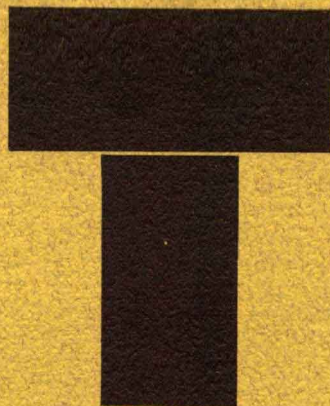




日本語テストハンドブック

付録：日本語能力認定試行試験問題

日本語教育学会―編 編集代表―林 大

日本語テストハンドブック

付録：日本語能力認定試験試験問題



日本語教育学会—編集代表—林

© The Society
for Teaching Japanese
as a Foreign Language 1991
日本語テストハンドブック

1991 年 4 月 10 日 初版発行 定価 4,300 円
(本体 4,175 円・税 125 円)

編 者 社団法人
日本語教育学会
代表 林 大

発 行 者 鈴 木 荘 夫

発 行 所 株式 大修館書店
会 社

101 東 京 都 千 代 田 区 神 田 錦 町 3-24
電話/販売部 03-3295-6231 編集部 03-3294-2356

印刷/壮光舎印刷 製本/関山製本 装幀/代田 奨
ISBN 4-469-02105-9 Printed in Japan

はしがき

日本語教育学会は、国際交流基金の助成を得て、昭和 55 年度から昭和 59 年度までの 5 年間「日本語能力認定試行試験」を実施した。この試験は、日本語能力試験の世界各地での実施可能性を知るための調査研究であり、現在実施されている「日本語能力試験」の基礎を固める役割を果たしたものである。

我々は、この 5 年間に蓄積された具体的な経験と豊富なデータをもとにして、日本語教育の研究者、教育測定の研究者、言語学の研究者という異分野の研究者の協力によって、さらに 6 年間、討論を重ねた。この『日本語テストハンドブック』は、そこで得られたテストに関するさまざまな知識・ノウハウを日本語教育に携わる者の共有の財産にしようという意図で企画されたものである。

教育の場で、テストは一般に意識されているよりも非常に大きな役割を果たしている。学校での意志決定の多くがテストの結果に基づいているといっても過言ではない。たとえば、教員の採用・不採用の決定、学生の入学許可・不許可の決定、教えるべき内容の決定、授業の進度の決定等々教育に関わる重要な決定には、その資料としてテストの結果が用いられる。ちょうど船における羅針盤のような役割を果たしているのが教育におけるテストであ

る。指導者の側でその指導技術を反省し、その改善、展開をはかる第一の手がかりとなる一方、学習者の側から見れば、テストの結果から何が自分の長所であり、何が自分の弱点であるか等を知ることによって、学習者自身が学習計画を決定することが可能となる。したがって、質の良い日本語のテストは、日本語教育の質の向上のために極めて重要な役割を果たすといえることができる。それにもかかわらず、このように重要なテストが安易に行われ、その結果が安易に利用されているのが現状なのではあるまいかと感じられる。

テストの作成は、容易なものではない。最初に述べた「日本語能力認定試行試験」の問題作成の際にも、一生懸命考えたすえに、成績上位の者よりも成績下位の者の方が多く正解するような問題項目を作った経験が何度かある。そのような不適切な問題項目の改訂に工夫を重ねたことはいうまでもないが、その結果として改悪になってしまった場合もなかったわけではない。良質のテストを作成するためには、テスト自身を評価する過程を熟知し、その結果に基づいて、改良を加えていくノウハウを身につけておくことが不可欠であることは、我々自身が痛切に感じているところである。本書のねらいの第一点は、5年間の経験を通じて得られた良質のテスト作成のためのノウハウを公にし、今後の日本語のテストをよりいっそう高度なものとする助けとなることである。

テストの質を評価し、改良を加えるために統計的な手法を利用することがいかに有効なものかもその機会に十分に知らされた。文科系出身者の数が圧倒的に多い日本語教育の世界では、このような手法が十分に活用されているとはいいがたい。本書のねらいの第二点は、基礎的な統計的手法の活用が日本語教育の世界で常⁴

