

# 項目反応理論から見た大学入試センター試験 「国語」「数学」「英語」の特徴

荘島宏二郎, 吉村 幸 (大学入試センター), 中畝菜穂子 (新潟大学),

本研究では, 項目反応理論 (item response theory, IRT; e.g. Lord, 1980; 池田, 1994) を用いて, 大学入試センター試験 (National Center Test for University Admissions, NCT) の「国語 I・国語 II (国 I・II)」「数学 I・数学 A (数 I・A)」「数学 II・数学 B (数 II・B)」「英語」テストを分析した結果を報告した。

## 1. はじめに

IRT は, テストの計量学として, 強力な統計ツールであり, 米国の SAT や TOEFL などでも実用化されている。我が国でも医学部共用試験 (東京医科歯科大学医歯学教育システム研究センター) や日本留学試験 (日本国際教育協会 AIEJ) などの大規模試験において IRT 化が達成されている。また, 我が国の大学入試センター試験は, 石塚 (2002), 林 (2003), 吉村 (2004) が IRT を用いた分析を行っているものの実用化にはいたっていない。

本研究では, IRT から見た「国 I・II」「数 I・A」「数 II・B」および「英語」テストの特徴の経年的な変化を考察することを目的とする。その際, 吉村ら (2003) の作成した NCT のデータベースを利用した。なお, 本研究で報告される項目母数などは, 年度を渡って等化されたものではないことを一言断っておく。

## 2. IRT の分析結果

主に以下の 2 点について報告する。共に IRT において重要なものであり, それらは

- (1) テスト反応関数 (test response function, TRF)
- (2) テスト情報関数 (test information function, TIF)

である。TRF は, 学力 ( $\theta$ ) が与えられているときのテスト得点の期待値であり, (1)式で与えられるものである。すなわち

$$\begin{aligned} TRF(\theta) &= E[T|\theta] \\ &= \sum_{j=1}^n w_j E[X_j|\theta] \\ &= \sum_{j=1}^n w_j P_j(\theta) \end{aligned} \quad (1)$$

である。ここで

$$T = \sum_{j=1}^n w_j X_j \quad (2)$$

であり,  $X_j$  は項目  $j$  の得点である。 $P_j(\theta)$  は能力値  $\theta$  が与えられた元での項目  $j$  に対する正答率であり, 項目反応関数 (item response function, IRF) と呼ばれる。

また, TIF は,  $\theta$  が与えられているときの当該  $\theta$  の推定精度 (標準誤差の 2 乗の逆数) であり, 次式によって与えられる。すなわち

$$\begin{aligned} TIF(\theta) &= \frac{1}{V[T|\theta]} \\ &= \sum_{j=1}^n I_j(\theta) \\ &= \sum_{j=1}^n \frac{\{\partial P_j(\theta)/\partial \theta\}^2}{P_j(\theta)\{1-P_j(\theta)\}} \end{aligned} \quad (3)$$

ここで,  $I_j(\theta)$  は, 項目情報関数 (item information function, IIF) と呼ばれ, 当該

項目の $\theta$ のレベルに対する推定精度を検討することができる。

### 2.1 テスト反応関数

図1-4に「国I・II」, 「数I・A」, 「数II・B」, 「英語」における1997年度から2004年度にわたるTRFを, 図の見易さのために4年度に絞って提示した。なお, 図中, 年度はシャッフルしてある。なお, 「数I・A」および「数II・B」は大問選択パタンの数だけTRFが描けるはずであるが, ここでは, すべての項目を用いてTRFを算出し, 得点が100点になるように調整した。なお, 「国I・II」と「英語」には3パラメタ・ロジスティックモデル(3PL)を, 「数I・A」, 「数II・B」には2パラメタ・ロジスティックモデル(2PL)を用いた。「国I・II」と「英語」は多肢選択式問題だからである。

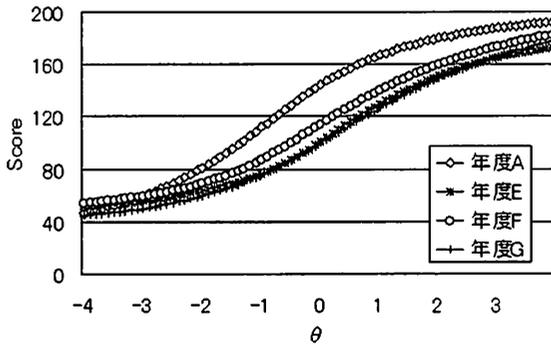


図1 「国I・II」のTRF (3PL)

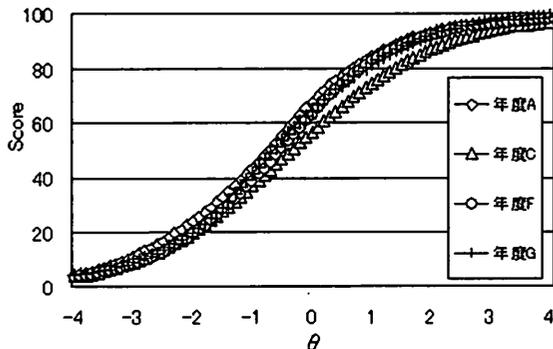


図2 「数I・A」のTRF (2PL)

