

<書評>

別府正彦 2015. 『IRT(項目反応理論)入門』
東京：河合出版 180頁

井澤 廣行*

Hiroyuki Izawa

本書の監修者として野口裕之・柴山直・熊谷龍一が挙げられており(p. 180)、「監修にあたって」(p. 1)の中で、日本でのIRT研究・適用における先導者の一人である野口裕之が次の言葉を記している。

『[別府正彦氏は、]2006年から実施されている「受験学力測定テスト」(学校法人河合塾)の開発に際しては中核メンバーの一人として実現に貢献されました。テスト受験者の高等学校3年間にわたる学力の実質的変化を把握するために、「偏差値」ではなく「IRTベースの共通尺度」を用いて表示するテストはわが国の受験界では嚆矢でした。』

又、「著者紹介」(p. 180)において、「2009年より、テストの開発・分析を主な事業とする株式会社ハピラル・テストソリューションズ代表取締役社長」ともあり、IRTの研究とその理論の現実適用への別府正彦による造詣の深さが窺われる。なお、IRTはItem Response Theoryの略語である。

日本人教員で、項目反応理論ないしは項目応答理論という言葉を目にしたことがない者は多くはない筈であるが、それが本質的に何であり、何を可能とするのかを理解している者はまだ少ないと思われる。著者は、『今後、国内でもIRTが盛んに使われるようになると思います。[中略]ですので、専門的なことまではわからないにしても、基本的なことは理解しておきたいと考えます』(p. 63)と述べており、それが本書の執筆出版目的である。評者において、Rasch項目分析モデルへの言及が僅か9行(pp. 117-118)でありそのモデル特異性への記述欠如が意識されるが、2パラメタ・ロジスティック・モデル(以降、2PLM)に関する平易で丁寧な説明により上記の目的は8割程度に果たされている。

第3章「従来のテスト得点」(pp. 38-63)において分かり易い例を挙げて問題点を記述しているが、この内容は、何らかの形で試験作成・得点表示に従事するすべての教員に容易に理解できるものである。特に、「偏差値」のあいまいさに関する記述(pp. 54-58)が秀逸である。3章のまとめとして記載されている5つの指摘は簡明直截であり、著者がその記述をIRT説明への導入としているので、以下にそれを引用する。

- ① 『出題された問題が異なるテストの場合、正答数得点や重みつき正答数得点を比較することは極めて難しい』(p. 59)。
- ② 『正答数得点では、出題された問題の難度や重要度が得点に反映されにくい』(p. 59)。
- ③ 『重みつき正答数得点の場合、合理的で客観的な配点の決め方が不明確である』(p. 59)。
- ④ 『異なる集団で求めた偏差値は比較することはできない』(p. 60)。
- ⑤ 『偏差値が上がったり下がったりする理由がわかりにくい』(p. 60)。

「従来のテスト得点の問題点」としての上記5つの指摘との関連で、著者は、①～④についてはIRTの有用性、⑤については本書執筆出版の目的として、以下の様に述べている。

- 1) 『出題された問題が異なるテストであっても[IRT適用等化(共通尺度化)により]テストの得点を相互に比較できる。[中略]ただし、この特長を活かすには、局所独立の仮定やテスト1次元性などのIRTを使うための条件を満たしておく必要があります』(pp. 60-61)。
- 2) 『IRTを使うと、問題の「難度」を合理的な方法で分析し数値で表すことができます。[中略、問題の「重要度」については、]IRTにおいても、学力の違いによって正誤にどのくらい差が見られるかを、その問題の特性の一つとして取り上げます』(p. 61)。
- 3) 『IRTを使って得点を求める場合、事前に配点を決める必要はありません。もちろん、テストのあとにも決める必要はありません。[中略]「パラメーター」はテストを実施して、IRTを使って分析することでその値が求まります』(p. 62)。
- 4) 『受験者集団が同じだろうが、違っていようが、そんなことには関係なく、IRTを使って得られたテストの得点は[IRT適用等化により]互いに比較することが可能です』(p. 62)。
- 5) 『そもそも、「IRTではどうやって得点を出すのか」が実はとてもわかりにくいのです。[中略]でも、IRTの基本的な考え方を、多くの方たちが理解することは土台無理な話なのではないでしょうか。もしそうだとしたら、まったくわからないIRTを使って算出された成績が、個人の人生に大きな影響を及ぼしてしまう。そんな世の中になるのでしょうか。それは絶対に避けるべきだと思います。ですので、専門的なことま

