

## 項目反応理論を用いた試験問題の検討：共通教育心理学の例

阿久津 洋巳\*・石亀 雅哉\*\*

(2012年3月5日 受理)

Hiromi AKUTSU and Masaya ISHIGAME

Examining Examination Questions Using Item Response Theory : Psychology in general education

### 要 旨

大学生の共通教育科目の学力は、合理的な手立てとして多くの場合試験によって評価される。受講生の人数が多いため、客観式テストによる学力評価が望ましいが、信頼性のあるテストを作成することは容易なことではない。本研究は、項目反応理論を用いて、適正な評価を行うために必要な試験問題の選び方とその採点の方法について検討した。そのため実際の共通教育試験問題のデータに対して項目反応理論を適用し、不適格な問題項目を見出し、それらの問題項目を取り除いた採点結果と初めの試験問題を使った通常の採点結果とを比較した。項目反応理論を適用する有効性がいくつかの側面で明らかとなった。

### 1. はじめに

近年、大学においては学士の質の保証が新たな達成課題として浮き上がってきた(中央教育審議会, 2008)。その中心は、いかにして学士の能力を保証するかである。学士が身につけるべき諸種の能力のうち学力だけを取り上げても、専門教育科目の学力、共通教育(かつての一般教養)科目の学力は必須の構成要素であり、これらを客観的かつ成績基準を明確にして評価する手立てが必要である。これらの学力の一部は、長い間試験によって検査されてきた(池田, 1992; 辰野・石田,

2003)。

多くの共通教育の授業では、受講生の人数が多く、受講生が100人を超える授業も希ではない。評価の労力と時間、評価の客観性と信頼性を考えあわせると、学生の学力の評価には客観式テストが望ましい(Downing, 2006; 池田, 1992)。何年もテストをしている教員ならば、作問においていくつかの指針を持っている。例えば、大まかに解答が容易な問題、ほどほどに難しい問題、解答がかなり難しい問題という難易度の分類がある。試験の範囲とトピック、重要事項、習得すべき考え方などもある。選択式問題の場合、選択肢の数を同じにするかもしれない。さらに、受講学生の学力水準や問題数と問題の配置も考慮に入れる。後は、経験と勘と専門的知識に頼って問題を作りあげ、試験を実施して、あらかじめ設定した配点に従って、
$$\text{総計} = \sum \{ \text{各問の正答} \times \text{各問の配点} \}$$
で受験者ごとに得点をつける。場合によっては平常点や特別点などを加算して最終得点を算出するが、多くの場合得点の統計量には頓着しない。これが、良心的な教員が実施する学力評価のシナリオの1例であろう。

これらの作問と採点の過程を概観すると、上手なテスト問題を作成するのは芸術(もしくは職人技)に近い。内容の専門的知識を持つことは、問題作成に必須の要件ではあるが、専門的な知

\*岩手大学教育学部、\*\*岩手大学教育学研究科

識だけでは十分な要件を満たさない (Downing, 2006)。例えば不適切に作られた問題項目は、試験が測定しようとする構成概念 (例えば, ○○の学力) と無関係な正誤答のばらつき (分散) を作り出すため, 測定値の信頼性を低下させる。(信頼性を, テストの信頼性係数 (真の能力の分散 / テスト得点の分散) から考えることができる。) 従来から, 不適格な問題を発見するために教育評価では項目分析や G-P 分析など様々な技術が使われてきた。新しいテスト理論と呼ばれる項目反応理論 (Item Response Theory, IRT) は, 正に不適格な問題を見つけるのに有効な道具であり, 良問のみから成る試験を準備するために使えるだけでなく, すでに実施した試験から良問を選び出し合理的な得点を作成するために役立つ (熊谷, 2002; 豊田, 2002; 野口・斉田・孫, 2005; 尾崎・濱本・佐藤・野須, 2007; 月原・鈴木・廣瀬, 2008; 長谷川・鶴沼, 2010)。

本研究は, 試験による学力検査を適正に実施するために必要な試験問題の選び方とその採点について, 実際のデータにもとづいて考察する。具体的には, 試験問題の中にもどの程度不適格な問題が含まれていたかを項目反応理論 (IRT) を適用して検討し, さらに, それらを取り除いた最終版試験問題による採点結果と初めの試験問題を使った通常の採点結果を比較することである。最終版試験問題による得点は項目反応理論を使って算出した。

