

新しい知能観と測定技術を踏まえた 個別式知能検査の開発可能性

筑波大学心理学系 海保 博之

宇都宮大学教育学部 服部 環

筑波大学大学院(博)心理学研究科 青山 征彦

Feasibility of a new type of individual test of intelligence

Hiroyuki Kaiho, Tamaki Hattori and Masahiko Aoyama(Institute of Psychology, University of Tsukuba, Tsukuba 305, Japan)

Feasibility of a new type of individual test of intelligence was explored from three points of view, that is, new concept of intelligence, recent development of test theory and computer technology. A proposed type of individual test of intelligence is expressed as the following three key words; inclusiveness, tailored and multi-media.

Key words: individual test of intelligence, test theory, multi-media.

はじめに

1994年に発行されたSternberg, R.J. (1994)編集の知能ハンドブックで紹介されている主な「個別式」知能検査は、わずかに次の6種類、その他として名前だけが挙げられているものでも13種類しかない。

- 1905年 Binetらによる「ビネー」系
- 1949年 Wechslerによる「WISC」系
- 1972年 McCarthyによる子どもの認知能力検査
- 1977年 WoodcockとJohnsonによる認知能力検査
- 1983年 Kaufmanによる「K-ABC」
- 1993年 BayleyによるBayley尺度-II

これに比して、「集団式」知能検査と銘打った検査は、数えるのが難しいほど、その種類が多い。「開発経費の回収」「診断より選抜目的の使用」「検査実施上のより厳格な条件統制」「多種類・多項目検査

による信頼性の向上」「集団式を個別に実施しても問題がない」といったことなどが「個別式」知能検査の開発意欲を鈍らせてきたように思われる。しかし、こうした不利な条件があるにもかかわらず、本稿で、あえて、「個別式」検査の開発可能性を考えてみようとするのは、2つの理由からである。

1つは、検査者と被検査者とは対面して行うことで、被検査者の多様な反応—身体的運動や発声などを採取できること、したがって、集団式ではとらえることのできない行動を指標とした知能要素を取り出すことができる。

2つは、パーソナル・コンピュータは、まさにその名前のとおり、人とコンピュータとの個人的な交流を前提にした機械である。マルチメディア化技術による表示の質の向上と、マルチモーダル入力技術による反応採取の多彩さは、個別式検査の新たな可能性を示唆していると思われるからである。しかも、テスト理論も、こうしたコンピュータ技術の革新と普及とを前提にして、新たな展開をみせている。そうした点も含めて、個別式検査の開発可能性を探っ

注1：本稿の執筆分担は、はじめにと1は海保、2は服部、3は青山である。全体の企画と調整は海保が行った。

てみる。

なお、ここで想定する被検査者としては、年齢段階では幼児から中学生まで、および、幼児から成人におけるある種の知的障害者である。

1. 新しい個別式知能検査の基本概念

ビネーが知能検査を開発してからほぼ1世紀がたつ。その間の知能検査の研究の流れは、次の3つに整理することができる。

(1) 新しい型の検査の開発

個別式知能検査と集団式知能検査の開発。選抜目的の検査と診断目的の検査の開発。

(2) 知能に関するモデルの提唱

もっぱら因子分析を駆使した知能の構造モデル。知的活動の過程や方略を想定した認知論的なモデル。知能概念を拡張した多面的知能モデル。

(3) 知能検査の実用的／社会的な問題

検査による選抜にまつわる諸問題—公平性、妥当性などをめぐっての議論。

知能検査のこうした研究の動向を踏まえて、ここで、提案する個別式知能検査の基本コンセプトは、次の3点である。

1-1) 知能を包括的に

知能に関する因子分析的な研究によって知能の構造モデルは精緻化され、認知論的アプローチによって、知的活動のプロセス(関与する過程や方略)が明らかにされてきた。そして、両者の成果を組み合わせた知能モデルが、たとえば、Caroll, J.B. (1976)によって提案されてきた。次に述べるように、多彩な目的に応ずることのできる検査を想定するなら、知能モデルはできるだけ包括的なほうがよい。そこで、ここでは、Sternberg, R. (1985)の三頭理論(要素、経験、文脈)を検査開発のマクロモデルとして採用し、その枠のもとで、「構造モデルと認知モデルとを組み合わせた実行モデル」(「検査開発を導くに足る具体性のあるモデル」の意)を構築してみる。