ではわからないにしても、基本的なことは理解しておきたいと考えます』(pp. 62-63)。

上記引用 2) について、Rasch 項目分析モデル(Rasch¹⁾, 1960, pp. 62-75)の特異性ととの区別の上で、注釈が必要である。2)において、「問題の重要度」として著者により解釈されている IRT 特性は、「識別力」ないしは「弁別力」と訳されているパラメタである。2PLM は次式で表わされる(本書、p. 70)。

$$P_j(\theta) = 1 / \{1 + \exp[-Da_j(\theta - b_j)]\} \quad (1)$$

θ : 受験者能力パラメタ値

D : 尺度因子であり、 $D=1.7$ のときに θ 全域にわたり正規累積 2PLM との違いが 0.01 以下になることが知られている(村木²⁾, 2011, p. 45)。

a_j : 項目 j の識別力パラメタ値

b_j : 項目 j の困難度パラメタ値

$P_j(\theta)$: 能力パラメタ値 θ を持つ受験者の項目 j への正答確率

式(1)での要点は $a_j(\theta - b_j)$ であり、項目識別力パラメタは、受験者能力と項目困難度の両パラメタの差に乗算(関連)されて、2PLM の正答確率 $P_j(\theta)$ に関係していることが明白である。この点に関して、3 つ以上の同数順序カテゴリから成る項目群への被験者群応答一次元多値カテゴリデータに対して被験者群位置母数・項目群位置母数・カテゴリ閾値母数の分離推定値を与える多値 Rasch 分析モデルとしての Rating Scale モデル(Andrich³⁾, 1978)を考案した Andrich⁴⁾ (1988) は次の様に述べている。

『2PLM は、受験者能力母数と項目困難度母数の Rasch 項目分析モデル上で明示されている母数推定値「不変性」(Invariance)を不能としている』(p. 67)。

2PLM における受験者能力と項目困難度の上記 Andrich⁴⁾ (1988, p. 67)指摘による母数推定値不変性不能は、受験者能力と項目困難度の母数推定値「分離」不能 (Wright⁵⁾, 1977, p. 220) と同義である。

本書著者は Rasch 項目分析モデルと 1 パラメタ・ロジスティック・モデルを同一とみなして次式を与えている(p. 118)。要するに、2PLM 式(1)における項目識別力パラメタを一定にするものである。

$$P_j(\theta) = 1 / \{1 + \exp[-Da(\theta - b_j)]\} \quad (2)$$

式(2)は、野口・大隅⁶⁾ (2014)によっても与えられており(p. 91)、本書及び野口・大隅⁶⁾ (2014)において言及が欠けているのは、Rasch 項目分析モデルの特異性としての受験者能力パラメタと項目困難度パラメタの「母数分離性」(Rasch¹⁾, 1960, pp. 171-178; Wright and Linacre⁷⁾, 1987, pp. 5-6; 静⁸⁾, 2007, pp. 197-206)である。Rasch (1960¹⁾, pp. 174-178)の証明により、テストに

