東京外国語大学
留学生日本語教育センター論集　21：29～48，1995

学部留学生の日本語能力試験開発のための基礎研究（１）

横田　淳子・伊東　祐郎・西郡　仁朗

（1994. 11. 1 受）

1　研究の意義と目的

文部省国費学部留学生の予備教育は、大阪外国語大学留学生日本語教育センターが学部留学生の予備教育に関わるまで、すべて東京外国語大学附属日本語学校（現在、留学生日本語教育センター、以下「センター」）で行ってきた。そのため、学内で学期毎に行う試験が、それぞれ初級、中級、上級の日本語レベルの試験になり、１年間の予備教育の最後に行う３学期最後の試験がそのまま予備教育修了試験となっていた。しかし、現在では、センターとほぼ同数の留学生が文部省の国費留学生として大阪外国語大学留学生日本語教育センターで教育を受けている。また、マレーシア、インドネシア、タイ政府派遣の学部留学生の数も増えており、文部省が受け入れている学部留学生だけを見ても、多様化が進んでいると言える。

私費の留学生に対しては「日本語能力試験１級」合格（程度）が大学入学のための必要条件になりつつあるが、「日本語能力試験」はもともと留学生だけを対象に作成されているものではなく、もっと広い範囲での日本語学習者を対象とした試験である。現在は留学生を対象とした統一試験（留学生に対する英語能力試験であるアメリカのＴＯＥＦＬのようなもの）がないために、「日本語能力試験１級」を流用しているのが現状である。国費の学部留学生はこの「日本語能力試験１級」を受験する必要はない。そのかわりに、各予備教育機関がそれぞれに修了試験を課し、大学に入学するのに必要な日本語能力を有しているかどうかを測定している。

上記のように、留学生の種類は、同じ学部留学生であっても、国費留学生／私費留学生、文部省奨学金留学生／外国政府派遣留学生といくつかに分かれ、そこで行われる予備教育も修了試験もそれぞれ異なる。しかし、これらの留学生は、予備教育修了後、学部に入学し、日本人の学生とともに勉学していくという点では同じである。このような現状を考えると、大学学部入学に最低限必要な日本語を有しているかどうかを測定する共通試験の開発が早急に望まれる。

学部入学予定留学生のための日本語能力試験の開発のためには、まず、学部留学生にとっての基礎日本語能力とはどのようなものであるかを明らかにする必要がある。この研究の目的は、基礎日本語能力を明らかにするために、まず、文部省学部留学生の予備教育機関として20余年の実績を持つセンターの学部留学生の試験を分析し、基礎日本語能力試験作成のための予備資料を提供することである。試験結果を分析し、研究を積み重ねることによって、基礎日本語能力試験を開発し、さらにはＴＯＥＦＬに匹敵するような学部予備教育修了試験の開発をもめざしたいと考えている。

2　研究方法

今年度は、センターの予備教育における日本語教育の中で1993年12月に学内で実施した２学期期末試験の分析を行った。２学期期末試験はセンターに４月に入学した留学生(1993年度は51人)が８か月（600時間)ほど学習した段階で受ける試験である。日本語のレベルは中級終了程度で、２学期に行った授業に基づく到達度テストである。1993年度には、日本語上級レベルにあたる３学期は、上級の共通の教科書のほかに留学生の専門分野によって理系と文系に分け、分野別に教材を選び、それぞれの分野で主に使われる文型や語彙等を教育する方法をとった。したがって、共通の基礎日本語力としては２学期期末試験が測定しているものを分析することが適当であると判断し、２学期期末試験を分析対象とした。

２学期末に行う日本語の試験は、文法、読解、漢字、聴解、作文、話し方の６種類であるが、そのうち、文法、読解、聴解を今回は分析の対象とした。作文、話し方は主観的なテストであり、漢字は再生記入式テストで、採点方法などにいろいろ考慮すべき点が多く、能力試験的なものになりにくいと判断したためである。

２学期期末試験のそれぞれの結果をコンピュータに入力し、各試験問題の項目分析を行い、試験問題一つ一つを検討した。試験全体に関してはこの試験が本来到達度テストであり、その意味で統計的な処理では分析しにくいという問題があった。

センターの同じ留学生に２学期期末試験とほぼ同時期に「1989年度日本語能力試験２級」のうち、文法・読解と聴解の２種類の試験を実施した。「日本語能力試験」では２級が日本語学習 600時間程度とレベルを設定しているためである。また、1989年度の試験を用いたのは、1990年からは「日本語能力試験」が公開され

ていて、学生が既に練習している可能性があるためである。2学期期末試験と「日本語能力試験2級」の相関関係を調べた。

次に本センターで使用されている教科書『中級日本語』に基づいて文法項目をリストアップし、『日本語能力試験出題基準』(以下、『出題基準』)と照合し、妥当性を検討した。『出題基準』は、読解、聴解に関しては技能的なものの基準のみを示したものであり、照合することは意味がなかった。文法については［文法的な＜機能語＞の類］が2級のサンプルとして提示されている。それらの［文法的な＜機能語＞の類］とセンターの教科書で扱われている機能語がどのくらい重なるかを検討した。その結果、2級の『出題基準』の約8割はセンターの文法項目のリストに含まれ、カバーされていることがわかった。

