

漢検のテスト特性調査

～ 項目応答理論による分析～

平成19年 1月11日

財団法人 日本漢字能力検定協会

調査のねらい

「検定試験制度」は、受検者の能力が、定められた審査基準に適合しているかを判定する試験制度である。したがって受検する年や月が異なっても、受検者の能力に変化がなければ検定試験の結果は同じでなければならない。同じ受検者に異なる評価が提示されることはあってはならないのだ。したがって、難易度や審査基準等が一定した、品質の高い検定試験を永続的に提供し続けることが検定実施団体に課せられた使命であり、受検者や社会に対する責任である。

その使命を遂行するため、漢検では、テスト実施後に結果を分析し、その平均点や偏差値、問題ごとの正答率などを算出している。問題作成の過程においては、これらのデータを参考にして検討を重ね、毎回のテスト難易度の平準化に努めている。このような分析手法は「古典的テスト分析」と呼ばれ、これまでも多くのテストで用いられてきたものである。

このたび漢検では「古典的テスト分析」を行うと同時に、受検者集団に依存せずにテストの項目(1問1問の問題)やテストの特性に関するデータの科学的分析が可能とされている「項目応答理論」による分析をすることによって、更にテスト品質を高める試みに着手した。本レポートでは、この項目応答理論を、平成17年度に実施した漢検のテスト結果に適用し、その分析結果から、漢検のテスト特性について考察する。

項目応答理論：受検者の能力推定値や項目(1問1問の問題)の難易度を、受検者集団とは独立して絶対評価として求めようとする理論。

調査対象

平成17年度第1回 日本漢字能力検定 2級・準2級・3級
平成17年度第2回 日本漢字能力検定 2級・準2級・3級
平成17年度第3回 日本漢字能力検定 2級・準2級・3級

受検者数(単位:人)

	第1回	第2回	第3回
2級	23,888	20,326	25,818
準2級	16,106	13,905	15,507
3級	16,379	16,049	15,135

2級・準2級・3級を調査対象とした理由

- ・項目応答理論による分析のために必要な受検者数(概ね1万人以上と言われている)を満たしている。
- ・2級・準2級・3級は、多くの高等学校や大学において、入試優遇や単位認定の評価基準として採用されており、社会的に認知されている。

調査方法

上記調査対象のテスト結果を項目応答理論によって分析し、「項目パラメータ」と「能力推定値(シート)」を算出し、200点満点の得点(従来の正答数に基づく加点方式)と、項目応答理論により算出さ

れた受検者の能力推定値 との相関係数を計算した。

その後、同じ級の複数テスト(例、第1回2級&第2回2級&第3回2級)のテスト特性曲線同士、およびテスト情報曲線同士を重ね合わせて描画し、それぞれのテスト特性について比較し、考察を行った。(テスト特性曲線やテスト情報曲線は、テスト特性を視覚的に考察するために描画するグラフである。)

なお、分析に際しては、株式会社日本統計事務センターが開発したテスト分析ツール「iTEMS」(integrated TEst Management System: アイテムズ)を使用した。

「iTEMS」integrated TEst Management System: アイテムズ

項目応答理論によるテスト分析(項目パラメータ算出、能力推定値算出、テスト曲線描画等)だけでなく、テストフォーム作成やスコア計算、成績レポート作成が可能なテスト統合管理ツール。

用語の解説

「相関係数」

2つのデータ列の間の相関(類似性の度合い)を示す統計学的指標。-1から1の間の実数値で表される。1に近ければ「正の相関がある」、-1に近ければ「負の相関がある」と言い、下表のような解釈をすることができる。

相関係数	解釈
-1.0 ~ -0.7	強い負の相関関係がある
-0.7 ~ -0.4	かなりの負の相関関係がある
-0.4 ~ -0.2	やや負の相関関係がある
-0.2 ~ 0.2	ほとんど相関関係がない
0.2 ~ 0.4	やや正の相関関係がある
0.4 ~ 0.7	かなりの正の相関関係がある
0.7 ~ 1.0	強い正の相関関係がある

「項目パラメータ」

テストを構成する1問1問の問題を「項目」と呼び、項目応答理論によって導き出される、この項目の難易度等の特性を示す値を項目パラメータと呼ぶ。よって項目パラメータの値は項目ごとに異なる。

