The Japan Association of College English Teachers (JACET)

大学生のための標準英語聴解力テスト作成に関する最終報告

——JACET・COLTD 共同研究——

共同研究小委員会

この研究は次の委員が中心となって行なわれた： 坪井忠二 (COLTD)，天野一夫 (JACET)，阿部 達 (JACET)，伊藤嘉一 (JACET)，河野 護 (JACET)，国吉丈夫 (JACET)，松居 司 (JACET)，森戸由久 (JACET)，佐藤秀志 (JACET)，島岡 丘 (COLTD・JACET)，鈴木 博 (COLTD・JACET)，若林俊輔 (JACET)，吉田正治 (JACET) 計 13 名。

本研究に必要なコンピューター用のプログラム開発は次の諸氏によって行なわれた。前田功雄 (独協大学)，高柳敏子 (独協大学)，河野 護 (JACET)。

はじめに

本テストの開発は 1972 年春に始まり，約2年半かかってやっと一応の完成を見たものである。その間，3次にわたるテストを実施してテスト項目を精選したため，後で述べるように信頼度の高いテストとなり，会員諸氏に安心してご利用いただけるテストが準備できたものと信ずる。

開発の過程については JACET の 1972 年および 1973 年の大会で中間報告を行ない，JACET「紀要」第4号にその記録が載っているので，本報告ではその後の経過および標準化の手続きを中心に述べることにする。

1. 本テスト開発の方針

本テストは日本で英語を学習してきた日本の大学生を対象としている。したがって，英語を母国語とする国に移住したり，留学する時に与えるテストとは目的が異なっている。まず，文化的背景に関する知識はテストの中に入れないようにしてある。 dime や quarter がそれぞれ 10 セント，25 セントのコインであることは米国では誰でも知らなければならないが，日本の英語教育では

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

あまり取り扱わない。本テストでは dime や quarter を知らないからといって聴解力がないとはいえないという考え方をとっている。

また，スタイルとか声の表情なども，日本の英語教育の現状から考えてそれをテストの中にとり入れるのは時期尚早であるということで本テストには入れてない。本テストは現在行なわれている英語の教授・学習で身につくと思われるごく基本的な聴解力を試すテストとして開発されたのである。

その他の方針については既に「紀要」第4号 p. 22 に述べてある通りである。

次に，本テスト開発に当って考慮したことは，言語学的な立場から見た妥当性よりも学生の聴解力の有無を識別できる弁別力のあるテスト項目を集めることに重点を置くということである。前者を中心にすると問題を作るのが非常にむずかしくなり，いつまでたっても弁別力のあるテスト項目ができない領域が残ってしまう恐れがある。そこで，我々はまずテストを繰返し行なって弁別力のあるテスト項目を数多く集め，その中から言語学的および内容的に片寄りのないように配慮しながら選択するという方向をとることにした。

もう一つ考えたことは，聴解力を3つの異った方向から試すということである。後で示す例にあるように，Type A では常識で考えて正しいか誤りかをきく true or false の形式，Type B では2つの文をきかせて内容の異同をきく same or different の形式，Type C では passage の中から必要な情報を聞き出す力を見る形式とした。

内容の上からも3つの形式が必要であるように思われる。すなわち，Type A では事実に基づくことに関しての問題が短時間に沢山与えられる。価値観の入る内容，感情や物の状態の描写などに関する場合は Type B でこなすことができる。少しまとまった内容のある英語を聞く力を見るのには Type C が必要となるが，従来のように passage を聞いたあとで質問と答がある方式では passage が長くなれば長くなる程記憶の要素が大きな影響を与えるので，質問を passage の前に与える形式にしたのである。

2. 第3次テスト

1973 年大会で報告した時点で，弁別力ありと判定されたテスト項目は Type A が 24，Type B が 20，Type C が 26 あり，各 Type 共 20 ずつとすれば数の上では確かに充分であった。しかし，Type A については，Negative-False のテスト項目が少なく，Type B については内容的に片寄りが見られたので新たに項目を追加する必要が感じられた。追加した項目は前からあった項

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

目と一緒にして予備テストをして弁別力の高いものだけを残し，第3次テスト問題を作成した。その結果，テスト項目は Type A が 28, Type B が 26, Type C が 26 となった。経費と労力を省くため，この第3次テスト問題を使って，弁別力のあるテスト項目を最終的に決定することと，テストの標準化の資料を集めるという2つの目的を持ったテストを行なうことになつた。最終的には各タイプ共テスト項目を 20 にする計画であったので，第3次テストではそれぞれ 6~8 題テスト項目が多かったことになる。

実際のテストでは，Type A の問題を Part 1, Type B を Part 2, Type C を Part 3 としてある。吹込みは，米国中西部出身で現在日本の大学で英語および言語学を教えている男女各1名に依頼し，Part 1 は男性に，Part 2 は女性に，Part 3 は passage の部分を男性に，問と答の部分を女性に担当してもらった。

吹込みはプロ級のスタジオを使用し，ミキシングは放送局のミキサーに依頼した。録音テープはスコッチ 206 番を 38 cm/sec の速度で使用した。原テープに直接はさみを入れて編集し，それから直接ソニ SLH テープに 19 cm/sec で複写して第3次テストテープとした。複写に使用したテープレコーダーは TEAC R740 である。テープの最初に 1,000 Hz と 7,000 Hz の純音を入れておき，テストに使う再生装置で再生してその2つの音の大きさが極端に違って聞こえたら設備がテストに適さないことが分るようにした。

第3次テストはこのようにして準備され，協力を申し出て下さった 16 の大学で実施されて 1479 名が受験した。実施大学学部学科学年は後述の通りである。送り返されてきたアンケートの記入によると，全受験生の約75% が，LL でテストを受けたか，イヤホンをテープレコーダーにつないで聞いたようである。また，周囲からの騒音はほとんどないが，時にうるさいことがあった程度というのがほとんどである。したがって，まず大部分の受験者がかなり良い状況の下でテストを受けたと考えられる。

答案を採点し，各テスト項目を分析検討した結果，すべての項目が弁別力ありと判定されたが，その中でも弁別力が大きく，通過率の良いテスト項目だけを残すことにし，計画通り，各 part 共 20 題ずつ選んだ。テープに入っている説明と問題の吹込時間は約 35 分で，途中3分間の休みを取るのでテストの所要時間は 40 分弱となる。

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

3. 本テストの内容

本テストは全て聴くだけで行なわれ，文字はいっさい与えられない。テープの最初に，前述の通り再生装置をおおまかに点検するための音とその説明が入っている。そのあと，テストについての全般的な説明があり，以下 Part 1 の解説と問題，Part 2 の解説と問題が入っている。Side 1 がこれで終り，Part 3 の解説と問題は Side 2 に入っている。テープをかけかえるのを機会に 3 分間休みをとらせることにしてある。

以下，各 Part の解説の部分を転載してそのテストの形式や内容の程度を紹介したい。

Part 1

Part 1 には問題が 20 あります。その各々の内容が常識で考えて正しい場合は Part 1 の解答欄の「正」の字を丸で囲んでください。 正しくない時は「誤」の字を丸で囲んでください。各々の問題は 1 回ずつ言います。 そのあと 7 秒程ポーズがありますからその間に解答をしてください。

では例題をやってみましょう。

Listen to the following examples.