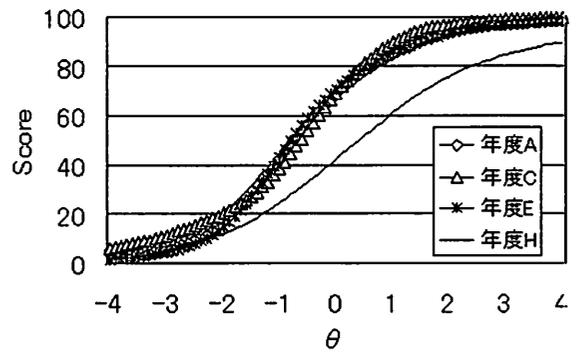


図3 「数II・B」のTRF (2PL)

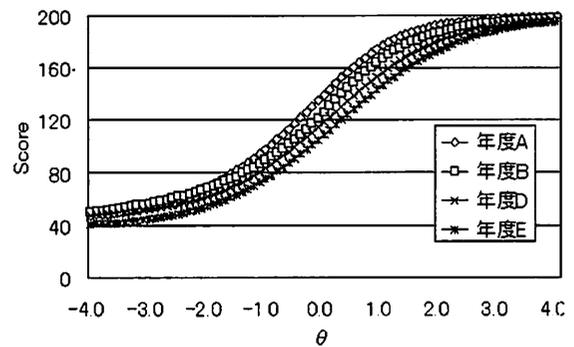


図4 「英語」のTRF (3PL)

どの科目も年度によって比較的難易度が安定しているといえるが, 「国I・II」の年度Aは他の年度に比べて, とても難易度が低い。曲線が全体的に左に位置しているからである。また, 「数II・B」の年度Hは, 他の年度に比べて難易度が高かった。

### 2-2. テスト情報関数

本節では, テスト情報関数の結果を報告する。図5-8に「国I・II」, 「数I・A」, 「数II・B」, 「英語」における1997年度から2004年度にわたるTIFを, 図示の見易さのために4年度に絞って提示した。

情報関数は, 各 $\theta$ のレベルにおける推定の精度をあらわしている量である。全般的に, 各科目, 各年度のテストは,  $\theta$ が0付近をよく推定するテストであるということが分かる。これはすなわち, 平均的な学生をよく識別し

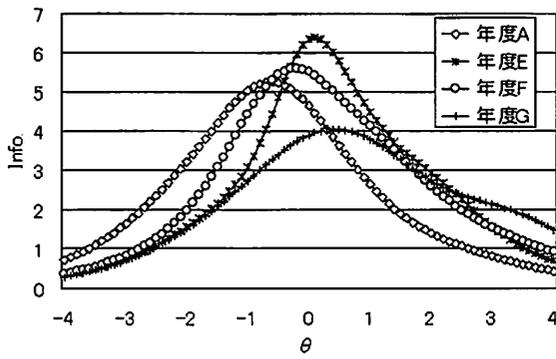


図5 「国 I・II」の TIF (3PL)

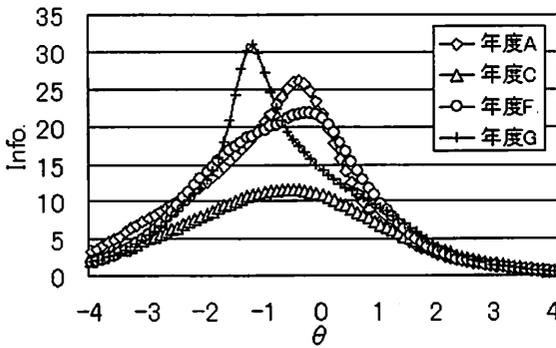


図6 「数 I・A」の TIF (2PL)

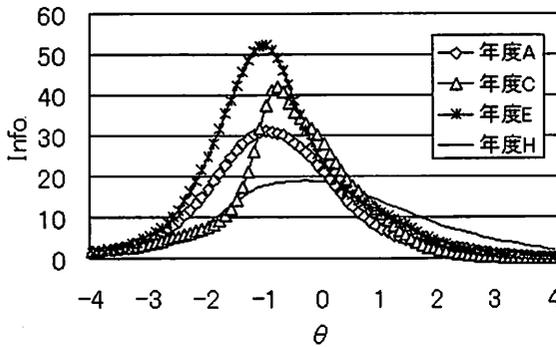


図7 「数 II・B」の TIF (2PL)

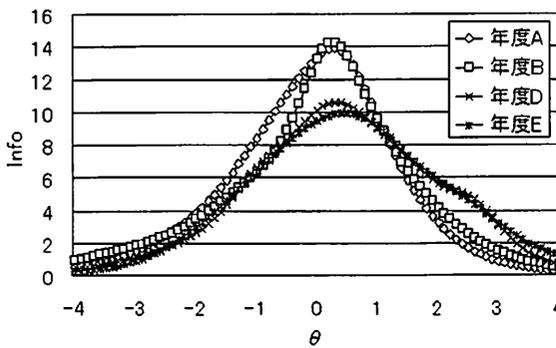


図8 「英語」の TIF (3PL)