1-2) 知的能力の評価の目的に応じて

知的能力を評価する実際的な目的には(研究目的は除く)、大きく次の2つがある。

- ・選抜／序列づけ目的：子どもの能力が所定の教育機関や組織での教育や訓練に耐えられるかを知りたいとき、あるいは、多数の希望者の中から選抜しなければならないときに使う。
- ・診断／予測目的：子どもの知的機能がどれくらいか、あるいは、子どものどこにどんな適不適、

得意不得意があるかを診断するときに使う。

1つの検査で、2つの目的のいずれもが果たせる方式の検査は寡聞にして知らない。しかし、コンピュータの呈示技術の進歩は、1つの検査で、目的に応じて、検査問題の柔軟な組合せ選択ができる「バッテリー方式の知能検査」を可能にする。

1-3) 検査自体が知的能力の向上になり、かつ、楽しくできる

心理検査も医学的検査などと同じく、診断・予測や選抜という目的のために行われる手段的なものに過ぎない。したがって、検査を受けることで知的機能が增強したり回復したりといったようなことは期待されない。しかし、子供を被検査者とした検査なら、検査すること自体にも知的機能の開発効果があったほうがよいことは言うまでもない。そして、できれば、コンピュータをゲーム機械としてもっぱら使ってきた世代であるので、楽しさを感じさせるような仕掛けを組み込んだものであれば言うことなしである。いわば、「ゲーム感覚で取り組める知能検査」である。

2. 新しいテスト理論を援用した個別式知能検査の開発可能性

項目反応理論は個人の特性値と各項目に対する反応との関係を確認率関数(これを項目特性関数と呼ぶ)を用いて説明し、個人の特性値を項目反応に基づいて推定する(Birnbaum, 1968; Lord & Novick, 1968; Weiss, 1983; Hulin, Drasgow & Parsons, 1983; Hambleton & Swaminathan, 1985; 芝, 1991; 池田, 1994)。ここが、根本的に古典的テスト理論と異なるところであり、この結果、古典的テスト理論では不可能であった新しいテスト方式が可能となった。ここでは、項目反応理論に準拠した個別式知能検査の方式およびその利点を考えてみる。

2-1) 受検項目が異なってもよい

力量検査の場合、古典テスト理論に準拠したテストは全被検査者に同一の問題項目を課す必要がある。そのため、特性値の低い者は過度に難しい項目を受検しなければならず、心理的な負担が大きい。逆に、特性値の高い者には容易に正答が得られる易しい問題が提示され、余計な時間を取る。しかし、項目反応理論は、受検項目が異なっても特性値を共通の尺度上で推定することができるので、個人の特性値水準に応じた困難度の項目を提示すればよい。そのため、被検査者の心理的な負担を軽減すること

ができる。

例えば、未提示項目の中で各個人に対して情報量が最大となる項目を提示する(例えば、柴山ら, 1987, 服部, 1989, 1990)。これは特性値の推定精度を最大にする(標準誤差を最小にする)ことが期待される項目を提示する方式である。ロジスティック・モデル(Birnbaum, 1968; Lord & Novick, 1968)は正答期待確率が最も0.5に近い項目、多値的採点モデル(Samejima, 1969; Masters, 1982; Wright & Masters, 1982; Masters & Wright, 1984; Muraki, 1992)は項目内の困難度あるいはステップ母数の値が特性値の近い範囲に散らばっている項目が選択される。

これらの方法は少数項目で特性値の推定精度を最大にするという観点からは望ましい選択基準であるが、正答期待確率が0.5となる項目は被検査者にとっては心理的な負担が大きいという実験報告もある(服部, 1990; 藤森, 1995)。さらに、別の観点からも項目選択基準を検討する必要がある。

2-2) 個人ごとに一定の測定精度を維持できる

項目反応理論は被検査者ごとに測定精度を定義するので、ある一定の精度を満たしたところでテストを終了することができる。しかし、一定数の基準を満たすまでに相当数の項目を必要とする被検査者が出現することがある。また、ごく少数の項目で推定精度の基準を満たした場合に被検査者が検査方式に疑念をいだくとも考えられる(Hambleton, Zall & Pieters, 1990)。やはり、最大提示項目数と最小提示項目数を設定する必要がある。

2-3) 規準値推定が可能である

事前に規準値を設定しておき、被検査者の特性値がその水準よりも高い水準にあるのかどうかを統計的観点から判断することができる。例えば、1項目の反応を得るごとに特性値の95%あるいは90%信頼区間を計算し、規準値がその範囲外になったところでテストを終了する。

