

も く じ

監修にあたって	1
はじめに	2
テスト用語一覧	5

第 1 章 なぜ、この本を書くことにしたのか 16

1 IRT に関連した仕事を通じて感じた必要性	16
(1) IRT はよくわからない?! 16	
(2) 大学入試センター試験がなくなる衝撃と戸惑いに答えるために	17
2 寄せられた IRT への疑問、不安、心配、不信	18
3 あえて IRT を一言で言えば	21
(1) 学力を数値化する方法の一つ 21	
(2) 異なる問題のテスト結果、異なる受験者集団の テスト結果を比較できる 22	
(3) IRT は魔法の杖ではない 23	
4 IRT を使う条件	24
(1) 問 1 の答えをふまえて問 2 に解答するような問題を 出題してはいけない 24	
(2) 一つのテストに出題する問題は すべて同じ学力を問うていること 26	
5 IRT への疑問や不安に答えると	27
(1) そういうことはありません、誤解です 27	
(2) 用意する問題の数は状況次第 28	
(3) 過去問題の利用や問題非公開は IRT を使うテストの 条件ではない 28	
(4) 学力向上と得点向上の関係は従来のテストと同じ 29	
(5) 誤解がある「～しなくてはならない」 29	

第2章 学力を数値化する測定理論 31

- 1 「学力を数値化する」とは 31
 - (1) “しかめ面さん” 31
 - (2) テストの得点は絶対か? 32
- 2 テストは非破壊検査に似ている 33
 - (1) 学力は外からわからない 33
 - (2) 頭の中はブラックボックス 33
 - (3) テストは仮定から始まる 34
- 3 学力を数値化する過程 35
 - (1) 採点と採点基準 35
 - (2) 得点化 37

第3章 従来の「テスト得点」 38

- 1 第1グループの得点の求め方 38
 - (1) それぞれの得点を一言で言えば 38
 - (2) 問題の難度が反映されない「正答数得点」 39
 - (3) 配点の決め方が重要な「重みつき正答数得点」 40
 - (4) 2通りの配点で起きたこと 42
 - (5) 事前にやっておくべきだったこと 44
 - (6) 100点満点の呪縛 45
 - (7) テスト得点に潜む「あいまいさ」 46
- 2 第2グループの得点の求め方 47
 - (1) 得点の比較ができる第2グループ 47
 - (2) 平均値 48
 - (3) 標準偏差 49
 - (4) 標準偏差の求め方 50
 - (5) 標準得点とその求め方 52
 - (6) 偏差値とその求め方 53

(7) 同じ集団でも比較できない偏差値	54
(8) 測定範囲が変わる偏差値	55
(9) 変化の仕方がわからない偏差値	56
③ 従来の「テスト得点」の問題点	59
④ IRT は「従来のテスト得点」の問題点を解決するのか	60

第4章 IRT による学力推定の方法・その1 64

① 学力を測る「ものさし」を事前に用意する IRT	64
② 「学力 θ 」と「尺度 θ 」	65
③ 視力検査に似ている IRT	66
④ 輪の大きさによって決まっている視力	68
⑤ 項目特性	68
(1) 学力によって変わる「正答できる可能性」	68
(2) IRT モデル	69
⑥ 2パラメーター・ロジスティック・モデル	70
(1) 困難度と識別力によって項目特性曲線の形が決まるモデル	70
(2) 困難度の違いによる項目特性曲線の変化	72
(3) 困難度は正答確率が0.5となる「学力 θ 」	73
(4) 「位置パラメーター」とも呼ばれる困難度	75
(5) 識別力の違いによる項目特性曲線の変化	76
(6) 識別力の大きさの制限	78
⑦ 仮定した「ものさし」	79
⑧ 測定値の意味づけ	80
⑨ これまでの質問例 ～ 簡単なまとめにかえて	82