「(シータ)」

項目応答理論によって導き出される受検者の能力推定値。概ね -4から4の間の実数値で示され、4に近い数値ほど能力が高いことを示す。

「テスト特性曲線」

ある能力推定値 の受検者であればどの程度の正答数が期待できるかを示すグラフで、そのテストの特性を示すと考えられる。ある能力推定値 の受検者が正答する確率を項目ごとに算出し、それをテストの項目の数だけ縦軸に加算したもの。分析の際、能力判定に不適切だと思われる項目があればそのデータは削除するため、漢検では通常1テストは120の項目で構成されているが、本レポートのテスト特性曲線の縦軸は110までとなっている。

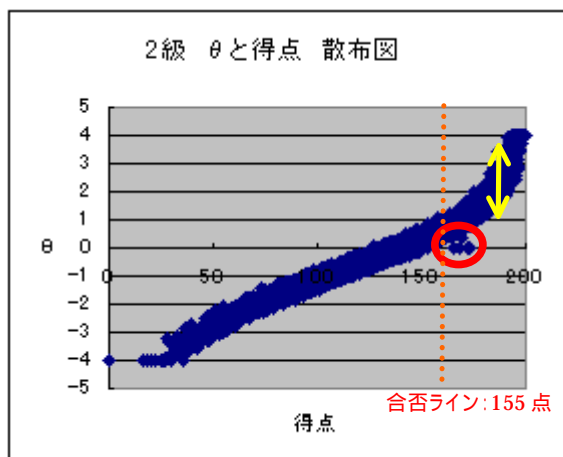
「テスト情報曲線」

どの程度の能力推定値 の受検者を測るのにふさわしいテストであるかを示すグラフ。グラフの頂点付近が最もテスト情報量が多く、そのあたりに該当する能力推定値の誤差が最も小さいとみなすことができる。言い換えれば、頂点付近における能力推定値はより信頼性が高いものになる。

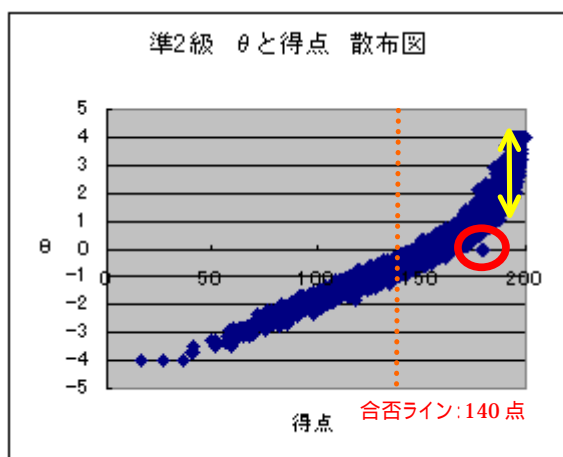
調査結果

1. 得点(200点満点)と、項目応答理論による能力推定値 の相関調査結果
(以下の表とグラフは平成17年度第1回の調査結果である)