## 2. 方法

第一著者が実際に2011年度に岩手県内の大学で実施した中間試験と期末試験の問題を分析対象とした。科目は「心の理解」と「心理学」であり, 心理学の入門と行動科学でよく取り上げられるテーマ (例えば, 学習, 認知, 性格, 感情など) について講義した。以下, 3つの試験を心の理解1, 心の理解2, 心理学とよぶ。3つの試験では, 一部同じ問題が使用されたが, 3つの試験の受験者に重なりはない。試験問題数と受験者数を表1に示す。試験問題の形式は, 大部分が4つの選択枝

から正答を1つ選ぶ多枝選択式問題であった。5つの選択枝や3つの選択枝がごく少数含まれていたが, 本研究では, 全ての問題が同一の形式であるとみなして分析した。

表1 試験科目と問題数および受験者数

試験科目	心の理解1	心の理解2	心理学
問題数	41	50	80
受験者数	163	46	55

## 分析方法

3つのテストを個別に分析した。各テストでは, 全ての回答 (反応) に対して, 正答を1, 誤答を0として, 受験者と問題項目からなる行列にデータ入力した。統計的分析にはRとlrmパッケージを使用した (CRAN, 2011; Rizopoulos, 2006)。項目反応理論のモデルに関しては, 豊田 (2002) と Rizopoulos (2006) を参考にした。

## 3. 結果と考察

心の理解1, 心の理解2, 心理学の順に結果の要点を以下に述べる。

### 【心の理解1】

問題項目41を screeplot と単純構造の因子数を参考にして, 3因子 (promax 回転), 2因子 (promax 回転と varimax 回転) で探索的因子分析を行ったところ, どの場合でも第1因子以外では10問以上の問題項目が選ばれなかった。そこで, 1因子を仮定して項目反応分析を行った。2パラメータモデルを使用して困難度と識別力を41全ての項目について求めた。全項目に対する Cronbach の  $\alpha$  は 0.692であり高くなかった。ついで, 困難度と識別力を考えあわせて, 不適切な問題項目を17問削除した。削除の基準は識別力が低い (0.35以下), もしくは困難度が著しく低い (-4以下) か著しく高い (+4以上) 項目とした。このようにして作成された最終試験項目 (24問題項目) の困難度と識別力を全問の2パラメータともに付録の表5と表6に掲載した。

項目反応理論においては項目反応モデルの選択

も重要である。1パラメータモデル、2パラメータモデル、3パラメータモデルの3つのモデルのうち、どのモデルがデータ（24項目データ）によく適合するかを比較するために、likelihood ratioを計算した。2パラメータモデルは、1パラメータモデルよりも有意に適合度がよかった（likelihood ratio= 54.84, df=23,  $p<0.001$ ）が、3パラメータモデルは、2パラメータモデルと違いがなかった（likelihood ratio= 28.91, df=23,  $p>0.22$ ）。最終的に2パラメータモデルを採用した。

オリジナル41問を使った通常のテスト得点（合計得点）、41問に2パラメータ項目反応理論を適用した得点（全問IRT）と最終試験項目24問に項目反応理論を適用した得点（最終IRT）の3種類の得点を計算して、それらの関連を検討した。3つの得点の相関は非常に高かった（表2）。加えて、合計得点と最終IRT得点を散布図に表して比較した。比較を容易にするために全て平均50、標準偏差10の得点に変換した（図1）。

