福岡教育大学紀要，第66号，第４分冊，1−13（2017）

自治体学力調査の再分析

Reanalysis of Local Student Achievement Test

川　口　俊　明
Toshiaki KAWAGUCHI
（学校教育講座）

松　尾　　　剛
Go MATSUO
（教育心理学講座）

礒　部　年　晃
Toshiaki ISOBE
（教育総合研究所）

樋　口　裕　介
Yusuke HIGUCHI
（学校教育講座）

（平成28年9月30日受理）

1. はじめに―何が問題なのか―

2007年に全国学力・学習状況調査が実施されるようになって以降，学力向上を目指して，独自に学力調査を実施する自治体が増えている。しかし，こうした自治体が実施する学力調査（以下，自治体学力調査と呼ぶ）は，出題されている問題の質や調査設計それ自体をはじめ，「学力調査の質」が十分とは言えないものが多い。本稿は，A市で実施されている自治体学力調査の再分析を通して，自治体学力調査の課題と，その改善策について検討するものである。

はじめに，なぜ「学力調査の質」を問う必要があるのか述べることにしよう。2000年以降，主として教育社会学領域の研究者を中心に，日本の学力格差の実態把握や，学力に影響を与える要因分析を行う研究が進められるようになった。とくに有名なものとしては，学力格差の実態を示した苅谷・志水（苅谷・志水編2004），学力格差を縮小する「学校教育の効果」について論じた志水（志水編2009）などが挙げられる。近年では，PISAやTIMSSといった国際学力調査を利用した研究も増えつつあり，学力調査に基づく研究がほとんど存在しなかった2000年以前とは，大きく状況が変わってきている。こうした研究は，教育政策にも影響を与え，「エビデンスベースド」が重視される昨今の風潮とも相まって，学力調査に基づいた学力政策が強調されるようになってきている（中室2015）。

一方で，考えなければならない問題がある。それは，これまで実施されてきた教育社会学の学力研究のほとんどは，学力格差の実態や学力格差を改善することに主眼があり，そもそも学力調査で測定されている学力とは何なのか，あるいは，測定したい「学力」を適切に計測できているのかといった「学力調査の質」の問題を，あやふやなままにしてきたという点である。たとえば苅谷・志水は，学力とは「ペーパーテストで測定した学業達成」であると定義することによって，「学力とは何か」「学力調査はどうあるべきか」といった問題を問うことを回避した（苅谷・志水編2004，pp.4-7)。また，志水のように「学校教育の効果」を扱う研究では，学力テストの点数は高い方が好ましいということが無条件に前提とされており，「学力調査の質」という問題は，そもそも検討されていない。

こうした態度は，もともとの学力調査が十分に質の高いものであれば，それほど問題にはならないかもしれない。たとえば，PISAやTIMSSといった国際学力調査であれば，数百ページに及ぶTechnical Reportが公開されているため，それが何を測っているのか，あるいは適切に測れているのかといった疑問につい

て，データをもとに検証することが可能である（OECD 2014，Olson. et al, 2008)。

しかし，日本で実施されている学力調査は，そもそもそれが何を測っているのか，あるいは適切に測れているのかといった点を十分に検討していない調査がほとんどである。典型的な例としては，全国学力・学習状況調査を挙げることができるだろう。国立教育政策研究所のホームページには，全国学力・学習状況調査の質問紙や，記述統計量，あるいは得点の高低に関わる分析こそ掲載されているものの，そもそもこの調査が何を測定しているのか，そしてそれは適切に測定できたのかに関する記述は存在していないのである。この点は，教育社会学者が独自に実施してきた学力調査も同様である。ほとんどの研究は，それが何を測定しているのか，あるいは適切に測定できたのかといった，「学力調査の質」自体には踏み込まず，測定された成績の高低について論じるだけにとどまっている（苅谷・志水編 2004，志水編 2009)。

その意味では，教育社会学を中心とした現在の日本の学力研究は，そもそも何を学力と定義するのか，そしてそれが適切に測れているのかという，根本的な問題を無視したまま進められてきたとさえ言えるのである。学力格差の実態を明らかにすることはもちろん重要な研究課題である。また，「学力格差」の縮小が政策的・実践的に重要であることも論を待たない。しかし一方で，そもそも問題としている「学力」が何なのかという問題を軽視することは，学力研究の適切な在り方とは思えない。

さらに，近年の学力研究の技術的進捗が，こうした教育社会学の学力研究の在り方に，変化を強いるようになってきている。とくに重要な変化が，厳密な因果関係を推定しようとする因果推論に関する研究の隆盛である。学力調査が繰り返され，データが蓄積されていくに従って，個々の児童生徒の学力が何の要因によって，どの程度変化するのかという，学力に影響を与える要因の推定を目指す研究が増加していくのは自然なことである[(1)]。ただし，こうした研究を実施するには，複数の時点で測定された成績が，同一尺度上に存在する（＝成績の変化が測定できる）という前提が，絶対に必要である。残念ながら，「学力調査の質」に踏み込まずに積み重ねられてきた，これまでの教育社会学の学力研究は，こうした状況に対処できない。

