福岡教育大学紀要，第64号，第４分冊，1−12（2015）

項目反応理論による学力調査の再分析

Reanalysis of Achievement Test using Item Response Theory

川　口　俊　明

Toshiaki KAWAGUCHI
学校教育講座

（平成26年９月30日受理）

1．はじめに

本稿の目的は，項目反応理論（Item Response Theory: IRT）による学力調査の再分析を通して，日本の教育社会学における学力研究の課題を改善することにある。

学力研究は，本来，教育学の諸分野が関わる研究課題である。たとえば，国際学力調査として有名な PISA や TIMSS を考えてみよう。これらの調査の Technical Report を読めば，そこには日本で言う教育社会学だけでなく，教育心理学・教育工学・教育方法学・教科教育といった教育学の諸領域の知見が採用されていることがわかる[(1)]。しかし，日本の学力調査において，こうした諸領域の協同はほとんどみられない。全国学力・学習状況調査は典型的な事例であるが，そこには教科教育や教育方法学の知見はともかくとして，教育心理学・教育工学のそれが引用されることはほとんどない。教育社会学者の実施する学力調査でも，事情は似たようなものである。とくに，近年の学力調査の根幹をなす，IRT に無頓着である点が，教育社会学の学力研究に大きな限界をもたらしている。本稿の目的は，これまでの教育社会学者の学力調査の限界を指摘した上で，IRT によるその改善について論じることである。

本稿の概要は，次の通りである。まず，２節で教育社会学者による学力調査の限界と項目反応理論の概要を示す。３節ではデータの概要について述べ，４節では分析結果を述べる。これらを５節でまとめ，今後の課題を示す。

2．教育社会学者による学力調査の限界と IRT

まずは，教育社会学者が実施する学力研究の限界を示すことから始めよう。2000 年の学力低下論争以後，教育社会学者が学力に関して調査を行い，それをもとに発言する機会は確実に増えている。中でも有名なのは，学力低下の実態を検証した苅谷ほか（苅谷ほか 2002），あるいは家庭の年収と学力の関連を示した耳塚（耳塚 2007）であろう。

こうした教育社会学者による学力調査には，いくつか大きな欠点がある。とくに大きな問題は，（1）標本抽出の方法が曖昧で知見の一般化可能性に乏しいこと，（2）IRT など近年のテスト理論の発展をほとんど参照していないこと，の２点であろう。前者の問題については別の機会に論じるとして，今回は後者の問題に絞って議論を行うこととする。

はじめに，古典的テスト理論という概念について説明しよう。いわゆる 100 点満点のテストを想像すれば，それが古典的テスト理論である。古典的テスト理論では，あるテストが 50 の項目（問題）で構成されていた時，うち 40 問を正答すれば 80 点，30 問を正答すれば 60 点といった具合に，総問題数に占める正答数の割合を学力の尺度としている。このままでは異なる集団・異なるテスト間での比較が困難なので，通常，この得点は標準化が行われる。日本の場合，平均 50，標準偏差 10 とした，いわゆる「偏差値」が馴染

み深いだろう。

こうした古典的テスト理論は，とくに専門的な知識が無くても理解できるというメリットをもつ一方，学力研究においては都合の悪い特徴がいくつかある。その一つが，まったく同じテストを実施しない限り，2つの異なる集団が受験したテストの点数を比較することができないというものである。全国学力・学習状況調査は，しばしば年度がちがうと点数の比較ができないと批判されることがある(2)が，これは古典的テスト理論の典型的な欠点である。

この欠点は，学力研究において深刻な問題になる。たとえば，小学校3年次・4年次・5年次の成績を調べ，その伸びを検証したいという状況を考えてみよう。この場合，3年次・4年次・5年次の学力調査の得点が同一尺度上で比較できるという保証がなければ，分析はできない。そして現実問題として，3年次から5年次までの異なる学年が同一のテストを受けることはほとんどありえないから，古典的テスト理論の枠組みに従う限り，こうした問題に挑むのはかなり困難である。

同一集団の成績の伸びという文脈でなくても，日本の子どもたちの学力が下がった／上がったという議論も，古典的テスト理論では処理しづらい問題である。もちろん，まったく同じテストを使えば，古典的テスト理論であっても，異なる年度の得点を比較することは可能である。しかし，現実問題としては学習指導要領の変化などもあるから，毎回，微妙に異なる問題を出題せざるを得ない。これまでの教育社会学者による学力調査では，志水ほか（2014）や苅谷ほか（2002）のように，複数のテストで共通の項目のみを抜き出し，共通項目から計算した正答率によって，学力の変化を推定してきた。ただ，この方法では分析に使用できる項目が，どうしても少なくなってしまう。共通項目以外の項目は，分析されず無駄になるのである。とくに，継続して調査を繰り返すほど，微妙な問題の変更に伴い共通項目が少なくなっていくため，推定の精度は減少していく。

その他，古典的テスト理論では，学力研究の従属変数となる得点に，正規分布が仮定できるとは限らないという問題もある。じっさい日本で実施される学力調査は，受験者の成績の分布を把握するには簡単すぎる傾向がある。典型的には全国学力調査のそれであるが，しばしば天井効果が発生し，成績上位者の数値が正確に推定できていない(3)。