表2 3種類の得点の相関行列

	合計得点	全問IRT	最終IRT
合計得点	1	0.88	0.88
全問IRT		1	0.99
最終IRT			1

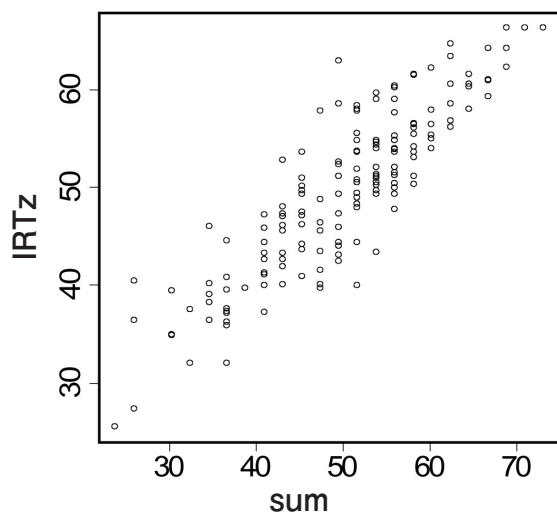


図1 全問の合計得点（横軸）と最終IRTの得点（縦軸）の関連

合計得点と最終IRT得点の間には強い相関（0.88）があるが、実際には同じ合計得点に対して幅広く最終試験IRT得点に対応している。予想されることであるが、全問試験と最短試験の情報量曲線はほぼ同じであった。最終試験問題全体の情報量曲線を図2に示す。この曲線から、最終試験は能力が比較的低い学生（ $z$ が0以下）の能力は適切に測定できるが、学力が高い学生を評価するのは向いていないことがわかる。

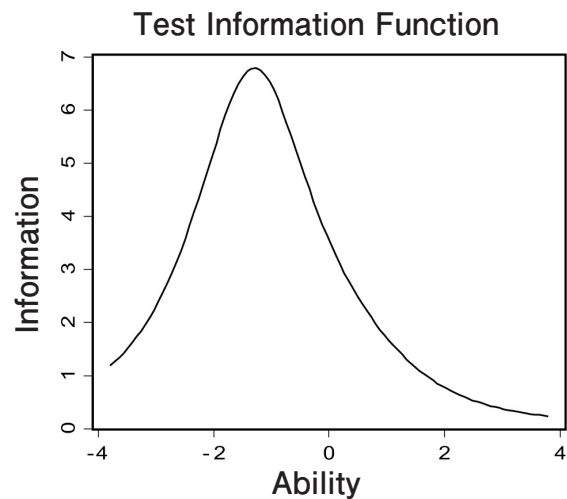


図2 最終試験の情報量

### シミュレーションによる能力値精度の推定

項目反応理論による得点（以下IRT得点）と通常の合計得点（以下素点）のどちらがテストが測定しようとする能力（以下 $\theta$ ）を適切に測定できるかを数値シミュレーションを行って検討した。シミュレーションでは、あらかじめ決められた $z$ 値を真の能力値（ $\theta$ ）とした。この値として実際のIRT得点163人分を使った。これらの受験者の正答と誤答をシミュレートするため0から1の間の一様乱数（ $R$ ）を各受験者の各問題に対して発生させた。この $R$ と試験問題の項目特性曲線が潜在能力を横座標とした $\theta$ を横切る位置の確率（ $P$ ）を比較して、 $P > R$ ならば正答（1）、 $P < R$ ならば誤答（0）として、受験者の反応パターンを作成した（月原、鈴木、廣瀬、2008）。したがって、反応パターンは受験者（163） $\times$ 問題（24）の行列にその要素が1もしくは0が記録された163行24

列からなっていた。反応パターンをもとに素点と項目反応理論によるIRT得点を計算した。IRT得点は標準化されたz得点であるが、素点は異なる。両者を比較可能な数値にするために、素点はその平均と標準偏差を使って、標準化得点(z)に変換した。このシミュレーションを100回実施して、素点とIRT得点の2つについて平均2乗平方根(RMSE)を計算し、能力値推定精度の指標とした。その結果は、項目反応理論では0.532、素点では0.484であった。わずかながら素点の方が得点の精度が高かった。

### 【心の理解2】

この試験の問題数は50問、受験者は45名であった。50問のうち、受験者全員が正答した問題が6問あったので、この問題を除いた44問を分析対象とした。(問題数に対して受験者が少なかったため因子分析は省略した。)まず、2パラメータ項目反応理論を適用して、各問題の困難度と識別力を調べた。困難度と識別力を考えあわせて、問題を削除し最終的に22問の試験問題を良問として残した。良問は、識別力が高く(0.5以上)、困難度はやや低かった(-3.67~-0.56)。最終試験項目22問の困難度と識別力に加えて全問の2パラメータを付録の表7と表8に掲載した。

心の理解1の項目分析と同様に1パラメータモデル、2パラメータモデル、3パラメータモデルのうちどのモデルが22問のデータに適合するかをlikelihood ratioを計算して調べたところ、2パラメータモデルは、1パラメータモデルよりデータに対する適合がよかった(likelihood ratio=36.64, df=21, p<0.02)。2パラメータモデルと3パラメータモデルの間に適合度に有意差はなかった(likelihood ratio=24.36, df=22, p>0.3)。