Example 1

There are seven days in a week, and there are twenty-four hours in a day.

1 週は 7 日，1 日は 24 時間と言っていますから正しいことを述べています。ですから解答用紙の「正」の字を丸で囲みます。もう一つ例題をやってみましょう。

Example 2

The sun rises in the east, but the moon rises in the west.

太陽は東から昇るが月は西から昇ると言っています。月も太陽と同じく東から昇りますから間違ったことを述べています。ですから解答用紙の「誤」の字を丸で囲みます。

Part 2

Part 2 には 20 組の文があります。その各々の組に (a) と (b) の 2 つ

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

の文があります。(a) の文の内容と (b) の文の内容がほぼ同じだと思ったらの解答欄の「同」の字を丸で囲んでください。違っていると思ったら「異」の字を丸で囲んでください。各々の文は1回ずつ言います。そのあと7秒程ポーズがありますからその間に解答してください。

では例題をやってみましょう。

Listen to the following examples.

Example 1

(a) Mary cooks better than any other member of her family.

(b) Mary cooks best in her family.

(a) の文は「Mary は家族の内の誰よりも料理が上手だ」と言っており，(b) の文は「Mary は家族の内で一番料理が上手だ」と言っています。この2つの文は内容がほぼ同じです。ですから「同」の字を丸で囲みます。

Example 2

(a) George is looking forward to his father.

(b) George is looking for his father.

(a) の文は「George はお父さんに会うのを楽しみに待っている」という内容ですが，(b) の文は「George はお父さんを探している」という内容です。この2つの文は違ったことを言っているわけですから「異」の字を丸で囲みます。

Part 3

Part 3 には問題が 20 あります。各々の問題の形式は次のようになっています。まず質問が聞こえます。例えば，

Do barbers in Japan usually work on Sundays?

この質問に答えるために次の文章を聞いてください。

We don't go to school on Sundays. Most people don't work on Sundays. But some people work even on Sundays. Those who work on Sundays usually have a day off on week-days. Usually, barbers in Japan work on Sundays and they don't work on Mondays.

ここで最初の質問をもう一度くり返します。

Do barbers in Japan usually work on Sundays?

この質問に対する答を次の a, b, c の 3 つの中から選んでください。

a. No, they don't.

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

b. They work on Mondays.

c. Yes, they do.

日本の床屋さんは普通日曜日に仕事をして月曜日に休むと言っていました。ですから正しい答はcです。解答欄のcを丸で囲むことになります。ではもう一度今の問題を通して聞いてください。（ここで英語の部分だけくり返す）

4. テストの実施

テストは集団で実施しても，個人で実施してもよい。テストは全てテープに録音されているので，テスト開始前の注意・説明は不用，所要時間 35 分，但し，始める前にテープ・レコーダーの音量・音質の点検をする。点検法は同じくテープの最初に録音されている。実施中の監督についても特に要する注意はない。明瞭かつ十分な音量で終始聞こえるよう配慮し，不正がなきよう気をつける。

5. テストの採点・結果の処理

採点は減点法により行なう。Type A, Type B は正答の数から誤答の数を引き2倍する。Type C は正答の数から誤答の数の2分の1を引き，その値を2倍する。その後，3つの下位テストの得点を合計して，テストの得点とする。満点は120点で，マイナス点もありうる。以上の採点法を式で示すと次のようになる。

Type A, Type B の採点方法

$$\text{subtest score} = (R - W) \times 2$$

Type C の採点方法

$$\text{subtest score} = \left(R - \frac{W}{2}\right) \times 2$$

R：正答の数　　W：誤答の数

特に，単純に正答数の合計を得点としないよう注意する。減点法によらないと，解答の選択肢が少いため，信頼度が小さくなってしまうからである。

なお，採点は各大学で行なってもよいが，実費を払うことにより，当本部がコンピューターで採点・処理をすることもできる。規定の OMR シートで実施し，答案を本部に郵送されると，大学全体の高点順リストと，クラス内高点

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

順リスト，及び，両者のヒストグラムを作成し，返送する予定。但し，常時このような処理は不可能で，実施するとしたら，年 1，2 回決まった期間に全国一斉に実施し，本部が一括採点，処理することになろう。

6. テストの標準化と現段階におけるテスト利用上の注意

テストの得点を正しく評定するための規準をそのテストが対象とする集団の中で設定していくことを標準化 (standardization) といい，このような規準によって作成されたテストを標準テスト (standardized test) という。従って，その規準となる集団の設定はテストの作成にあたって極めて大切な問題となる。我々の場合，大学生の英語の聞き取り力を測定する標準テストを作ることが目的であれば，日本の全大学生が対象になり，この母集団を正しく代表するような標本を抽出しなければ，大学生のための標準テストとはいえない。その意味において，今回のテスト，及び，その結果にもとづいて設定した諸統計的規準は，あくまでも協力してくれた大学の被験者を規準集団 (norm group) とし，この集団について標準化したのであって，大学生のための標準テストとは，現段階ではいえない。このテストをこれから使用される方は，ここに掲げられている統計的規準・尺度は上記集団内のものであることを銘記し，上記集団をいわば準拠集団 (reference group) と見なして，テストを実施する個人又は集団の結果を解釈していただきたい。換言すれば，実施して得られた結果をここにあるデータと比較し評定するということであって，得られた得点が，例えば，平均以上であるから日本の大学生の平均より上位である，といってはならない。あくまでも準拠集団の平均よりも良いということである。その意味から，今回協力してくれた大学については，現段階でも利用価値が高まるよう，被験者の専攻・学年・人数まで，支障のない限り詳しく公表することにした。このような学生の集団に較べて，自分の大学の学生はどの程度の力をもっているのかを比較評定していただくことになる。

今後の我々の課題は，標本抽出法を十分研究し，多数の大学の協力を得て，真の標準テストといえるような，日本の大学生のための尺度を作成することである。

7. テストの結果

7—1 各種平均点

各種平均点を表にまとめた。全被験者 1479 名の平均は 32.08 で標準偏差

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

協力大学一覧

大学名	学部	学科	学年	人数
国公立大学				
福島県立医科大	医	医	1	50
東京工業大	工	第6類	1	44
九州大	法	法	1	43
	法・経	(混合)	2	50
	工	建築	1	52
	工	応用原子核	1	50
筑波大	第1一学群	人文学群	1	37
		自然学系	1	38
	医学専門		1	30
千葉大	教育	英・文(混合)	1	15
	〃	英・文(混合)	2	18
	〃	英・文(混合)	3	31
宇都宮大	教育	英語	1	11
	〃	甲類国語	2	22
	〃	甲類社会	1	37
東京大	理科Ⅱ類		1	30
	文科Ⅰ類	Ⅱ類(混合)	2	46
徳島大	教育	小学校	1	30
	工	精密機械工	1	29
	〃	応用科学	1	29
	〃	情報工	1	20
	医	医	1	36
	〃	栄養	1	42
私立大学				
早稲田大	理工	機械工	1	74
	〃	応用物理	1	29
大谷女子大	文	英文	1	130
	〃	〃	2	73
東洋女子短大		英文	1	62
		〃	2	31
慶応大	経済	経済	1	30
	〃	〃	2	32
成城大	文芸	英文	2	31
目白学園女子短大		英語英文	2	74
福岡大	人文	英語	2	33
日本女子大	文	社会福祉	1	22
	家政	住居	1	25
	文	国文	2	44