できる子とできない子が2極化しているという「フタコブラクダ化（苅谷ほか 2002)」も，興味深い現象ではある。しかし，従属変数に正規分布を仮定できないという状況は，統計技法の選択や検定の妥当性という観点から言うと，あまり好ましいものではない。これまでの研究では，こうした特徴を無視した分析が行われることがあったが，これは推定値に何らかのバイアスを生む可能性が高い。

ここまでに述べたような問題を総合的に解決することを可能にするのが，心理学領域で近年もっとも研究の活発な分野の一つであるIRTである。 IRTについては，日本でも2000年頃から活発な研究が行われ，豊田（2012）による入門書・事例集をはじめとし，村木（2011）等，良書が少なくない。IRTの詳細については，こうした書籍を参照してほしい。

以下では，ごく簡単にIRTの概要について述べる。IRTでは，横軸に受験者集団には依存しない受験者の能力値θを想定し，縦軸に個々のテスト項目への正答率を配置する。さらに，θと正答確率の関連として，ロジスティック分布の分布関数を利用することで，項目への正答確率Pと能力値θの関連を，次の式のように考える。

$$P_j(\theta_i) = \frac{\exp[a_j - b_j]}{1 + \exp[a_j(\theta_i - b_j)]}$$

この式は，受験者iが項目jに正答する確率を数学モデルで示したものである。この式には，受験者の能力値θと，項目難易度bおよび，項目の識別度aが含まれている。これを，2PLモデルと呼ぶ。その他，項目の識別度を考えない1PLモデルや，当て推量で正答する確率を考慮した3PLモデルなども存在するが，本稿では扱わない。

さて，2PLモデルに則って学力調査を分析した場合，項目の難易度と受験者の能力θは，図1のように考えることができる。図1には，今回の分析に使用する学力調査の問題のうち，かなり易しい問題である問1-4と，やや難しい問3-3の2つを示している。図の横軸は受験者の能力値であり，IRTでは受験者の能力値に標準正規分布を想定するため，-2から2のあいだに全体の95%の受験者が入ることになる。図の縦軸

は，受験者が項目に正答する確率であり，0 から 1 までの値をとる。

図 1 の曲線を ICC（項目特性曲線）と呼ぶ。たとえば問 1-4 の ICC を見てみよう。横軸の能力値 -2（下から 2.5% 程度の能力値）のところを見ると，この受験者は問 1-4 に 50% 程度の確率で正答することがわかる。これは，問 1-4 が非常に易しい（難易度 b が小さい）問題であることを示している。続いて，問 3-3 に移ろう。問 3-3 は問 1-4 に比べると難しい（難易度 b が大きい）問題であり，能力値 -2 での正答率はほぼ 0 である。能力値 0 であれば，問 3-3 の問題の正答率はほぼ 50% になる。さらに能力値が高くなると正答率も上昇し，能力値 2 の受験者（上から 2.5% 程度の能力値）であれば，ほぼ 100% 正答することになる。なお，ここまで横軸の能力値を基準に話を進めたが，項目特性曲線は縦軸の正答確率を基準に解釈してもかまわない。つまり縦軸を先に見て，問 1-4 に 50% の確率で正答する受験者の能力値は，おおよそ -2 といった具合に解釈することも可能である。

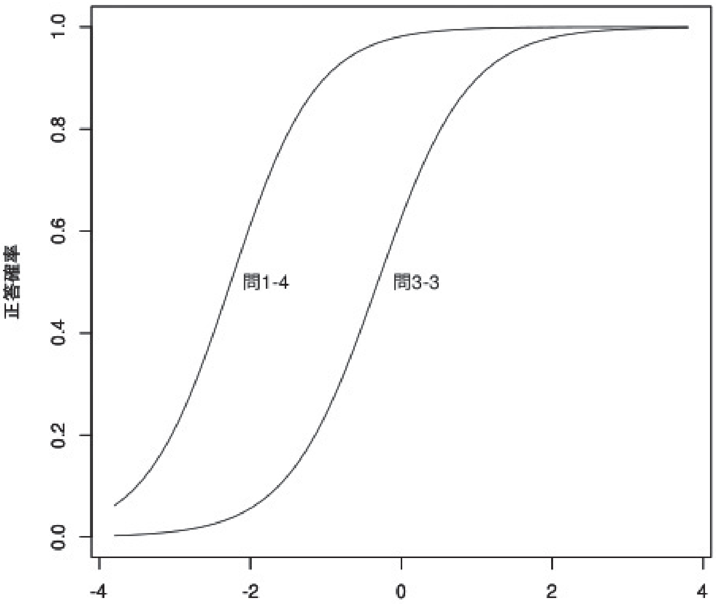

図 1. 項目特性曲線の例

図 1 からは，それぞれの項目が，どのレベルの受験者をもっとも弁別しやすいか把握することもできる。たとえば，問 1-4 は，能力値が -2 付近で急激に曲線が立ち上がり，正答率が上昇している。これは，問 1-4 が能力値 -2 前後の受験者を弁別するのに適切であることを示している。同様に，問 3-3 は能力値 0 前後の受験者を弁別するのに適切であることもわかる。識別度 a は，個々の項目がどの程度ハッキリと受験者を識別する能力を持つかを示した値であり，この値が高い項目ほど，急激に ICC が立ち上がることになる。