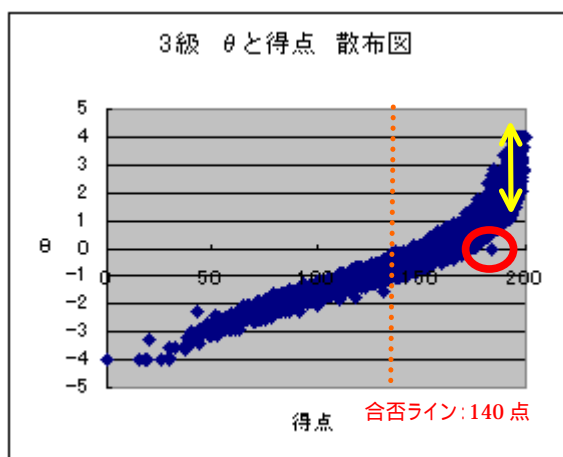
	2級	準2級	3級
相関係数	0.974	0.952	0.928



相関係数:0.974



相関係数:0.952



相関係数:0.928

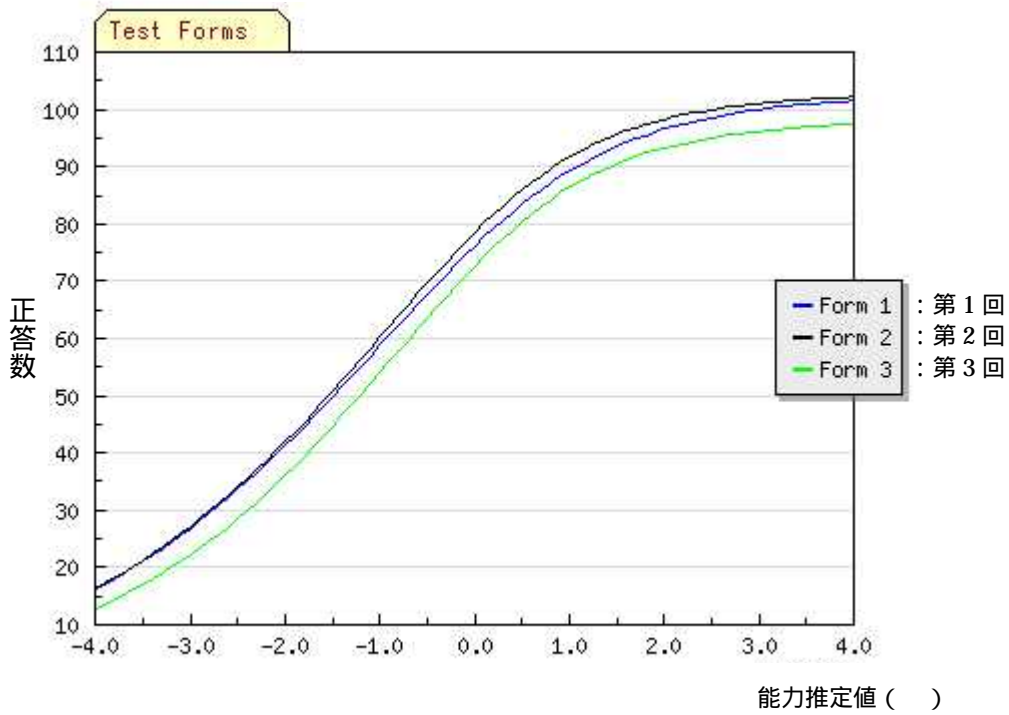
いずれも平成17年度第1回

考察:

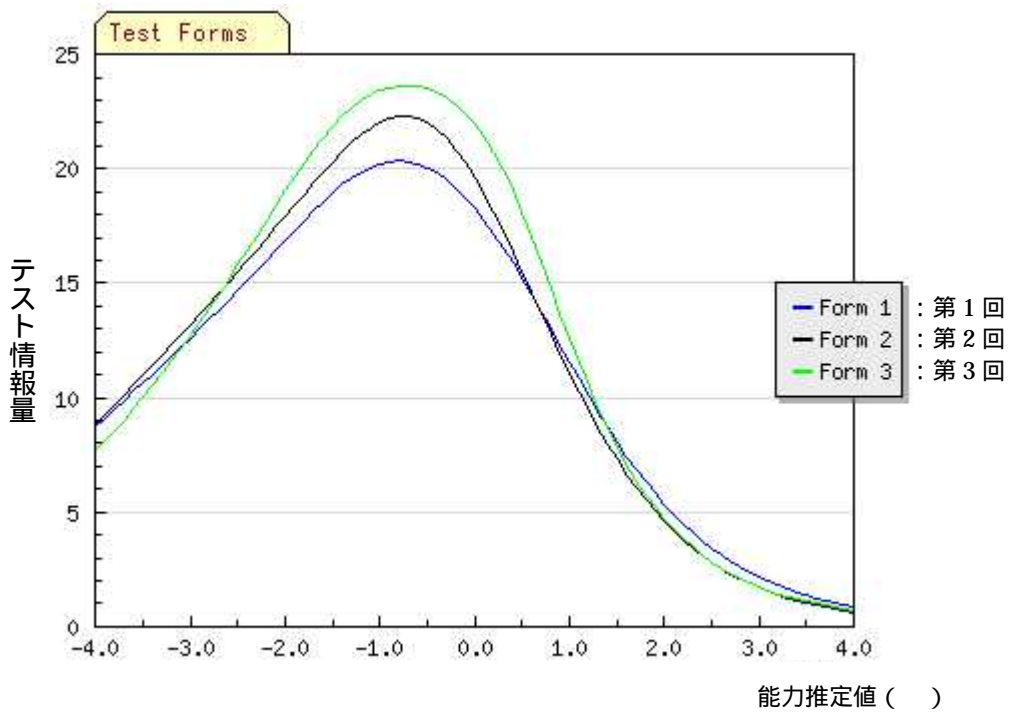
すべての級において、得点が180点から200点付近の受検者(高得点での合格者)層における の幅が広がっている(グラフの黄色い矢印の部分。同じ得点でも の差が大きい)点や、 が0付近での集団右側に「外れ値」が各級に観察されている点(グラフの赤い丸の部分。 が0付近でも、高い得点を得た受検者がいる)については、その原因の調査が必要だと考えるが、「得点」と項目応答理論分析により算出された「能力推定値」は、極めて高い相関を示している(相関係数が1に近い)。つまり、従来の加点方式による200点満点での得点においても、項目応答理論による能力推定値とほぼ同程度に受検者の能力を精度高く測ることができていたと考えられる。

