

一般知的能力検査における紙筆版と CBT 版の項目特性の比較

○ 前田 純子 藤田 彩子 舩田 博之
HRR 株式会社

1. 研究の背景

近年、コンピュータ技術の発達により、従来からの紙筆版テストの CBT 化が急速に進みつつある。紙筆版テストから CBT 版への変化は、単なる媒体の変化にとどまらず、項目提示方法や回答時間の管理方法などテスト実施方式の違いによって、項目特性に差が生じる可能性が考えられる。熊谷 (2002) は、特定の問題領域に関して紙筆版と CBT 版で同等性を確認しているが、Puhan&Boughton (2004) は、一部の問題領域に差が生じたがその原因を特定できないことを報告しており、CBT 化の影響についてははっきりとした結果が得られていない。本研究では、一般知的能力検査において紙筆版と CBT 版の両方で使用される項目の特性を比較し、テスト実施方式によって違いが見られるかを確認する。

2. 方法

(1) 使用した一般知的能力検査

本研究では HRR 株式会社の基礎能力検査 GAT (General Ability Test) の紙筆版と CBT 版を用いた。GAT は言語検査 (語の意味の把握、文章の構成や要旨の理解) と非言語検査 (数的処理、論理的思考) で構成されている。

(2) 手続き

I. 項目特性値の算出

BILOG-MG を使用し IRT の 2 パラメタ・モデルにおける項目特性値を算出した。BILOG-MG では被験者集団の θ 分布を想定して項目特性値を算出することが可能なため、標準化され等化済みである紙筆版 GAT の標準得点を使用して被験者集団の事前分布を設定し、項目特性値の算出を行った。これにより紙筆版・CBT 版の項目特性値の共通尺度上での比較が可能となる。

1) CBT 版

2004 年 2 月から 5 月までに紙筆版と CBT 版の両方を受験した 32039 人のデータを用いた。CBT 版は紙筆版の項目を用いて項目プールを構成しているが、適応型テストであるため被験者毎に出題項目が異なり、紙筆版と全く同じ項目セットが出題されることはない。CBT 版の項目特性値を算出する際には、被験者約 3 万人の紙筆版の標準得点を元に θ の事前分布を設定した。

2) 紙筆版

CBT 版で使用されている項目は紙筆版の 20 版に散在している。そのため版毎に約 2000~8000 名の回答データを用いて、当該項目の項目特性値を算出した。このとき、各版の被験者集団の θ 分布が微妙に異なるため、標準得点を元に版毎の θ の事前分布を設定した。

II. 項目特性の比較

紙筆版で算出した項目特性値と CBT 版で算出した項目特性値に違いが見られるならば、その項目は紙筆版と CBT 版の間で項目バイアスがあると言える。

項目特性値を比較するには、項目特性値の推定誤差を考慮した標準バイアス指数を使用する方法があるが、サンプル数が大きい場合に過剰に差異ありと見なされるおそれがあるため、本研究では 2 つの項目特性曲線に挟まれた面積を計算するという総面積法 (total-area method) を用いた。この方法では、項目特性が異なるときには、2 つの項目特性曲線に挟まれた面積 Δ が大きくなる。本研究では Δ が 0.5 以上の場合に項目バイアスがあると判断した。

3. 結果および考察

表 1 に各検査の Δ の値が 0.5 以上の項目数の全体に占める割合を示す。言語・非言語検査ともに 7 割以上で項目バイアスがあるという結果が得られた。また、各検査を構成している下位領域でも、全ての領域において過半数の項目にバイアスが認められた。言語検査では特に「並べ換え」と「文章理

解」の領域において、ほとんどの項目にバイアスが確認され、大半の項目に顕著な差 ($\Delta \geq 1.0$) が見られた。この 2 つの領域は他の領域と比較すると表示される文字数が多く、画面で文章を読みながら問題を解くことの困難さや画面上に書き込みができない不便などを、特に反映しているものと思われる。非言語検査においては、どの領域においてもバイアスが見られる項目が 6 割以上であった。

表1: $\Delta \geq 0.5$ の項目の割合

		言語《291》	0.70(0.26)
		非言語《151》	0.74(0.20)
言語検査	並べ換え《20》	1.00(0.85)	
	文章理解《27》	0.89(0.59)	
	語の関係把握《62》	0.68(0.24)	
	文法《64》	0.75(0.20)	
	語彙《118》	0.59(0.14)	
非言語検査	推論《38》	0.79(0.26)	
	数量《39》	0.64(0.05)	
	表の計算《27》	0.78(0.26)	
	確率・組合せ《47》	0.74(0.23)	

