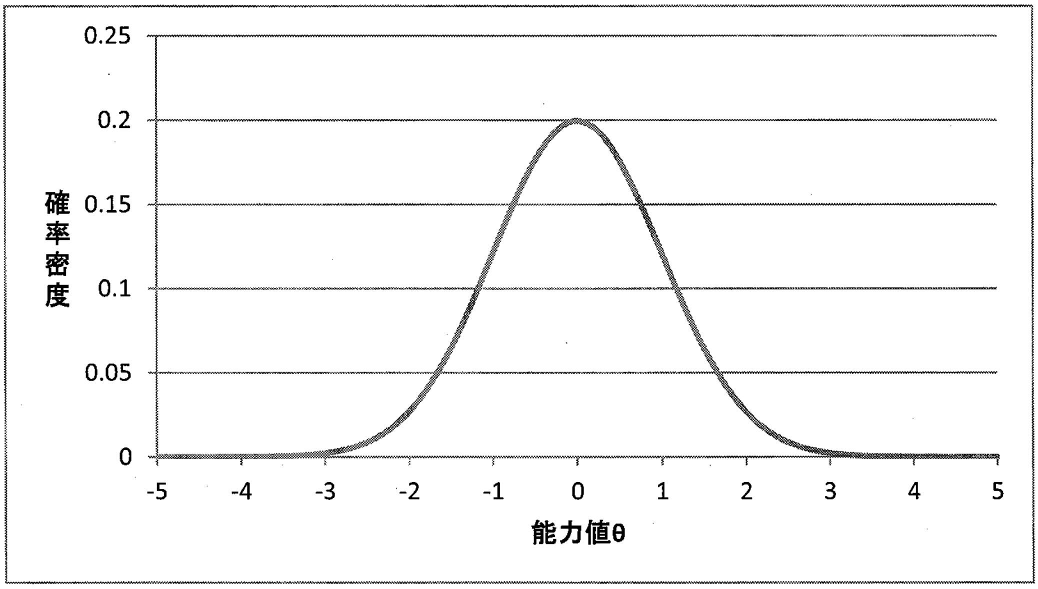
第３章　IRT（項目反応理論）による調査結果の分析

３．１　IRT とは

IRT（項目反応理論）とは，Item Response Theory の略で，項目応答理論とも呼ばれるテスト理論である．IRT にはいくつかのモデルがあるが，ここでは識別力と困難度の 2 つのパラメータを用いて受験者の能力値を推定する 2 パラメータロジスティックモデル（2PL モデル）を適用し分析を行う．

能力値とは，各受験者の能力を表す数値で，これまでのこの基礎学力調査の結果からほぼ標準正規分布に従うことが分かっている（ただし，一般に IRT 分析の結果の能力値分布は，標準正規分布になるとは限らない）．0 が平均であり，1 以上の割合は約 16%，2 以上の割合は約 2%である．偏差値と同等のイメージで考えると分かりやすい．能力値 0＝偏差値 50，能力値 1＝偏差値 60 である．

図 3.1　　標準正規分布の確率密度関数



識別力とは，その問題が受験者の能力を推定する際の効力を表す数値であり，値はおよそ 0～2 の範囲で大きいほど効力が大きい．困難度とは，問題の難易度を表す数値であり，値はおよそ－2～2 の範囲で大きいほど難しい．

困難度と識別力 2 つのパラメータを推定すると，問題ごとの項目特性曲線（ICC, Item Characteristic Curve）と呼ばれるグラフが得られる（図 3.2）．すなわち，θ を変数として，つぎの関数で与えられるグラフである。

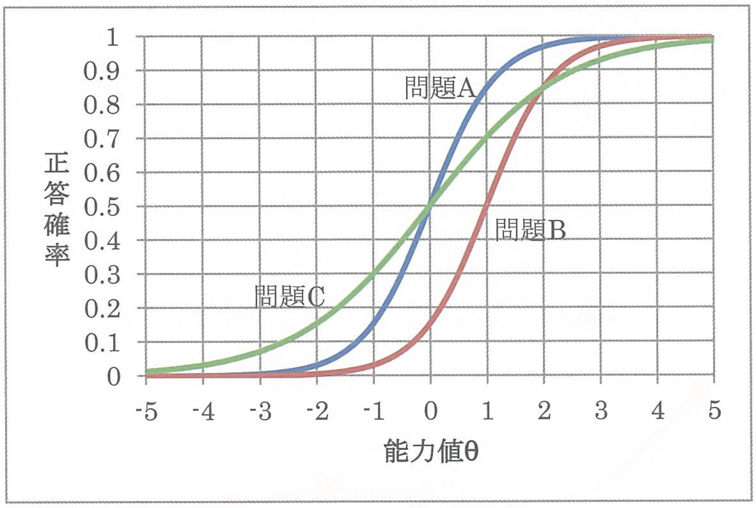
$$p_{a,b}(\theta) = \frac{1}{1 + e^{-1.7\times a(\theta - b)}}$$