3　試験の分析の方法

3－1　プログラム

国際基督教大学教養学部教育学科石本菅生教授作成のテスト分析用プログラムパッケージを使用し、2学期期末試験文法、2学期期末試験読解、2学期期末試験聴解の3種類の試験に関して以下のデータを得た。

個人得点の採点

統計量の計算（平均、標準偏差、最高点、最低点、得点分布）

各項目の（成績上位群、成績下位群、全体別）通過率と弁別指数

各項目の選択肢別回答数（率）

成績グループ別各項目選択肢別回答数（率）

次に、ＳＰＳＳを使って、2学期期末試験文法と「日本語能力試験1989年2級読解・文法」、2学期期末試験読解と「日本語能力試験1989年2級読解・文法」、2学期期末試験聴解と「日本語能力試験1989年2級聴解」の各相関関係を調べた。

3－2　手順

1. 全学生に6桁の番号を与え、学生名簿を作成する。
2. 試験問題に通し番号をつけ、各項目ごとに正解を記入する。
3. 自由記入の項目に関しては、全学生の答案を調べ、誤答をすべて書き出し、誤答の中から解答の多いものにｂ，ｃ，ｄの記号を与え、その他の解答はすべてまとめてｅとする。正解にはａの記号を与える。
4. 正解の記号をインプットする。

5. 学生の解答を記号でインプットする。

6. 上記テスト分析用プログラムを起動させ、処理する。

7. 学生の各試験の得点をインプットし、ＳＰＳＳを作動させる。

4　分析結果

4－1　２学期期末試験文法（100点、70分、70項目）

a　試験の内容

問題Ⅰは８項目で、文の中の空所に助詞を自由記入させるものである。書き言葉で多く使われる表現の中の助詞である。ひらがなの数の指定は下線で示されている。ひらがな２文字が入るものは両方とも正しいときのみを正解とし、片方が正しくても点は与えられない。

問題Ⅱは８項目で、文の中の空所に、与えられている動詞を適当な形に変えて入れさせるものである。「べきだ」「かねない」などの文末表現と動詞のつながり方を問うものである。

問題ⅢはＡ、Ｂ、Ｃ、Ｄの４つの部分に分かれ、Ａ、Ｂ、Ｃの項目は各々６項目、Ｄは７項目ある。文の中の空所に与えられた語群の中から一つの語を選び、記入させている。各語群には８つの語が提示されているが、各語は１回しか使用できないと断ってある。語は形式名詞やそれに準じるもの、格助詞相当句、ムードを表す文末表現などである。

問題Ⅳは５項目で、文中に入る適切な語句を四肢から選択する形になっている。単文の内容とそれに合った文中に入る語の語彙的知識を問う。

問題ⅤはＡ、Ｂ、Ｃ、Ｄの４つの部分からなる。各部分のはじめに４～６文からなる文章がある。文中の空所に適切な言葉を四肢選択により挿入するものや、助詞を自由記入するもの、与えられた動詞を適切な形に変えて書き入れるもの、文中の１文の意味を四肢選択で確認するものなど、合計24項目ある。

以上を解答方法から分類すると次のようになる。

・自由記入問題：27問（項目1～16、50、51、52、53、57、58、59、63、67、68、70）

・選択記入問題：25問（項目17～41）

・四肢選択問題：18問（項目42～49、54、55、56、60、61、62、64、65、66、69）

実際の２学期期末試験の得点算出時には、各項目の重要度に従って、１点と２

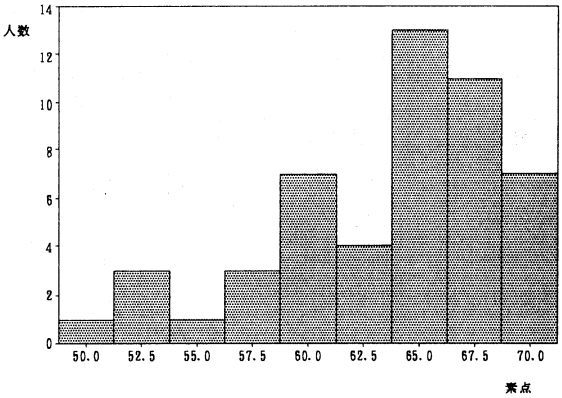
点の配点であったが、以下の統計的処理では全項目とも正答は１点、誤答は０点とした。

b　全体像

この試験の結果の概略は以下の通りである。

有効数　：50
総項目数：70
平均　　：63.56
標準偏差：4.94
最高点　：69.00
最低点　：50.00

ヒストグラム－１（２学期期末試験【文法】）



試験全体の分析結果を見ると、まず平均が、70点満点の63.56点（90.80％）と非常に高いことがあげられる。80％以上の得点（56項目以上正解）の学習者は全部で46名いる。完全学習の場合は、80％習得が一つの目安になっているから、46名については試験範囲の日本語の文法を習得できたと考えてよいであろう。項目別に見ると、70項目中、全員正解（100％）の項目は23項目あり、80％以上の正解率の項目は59項目ある。「日本語能力試験２級読解・文法」の試験のヒストグラムと比較しても、この試験結果は正規分布しておらず、極端に右あがりの分布になっていることがわかる。

c　項目分析

正答率が0.2～0.8の問題項目は13、16、34、42、43、49、50、54、56、68、70の11項目である。

弁別指数が0.4以上の項目は13、16、42、43、49、54、56、68、70の９項目である。

正答率が0.2～0.8の間にあり、弁別指数が0.4以上の項目の問題が良問であることが多いという基準（池田、1992、pp.114-115）から良好と思われる問題を選ぶと、13、16、42、43、49、54、56、68、70の９項目が該当することになり、全体