※1《 》内は項目数

※2《 》内は $\Delta \geq 1.0$ の割合

表2: CBT版と紙筆版の項目特性値の平均と差

a値=識別力
b値=困難度

		CBT		紙筆		差(CBT-紙筆)	
		a値	b値	a値	b値	a値	b値
言語検査	並べ換え《20》	0.50	-0.14	0.44	-1.41	0.06	1.27
	文章理解《27》	0.49	-1.13	0.52	-1.63	-0.03	0.50
	語の関係把握《62》	0.63	-0.81	0.53	-1.44	0.10	0.63
	文法《64》	0.61	-0.67	0.53	-1.36	0.08	0.69
	語彙《118》	0.58	-0.09	0.55	-0.43	0.03	0.34
非言語検査	推論《38》	0.61	0.92	0.54	0.20	0.07	0.72
	数量《39》	1.00	0.38	0.74	-0.14	0.26	0.52
	表の計算《27》	1.02	0.17	0.75	-0.50	0.27	0.67
	確率・組合せ《47》	0.87	-0.05	0.71	-0.85	0.16	0.80

※《 》内は項目数

更に詳細に確認するために、CBT版と紙筆版の特性値の平均とその差を示したのが表2である。言語検査の中では特に、「並べ換え」領域において困難度 b に 1.27 と顕著な差が見られた。これは「ばらばらの文節を正しく並べ換える」という領域のため、画面上で問題を解くことの難しさをより反映しているものと思われる。

非言語では全ての領域で、CBT版の方が困難度 b において 0.5 以上高くなっていた。これは、画面上で問題を解くことの難しさだけでなく、特性値を算出する際の無答の扱い方の違いが一因であると思われる。紙筆版では、無答は特性値算出には用いず、回答があった項目のみを使用して算出しているが、CBT版では無答は誤答として扱い、特性値算出に用いている。ここで、無答の出現の仕方については次のことがわかっている。紙筆版はテスト全体で制限時間を設けており、被験者は最初から順番に解くことや易しいと思われる問題から解くことが多く、言語検査ではほとんどの被験者が制限時間内に回答が完了するが、非言語検査では制限時間内に回答が完了しない割合が高いことが確認されており、最後の方に配置されていたり難度が高い問題は無答になる可能性が相対的に高くなる。この場合、能力の高い被験者の回答によって項目特性値が算出され、b 値が低くなる可能性が高い。以上のように、紙筆版の非言語検査では、制限時間内に回答が完了できず無答になった項目の b 値が低くなる可能性がある。それに対して、CBT版では 1 問毎に制限時間を設けており、制限時間内に回答できなかった場合には誤答となるため、b 値が高く算出される可能性がある。

非言語検査の中で、「数量」と「表の計算」領域は b 値の高さに加えて a 値が若干高く出ている。この 2 つ領域では、ある程度の時間をかけると解くことが可能な項目が多い。CBT版では 1 問毎に制限時間があるため、時間内に解くことができない被験者が比較的多く発生し、個人差がより明確になった影響ではないかと推測される。

4. 今後の課題

紙筆版と CBT 版では、項目特性が変化することが確認され、その原因についていくつかの観点で考察を行った。更に原因を明らかにするためには、制限時間の設定を変更する、制限時間を設けない、領域を特定して実施を行うなど、種々の条件を統制して確認する必要がある。

また、項目特性が変化することを考慮した上で、紙筆版と CBT 版での個人特性 θ を比較することにより、テストとしての同等性を詳細に確認することも必要である。

《参考文献》

- 熊谷龍一 2002 語彙理解尺度における CBT 版と紙筆版の同等性の検証—項目反応理論によるテスト作成・分析を通じた検討— 名古屋大学大学院教育発達化学研究科紀要 (心理発達科学), 第 49 巻, 47-54.
- Puhan, G., & Boughton, K.A. 2004 Evaluating the Comparability of Paper and Pencil Versus Computerized Versions of a Large-Scale Certification Test ETS
- Ronald K. Hambleton 野口裕之 訳 1992 第 4 章 項目応答理論の基礎と応用 池田央/藤田恵麗/柳井晴夫/繁樹算男 (編) 教育測定学原著第 3 版 C.S.L 学習評価研究所