1 つの問題に対して，その識別力（a），困難度（b）が定まったとすると，θ という能力をもった受験者がこの問題に正答する確率が $p_{a,b}(\theta)$ で得られるという意味である。図 3.2 の問題 A は，能力値－1 の受験者の正答確率はおよそ 0.15 であるが，能力値 0 の受験者の正答確率は 0.5 である．

このグラフの形状からその問題の特徴を見て取ることが容易になる．図 3.2 の問題 A と問題 B は，識別力はともに 1 であるが，困難度が問題 A は 0，問題 B は 1 である。正答確率が 0.5 の能力値を見ると，

Aが0であるのに対しBが1であり，Bの方が難しい問題である．問題Aと問題Cは，困難度はともに0であるが，識別力はAが1，Bが0.5である．Aの方が困難度0付近のグラフの立ち上がりが急で，能力値が0以上か否かで，その問題に正解か不正解かが決まりやすいといえる．

図3.2　項目特性曲線



3．2　IRTの適用

IRTの適用手順については下記のⅠ～Ⅲに従う。

Ⅰ．年度ごとに各問題の識別力と困難度の2つのパラメータを推定

2005～2013年度に出題された82問（項目）の全受験者38,187人の試験結果から2PLモデルの項目パラメータ（識別力パラメータと困難度パラメータ）を年度ごとに推定する．

Ⅱ．共通項目法による項目パラメータの等化

各年度共通して出題されている問題が多数あるので，受験者数が比較的に多かった2007年度（4,575人）の44項目の推定値を基準に，共通項目法により全82項目のパラメータの等化を行う．これにより異なる年度間でも同一尺度（能力値）での比較が可能になる。

Ⅲ．最尤推定法による能力値推定

等化された項目パラメータをもとに最尤推定法により全受験者の能力値の推定を行う．最尤推定法を用いると全問正解と全問不正解の受験者の能力値は，前者は正の無限大に，後者は負の無限大となってしまうので，これらを除く36,216人について能力値を推定する．

以下の分析は，この全問正解と全問不正解を除いた受験者について行う．

3．3　平均値の推移

この調査では，各年度4つの問題セットを用意し，各セットは2005年が10問，それ以外は11問を出題した．表3.1および図3.3は年度ごとの平均得点率の推移を示したものである．ここで得点率とは，セットの全問題数のうちの正解した問題数の割合(%)のことで，ある集団（全受験者や共通10校の受験者）の得点率の平均を，その集団の「平均得点率」と記す．共通10校とは，9年間すべて受験した10校のことである．

2013年度は9年間で最も低くなっていることが見てとれる．注目すべきは，満点の割合で平均得点率との相関が高く(2005～2013年度で相関係数は0.92，平均得点率と満点の割合が特に高い2005年度

を除いた 2006～2013 年度の相関係数は 0.79)，2013 年度はやはり 9 年間で最も低い 2.7%であった．

表 3.1　平均得点率の推移

	2005年度	2006年度	2007年度	2008年度	2009年度	2010年度	2011年度	2012年度	2013年度	全年度
全受験者数	2,237	3,365	4,575	5,300	2,900	3,792	4,084	5,902	6,032	38,187
全受験者の平均得点率(%)	68.2	58.2	55.6	56.9	58.1	56.3	55.9	55.2	53.1	56.6
共通10校の平均得点率(%)	63.0	58.2	55.3	54.4	53.3	52.7	53.5	53.6	51.5	54.5
全受験者の満点の割合(%)	9.0	4.4	3.3	6.0	6.1	3.9	4.0	3.5	2.7	4.4
全受験者の0点の割合(%)	0.1	0.5	0.6	1.3	0.9	0.6	0.7	0.8	0.9	0.8