2. テスト特性曲線とテスト情報曲線の考察

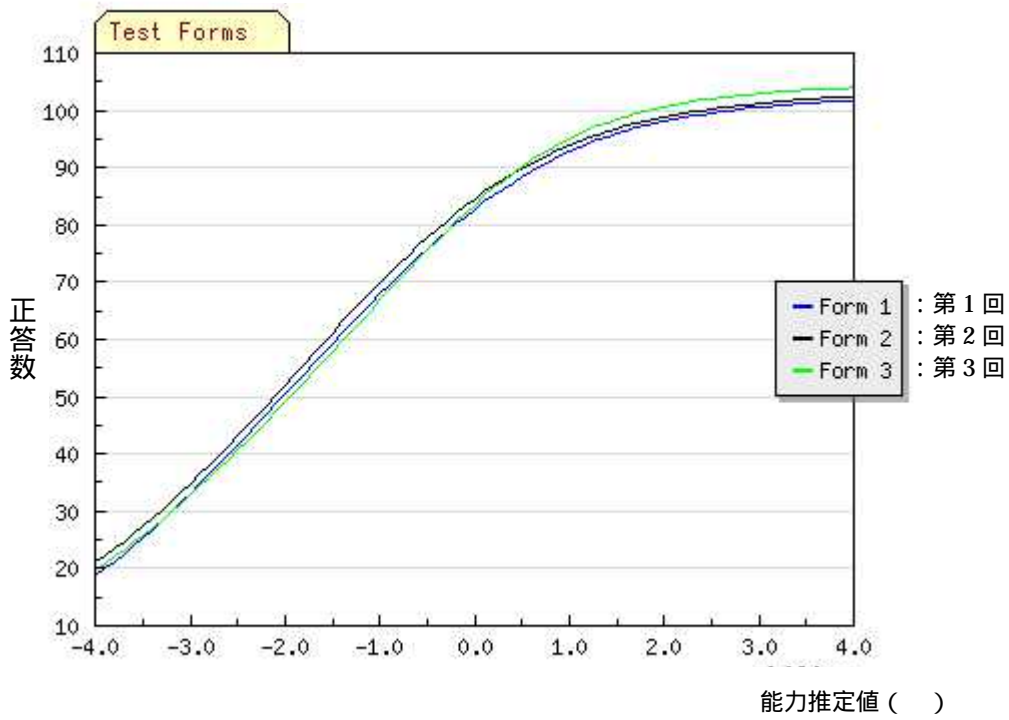
2級 テスト特性曲線



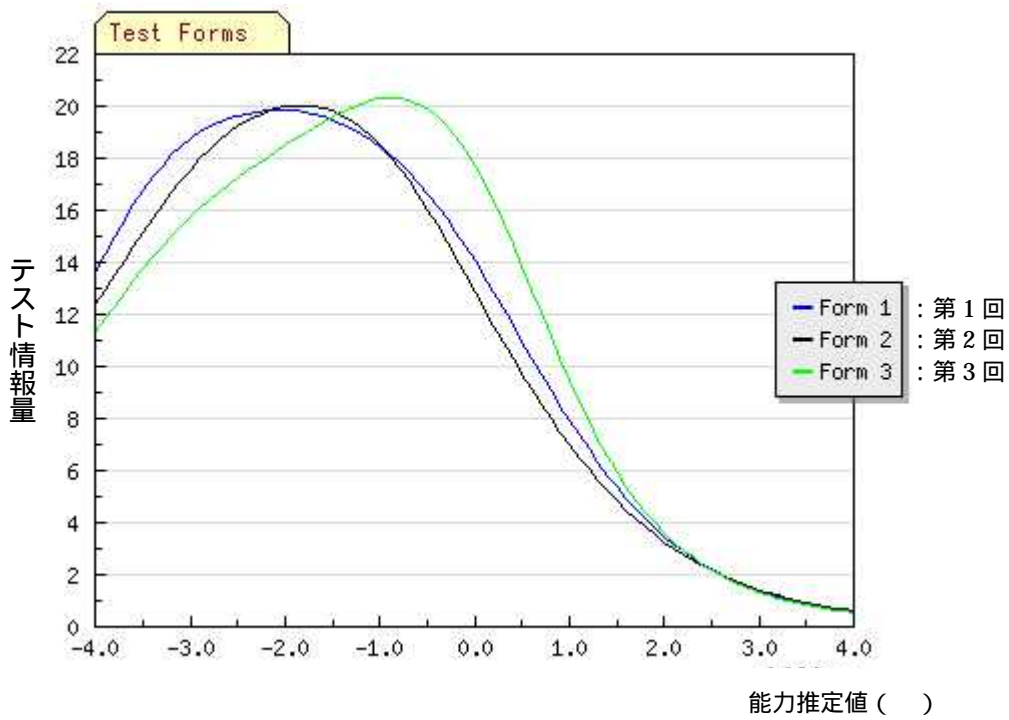
2級 テスト情報曲線



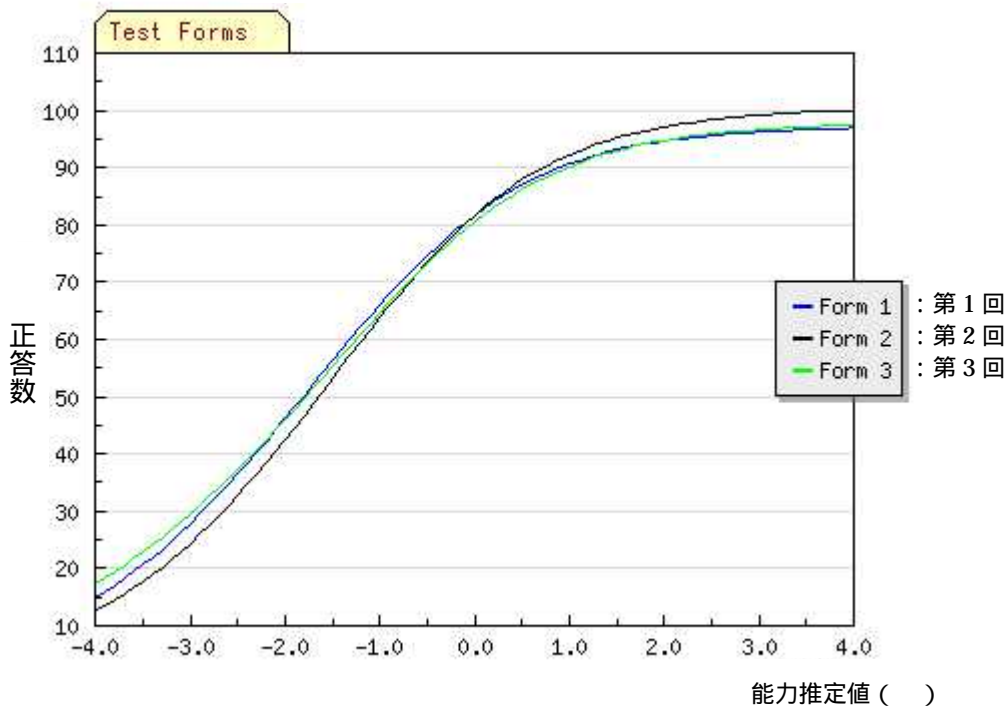
準2級 テスト特性曲線



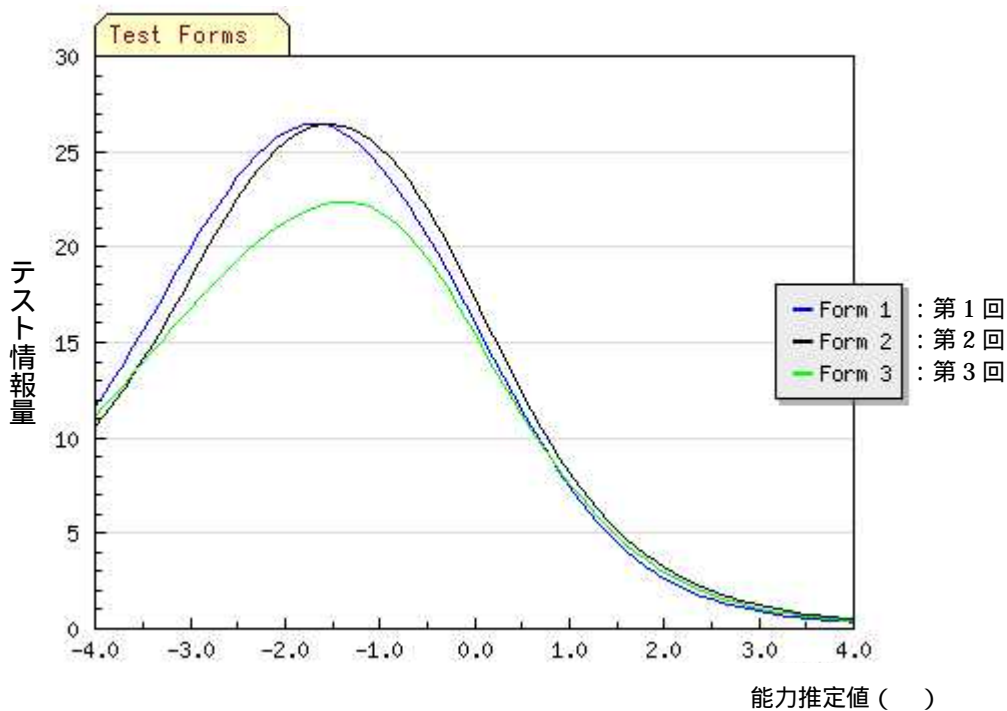
準2級 テスト情報曲線



3級 テスト特性曲線



3級 テスト情報曲線



考察:

テスト特性曲線からの考察: ほぼ同様の曲線を描いていることから、複数テスト間においても、テスト特性はほぼ同じであったと考えられる。

テスト情報曲線からの考察: 2級と3級とはほぼ同様の曲線を示しているが、準2級に関しては、第3回のみ曲線の頂点が右にずれている。これは、能力がやや高い受検者にとってはテスト情報量の多い(誤

差の少ない)項目が多く含まれていたテストであった、ということの意味する。また合否を判定する検定試験である漢検では、合否判定ラインがテスト情報曲線の頂点付近になることが理想であるといえる。

最後に

今回の調査では、テスト情報曲線における頂点のずれ、頂点付近の値と合否ラインの関係等、研究を進めていくべき課題が見えた。

その一方で、得点は値との間に極めて高い相関関係が認められた。調査にあたって、「強い正の相関関係がある」とされる、相関係数:0.7 を超える数値を目標値としていたが、結果はすべて0.9以上という、目標値を大きく上回るものであった。また、テスト特性曲線の観点からは、それぞれのテストがほぼ同一の特性を持っていると推測される分析結果が得られた。これらのことは、項目応答理論という科学的分析によっても、漢検のテスト品質が高いレベルで維持され続けていることを裏付ける結果だと考えている。今後もより品質の高い、受検者に信頼されるテストにするために更に詳細な研究を進めていきたい。

以上