.5	70.4	54.6	27	20.6	41.4
49	69.8	54.3	26.5	19.3	41.1
48.5	69.2	54.0	26	17.8	40.8
48	68.5	53.7	25.5	17.2	40.5
47.5	67.5	53.4	25	15.2	40.2
47	67.1	53.1	24.5	13.2	39.9
46.5	66.4	52.8	24	12.0	39.6
46	66.0	52.5	23.5	10.5	39.3
45.5	64.0	52.2	23	9.4	39.0
45	63.0	51.9	22.5	7.8	38.7
44.5	62.7	51.6	22	6.6	38.4
44	61.1	51.4	21.5	5.6	38.1

21	4.7	37.8	10.5	0	—
20.5	4.1	37.5	10	0	—
20	3.3	37.2	9.5	0	—
19.5	2.6	36.9	9	0	—
19	1.9	36.6	8.5	0	—
18.5	1.6	36.4	8	0	—
18	1.3	36.1	7.5	0	—
17.5	1.2	35.8	7	0	—
17	1.0	35.5	6.5	0	—
16.5	0.7	35.2	6	0	—
16	0.4	34.9	5.5	0	—
15.5	0.3	34.6	5	0	—
15	0.2	34.3	4.5	0	—
14.5	0.1	34.0	4	0	—
14	0.1	33.7	3.5	0	—
13.5	0.1	33.4	3	0	—
13	0.1	33.1	2.5	0	—
12.5	0.1	32.8	2	0	—
12	0.1	32.5	1.5	0	—
11.5	0	32.2	1	0	—
11	0	—	0.5	0	—

B. 各部別 標準表

① 第一部

問項數 20 개, 10 點을 滿足으로 하였다.

原點數 (Raw Scores)	百分位點數 (Percentile Scores)	原點數 (Raw Scores)	百分位點數 (Percentile Scores)
10	—	4.5	55.3
9.5	99.3	4	44.3
9	98.3	3.5	33.0
8.5	97.0	3	21.0
8	94.3	2.5	12.5
7.5	91.5	2	4.0
7	86.5	1.5	1.3
6.5	83.3	1	—
6	76.3	0.5	—
5.5	69.5	0	—
5	62.0		

② 第二部

問項數 20 개, 20 點 滿點으로 하였다.

原點數 (Raw Scores)	百分位點數 (Percentile Scores)	原點數 (Raw Scores)	百分位點數 (Percentile Scores)
20	—	9	43.5
19	99.5	8	35.5
18	98.8	7	26.8
17	96.3	6	18.0
16	92.8	5	10.3
15	90.5	4	5.0
14	85.5	3	1.3
13	79.3	2	—
12	71.3	1	—
11	63.0	0	—
10	54.8		

③ 第三部

問項數 35 개, 35 點 滿點으로 하였다.

原點數 (Raw Scores)	百分位點數 (Percentile Scores)	原點數 (Raw Scores)	百分位點數 (Percentile Scores)
35	—	17	64.0
34	99.8	16	58.5
33	99.0	15	51.5
32	98.8	14	47.3
31	98.5	13	40.8
30	98.3	12	30.3
29	97.5	11	21.0
28	95.0	10	16.0
27	93.8	9	10.5
26	93.0	8	6.8
25	90.8	7	3.5
24	87.8	6	1.5
23	85.0	5	0.8
22	81.5	4	—
21	79.5	3	—
20	76.8	2	—
19	73.3	1	—
18	69.0	0	—

④ 第四部

問項數 20 개, 20 점을 滿點으로 하였다.

原點數 (Raw Scores)	百分位點數 (Percentile Scores)	原點數 (Raw Scores)	百分位點數 (Percentile Scores)
20	99.5	17	94.8
19.5	99.5	16.5	93.8
19	99.0	16	91.3
18.5	98.5	15.5	90.3
18	96.5	15	88.3
17.5	96.5	14.5	87.8
14	85.3	6.5	45.0
13.5	81.8	6	38.8
13	79.0	5.5	36.5
12.5	76.8	5	31.0
12	75.0	4.5	29.3
11.5	72.8	4	23.3
11	70.3	3.5	22.3
10.5	68.3	3	12.0
10	65.0	2.5	11.5
9.5	63.3	2	6.0
9	60.0	1.5	5.3
8.5	58.3	1	1.3
8	53.8	0.5	1.3
7.5	52.3	0	—
7	48.5		

⑤ 第五部

問項數 15 개, 15 점을 滿點으로 하였다.