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

表1　平均点一覧

大学	学年	英語・英文	文科系	理科系	合計
国公立	1年	53.57 (19)	26.60 (154)	38.08 (450)	35.94 (628)
	2年	92.56 (9)	44.48 (126)		47.48 (134)
	3年	72.60 (10)	64.48 (21)		68.00 (27)
	小計	67.81 (38)	36.73 (301)	38.08 (450)	38.99 (789)
私立	1年	10.42 (192)	31.76 (77)	35.79 (103)	21.86 (372)
	2年	23.85 (242)	36.46 (76)		26.86 (318)
	小計	17.91 (434)	34.10 (153)	35.79 (103)	24.17 (690)
合計		21.93 (472)	35.84 (454)	37.65 (553)	32.08(1,479)

( ) 内被験者数

は 28.96 であった。( )内は各セルに含まれる被験者の数である。

国公立大学の英語・英文，及び，文科系の3年のところは，数が少なすぎるため，英語・英文と文科系を一つにまとめるべきと考えたが，比較の対象として役立つと判断し，あえて別記した。空欄は被験者が得られなかったところである。合計の欄についてだけ説明を加えると，右上から，国公立1年・2年・3年の各平均，次が国公立全体の平均である。その下が私立1年・2年の平均と，

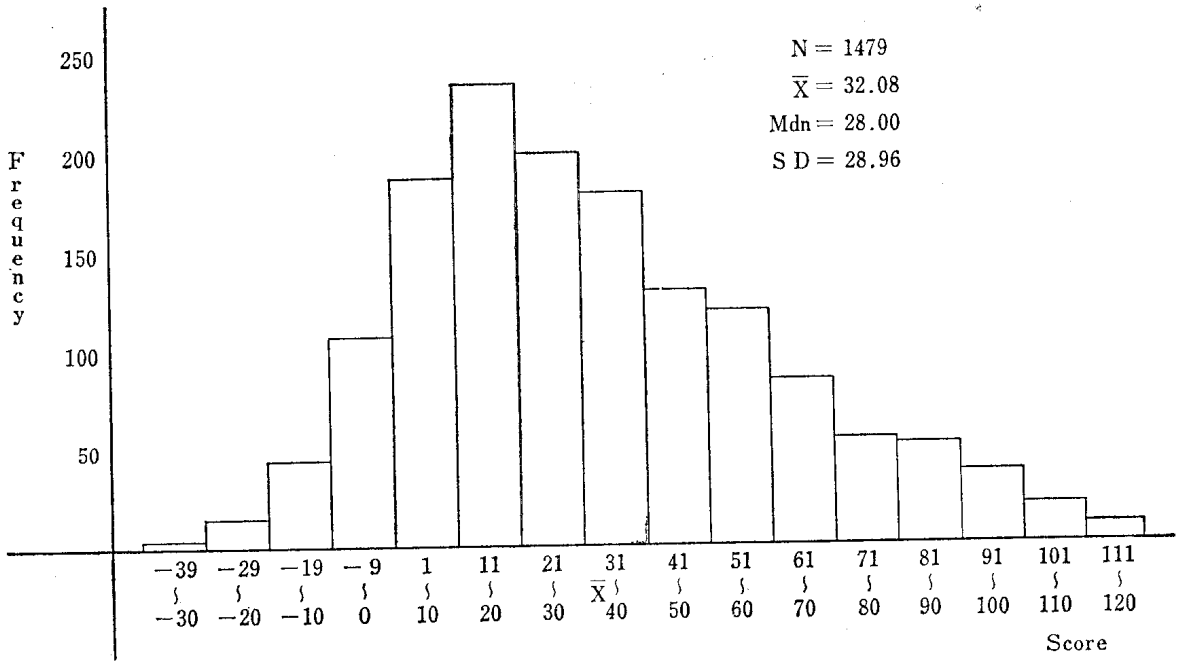


図1　得点分布

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

両者を加えた平均。そして，一番下が全体の平均である。最下列の合計については説明不要と思う。この表に含まれない平均点も算出してみた。その内，性別における平均点は男 39.82 (N=747)，女 24.17 (N=732) であった。

7—2 得点分布

全体の得点分布をヒストグラムで表わすと図1になる。

一見して正に歪曲していることがわかる。度数分布表からχ検定により正規性の検定を行なってみたところ，1% レベルで有意であった。従って，この分布は正規分布しているとはいえない。