図 3.3　平均得点率の推移

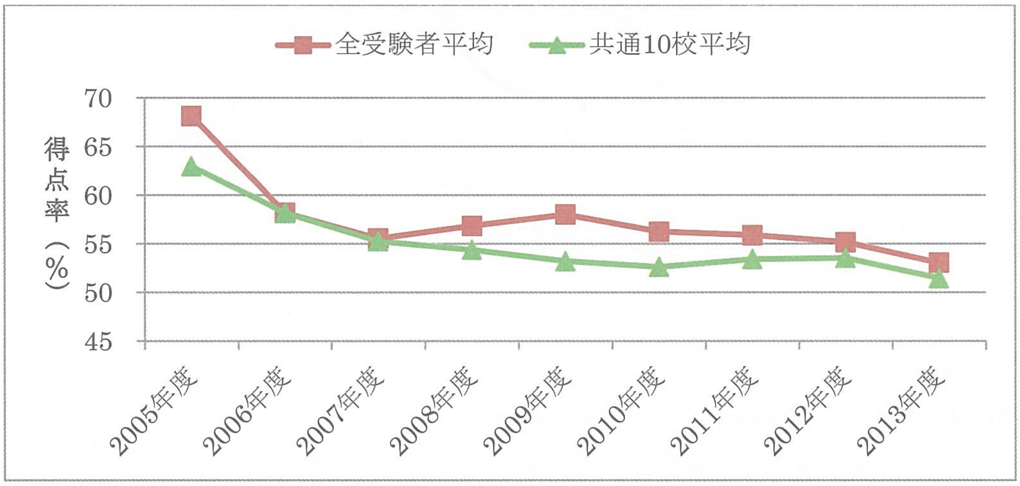
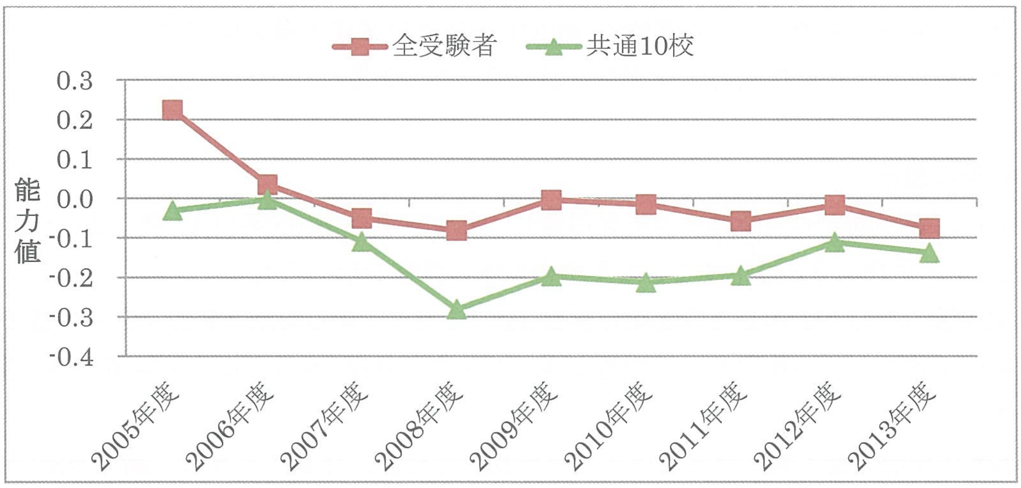


表 3.2 および図 3.4 は IRT を用いた平均能力値の推移である．

表 3.2　平均能力値の推移

	2005年度	2006年度	2007年度	2008年度	2009年度	2010年度	2011年度	2012年度	2013年度	全年度
全受験者数（満点と0点は除く）	2,033	3,199	4,393	4,913	2,698	3,622	3,894	5,648	5,816	36,216
全受験者の平均能力値	0.225	0.036	−0.049	−0.080	−0.003	−0.014	−0.056	−0.016	−0.075	−0.023
共通10校の平均能力値	−0.030	−0.002	−0.108	−0.280	−0.196	−0.212	−0.194	−0.110	−0.136	−0.148

図 3.4　平均能力値の推移



ここで，能力値分布の正規性を検討するために Kolmogorov -Smirnov 検定を行った．その結果，2005～2013 年度の全受験者の能力値が正規分布に従うとする帰無仮説に対する有意確率は$p = 0.099$で，有意水準 5%で帰無仮説は棄却されない（正規分布でないとはいえない）．同様に各年度の受験者の能力値分布の正規性を検定すると，表 3.3 の通りになる．「正規性」欄で「×」は有意水準 5%で能力値分布が正規分布であるという帰無仮説が棄却されることを表し，「○」はこの帰無仮説が棄却されないことを表す．2005 年度と 2010 年度の分布は，正規分布になっているとは言い難く，2005 年度は特に正規分布からの逸脱が著しい．2005 年度はこの数学基礎学力調査の初年度であり，この年度の受験者集団には能力値に偏りがあった可能性も考えられる．