の問題70項目のうちの１割強に過ぎないことになる。これら９項目のうち、四肢選択の問題は42、43、49、54、56の５問である。問題別にまとめると以下の通りである。

項目分析表－１（２学期期末試験【文法】）

	解答形式	正答率20%未満	正答率20%以上80%未満		正答率80%以上	項目合計
			弁別指数0.4未満	弁別指数0.4以上		
問題１	自由記入	0	0	0	8	8
問題２	自由記入	0	0	2	6	8
問題３	選択記入	0	1	0	24	25
問題４	四肢選択	0	0	2	3	5
問題５Ａ－１	四肢選択	0	0	1	2	3
２	自由記入	0	1	0	0	1
３	自由記入	0	0	0	3	3
Ｂ－１	四肢選択	0	0	1	1	1
２	四肢選択	0	0	1	0	1
３	自由記入	0	0	0	1	1
４	自由記入	0	0	0	1	1
Ｃ－１	自由記入	0	0	0	1	1
２	四肢選択	0	0	0	3	3
３	自由記入	0	0	0	1	1
Ｄ－１	四肢選択	0	0	0	3	3
２	自由記入	0	0	1	1	1
３	四肢選択	0	0	0	1	1
４	自由記入	0	0	1	0	1
		0	2	9	59	70

次に、四肢選択の問題で、正答率、弁別指数の点から良好と思われる５問(42、43、49、54、56の項目）についてそれぞれの解答パターンを見てみる。受験者を得点によって４つのグループに分ける。上位27%を最上位グループ、次の23%を上位グループ、その次の23%を下位グループ、一番下の27%を最下位グループと

し、それぞれのグループがどの選択肢を選んでいるかを百分率および人数で出してみる。

四肢選択の解答パターンの理想的な形は、最上位のグループの受験者が正答である選択肢を多く選択し、下位に近づくにつれて、正答の選択肢を選ぶ割合が減り、最下位のグループではどの選択肢をもほとんど均等に選んでいるといった形である（池田、1992、 p.110)。

問題項目４２は、正解肢であるａを選択した受験者は最上位グループでは76.9%、上位グループでは50.0%、下位グループでは25%、最下位グループでは23.1%となっている。選択肢ｂは反対に、下位２グループで選択したものが多く、上位２グループで選択したものが少ない。選択肢ｃは上位の３グループでは選択したものが一人もおらず、最下位グループの者（２人）だけが選択している。選択肢ｄでは上位グループに属する者（１人）だけが選択している。選択肢ｃとｄは選択肢としてあまり機能しておらず、したがって、42の問題は実質的には２肢選択に近い問題となっている。

同様に、問題項目43、49、54、56においても、正解肢であるａは最上位グループでは約70%から100%が選択している。上位グループでは項目49を除いて(33.3%が選択)、50%以上が選択している。下位グループでは項目56を除いて(66.7%が選択)、50%以下の選択となっている。最下位グループでは正解肢を選択したものはどの問題項目でも50%以下である。選択肢ｂは、下位、最下位グループで選択したものが多くなっている。選択肢ｃは最上位グループでは選択したものは一人もおらず、上位、最下位、下位グループで選択した者も少ない。選択肢ｄを選択したものは、問題項目43、49では一人もいない。問題項目54では選択肢ｄを選択したものが、最下位グループに３人、下位グループに１人いる。問題項目56では選択肢ｄを選んだ者は下位グループに１人いるだけである。

解答パターンからは、問題項目54がどの選択肢も使われており、また、正解の選択肢を最上位グループが多く選択している形で、理想に近い形になっている。ほかの問題項目では、だれも選択していない選択肢や選択している受験者数が極端に少ない選択肢があり、多肢選択の選択肢として機能していないわけであるから、作り方の工夫が必要であると思われる。

4－2　2学期期末試験読解（100点、70分、38項目）

a　試験の内容

38問の試験は、3つの部門（問題I～Ⅲ）から構成されている。文科・理科の全学生が受験するものであるが、各部門の内容は、問題Iが文科理科共通のトピック、問題Ⅱは文科的内容、問題Ⅲは理科的な内容となっており、学生の専門による得点の偏りが出ないように配慮されている。

問題Iは、日本人と欧米人の文化的な思想を対比させた文章をもとにした問題で、質問は17項目ある。問題Ⅱは、心情についての文章をもとにした問題で8項目よりなる。問題Ⅲは、科学的記録に関する文章をもとにした読解問題で、13項目よりなる。出題内容は各部門とも、文章中の空欄に適当な副詞や接続表現などを入れるもの、文章中の表現の文脈の中での意味や同意の表現を問うものなどである。

各項目を解答法の形式面からみると以下のように分類された。

- 記入　問題　：4問（項目4, 9, 10, 25）
- 四肢選択問題　：14問（項目1, 2, 3, 8, 14, 17, 18, 19, 20, 21, 22, 23, 24, 38）
- 三肢選択問題　：8問（項目6, 7, 26, 27, 28, 29, 30, 37）
- 二肢選択・正誤問題　：10問（項目5, 11, 12, 13, 31, 32, 33, 34, 35, 36）
- 四肢択二問題　：2問（項目15, 16）