原點數 (Raw Scores)	百分位點數 (Percentile Scores)	原點數 (Raw Scores)	百分位點數 (Percentile Scores)
15	99.5	7	63.3
14	98.8	6	54.3
13	97.3	5	40.8
12	95.0	4	25.0
11	90.8	3	11.8
10	87.3	2	4.0
9	79.8	1	1.5
8	73.0	0	—

II. 學級規準表 (class-norm table)

31個 學級의 平均點으로 學級規準表를 만들었다.

原點數 (Raw Scores)	百分位點數 (Percentile Scores)	原點數 (Raw Scores)	百分位點數 (Percentile Scores)
63~65	96.7	39~41	51.6
60~62	93.4	36~38	29.0
57~59	87.0	33~35	12.8
54~56	77.3	30~32	12.8
51~53	77.3	27~29	6.5
48~50	70.9	24~26	—
45~47	70.9	—	—
42~44	64.4	—	—

第三節 結果의 解釋

- 이 檢查의 結果(點數)는 아래에 規定된 바에 의하여 解釋되어야 하며 學級의 平均成積이나 原點數(raw score)에 의해서 任意로 解釋 判斷되어서는 안된다.

1. 百分位點數(Percentile Rank)

이 檢查에서는 檢查 總點과 5個의 各 部(part)別 點數의 規準을 百分位點數로 表示해 두었다. 卽, 이 檢查의 解釋에 있어서는 總點이거나 各部別 點數이거나를 莫論하고 正答數를 세어서 얻은 原點數(raw score)를 가지고 그대로 해석하는 것이 아니고 반드시 百分位點數 規準에 의해서 해석되어야 한다.

그러면 이 檢查에서 原點數를 그대로 利用하지 않고 구하여 百分位點數를 使用하는 理由는 무엇인가? 우리는 非公式的 檢查에서 대개 原點數만을 가지고 해석하고 使用했지만, 엄밀한 測定理論的立場에서 보면 이 原點數는 여러가지의 缺點을 가지고 있다. 예컨대 어떤 學生이 한 考查에서 60점을 얻었다고 하더라도 우리는 正確하게 이 學生의 成績을 짐작할 수는 없다. 왜냐하면 考查의 쉽고 어려움에 따라서 이 60점의 意味가 달라질 수 있기 때문이다. 또 어떤 考查에서 0點이 나왔다 하더라도 이 0點이란 것이 그 分野의 實力이 全然 없다는 것을 指示해 주지는 않는다. 마찬가지로 두가지의 다른 考查, 예컨대 數學考查의 50점과 英語考查의 50點도 각각 똑같은 程度의 實力を 가리키는 것으로는 보기 어렵다. 따라서 檢查點數를 정말로 업격하고 意味있게 해석하기 위해서는 이 原點數를

다른 點數體制로 換算하여 使用해야 한다. 이렇게 換算된 點數體制 중의 한가지가 本檢查에서 使用한 百分位點數다.

이 百分位點數란 소위 百分率(percentage)에서 나온 말이다. 卽 “原點數가 80 點이다”라고 해서는 엄밀한 해석이 不可能하기 때문에 이 特定한 原點數를 한 集團內의 席次로 바꾸되 이 席次를 “全體人員數를 100 名으로 보았을 때의 相對的인 席次”로 바꾼 것이 百分位點數인 셈이다. 따라서 百分率이 “100 名 中의 比率”을 나타내는 것이라면 이 百分位點數는 “100 名中의 席次”를 나타내게 된다. 그런데 이 百分位點數가 소위 席次와 顯著하게 다른 點은, 席次에서는 成績이 우수한 사람부터 1 등, 2 등, 3 등……으로 표시되지만, 百分位點數에서는 成績이 우수한 사람부터 99 등, 98 등, 97 등……으로 表示된다는 것이다. 예컨대 百分位點數로 50 點은 한 集團의 꼭 中央에 해당되는 點數이며, 百分位點數 75 點이란 것은 이 集團의 總人員數를 100 名으로 보았을 때에 이 학생의 點數보다 낮은 點數를 얻은 사람이 75%(100 명 中의 75)란 말이 되므로 비교적 우수한 成績을 나타낸다. 마찬가지로 한 학생이 百分位點數로 30 點을 얻었다면 이 학생보다 낮은 점수를 얻은 학생들은 全體의 30%(100 名 中의 30名)라는 말이므로 비교적 不良한 成績임을 나타낸다. 이 檢查의 實例를 들어보면 다음과 같다.