このように，IRT の 2PL モデルでは，項目ごとに難易度（b）や受験者の識別しやすさ（a）があると考えたうえで，テストに使用した全項目に対する受験者の反応から，最終的な能力値を推定することになる。この技法のメリットは次のとおりである。先程の例では，問 1-4 における正答率 50% の値は能力値 0 と解釈できた。そして，一度 ICC を推定してしまえば，今後はどのような集団に実施したとしても，問 1-4 における正答率 50% という反応は，能力値 -2 として解釈することができる。つまり，受験者の質によらず，受験者の能力を推定できるようになる。

さらに，異なる複数のテストであっても，一部に共通項目を含んでさえいれば，その共通項目の難易度を知ることで，複数のテストの難易度を等化（equating）することができる。とくに，難易度が大きく異なる共通項目を複数配置していれば，少ない項目であってもテストを十全に等化することが可能である（豊田 2012）。もちろん，それぞれのテストにおける被験者の能力推定は，試験問題のすべてを使用して行われるから，共通項目以外の項目が無駄になることもない。

最後に，IRT では受験者の能力に標準正規分布を仮定しているため，推定された能力値は正規分布に近づく。そのため，学力を従属変数としたときに，各種の統計技法を当てはめることが容易になる。

以上のように，IRT は学力研究において有効な特性をいくつも備えている。心理学領域で発展したとはいえ，容易に分析を実行可能な分析ソフトがいくつも存在するため，他領域の研究者であっても，基本的な分析を行うことは難しくない。本稿では，フリーの統計ソフトである R を使用した推定を行う。

３．データの概要

本稿では，大阪府下で 1989 年，2001 年，2013 年に，小学 5 年生および中学 2 年生を対象に実施された学力調査の結果を再分析し，IRT の有用性について検討する。各年度の調査には，子どもたちの生活習慣について質問した生活実態調査が付随しており，89 年から 13 年までの子どもたちの生活実態の変化が把握できるという点でも貴重なものである。調査の詳しい概要については，志水ほか（2014）を参照してほしいが，89 年・01 年のデータは，苅谷らによる『学力低下の実態（苅谷ほか 2002)』と同一のものであり，日本の教育政策に与えた影響などを考えても，再分析に値する価値を持っていると言える。ただし，対象校は全国から無作為に抽出されたものではないので，知見の一般化可能性には留意する必要がある。今回は，このうち，3 回の調査に全て参加した中学校 11 校のデータを使用して分析を行う。生徒数は，89 年から順に 2222 人，1281 人，1226 人である。分析する項目は，各年度で実施された国語および数学の学力調査の全項目である。

さらに，IRT と正答率のそれぞれを従属変数とした場合の分析結果を比較するため，後の分析で回帰分析を行う。このとき，独立変数として，生活実態調査の設問をいくつか使用する。具体的な項目は，3 回の生活実態調査で共通している，勉強時間（単位 : 時間），通塾しているか否か（通塾している =1，していない =0 のダミー変数），性別（男子 =0，女子 =1 のダミー変数）の 3 つである。それぞれの変数の記述統計量は，表 1 の通りである。

表 1. 記述統計量

	89 年	01 年	13 年
勉強時間	0. 70 (0. 74)	0. 46 (0. 69)	0. 53 (0. 71)
通塾	0. 54 (0. 50)	0. 51 (0. 50)	0. 50 (0. 50)
性別	0. 49 (0. 50)	0. 51 (0. 50)	0. 50 (0. 50)

() 内は標準偏差

４．分析結果

分析結果を示す前に，本稿で使用する IRT の前提を確認しておこう。まず，各項目への受験者の反応は，正答 1，あるいは誤答 0 の 2 値のみとして扱う。受験者の反応を多値に拡張した IRT も存在するが，今回利用した学力調査には，基本的に正答・誤答のみしかデータが存在しないため，こうした拡張を行うことはできない。

また，推定が不安定になることを防ぐため，あまりに正答率の高すぎる／低すぎる問題は分析対象から除外した。具体的には，90% を超える正答率，および 10% を下回る正答率の問題を除外している。本来，こうした項目は，受験者の能力を測定するためには適切でないから，予備調査の段階で分析から取り除くべきものである。注の表 4 および表 5 には，今回の分析対象とした全項目の概要を示しているが，これを見ると，今回使用する学力調査には，こうした不適切な項目がいくつか存在する。こうした項目がテストに含まれてしまっていることも，テスト設計を専門としない教育社会学者による学力調査の欠点と言えるだろう。

さて，IRT を分析に使用する際には，重要な前提が存在する。それが，項目の局所独立の仮定，および項目の一元性の仮定である。前者は，前の問題の正誤が後ろの問題に影響を及ぼさないというものであり，たとえばセンター試験でしばしば見られる，「大問形式（前の問題に正答することが，次の問題に正答する条件となるような問題)」を含めてはいけないというものである。今回分析する範囲には，該当する項目は存在しなかったため，この点は考慮しなくてかまわない。

