

# J-CATと既存のテストの関連付けの方法

今井 新悟 黒田 史彦

## 要 旨

J-CAT® (Japanese Computerized Adaptive Test) を早稲田大学においてプレースメントテストとして導入するにあたり、従来の紙によるプレースメント用テストとJ-CATの両テストを受験させた103人のデータをもとに両テストの関連付けを行った。両テストの総合得点間には統計的に有意な ( $P<0.01$ ) 強い相関関係 ( $r=0.903$ ) が認められた。この結果、両テストの併存妥当性が確認された。その後、J-CATの点数区分を8レベルに関係付ける作業を人数の分布に着目して行い、J-CATと従来のテストの対応表を作った。これにより以後、J-CATをレベル分けのプレースメントテストとして使えるようにできた。

【キーワード】 J-CAT プレースメントテスト 併存妥当性

## A Method of Association of J-CAT with Other Tests

IMAI Shingo, KURODA Fumihiko

【Abstract】 When Waseda University replaced its paper-based placement test for Japanese learners with the J-CAT (Japanese Computerized Adaptive Test), the two tests were compared and related to each other by using the scores of both tests taken by 103 students. We found a strong correlation between the two tests ( $r=0.903$ ), which is statistically significant ( $P<0.01$ ). This result affirmed the concurrent validity between the two tests. As a following step, we decided the cutting scores for eight levels based on the distributions of population and made a correspondence table, which enabled us to use J-CAT as a placement test for the replacement of the former paper test.

【Keywords】 J-CAT, placement test, concurrent validity

## 1. はじめに

J-CAT (Japanese Computerized Adaptive Test) は項目応答理論を用いた適応型テストであり、インターネット上で無料で受験できる。筑波大学留学生センターが平成22度から文部科学大臣認定「日本語・日本事情遠隔教育拠点」に指定されたことを受け、当該事業の一環として運用することになった。当該事業費でサーバーレンタル料、保守費用、事務人件費等を賄っている。また、開発・改良を継続しており、これには大学からの資金援助および各種競争資金等を得て充てている。<sup>1</sup>

J-CATは国内外で主に二つの用途で用いられている。一つは日本語学習者個人が自分の日本語能力を測定するためであり、もう一つは、教育機関におけるプレースメントテストである。

## 2. J-CATの概要

J-CATは文字・語彙、文法、読解、聴解の4セクションからなる、適応型（アダプティブな）テストである。受験者の能力に合った難易度の問題が、問題アイテムプールから選ばれて出題される。このアダプティブテストの仕組みを説明するには視力検査のメタファが適当である。視力検査ではまず、いくつか難易度を変えて探りを入れるように環や文字を指し示して、見えるか見えないかを確認する。その後、次第に被験者の視力でようやく見える程度と見えなくなるところの境を探す。そして、見えるもので最小の程度が視力と判定される。アダプティブテストも同じ原理で、まず大まかに受験者の能力を探るようにして出題する問題の難易度を上下させ、次第に受験者の能力でようやく解けそうな、つまり、能力に合った難易度の問題を提示することによって、能力を判定する。これにより、紙テストの場合のように無駄に能力に合わない問題に解答させることがなく、精度を上げながら受験時間の短縮も実現している。

また、項目応答理論を用いることにより、受験者集団や受験した問題項目によって成績が影響を受けないようにしている。従来の古典的テスト理論、いわゆる平均値や標準偏差を使うものでは、受験者の集団により点数が変わる。また、テストの難易度によっても点数が変わる。よって、異なる受験者集団同士の比較や、異なるテストセット間での比較ができない。つまり、受験者が成績として受け取る結果がそのたびに揺れるわけで、信頼性が確保できない。これらを解決するのが、項目応答理論による等化という方法である。J-CATはこれにより、成績の信頼性を確保している。

テスト終了と同時に結果が即座に示される。よって、リアルタイムで能力の測定ができる。成績証を保存および印刷できるとともに、団体受験の場合には、テスト監督者に受験者全員の成績一覧表が送付される。これは、Excel等で扱えるCSV形式のファイルなので、事後の分析用データとして利用できる。