もちろん，ここまでに指摘したような問題は，教育社会学の学力研究にのみ当てはまる問題ではない。すでに触れたように，全国学力・学習状況調査を始めとし，日本で実施されている学力調査は，ほぼすべてが「学力調査の質」の問題を無視しているように見える。このような状況では，まともな学力研究・学力政策など実施できるはずがない。一刻も早く，「学力調査の質」を改善し，より適切な学力研究・学力政策を進めていく必要がある。

以上が本稿の持つ問題意識である。以下では，ある自治体の学力調査を事例にとり，その課題と改善の方向性を検討する。分析対象とするのは，A 市で行われている学力実態調査である。この調査は，A 市教育委員会が，数年前から市内の全公立小中学校の児童生徒を対象に実施しているもので，国語，算数・数学など複数の教科に関するテストが行われてきた。ただ，その結果については，テストの平均点や，各学校のおおよその位置（平均より上など）が示されるだけで，テストの質に関する分析は行われていない。

なお，A 市が保有する学力データは数年分，複数教科に渡る膨大なものであり，これらをすべて分析対象にするのは無理がある。そこで今回は，後々「学力の変化」を分析したいという思惑もあって，分析対象を 2011 年度の小学 4 年生・2013 年度の小学 6 年生・2014 年度の中学 1 年生に限り，さらに教科を算数・数学に限定することにした。分析対象をこのように限定することで，同一の児童生徒が，小学 4 年生から中学 1 年生まで，どのように成績が変化するのか議論することができる。また，算数・数学に限定したのは，低学年から高学年まで，学習内容が連続していることもあって，もっとも「成績の変化」を議論しやすい教科だと判断したためである。

分析には，個々の児童生徒のテスト項目への回答を正答＝ 1，誤答＝ 0 の 2 値情報に変換し，さらに全体から 1000 人を無作為抽出したデータセットを用いた。また，データセットから個人情報を類推できないようにするため，児童生徒の性別その他の情報を含め，分析に直接関わりのない情報は，データセット作成時にすべて削除している。

以下，本稿の構成を述べる。まず 2 節では，A 市の学力調査が，どのような能力を測定しているのか，因子分析と呼ばれる技法を用いて検証する。次に 3 節では，項目分析と呼ばれる技法により，各項目の特徴について検討する。最後に 4 節では，これらの情報をまとめ，自治体学力調査の現状と課題，および改善の方向性について議論する。

2. 自治体学力調査は何を測定しているのか

2.1. 因子分析を用いたテストが測定している概念の検討

学力調査の過程においては，実施前の問題作成や調査対象の抽出方法等に配慮を行うことも重要だが，実施後に調査結果について十分な分析を行うことも同様に重要である。その内容としては，調査対象者の学力実態を把握するための分析だけでなく，実施した調査（使用したテスト）の質についての評価を行うための分析も含まれる。このような分析を行うことで，次回以降の調査を実施する際に必要となる多様な情報を引き出すことが可能になる。本節では特に，調査の質についての評価を行うための分析という観点から論じていく。

学力調査において測定の対象としている「学力」とは，直接観察したり，触れたりすることができない構成概念である。そのため，学力を検討するためには，テストという道具を用いて学力を客観的な数値に置き換える作業，すなわち測定が必要となる。より良い測定のためには，検討の目的となる「学力」を正しく捉えることができるようなテストを作成することが不可欠である。

テストが測定したい構成概念を正しく測定できているという性質は「妥当性」と呼ばれる。テストの妥当性を高めるための一般的な手続きとしては，最初に調査によって測定しようとしている構成概念がどのようなものであるかを理論的に明確にしなければならない。自治体学力調査の文脈で述べるならば，測定しようとしている「学力」とはどのような能力であるのか，国語の学力とはどのようなものをさすのか，算数の学力とはどのようなものをさすのか，それらは単一の能力なのか，それとも複数の独立した下位能力（基礎的・基本的な知識の習得，知識を活用する力，など・・・）によって構成されているのか，といった事を理論的に明確にしておかなければならない。その上で，そのような「学力」であれば，どのような問題とどのように関連するか，テストに含まれている問題によって，構成概念のすべての側面（下位能力）を網羅的に測定可能か，といった点などを検討しながらテストが作成されることになる。したがって，そもそも何を測定しようとしているのか，という点を十分に検討しないままテストを作成したとすれば，調査の結果によって何らかの点数が得られたとしても，その点数が何を意味するか，という点については言及することができず，結局はその点数から調査対象者の「学力」について有益な情報を得る事はできないということになる。

上記はテストの妥当性について，事前の手続きの中で配慮すべき点である。テストの実施後においては，例えば因子分析といった方法によって，テストの結果に基づいて測定されている構成概念についての示唆を得ることができる。以下では自治体学力調査について実際に因子分析を実施し，その結果から学力調査が測定している能力について考察する。

まず，因子分析について簡単に説明をしておく。因子分析は一般に図１のようなモデルを前提としている。図中の中央に四角囲みで「項目１～５」と書かれているのが，例えばテストであれば各問題の得点，アンケート調査であれば各質問項目への回答などを指す。我々が調査によって実際に手にすることができるのはこの部分のデータである。因子分析のモデルでは，この項目の点数の変動に「共通因子」と「独自因子」という２つの潜在変数が影響すると考える。共通因子は図中の一番左に「因子 1」「因子 2」と記載している変数である。例えば，調査しようとしている学力が，単一の能力であれば１つの共通因子が想定されることになるし，「知識の習得」「知識の活用」といったように２つの下位能力によって構成されるとすれば，２つの共通因子が想定されることになる。独自因子は図中の右に記載されている（e1 ～ e5）という潜在変数であり，共通因子では説明できない個々の項目の点数の変動を説明する因子である。テストを例にして考える