識となるようにすることである。

「日本語能力とは何か」ということが討論に加わった執筆者全員にとって最大の難問であった。日本語能力観は、現在にいたってもしっかりとしたものにはなっていない。最近になって、さらに新しい能力観が出現してきてさえいる。そのようなわけで、本書が日本語能力を測るテストの作成法の究極的な姿を示したものであると主張するつもりはない。それは、テスト理論の面からも同様である。編集代表者としては、この有意義な研究作業に多年参加してその結果をここまでまとめあげられた、大坪一夫氏を中心とする委員諸氏の労を多とするとともに、改めて、読者諸氏の建設的な御批判を願ってやまない次第である。

最後に、本書の基礎となった5年間の調査研究を可能にしてくださった国際交流基金、「日本語能力認定試行試験」実施に関係された前日本語教育学会会長（認定委員会委員長）故小川芳男先生をはじめ、出題委員、協力委員、日本語教育学会の事務担当者等多くの方々、同試験に積極的に参加してくださった受験者諸氏、本書の編集委員の無理難題に辛抱強く耳を傾け、現在の形に整えてくださった大修館書店編集部の岡田耕二氏、富永七瀬氏に心から感謝する。

平成3年3月

編集代表者 日本語教育学会会長

林 大

編者・執筆者一覧

【編者】

編集代表	林 大
編集委員長	大坪 一夫
編集委員	福地 務
	村上 隆

【執筆者】

1章	野田 時寛
2章	村上 隆
3章	斎藤 修一・大坪 一夫
4章	福地 務・河原崎 幹夫・大坪 一夫・清田 潤
5章	村上 隆
6章	角田 太作
7章	村上 京子・藤原 雅憲・酒井 たか子
8章-1.	村上 隆
-2.	三枝 令子
9章-1.	高見沢 孟
-2.	菊池 康人

日本語能力認定試行試験問題

【日本語能力調査認定委員および協力委員】

今榮 國晴	椎名 和男
越前谷 孝子	富田 隆行
小川 芳男	野沢 素子
大坪 一夫	福地 務
加藤 容子	藤原 雅憲
河原崎 幹夫	水谷 信子
才田いずみ	村上 隆
斎藤 修一	(五十音順)

目 次

はしがき	iii
編者・執筆者一覧	vi

第1章 何のためのテストか

1. テストとは.....	2
(1) はじめに	
(2) テストの二大別	
2. テストの目的	5
3. まとめ	8

第2章 良いテストはどのような性質をもつか

1. はじめに	10
(1) テストの理論	
(2) この章の目的と読み方	
2. 測定誤差の理論	13
(1) テストの信頼性と妥当性	
(2) 測定誤差の理論	
(3) データとモデル	
3. 個人差測定の特徴と相関係数の役割	31
(1) 個人差測定におけるさまざまな制約	
(2) 相対評価と標準得点	

(3) 相関係数と基準関連妥当性	
(4) 信頼性と相関係数	
4. 信頼性係数とその意味	48
(1) 基本モデルと信頼性係数	
(2) 信頼性係数と平行テスト	
(3) 信頼性係数の意味	
(4) 相関係数の希薄化	
5. テスト項目の統計と α 係数	64
(1) テスト得点と項目得点	
(2) α 係数と内的整合性	
(3) 信頼性係数としての α 係数	
(4) 内的整合性の意味——二つの架空のテスト	
6. 項目分析	85
(1) 項目分析とは何か	
(2) 項目分析の実際	
(3) 結果を解釈するにあたっての留意事項	
7. テストの妥当性	98
(1) 妥当性の概念	
(2) テスト得点の性質	
(3) 内容的妥当性	
(4) 発想の転換	
(5) 構成概念妥当性	
(6) おわりに	

第3章 日本語能力のシラバス試案

1. シラバス	116
(1) シラバスとは何か	
2. 日本語能力とシラバス	121
(1) 日本語能力のシラバスの例	
(2) より充実したシラバスを求めて	

第4章 テストはどのように作られるべきか

1. テストの形式 132
 - (1) はじめに
 - (2) 主観テストと客観テスト
 - (3) 客観テストの形式
2. 読解テスト 144
 - (1) はじめに
 - (2) 問題形式の種類
 - (3) 読み物のテキストタイプ
 - (4) 問題作成の実際
 - (5) 長文読解の問題点
3. 聴解テスト 163
 - (1) 何を測るのか
 - (2) 問題形式の種類
 - (3) 問題作成の実際

第5章 大規模なテスト —— データの分析の一例

1. はじめに 182
2. データ分析の目的 182
 - (1) データの統計的処理
 - (2) データの保存とその意義
3. データ分析の一例 188
 - (1) 基本的な分析
 - (2) テストの内部構造の分析
 - (3) テスト得点の等化の試み
 - (4) データ分析の妥当性検討への含意