オリジナル50問を使った合計得点、50問に2パラメータ項目反応理論を適用した全問IRTと最終試験項目22問に項目反応理論を適用した最終IRTの3種類の得点を計算して、それらの関連を検討した。合計得点と最終IRTの相関は高かった(表3)。また、合計得点と最終IRT得点を散布図に

表して比較した。比較を容易にするために全て平均50、標準偏差10の得点に変換した(図3)。合計得点と最終IRT得点の間には強い相関(0.81)があるが、実際には同じ合計得点に対して幅広く最終試験IRT得点に対応しており、この傾向は合計得点が50以上で顕著であった。

表3 3種類の得点の相関行列

	合計得点	全問IRT	最終IRT
合計得点	1	0.69	0.81
全問IRT		1	0.92
最終IRT			1

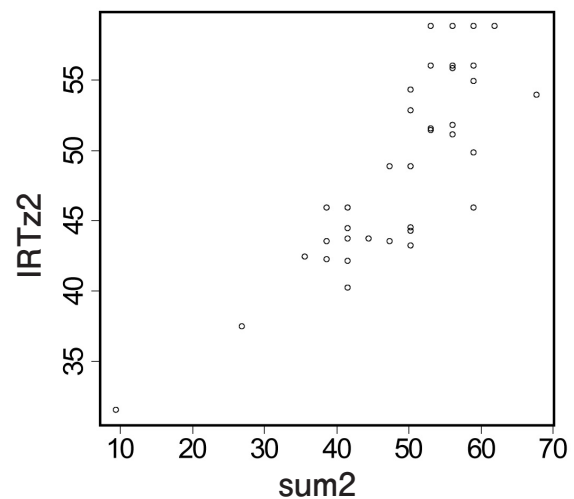


図3 全問の合計得点(横軸)と最終IRT(縦軸)の関連

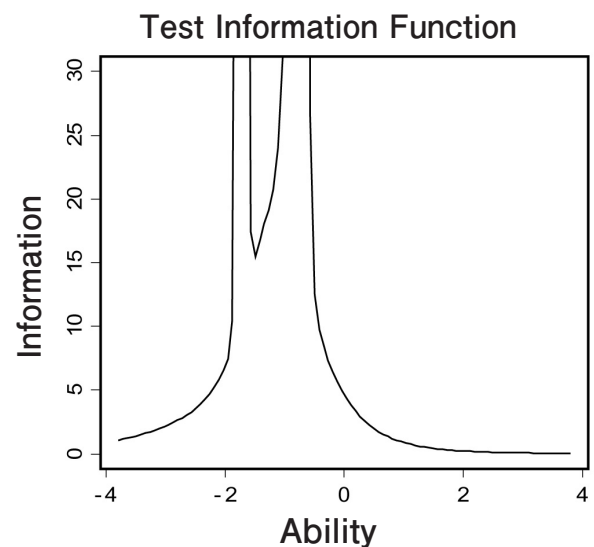


図4 最終試験の情報量



最終試験の情報量を見ると（図4）非常に識別力が高い項目が4問あったため、それらに対応する能力値付近で情報量が大きい。この点を除けば、心の理解1の試験と同じく能力測度が0以下の能力に対して試験の精度が高いが、0以上では精度が悪く、比較的学力が低い学生向きの試験であった。

#### シミュレーションによる能力値精度の推定

実際のIRT得点（z値）を45人分使って、それぞれに先ほどの正誤反応を発生させる方法で反応パターンを作り、シミュレーションを100回実施した。RMSEはIRT得点では0.411、素点では0.606であり、項目反応理論の方が多少精度が高かった。

#### 【心理学】

問題数は80、受験者は55名であった。50問のうち、受験者全員が正解した問題が1問あった。この1問を除いた79問題を分析対象とした。（問題数に比べて受験者数が小さいので因子分析はできなかった。）まず、項目反応理論を適用して、各問題の困難度と識別力を調べた。困難度と識別力を考えあわせて、問題を削除し、最終的に49問の試験問題を良問として残した。

49問の試験問題を対象にして、1パラメータモデル、2パラメータモデル、3パラメータモデルから最適なモデルを選んだ。予想に反して1パラメータモデルと2パラメータモデルの適合度に有意差はなかった（likelihood=45.27, df=48, p>0.5）ため、1パラメータモデルを採用した。