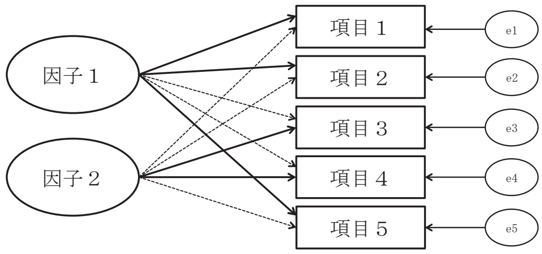

図 1. 因子分析のモデル

と，各項目（問題）の点数や正答確率は，共通因子の影響だけでなく，問題の出題順序，その問題文で使用されている具体例になじみがあるか，など，その他の様々な要因によって影響を受けていると考えられる。このように因子分析モデルでは，項目の点数の変動に影響を与える潜在変数のうち共通因子では説明ができないものを独自因子として想定する。

各因子は各項目の得点の変動に対して影響を与えており，図中では各因子からの矢印がこの影響関係を表現している。共通因子は基本的には全ての項目の得点に影響していると想定されるが，その影響力の強さは項目によって異なる。例えば，因子１が「知識の習得」に対応する能力であれば，学校でならった知識をそのままあてはめて早く正確に答えを導くことを求めるような問題との関係性は強いかもしれないが，現実の文脈における問題解決をもとめるような問題との関係性は弱いかもしれない。このような関係性の強さの違いを図では実線（関係性の強い項目）と破線（関係性の弱い項目）で表現している。例えば，因子１は問題

表1．各学年のテストに関する因子分析の結果 [(3)]

小学４年生		小学６年生		中学１年生	
問題番号	因子負荷	問題番号	因子負荷	問題番号	因子負荷
1	0.62	1	0.39	1	0.56
2	0.58	2	0.54	2	0.74
3	0.54	3	0.57	3	0.71
4	0.55	4	0.70	4	0.49
5	0.35	5	0.62	5	0.77
6	0.41	6	0.72	6	0.74
7	0.50	7	0.58	7	0.85
8	0.58	8	0.65	8	0.79
9	0.60	9	0.64	9	0.66
10	0.69	10	0.63	10	0.79
11	0.66	11	0.75	11	0.57
12	0.43	12	0.81	12	0.63
13	0.60	13	0.56	13	0.71
14	0.54	14	0.69	14	0.52
15	0.58	15	0.62	15	0.78
16	0.57	16	0.71	16	0.78
17	0.62	17	0.65	17	0.74
18	0.55	18	0.57	18	0.73
19	0.59	19	0.79	19	0.68
20	0.65	20	0.44	20	0.53
21	0.62	21	0.67	21	0.67
22	0.58	22	0.65	22	0.59
23	0.66	23	0.70	23	0.64
24	0.50			24	0.74
25	0.40			25	0.90
				26	0.74
				27	0.79

の1，2，5と関連が強く，因子2は問題の3，4と関連が強い。このような関係を知る事によって，問題の特徴から，このテストで測定されている因子がいかなるものであるのか，作成時の仮説に合致したものであるのか，といったことを探っていくのが因子分析と呼ばれる分析手法である。

以下では，まず，分析対象となるテストに含まれている問題の正解，不正解のパターンをもとにして，共通因子を探索することで，このテストが測定している「学力」の質や構造（単一の構成概念であるのか，複数の下位能力の複合体であるのか，など）を検討する。小学4年生，小学6年生，中学1年生の算数・数学のテスト結果について因子分析[2]を行った結果，全ての学年において，単一の因子を測定しているというモデルを想定することで，データを最も適切に説明できると判断された。仮に，テスト設計の段階で算数・数学の「学力」を単一の能力として想定していたのであれば，因子分析の結果はその想定を支持するものであるし，仮に，多様な能力を想定していたのであれば，今回のテストはそのような「学力」を十分に測定できていない可能性が考えられる。

次に因子と各項目の関連性の強さである。因子分析においては，この関連性の強さは因子負荷という指標で与えられる。因子負荷の数値が大きいほど，その項目と因子との関係が強いということになる。表1に各学年のテストに関する因子分析の結果として，各問題の因子負荷の値を示した。例えば小学校4年生のテストであれば，問題番号10，11，23などは測定されている「学力」と関連の強い問題であり，逆に問題番号5，25などの問題は関連の弱い問題であると考えられる。

2.2. 問題内容の分析

以下では，上記の因子分析結果を踏まえ，因子負荷の高い問題と低い問題を比較しながら各問題の内容を精緻に検討することで，分析対象としている学力調査が測定しているであろう「学力」とはいかなるものであったのかを検討する。