後者の項目の一元性の仮定は，すべての項目が同一の能力を測定しているというものであり，因子分析等によって，全体が 1 因子構造になっているか否かを確認する必要がある。今回は，テスト全体の正答率との相関係数が 0.3 を下回っているものを除去した上でスクリープロットを描くことで，全体が一因子構造であることを確認した。なお，これらの判断は，豊田（2012）に倣っている。

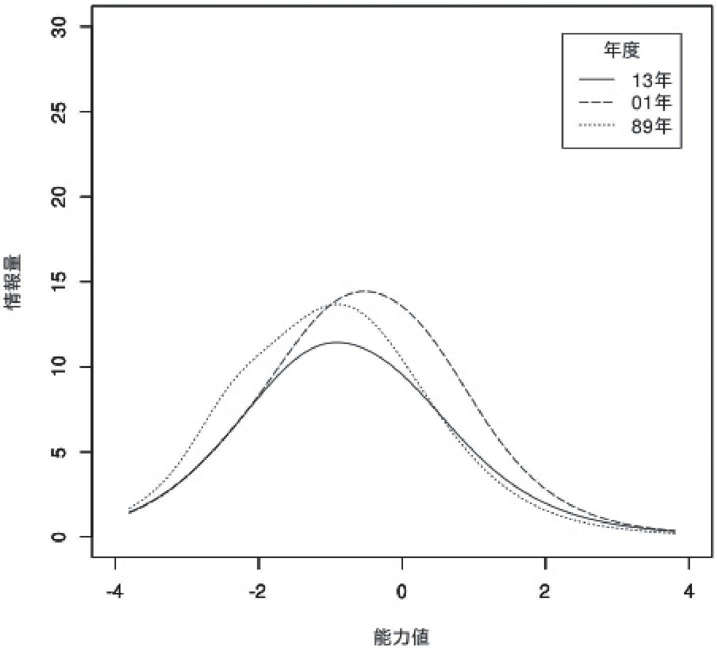

図 2.　テスト情報関数（国語）

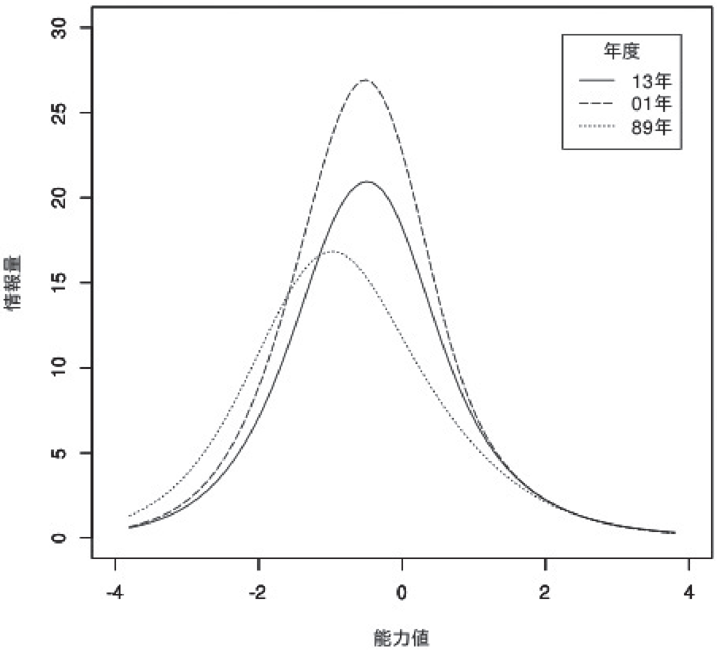

図 3.　テスト情報関数（数学）

IRT の推定方法はいくつか考案されているが，今回は最尤法（周辺最尤推定法）を用い，成績の等化には Stoking ＆ Lord 法を使用した。推定には，R の ltm および plink を利用している [(4)]。最終的に，3 回の学力調査のそれぞれで，個々の項目の難易度が -4 から 4 のあいだにあること，および識別度が 0.2 を越えて

いたことから，今回の推定には，一定の妥当性があると考えられる(5)。

さて，IRT では，テスト情報関数と呼ばれる曲線を描くことができる。これは，当該テストが，どの程度のレベルの受験者をもっとも弁別できるかを視覚的に示した図と言うことができる。はじめに，この点について確認しておこう。

図 2 や図 3 を見ると，国語・算数ともに，どの年度のテストも，能力 -1 前後の位置に最大値が来ていることがわかる。これが意味するのは，今回の調査で使用したテストは，かなり簡単だということである。つまり，これらのテストでは，能力の高い受験の能力値は正確に測定できていない（＝標準誤差が大きい）という点には留意が必要となる。

続いて，IRT と正答率を用いた場合，それぞれの教科の得点がどのように推定されるか見ていくことにしよう。表 2 が，IRT と正答率を用いた場合の得点である。なお，この値は 1989 年の全受験生の平均を 50，標準偏差を 10 になるように変換したものである。正答率については，89 年・01 年・13 年で共通する問題の正答率を表示している。