(1) 學生 A——原點數(總點) 38.5 點

學生 A 는 이 檢查의 總點이 38.5 點이다. 總點 規準表를 보면 原點數 38.5 點에 해당하는 百分位點數는 50.2이다. 따라서 이 학생 A 보다 낮은 성적을 얻은 학생을 全國 學生數를 100 명으로 보았을 때 約 50 名이다. 대체로 平均程度의 成績임을 알 수 있다.

(2) 學生 B——原點數(總點) 86.5

學生 B 는 總點이 86.5 點이다. 역시 規準表를 보면 이에 해당하는 百分位點數는 99.3이다. 이 학생의 성적보다 낮은 성적을 얻은 학생들은 전체의 99.3%나 된다. 즉, 매우 우수한 成績이라고 할 수 있다.

(3) 學生 C——原點數(總點) 28.5 點

學生 C의 原點數가 28 點이라면 그 百分位點數는 24.8임을 알 수 있다. 전체 학생 중 上位로 부터 約 $\frac{3}{4}$ 에 해당하는 成績이므로 대체로 不良한 成績이다.

2. T 點 數

앞에서도 말한 바와 같이 正答數를 세어서 바로 얻은 原點數에 대해서는 意味 있는 해석을 내릴 수가 없으므로 이 原點數를 百分位點數의 尺度로 바꾸어 使用하게 된다. 그런데 이 百分位點數 自體도 완전한 것이 아니라 그 自體에 허다한 弱點을 갖고 있다. 그中の 중요한 약점은 點數尺度에 同間性이 缺如되어 있다는 点이다. 예컨대, 百分位點數 尺度에서 (90點~80點)은 (60點~50點)과 현저하게 다르다. 數字上의 차이는 다 같이 10點씩이지만 이 두개의 10點이 나타내는 實際의 實力上의 差는 非常히 크다. 말하자면 百分位點數의 尺度를 100cm의 자(尺度)와 같은 것으로 생각하면, 이 자의 간눈 사이의 間隔이 서로 같지가 않다. 간눈 사이의 간격이 一定하지 않은 100cm 尺度는 아무짝에도 쓸모가 없듯이, 우리는 이런 결함을 가진 百分位點數의 尺度보다 좀 더 완전한 尺度에 의하여 點數를 표시하고 해석하고 싶다. 여기에 새로 考案된 尺度가 소위 T點數 尺度라는 것이다.

이 T點數란 것은 소위 統計에서의 平均值 및 標準偏差의 概念을 動員해서 만들어진 尺度이다. 標準偏差란 것은 한 分布를 이루는 點數들이 얼마나 分散되어 있느냐 하는 程度를 나타내는 統計的인 方法이다. 한 原點數에 대한 解釋에 있어서 이런 平均值 및 標準偏差가 動員되어야 하는 理由는 다음과 같다. 예컨대, 한 학생의 原點數가 60點이라는 原點數에 대해서는 아무런 의미있는 해석이 下可能하다. 그런데 이 경우에 이 학생이 屬하는 集團의 原點數 平均이 50點이라면 우리는 이 60點의 意味를 어느 程度 뿐만이 알 수 있다. 卽 原點數 60點에 대한 해석이 어느 程度 可能해진다. 그러나 이 原點數 60點이 平均點인 50點으로부터 얼마나 離脫된 點數냐 한다면, 간단히 $60 - 50 = 10$ 점이라고 만족해 버릴 수는 없다. 왜냐하면 이 10점이라는 것은 絶對的인 것이 아니고 檢查點數分布의 變散度(分散된 程度)에 따라 10點의 크기가 주는 意味가 달라질 수 있기 때문이다. 따라서 어떤 原點數에 대해서 완전한 해석을 하기 위해서 이와같이 그 分布平均值 및 標準偏差를 기초로 하여 해석하지 않으면 안된다. 原點數로부터 T點數를 算出하는 公式은 다음과 같다.

$$T = \left(\frac{X - M}{SD} \right) 10 + 50$$

여기서 $T \cdots T\text{-點數}$

$X \cdots \text{原點數}$

M … 原點數分布 平均值

SD …原點數分布의 標準偏差

이 公式에서 가장 重要한 部分은 等式의 右邊 第一項의 $(\frac{X-M}{SD})$ 이다. 여기에 곱하는 10은 단순히 小數點을 한 자리만큼 줄이는 役割을 할뿐이고, 50點을 더하는 것도 尺度의 中央點이 0이 되는 것을 막고 50이 되도록 하기 위한 것이다.