こうした解答方法の不統一は統計的処理を行う上でさまざまな問題を生む。選択肢数の違いはチャンス・レベルの項目間の違いとなるし、記入問題では通常の統計的誤答分析ができない。前者についてはチャンスレベルを無視して処理し、後者については学生の誤答を分類し、これを仮の選択肢として処理せざるを得なかった。こうした修正などによって一応の数量化は可能となったが、分析の精度には多くを期待できない。

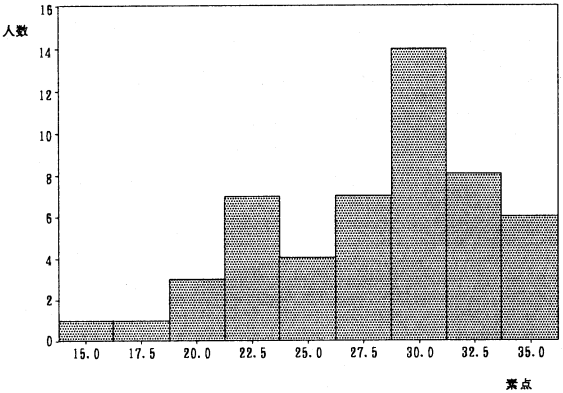
また、実際の試験での得点算出時には、各項目の重要度に従って1点から4点までの重みづけがあったが、以下の統計的処理では全項目とも正答：1、誤答：0となっている。

b　全体像

この試験の結果の概略は以下の通りである。

有効数　：51
総項目数　：38
平均　：28.0
標準偏差　：4.88
最高点　：36.00
最低点　：14.00

ヒストグラム－２（２学期期末試験【読解】）



能力試験という観点から見ると、平均点（73.6%）がやや高く上詰まりの傾向はあるが、ばらつきに大きな偏りは見られず（ヒストグラム－２および標準偏差参照）、正規型の分布から大きくはずれてはいないと考えることができる。

これは、従来（1991年度以前）の読解試験や文法等の成績とは大いに異なる結果であるが（平均点は常に80%以上であった）、試験問題作成時点での以下に述べる配慮が利いていたのではないかと思われる。

この試験問題は、東京・大阪両外国語大学の担当者の事前打ち合わせのもとで作成されたが、その合意事項として以下の２点があった。

１．東京外国語大学で作成していた従来の試験問題は、学習済みの文型・語彙だけをもとにした到達度型のテストであったが、前年度より未習事項を含めた能力試験型の部分を加えた。今回もこの能力試験型の部分を保存する。

２．試験全体を３つの問題に分け、問題Ⅰ（一般的な内容）では、従来通りの到達度試験の性格を残す。問題Ⅱ・問題Ⅲは、それぞれ文科的な内容・理科的な内容で、読解の「能力」を測るものに近づけ、未習の語彙・文型を適度に含める。（ただし、未習の要素を読解テキストに含める場合、そのキーワードやキーセンテンスにはならないよう配慮された）。

この問題間の性質の相違は試験結果に明確に現れており、

問題Ⅰ　（項目　1～17）の総平均：79.58%

問題Ⅱ,Ⅲ（項目18～38）の総平均：68.90%　となっている。
統計的な分布だけからみれば、問題Ⅱ,Ⅲのような未習要素を含む内容の方が、上詰まりの分布が避けられ、能力を測るのに適当な分布に近いようである。

c　項目分析

項目分析によって算出された正答率と弁別指数の２つの値から、各項目(設問)を難易度と弁別指数面から分類すると以下のようになった。解答形式については、ほぼ項目毎に異なっているため記していない。

項目分析表-２（２学期期末試験【読解】）

問題部門	正答率20%未満	正答率20%以上80%未満		正答率80%以上	項目合計
		弁別指数0.4未満	弁別指数0.4以上		
Ⅰ	0	6	4	7	17
Ⅱ	0	3	3	2	8
Ⅲ	0	3	6	4	13
計	0	12	13	13	38

この表を見ても分かるように、能力測定上適当な問題水準であるとされる「正答率20%以上80%未満、弁別指数0.4%以上」の項目数は三分の一程度にすぎない。また易しすぎる問題（正答率80%以上）の項目数は三分の一を超えている（問題Ⅰにおいて特に多い)。

正答率20%以上80%未満となったもののうちでも少なからぬ項目が弁別指数が0.4%未満になってしまっているが、こうした項目の問題点のいくつかは以下のように分析可能であった。

[項目6] 選択課題が他の項目と連動している。他の項目で間違えた者はこの項目6で正答しても誤答となってしまうこと、また、他の項目5で間違えた者がこの項目6で再び間違えても正答となる可能性があるために弁別指数が低くなったと思われる。選択肢の提示方法に問題がある。

[項目15と16] これら項目は四肢択二課題となっている。項目分析の際には、正答となる２選択肢を別々に扱って分析を行ったが、本来、この出題形式は通常の項目分析で扱えない。統計的には出題形式に問題がある。