表 2. 記述統計量

	国語（IRT）	国語（正答率）	数学（IRT）	数学（正答率）
89 年	50.2(10.1)	50.2(10.0)	50.4(10.0)	50.4(9.9)
01 年	47.3(10.0)	47.3(10.6)	46.9(11.2)	47.6(11.6)
13 年	50.3(9.0)	49.9(9.4)	46.9(10.0)	47.7(10.5)

() 内は標準偏差

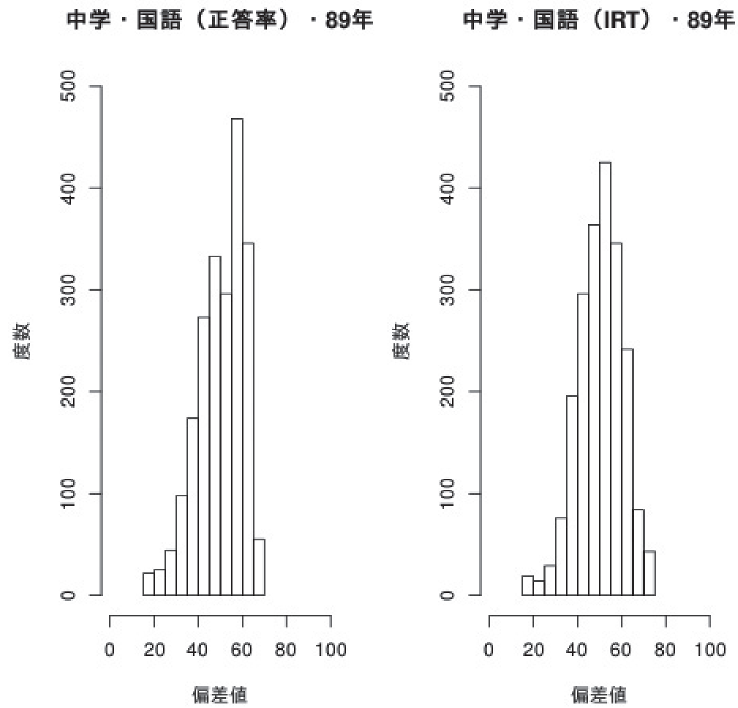

図 4. ヒストグラム（89 年）

表 2 を見てわかることは，平均点については IRT を使用しても正答率を使用しても，それほど大きく変化しないという点である。一方，標準偏差については増減があり，年度によって大きく推定されたり小さく推定されたりしている。これは，IRT が受験者の能力が正規分布することを仮定しているからである。図 4

から図 6 を見るとわかるように，正答率を使用した場合に比べて，IRT の方が正規分布に近く，天井効果が軽減されている。

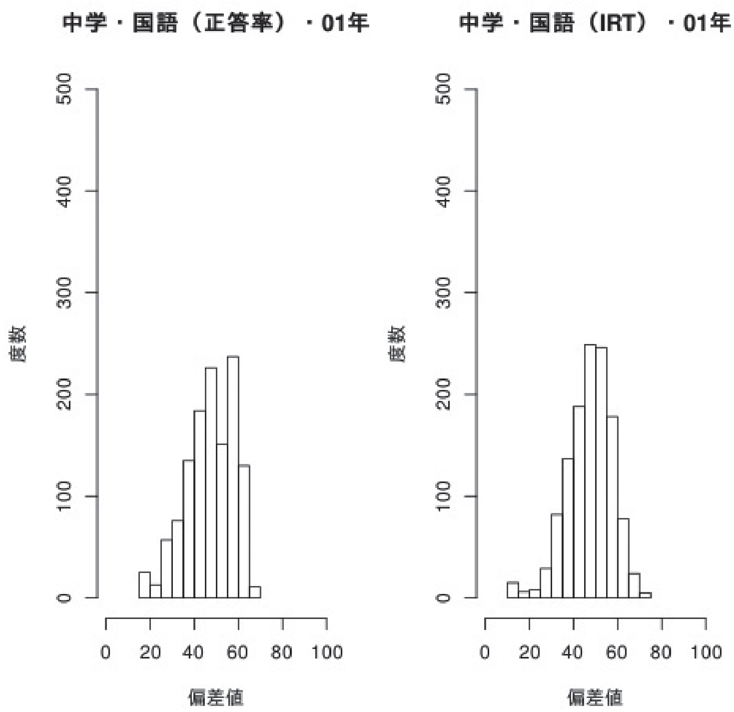

図 5. ヒストグラム（01 年）

それでは，IRT を使用した場合，推定値はどのように変動するのだろうか。以下では，回帰分析を例に取り，推定値の変化についてみていくことにしよう。なお，回帰分析の推定には OLS（最小二乗法）を使用し，欠損値はリストワイズ法により削除している。

表 3 が国語，表 4 が数学の分析結果である。IRT と正答率による推定値を比較したときにわかることは，IRT を使用した場合，全体的に推定値が大きくなる傾向があるということである。また，R^2 値も上昇し，モデルの精度が向上することもわかる。