成績証 Score report		J-CAT® Japanese Computerized Adaptive Test <a href="http://www.j-cat.org/">http://www.j-cat.org/</a>	
あなたのJ-CAT スコアは以下の通りです Your J-CAT scores are as follows.			
名前 / Name	鈴木 一朗	聴解 / Listening	61
所属 / Affiliation	Sakura University	語彙 / Vocabulary	52
生年月日 / Birth date	1983/08/15	文法 / Grammar	60
登録日 / Register date	2010/08/19	読解 / Reading	30
受験日 / Exam date	2010/08/19	合計 / Total	203

(watermark: 25b4f63e28f50c7a58a9031815205db0)  
J-CAT.org

Interpretation of J-CAT score		
J-CAT score	Proficiency Level	JLPT 日本語能力試験
- 100	Basic 初級	Level 4
101 - 150	Pre-Intermediate 中級前半	Level 3
151 - 200	Intermediate 中級	
201 - 250	Intermediate-high 中級後半	Level 2
251 - 300	Pre-Advanced 上級前半	Level 1
301 - 350	Advanced 上級	
350-	Near native 日本語母語話者相当	

- "Basic" learners can exchange basic ideas.
- "Intermediate" learners can manage daily communication.
- "Advanced" learners can manage academic and professional communication.

J-CAT.org

図1 成績証のサンプル

以下はJ-CATのトップ画面である。インストラクションは、日本語の他、英語、中国語、および全ひらがなで提供される。



図2 J-CATトップ画面（日本語）



图3 J-CATトップ画面（中国語）

### 3. 早稲田大学における従来のプレースメントとその課題

従来の早稲田大学プレースメントテストは紙に印刷されたものである。文法と会話のセクションで構成され、それぞれ初級、中級、上級の順に問題が並べられている。文法セクションは単文レベルの出題で、助詞、助詞相当語句、動詞のフォーム、その他文法項目について、その一部ないし全部を空所補充形式で解答させる。会話セクションは対話による場面設定を行い、会話的表現を中心として出題される。音声はなく、文法セクションと同様の空所補充形式で解答させる。テストの配点は、文法セクションは初級35点、中級25点、上級25点、会話は初級5点、中級5点、上級5点であり、合計100点満点となっている。

早稲田大学では、初級の1レベルから超上級の8レベルまでが設定されている。日本語科目のプレースメントテストを上記のペーパーテストで行っていたが、近年の留学生数の増加に伴い、従来の方法による実施の負担が増大していた。さらに渡日前にプレースメントテストを実施してクラス分けを行い、教室や講師の割り振りを決めたいという希望もあったが、各国で受験会場を確保してテストを行うということは現実的ではなかった。

そこで、インターネット上で受験でき、場所・時間が自由に設定できるJ-CATを導入することとなった。これにより、渡日前に学習者の能力が把握できるため、クラス割りなどの作業の効率化が期待された。学習者の能力については、既成のテスト、例えば日本語能力試験などの利用も検討されたが、必ずしも全員が受けていないこと、また、受験時と来日時の時間的なずれのため、来日時の能力を反映していないことが予想された。なるべく来日時に近い時期に受験できて、来日時の能力を反映したものが求められた。これらの条件を満たすものは、J-CAT以外に見当たらなかった。以上の理由で2010年からプレースメントテストとして従来のペーパーテストに代えてJ-CATを導入するべく準備を開始した。特に、J-CATの点数を使ってどのようにして早稲田大学におけるレベル分けを行うべきかについて検討が行われた。

### 4. J-CATの点数とレベル分け

J-CATの点数は、各セクション100点満点で、計400満点のワンスケールで与えられる。よって、点数そのものは、能力の程度を表しているが、それぞれの点数が教育機関でどのレベルなのかということは一概には決められない。そもそも、レベルの分け方が教育機関ごとに異なっている。よって、各機関でJ-CATをプレースメントテストとして導入する場合、何点から何点までをどのレベルとするかは各機関で判断しなくてはならない。例えば、すでにレベルの分かっている学習者にJ-CATを受験させて、その結果と受験者のレベルを対応させることができる。しかし、この方法の場合、受験時のレベルが分かっていると異なる。つまり、別途レベル分けの基準を設けて、その基準に従ってレベル分けが済