[項目19] 伝聞の「そう」が求められている問題であるが、その正解を規定する「～

によると」も項目18で問題になっているために弁別指数が低くなったと思われる。これも他項目との連動問題であろう。

［項目21］適当な副詞を求める問題であるが、実は正答となる語は未習語である。未習語はキーワードにしないのが原則であった。従って問題として取り上げることに問題があった。他の選択肢は既習語であるが、単語単位の問題では、消去法によっては唯一の正解に至ることができないということであろう。

総括として言えることは、本試験を能力を測定するものとして統計的な見地から見た場合、大いに改善されなければならないということである。試験作成の目的は様々であるが、この読解試験を能力測定目的で用いるのは明らかに危険である。これは解答方法が整理されていない点、項目一つ一つの科学的吟味が不足している点などが主な原因であろう。能力を測定するものとしての試験なのか、到達度を測るためのものなのかの議論を出発点として、問題点を継続的に点検し、試験の妥当性と信頼性とを高めていくことが強く望まれる。

4－3　2学期期末試験聴解（100点、70分、50項目）

a　試験の内容

50項目の試験は、4つの部門（Ⅰ～Ⅳ）から構成されている。問題ⅠはA（10項目）とB（9項目）の二つのセクションからなっており、A・B共に問題は、テープの音声を一度だけ聞いた後に、文字（ひらがな、カタカナ、アルファベット、数字）による再生を必要とするディクテーション形式のものである。

問題ⅡもA（6項目）とB（6項目）の二つのセクションからなっており、ここでは各セクション異なったビデオテープを見て、ビデオの内容の適否を、与えられた文中の二者（A，B）択一問題を解きながら文を完成させるものである。

問題Ⅲは6項目で、それぞれ短い会話を一度聞いた後、内容に関する説明文が音声で4問与えられ、正しいものを一つ選ぶ四肢選択形式のものである。

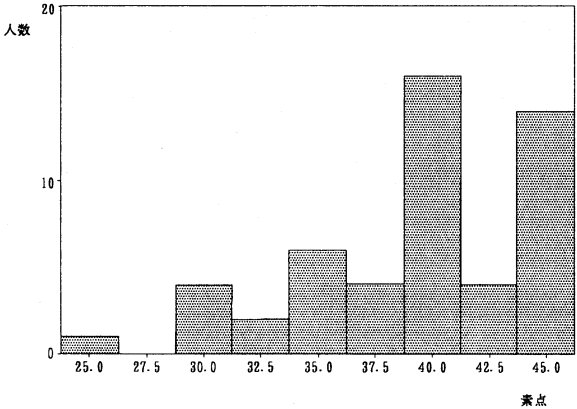
問題Ⅳは13項目で、短い会話を二度聞いた後、音声によって与えられる内容の適否を○×で解答する形式のものである。

b　全体像

この試験の結果の概略は以下の通りである。

有効数　：51
総項目数：50
平均　　：39.2
標準偏差：5.02
最高点　：46.00
最低点　：24.00

ヒストグラム－３（２学期期末試験【聴解】）



100点満点での平均点は78.4点で、かなり高いことが特徴として挙げられる。また、得点のヒストグラム（ヒストグラム－３参照）は大きな段差を示し、満点に近い高得点をとっている学生集団と平均点レベルに集中している二つの学生集団が存在することがわかる。いずれも極端な右偏り傾向になっており、問題の易しさを示していると判断できよう。このことは、項目分析からも明らかで、正答率100％の項目が50項目中10項目もあり、また､80％以上のものを含めるとその数は28項目（56％）に及ぶ。

c　項目分析

項目の難易度と弁別指数を使用して、項目分析を行った。問題として適切なものを選ぶ基準として、正答率が20％から80％の間にあり、かつ弁別指数が 0.4以上の項目を選び出した。ただし、解答が再生形式で、解答パターンが多数にわたり項目分析不可能である問題Ｉ－ＡとＩ－Ｂの19項目は除いた。その結果、残り31項目中、７項目が望ましい項目として選び出された。その他に、難易度は適切ではあるが、弁別指数がない項目が５項目、残りの19項目は､すべて易し過ぎて、弁別指数なしという結果を得た。問題別にまとめると以下の通りである。

項目分析表－３（２学期期末試験【聴解】）

	解答形式	正答率20%未満	正答率20%以上80%未満		正答率80%以上	項目合計
			弁別指数0.4未満	弁別指数0.4以上		
問題Ⅰ－A	再生記入	＊	＊	＊	＊	１０＊
Ⅰ－B	再生記入	＊	＊	＊	＊	９＊
問題Ⅱ－A	二肢選択	0	0	0	6	6
Ⅱ－B	二肢選択	0	2	1	3	3
問題Ⅲ	四肢選択	0	2	1	3	6
問題Ⅳ	真偽○×	0	1	5	7	１３
		0	5	7	１９	５０

＊解答が再生記入形式で、解答パターンが多岐にわたり項目分析不可能

問題Ⅰ－A：

この問題は、音声テープを聞かせ、内容の一部をひらがなで（　）の中に記入させるディクテーション形式の問題である。一種の再生形式のため、誤答パターンが５種類以上におよぶ場合があって、本来の項目分析は不可能となった。したがって以下の項目分析では、誤答パターンの数を明示し、誤りの原因となった要素を考察してみる。