オリジナル79問を使った合計得点、79問に2パ

ラメータ項目反応理論を適用した全問IRTと最終試験項目49問に項目反応理論を適用した最終IRT（2パラメータ）および最終IRT（1パラメータ）の4種類の得点を計算して、それらの関連を検討した。4つの得点の相関は非常に高かった（表4）。加えて、合計得点と最終IRT得点を散布図に表して比較した。比較を容易にするために全て平均50、標準偏差10の得点に変換した（図5）。これまでの2つの試験に比べると、合計得点に対応する最終IRT得点のばらつきの幅は小さかった。

テスト情報量（図6）を見ると、能力0の付近でも相対的には情報量が多い点これまでの2つの試験と異なる。大まかに見ると能力測度1付近まで高い精度で学力の評価ができそうである。

#### シミュレーションによる能力値精度の推定

実際のIRT得点（z値）を55人分使って、それぞれに上記の正誤反応を発生させる方法で反応パターンを作り、シミュレーションを100回実施した。RMSEはIRT得点では0.418、素点では0.322であり、素点の方がわずかに精度が高かった。

この試験問題は、Raschモデルが適合した例である。Raschモデルは、全ての問題の識別力が一定であり項目特性曲線が項目間で交わらない特異な項目セットである（図7）。Raschモデルが適合する場合は、素点とIRT得点に実質的の違いがない。2得点間の相関は、0.995であった（図8）。言い換えれば、素点の採点で十分能力評価ができるため、試験を実施、採点する立場からは都合がよい。

表4 4種類の得点の相関行列

	合計得点	全問IRT	最終IRT (2パラメータ)	最終IRT (1パラメータ)
合計得点	1	0.9	0.9	0.94
全問IRT		1	0.99	0.98
最終IRT (2パラメータ)			1	0.98
最終IRT (1パラメータ)				1