(1) 小学4年生のテストの検討

①因子負荷の高い問題例1（小学校第4学年：問題番号10，因子負荷0.69）

ア）問題：花子さんは，360円をもって買い物に行きました。1本25円のあめを買うと，何本変えますか。式と答えを書きましょう。

イ）問題の概要：包含除の場面を理解し，立式し解を求めることができるかどうかをみる。

②因子負荷の高い問題例2（小学校第4学年：問題番号11，因子負荷0.66）

ア）問題：図かんの重さは，物語の本の重さの5倍で1700gです。物語の本の重さを□gとして，かけ算の式に書きなさい。

イ）問題の概要：2つの数量の関係を理解し，□を用いた乗法の式に表すことができるかどうかをみる。

③因子負荷の高い問題例3（小学校第4学年：問題番号23，因子負荷0.66）

ア）問題：下の折れ線グラフは，1年間の気温の変わり方を月ごとに表したものです。ひと月ごとの変わり方がいちばん大きいのは，何月と何月の間か答えなさい。

イ）問題の概要：折れ線グラフの傾きに着目して，グラフの特徴を読むことができるかどうかをみる。

上記の①〜③の問題は，大きく2段階の思考が求められる問題である。まず，第1段階は，示されたテクストやグラフから数量の関係や変化の特徴を読み取ることが求められている。次に第2段階では，把握した数量の関係やグラフの特徴を基に立式したり，データを取り出したりすることが求められている。上記の問題例以外にも，第1段階として複数の単位間の関係を捉えたり，示された図を観察して構造をとらえたりして，第2段階で問題の解決を実行する問題（問題13，20，21）では，因子負荷が高い傾向にある。

次に，因子負荷の低い問題の特徴について検討を行う。

④因子負荷の低い問題例1（小学校第4学年：問題番号5，因子負荷0.35）

ア）問題：次の計算をしなさい。

4306 − 267

イ）問題の概要：整数の減法（4 位数－ 3 位数）の計算ができるかどうかをみる。

⑤因子負荷の低い問題例 2（小学校第 4 学年：問題番号 6，因子負荷 0.41）
ア）問題：次の計算をしなさい。
846 × 53
イ）問題の概要：整数の乗法（3 位数× 2 位数）の計算ができるかどうかをみる。

⑥因子負荷の低い問題例 3（小学校第 4 学年：問題番号 12，因子負荷 0.43）
ア）問題：□にあてはまる数を答えなさい。
3 L = □ mL
イ）問題の概要：体積の単位間の換算ができるかどうかをみる。

⑦因子負荷の低い問題例 4（小学校第 4 学年：問題番号 25，因子負荷 0.40）
ア）問題：ある学級で，先週と今週の図書館の利用の様子を調べ，下のような表にまとめました。
この学級の人数は，全部で何人か答えなさい。

図書館の利用の様子（人）

		先週	
		行った	行っていない
今週	行った	13	5
	行っていない	10	4

イ）問題の概要：二次元表の読みができるかどうかをみる。

④～⑦の問題は，その解決に際して，複数段階の思考を要するものではないことが特徴である。具体的には，④，⑤は計算の技能の確実な習得がなされているかどうかをみる問題であり，⑥，⑦は単位や二次元表の読みに関する知識が確実に習得されているかどうかをみる問題である。このように，1 段階の思考で解決できる知識及び技能の定着に関する問題については，因子負荷は低い傾向にある。

(2) 小学 6 年生のテストの検討
①因子負荷の高い問題例 1（小学校第 6 学年：問題番号 11，因子負荷 0.75）
ア）問題：（図で半径 10 cm の円が示された上で）下の円の円周の長さは □ cm です。
イ）問題の概要：半径 10 cm の円の円周の長さを求めることができるかどうかをみる。

②因子負荷の高い問題例 2（小学校第 6 学年：問題番号 12，因子負荷 0.81）
ア）問題：1000 円の服を 25% 引きの値段で買いました。（消費税は考えません。）代金を求めるために，
太郎くんは次のように計算をしました。次のア～ウの中に入る数を答えなさい。
同じ記号には同じ数が入るものとします。

太郎くんの考え

$\boxed{ア}-0.25=\boxed{イ}$
$1000\times\boxed{イ}=\boxed{ウ}$
答え$\boxed{ウ}$円

イ）問題の概要：問題の構造と解法とを対応させて，解法を完成させることができるかどうかをみる。

③因子負荷の高い問題例 3（小学校第 6 学年：問題番号 19，因子負荷 0.79）
ア）問題：次の場面で，x と y の関係を表した式を，ア～エの中から選び，記号で答えなさい。（完答）
(1) 縦の長さが x cm，横の長さは y cm の長方形があります。面積は 20 cm^2 です。

(2) 兄の持っている折り紙は x 枚です。弟の持っている折り紙は y 枚です。合わせると 20 枚です。

ア　20 + x = y　　イ　20 − x = y　　ウ　20 × x = y　　エ　20 ÷ x = y

イ）問題の概要：2 つの数量の関係を理解し，的確な式を選択することができるかどうかをみる。

上記の①～③の問題は，複数段階の思考が求められる問題である。例えば②の問題では，まず，示された問題の構造を理解することが求められる。次に，示された「太郎くんの考え」が，どのような方略に基づく解法なのかを理解することが求められる。最後に，問題の構造の理解，解法の理解に基づいて，解法を完成させることが求められる。また，上記の問題以外にも，第 1 段階として安定した位置に置かれていない平行四辺形の底辺と高さを取り出し，第 2 段階で面積を求める問題（問題 16）でも，因子負荷が高い傾向にある。このように，小学 6 年生のテストも，小学 4 年生のテストと同様に，複数段階における思考が伴う問題においては，因子負荷が高い傾向にある。