とくに，通塾や性別といった変数の係数が大きくなっていることは興味深い。これは，IRT を利用することで天井効果の影響が抑制されることが大きいと思われる。たとえば，一般的に通塾している子どもの方が成績が高く，満点を取りやすいから，天井効果が発生する可能性が高い。そのため，正答率による分析は，通塾している子どもと通塾していない子どものあいだにある成績の差を，やや過小評価することになる。IRT を使用することで，こうした過小評価が補正されたのではないだろうか。教育社会学の分析では，通塾しているか否かといった社会的属性による成績の差，いわゆる学力格差を分析対象とすることが多い。IRT を使用することで，より学力格差を捉えやすくなると考えることができるだろう。

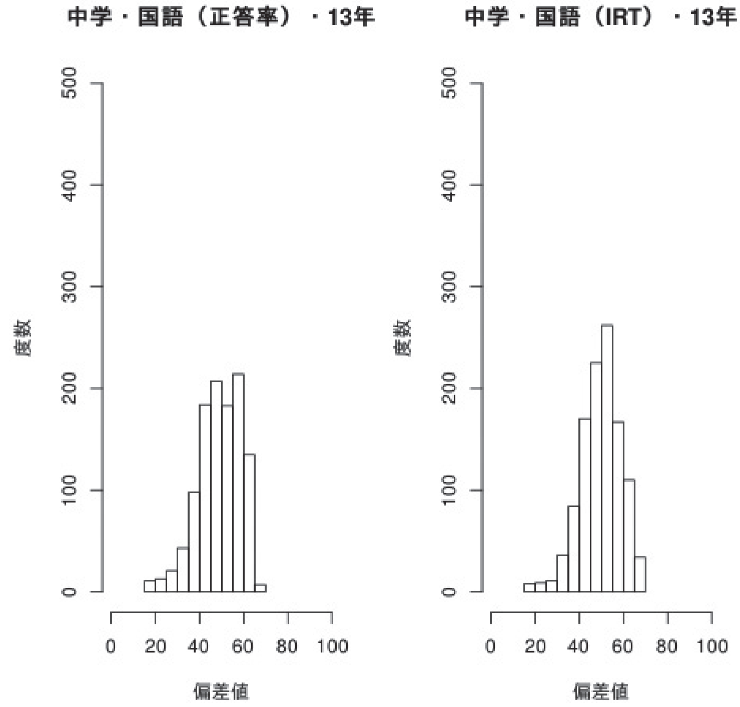

図 6. ヒストグラム（13 年）

表 3. 回帰分析の推定値（国語）

	89/正答率	89/IRT	01/正答率	01/IRT	13/正答率	13/IRT
切片	45.8(0.4)	45.5(0.4)	43.4(0.5)	43.2(0.5)	47.5(0.5)	48.0(0.5)
勉強時間	2.4(0.3)	2.6(0.3)	0.2(0.4)	0.3(0.4)	1.9(0.4)	1.9(0.4)
通塾	3.1(0.4)	3.2(0.4)	4.7(0.6)	4.8(0.6)	1.4(0.6)	1.5(0.6)
性別	2.4(0.4)	2.7(0.4)	3.5(0.6)	3.5(0.6)	0.0(0.0)	0.0(0.0)
R^2	7.4	8.5	7.5	8.6	3.1	3.7

()内は標準誤差

表 4. 回帰分析の推定値（数学）

	89/正答率	89/IRT	01/正答率	01/IRT	13/正答率	13/IRT
切片	45.4(0.4)	45.2(0.4)	42.5(0.6)	42.0(0.6)	42.9(0.5)	42.6(0.5)
勉強時間	2.7(0.3)	2.8(0.3)	1.5(0.4)	1.3(0.4)	2.9(0.5)	2.6(0.4)
通塾	5.5(0.4)	5.8(0.4)	8.7(0.6)	8.8(0.6)	5.8(0.6)	5.6(0.6)
性別	0.5(0.4)	0.2(0.4)	0.5(0.6)	0.4(0.6)	0.0(0.0)	0.0(0.0)
R2	12.3	14.1	15.8	16.8	14.0	14.1

()内は標準誤差

5. まとめと今後の課題

本稿では，教育社会学における学力研究の問題点として，古典的テスト理論に基づく学力調査の欠点について示した。具体的には，異なる集団に異なるテストを実施した場合，同一尺度上で成績を評価できないと

いう点が，学力研究にとって非常に大きな壁となるということである。本稿では，この壁を越えるための手法として，心理学領域で発展しているIRTを導入し，苅谷ほか（2002）・志水ほか（2014）が分析した学力調査を再分析することを試みた。結果として，(1) これらの学力調査が易しい問題から構成された難易度の低いテストであり，能力値の高い受験者については標準誤差が大きいこと，(2) IRTを使用したとしても，得点の平均値は大きく変わらないこと，(3) IRTを使用することで，R^2値をはじめとして全体の推定値が大きくなること，がわかった。これらを踏まえると，先行研究の知見は学力低下については妥当だが，学力格差についてはやや格差を過小評価している可能性を指摘できる。学力格差は，教育社会学の学力研究において，中心となるテーマの一つである。本稿の分析を鑑みれば，IRTを使用した結果の再分析や，今後の学力調査におけるIRTの採用が，今後の教育社会学の学力研究にとって重要なものとなる。