国公立と私立における分布をコンピューターが打ち出したままのヒストグラムを二つ重ねて図2にまとめてみる。※や○は一つにつき3人を表わす。○印の方が私立大である。

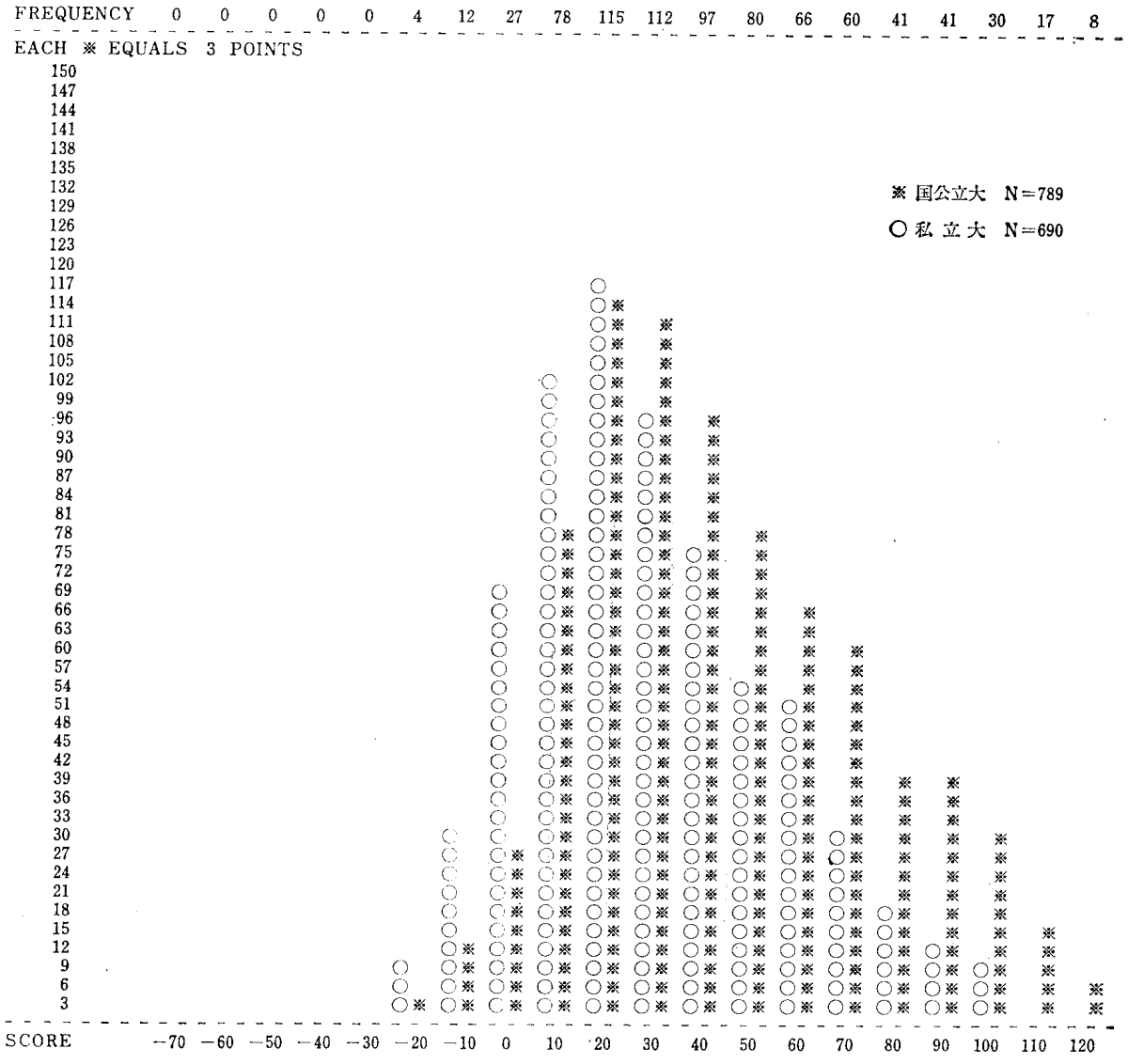


図2 国公立大・私立大別分布

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

各級間における度数は全て右寄せにしてある。即ち，例えば，71 点から 80 点の者は 80 のところに表わされている。※及び○一つは3人を表わすが3で割って余った数は除かれている。それ故国公立大の111点から120点までの実際の度数は7であるのにグラフでは※二つで表わされている。

8. テスト結果の解釈と評定

協力いただいた各大学には，(1) 大学内における高点順リストとそのヒストグラム (2) 各クラス内高点順リストとそのヒストグラムの 2 種の結果を返送した。その資料により，各被験者個人は (1) 下位テスト別の正答数・誤答数・得点，及び，合計得点，(2) クラス内，大学内，及び全被験者内の順位及び，この報告を見ることにより，(3) 全体におけるパーセンタイル順位，各種標準得点，各種評定等知ることができる。これから利用される方はこの報告を参照して，比較評定していただくことになる。

8—1 パーセンタイル順位と正規化得点

まず，自分の得点を表より見つけ，自分の全体におけるパーセンタイル順位を読みとる。この値は実測値による。先にも述べたように，得点は正規に分布しているとはいえないので，このまま SD により各評定区分を決めると，下位の評定が異常に多くなるため，及び，表3を使って簡単に各種標準得点や評定が得られるようにするため，その隣の欄に正規化した z 得点，及び，T 得点（日本では偏差値といわれる）を併記した。表2により z 又は T 得点を読み取って，表3の中よりその値のところを横に見ていけば，各種の評定がすぐに見つけられる。例えば，得点が 65 点の者は表により 84% タイルであることがわかる。正規化 z 得点は 1.0 であるから，表3の 1.0 のところを横に読み取ると，偏差 IQ は 115，パーセンタイルは 84（偶然正規分布の値と実測値とが一致している）であり，スタナイン尺度では 7，5 段階評定では 4 ということになる。なお，スタナインとは standard nine scale の略で，9 段階であるため，コンピューターのパンチカードに1行1穴で処理できるので便利である。

得点の正規性を調べるため，及び，正規化得点変換尺を作るため，次のようなグラフを作った。正規分布しているならば，点は直線になる。読み方は矢印で示してあるように，例えば，粗点 50 点は下の得点 50 のところから上に見て，直線に当ったところの左右を見れば，正規化 z 得点は 0.5，T 得点は 55

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

表 2 標準得点変換表

obt. score	% ile	z^*	T*	obt. score	% ile	z^*	T*
115 ～	100	2.9	79	41 ～ 43	67	0.2	52
112 ～ 114	100	2.8	78	39 ～ 40	64	0.1	51
110 ～ 111	100	2.7	77	36 ～ 38	61	0.0	50
107 ～ 109	99	2.6	76	33 ～ 35	57	–0.1	49
104 ～ 106	99	2.5	75	30 ～ 32	53	–0.2	48
101 ～ 103	98	2.4	74	28 ～ 29	51	–0.3	47
99 ～ 100	98	2.3	73	25 ～ 27	47	–0.4	46
96 ～ 98	97	2.2	72	22 ～ 24	42	–0.5	45
93 ～ 95	97	2.1	71	19 ～ 21	38	–0.6	44
91 ～ 92	96	2.0	70	17 ～ 18	35	–0.7	43
88 ～ 90	95	1.9	69	14 ～ 16	31	–0.8	42
85 ～ 87	94	1.8	68	11 ～ 13	25	–0.9	41
82 ～ 84	93	1.7	67	8 ～ 10	21	–1.0	40
80 ～ 81	92	1.6	66	6 ～ 7	18	–1.1	39
77 ～ 79	91	1.5	65	3 ～ 5	15	–1.2	38
74 ～ 76	90	1.4	64	0 ～ 2	11	–1.3	37
71 ～ 73	89	1.3	63	–3 ～ –1	9	–1.4	36
69 ～ 70	87	1.2	62	–5 ～ –4	7	–1.5	35
66 ～ 68	86	1.1	61	–8 ～ –6	5	–1.6	34
63 ～ 65	84	1.0	60	–11 ～ –9	4	–1.7	33
60 ～ 62	82	0.9	59	–13 ～ –12	3	–1.8	32
58 ～ 59	81	0.8	58	–16 ～ –14	2	–1.9	31
55 ～ 57	78	0.7	57	–19 ～ –17	1	–2.0	30
52 ～ 54	76	0.6	56	–22 ～ –20	0	–2.1	29
49 ～ 51	74	0.5	55	–24 ～ –23	0	–2.2	28
47 ～ 48	72	0.4	54	–27 ～ –25	0	–2.3	27
44 ～ 46	70	0.3	53	～ –28	0	–2.4	26

z^*, T* は正規化された標準得点

% ile は実測値におけるパーセンタイル

と知ることができる。その上の点は実測値で，実測値の分布においては z 得点は 0.65 であったことがわかる。このように読み取って，表3により各評定を下してもよい。

なお．図の中の一次式 $z=0.0365X-1.3399$ は直線の方程式で，粗点 X から z 得点への換算式である。その下の公式は z 得点から T 得点への換算式である。