ているということである。このあたりの事情は吉村・荘島・鈴木 (2004) に詳しい。

また、「国語」「英語」のような文系科目では、年度によって TIF の形がよくまとまっているといえるが、「数 I・A」および「数 II・B」の理系科目では、TIF が年度によって落ち着きがない。これは、理系科目では、極端に識別力の高い項目が 2, 3 個散見されるためである。識別力母数が高い項目が混入していると、2 乗のオーダーで情報関数が大きくなるために起こっていることである。決して悪いことではないが、項目レベルでの IRT モデルのフィットが悪いという可能性も含めて、解釈には慎重になるべきであろう。

### 3. まとめ

本研究では、IRT の観点から、「国 I・II」「数 I・A」「数 II・B」「英語」の 4 科目について分析した。そこでは、主に TRF と TIF に焦点化した。また、用いた IRT モデルは、2 パラメタ・ロジスティックモデルと 3 パラメタ・ロジスティックモデルであった。

吉村ら (2004) の報告によると、「国 I・II」と「英語」については、局所独立の仮定をバイオレートしているにもかかわらず、3 パラメタ・ロジスティックモデルを用いることについて、一定の知見を得ることができそうである。ただし、2 パラメタ・ロジスティックモデルを用いることには警鐘を鳴らしている。

また、「数 I・A」「数 II・B」については、2 値モデルを用いることがそもそも可能でない場合がある。しかしながら、多値モデルの安易な適用にも注意すべきである。もっとも重要な問題は、局所独立の仮定のバイオレートであろう。いずれにしても実務レベルでの今後の検討が待たれる。

IRT は、主に米国で発展してきたものであり、我が国のテストスタイルの伝統を色濃く反映しているセンター試験に適用するには今なお、検討が必要であろう。

ただし、IRT の豊かな分析力を用いれば、多次元モデル (McDonald, 1997; Reckase, 1997) の適用や局所独立の仮定をリラックスさせたモデル (Ueno, 2002) の適用など、上述の問題をクリアすることが可能であるかもしれない有望なモデル群も存在する。さらに、異なる項目形式ごとに異なる IRT モデルを適用することも可能である (荘島・豊田, 2004) ので今後の課題となるだろう。

また、IRT による信頼性係数を推定 (荘島・豊田, 2002) することにより、テストの品質管理を行ったり、さらに、多層モデルの適用 (Mislevy & Bock, 1989; 荘島, 2003) などにより、テストを現象学的に眺めていくことも重要な作業である。

#### 文献

- 林 篤裕 (研究代表者), 2003, 大学入試センター試験の本試験と追試験の難易度比較に関する調査研究と標準化の試み, 大学入試センター研究開発部.
- 池田 央 (1994), 現代テスト理論, 朝倉書店.
- 石塚智一 (研究代表者), 2002, 大学入学者選抜における評価の標準化に関する研究, 大学入試センター研究開発部.
- Lord, F. M. (1980), *Applications of Item Response Theory to Practical Testing Problems*, Lawrence Erlbaum Associates.
- McDonald, R. P., 1997, Normal-Ogive Multidimensional Model., In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, Springer-Verlag (pp. 257-269).
- Mislevy, R. J. & Bock, R. D., 1989, A hierarchical item-response model for educational testing. In R. D. Bock (Ed.), *Multilevel analysis of educational data*, (pp.57-74).
- Reckase, M. D., 1997, A Linear Logistic Multidimensional Model., In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, Springer-Verlag (pp. 271-286).
- 荘島宏二郎, 2003, 実験計画のための項目反応モデル, 心理学研究, 73(6), 506-511.
- 荘島宏二郎・豊田秀樹, 2002, 項目反応理論における Cronbach の  $\alpha$  係数の推定, 心理学研究, 73(3), 227-233.
- 荘島宏二郎・豊田秀樹, 2004, テストが複数の項目形式を含むときの項目母数の推定, 教育心理学研究, 52(1), 61-70.
- Ueno, M., 2002, An extension of the IRT to a network model, *Behaviormetrika*, 29, 59-79.
- 吉村 幸 (研究代表者), 2004, 大学入試センター試験モニター調査研究報告, 大学入試センター研究開発部.
- 吉村 幸・中畝菜穂子・荘島宏二郎・石岡恒憲, 2003, 大学入試センター試験統計情報 データベース作成の試み, 日本テスト学会第1回大会発表論文抄録集, 43-46.
- 吉村 幸・荘島宏二郎・鈴木敦命, 2004, 多肢選択によるテストの分析に 2 パラメータ IRT モデルを適用した場合に起こること, 日本行動計量学会第 32 回大会発表論文抄録集, 372-375.