

プレースメントテスト

小川 貴士 ・ 嶽肩 志江

1. はじめに

国際基督教大学の日本語教育課程（以下 JLP）では、毎年 9 月と 4 月に日本語履修学生に対してプレースメントテストを実施している。そのテストの結果を基に、最も適当であると考えられるレベルにプレースする。対象となっているのは、全く日本語ゼロから始める留学生を除く、Japanese 2, 3, 4, 5, 6 それに並行する Intensive Japanese 1, 2, 3 及び上級日本語 1, 2 の計 7 レベルと、帰国日本人学生を対象とする Special Japanese A, B, C の 3 レベルである。

この稿では、JLP40 周年以降の 1994 年度から 2003 年度まで 10 年間のプレースメントテストの変遷を概観し、今後の課題と新たな取り組みを展望する。

2. プレースメントテストの意義と実施時期

アカデミックジャパニーズの 4 技能を総合的に伸ばすことを目的としている JLP では、入学時の対象学生の各技能を十分に把握することが肝要である。そのためプレースメントテストでは、3. で後述するように複数のテストを組み合わせて総合的な判断を行う。また、ICU のプログラムの進捗スケールに合致する測定ができるように、JLP で使用している教材に基づいた、またはプログラム内容を考慮したテスト問題が作成されている。

プレースメントテストはいわゆる 9 月生（外国人留学生、および日本人帰国学生）が入学する 9 月に通常行われる。また、学生派遣プログラムの開始時期の関係で、例えば 4 月入学の国費留学生や 4 月入学の国際基督教大学高等学校からの帰国生の入学生などについては、春学期授業登録日前の 4 月初めに実施している。

3. プレースメントテストの構成

3-1. 外国人留学生用プレースメントテスト

このテストには、4 つのタイプがある。Comprehensive Test（以下 Comprehensive）、Reading Writing and Vocabulary Test（以下 RWV）、Aural Comprehension Test（以下 Aural）、Interview Test（以下 Interview）である。Comprehensive と RWV は問題用紙に印刷された

もので、多肢選択式の問題をマークシートに解答する形で行っている。Aural は録音された音声聞いて答える多肢選択式の問題をマークシートに解答する形で行っている。一方、Interview は教員と学生の対一の面接方式で行われる。

3-1-1. Comprehensive Test

このテストは初級から中上級までの文法項目をカバーしており、構文や表現の知識がどのレベルにあるかを測ろうとするものである（単文形式の問題もあり、また会話の受け応えの中に問題が設定されているものもある）。また、同時に語彙力、読解力も測れるよう、新聞記事、雑誌記事や随筆、パンフレット類など様々なタイプの読み物から構成される部分もある。問題の指示は基本的に英語で示されているが、上級の力を持った学生を想定した読解問題では日本語の指示になっている。

3-1-2. Reading, Writing and Vocabulary Test

RWV は漢字の読み方の知識、反対語や類義語などの関連語の知識、熟語の意味の知識などを測るテストである。マークシート方式であるため、受験者が漢字を書く質問がないが、ある文脈の中で漢字が特定できる語を仮名書きにし、選択肢の中から正しい漢字表記を選ぶ問題が設けられている。問題の指示文は、日英両語の併記になっている。

3-1-3. Aural Comprehension Test

このテストは聴解力を測るためのものである。2 発話の簡単な会話文からやや長めの会話文までその内容について解答するもの、説明文の内容について解答するもの、簡単な講義形式の意見文の内容について解答するもので構成されている。指示文、問題ともすべて音声形式で行われるが、指示文は英語で示される。

3-1-4. Interview Test

学生の口頭表現能力を測るために、教員と学生の対一の面接形式で行われる 5 分間のテストである。JLP のシラバスに基づいた初級、中級、上級のプレースができるように質問項目が設けられている。先行して行われた Comprehensive、RWV などの結果を可能なかぎり手元に置きながら面接を行い、JLP のスケールの中での学生の位置付けを行う。