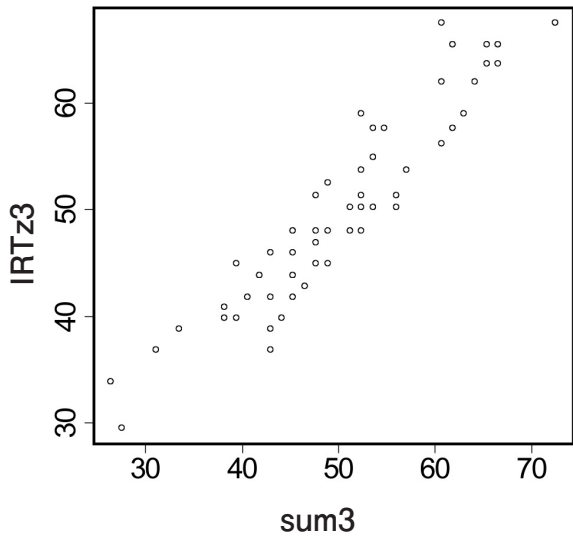


図5 合計得点（横軸）と最終試験による1パラメータモデル得点（縦軸）の関連

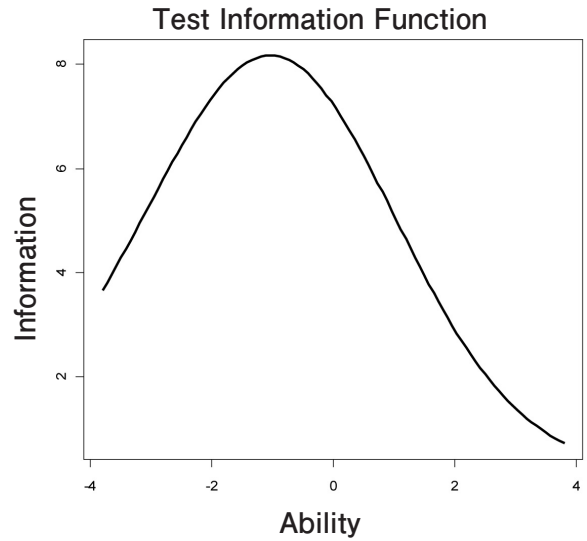


図6 適合モデルのテスト情報量を潜在特性の関数として表す。テスト全体（49問）の情報量は、-1付近が最大である。

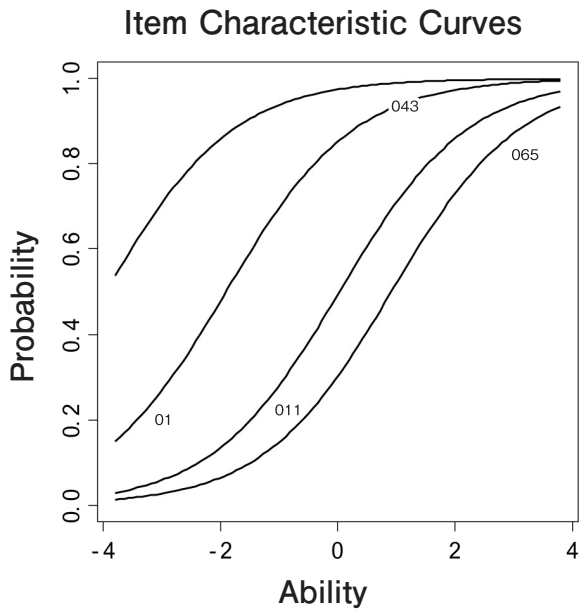


図7 項目特性曲線を4つの代表的な問題について示した。Raschモデルが適合するので4つの曲線は全て傾き（識別度）が等しい。1番易しい問題から中程度の問題，やや難しい問題，1番難しい問題の4つを選んで，その項目反応曲線を示した。

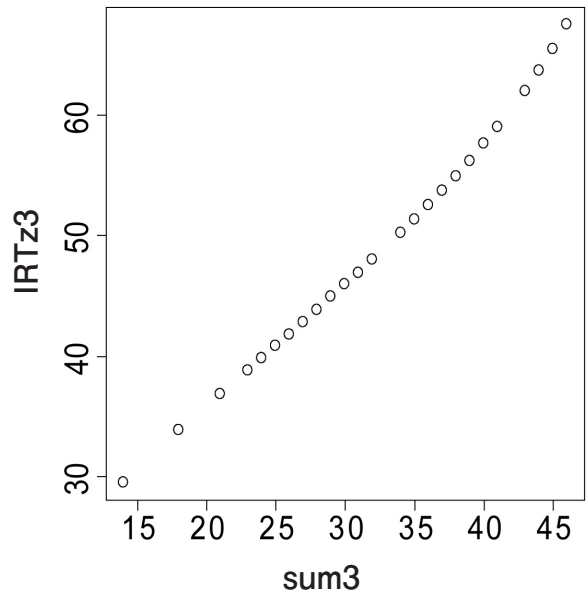


図8 素点とIRTzの関連を示す。Raschモデルが適合するので，素点とIRTzの間には1対1の対応関係がある。

#### 4. まとめ

項目反応理論を実際の試験問題に適用することにより、いくつかの問題点が明確になった。

- (1) 試験問題の選択 学力評価に使えない試験項目が多数検出された。試験問題を作成した第一著者が特別に作問能力が劣るのかもしれないが、他の教員の試験問題に対しても本研究で使われたのと同様な方法で問題の性質を検討すべきであろう。成績評価に有効ではないことが判明した試験項目は修正する必要がある。
- (2) 試験問題と学力水準 良問だけを残した最終版試験問題が高い精度で弁別できる学力水準を明確にできた。テスト情報量関数から、3つの試験はどれも得点が低い方で学力の判別が適切にできることが分かった。学力が高い学生を適切に評価しようとした問題項目が適切に機能せず、識別力が0に近いが、マイナスとなる項目特性曲線が多かった。したがって、学力が高い学生を適切に評価できる問題項目を開発することが、実践的課題として明らかになった。
- (3) 素点利用への道 3番目の試験問題（心理学の試験）にはRaschモデルが適合した。この最終版試験問題については、項目反応理論を適用して採点しても、素点でも実質的な違いがなかった。Raschモデルが適合する問題項目の集合を用意するならば、実施と採点が容易な試験問題を作ることが確認できた。採点の簡略化を考慮するならば、Raschモデルが適合する試験問題の作成を目指すべきであろう。

#### 6. 引用文献

- CRAN R Development Core Team (2011)  
URL <http://www.R-project.org/>.
- 中央教育審議会 (2008) 学士課程教育の構築に向けて (答申)
- Downing, S.M. (2006) in *Handbook of Test Development*, Downing, S.M., & Haladyna, T.M.,