次に，因子負荷の低い問題の特徴について検討を行う。

④因子負荷の低い問題例 1（小学校第 6 学年：問題番号 1，因子負荷 0.39）

ア）問題：次の計算をしなさい。

5.25 × 1.6

イ）問題の概要：小数の乗法の計算ができるかどうかをみる。

⑤因子負荷の低い問題例 2（小学校第 6 学年：問題番号 20，因子負荷 0.44）

ア）問題：右の図は線対称でもあり，点対称でもある図です。（図は省略）

図を見て次の問題に答えなさい。

(1) 直線 AB を対称の軸とすると，頂点アに対応する頂点を答えなさい。

イ）問題の概要：線対称な図形の性質を理解しているかどうかをみる。

④，⑤の問題は，その解決に際して，複数段階の思考を要するものではないことが特徴である。具体的には，④は計算の技能の確実な習得がなされているかどうかをみる問題であり，⑥は線対称な図形の性質を確実に理解しているかどうかを見る問題である。このように，1 段階の思考で解決できる知識及び技能の定着に関する問題については，小学 4 年生のテストと同様に因子負荷は低い傾向にある。

(3) 中学校第 1 学年の調査の検討

①因子負荷の高い問題例 1（中学校第 1 学年：問題番号 7，因子負荷 0.85）

ア）問題：a = − 3 のとき，4a + 15 の値は $\boxed{}$ である。

イ）問題の概要：代入して，文字式の解を求めることができるかどうかをみる。

②因子負荷の高い問題例 2（中学校第 1 学年：問題番号 8，因子負荷 0.79）

ア）問題：方程式　3x = (5x − 7) ／ 4 を解くと，$\boxed{x = }$ である。

イ）問題の概要：一元一次方程式を解くことができるかどうかをみる。

③因子負荷の高い問題例 3（中学校第 1 学年：問題番号 25，因子負荷 0.90）

ア）問題：兄と妹が，同じマラソンコースを同時に出発して，それぞれ一定の速さで 2000 m 走りました。右の図（省略）は，兄と妹が走った時間と道のりの関係をそれぞれグラフに表したものです。(3) 兄と妹の走った道のりの差がはじめて 300 m になるのは，2 人が出発してから $\boxed{\text{分後}}$ である。

イ）問題の概要：2 つの事象の速さを求め，問題を解決することができるかどうかをみる。

④因子負荷の高い問題例 4（中学校第 1 学年：問題番号 27，因子負荷 0.79）

ア）問題：右の図（省略：底面が a の正方形で，高さが b の四角柱）は底面が正方形である四角柱です。この図で，体積を表す式を a，b を使って表すと $\boxed{}$ である。

イ）問題の概要：四角柱の体積を文字式で表すことができるかどうかをみる。

上記の①，②，④の問題は，多段階の思考を要するものではないことが特徴である。具体的には，計算の技能の確実な習得がなされているかどうかをみる問題であり，②は式変形が確実にできるかどうかをみる問題であり，④は，小学校で学習した柱体の体積の公式に文字をあてはめて立式できるかどうかをみる問題である。また，③の問題は一見すると，図の解釈，立式，解を求めるといった３段階の思考を要するように見えるが，図から４分間当たりの距離の差を読み取ればほぼ１段階の思考で解決できる問題である。

⑤因子負荷の低い問題例１（中学校第１学年：問題番号 4，因子負荷 0.49）
ア）問題：次の計算をしなさい。
2.9×1.8
イ）問題の概要：小数の乗法の計算ができるかどうかをみる。

⑥因子負荷の低い問題例２（中学校第１学年：問題番号 14，因子負荷 0.52）
ア）問題：5 人の班で，班長と副班長をそれぞれ１名ずつ選ぶとき，選び方は全部で $\boxed{}$ 通りである。
イ）問題の概要：簡単な場合の組み合わせを求めることができるかどうかをみる。

上記の⑤，⑥の問題は，それぞれ小学５年生，６年生の内容の定着が確実になっているかどうかをみる問題である。このように，中学１年生のテストでは，小学校の内容に関する問題については因子負荷が低い傾向にある。ただし，単純な異分母分数の減法（問題番号 3，因子負荷 0.71）や基準量と割合を基に比較量を求める問題（問題番号 10，因子負荷 0.79）といった例外もみられるため，因子負荷の低い問題については，今後さらなる検討が必要になる。

（1）～（3）の具体的な問題に基づく検討から，本調査において測定しようとしている学力とは，小学校においては当該学年の学習の定着度の中でも，特に２段階の思考を要する問題解決能力であり，１段階の思考で即再現ができる知識及び技能の確実な定着ではないと考えられる。一方，中学校では，小学校とは逆に，１段階の思考で再現ができる知識及び技能の確実な定着を調査しているものと考えられる。3 つの調査は，それぞれ実施時期が異なるため，一概には結論づけることは難しいが，本調査の表題にある「学習定着度に関する調査」の「学習定着度」に関するターゲットが，小学校と中学校では異なることは興味深い。

しかしながら，問題形式だけに着眼すると，本調査は，算数・数学を活用して思考・判断する問題は出題されておらず，調査の性質としては，当該学年段階までのカリキュラムの在り方を考察の対象としている TIMSS 調査や全国学力・学習状況調査の A 問題の調査に近いと考えられる。