8—2 相対評価と絶対評価

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

表3　各種標準得点と人数の割合

正規化された z 得点	T 得点	偏差IQ	それ以下の人の割合 パーセンタイル順位	スタナイン	五段階評定値	正規化された z 得点	T 得点	偏差IQ	それ以下の人の割合 パーセンタイル順位	スタナイン	五段階評定値
−3.0	20	55	0			0.0	50	100	50		
−2.9	21		0			0.1	51		54		
−2.8	22	58	0			0.2	52	103	58		
−2.7	23		0			0.3	53		62		
−2.6	24	61	0		1	0.4	54	106	66	6	
−2.5	25		1		(7%)	0.5	55		69	(17%)	
−2.4	26	64	1	1		0.6	56	109	73		
−2.3	27		1	(4%)		0.7	57		76		
−2.2	28	67	1			0.8	58	112	79		
−2.1	29		2			0.9	59		82		4
−2.0	30	70	2			1.0	60	115	84	7	(24%)
−1.9	31		3			1.1	61		86	(12%)	
−1.8	32	73	4			1.2	62	118	88		
−1.7	33		4			1.3	63		90		
−1.6	34	76	5	2		1.4	64	121	92		
−1.5	35		7	(7%)		1.5	65		93	8	
−1.4	36	79	8			1.6	66	124	95	(7%)	
−1.3	37		10			1.7	67		96		
−1.2	38	82	12		2	1.8	68	127	96		
−1.1	39		14	3	(24%)	1.9	69		97		
−1.0	40	85	16	(12%)		2.0	70	130	98		
−0.9	41		18			2.1	71		98	9	5
−0.8	42	88	21			2.2	72	133	99	(4%)	(7%)
−0.7	43		24			2.3	73		99		
−0.6	44	91	27	4		2.4	74	136	99		
−0.5	45		31	(17%)		2.5	75		99		
−0.4	46	94	34			2.6	76	139	99		
−0.3	47		38			2.7	77		100		
−0.2	48	97	42			2.8	78	142	100		
−0.1	49		46	5	3	2.9	79		100		
0.0	50	100	50	(20%)	(38%)	3.0	80	145	100		
				平　均		0	50	100	50強	5	3
				標準偏差		1	10	15	29弱	2弱	1

（　）内の%はその得点を得る人の割合．
表2と順序が逆になっていることに注意．

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

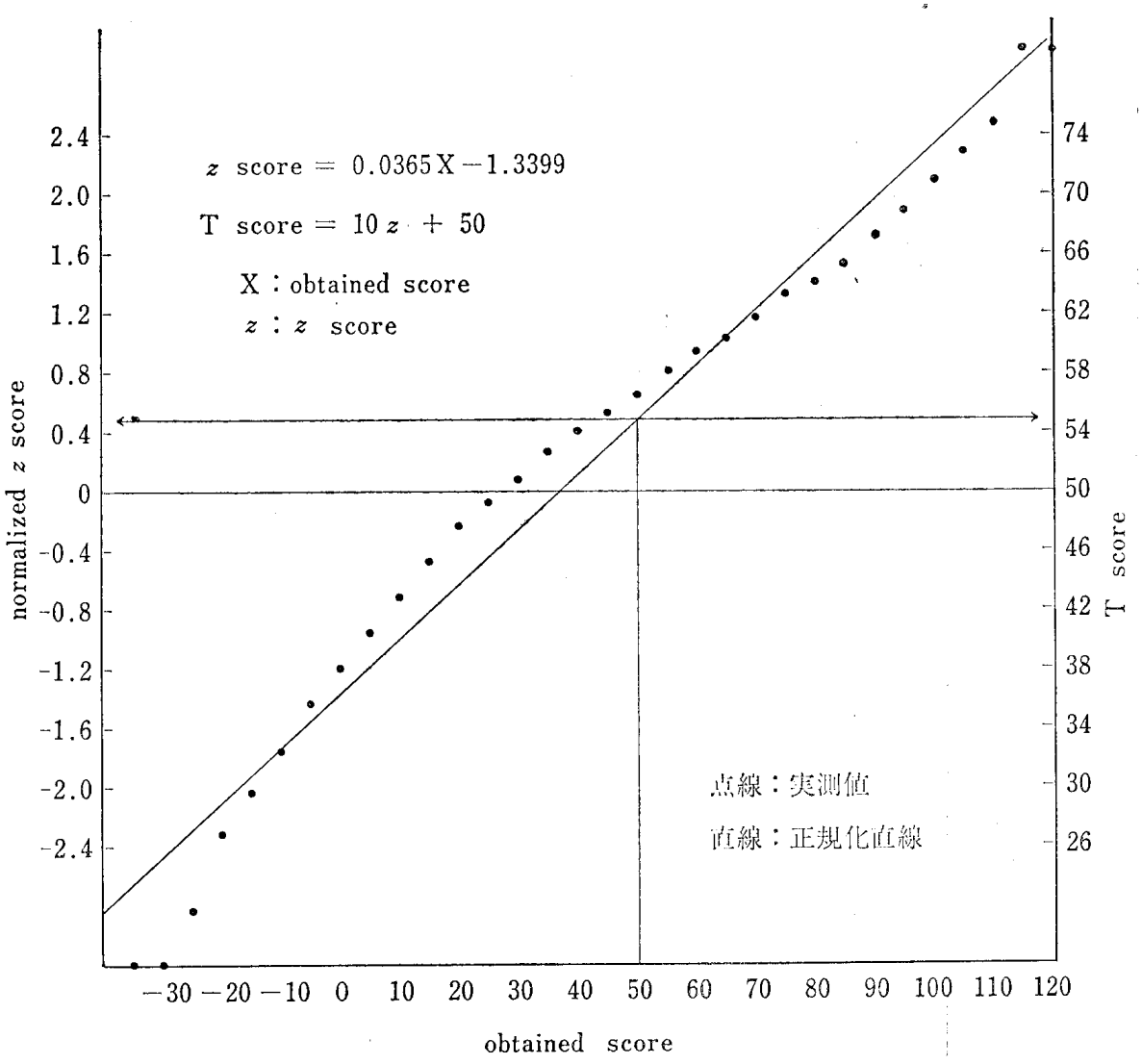


図3　正規化得点変換図

評価とは目標追求—評価—調整という単位での，目標追求活動における部分活動であって，追求活動の実績と目標の関係をチエックし，調整活動のために，フィード・バック情報を提供するものである。評定はそのようなサイクルの中で考えるのではなく，評定とは対象についての客観的判断であり，そこにおいては公正さが必要とされ，評定結果そのものが重要な目的になっている。

標準テストの結果の解釈のための尺度は全て規準集団（norm group）内における相対的位置を求めることによって，ある評定を下すのである。表2・表3に載せた諸尺度も，すべて今回受験した被験者集団内における相対的位置を得ることが目的であった。