(Eds.) テスト作成における解答選択式問題の形式 池田央監訳 テスト作成ハンドブック 教育測定研究所

- 長谷川桐・鶴沼秀行 (2010) 心理学における統計教育のための評価項目の選択 項目反応理論による分析 日本教育心理学会総会発表論文集 (52), 730
- 池田央 (1992) テストの科学 —試験にかかわるすべての人に— 日本文化科学社
- 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科学研究科紀要, 心理発達科学49, 47-54
- 野口裕之・斉田智里・孫媛 (2005) 項目応答理論の基礎と応用 —学力テストの分析を通して— 教育心理学年報44, 32-34
- 尾崎将範・濱本和彦・佐藤実・野須潔 (2007) 項目反応理論を用いた信頼性のある能力評価システムの研究 情報科学技術フォーラム一般講演論文集6 (3), 595-598
- Rizopoulos, D. (2006) ltm: An R package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, 17, 1-25.
- 辰野千尋・石田恒好 (2003) 教育評価法概説 図書文化
- 月原由紀・鈴木敬一・廣瀬英雄 (2008) 項目反応理論による評価を加味した数学テストと e-learning システムへの実装の試み 九州工業大学学術機関リポジトリ
- 豊田秀樹 (2002) 項目反応理論 [入門編] 朝倉書店

#### 付録

1. 全試験問題の困難度と識別力パラメータと最終版試験問題の困難度と識別力パラメータを表にして示す。

## (1) 心の理解 1

表 5 全試験問題

問題番号	困難度	識別力
Q1	-1.467	1.971
Q2	-0.927	1.358
Q3	-1.963	0.964
Q4	-1.163	2.473
Q5	-0.833	1.303
Q6	-1.224	0.439
Q7	-1.954	1.792
Q8	-2.419	1.008
Q9	-0.931	0.793
Q10	-8.907	0.278
Q11	36.707	0.018
Q12	-4.449	0.459
Q13	-6.722	0.386
Q14	-1.597	0.461
Q15	54.948	-0.031
Q16	-0.843	0.633
Q17	-1.192	-0.156
Q18	-4.816	0.463
Q19	-2.475	0.673
Q20	-1.388	1.459
Q21	-6.361	0.533
Q22	-0.653	1.167
Q23	-6.004	0.206
Q24	2.036	0.370
Q25	4.125	0.245
Q26	8.250	-0.125
Q27	-3.114	0.360
Q28	-1.724	0.894
Q29	-0.398	0.737
Q30	4.810	0.113
Q31	-0.987	1.285
Q32	-25.308	0.047
Q33	-3.124	0.183
Q34	-1.967	0.683
Q35	-1.152	1.467
Q36	0.365	1.417
Q37	43.638	-0.063
Q38	9.272	-0.298
Q39	-0.300	0.688
Q40	2.206	-0.176
Q41	-1.749	0.492

表 6 最終版試験問題

問題番号	困難度	識別力
Q1	-1.53	1.795
Q2	-0.95	1.313
Q3	-2.02	0.93
Q4	-1.23	2.123
Q5	-0.82	1.332
Q6	-1.1	0.493
Q7	-1.95	1.813
Q8	-2.35	1.049
Q9	-0.93	0.792
Q14	-1.66	0.442
Q16	-0.87	0.606
Q19	-2.48	0.672
Q20	-1.41	1.423
Q22	-0.65	1.163
Q24	2.113	0.356
Q27	-3.16	0.355
Q28	-1.61	0.983
Q29	-0.4	0.722
Q31	-0.98	1.305
Q34	-1.87	0.729
Q35	-1.13	1.523
Q36	0.366	1.426
Q39	-0.3	0.682
Q41	-1.74	0.496

## (2) 心の理解 2

表 7 全試験問題

問題番号	困難度	識別力
Q1	-0.878	2.922
Q2	-0.77	2.437
Q3	-1.098	1.779
Q4	-0.852	2.437
Q5	-0.675	1.245
Q8	-1.721	48.674
Q9	-1.161	0.001
Q10	2.795	-0.19
Q13	2.69	-2.154
Q14	-2.316	1.104
Q15	-1.435	2.334
Q17	-0.774	3.266
Q18	-1.289	1.377
Q19	-0.711	34.624
Q20	-2.062	1.532
Q21	-2.207	0.448
Q22	-6.009	-0.228
Q23	1.414	-16.483
Q24	-6.374	0.603
Q25	-4.187	0.956
Q26	17.157	-0.181
Q27	-13.76	0.168
Q28	24.229	-0.097
Q29	-2.782	0.689
Q30	-1.321	3.852
Q31	-5.932	0.349
Q32	-0.682	0.818
Q33	-0.761	27.699
Q34	-1.262	4.468
Q35	-2.268	-0.868
Q36	-1.682	1.469
Q38	2.2	-1.465
Q39	-1.987	1.017
Q40	-2.373	1.538
Q41	-1.721	48.674
Q42	6.345	-0.129
Q43	8.416	-0.096
Q44	6.9	-0.168
Q45	3.627	-0.369
Q46	2.064	-25.226
Q47	-1.495	-0.082
Q48	-1.043	0.351
Q49	4.632	-0.349
Q50	5.402	0.328