第6章 日本語教育のための基礎資料としてのテスト

1. はじめに222
2. 母語の影響222
 - (1) 母語の影響の実例
 - (2) 母語とテスト成績の関係
3. 語順とそれに関連した現象228
4. おわりに237

第7章 事前的評価——日本語習得適性テストの実践

1. 事前的評価の目的と内容 242
 - (1) 事前的評価の目的
 - (2) 既習者に対する事前的評価
 - (3) 未習者に対する事前的評価
2. 事前的評価の対象としての適性 246
 - (1) 外国語習得適性テスト
 - (2) 適性テストの目的と特性
3. 日本語習得適性テストの作成 250
 - (1) 日本語習得適性テストの作成
 - (2) 日本語習得適性テストの信頼性と妥当性
 - (3) 日本語習得適性テストの利用と問題点

第 8 章 形成的評価——授業における評価活動の実例

1. 形成的評価の理論 266
 - (1) はじめに
 - (2) 評価の機能
 - (3) 形成的評価と S-P 表
2. 授業における評価活動の実例 273
 - (1) はじめに
 - (2) コース概要
 - (3) コース中の評価
 - (4) コース中のテスト
 - (5) S-P 表の利用
 - (6) その他の評価
 - (7) コース後の評価

第 9 章 対面テストと作文の出題と採点

1. 対面テストの作成と採点 292
 - (1) はじめに
 - (2) 対面テストの観察対象
 - (3) 対面テストの目的
 - (4) テスト方式の種類
 - (5) 対面テストにともなう問題点
 - (6) 対面テストの能力測定
 - (7) まとめ
2. 作文——その性質と出題・採点 305
 - (1) 作文というものの性質
 - (2) 出題の仕方
 - (3) これまでに提案されてきた採点方法（「項目細分方式」）の検討
 - (4) 本稿の採点方法

*

*

〈統計の基礎 I — 正規分布〉	20
〈統計の基礎 II — 平均値〉	23
〈統計の基礎 III — 分散と標準偏差〉	25
〈統計の基礎 IV — 相関係数の定義〉	40
〈統計の基礎 V — 相関係数の意味〉	43
〈統計の基礎 VI — 回帰直線による予測〉	56
〈統計の基礎 VII — 予測の実例〉	58
〈統計の基礎 VIII — 予測の精度〉	60
〈統計の基礎 IX — 2 値変数の統計測度〉	66
〈統計の基礎 X — 合計点の統計測度〉	71
〈統計の基礎 XI — 2 項分布〉	81
〈統計の基礎 XII — 相関係数に影響を与える諸要因〉	96
〈統計の基礎 XIII — 標本誤差と推定の精度〉	183
〈統計の基礎 XIV — 因子分析〉	200
〈テストの話題 A — まぐれあたりの効果〉	15
〈テストの話題 B — 得点の絶対的意味〉	36
〈テストの話題 C — ガットマン・スケール〉	83
〈テストの話題 D — 目標準拠テスト〉	103
〈テストの話題 E — 信頼性と測定誤差の具体例〉	192

付録 日本語能力認定試行試験問題

日本語能力認定試行試験について	326
日本語能力認定試行試験問題 1983 年版	336
1984 年版	387
試験問題の分析について	420

*

*

参考文献一覧	437
索引	442
執筆者紹介	447

第1章

何のためのテストか

1. テストとは

(1) はじめに

これからテストについて考えていく。何かについて考える場合、まず、考える対象をはっきりとつかむことが大切である。言い換えれば、その定義から始めなくてはならない。すなわち、「テストとは何か」。

テストといっても、体力テストから、心理テスト、製品検査などいろいろな種類があるが、この本でのテストとは、言うまでもなく「外国語としての日本語」のテストである。

日本語のテストは、たとえば「これは漢字のテストです」というように何枚かの印刷された紙として、「もの」としての形もあるが、より広く、一つの行動として、「テストをする」という形で考えたい。5 W 1 H というものがある。上の短すぎる問いを言い換えると「テストとは、いつ、どこで、だれが、なぜ、どのように、何をするものなのか」ということになる。

