

大規模入学者選抜における IRT-CBT への懸念

林 篤裕 (名古屋工業大学 工学研究科)

1 はじめに

情報化社会と言われるようになってもう何年経つのであろうか。ICT 機器やインターネットが生活の一部としてあらゆる世代に広く普及し、全ての機器がネットワークで連携する IoT の導入も進んでいる。また、このような環境を活用して種々の行政手続きを簡便に実現するべく、政府でもデジタル庁の創設に向けての準備が進んでいると聞く。そうした中で、一般的な試験だけでなく大学入試の実施にも ICT を活用した試験環境の導入が検討される段階になってきた。そこで、大学入試の中でも特に共通テストのような大規模試験における ICT の利活用について検討することになり、(独)大学入試センターの大学入学共通テスト企画委員会の下に「CBT 活用検討部会」が設置され、その準備会合であった有識者会議を含めると約 2 年間に亘っていろいろな視点から検討された。

この検討部会において、筆者は大規模試験と CBT の親和性や、試験のスコアを潜在変量で表現する手法を利用することの懸念について提起した。本稿はこれらの事項について再録するものである。

2. 大学入試と CBT

コンピュータが比較的容易に使えるようになった時期から、その応用分野として教育への利用も提案されてきた。学習の支援に続いて学習成果を確認する試験に対してもコンピュータを用いた評価方法が検討され、従来からの紙と鉛筆による試験(Paper Based Testing, PBT)からコンピュータを用いた試験(Computer Based Testing, CBT)への移行が徐々に進みつつあり、欧米の入学試験や日本における資格試験の一部では導入事例も見られるようになってきた。

PBT および CBT のメリット・デメリットについては上記検討部会が本年 3 月に公表した報告書に詳しくまとめられているので詳細はそちらに譲るが、PBT では試験問題の提示に静的なものしか出題できないのに対して、CBT では動的なマルチメディアも利用でき、個々人の解答に応じて出題内容を臨機応変に切り替えられることが最大の利点であり、特に大規模試験の場合には、試験問題の提供も従来より安価でかつ迅速に行えるというメリットも有る。

しかし一方で、2006 年から大学入試センター試験に導入された英語リスニングテストの例をみても判るように、既に導入から 15 年ほどが経過しているにも関わらず未だにリスニング機器トラブルによる再開テストは根絶できていない。工学的見地から類推すると、リスニング機器より部品点数が格段に多くなって複雑化しているパソコンでは、そのトラブル発生率がリスニング機器のそれよりも遥かに高くなることが予想され、公平な環境を求める入試にパソコン類を利用するには慎重な検討を要すると思われる。

また、欧米では入試に CBT が導入されている事例があるにも関わらず、日本ではなかなか導入に踏み切れていない理由は幾つかあるが、例えば入試問題が問おうとしている内容・質に対する要求度や、これまで長年に亘って改良が続けられてきた多肢選択式設問の作題技術等といった試験文化の違いが挙げられるであろう。他にも競争性の違いも関係していると思われ、欧米では大学終了段階(出口)の学修度合いを厳密に評価することによって卒業の可否を認定しており、入り口にあたる大学入試にはあまり厳密さを求めておらずローステークスとなっている。一方、日本の大学入試は言うまでもなくハイステークスであり 1 点(時にはそれ未満)で合否が分離されることが一般的である。このことへの対応は先の高大接続答申(2014)においても、1 点の取り扱いの厳密さを緩和する方向性を提言してはいるものの、その具体的な方策に名案があるとは言い難い。

3. IRT を日本の大学入試に用いる際の注意点

CBT との親和性の高い試験技術に IRT (Item Response Theory)がある。プレテストを事前に行い統計量を把握した多数の Item(設問)を準備することに依って、受験者の能力を少数設問で正確に推定することができ、もしくは、複数回の試験のスコアを比較することも可能になる。しかし、これを実現するには幾つかの守るべき仮定が設けられており、また、モデルのパラメータの推定にも懸念点がある。

因子分析(Factor Analysis, FA)は顕在変量として測定されたデータを、モデルに基づく因子負荷量と共通因子の積に分解することによってデータ構造を明らかにしようとする統計手法である。この手法に

において、因子負荷量と共通因子の積の組み合わせは数学的に無限に存在する(回転の不定性と呼ばれる)が、これらは何れも潜在変量と呼ばれ実際に測定することができない。顕在変量は文字通り測定値として確かめることができるが、潜在変量はそれができず、加えて、無数に存在するとの組み合わせを採用するかは分析者の判断が関与することになり、このことをどのように捉えるかが一部研究者間で議論になることがある。