んでいなくてはならない。J-CATの点数はア・プリアリに教育機関のレベルを示すことはできない。J-CATの点数と当該教育機関のレベル分けの基準の関連付けの作業を行わないといけない。以下では、そのケーススタディとして、早稲田大学の例を検討する。

## 5. 方法と結果

早稲田大学では、J-CATをプレースメントテストとして導入した際、従来の紙によるプレースメント用テストとJ-CATの関連付けを行うため、両テストを104人に受験させた。テストを完了しなかった1人を除く103人のデータをもとに分析を行った。

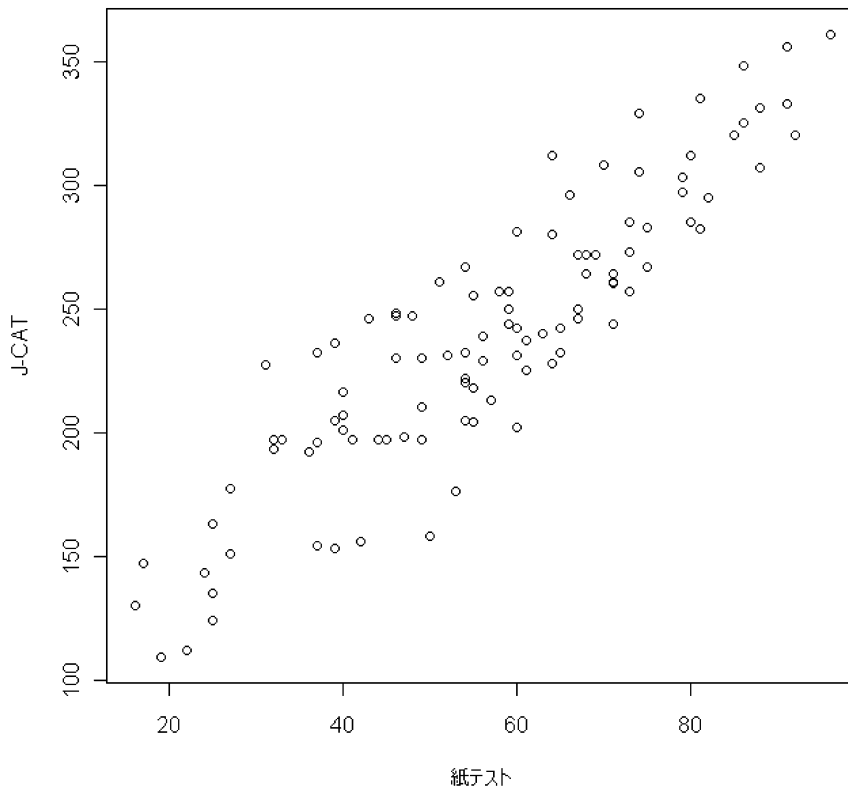


図4 J-CAT総合得点と早稲田テスト（紙テスト）総合得点の相関図

両テストの総合得点間（ペーパーテストは100点満点、J-CATは400点満点）には統計的に有意な（ $P < 0.01$ ）強い相関関係（ $r = 0.903$ ）が認められた。（分析には統計解析ソフトウェアRの2.12.0版を使用した。）

この結果は併存妥当性の検証として採用でき、妥当性係数として0.903が得られたこととなり、両テストの併存妥当性が確認された。これによりレベル分けの指標としてJ-CATの総合点を使うことの妥当性が確認された。