3. テストに含まれる各問題はどのような特徴を持っているか

3.1. 項目分析の考え方

テストに含まれる個々の問題は，妥当性の他にも，難易度（学力調査を受検した集団にとって，各問題がどのくらい難しかったのか），識別力（各問題に受検者の学力水準の違いを区別する機能がどの程度備わっているのか），信頼性（テストに含まれる各問題の一貫性や安定性など）といった特徴を持っており，これらの側面から問題の質を評価することができる。このような問題の特徴を把握するための分析を項目分析と呼ぶ。今回は，古典的テスト理論という考え方に基づく項目分析（具体的な手順は豊田（2002b）に従う）によって学力調査に含まれる問題の特徴を検討する。

最初に，各問題の難易度の指標として正答率を算出した。次に識別力について検討した。識別力とは，各問題がテスト全体で測定している構成概念を適切に反映しており，かつ，受検者の能力の違いを区別することができているかどうか，という性質を表す指標である。識別力の指標としては I-R 相関を計算した。I-R 相関とは，テストの総点から検討対象となる問題の点数を引いた得点と，検討対象となる問題の得点との積率相関係数である。この値は理論上 −1.0 から 1.0 の値を示し，絶対値が大きいほど２つの得点の関係性が強い事を示す。また，正負の符号は関係の方向を示しており，符号が正の場合には一方の得点が高くなると他方の得点も高くなり，符号が負の場合には一方の得点が高くなると他方の得点は低くなる，といった関係

にある。つまり，テスト全体の得点が高い受検者ほど正答できるといった関係が明確な問題ほどI-R相関の値は正の符号で，かつ絶対値が高くなる。また，テスト全体の得点が高い受検者でも間違っている人がある程度の割合で存在するし，逆にテスト全体の得点が低い受検者の中にも正答できている人がある程度の割合で存在するといったような問題の場合には，符号は正であったとしても絶対値は低くなる。I-R相関の絶対値が低い問題は，一般に，そのテスト全体として測定している能力の違いについての識別力が低い項目だと考えられるため，短縮版のテストを作成する際などには，優先的に削除する項目の候補として考えることができる。特に注意すべきは，I-R相関の値の符号が負になるような問題が見られた場合である。その問題に正解した人ほどテスト全体の得点が低くなるという関係があるということになる。反応の一貫性という観点からの信頼性が著しく低い項目と言える。このような事態が生じる原因としては，そのテスト全体として測定している潜在概念とは明らかに異質な潜在概念を測定するための問題が混ざり込んでいる，学力が高い受検者ほど混乱して誤答が多くなるような選択肢が紛れ込んでいる，などの不備が考えられるため，問題の再検討が求められる。

3.2. テスト問題の項目分析

図2，3，4は小学4年生，小学6年生，中学1年生の学力調査に含まれていた問題について，横軸を正答率，縦軸を識別力とする平面上に各問題をプロットした散布図である。正答率が低い（相対的な難易度が高い）問題の場合は，学力が高い受検者であっても不正解になる確率が高いために識別力は低くなりがちである。逆に，正答率が高い（相対的な難易度が低い）問題の場合は，学力の低い受検者であっても正解できるために識別力は低くなりがちである。正答率が中程度の項目であれば，学力が高い子は正解できるが，学力が低い子は不正解になるという関係が明確になりやすいため，識別力も高くなる傾向がある。したがって逆U字型の散布図になることが一般的である。今回の学力調査においては全ての学年に共通して，難易度が低い（正答率が高い）ために識別力が低くなっていると考えられる問題が複数見られる。特に，小学校4年生の問題には他の学年と比べて識別力の低い項目が多く含まれている。受検者の能力を識別するという目的に重点を置いたテストを作成する場合には，このような識別力の低い問題は優先的に検討されるべきだと言える。ただし，全ての受検者に一定の点数を取らせることで自信を持たせたい，単位認定のために全員がある程度の点数をとれるようなテストを作成したい，というような目的がある場合にはこのような項目も必要となる。また，詳細は後述するが，測定対象となる集団の質が異なると，そのテスト項目の難易度や識別力が変わるということもあり，必ずしもその問題が常に不適切であるとは限らない点にも注意が必要である。

3.3. 項目特性図を用いた項目分析

上記の識別力に関する議論は今回のテストの受検者全体を検討の対象としたものであるが，受検者集団の中には様々な能力を持つ受検者が含まれている。そこで，どのような能力水準の受検者にはどのような問題が有効であるかという観点から問題についての検討を行うことも必要である。難易度が低い問題は，受検者全体については識別力の低い問題となる傾向が見られるが，能力が低い受検者を主な対象とした調査の場合には，中程度の難易度を持つ問題よりも高い識別力を示すことがある。なぜなら，受検者全体にとって難易度が中程度の問題というのは，能力が低い受検者にとっては高難度の問題となるため，正答率が低くなってしまい，識別力も低くなる可能性が想定されるからである。

このような，受検者の能力水準との関連において問題の識別力を把握する方法の1つに項目特性図を描くという方法がある。作図の方法は，受検者全体を成績順に複数のグループに分け，そのグループを横軸，各グループの正答率を縦軸とする折れ線グラフを作成する。今回は中学1年生の学力調査に含まれている問題の中から，特に特徴的なパターンを示した4つの問題についての項目特性図を作図した（図5，6，7，8）。なお，今回の項目特性図の作成にあたっては，分析対象となった1000人の子どもを200名ずつの5グループ（成績上位のグループからG1，G2，G3，G4，G5）に分類した。