第5章 IRT による学力推定の方法・その2 84

① 「学力 θ 」の推定法	84
----------------------	----

2 「学力 θ 」を推定する確率計算	86
3 重要な「局所独立の仮定」	88
4 A 君の「学力 θ 」	89
5 最尤推定法	90
6 全問正答、全問誤答の場合	92
7 テストの測定精度	92
(1) テスト情報量	92
(2) テストの実施目的に合ったテスト情報量	94
8 項目パラメーターの推定法	96

第 6 章 IRT による等化 99

1 共通尺度	100
2 等化	103
3 等化係数	107
4 等化デザイン	108
(1) 共通項目デザイン	108
(2) 共通受験者デザイン	112
(3) 等化は必ず成功するのか	115

第 7 章 その他の IRT モデル 116

1 二値型モデルと多値型モデル	116
2 1 パラメーター・ロジスティック・モデル	117
3 3 パラメーター・ロジスティック・モデル	118
4 段階反応モデル	119
(1) 項目反応カテゴリー特性曲線	119
(2) 段階反応モデルを使うときの注意点	122
5 名義反応モデル	124

第8章 IRT を使ったテストの開発と実施 127

1	テストの実施目的の明確化	128
(1)	何のためのテストなのか	128
(2)	何を測るテストなのか	129
(3)	どんな問題を出題するのか	129
(4)	IRT を使うなら	130
(5)	なぜIRT を使うのか	131
2	テストの出題企画の立案	132
(1)	どのようなテストにするのか	132
(2)	平均点の設定	133
(3)	テストの実施目的によって異なる得点分布	134
3	テストの等化計画の立案	135
4	テストの実施計画の立案	136
(1)	コンピュータを使うテスト	137
(2)	CAT の場合はIRT が必要	137
(3)	必要な項目数	139
5	テストの収支計画の立案	140
6	テストの分析企画の立案	141
(1)	分析は2段階	141
(2)	統計的な手法で分析	141
7	テストの活用計画の立案	142
(1)	教育の実践者と検討	142
(2)	評価基準作り	142
8	テスト冊子の作成	143
9	テストの実施	144
10	テスト結果の分析	145
(1)	第0段階の分析	145
(2)	IRT の第1段階の分析	146
(3)	スクリー・テスト	146

(4) 信頼性係数の推定	148
(5) 項目分析	149
(6) 正答率（通過率）	150
(7) 点双列相関係数	150
(8) G-P 分析（Good-Poor analysis）	152
(9) G-P 分析 1 ～ うまくいった例	153
(10) G-P 分析 2 ～ うまくいかなかった例	153
(11) G-P 分析 3 ～ 注意喚起の例	156
(12) IRT の第 2 段階の分析	158
11 テストの評価基準の作成	159
12 テスト結果の活用	160

第 9 章 項目プールの開発 161

1 項目プール（アイテムプール）	161
2 項目プール開発に必要な役割と体制 1	163
3 項目プール開発に必要な役割と体制 2	165

第 10 章 良質な問題は国の宝だ 170

1 IRT を使ったテストの実施状況	170
2 IRT を使ったテストの実現に向けて	171
(1) 項目作成	171
(2) 等化	172
(3) 専門性	172
3 項目プールを「みんなで作って、みんなで使う」	173
(1) みんなで作る	174
(2) みんなで使う	175
4 知のバックボーン ～「おわりに」にかえて	177

IRT では、テストの実施後に、1次元性が保たれているか（成り立っているか）を確認します。そして、1次元性の確認ができて、初めてIRTの分析に取りかかるのです。仮に1次元性が十分に成立っている場合でも、1次元性に悪い影響のある問題（＝一つのテストで問うている「同じ学力」とは別の学力を問うている問題）は除外したうえでIRTの分析に取りかかります。そのくらい重要なことだと捉えてください。

なお、IRTの中には、「多次元IRTモデル」と呼ばれる分析方法もあります。このモデルでは、1次元性が保たれていなくてもIRTを使うことが可能です。ただし、現時点では「多次元IRTモデル」が使われる場面は決して多くはないと判断し、本書ではIRTが使える条件としてテストの1次元性を取り上げました。