なお、両テストで共通するセクションである「文法」間の相関は $r=0.813$  ( $P<0.01$ ) であり、やはり強い相関が認められた。しかしながら、両テストの総合点間の相関に比べるとわずかながら低くなっている。これは、両テストにおいて、「文法」と呼ばれる共通のセクションはあるものの、その中で測定している構成概念が少しずれていることを示唆している。もちろん、異なるテストであるので、構成概念の違いがあっても当然ではある。しかし、今回の目的である関連付けにおいては、両テストの併存妥当性がより高くなる総合得点を採用すべきであると判断される。

## 6. レベル分けの方法

両テストの総合得点の相関が高いことが確認されたので、続いてJ-CATの点数区分を8レベルに関係づける作業を行った。従来のプレースメントテストでのレベル分けをした場合の結果を参照しながら、人数の分布に着目してカッティングスコアを調整した。その結果、レベルの最も低い区分と最も高い区分では点数の幅を広く取り、それ以外は30点の区分を設定するのが適当であると判断された。レベル1では0点から120点まで120点の幅があるが、これはJ-CATが四肢択一形式であることから、まぐれ当たりが25%の割合で発生する可能性があることの影響である。さらにレベル2も50点の幅があり、他のレベルより広い。これも、能力が下の場合、まぐれ当たりの可能性の影響を受けて、精度が甘くなっていることによると考えられる。また、レベル8の幅が広いのは、それが最上のレベルであり、いわば能力は青天井になっているためであろう。以上の考察を踏まえれば、J-CATが概ね間隔尺度として機能していることも示唆される。

以上により、J-CATの点数による8レベルのクラス分けの基準を作成し、従来のテストに代えてJ-CATを使ったプレースメントテストが実施できるようになった。

表1 J-CATと早稲田プレースメントテストの比較とレベル分けの基準

レベル	J-CAT 点数 区分	人 数	J-CAT						早稲田テスト	
			合計平均	標準偏差	聴解平均	語彙平均	文法平均	読解平均	合計平均	標準偏差
1	0-120	2	110.5	2.1	42.5	23.0	17.0	28.0	20.5	2.1
2	121-170	11	146.7	12.4	44.7	38.2	31.5	32.3	29.7	10.8
3	171-200	12	192.8	7.8	54.8	50.8	39.5	47.8	39.7	8.0
4	201-230	18	216.2	10.6	65.5	51.4	46.0	53.3	50.2	9.1
5	231-260	25	244.1	9.0	70.7	55.7	53.8	64.0	57.4	9.6
6	261-290	16	273.1	8.6	80.7	60.4	58.1	73.8	68.8	8.3
7	291-320	11	306.8	8.8	84.4	75.6	64.7	82.1	78.1	8.9
8	321-400	8	339.8	13.4	89.8	80.3	83.5	86.3	86.6	6.8

## 7. まとめ

本稿では、ケーススタディとして1例のみを取り上げた。同様の手順は他の教育機関においてもJ-CATを従来のプレースメントテストに代えて使う場合には必要となる。ただし、従来のプレースメントテスト自体に信頼性がない場合、この手順は使えない。例えば、従来のプレースメントテストを毎年新たに作成していた場合、ある特定の年度の結果がプレースメントテストとして正しく機能したという保証がない。それは、毎年異なったテストを行うことにより、その困難度が揺れるからである。現場ではそういう場合に、テストの成績よりも、その後のクラスパフォーマンスや教師の判断でクラスの変更を認める場合が多いだろう。だとすれば、そのようなテストの点数とJ-CATの点数の関連付けをそのまま行ったとしても、J-CATの点数によるレベル分けも信頼できるものにはならない。そういう場合には、関連付けを行う前に、従来のプレースメントテストのレベル分けの点数の調整を行う必要がある。実際にはレベル分けの区切りの点数を変更したとしてもそれに伴って学生のクラスを移動させることはしないだろうが、仮に、そうすることが許されると仮定して、本来ならばどの点数で線引きすべきであったかを求め、その後J-CATとの関連付けを行うべきである。この際、異なる年度のテスト成績を混ぜてはならない。受験者数が少ないからという理由で、複数年度の成績を混ぜてしまうと、異なる困難度のテストが混ざってしまい、関連付けが正しく行われぬ。