全問正解、かつ誤答パターンが１つのみの問題項目はゼロで、誤答パターンが２つのものは、４項目（項目２、７、８、10)、３つのものは、４項目（１、３、５、６)、５つのものは、１項目（４)、そして誤答パターンが最多18にも及んだものは、１項目（９）となっている。文字による再生形式のため、誤答を分析してみると、いくつかの項目で、「い」と「ひ」の混同、「て」と「で」や、「か」と「が」などの有声音・無声音の認識で問題をもつ被験者のいることがわかる。また、拍感覚が十分把握されていない点や、拗音の認識で問題が存在していることが推測される。また、新出単語／表現の音声上の認識とその単語／表現の表記にズレのあることが確認できる。これらの直接の原因が、聴解力に問題があるのか、表記上の単純な記入ミスによるものなのかは分析できない。しかし、今後の解答形式のありかたを検討する必要は十分あろう。

問題Ⅰ－Ｂ：

この問題もⅠ－Ａ同様、音声テープを聞かせ、内容の一部を漢字以外の文字で（　）の中に記入させるディクテーション形式の問題である。分析方法は、Ⅰ－Ａと同様であるが、誤答パターンの数は、10項目中８項目が10から24にも及び、問題の難易度はＡより高くなっていると考えられる。

誤答パターンの少ない順からみてみると、４つのものが項目17、次いで８つのものが項目14、10のものが項目18、12のものが項目16とそれぞれ１項目ずつある。また、誤答パターンが14のものが２項目（12と15）、20のものが１項目(11)、24のものが１項目（19）、最多25のものが１項目（13）となっている。Ｂでは解答を固有名詞で記入させるものが多い。誤答パターンを分析してみると、誤答の多くが明らかに固有名詞の表記ミスであろうと考えられるものが目立つ。ここでも、有声音・無声音の区別、正しい拍感覚、拗音の正確な認識に問題があることが推測される。また、固有名詞の音声上での認識と表記上での正しい理解の度合いによって、正答率が左右される可能性も考えられる。誤答の要因が、聴解力にあるのか記入ミスにあるのか直接的な要因分析はできないが、Ⅰ－Ａ同様、解答形式の検討は課題として挙げられよう。

問題Ⅱ－Ａ：

ビデオを見た後で、ビデオの内容の適否を問う問題であるが、６項目中５項目（20～24）は、全員正解という結果を得ている。また残り１項目（25)も下位・最下位群６％を除き94％が正解している。二者択一の形式であるが、「まよわし」がその機能をほとんど果たしていないことが指摘される。全体として非常に易しすぎる項目で占められており、今後の選択肢の作り方に工夫が必要であると思われる。

問題Ⅱ－Ｂ：

Ⅱ－Ａ同様、ビデオを見た後で、ビデオの内容の適否を、文中の二者（Ａ，Ｂ）択一問題を完成させながら内容理解度を測る問題である。６項目中２項目（26、28）は難易度ゼロで全員正解である。「まよわし」は無意味だったことが挙げられる。項目30では、下位群、最下位群では、各１名「まよわし」を選んではいるが、難易度としては易しすぎる項目となろう。一方、難易度は適切だが、弁別指数のない項目が２項目（27、31）ある。下位群にいく程、「まよわし」を選んでいるものが多くなるような望ましい項目は、１項目（29）のみである。

問題Ⅲ：

全６項目で、それぞれ短い会話を一度聞いた後、内容に関する説明文が音声で４問与えられ、正しいものを一つ選ぶ四肢選択形式のものである。６項目中半分の３項目（34、35、37）が、正答率90％前後で易しすぎる項目である。他の２項目（32、36）は、難易度は適切であるが、弁別指数のない項目で、難易度、弁別指数共に望ましい項目は、１項目（33）のみである。

問題Ⅳ：

短い会話を二度聞いた後、音声によって与えられる内容の適否を○×で解答する形式のものである。全13項目中７項目が易しすぎる問題で、その中でも３項目（40、42、48）は、全員が正答を選んでいる。難易度は適切であるが、弁別指数のない項目は１項目（44）で、難易度、弁別指数共に望ましい項目は、５項目（39、41、45、46、50）ある。望ましい項目数の割合からみると、問題Ⅳは適切な問題項目が最も多いことがわかる。

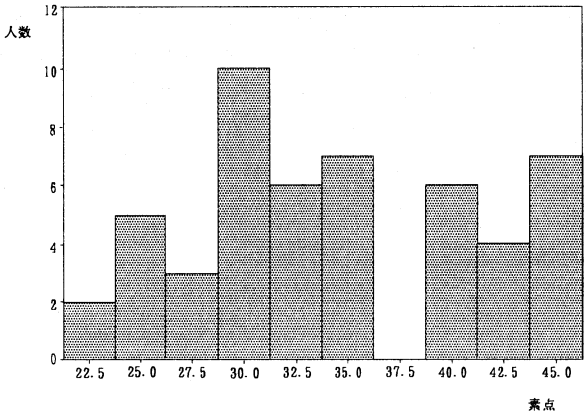
４－４　相関

今回は、調査研究の対象になった２学期期末試験の「文法」「読解」「聴解」の３種類と、日本語能力試験の「読解・文法」（「ヒストグラム－４」参照）と「聴解」（「ヒストグラム－５」参照）との相関を調べてみた。また、同時に試験の均質性、客観性、信頼性等に関する資料、及び２つの異なる試験の内容的妥当性の資料の一部を得ることもその目的とした。