이 T 點數는 앞에서 말한 百分位點數尺度의 非同間性이라는 결함을 解消시키며, 여러 檢查의 點數들을 直接 相互比較할 수도 있고 이들 點數를 마음대로 加減할 수도 있다.

이 T 點數의 한가지 弱點은 이 T 點數의 正確한 意味를 알려면 먼저 平均值 및 標準偏差等의 統計的 概念을 完全히 理解하여야만 한다는 點이다. 그러나 尺度面에서 볼 때의 이 T 點數의 優秀性이 널리 알려져 있기 때문에 이 檢查에서도 T 點數의 規準表를 만들어둔 것이다.

3. 個人規準表(Individual norm Table)

이 檢查에서 作成한 規準表 中에는 大別해서 個人規準表와 學級規準表의 두가지가 있다. 이 中의 個人規準表는 다시 總點 規準表와 5個의 各部別 點數 規準表들로 나누어 진다.

이 檢查를 實施한 結果로 얻은 個人的 成績을 解釋하려고 할 때에는 이 個人規準表를 보도록 해야 한다.

이 個人規準表는 이 檢查의 檢準化集團과 비슷한 條件에 있는 學校의 학생들의 得點을 해석하는 데에 가장 適合하며一般的으로 水準이 떨어지는 學校의 學生들에게는 不利한 해석을 나이게 되기가 쉽다.

이 檢查를 實施하면 우선 각 學生의 總點 得點을 이 個人別 規準表에 비추어 해석해 보고 다음에 各部別 成績을 檢討해 보는 것이 좋다. 여기서 한가지 注意해 둘 것은 이 檢查에서의 個人的 得點도 測定上의 誤差의 영향을 어느 程度까지 받고 있다는 點이다. 그러나 나중에 信賴度에 관한 說明에서 다시 말하겠지만 檢查總點의 信賴度는 极히 良好하기 때문에 誤差의 作用이 아주 적은 것으로 볼 수 있지만, 各部別 檢查에서는 그 作用이 相對적으로 꽤 커지고 있다는 事實을 언제나 忽視에 두어야 할 것이다. 即 各部別 檢查의 個人規準表는 總點規準表에 비추어 절 正確한 셈이다.

4. 學級規準表(Class Norm Table)

i) 學級規準表는 個人的 得點이 아니라 한 學級全體로서의 成績을 意味있게 해석할 수 있도록 하기 위해서 만든 것이다. 이 檢查를 통해서 우리는 學生個個人의 成績을 正確하게 診斷할 수 있을 뿐만 아니라, 한 學級을 통털어 보았을 때 이 學級의 英語實力의 程度가 어느 程度인지도 이 學級規準表를 通해서 알 수 있는 셈이다.

ii) 學級規準表는 i) 檢查의 學級平均點을 기초로 하여 만든 것이다. 따라서 어떤 學級에서 ii) 檢查를 實施할 경우에는 個人別 得點에 대한 해석이 끝난 뒤 學級全體로서의 實力에 대한 해석을 해보는 것이 바람직하다. 이런 해석의 結果로서 흔히 教育課程의 展開計劃, 學習指導方法에 重要한 變革을 가져올 수 있다.

해석의 實例를 들어보기로 하면 다음과 같다. 한 學級에 속하는 여러 학생들의 總點 原點數를 合計하여 平均點數를 밝힌다. 그 平均點數가 가령 46點이었다면 그 百分位點數는 70.9이다. 따라서 全國의 學級中에서 이 學級의 平均點數인 46點보다 낮은 平均點數를 가진 學級이 70.9%나 된다는 말이다. 이 學級의 學生들은 全體的으로 상당히 좋은 成績을 갖고 있다고 할 수 있다.

iii) 學級規準表에 의한 해석에서 한가지 注意해야 할 것은 學級平均點數라는 것이 그대로 그 學級 指導教師의 指導能力을 가리키는 것으로 해석해서는 絶對로 안된다는 點이다. 한 學級의 成績은 教師의 指導能力以外의 數많은 外的 內的 條件에 의해서도 變化될 수 있는 것이다. iii) 規準表는 어디까지나 指導方法의 反省, 改善을 위한 資料로서만 使用되어야 하며, 指導能力의 評價를 위해서 사용하는 것은 絶對로 禁止해 두는 바이다.