表 3.3　年度別能力値分布の正規性

年度	有意確率(p)	正規性
2005	0.0002	×
2006	0.1658	○
2007	0.0680	○
2008	0.2945	○
2009	0.2973	○
2010	0.0101	×
2011	0.3569	○
2012	0.1101	○
2013	0.2270	○
2005-2013	0.0761	○
2006-2013	0.0540	○

2005 年度を除いて各年度の平均能力値と全体分布の平均能力値の差を分析するため，能力値の分散が異なる可能性も考慮し，Welch の t 検定を行った．能力値が正規分布に従うとき，各年度の能力値の標本平均を$\bar{X}$，全体分布の標本平均を$\bar{Y}$とすると

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{s_X^2/m + s_Y^2/n}}$$

は，次式で近似される自由度νの t 分布に従う．

$$\nu = \frac{(s_X^2/m + s_Y^2/n)^2}{s_X^4/\{m^2(m-1)\} + s_Y^4/\{n^2(n-1)\}}$$

表 3.4 の通り各年度とも$\nu > 2000$であり，tは標準正規分布に従うとみなせる．従ってtの 95%確率区間は$[-1.96, 1.96]$であり，帰無仮説を該当年度の平均能力値と全体分布の平均能力値が等しいとすると，

表 3.4　平均能力値の増減

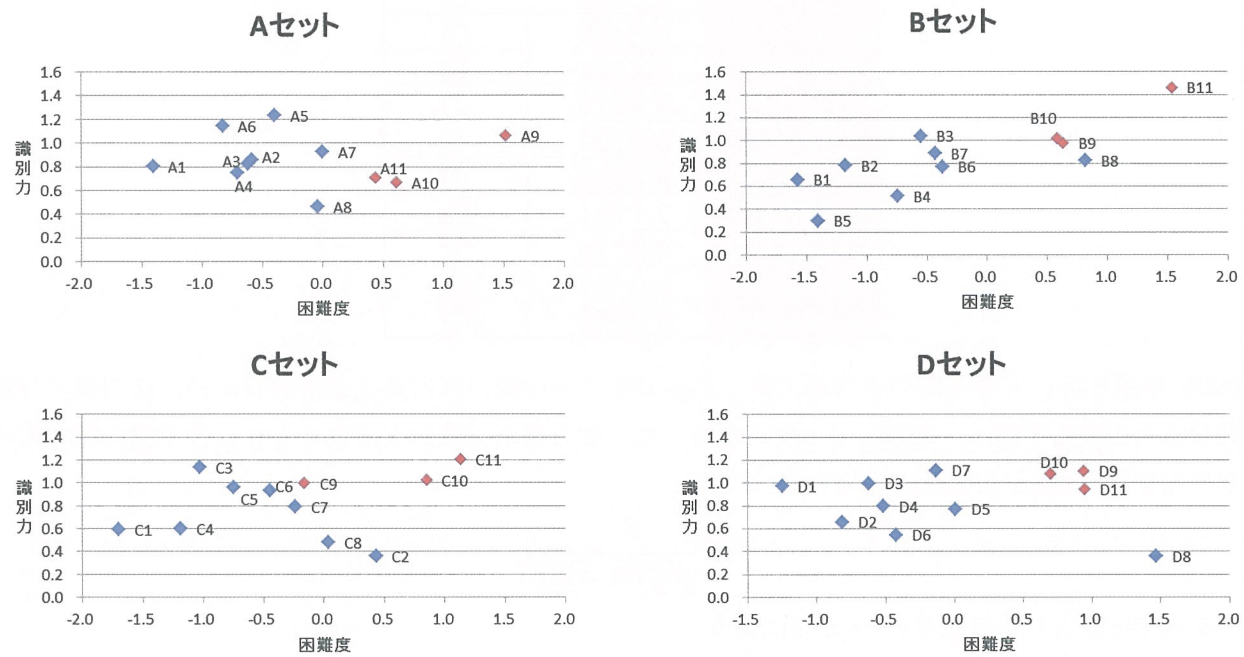
年度	データ数	平均	標準偏差	自由度 v	t	平均比較
2006	3,199	0.036	1.045	3,773	1.88	＝
2007	4,393	−0.049	1.050	5,478	−2.91	小
2008	4,913	−0.080	1.033	6,324	−5.10	小
2009	2,698	−0.003	0.997	3,146	−0.13	＝
2010	3,622	−0.014	0.990	4,449	−0.81	＝
2011	3,894	−0.056	0.991	4,851	−3.34	小
2012	5,648	−0.016	0.972	7,779	−1.13	＝
2013	5,816	−0.075	0.975	8,063	−5.39	小
2006～2013	34,183	−0.038	1.005	−	−	−