図5は，中学校1年生の調査で用いられた問題の中でも正答率が最も高く（94.8%），識別力が最も低かった（0.23）問題の項目特性図である。G1の生徒もG5の生徒も高い割合で正答できており，全体としての正答率が高くなるためにグラフが上側に位置しており，グループ間で正答率に違いが見られないためにほぼ横ばいの状態になっていることがわかる。あらゆる能力水準の受検者に対して識別力の低い問題であると言える。このような問題は図4の散布図上では右下あたりに位置することになる。図6は，正答率が中程度

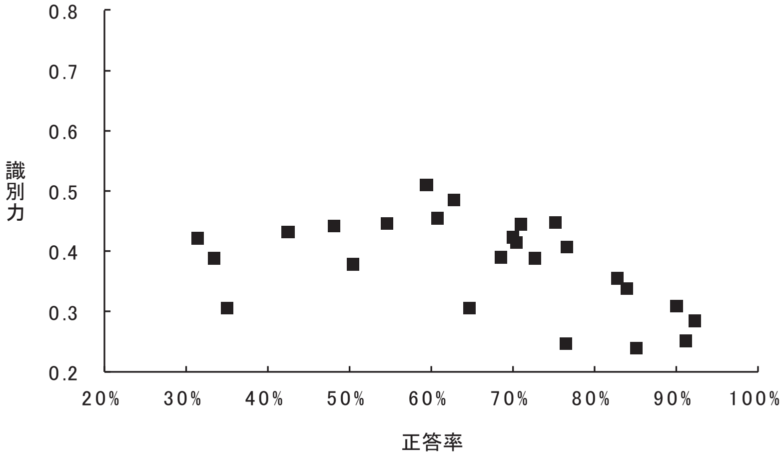

図 2. 小学 4 年生のテストにおける各問題の正答率と識別力の散布図

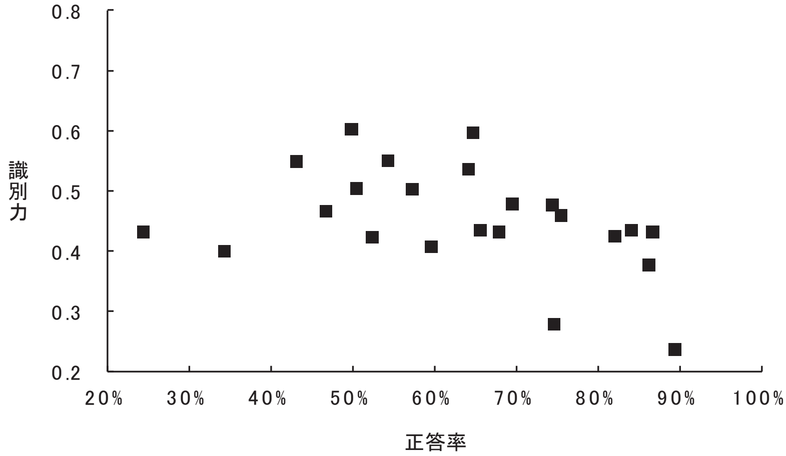

図 3. 小学 6 年生のテストにおける各問題の正答率と識別力の散布図

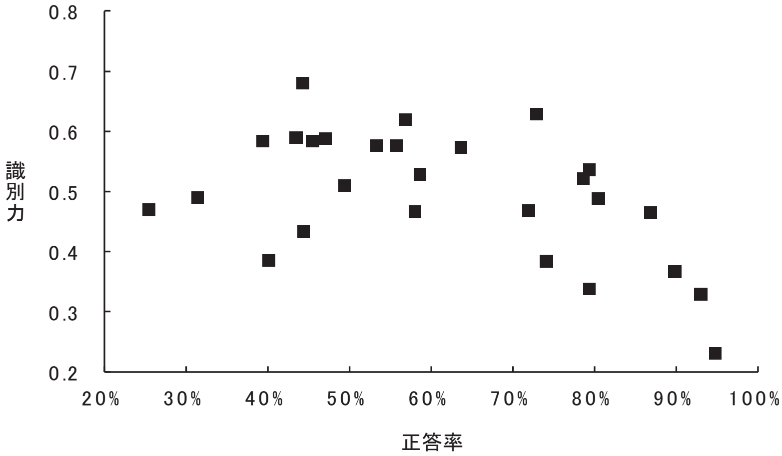

図 4. 中学 1 年生のテストにおける各問題の正答率と識別力の散布図

（56.8%）で，識別力が高い（0.62）問題の項目特性図である。テスト全体の得点が高いグループほど正答率も高くなるという右上がりの直線になっている。あらゆる能力水準の受検者にとって識別力の高い問題であると言えよう。図 7 も図 6 の問題と同程度に識別力が高い（0.63）問題の項目特性図だが，図 6 とは形状が