3-2. 日本人帰国学生用プレースメントテスト

このテストには3つのタイプがある。Special Japanese Placement Test (以下 SJPT)、論述テスト、漢字書き取りテストである。SJPT は問題用紙に印刷されたもので、多肢選択式の問題をマークシートに解答する形で行う。一方、論述テストと漢字書き取りテストは受験者が実際に記述する形式である。なおこれらのテストは定期的にある程度まとまった数の4月入学の一般日本人学生にも試験的に受けてもらい、帰国学生がこのグループのテストスコアと同等かそれ以上のスコアを出した場合は JLP コースを免除している。

3-2-1. Special Japanese Placement Test

漢字の知識、語彙力、読解力を総合的に測るテストである。漢字の正しい読み方を選ぶもの、ある文脈の中で漢字が特定できる語の仮名書きになっているものに適当な漢字を選ぶもの、四字熟語の中の正しい漢字を選ぶもの、手書きの手紙を読むもの、新聞記事や雑誌記事の内容について解答するものなどから成っている。

3-2-2. 論述テスト

具体的なテーマを与え(例えば、「今までに自分が受けてきた教育について述べてよ」など)、書き言葉における適当な漢字熟語や表現を使って書くことができるかどうかを測るものである。加えて、論の構成力・論理性や説得力も見る。当該年度に Special Japanese A, B, C を担当する3名以上の教員が読み、クラス分けを行う。

3-2-3. 漢字書き取りテスト

漢字は読めるが書けない、あるいは漢字は見ると大体意味が分かるが書けないという学生を把握するために行っているテストである。Special Japanese のコースで使っている漢字教材から取った12個の漢字の書き取りテストである。

4. プレースメントテストの変遷

外国人留学生用プレースメントテストについては、1994年から1998年まで Comprehensive、RWV、Aural の3種類のテストの組み合わせで行われた。しかし、90年代後半から、特に漢字圏出身の学生の傾向として、この3つのテストで比較的高いスコアを

出しながらも口頭表現能力が低い学生が上のレベルにプレースされる事例が少なからず現れ、プレースメントテストの見直しが行われた。その結果、1999年からはAuralを行わず、その代わりにInterviewを行うことで口頭表現能力を測り、全体的により適当なプレースメントが可能になった。Comprehensive、RWV、Interviewという組み合わせは現在まで続いている。

一方、日本人帰国学生用プレースメントテストは、1994年から1997年までSpecial Japanese Placement Test（当時は「Exemption Test（免除テスト）」と呼んでいた）のみの実施であったが、書き言葉の表現力についてクラス内でのかなりのばらつきが指摘されていたため、1998年から論述テストが加わり、さらに1999年から漢字書き取りテストも加わって、現在まで3種類のテストの組み合わせで総合的な判断を行っている。論述テストの採点は時間がかかる作業であるが、1997年以前より格段に的確なレベル分けが可能になり、その結果、クラス運営も効果的になった。

5. 分析と評価

5-1. プレースメント会議

前述の通り、例年9月が大規模なプレースメントテストの実施時期であるが、テスト実施翌日にプレースメント会議を行い、JLP全体で各学生のレベル分けの検討作業を行う。その際に、各学生の得点一覧、Comprehensive、RWV、Special Japanese Placement Testの基本統計量（平均値、最高点、最低点、標準誤差、中央値、最頻値、標準偏差など）と得点分布（ヒストグラム）といった統計資料、各学生の希望コース（JapaneseコースかIntensive Japaneseコースか）、レベル分けの判定基準、さらにInterview担当者のコメントや、サマーコースまたは前年度までのJLP履修生であればその時の担当者のコメントや成績なども検討材料としている。

プレースメント会議では学生のレベルを適格に判断し、適切なコースにプレースできるよう慎重に検討を行っているが、実際にコースに参加してみて他の学習者とのレベル差が明らかになるケースもある。その際には、コース担当者とJLP主任の判断のもと、できるだけ早い時期にコースの移動をするなどの措置を取っている。

5-2. プレースメントテスト改訂作業

9月からの秋学期が終了した後、プレースメントテスト実施手順やレベル分けが妥当なものであったか反省会を行うとともに、次年度のプレースメントテスト実施に向けて、各

テストの改訂作業が行われる。この作業は、1) 今年度テスト結果の統計資料の検討、2) 手直しまたは差し換えが必要な問題の検討、判定基準の妥当性の検討、3) JLP 全体での改訂案の検討と承認、という手順で行われる。1) 2)の作業は、Comprehensive、RWV、Interview、Special Japanese Placement Test の各班に分かれ、2～3名の担当者が行う。