第四節 檢查의 良好度

1. 檢查의 信賴度(Reliability)

한 考查나 檢查가 훌륭한 測定方法이 되기 위해서는 그 考查나 檢查가 실제로 測定하는結果가 信賴로운 것이어야 할 것이다. 만약에 同一한 檢查를 칠 때마다 同一한 學生의 點數가 다르게 나온다면 우리는 이 檢查의 結果를 믿을 수 없는 것이다. 아무리 엄밀하고 세밀한 測定에 있어서도 어느 程度의 測定의 誤差가 따르는 것을 全的으로 막아낼 도리는 없다.

우리가 만들어 사용하는 여러 非公式的 檢查, 標準化検査의 點數도 그건 작건 이런 誤差의 作用을 받고 있는 셈이다. 좋은 檢查란 이런 誤差의 作用을 最少限度로 줄여서 信賴로운 結果 解釋을 할 수 있는 檢查를 말한다.

標準化検査의 作成過程에 있어서는 이런 誤差의 作用을 最少限度로 줄이기 為한 여러가지 技術的인 配慮 措處가 주어지지만 이런 誤差의 作用을 完全히 排除할 수는 없기 때문에 檢查가 만들어진 후엔 그 檢查의 實際의 誤差界限를 明確하게 밝혀서 得點을 해석할 때는 이런 誤差界限에 留意하면서 하게 된다. 勿論 여기서 말하는 誤差란 것은 단순히 採點上의 錯誤만을 말하는 것이 아니고 檢查 實施上의 여러 條件의 變動이 끼치는 영향과 被檢者自身의 身體的 心理的 條件의 如何 때문에 발생되는 得點上의 變動까지도 다 包含하는 개념이다.

한 檢査의 信賴度는 信賴度係數에 의해서 表示되는데 이런 信賴度係數를 밝히는 方法에는 여러가지가 있다.

i) 檢査에서는 소위 奇偶半分法(split-half method)에 의해서 信賴度係數를 確定시켰는데 그 結果는 檢查總點에 있어 .93이다. 이 數値는 이 檢査의 信賴度가 극히 만족스럽다는 것을 나타내고 있다.

ii) 信賴度係數를 使用하여 이 檢査의 點數에 内包된 誤差點數의 크기를 짐작해 볼 수 있다. 우선 測定의 標準誤差(standard error of measurement)라는 것을 計算해 보면

$$SD_{meas} = SD \sqrt{1 - r_{tt}}$$

SD meas.....測定의 標準誤差

SDi) 檢査의 標準偏差

r_{tt}i) 檢査의 信賴度係數

$$SD_{meas} = 16.98 \sqrt{1 - .93}$$

$$= 4.40 i)$$

i) 測定의 標準誤差는 다음과 같이 解석할 수 있다. 即, 한 個人이 i) 檢査에서 60 點을 얻었다면, 誤差를 除去한 i) 사람의 정밀 點數(true score)가 $60 \pm 1 \times 4.4$ 即 55.6~64.4 사이에 있을 可能性이 約 68%이며, $60 \pm 2 \times 4.4$ 即 51.2~68.8에 있을 可能性이 約 95%이다. 即 한 個人이 60 點을 得點했을 경우에 i) 사람의 정밀點數가 51.2~68.8 사이에 있을 것이라고 말한다면 i) 말의 正確性은 약 95%이다. 이런 解석은 i) 測定의 標準

誤差와 소위 正常分布曲線(normal distribution curve) 間의 數字的 關係에 의해서 할 수 있는 것이다.

測定에 크건 작건 誤差가 따라다니는 限 그 得點을 융통성 있게 取扱해야 한다. 卽 한 學生이 45 點을 얻었다고 해서 이 사람의 實力이 正確하게 45 點程度라고 말할 수는 없고 앞에서와 같이 測定의 標準誤差라는 誤差界限에 留意해서 大體的인 解석을 해야 한다.

2. 檢查의 妥當度

信賴度는 어디까지나 “正確하게 測定하는 程度”만을 문제삼는 反面에, 妥當度는 測定하려고 하는 것을 充實하고 올바르게 測定하고 있느냐 하는 문제와 關聯된다. 어떻게 생각하면 檢查의 良好度란 문제에 있어서 妥當度는 信賴度보다도 훨씬 重要하고 深刻한 문제이다. 이 標準化英語學力檢查는 高等學校의 英語科 教育課程에서의 重要한 學習成果를 測定하는 檢查이다. 따라서 이 檢查는 어디까지나 이런 檢查目的에 充實한 문제들로서 檢查가 構成되어야 할 것이다.