日本の学力調査は，今のところ特定の領域の研究者が実施するものになりがちであるが，本稿の知見は，複数領域の知見を総合した学力調査の重要性も示している。もちろん個人が複数の領域に精通することは困難だろうから，教育学諸領域の連携と，複数の研究者が関わる学力調査の実施が求められていると言えるだろう。

＜注＞

⑴ PISAについてはOECD（2012），TIMSSについてはOlson et al（2008）を参照のこと。
⑵たとえば，川口（2011）を参照のこと。
⑶全国学力・学習状況調査の結果については，国立教育政策研究所のHPからダウンロード可能である。
⑷ Rのバージョンは3.1.1である。なお，plinkパッケージは2014年9月現在，cranから削除されている。
⑸学力調査の各項目の概要およびIRTによる推定値は，表5および表6の通りである。なお，「正答率90%超」は正答率が90%を超えた問題であり，「全体との相関低」は全体との正答率との相関係数が0.3を下回る項目である。これらの項目は推定から除外している。また，「－」となっているのは，指導要領の変更などの理由で項目が変更された箇所である。

　なお，2PLだけでなく3PLによる推定も行ったが，推定結果はほとんど変わらなかったため，2PLによる推定値を掲載する。

表 5. 問題概要、および 2PL における推定値（国語）

問題の概要	問題番号	13 年		01 年		89 年	
		困難度	識別力	困難度	識別力	困難度	識別力
漢字の読書	問 1_1	−0.95	1.62	−0.63	1.63	−0.81	2.16
漢字の読書	問 1_2	—		−2.51	1.93	−1.93	1.78
漢字の読書	問 1_3	正答率 90 超		正答率 90 超		正答率 90 超	
漢字の読書	問 1_4	−2.26	1.77	−1.31	1.85	−0.90	1.82
漢字の読書	問 1_5	—		0.93	1.37	—	
漢字の読書	問 1_6	−0.87	1.34	−0.51	1.35	—	
漢字の読書	問 2_1	0.16	1.09	0.38	1.47	−0.24	1.48
漢字の読書	問 2_2	−2.03	1.20	−1.40	1.44	−1.42	1.65
漢字の読書	問 2_3	—		0.26	1.53	0.30	1.20
漢字の読書	問 2_4	—		−0.56	1.98	−0.71	1.80
漢字の読書	問 2_5	0.15	1.34	0.46	1.68	—	
漢字の読書	問 2_6	−1.22	0.96	0.01	1.00	—	
漢字の読書	問 3_1	—		−0.67	1.93	−0.85	2.05
漢字の読書	問 3_2	—		全体との相関低		全体との相関低	
漢字の読書	問 3_3	−0.31	1.67	0.40	2.06	0.52	1.73
漢字の読書	問 3_4	−0.01	1.74	−0.03	2.04	−0.27	2.15
漢字の読書	問 3_5	0.24	1.48	0.55	1.75	0.04	1.84
漢字の読書	問 3_6	0.15	1.18	0.37	1.31	−0.13	1.05
文法	問 4_1	—		正答率 90%超		正答率 90%超	
文法	問 4_2	—		正答率 90%超		正答率 90%超	
文法	問 4_3	—		正答率 90%超		正答率 90%超	
文法	問 4_4	—		正答率 90%超		正答率 90%超	
文法	問 4_5	—		正答率 90%超		正答率 90%超	
文法	問 4_6	—		全体との相関低		全体との相関低	
文法	問 5_1	全体との相関低		全体との相関低		全体との相関低	
文法	問 5_2	−1.48	1.13	−1.28	1.27	−1.31	1.20
文法	問 5_3	−2.06	1.41	−1.82	1.17	−1.48	1.08
文法	問 6_1	−1.51	1.51	−1.28	1.65	−1.65	1.76
文法	問 6_2	−0.51	1.16	−0.29	1.53	−0.74	1.35
文法	問 6_3	—		−0.93	1.87	−1.42	1.72
文法	問 6_4	−0.02	0.77	−0.19	0.94	−1.03	0.71
文法	問 6_5	−1.08	1.24	−1.11	1.54	−1.71	1.67
文法	問 6_6	—		−1.09	1.15	−1.79	1.02
文法	問 7_1	−2.37	1.32	−2.21	1.08	−2.58	1.56
文法	問 7_2	—		−2.23	1.01	−2.44	2.85
文章構成	問 8	0.21	1.20	0.02	1.32	0.06	1.31
長文読解	問 9_1	−1.06	1.06	−0.88	1.31	−1.19	1.36
長文読解	問 9_2	−1.16	1.70	−1.25	1.52	−1.34	1.56
長文読解	問 9_3	−1.47	1.81	−1.66	1.61	−2.09	1.51
長文読解	問 9_4	0.01	0.57	−0.04	0.57	−0.31	0.72
長文読解	問 9_5	−0.68	0.84	−0.81	0.80	−1.36	0.60
長文読解	問 9_6	0.96	0.94	0.92	0.98	0.51	1.11
長文読解	問 9_7	0.49	1.19	0.64	1.30	0.12	1.23
長文読解	問 9_8	−0.03	1.88	0.13	1.61	−0.47	1.77
長文読解	問 9_9	0.12	0.82	0.30	0.62	−0.11	0.58
長文読解	問 9_10	−1.13	1.78	−1.08	1.70	−1.46	1.46
B 問題	問 10_1	−1.48	1.85	—		—	
B 問題	問 10_2	−0.06	1.24	—		—	
B 問題	問 10_3	−0.87	1.87	—		—	