$|t| \leq 1.96$の場合，有意確率 5%で帰無仮説は採択される．表 3.4 では，この場合に「平均比較」欄で“＝”と記し，$t > 1.96$の場合は当該年度の平均能力値は全体分布の平均能力値より大きいと考えられ“大”，$t < -1.96$の場合には全体分布の平均能力値より小さいと考えられ“小”と記す．
この結果から，全体分布の平均能力値に比べ，2007 年度，2008 年度，2011 年度および 2013 年度の平均能力値は小さくなっていると考えられる．

３．４　問題セットごとのパラメータ

図 3.5 は問題セットごとの識別力パラメータと困難度パラメータの散布図である．各セットとも困難度においてバランスよく出題されているといえる．赤で示したものは各セット 3 問ずつ出題された記述式の問題で，総じて識別力と困難度の双方が高い．

図 3.5　問題セットごとのパラメータ

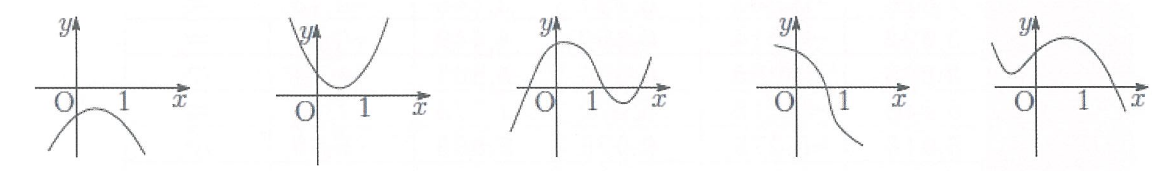


３．５　数値を変えた同種の問題のパラメータ

今年度の出題問題の中に，数値を変えた同種の問題を 2 組出題した．ここで上げたものは A 問題 7 と D 問題 7 であるが，もう 1 組は B 問題 3 と C 問題 6 である．

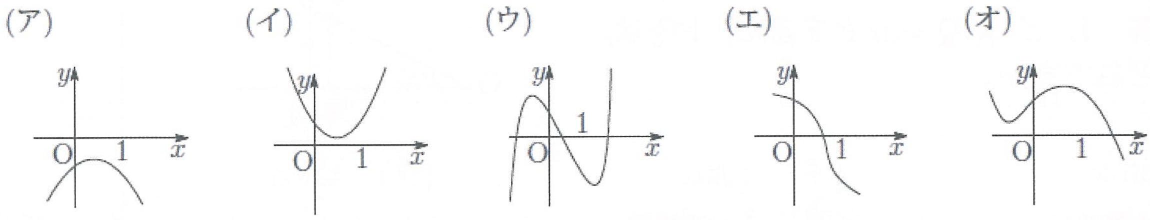
A 問題 7　　　正答率 48.4%　　識別力：0.93　　困難度 0.00

> 関数 $f(x)$ について，「$f'(0) > 0$ ，$f'(1) < 0$ かつ $f''(x)$ は定義域のすべての x に対して負」という条件が与えられているとき，下のグラフの中で，この条件を満たすものはどれですか。
>
> (ア)　(イ)　(ウ)　(エ)　(オ)