「いつ、どこで、だれが」ということは、実際にテストを実施することを考えると、非常に重要な側面であり、テストの内容にも大きく影響することになるが、とりあえず今は考えないことにする。残るのは「なぜ、どのように、何を」である。「なぜ」はより狭く「何のために」と言っていいた**であ**う。まず、「何を」するのかということに一応の答えを付けておこう。「日本語のテストとは、ある課題を受験者に与えて、それをさせることによって、受験者の日本語の能力を測るものである」。

ここで、「日本語能力」とは何か、それはそもそも測りうるものなのか、という根本的な問いがまた生まれてくるが、それは今は問わないことにする(⇒2章「良いテストはどのような性質をもつか」、3章「日本語能力のシラバス試案」)。

「何のために」の答えとして、「能力を測るために」と答えることも

出来るが、それはまた、「何のために能力を測るのか」という問いを生む。ここがはつきりしないと、「どのように」すればいいのかという問いにも答えられない。目的によって、最も適切な方法が決まってくるからである。日本語教師は、いったい何のためにテストをしているのだろうか。

それを考える前に、テストの性質についての基本的な問題を次節で考えておくことにする。

(2) テストの二大別

テストは、その結果をどのように使うかということから、大きく二つに分けられる。その結果によって、受験者を選別するのかもしれないのか、という点である。選別試験の代表的な例は入学試験である。能力認定試験も選別試験の一種である。ある水準の能力を持つと考えられる人をほかの人と区別するためのものである。たとえば「日本語能力試験」の1級は、それ自体が日本語能力認定試験であると同時に、大学入学のための日本語の試験を兼ねている。また、クラス分けのために期末テストの結果を使えば、そのテストは選別試験の面を持つことになる。一方、普通の学校での定期試験や毎日の小テストなどは非選別試験である。

選別試験にも二種類あって、選ばれる人数が決まっているものと、そうでないものがある。入学試験はその人数枠が厳しい。一方、日本語能力試験は合格者が何人いても構わない。ある水準に達すると見なされれば、皆合格になる。

また、選別試験の使い方も、得点の高いものから選ぶとは限らない。各クラスを平均化するために使う場合もある（Aクラスは1, 3, 5番…, Bクラスは2, 4, 6番…を集める）。

選別試験と非選別試験の大きな違いは、選別試験ではその目的から当然のこととして、受験者の間に何らかの適当な差が付かなければならないが、非選別試験ではそうではないという点にある。具体的に言えば、入学試験で平均点が95、ほとんどの受験者が90点以上、などという結果になっては困る。96点から上は合格で、95点では不合格、では批判が起こるだろう。その1点の差が、本当に受験者の学力の差を

反映しているかどうかは怪しいからである。また、クラス分けのための試験で半分以上が 100 点を取ってしまっても困る。しかし、教科書の各課ごとの復習テストではそのような結果になってもかまわないし、むしろ、なぜ 100 点を取れなかったかを考えねばならない。

一般に、「良いテストとは」とか、「テストをテストする」などの表現の中に現れる「テスト」は、多くの場合、人数の限られた選別試験の話である。そのような試験では、テストの結果として個人の能力差が正確に表れている必要がある。平均点 95 ではそのテストは「やさしすぎる」と言われる。100 点を取った受験者たち、つまり、このテストでは差が見られなかった受験者たちが、他のテストでは 90 点と 70 点に分かれるかもしれないのに、その差がこのテストでは測れなかったことになる。また、98 点と 97 点の「差」も疑わしい。同じ受験者が別のテストでそれぞれ 90 点と 80 点を取ったなら、その方が信頼できそうである（無論、そのテストのやり方によるので、一概にそうは言えないが）。

それに対して、日本語教師が日常的に行うのは非選別試験である場合が多い。学習したことをどの程度理解できたか、到達度を測るものである。その結果によって合格不合格を決めるわけではない。したがって、受験者の間で差が出る必要はない。毎日の漢字のクイズ、あるいは文法の復習テスト、教科書の課ごとのまとめのテスト、いわゆる期末試験なども、全員が 90 点以上取るならそれにこしたことはない。現実には、残念なことにさまざまな理由からそうならないというだけのことである（ただし、全員が楽に満点を取ってしまうようならば、そのコースの進度が遅すぎたのではないか、もっと早く進んでも学生はついて来られるのではないか、というコースに対する評価になってくる）。

上の二つのテストの違いは、テストの問題の形式、採点の方法など（「どのように」の問題）のさまざまな点に影響する。日本語教師はその違いをはっきり意識して、今自分がしようとしているテストはどんな性質のテストなのかということをよく考えることなしに、テストを作り、実施してはならない。