まず、２学期期末試験「文法」と能力試験「読解・文法」の信頼係数は .6129 でかなり相関があることが分かる。また、２学期期末試験「読解」と能力試験「読解・文法」の相関係数も .5721 で文法同様かなりの相関がある。一方、２学期期末試験の「聴解」と能力試験の「聴解」の相関係数は、上記の２つと比較して若干低くはあるが、相関のあることが確認できる。（「相関係数表」および「試験間相関－１～３」参照）

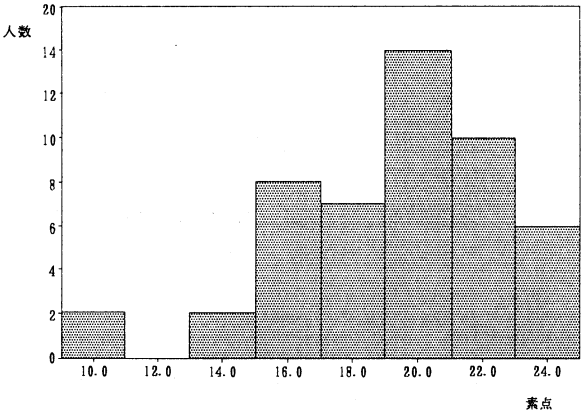
ヒストグラム－4（日本語能力試験【読解・文法】）

有効数　：50
総項目数：50
平均　　：34.36
標準偏差：6.95
最高点　：46.00
最低点　：22.00



ヒストグラム－5（日本語能力試験【聴解】）

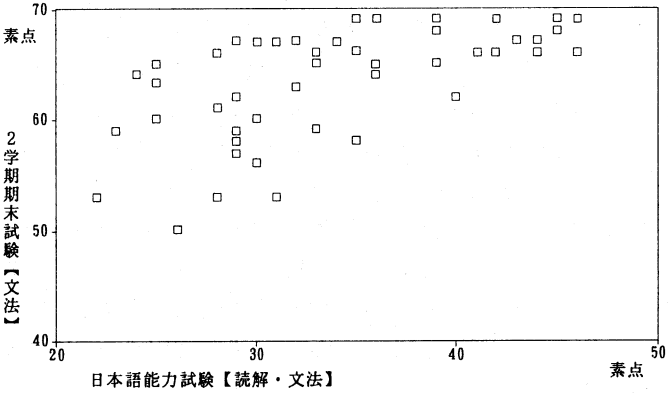
有効数　：49
総項目数：23
平均　　：18.73
標準偏差：3.17
最高点　：23.00
最低点　：10.00



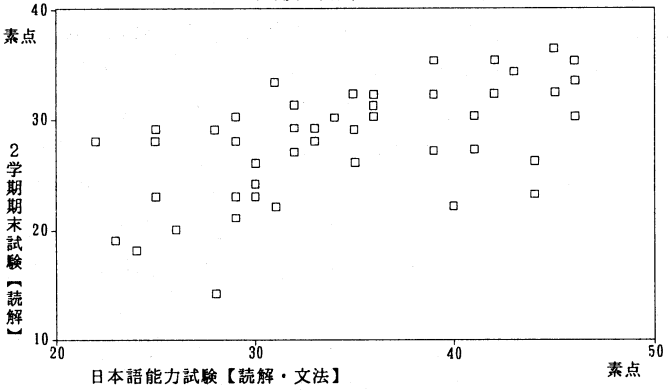
相関係数表

試　験	期末試験「文法」	期末試験「聴解」	期末試験「読解」	能力試験「聴解」
期末試験　「聴解」	.6392			
期末試験　「読解」	.5525	.5523		
能力試験　「聴解」	.5345	.4995	.2479	
能力試験　「読解・文法」	.6129	.6619	.5721	.4966

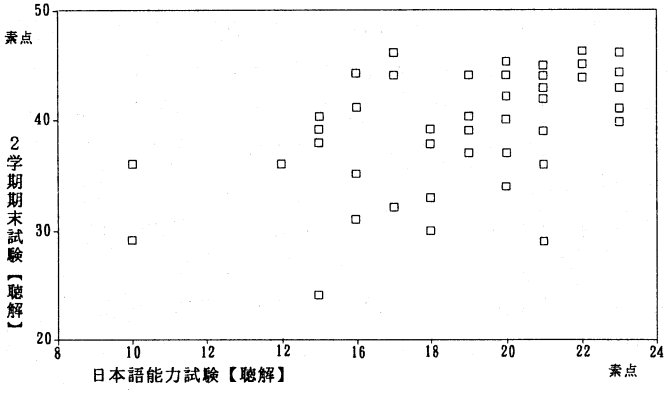
試験間相関－1



試験間相関－2



試験間相関－3



5　問題点

今まで学内試験に関しては信頼性や妥当性が検討されてこなかったが、今回の調査研究により、試験問題の内容、設問、選択肢、解答用紙の作り方など今後の試験作成にあたって参考とすべき点が多々明らかにされた。

「文法」の試験は、今回分析対象としなかった「漢字」の試験と同様に、2学期期末試験は2学期の目標に達しているかどうかを見るための到達度テストである。その目標は2学期後半の授業で学んだことという狭い範囲のものになっている結果、現行の試験結果を見ると、ほとんどの普通の学生にとっては教えられたことを記憶してさえいれば高い点が得られる試験であるように思える。