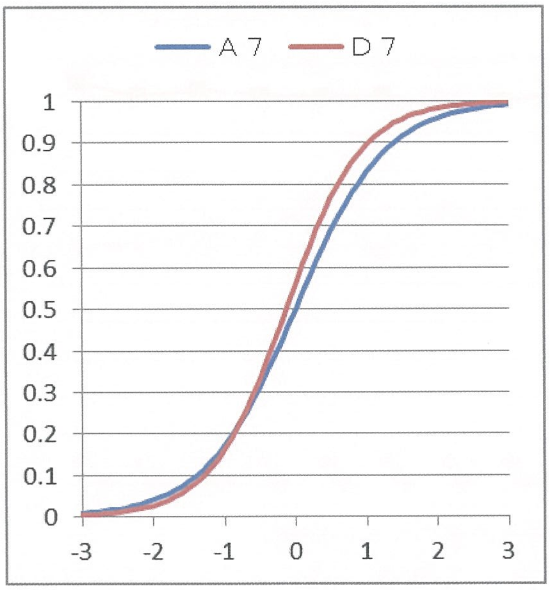
D 問題 7　　　正答率：54.8%　　識別力：1.11　　困難度：−0.13

関数 $f(x)$ について，「$f'(0) > 0$，$f'(1) < 0$ かつ $f''(x)$ は定義域において同一の符号ではない」という条件が与えられているとき，下のグラフの中で，この条件を満たすものはどれですか。



D7 は A7 より正答率が高く，受験者は D7 の方が取り組みやすかったのであろうが，パラメータを見ると D7 の方が識別力が高い．図 3.6 の 2 問の項目特性曲線を見てもほぼ変わらないことから，これらの問題は問題を入れ替えても能力値の測定にほぼ影響がないといえる．

図 3.6　A7 と D7 の項目特性曲線



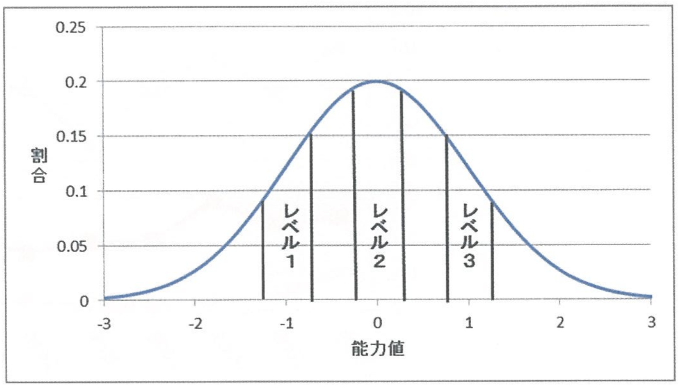
3．6　能力値の階層別正答率の比較

受験者の能力値でレベル 1〜3 の階層に分け，階層別に年度ごとの各問題の正答率の推移を調べた．能力値−1．25〜−0.75 をレベル 1，−0.25〜0.25 をレベル 2，0.75〜1.25 をレベル 3 とした．レベル 1 は下位 20%付近の受験者，レベル 2 は平均的受験者，レベル 3 は上位 20%付近の受験者である．

表 3.5　能力値の階層

階層	能力値
レベル1	−1．25〜−0．75
レベル2	−0．25〜0．25
レベル3	0．75〜1．25

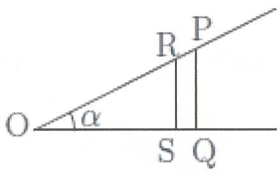
図 3.7　能力値の階層



★平均以上の受験者と下位層の差の大きい問題

C 問題 3　　識別力：1.13　　困難度：－1.02

右の図で，PQ ⊥ OQ および RS ⊥ OQ です。OQ=OR= 1，∠POQ $= \alpha$ とすると，PQ は，つぎのどれですか。

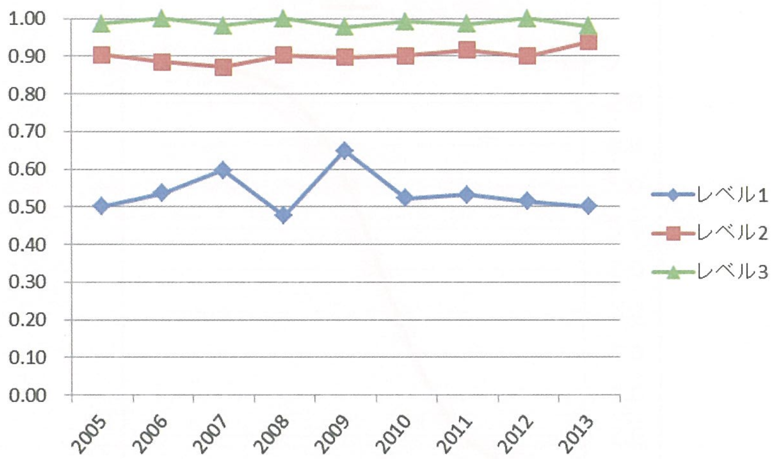


(ア)　$\sin\alpha$　　(イ)　$\cos\alpha$　　(ウ)　$\tan\alpha$

(エ)　$2\sin\alpha$　　(オ)　$1-\cos\alpha$

C 問題 3 は三角比の基本を問う，困難度が－1.02 と比較的易しい問題である．平均以上の受験者は 90%以上の正答率であるが，レベル 1 の受験者は 50%程度で差が大きい．識別力が 1.13 と比較的高いことから能力値－1 付近を境に正答率が跳ね上がる問題である．