2-4) 力量検査とスピード検査を混在させることができる

従来の個別式知能検査は力量検査が普通である。一方、集団式知能検査はスピード検査方式が普通である。もともと項目反応モデルは力量検査を念頭に置いて提案されているが、多値的採点モデルをスピード検査に適用することができる。例えば、スピード検査の得点を順序データへ変換し、これを項目得点とみなして特性値を推定する。したがって、力量検査とスピード検査を混在させて同一特性を測定す

ることができる。

2-5) 項目の削除と追加が容易である

古典的テスト理論に準拠して新しい項目を追加したり代替検査を作成する場合、改めて標準化の手続きを経る必要がある。しかし、項目反応理論に基づいて1つの項目プールを作成しておけば、大がかりな手続きを取らずに新しい項目を追加できる。また、何らかの理由で使用できない項目が出現しても、少ないコストで項目プールを改良していくことができる。例えば、既存の項目と新項目とを同一標本に実施して、新項目の母数推定値を既存の項目尺度へ等化する。

3. マルチメディア型の個別式知能検査の開発可能性

検査理論が進展を見せている一方で、近年のコンピュータの発達にも著しいものがある。その結果、デジタル技術を中心にしたメディアの統合、いわゆるマルチメディア化が、すでにパーソナル・コンピュータのレベルでも実現しつつある。こうした技術革新が、コンピュータが処理できる情報の質を飛躍的に拡大していることは、言うまでもない。このように、多様な情報を一元的に管理、計測、処理、および呈示できる装置としてコンピュータを考えると、同様に多様な情報が用いられる知能検査をコンピュータで支援するという利用法が考えられよう。また、コンピュータを利用することで、従来用いることのできなかった情報を処理することが可能になれば、従来の知能検査ではできなかったような検査が実現できるだろう。そこで本節では、こうした新しい技術をもとにした知能検査の開発可能性について考えてみたい。

とはいえ、新しい技術を導入しさえすれば従来より良い検査ができる、というわけではない。新しい技術によって従来の検査より優れたものを作れるのでなければ、技術を導入する意味がない。また、新しい技術の導入によって従来の検査とのつながりが失われるならば、利用者はそれまでの蓄積が生かされないことになり、導入に負担がかかることが予想される。それゆえ、新しい検査が十分に有用なものであるためには、次の2つの条件が充足されることが必要と考えられる。

(1) 継承性の条件

従来の知能検査のなるべく多くの検査形態を継承した上で、それらの検査をより容易に、あるいはより適切に実施できるようにすること。

(2) 発展性の条件

マルチメディア機器を用いることで、新たに可能になるような検査形態を実現すること。

3-1) 知能検査のインタフェース

実際にマルチメディア型の知能検査を作成する上では、適切なインタフェースの構築が必要となることが予想される。そこで、新しい知能検査について考えるにあたって、従来の知能検査が、どのようにして検査を構築してきたかを、インタフェースの観点から考え直してみることにしたい。

従来の個別式知能検査では、検査の全ての場面において、検査者が大きな役割を担ってきた。その役割を大別すると、「材料の呈示を管理する」「それに対する回答を得点化する」「得られた結果を解釈する」の3つが指摘できる。なかでも、幼児や知的障害を持つ被検査者を対象にする場合の検査者の介入は、検査を支える上で非常に重要である。その意味において、従来の検査者は「人的なインタフェース」として検査装置の一端を担ってきたのであり、こうした検査者の介入こそが、後述するような多様なインタフェース設計を可能にしてきたと言えよう。以下では、こうした従来の検査者の役割に対する認識をふまえた上で、検査場面の中心をなす呈示と回答という2つの場面について、インタフェースの観点から検討することにした。

・呈示

従来の個別式知能検査が呈示に用いてきたインタフェースは、多岐にわたる。ビネー系やWISC系では、発問や文章の朗読、数列の読み上げなどの音声による呈示法に加えて、絵や文字の書かれたカードを用いた呈示法が用いられているし、K-ABCでは、検査者が手の動作を呈示する、という実演による呈示法も採用されている。

呈示におけるマルチメディア型の検査の特長として挙げられるのが、従来よりも多様な呈示方法が可能になるということである。マルチメディア型とすることで、音声や画像のデータを利用できるため、これらの呈示方法のほとんどを継承できるばかりでなく、動画像のような従来は用いられなかった刺激も呈示できるようになることが期待される。

また、呈示の質を統制しやすくなる点も、特長として挙げられる。例えば、従来の音声呈示では、検査者によるばらつきが生じざるを得なかったが、コンピュータを利用することで均一性を保証できる。視覚的な刺激の呈示についても、絵カードなどの方法に比べ、呈示時間などの点でより厳密な制御が可能になる。