同様に、IRT も潜在変量を扱う統計モデルと言える。IRT の場合はその解法において尤度関数の最大化を行うパラメータを推定しているが、この作業は多変数多峰性関数のグローバルマックスを求めることを意味する。これは最適化の問題に帰着するが、探索の初期値とメッシュの粒度に依存した解が求まることになり、たとえ高速なコンピュータを用いたとしてもローカルマックスは求まるとしても、グローバルマックスが求まる保証はない。つまり、求まったと思われる解は多数ある解の一例にしか過ぎず、それより尤度の大きい解が存在することが否定できず、しかし、その存在を確かめることも不可能である。加えて IRT にはモデルとして 1 パラメータから 3 パラメータまでが存在し、それぞれで解も変わってくるが、そのどれを採用するかも FA 同様、分析者に任されている。

このように IRT で求めた試験のスコアには分析者が関与するブレが関係すると言えるが、前節で述べたように日本の大学入試においては 1 点の取り扱いが極めて重要になるものの、全ての受験者に有利に働く解は存在しないため、解に不確実性が残る手法を大学入試における個人の能力の推定に用いることは現状では理解が得られ難く馴染まないと考える。

4. IRT の活用場面

前節で IRT の利用についての懸念を述べたが、IRT 自身は試験のスコアを数理的に表現するモデルとして理解出来るため、IRT は大規模入学者選抜より適した活用場面があると感じている。IRT の利用には Item Pool が必要で、Item 数も相当に多いことが期待される。日本は初等中等教育において学習指導要領という教育スタンダードが存在し、全国の各学校でそれに則った教育が行われ、その教育成果を評価する試験が実施されている。ほぼ 10 年ごとに改訂が行われるとは言え、学習指導要領を固定すると、これに準じた Item が多数存在し、それが毎年全国で利用されている状況を考えれば、各学年の学習達成度を評価する方策として IRT を用いることは有効に機能すると考えられ、例えば数年前から実施されている「高校生のための学びの基礎診断」等で応用できると思われる。また、初等中等教育の学校には児童・生徒の横に教員が寄り添っているの

で、試験スコアの多少のブレは教員の指導に依って補正でき、教育に影響を及ぼさないと考えられる。日本においては教育評価の振り返り場面に IRT 技術を利用することは十分利用価値があると思われる。

5. まとめに代えて

一口に大学入試と言っても、個々の大学が実施する個別入試もあれば、全国規模で実施される共通テストもある。その違いの一番は受験者数であるが、トラブルが起こった際の影響が全国に波及してしまうという点で、大規模入学者選抜の実施環境は頑健に構築する必要がある。そう考えるとこれには長年の蓄積が有る PBT の方に一日の長があり、CBT や IRT を大規模試験に適用するにはまだ解決すべき課題が残されていると考えられ、ひとまずは選抜への貢献は置いておいて、教育評価の場面を想定しての研究を進めて行ってはどうであろうか。もしくは、日本も出口管理の国に制度を移行することにより、入り口段階の厳密さが軽減されれば、これら技術の導入へのハードルが下がるのではないかと予想される。

6. 参考文献

- 1) 丘本 正(1986)、「因子分析の基礎」、日科技連出版社。
- 2) 芝 祐順編(1991)、「項目反応理論—基礎と応用」、東京大学出版会。
- 3) 浅野 長一郎 & 江島 伸興(1993)、「潜在構造分析論の現状」、日本統計学会誌 第 22 巻 第 3 号、PP357-373。
- 4) 池田 央(1994)、「現代テスト理論」、朝倉書店。
- 5) 中央教育審議会 (2014)、「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について～すべての若者が夢や目標を芽吹かせ、未来に花開かせるために～」、中央教育審議会答申(中教審第 177 号)。
https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1354191.htm (2021 年 7 月 31 日現在)。
- 6) 宮澤 芳光(2019)、「Computer Based Testing (CBT)を用いたテストの出題」、大学入試センター・アドミッションリーダー研修「入試問題の作成・分析と CBT 入門」配布資料、2019 年 7 月 20 日。
- 7) 寺尾 尚大(2019)、「CBT・CAT とは何か」、日本テスト学会第 17 回大会、公開シンポジウム 2 「多面的総合的評価・CBT・アクティブ・ラーニング」配布資料、2019 年 8 月 29 日。
- 8) 独立行政法人 大学入試センター(2021)、「大規模入学者選抜における CBT 活用の可能性について(報告)」、https://www.dnc.ac.jp/research/cbt/cbt_houkoku.html (2021 年 7 月 31 日現在)。