表 6. 問題概要、および 2PL における推定値（数学）

問題の概要	問題番号	13 年		01 年		89 年	
		困難度	識別力	困難度	識別力	困難度	識別力
数	問 1_1	—		—		0.61	0.91
数	問 1_2	—		—		−1.28	1.40
数	問 1_3	—		—		−0.41	1.33
数	問 1_4	—		—		0.52	1.42
数	問 1_5	—		—		0.19	1.00
数	問 1_6	—		—		全体との相関低	
数	問 2	0.13	0.81	−0.29	0.95	−0.92	1.22
計算	問 3_1	−1.89	1.52	−1.75	1.60	−1.95	1.40
計算	問 3_2	—		−1.02	1.83	−1.29	1.59
計算	問 3_3	−1.71	1.86	−1.58	2.03	−1.81	1.74
計算	問 3_4	—		−0.82	1.44	−1.25	1.28
計算	問 3_5	正答率 90 超		正答率 90 超		正答率 90 超	
計算	問 3_6	—		−0.52	1.82	−0.91	1.56
計算	問 3_7	−0.92	1.39	−0.75	1.75	−0.89	1.29
計算	問 3_8	−0.36	1.90	−0.29	1.85	−0.60	1.46
計算	問 3_9	—		−0.25	1.12	−1.21	1.33
計算	問 4_1	—		−1.31	1.45	−1.42	1.40
計算	問 4_2	−1.06	1.43	−1.47	1.78	−1.72	1.56
計算	問 4_3	−0.85	1.59	−0.93	1.33	−1.22	1.45
計算	問 4_4	—		−0.71	1.95	−1.06	1.53
計算	問 4_5	1.32	0.83	0.86	1.26	0.75	1.29
計算	問 5_1	−0.46	1.69	−0.48	1.77	−0.79	1.49
計算	問 5_2	−0.10	1.04	−0.20	1.16	−0.34	0.88
計算	問 6_1	0.58	1.12	0.46	1.20	−0.39	1.20
計算	問 6_2	−0.19	0.76	−0.23	0.90	−0.47	0.80
計算	問 7_1	−1.15	2.24	−1.07	2.03	−1.55	1.74
計算	問 7_2	−0.78	2.14	−0.82	2.39	−1.43	1.78
計算	問 7_3	—		−0.43	2.23	−0.85	1.96
計算	問 7_4	−0.25	1.92	−0.27	2.18	−0.66	1.78
計算	問 7_5	−0.63	2.10	−0.54	2.55	−1.05	2.09
計算	問 7_6	−0.08	1.74	−0.07	2.31	−0.62	1.60
方程式	問 8	−0.67	1.70	−0.57	1.83	−0.73	1.76
方程式	問 9	0.10	1.23	0.19	1.18	−0.13	1.27
方程式	問 10	−1.22	1.79	−1.22	1.57	−1.37	1.52
方程式	問 11_1	−0.31	1.64	−0.37	1.60	−0.44	1.25
方程式	問 11_2	0.56	1.41	0.87	1.35	0.54	1.23
図形	問 12	−0.33	1.33	−0.63	1.43	−0.78	1.63
図形	問 13	0.36	1.50	0.01	1.16	−0.29	1.15
数	問 14	0.15	0.90	0.46	0.95	—	
図形	問 15	0.39	1.09	0.59	1.24	—	
計算	問 16_1	−0.75	2.02	−0.65	2.01	—	
計算	問 16_2	−0.52	2.10	−0.41	2.09	—	
計算	問 16_3	−0.91	2.82	−1.01	2.72	—	
計算	問 16_4	0.19	2.78	0.21	3.20	—	
計算	問 16_5	−0.23	3.13	−0.33	3.13	—	
計算	問 16_6	−0.50	2.10	−1.58	2.24	—	
B 問題	問 17	−0.80	1.35	—		—	
B 問題	問 18	0.59	1.56	—		—	
B 問題	問 19	全体との相関低		全体との相関低		—	
図形	問 20	—		—		−0.17	0.99
図形	問 21	—		—		0.13	1.09

<参考文献>

苅谷剛彦・清水睦美・志水宏吉・諸田裕子，2002,『調査報告「学力低下」の実態』岩波ブックレット.
川口俊明，2011,「日本の学力研究の現状と課題」『日本動労研究雑誌』614: 6–15.
耳塚寛明，2007,「小学校学力格差に挑む」『教育社会学研究』80: 23-39.
村木英治，2011,『項目反応理論』朝倉書店。
OECD, 2012, *PISA 2009 Technical Report*, OECD Publishing.
Olson, J., M. Martin, & I. Mulliseds., 2008, *TIMSS 2007 Technical Report*, TIMSS & PIRLS International Study Center.
志水宏吉・伊佐夏実・知念渉・芝野淳一，2014,『「学力格差」の実態』岩波ブックレット.
豊田秀樹，2012,『項目反応理論【入門編】（第 2 版）』朝倉書店 .