檢査의 妥當度를 밝히는 方法에는 여러가지가 있으나 大別해서 論理的 方法과 經驗的方法으로 나눌 수 있다.

學力檢查의 論理的 妥當度는 흔히 教育課程에 指示된 教育目的과 檢查問項의 内容을 比較 對照함으로써 檢查한다. 가령 英語科 教育課程에서 英文讀解力を 重要한 目的으로 指示하고 있는데에도 실제의 英語學力檢查속에 그런 讀解力を 測定하는 問項이 缺乏해 있다면 이 檢查는 英語教育目的에 비추어 妥當하지 못한 셈이다.

經驗的 妥當度는 實證的인 基礎에서 어디까지나 實際的인 資料와 對照하여 檢證된다. 예컨대 한 英語檢查의 結果는 학생들의 平素의 英語實力を 잘 反映시키는 것이 되어야 할 것이다. 勿論 약간의 差異는 있을 수 있지만 平素의 成績과 지나치게 동떨어진 結果가 나온다면 이 檢查의 妥當度는 어딘가 잘못되어 있다는 것을 짐작할 수 있다. 또한 檢查의 結果는 將來의 成績傾向을 잘 豫言할 수 있어야 할 것이다. 現在의 檢查成績에는 個人의 能力, 學習方法, 學習動機들의 條件이 다反映되어 있을 것이므로 이를 條件에 어떤 劇的인 變化가 오지 않는 限 現在의 檢查成績을 基礎로 하여 將來의 成績을 얼마만큼 推定할 수 있어야 할 것이다. 現在의 檢查成績만을 알 수 있을뿐, 이 成績이 將次 어떻게 變動할 것인지 전혀 짐작할 수 없다면 이 檢查點數의 妥當度는 그만큼 낮은 셈이다.

이 標準化英語學力檢查의 經驗的 妥當度로서, 이 檢查 得點과 高校에서의 英語 成績間

의相關係數를 밝혀 보았다. 그結果는 다음과 같다.

高校英語成績과 本検査得點間의 相關度

	N	r
D 高校	54	.83
P 商高	58	.65
E 女高	29	.85

이結果로 보아서 高校 英語成績과 이 標準化検査의 得點間에는 大體로 상당히 높은 相關關係가 存在한다는 것을 알 수가 있다. 이러한 事實은 이 標準化検査의 妥當度가 良好한 것을 나타내 준다고 할 수 있다.

第五節 結 語

以上으로 本所에서 製作 實施한 檢査의 經過와 結果의 解釋 및 그 良好度를 詳說하였다. 우리가 보는 바로는 1965 年度의 서울大學校 入學試驗의 英語考査問題는 入試問題가 갖는 制限內에서 이러한 檢査가 示唆하는 경향이 많이 참작 반영되었다고 여겨진다. 이에 새삼스럽게 느껴지는 것은——特히 1965 年度의 其他大學入試問題를 概觀하고 난 뒤의 느낌은——이러한 標準化的 性格의 入試出題에의 導入의 必要性이 切感되는 點이다.

本所製作의 本検査가 비록 서울大學校 志願者를 그 直接對象으로 하였다는 制限된 性格을 지니고 있지만 本検査의 利用可能性에 簡單히 言及함으로서 이 報告를 끝맺으려 한다.

于先 本検査를 實施함으로서 教師는 學生個個人의 弱點을 파악할 수 있고, 高校上級班學生의 指導方法을 向上할 수 있고, 學校成績에 對한 客觀的이고 信賴性 있는 補充評價를 할 수 있고, 各 教師가 담당한 學級全體가 全國的인 規模로 보아 他校와는 어떻게 比較될 수 있는가를 알 수 있고, 學生들 각者로 하여금 學科에 對한 關心을 북돋아 주는 結果를 얻을 수 있다.

이러한 여러 點을 考慮하여 本研究所에서는 全國高校英語教師에 依한 本検査의 廣範한 利用을 可能케 하는 方法을 考慮中에 있다. 關心이 있는 教師諸位는 書信으로 問議하시기를 바라며 이 報告를 끝 맺는다.