一方，我々英語教育関係者は，日本における大学生の英語の聴解力に関する実態を知り，標準化して，学生の相対的位置を知るだけでは，教育者として，

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

表 4　20 点間隔における通過率と評定

単位 %

		120～101	100～81	80～61	60～41	40～21	20～1	0～-19
Type A	1	88	84	70	63	56	50	34
	2	100	86	75	61	42	30	20
選	3	100	86	67	66	58	47	44
択	4	79	72	54	42	47	34	32
肢	5	88	73	46	32	28	23	14
2	6	100	77	71	64	60	47	41
	7	100	100	88	89	76	63	56
	8	100	68	45	35	34	25	23
	9	96	81	62	42	37	30	24
	10	100	80	60	46	46	39	38
	11	88	70	64	49	40	37	41
	12	100	100	96	89	71	56	33
	13	100	95	84	80	62	48	41
	14	100	98	95	94	79	55	33
	15	92	91	78	74	57	44	43
	16	100	83	79	73	62	50	48
	17	100	81	65	50	45	42	34
	18	96	94	78	60	51	46	41
	19	100	98	90	81	67	51	44
	20	100	91	70	54	46	37	31
Type B	1	100	94	89	85	64	54	48
	2	92	78	54	50	38	35	34
選	3	96	65	58	41	33	27	21
択	4	100	95	90	83	69	55	35
肢	5	100	94	91	86	73	57	46
2	6	83	86	79	69	54	41	29
	7	100	91	87	78	65	56	49
	8	96	76	69	54	43	37	22
	9	96	76	65	56	37	33	23
	10	100	91	86	85	75	62	52
	11	96	78	64	52	50	37	31
	12	100	70	60	59	53	41	34
	13	100	94	86	65	50	36	30
	14	100	100	96	91	77	58	51
	15	70	65	59	58	53	40	37
	16	100	94	86	84	78	69	56
	17	100	89	69	56	43	35	43
	18	96	66	59	61	60	52	29
	19	100	89	74	64	53	40	29
	20	92	72	74	57	53	47	46

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

		120～101	100～81	80～61	60～41	40～21	20～1	0～−19
Type C	1	88	80	49	38	18	20	14
	2	92	82	76	52	38	27	21
選	3	100	88	88	68	54	37	31
択	4	75	59	36	30	21	19	11
肢	5	100	75	64	52	40	32	31
3	6	100	94	87	84	79	72	60
	7	100	94	81	75	68	60	49
	8	96	98	82	73	55	42	36
	9	100	84	70	42	29	22	16
	10	100	80	74	56	40	29	21
	11	100	95	83	58	48	30	32
	12	100	85	70	44	29	22	20
	13	100	75	64	50	36	28	17
	14	100	88	76	56	46	37	32
	15	100	97	93	77	58	38	29
	16	100	99	93	75	60	45	41
	17	100	94	93	79	58	36	19
	18	100	93	86	70	51	36	26
	19	92	92	83	67	49	38	27
	20	92	70	45	39	32	26	26
人数		25 (2%)	89 (6%)	139 (9%)	249 (17%)	378 (26%)	421 (29%)	161 (11%)
60問中における70%以上の問数		60	55	38	21	8	他−20 点以下 17 人 1	0
評定		Excellent		Good	Fair		Poor	

又被教育者としても，不十分と考える。ある教育目標を設定し，それへの到達度を何んらかの形で知り，その評定結果を評価のために，換言すれば，フィードバック機能のために，客観的データを求めたいと考えるのが常である。筆者はそこで次のような簡単な分析を試み，到達度を評定するための尺度を設定してみた。

被験者すべてを得点によって，120 点満点から下に 20 点間隔で区分し，その区分内における通過率を求めた。その結果が表 4 である。

ところで，項目の作成基準はこの報告の前半，及び，JACET「紀要」第4号にあるとおり，全て常識で容易に判断できる内容である。英語としても在日アメリカン・ハイスクールの学生が自然に受け入れ，簡単に正解を選んだ項目ばかりであるから，満点の者こそ十分に聞き取る力がある者ということにな

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

る。しかし，多少の誤りは不注意からもありうるし，多少の誤り程度の力であれば，実際の言語活動は冗長度があるもので，聞き返すなりすれば，相手の言うことを理解できると考えられる。そこで，操作的に 70% 以上の通過率であれば合格とした。すると，101 点から 120 点までの間の者は，60 問全てにおいてその値を越えていることがわかった。このレベルの学生であれば，このテストの内容程度の英語は殆んど理解できるし，飛躍しすぎる推論になろうが，実際の英語による生活において何んとかやっていけるレベルに達していると考えてもよいのではないか。そう考えて，この範囲を合格としたい。勿論，これは今後の研究によらねばならぬことではある。一応，このレベルを大学における英語教育の到達目標と設定することを提案する。

逆に，下位の方を検討してみよう。このテストで 20 点以下の者は，70% 以上に達している項目は一つしかなく，マイナス点を取った者については全くない。又，33 ページの得点分布を見ると，マイナス得点を取った者が相当数いることがわかる。これらの学生たりとも意図的にマイナス点を取ろうとしたとは考えられないので，これは偶然にこのような結果になった測定の誤差と考えれば，プラス方向にもこの部分と同数の誤差が働らいていると考えられるので，マイナス点を取った者178 名，及び，同数の 30 点台までの者は全く理解できなかった者と考えてよい。この理由，及び，操作的 70% 基準からみて，20 点以下は“殆んど理解できない”範囲と定める。

中間の評定については，60 問中越えている項目が約 70% 程度以上であるところは 61 点以上であるから，61 点から 100 点まで (100 点はその上の合格の範囲に入れてもよい) を“相当理解できる”範囲とする。残った 21 点から 60 点までは 8 から 21 くらいである。“多少理解できる”範囲としたらよい。

以上，整理すると，表 4 の下に記してある範囲の限界点を修正して，100 点以上を“Excellent”又は“合格”，60～99 点までは“Good”又は“相当理解できる”，20～59 点までは“Fair”又は“多少理解できる”，19 点以下は“Poor”又は“殆んど理解できない”と評定の尺度を決めたらよい。この評定区分に従って今回の被験者を評定すると，1479 名中，2% 弱が“Excellent”で，“Good”は 15%，“Fair” 43%，“Poor”は 40% の比率になる。この尺度を表にすると表 5 になる。

これで，8—1 により相対的位置を知り，8—2 によって絶対的尺度で自己の到達レベルを知ることができるようになった。

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

表5 絶 対 尺 度

Pass	Non-Pass		
	Good	Fair	Poor
120～100	99～61	60～21	20～

8—3 解釈にあたっての注意

(このテストが現段階では真の意味の標準テストとしての機能は もっていないことは前述したので，ここでは省略する。) よく教師は，又，学生も，得られた得点が真に学生の力を測定しているのか，又，他の専攻学生や他の大学の結果と較べて，どちらができるか知りたがるものであるが，前者についてはテストの分析のところで述べることにして，後者については，ここで特に述べておく必要があると考える。

このテストは120点からマイナス数 10 点まで，広い範囲に得点は分布し，標準偏差も 28.96 と大変大きい。従って，簡単に2つの集団の平均点を比較して，一方が他方よる高いからより優れていると考えないでほしい。集団に含まれる被験者の数によって異なるが，30 人くらいの グループでは，平均点において 10 点くらい差があっても，有意な差があるとはいえない。普通，平均の差の有意性の検定では t 検定や F 検定を使用して調べる。しかし，その場合母集団の分布が正規でないと，そのままでは適用できない。比較する必要がある場合は，表2により粗点を T 得点などの標準得点に変換し，その後平均と SD を算出して検定すべきである。その他，分布の型にかかわらず検定ができる，ノンパラメトリック法の U 検定によって比較してもよい。