5-2-1. Comprehensive、RWV、Special Japanese Placement Test の検討

1)で使用する統計資料には、Comprehensive、RWV、Special Japanese Placement Test 各テストの基礎統計量（学生数、平均値、標準偏差、最高点、最低点）とその得点分布、成績別にみた各設問の正答率・弁別指数・選択肢別回答者数とその割合）といった項目が含まれる。

この作業では主に弁別指数と選択肢別回答者数と割合に着目し、弁別力が低い設問や誰も選ばなかった選択肢については積極的に改めるようにしている。しかし、実際には毎年、ほとんどの受験者が入れ替わるわけで、対象が変わればその成績や弁別指数などが変化する可能性は大である。この点を考慮し、データ上のみならず、各設問や選択肢の中身を個別に検討し、改訂を加えるべきかあるいは据え置いて様子を見るか判断する必要がある。テストごとに担当班を構成することは、各テスト問題とその結果をていねいに検討していくうえでも有効であると考えられる。

5-2-2. Interview の検討

Interview では、JLP のシラバスに基づいた初級、中級、上級向けの質問項目が決められており、その質問に対する学生の発話を、発音、語彙、文法、テキストの型（長さ）、ストラテジーなどの面から総合的に判断する。この方法は1999年のInterview 導入以来、継続して行われてきているが、各担当者間の判断基準を統一するために、ある Interview の様子をビデオに録画し、JLP 全員でレベルチェックを試みるなどの試みを行っている。

また、今後は、各担当者による Interview 結果、プレースメント会議で実際にプレースされたコース、そしてその後のコース移動の有無などをデータ化し、比較・検討することで、より精度の高い Interview が行えるよう改善していくつもりである。

6. 今後の課題とそれに向けた取り組み

今後、JLP として取り組むべき課題として次の2点が挙げられる。

課題1) 複数バージョンのプレースメントテストの作成

現行のプレースメントテストは、入学時と1年間のJLPコース終了時の2回、実施している。1年という間隔が空くとは言うものの、同じ試験問題を2回解くため、記憶効果が生じている可能性は否定できない。

また、サマーコース履修生が、再度9月にプレースメントテストを受ける必要が生じる場合がある。現在は、サマーコースでは前年度のテスト問題でプレースメントテストを実施し、9月の新学期から改訂版を使用しているため、厳密には全く同じ試験問題とは言えないが、全く別バージョンのテストであるとも言えない。

このように、プログラム評価の観点からコース開始時と終了時の二回テストを行いたい場合、また短期間に二回プレースメントテストを受験しなければならない学生がいる場合には、同質の複数バージョンのテスト問題を持つ必要がある。

課題2) テスト結果の分析方法の検討

5-2-1.でも指摘したように、これまで10年以上の間使用してきたテスト結果の統計分析ソフトは、毎年、受験者が入れ替わることまで考慮に入れたテスト分析が行えるものではない。その点は、人間の目で細かく観察していくことで補ってきたが、複数年にわたるテスト問題の妥当性を検討するうえでは、問題があると言わざるを得ない。

また、従来統計分析ソフトを使用するためには、非常に古い機種のコピーを維持していかなければならず、それが困難な状況に来ている。

こうした事情を踏まえ、新たな統計分析ソフトの導入を検討している。

6-1. 項目反応理論 (Item Response Theory: IRT) に基づいたテスト分析

ICU 教育学科心理学専攻の栗山容子教授の協力を仰ぎ、上記の2つの課題を克服するために、これまでの古典的テスト理論に基づくテスト分析だけでなく、新しいテスト理論である項目反応理論(以下、IRT)に基づくテスト分析を導入していくことを検討している。

従来の古典的理論(を用いたテスト)の特徴とIRTを用いたテストの特徴を比較したのが表1である。

<表1>

古典的テスト理論 (を用いたテスト) の特徴	1) 観測値は真の得点(実際得点) + 誤差得点と定義される。(誤差得点というのは、予測不可能な要因) 2) 代表値や散布度(ばらつき)によって分布の特徴を把握する
------------------------------	---