このような易しい試験の問題点は、上位のものの実力が天井効果により測定できず、試験結果に反映されない点である。その結果、ちょっとした不注意による間違いや勘違いが得点を左右することになり、信頼性を低くすることになる。

予備教育修了試験は、学生が試験実施の時点で大学での勉学に必要な日本語能力を有しているかどうかを測定するものであると同時に、将来4年間の大学生活の中で日本語力が伸びていき、勉学に必要な日本語力を獲得していく可能性をも予測するものでなければならないだろう。教えられた日本語が学生の中で能力として定着し、それが応用能力にまで発展していく予測的可能性を測定するような能力試験が必要である。このような応用能力は「読解」「聴解」「話し方」「作文」の試験で測定されるべきものであろう。能力試験の内容の妥当性の検討は今後の大きな課題である。

大阪外国語大学留学生日本語教育センターでも国費学部留学生の予備教育を行うようになった現在、何らかの形で両センターの学生の日本語力を客観的に測定するものが必要となってきている。学部入学予定の留学生の基礎日本語能力試験としてどのようなものが適切であるのか、試験実施ごとに試験問題を分析し、内容を検討して、良問を積み重ねていくようにするのが望ましいと思われる。今回はその第1回目ということで内容面よりも形式面での分析が主になった。分析に耐えられるように、試験問題を形式面で整えることがまず必要なことであると考えたためである。

今後の試験問題作成に関して、いくつかの形式面での提言を行いたい。

1　試験問題に通し番号をつけ、分析しやすいようにする。

2　なるべく選択問題にする。そのさい、記号を選択させるようにし、書かせることは避ける。試験は測定・評価するための道具であり、練習のための

道具ではない。これが授業の中で日々行うクイズなどの形成的評価と期末試験や予備教育修了試験などの総括的評価との違いである。

3　選択肢を工夫し、だれも選択しないような無意味な選択肢は極力避けるようにする。

4　同一問題の中にさまざまな解答方式（自由記入、選択記入、四肢選択、正誤問題など）が入ることを避ける。

5　学生番号をつける。届けてある正式な名前を試験の答案に書かず、クラス内での呼び名を書いているものが何人かいた。同じような名前の学生がいる場合もあるので、学生番号を与え、名前と一緒に書かせる。

6　答案用紙は別にすると分析には便利である。

最後に、今回行ったような統計分析だけでは良好問題、不良問題の判断を直ちに下すことはできず、試験問題をよりよい問題にしていくためには、設問内容の詳しい検討が必要であることは言うまでもないことを付け加えておく。しかし、良好問題とされた問題については、その内容を分析し、良問形式を今後の試験作りの参考とする、不良問題とされた問題については、内容に問題があったのか、形式面での工夫が必要であったのかを分析し、改善していくという経験を積み重ねて行くことによって、将来の能力試験開発につながって行くと考える。

参考文献

池田央『テストの科学』日本文化科学社、1992年。
石田敏子『入門日本語テスト法』大修館書店、1992年。
外国人日本語能力試験実施委員会企画小委員会『日本語能力試験の概要1992年版』国際交流基金、1993年3月。
竹谷誠『新・テスト理論』早稲田大学出版部、1991年。
東京外国語大学留学生日本語教育センター編著『中級　日本語』凡人社、1994年。
日本語能力試験企画小委員会　出題基準作成会議『日本語能力試験　出題基準（外部公開用）』国際交流基金、1993年3月。
姫野昌子「学部留学生に求められる学力を考える」『留学交流』　ぎょうせい、1991年7月。

本研究調査は平成５年度東京外国語大学教育研究学内特別経費による助成を受け

た共同研究であるが、執筆にあたっては、横田が１、２、３、４－１、５を、伊東が４－３、４－４を、西郡が４－２を担当した。

謝辞：本研究の試験の項目分析のために、快く自作項目分析用プログラムの使用を許可し、さらに指導してくださった石本菅生先生（国際基督教大学教授）にお礼申し上げます。

A Preliminary Study for Development of the Japanese Proficiency Test for Pre-College Students (1)

YOKOTA Atsuko, ITO Sukero & NISHIGORI Jiro

There is a growing need that a standard Japanese proficiency test for pre-college students should be produced. The Japanese Language Proficiency Test (JLPT) or Nihongo Noryoku Shiken, by the Japanese Association of International Education and the Japan Foundation, is a well- known examination to identify the individual's Japanese proficiency in the areas of vocabulary, reading/grammar and listening. However, the JLPT is aimed to evaluate general Japanese language proficiency but not pre- college students'. This research report describes a preliminary study of the Japanese language tests given to the international students prior to their entry into Japanese universities.

The investigation was conducted to provide information that could guide a decision on the ways of developing contents, formats and administrative factors of a standard Japanese proficiency test for pre- college students. The approach taken was to analyze the student scores and each test item of the term- end Japanese tests (grammar, reading, and listening) developed by the Japanese Language Center for International Students, Tokyo University of Foreign Studies. The additional investigation, analyzing the result of the JLPT taken by the same group, was undertaken to examine if there were satisfactory correlations with the term-end tests.

It is concluded that further modifications in the contents and the directions of the term-end tests should be taken in order to increase a test value as a Japanese language proficiency test. Other possible suggestions are also discussed.