9. テストの分析

9—1 妥 当 性

内容的妥当性： 内容的妥当性はテストの 内容が，測りたいと 考える場面や素材の全体から抽出された標本とみなすことができるかどうかで決まる。このテストも一つの学力テストであるから，学習すべき内容をかたよりなく適切に網羅していなければ，内容的妥当性があるとはいえない。しかし，英語の聴解力のテストでは，他の教科（特に高校以下の教科）のように教材の体系があるわけでもなく，抽出する母集団としての教材がなかったので，数回にわたる討議の後，JACET「紀要」第4号 p. 22 にあるような項目作成にあたっての基

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

本方針を決めた。その後，文の表現形式については，affirmative true, affirmative false, negative true, negative false のそれぞれの数を同数にするなど，妥当性を高めるよう考慮した。

何故に Type A (True or false), Type B の (Same or different), Type C (Short passage) の三つの形式にしたかの理由については，この報告の前半に述べてある。ここでは結果より数量的に妥当であるかを検討する。

まず，各下位テスト間の相関については，正答数，誤答数，及び，減点法による得点の各四つずつ，計 12 の間の相関行列を算出した。

表 6　下位テスト間相関行列

		No. of Correct Ans.				No. of Wrong Ans.				Score			
		Type A	Type B	Type C	Total	Type A	Type B	Type C	Total	Type A	Type B	Type C	Total
C・A	Type A												
	Type B	0.62											
	Type C	0.64	0.64										
	Total	0.86	0.86	0.89									
W・A	Type A	−0.40	−0.16	−0.30	−0.33								
	Type B	−0.16	−0.53	−0.30	−0.37	0.51							
	Type C	−0.24	−0.26	−0.60	−0.43	0.56	0.58						
	Total	−0.31	−0.37	−0.50	−0.46	0.81	0.82	0.88					
S	Type A	0.88	0.49	0.58	0.75	−0.79	−0.37	−0.46	−0.63				
	Type B	0.46	0.89	0.55	0.72	−0.37	−0.85	−0.47	−0.66	0.50			
	Type C	0.57	0.96	0.96	0.82	−0.42	−0.43	−0.80	−0.68	0.60	0.58		
	Total	0.75	0.83	0.83	0.91	−0.62	−0.66	−0.68	−0.78	0.83	0.83	0.86	

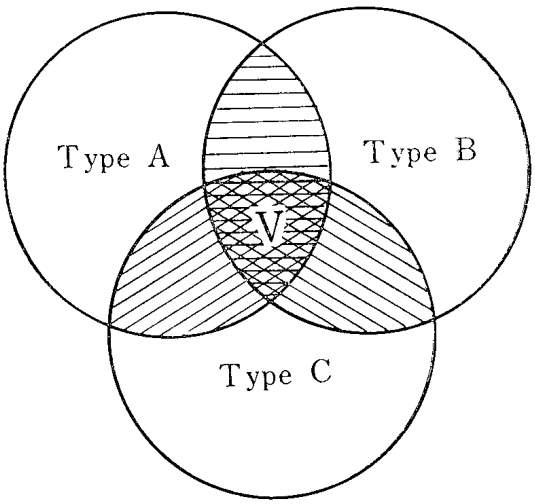
表 6 のうち，左上段の 枠内に入る四つの 相関係数は，減点しない場合の各下位テスト間の相関で，，その内 3 行目の三つの値は減点しない場合の合計得点との相関である。これは合計得点への寄与率と考えてもよい。右下段の値はこのテストの採点法で減点した場合の三つの 下 位テスト間 の 相 関 で あ る。0.50, 0.60, 0.58 といずれも高い相関とはいえず，このことは三つの下位テストが，それぞれ同程度の相関を保ちつつ，異った角度より，聞き取りの力を測定していることを意味し，問題として好ましいことを実証する。

この三つの変数による共通因子分散を図示すると図 4 のように表わせる。

Type A と Type B においては，25% は二つとも同じ側面を測定し，75%

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)



$r_{AB}^2 = 0.25$

$r_{AC}^2 = 0.36$

$r_{BC}^2 = 0.34$

V(A∩B∩C), or Vco, represents whatever the three variables or measures share in common.

図 4　共 通 因 子 分 散

は異ったものを，夫々測っている。同様に Type A と Type C では共通因子は 36% で，他は夫々異ったものを測定していると解釈する。そして三つの円に含まれる部分 V は三つの変数の共通因子分散といい，三つの下位テストが共通に測定している領域であることを意味する。

表 6 については更にいろいろなことがいえるが，紙面の関係で割愛する。

項目分析：　各ひとつひとつの項目の妥当性については，上下 27% による GP 分析の結果と，全被験者における通過率を表 7 にまとめた。

通過率については前述したように現在の大学生の実態に合せて易しくするようには配慮せず，弁別力の方を重視した。 χ^2 検定の結果はいずれも，1% レベルの値を大きく上回っている。但し，χ^2 の大きいことは必ずしも弁別力の大きいことを示さず，N が大きいと，値は非常に大きくなるので，ϕ 係数 (phi coefficient) も算出した。この値は大きければ大きいほど項目の弁別力の強いことを示す。いずれも 1% で有意な値であった。結論として，いずれの項目も，通過率の点で難かしすぎるものもあるが，弁別力は十分あり，妥当な問題であるといえる。

併存的妥当性：　併存的妥当性は別の方法で得られた測定や判定の結果とテストの結果がどの程度よく一致するかによって決定される。現在まで三つの妥当性係数を求めてみた。第 1 回は第 3 次予備調査段階でテスト項目の妥当性を調べるために，1973 年夏期特別プロジェクト (JACET『紀要』第 4 号，p.26 参照) の時に，H.D. Harris や L.A. Palmer が開発した A Comprehensive

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

表7　通過率 χ^2 ϕ係数

Type A				Type B				Type C			
No.	P	χ^2	ϕ	No.	P	χ^2	ϕ	No.	P	χ^2	ϕ
1	57.6	59.4	0.270	1	67.3	162.5	0.446	1	28.8	136.5	0.408
2	45.3	221.8	0.521	2	43.1	63.4	0.278	2	42.2	184.0	0.474
3	57.6	68.0	0.288	3	36.5	83.8	0.320	3	54.6	204.9	0.500
4	43.3	42.4	0.228	4	67.0	200.3	0.495	4	25.2	66.6	0.285
5	30.8	95.8	0.342	5	70.4	153.3	0.433	5	43.5	108.4	0.364
6	57.3	66.4	0.285	6	54.2	153.1	0.433	6	77.2	51.9	0.252
7	74.8	119.4	0.382	7	66.5	108.7	0.365	7	67.9	82.6	0.318
8	34.2	65.6	0.283	8	45.9	104.1	0.357	8	57.3	181.4	0.471
9	40.0	106.4	0.361	9	43.2	128.6	0.397	9	35.8	166.2	0.451
10	47.1	45.9	0.237	10	71.9	112.3	0.371	10	43.3	193.9	0.487
11	45.3	52.8	0.254	11	47.9	81.2	0.315	11	49.4	208.2	0.505
12	69.3	255.5	0.559	12	50.4	71.5	0.296	12	36.6	192.5	0.485
13	63.1	159.7	0.442	13	52.5	189.0	0.481	13	39.8	142.9	0.418
14	71.4	273.0	0.578	14	73.7	210.2	0.510	14	49.7	123.2	0.388
15	58.8	127.0	0.394	15	49.4	51.8	0.252	15	57.9	279.5	0.585
16	61.9	87.8	0.328	16	74.9	88.9	0.330	16	62.0	206.9	0.503
17	48.4	72.9	0.299	17	48.5	94.5	0.340	17	56.6	308.8	0.614
18	55.6	80.8	0.314	18	54.6	58.3	0.267	18	53.3	254.7	0.558
19	66.5	191.4	0.484	19	52.9	151.4	0.430	19	52.4	196.5	0.490
20	48.5	110.5	0.367	20	54.2	49.6	0.246	20	35.2	52.1	0.252