・回答

従来の個別式知能検査は、回答時においても多様なインタフェースを用いてきた。例えば、ビネー系やWISC系では、発話、指さし、筆記、およびブロックや紙片、絵カードなどの配列操作といった方法が採られてきた。さらに、K-ABCでは、手の動きを模倣したり、検査者の指示通りの行動をしてみせるような、動作による回答法も導入している。

こうした多様性を支えてきたのは、検査者という人的なインタフェースである。したがって、マルチメディア型の知能検査の開発可能性を考えるにあたって、こうした検査者の役割がコンピュータでどこまで代行可能かが問題になる。そこで以下では、従来の様々なインタフェースについて、この点を検討することにする。

まず、発話や数唱のような言語的な回答法は、検査者の介入なしには導入できない。コンピュータが検査者を代行することになれば、被検査者の音声入力をコンピュータ側で解釈するか、キーボードやタブレットなどによる文字入力を被検査者に行わせるか、どちらかの方法しかない。しかし、現時点では、前者は実用にはほど遠い。後者も、入力装置に慣れていないとインタフェースの透明性が確保できず、回答に支障が生じよう。知的障害を持たない成人でも、こうした装置を使いこなせない場合が多いことを考えると、幼児や子ども、さらには知的障害を持つ被検査者に自ら文字を入力させるのは、負担が重すぎる。したがって、言語的な回答を指標に用いるならば、人的なインタフェースの介入が必要と考えられる。

同様に、検査者の役割を代行しきれないのが、動作的な回答である。動作は、直接にはコンピュータに入力しにくい。また、動作の適切さを評価して得点化するという手法も、検査者の介入によって達成されてきた側面が強い。

一方、指さしは、画面に直接触れて反応を返すタッチパネル方式を用いれば、そのまま代行できよう。使用方法もわかりやすく、様々な被検査者に適用できると考えられる。また、筆記や配列操作も、インタフェースの工夫次第では可能である。例えば、迷路などの筆記課題には、画面が表示されるタイプのタブレットが適している。各種の配列課題も、タッチパネルで対応できると考えられる。

このように検討していくと、従来の検査者が回答時に果たしてきた役割には、コンピュータで代行可能なものと、代行不可能なものが混在していることがわかる。ただし、後者には、従来の知能検査の中心をなしてきた発話による回答方法が含まれるこ

と、新しい知能検査は従来の検査に対して継承性を持つ必要があることを考えると、マルチメディア型の知能検査においても、検査者を置くべきであろう。

ここまでは、継承性の条件についてのみ述べてきたが、発展性の条件について、マルチメディア型の知能検査で初めて実現される検査方法も考えられる。その1つとして、従来は実現できなかった反応時間の測定や、課題に対する習熟過程の分析が可能になることが挙げられる。こうしたミクロなデータは、従来の検査では得られなかったものである。

3-2) マルチメディア援用型検査の利点

このように、従来の知能検査に対する継承性という観点から考えると、コンピュータだけでも呈示は代行できるが、回答に関しては検査者の介入が必要になる。それゆえ、マルチメディア型の知能検査の構築にあたっては、検査者を置いた上でコンピュータを支援的に用いるのが、最も実現可能な方法であろう。それによって、コンピュータを導入することの利点を確保できる一方で、上述した問題点の多くが回避可能になる。こうした新しい検査形態は、「マルチメディア型」から一歩進めて「マルチメディア援用型」と言うことができよう。以下では、このマルチメディア援用型の知能検査について、上述した以外の利点を指摘したい。

まず第1に挙げられるのが、材料の呈示の管理と、回答結果の解釈において、検査者の負担を大きく軽減できる点である。これによって、それほど訓練を積まない検査者でも十分な検査が行えることになる。こうした特長は、実際に検査を必要とする現場で、大きな力となる。

第2に、多様な呈示法を用いることで、検査のゲーム性が増すことが挙げられる。知能検査は、長時間にわたることが多い。幼児など、時間の負荷が負担になる被検査者にとっては、ゲーム的な感覚が持てる検査形態を採用することで、負担感を軽減できる。

第3に、コンピュータを利用することにより、データを保存・活用しやすいことが挙げられる。検査結果を前回の結果と比較することや、被検査者集団全体を分析することが容易になる。

第4に、検査自体をソフトウェアにすることで物理的に小型化できるため、複数の検査をまとめてパッケージ化しやすくなることが挙げられる。様々な検査形態をまとめてパッケージ化し、目的に応じて、また、対象や年齢にあわせて必要な検査を選んで使用するようにすれば、同一環境で様々な検査形態が実現できる。