一方、毎年同じテストを使っていてカッティングスコアの信頼性が確保されている場合には、そのままJ-CATとの関連付けが行える。本稿で扱った早稲田大学の例がそれである。ただし、その際には、同じ受験者が両テストを受けることが条件であり、また、両テストを時間を空けずに受験することが理想である。本稿で扱った例はこの条件も満たしている。さらに実験計画法の観点からすれば、受験者集団を半分に分けて、両テストを与える順番を変えることにより、疲れによる影響などのカウンターバランスを取ることができればさらに望ましい。

以上の条件を満たしたとしても、両テストの結果の相関が、本稿の例のように強い相関を示さない場合もあるかもしれない。その原因は、両テストが測っているもの、つまり構成概念が異なる可能性が高い。本稿の例では、「文法」の相関が「総合得点」の相関に比べてやや劣ることを示した。それでも今回の例では「文法」においてもそれほど相関が低くはなっていないので、即座にそれが問題になることはないが、もし、両テストの相関が相当程度に低い場合には、両テストの関連付けを行うことに慎重にならなくてはならない。その場合には別の方法も考えるべきだろう。こういうケースでは、過去のプレースメントテストは一切使わず、J-CATの成績のみでレベル分けを一度行い、その後、教師の意見などを聞きながら各自のレベルの調整を行い、それに従ってJ-CATにおける各レベルのカッティングスコアを決めるという方法が適切であろう。ただし、そもそもJ-CATの点数を使



うべきなのかどうかということは事前に考察されるべきである。J-CATの構成概念と授業とが著しく異なっている場合には、慎重な判断が必要である。例えば、その授業が会話である場合、現行のJ-CATには会話テストはないのであるから、J-CATの点数をもって、会話能力の推定値としてよいかどうかの判断が必要になる。例えば、会話能力と聴解テストの相関は高いという報告もあるので、それを根拠に、総合得点よりも聴解セクションの得点で会話クラスのレベル分けを行うなどの判断が求められる。そして、この場合も事後の調整が必須である。

以上、J-CATを各教育機関でプレースメントテストとして利用する場合の注意点について述べた。最後に以下のことを繰り返し強調しておきたい。J-CATの点数がそれぞれの教育機関においてどのレベルに対応しているかという質問は意味を持たない。J-CATの点数は、初級から上級までのワンスケールの点数の並びである。何点をカッティングスコアとして設定して、いくつにレベル分けをするかは各教育機関の実態に則して、個々に決められるべきものである。

## 注

1. 2012年のシステム強化にあたって早稲田大学にも費用を分担していただいた。

## 参考文献

今井新悟・伊東祐郎・中村洋一・菊地賢一・赤木彌生・中園博美・本田明子 (2010)

『J-CAT Japanese Computerized Adaptive Test—日本語能力をコンピュータで測る—』

山口大学留学生センター

J-CAT URL <http://www.j-cat.org>

## 謝辞

J-CATは実にたくさんの人・機関の協力を得て今日を迎えている。当初の開発の中心メンバーは著者を含む以下の各氏であり、御礼申し上げる。なお、所属は現在のものである。伊東祐郎（東京外国語大学）、中村洋一（清泉女学院短期大学）、菊地賢一（東邦大学）、赤木彌生（山口大学）、中園博美（島根大学）、本田明子（立命館アジア太平洋大学）、平村健勝（アクセンチュア）、浅田岐依（フリー）。筑波大学の日本語・日本事情遠隔教育拠点に運用を移管してからは、李在鎬、石川浩一郎、古川雅子、足立麻里子の各氏に特にお世話になっており、感謝する。プレテストでは世界中の学習者・機関に協力を得ているが、そのすべてを列挙すると相当の紙幅を費やすので、申し訳ないが割愛させていただく。システムが稼動してからもたくさんの方々モニターしてもらい、そのフィードバックを得て、改良が進んだ。現在、年間約8000人の受験者があるが、そのすべてが有益なデータ

を提供してくれるので、すべての受験者にも感謝申し上げます。