表 8 最終版試験問題

問題番号	困難度	識別力
Q1	-0.861	2.968
Q2	-0.785	2.23
Q3	-1.079	1.8
Q4	-0.838	2.443
Q5	-0.56	1.598
Q8	-1.721	48.822
Q14	-2.493	1.003
Q15	-1.42	2.355
Q17	-0.798	2.786
Q18	-1.299	1.347
Q19	-0.709	35.394
Q20	-2.066	1.517
Q21	-1.403	0.714
Q29	-3.67	0.511
Q30	-1.31	3.922
Q32	-0.826	0.664
Q33	-0.86	12.591
Q34	-1.277	4.278
Q36	-1.635	1.519
Q39	-2.056	0.97
Q40	-2.658	1.303
Q41	-1.721	48.822



## (3) 心理学

表9 全試験問題

問題番号	困難度	識別力
Q1	-3.659	0.431
Q2	1.148	0.284
Q3	-1.943	0.549
Q4	-5.262	0.208
Q5	-1.834	0.802
Q6	-0.962	1.530
Q7	-3.908	0.401
Q8	0.203	0.802
Q9	-2.486	0.416
Q10	-2.227	1.033
Q11	0.045	0.432
Q12	0.491	1.33
Q13	0.799	0.619
Q14	1.995	0.09
Q15	40.237	0.02
Q16	-7.153	0.411
Q17	-0.234	0.651
Q18	-28	0.103
Q19	-1.84	0.799
Q20	-1.338	1.94
Q21	-1.969	1.595
Q22	-2.009	2.16
Q23	-0.332	-0.732
Q24	-0.997	-1.037
Q25	0.185	1.858
Q26	-0.567	0.951
Q27	-1.73	1.073
Q28	-59	0.018
Q29	1.344	0.242
Q30	-3.824	0.263
Q31	0.074	0.955
Q32	2.538	0.252
Q33	-2.482	0.669
Q34	-0.652	1.244
Q35	-1.105	1.178
Q36	22.609	0.052
Q37	-14	-0.091
Q38	-0.824	0.336
Q39	0.57	0.576
Q40	-5.168	-0.252
Q41	-1.128	-0.31
Q42	-0.689	0.706
Q43	-2.997	1.325
Q44	1.332	0.425
Q45	-0.453	1.584
Q46	-11	0.211
Q47	15.566	-0.036
Q48	-5.576	-0.162
Q49	-2.208	0.084
Q50	-22	0.062

表10 最終版試験問題

問題番号	困難度
Q1	-1.909
Q2	-1.267
Q3	-1.632
Q4	-1.267
Q5	-1.909
Q6	0.2173
Q7	-1.267
Q8	-2.412
Q9	0.028
Q10	0.7061
Q11	0.6054
Q12	-0.16
Q13	-1.631
Q14	-2.062
Q15	-2.854
Q16	-3.485
Q17	0.409
Q18	-0.54
Q19	-1.909
Q20	0.1228
Q21	-1.909
Q22	-0.737
Q23	-1.267
Q24	0.4091
Q25	-0.54
Q26	-3.966
Q27	0.7062
Q28	-0.54
Q29	-1.632
Q30	-0.638
Q31	-0.737
Q32	-1.505
Q33	-2.854
Q34	-1.046
Q35	-1.046
Q36	-0.737
Q37	0.9139
Q38	0.7061
Q39	-1.046
Q40	0.7061
Q41	-0.349
Q42	-1.383
Q43	-2.228
Q44	-0.444
Q45	-3.134
Q46	-1.267
Q47	-2.062
Q48	-3.485
Q49	-0.941
識別力	0.9157