以上のような利点に対して、一方で、様々な課題

も残されている。例えば、コンピュータを導入すること自体が検査結果に影響を与えてしまう可能性が懸念される。教室場面でのコンピュータの導入が、教師-生徒関係を変質させることが指摘されている(例えば、佐伯ら, 1993)が、これと同様のことが知能検査場面でも起きるかもしれない。従来の知能検査場面に潜んでいた検査者-被検査者関係とはどのようなものだったのか、新しい検査形態はその関係性をどう変えるのか。開発と実際の運用にあたっては、こうした点からの検討も必要になろう。

さらに、本稿では触れなかったが、ハードウェアの構成や、ソフトウェアの開発^(注2)など、実際にシステムを構築する際に考えなければならないことは、他にも多くある。実現に向けては、さらに詳細な検討が必要であろう。

4. おわりに

コンピュータは出現は、社会の至るところの仕事を変容させている。本稿では、最新のコンピュータ技術を踏まえて、新たな個別式知能検査の開発可能性を探ってみた。そして、従来型の検査では実現されていない(できなかった)いくつかの新しい仕掛けを組み込んだ個別式検査の開発可能性を提案することができた。しかし、逆に、コンピュータ援用による新たな制約もある。そうしたものに配慮しながら、今後、本稿で提案した構想を実際の検査開発につながる努力をしてゆくことになる。

引用文献

- Caroll, J.B. 1976 Psychometric tests as cognitive tasks: A new "structure of intellect" In L. Resnick (ed) The Nature of Intelligence. Hillsdale, NJ: Erlbaum.
- Birnbaum, A. 1968 Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M., & Novick, M.R. 1968 Statistical Theories of Mental Test Scores. Reading, Mass.:

注2：ソフトウェアの媒体について

近年の記憶媒体の発達にはめざましいものがあるが、なかでもCD-ROMは、ドライブ、媒体ともに安価であり、大容量をコンパクトに扱えることから、知能検査の構築に適していると考えられる。ただし、検査結果の保存にあたっては、通常のドライブではCD-ROMに追記できないので、追記型ドライブを使うか、データをフロッピーディスクなど他の記憶媒体に書き出すようにする必要がある。後者は実現の容易さと、ファイル保護をしやすい点が評価できる。

- Addison-Wesley.
- 藤森 進 1995 テスト項目の心理的に最適な困難度水準の研究. 心理学研究, **65**, 446-453.
- Hambleton, R.K., & Swaminathan, H. 1985 Item Response Theory. Boston: Kluwer・Nijhoff.
- Hambleton, R.K., Zaal, J.N., & Pieters, J.P.M. 1990 Computerized adaptive testing: Theory, applications, and standards. In Hambleton, R.K., & Zaal, J.N. (eds.) 1990 Advances in Educational and Psychological Testing. Norwell, Massachusetts: Kluwer Academic Publishers.
- 服部 環 1989 調整テスト方式により中学生の語彙理解力を測定する試み. 日本教育工学雑誌, **13**, 129-137.
- 服部 環 1990 個人差に応じたテスト方式による語彙理解力の測定. 教育心理学研究, **38**, 445-454.
- Hulin, C.L., Drasgow, F., & Parsons, C.K. 1983 Item Response Theory: Application to Psychological Measurement. Homewood, IL: Dow Jones-Irwin.
- 池田 央 1994 現代テスト理論 朝倉書店.
- Lord, F.M., & Novick, M.R. 1968 Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley.
- Masters, G.N. 1982 A Rasch model for partial credit scoring. Psychometrika, **47**, 149-174.
- Masters, G.N., & Wright, B.D. 1984 The essential process in a family of measurement models. Psychometrika, **49**, 529-544.
- Muraki, E. 1992 A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, **16**, 159-176.
- 佐伯 胖, 佐藤 学, 荻宿俊文, NHK取材班 1993 教室にやってきた未来 NHK出版
- Samejima, F. 1969 Estimation of latent trait ability using a response pattern of graded scores. Psychometric Monograph, No.18.
- 芝 祐順(編) 1991 項目反応理論—基礎と応用—東京大学出版会.
- 柴山 直・野口裕之・芝 祐順・鎌原雅彦 1987 最適化テスト方式による語彙理解力測定の測定. 教育心理学研究, **35**, 363-367.
- Sternberg, R.J. (Ed.) 1994 Encyclopedia of Human Intelligence. MacMillan Publishing Com.
- Weiss, D.J. (eds.) 1983 New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing. New York: Academic Press.
- Wright, B.D., & Masters, G.N. 1982 Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press.

—1995.9.30受稿—