含まれた受験者群と項目群のそれぞれについて、各正答数得点一群が各正答数得点一群の母数推定値算出への十分統計量であり、十分統計量としてのそれぞれの正答数得点一群の存在が受験者群と項目群の母数分離を保証するものとなる。Fisher Jr.⁹⁾ (1997)により、受験者群と項目群の母数分離のためには項目群識別力母数と受験者群識別力母数のそれぞれを同値とするのが必然と説明されている(p. 566)。Linacre¹⁰⁾ (1999)によっても、Rasch モデルのみが母数分離を可能として、構成された推定値尺度の安定性・再現性のためには項目群識別力を一定にする必要があると指摘されている(p. 692)。構成された推定値尺度の安定性・再現性とはすなわち受験者群と項目群についての母数推定値不変性である。従って、『ラッシュモデルに合っている場合のみ、客観的な測定が可能である』(静⁸⁾、2007, p. 217)と理解される。本書における Rasch 項目分析モデルについての説明箇所(pp. 117-118)に 2PLM と区別される基本必須事項として、上述の母数分離性に関する記述を付加する必要がある。その上で、2PLM の適用に精通・熟達している著者により、数千・数万の受験者群データへの Rasch 項目分析モデル適用に勝る 2PLM 適用上での格段に少ない能力同一推定値群を含む実用的な有用性に関する記述がなされて然るべきである。

第 10 章「良質な問題は国の宝だ」(pp. 170-179)における「良質な項目をみんなで作って、良質な項目をみんなで使う」(pp.173-177)との著者による提案は、その発想に IRT 適用等化による独自性がある。著者によるその「良質項目プール」という理想は、小学 1 年から高校 3 年までの各教科に関する能力測定正当性・一貫性への真摯なこだわりと拝察する。著者による日本での能力測定の正当化と教育の充実・進展へのこだわりは、本書執筆出版目的としての上記引用 5)を含めて本書のいたるところに窺われる。

『テストは、問題を作るよりも、試験を実施するよりも、採点するよりも、分析するよりも、何よりもテスト結果を活用することが最も難しいと思います。テストの結果が出ると、そこに何らかの課題が見出されます。課題が見出されれば、その解決のために何らかの行動を起こさなくてはなりません。何らかの行動を起こして、変化を生じさせなくてはなりません。[中略]少し大げさな言い方ですが、変える勇気をもって、テストの結果を活用する必要があります』(著者による太字強調、p. 160)。

上記引用文においても著者によるそのこだわりが顕示されており、IRT 適用の如何を問わず、テスト実施主体によるテスト結果の活用を促すものである。本書は、教員にとって思考の糧となる書でもある。

【参考文献】

- 1) G. Rasch. 1960. *Probabilistic models for some intelligence and attainment tests*. The Danish Institute for Educational Research. (Reprinted in 1980 by the University of Chicago Press with

a Foreword and Afterword by Wright, B. D.)

- 2) 村木英治 2011. 『項目反応理論』 東京： 朝倉書店
- 3) D. Andrich. 1978. A rating formulation for ordered response categories. *Psychometrika*, 43, 4, pp. 561-573.
- 4) D. Andrich. 1988. *Rasch models for measurement*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-068. Beverly Hills: Sage Publications.
- 5) B. D. Wright. 1977. Misunderstanding the Rasch model. *Journal of Educational Measurement*, 14, 3, pp. 219-225.
- 6) 野口裕之・大隅敦子 2014. 『テストインテグレーションの基礎理論』 東京： 研究社
- 7) B. D. Wright & J. M. Linacre. 1987. Rasch model derived from objectivity. *Rasch Measurement Transactions*, 1, 1 in Linacre, J. M. (Ed.), 1995, *Rasch Measurement Transactions, Part 1* (pp. 5-6). Chicago: MESA Press.
- 8) 静 哲人 2007. 『基礎から深く理解するラッシュモデリング — 項目応答理論とは似て非なる測定のパラダイム』 大阪： 関西大学出版部
- 9) W. P. Fisher Jr. 1997. Blind guides to "measurement." *Rasch Measurement Transactions*, 11, 2, pp. 566-567.
- 10) J. M. Linacre. 1999. Comment on *the New Rules of Measurement* (edited by Embretson, S. E. & Hershberger, S. L. 1999). *Rasch Measurement Transactions*, 13, 2, pp. 692-695.