	<ul style="list-style-type: none"> 3) わかりやすく実用的である 4) 受験者の条件（特性）が考慮に入られていない 5) テスト項目の特徴を示す困難度や識別力は、ある特定の集団に対して定義されている（対象が変わると比較できない） 6) 解答する項目が受験者間で異なる場合、比較はできない
IRT を用いたテストの特徴	<ul style="list-style-type: none"> 1) 複数のテスト間の結果の比較が容易である（異なる受験者が、異なる問題を、異なる日時に、異なる場所で受けたとしても比較が可能） 2) テストの測定精度を尺度のレベルごとに明示し、細かく確認できる（受験者個人ごとに測定の精度を評価することが可能。受験者の特性と問題の特徴（正答率）から判断） 3) テスト実施前に平均点をコントロールできる（問題を入れ替えても同じ平均点を維持したテストを作成することが可能） 4) テスト得点の対応表が作成できる（例えば昨年のテストの70点が今年のテストの何点に相応するかという比較が可能） 5) 受験者ごとに最適な問題を瞬時に選び、その場で出題することができる（コンピュータを使用したテストが実施できる場合、限られた時間でレベルを測るのに有効である）

表1の IRT を用いたテストの特徴1) からわかるように、今後、この理論を応用したテスト分析・作成を行っていくことで、毎年 of プレースメントテストの結果を比較し、その妥当性を検証していくことが可能になる。さらに、3)、4) の特徴を生かし、今後、様々な問題をデータベースによってプール、それらを各問題の特徴や出題パターンを考慮しながら組み合わせることで、等質のさまざまなバージョンのテストを作成することが可能になる。

もちろん IRT が全てのテストに応用できるわけではなく、クリアしなければいけない条件がある（表2）。条件の第一点目については問題がない。第二点目については、長文読解の問題などで独立性が失われる可能性があるが、過去の問題からその影響がどの程度であるかを把握し、問題作成時に考慮していく必要があるだろう。第三点目は、JLP のプレースメントテストの受験者数（外国人留学生、日本人帰国学生各 100 名以内）がサンプル数として十分なものであるかどうかは未知数である。実際に分析を行ってみて、データの出方から読み取っていく必要がある。

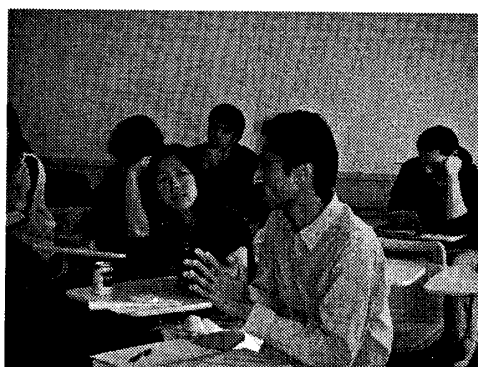
<表2>IRT 適用の条件

- ・テストが次元性を持っていること(問題が複数あっても目標が一つであること。例えば、「日本語能力を測る」ことのみが目標であればよい。)
- ・テスト項目が局所独立であること。(ある問題への解答が他の問題の解答に影響しない)
- ・パラメータの推定を行うために、ある程度まとまったサンプルが必要

現在、JLP では IRT に基づく分析ソフト(「BILOG-MG」)導入を目指して、試験的なデータ分析を始めている。この分析ソフトは、各問題の識別力、困難度、当て推量の3つのパラメータからその問題の適切さを評価するものであるが、今後、これまでのテスト結果を分析し、JLP プレースメントテストの特徴を客観的に把握し評価することで、より適切なレベル分けのできるテスト作成を目指して行きたい。

参考文献・資料

- 栗山容子(2003)「項目反応理論(IRT)に基づいたテスト分析の実際-日本語教育における、IRT を利用したテスト分析の可能性」2003年10月31日、JLP プレースメントテスト勉強会
- 豊田秀樹(2002)『項目反応理論 [入門編] -テストと測定の化学-』朝倉書店
- 渡辺直登・野口裕之(1999)『組織心理測定論-項目反応理論のフロンティア-』白桃書房



小川 貴士



嶽肩 志江

授業風景 (2003 年秋学期)