図 3.8　C 問題 3 の階層別正答率の比較

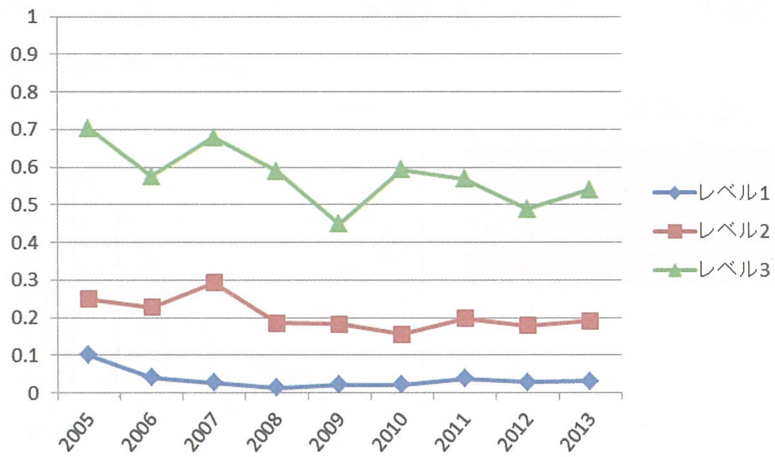


★上位層と平均以下の受験者との差の大きい問題

D 問題 10　　識別力：1.08　　困難度：0.69

座標平面上の 2 円 $(x-2)^2+(y-14)^2=125$ ， $x^2+y^2=25$ の共通な弦と原点との距離を求めなさい。

図 3.9　D 問題 10 の階層別正答率の比較



D 問題 10 は記述式の問題で，困難度が 0.69 と比較的難しい問題である．平均的な受験者の正答率も 20%程度と低く，レベル 1 の受験者はほぼ正解できていない．上位のレベル 3 の受験者でも近年は正答率が 50%程度で，2005 年，2007 年は 70%近い正答率であったことを見ると気になる下がり方である．

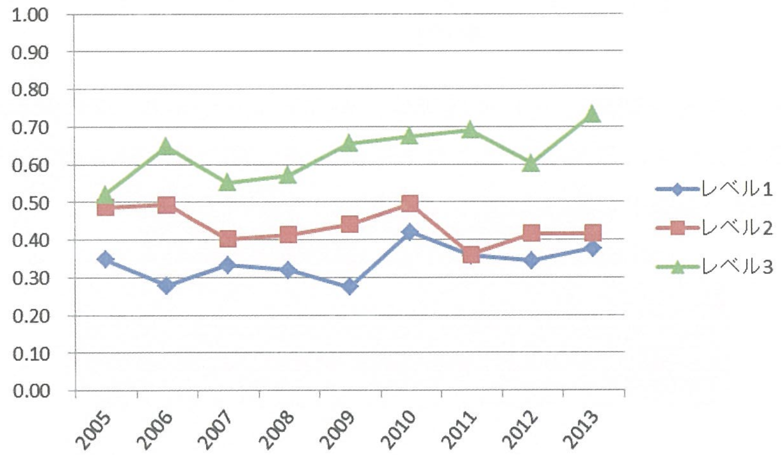
★上位層の受験者正答率が上がった問題

C 問題 8　　識別力：0.48　　困難度：0.03

$$\int_0^1 \frac{12x}{(2x^2+1)^2}\,dx$$ の値はつぎのどれですか。

(ア) -2　(イ) -1　(ウ) 2　(エ) $\log 2$　(オ) $3\log 3$

図 3.10　C 問題 8 の階層別正答率の比較



C 問題 8 は定積分の問題で，困難度が 0.03 と平均的な難易度の問題であるがあるが，識別力は 0.48 と低い．このことは平均的なレベル 2 の受験者とレベル 1 の受験者の正答率が近年変わらなくなってきていることからもうかがえる．