以上、IRTとは何か、IRTの特長、IRTを使う条件を取り上げました。これらをふまえて、あらためて、皆さんから寄せられた疑問や不安などを検討してみましょう。

5 IRT への疑問や不安に答えると

(1) そういうことはありません、誤解です

- ① IRT を使うには、一問一答式の簡単な問題でなければならない。
- ② IRT を数学のテストで使うときには、必要な公式を提示しなくてはならない。
- ③ IRT を使うと、思考力、判断力、表現力が問えるようになる。

どういう内容の問題をどういう形式で出題しようとも、解答が正確に合理的な基準に則って採点されていれば、原則としてIRTでの分析は可能です。ですから、①・②のようなことはありません。③については、IRTを使いさえすれば、思考力、判断力、表現力が測定できるということはありません。思考力、判断力、表現力を測るには、どういう問題が適しているのか。そして、いかに採点すべきなのか。これはIRTを使うかどうかと



図2 テストは学力を仮定することから始まる

クボックスの中にあると仮定した学力を、おそらくこうではないか、と推定するのです（図2の④）。このように、「テストは仮定することから始まる」と思うのですが、いかがでしょうか。

3 学力を数値化する過程

学力を数値化する過程は、「採点」の過程と「得点化」の過程から成ります。ここでの採点とは、受験者の解答を、正答、部分的に正答、誤答等に分類することとします。そして、採点結果から得点を求める過程を、「得点化」として、両者を区別したいと思います。

(1) 採点と採点基準

まず、**採点**について考えましょう。採点をおこなう前に、解答を分類するための基準を作っておかなくてはなりません。それが、**採点基準**です。採点基準は、正答として求める条件（ただし、一つとは限りません）、とも

ある項目の項目特性とは、「学力 θ 」の違いに応じて、その項目に正答できる確率（＝「正答確率」）がどのような値になるのか

ということです。「学力 θ 」の違いに応じてということですから、受験者の学力の高い低いの違いに応じて、その項目に正答できる確率はどのように違ってくるのか、ということになります。

この様子を、図で表わしたものを**項目特性曲線**と言います。項目特性曲線はどのような形になるのでしょうか。学力が低い受験者から高い受験者へとなるにつれて、正答できる可能性は徐々に高くなり、ある学力を超えると一気に高まって、再び緩やか上昇となると考えるのが、一般的だと思います。図6に示したような緩やかなS字カーブの曲線です。

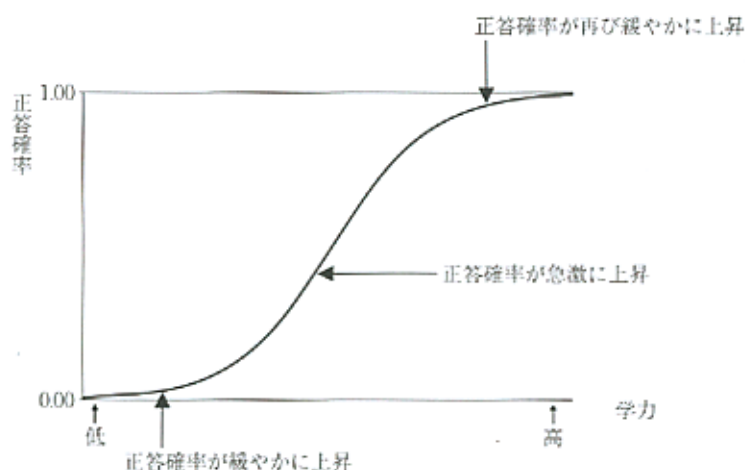


図6 学力の違いによる正答確率の違い

(2) IRT モデル

項目特性は、テストを実施した結果をIRTで分析することで得る

ことができます。分析によって得られた項目特性は、個々の項目に固有のものとなります。そのため、基本的に、項目ごとに項目特性曲線の形は