English Language Test for Speakers of English as a Second Language の test battery のうち，Listening Test, Form L-A（以下，CELT と略す）を同時に実施して一致度を調べた。その後の項目分析の結果，有効と決定した 174 問中 70 項目の合計点と CELT の値との妥当性係数は被験者数が 30 と小さいにもかかわらず，0.79 あった。二つ目は 1974 年度ある大学のヒアリングを中心とする授業において行なわれた 4 回の試験の合計点（250 点満点）とこのテストとの妥当性係数を求めてみた。値は 0.604 であった。最後に，同じく 1974 年夏 COLTD 主催の特別講習 ITC の参加者 49 名に CELT と同時に実施した。この妥当性係数は 0.701 であつた。以上，三つのいずれの値も，このテストが妥当であることを証明してくれるものである。

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

9—2 信 頼 性

信頼度係数については四つ算出してみた。表 8 にある。

表 8 信 頼 度 係 数

	非減点法	減 点 法
K-R 公式 20	0.864	0.983
K-R 公式 21	0.699	0.963

減点法を採用しない場合と，減点法による場合の，二つの採点法につき，夫々 Kuder-Richardson の第 20 公式と第 21 公式の 2 公式を使い計算した。両者の公式の差異は前者がよく使用されるもので大きい値が出るが，後者は項目の通過率が同一と仮定した場合に使われるもので，項目の通過率が異なるときにこの公式を用いると前者より小さい値が得られる。普通，信頼度係数はこの値よりは小さくならないと考える下限値として受け入れられている。このことを念頭に置き，表8の結果を検討すると減点しない場合は 0.864，最低 0.699 まで落ちることが考えられるのに対し，減点した場合の係数はどちらにおいても非常に高い。この採点法による差は，項目の数と信頼度係数との関係は $k=R(1-r)/r(1-R)$ であるから，約 9 倍の テストを実施したのと同じ効果をもつことになる。 非常に高い信頼度係数が出てきたが， これは得点の幅が 120 点からマイナス数 10 点までと広く，標準偏差も大きいため，力を長い尺度で測っていることによるもので，不思議ではない。以上の結果より，信頼性についても，非常に高く，安定したテストであるといえる。

測定の誤差： 測定の標準誤差は小さい方の信頼度係数 0.963 を用いた場合，$S_e=4.36$ になる。従って，その人の真の得点の合理的限界は $X_i \pm 2S_e$，即ち，$X_i \pm 8.7$ となる。換言すれば，得られた得点の ± 8.7 の範囲に真の値はあると考えられる。

10. 今 後 の 課 題

2 年半の作成期間を経て，ここにようやく利用可能なテストが完成した。今まで何回にもわたって協力下さった多数の大学の先生方や，貴重な資料を提供するため受験してくれた学生諸君，更に，さまざまな形で協力し，声援下さった方々に，心からの感謝の意を表わすとともに，現段階までこられたことを皆で喜び合いたい。共同研究小委員会のメンバーが，全く無償で貴重な時間とエネ

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

ルギーをこのテストのために捧げえたのも，上記の方々の協力と声援によるものであった。このテストが，このように大学の英語教育に携わっている者の日頃の教育・研究活動の中から生まれてきたことは，大変意義深いものがあると考える。

新生児ともいえるこのテストは，暖かく迎え入れられることによってはじめて世に出られるものである。各大学で利用されることによって，この新生児は澄んだ瞳で学生を見つめ，偽わることなく正直に，その瞳に映った学生の姿を語ってくれるであろう。一方，この新生児たりとも，そのように暖かく受け入れられると，いつまでもその澄んだ瞳を保ち，育ててくれる方々により一層信頼されるよう，自らその姿を調整していくことであろう。生みの親である委員会は，そのような成長を願ってやまないのである。

我々に残された仕事は，本文中にも述べてあるが，表1の各セルを妥当に代表するような大学を選択して，このテストを実施し，その結果をもって表1の値を修正することである。そして，日本の大学生のための真の標準テストとしての諸資料を，ここに提出したのと同様な形式で，再び作り代えることである。幸い，今回のデータ処理にあたっては，そのことを予期し，そのために使えるようにコンピューター・プログラムを作ってある。多数の大学の協力と予算が許されれば，実現することであろう。

参 考 文 献

Clark, J.D., *Foreign Language Testing: Theory and Practice.* 1972. Center for Curriculum Development.

Educational Testing Service, Multiple—Choice Questions: A close book. 1963. Educational Testing Service.

Fisher, R.A., *Statistical Methods for Research Workers.* 13th ed. 1958. 遠藤健児・鍋谷清治共訳. 森北出版. 1971.

Guilford, J.P., *Fundamental Statistics in Psychology and Education.* 4th ed. 1965. McGraw-Hill.

肥田野　直編,『テストⅠ』心理学研究法 7, 東京大学出版. 1972.

肥田野　直他,『心理・数育統計法』培風館. 1961.

池田　央『テストⅡ』心理学研究法 8, 東京大学出版. 1973.

岩原信九郎『新教育統計法』日本文化科学社. 1960.

Karmel, L.J., *Measurement and Evaluation in the Schools.* 1970. Macmillan.

Kerlinger, F.N., *Foundations of Behavioral Research.* 1973. Holt, Rinehart & Winston.

NII-Electronic Library Service

The Japan Association of College English Teachers (JACET)

Lado, R.L., *Language Testing: The Construction and Use of Foreign Language Test.* 1967. Lowe & Brydone.

Marascuilo, L.A., *Statistical Methods for Behavioral Science Research.* 1971. McGraw-Hill.

Mouly, G.J., *The Science of Educational Research.* 2nd ed. 1970. Litton Educational Publishing.

日本標準テスト研究会,『教育評価』現代教育研究会 13, 日本標準テスト研究会. 1968.

Standards for Educational and Psychological Tests and Manuals. 1966. American Psychological Association.

続　有恒,『教育評価』教育学叢書 21, 第一法規出版. 1971.

Wood, D.A., *Test Construction.* Charles E. Merrill Publishing. 1960. 池田　央訳, 日本文化科学社. 1970.

付記　本報告の執筆には，鈴木委員が第 1～3 節を，河野委員が第 4 節以降を担当した。

NII-Electronic Library Service