３．７　テスト情報量

IRT を用いると，問題セットごとにどのあたりの能力値の受験者の能力値をより精度よく判定できるかの指標を得ることができる．これをテスト情報量といい，図 3.11 のように表される．

問題セット A は平均よりやや低い能力値の受験者の測定にもっとも適しているといえ，問題セット D は平均よりやや高い能力値の受験者の測定にもっとも適しているといえる．

問題セット C は平均よりやや低い能力値の受験者の測定に適しているといえるが，問題セット A に比べるとやや精度が落ちる．

問題セット B はおよそ能力値が 1.5 以下の受験者の測定においては他セット比べ精度が落ちるが，1.5 以上の受験者の測定にはもっとも適している．

図 3.11　問題セット別テスト情報量

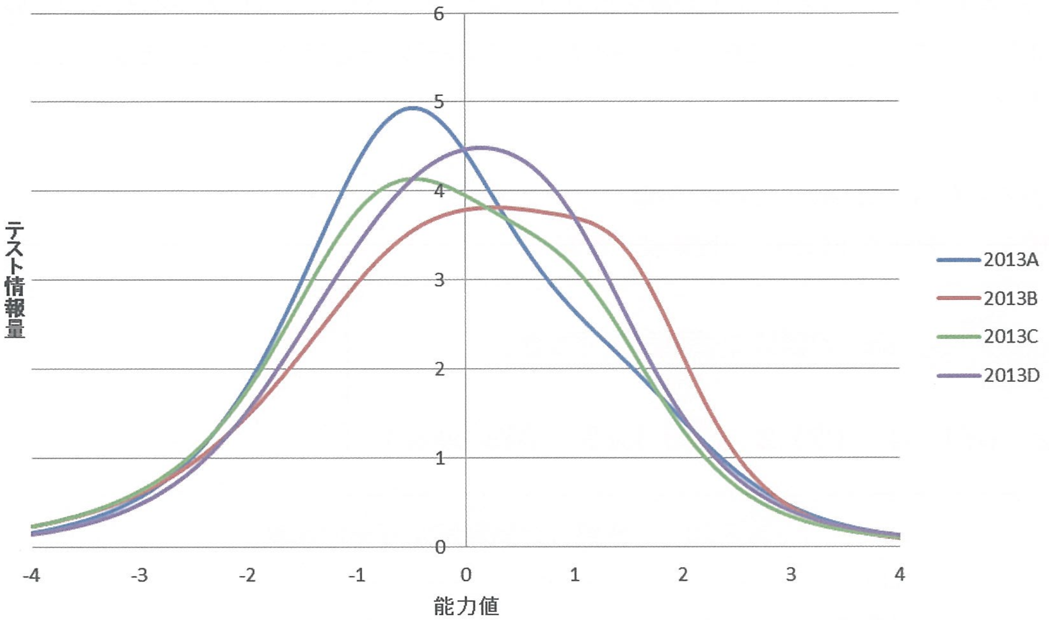


表 3.6 のとおり問題セット A と問題セット B は平均点が同じ 5.9 であり，平均能力値も両者ともにほぼ 0 である．しかし問題セット A は平均能力値よりやや低い能力値の受験者の測定に適し，問題セット B は平均能力値よりやや高い受験者の測定に適しているという違いがある．

表 3.6　問題セット別の平均点と標準偏差

	問題A	問題B	問題C	問題D
平均点	5.9	5.9	6.0	5.6
標準偏差	2.57	2.51	2.49	2.61

３．８　まとめ

IRT を用いて分析を行うと次のような利点があげられる．

Ⅰ．異なる年度間でも同一尺度上での能力値の比較が可能である．

このような学力調査においては年度ごとの推移を見るうえで非常に有用である．

Ⅱ．識別力と困難度の 2 つのパラメータから問題の特徴をとらえることが容易になる．

識別力と困難度の 2 つのパラメータを知ることにより項目特性曲線が描け，ビジュアル的にも問題の特徴を知ることができる．

Ⅲ．より適切な問題セットの作成が可能になる．

IRT により求めた 2 つのパラメータと能力値分布から，平均得点や得点の分散を推定することができる．出題される問題はすでにパラメータの分かっている問題が大半なので，能力値分布を仮定することにより，テスト実施前の問題作成段階においても平均得点や分散を知ることができる．これは適切な問題セットの作成に非常に有用である．