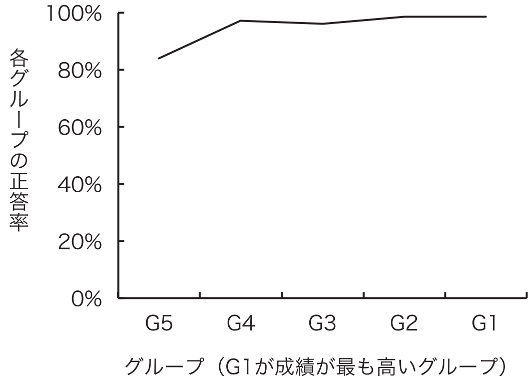

図 5. 問題 1 の項目特性図
（正答率 94.8%，IR 相関＝ 0.23）

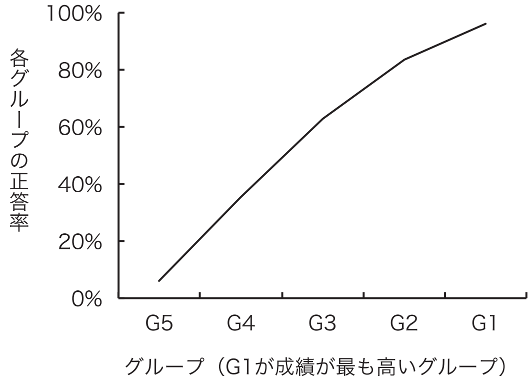

図 6. 問題 10 の項目特性図
（正答率 56.8%，IR 相関＝ 0.62）

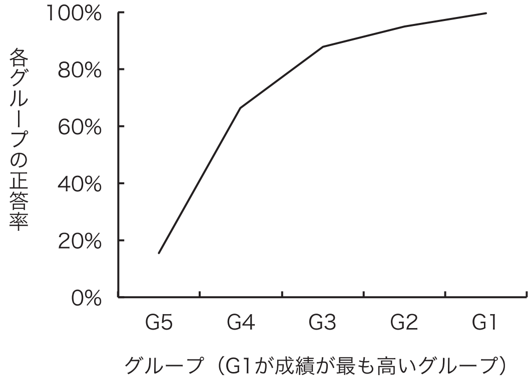

図 7. 問題 7 の項目特性図
（正答率 72.9%，IR 相関＝ 0.63）

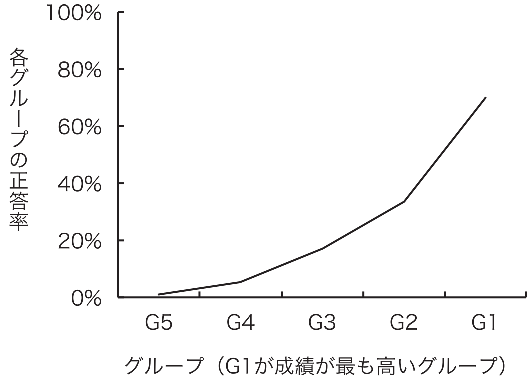

図 8. 問題 17 の項目特性図
（正答率 25.4%，IR 相関＝ 0.47）

異なっている。図 7 では G5 と G4 の間で正答率に大きな違いが見られるが，その他のグループ間での変化は比較的なだらかになっている。正答率は 72.9% と受検者集団全体にとってはさほど難易度の高くない問題だが，能力水準の低い受検者にとっては，それなりに難しい問題であったと考えられる。そのため，G5 に含まれる受検者に対しては高い識別力を示すが，G2 や G1 に含まれる受検者に対しては識別力が低くなる問題であったことがわかる。逆に，図 8 は能力水準の高い受検者に対して高い識別力を示す問題の項目特性図である。全体的な識別力もある程度高い（0.47）が，問題の難易度が高い（正答率 25.4%）ため，特性図としては G2 や G1 にかけての正答率の変化が急になっている。

このように受検者集団全体を対象とした識別力という性質とは別に，どのような学力水準の受検者に対して高い識別力を持つかといった点からも問題の特徴を検討することが可能である。受検者の能力水準と問題の識別力の関係は，受検者集団全体を対象として得られた値には反映されにくいため，項目特性図を利用した分析が重要である。このような情報を把握しておくことで，学力が低いと想定される受検者集団に対して識別力が高くなる問題を集めたテストや，学力が高いと想定される受検者集団に対して識別力が高くなる問題を集めたテスト，といったように対象者に応じたテスト設計を行うための基礎的な情報を得ることも可能となる。

4. まとめと今後の課題

本稿では，因子分析・項目分析といった技法を使って，A 市の学力調査を分析してきた。それでは，こうした分析から見えてくる，自治体学力調査の課題とは何だろうか。

もっとも大きな課題は，A 市教育委員会が，この算数・数学の学力調査を通して，「一体何の能力を測ろ

うとしていたのか」という点が，はっきりしていないという点である。テストで測定しようとした構成概念が，せいぜい算数・数学の学力というレベルにとどまっており，PISA や TIMSS といった国際調査に比べると，非常にあやふやなのである。

因子分析の結果から，今回のテストが，何らか一つの能力を測ることができていたということはわかる。また，個々の項目の因子負荷から，小学校では２段階の思考を要する問題解決能力，中学校では１段階の思考で再現できる知識および技能の定着を調査する設計になっていたということもわかる。ただ，これらが，教育委員会の意図通りだったのか，それとも意図とは違っていたのか，こうしたことを計量分析から明らかにすることはできない。学力調査を実施する側には，「そもそも何を測りたいのか」という点を明らかにすることが求められるのである。

もう一つの問題は，「何のためにテストをするのか」が明らかではないという点である。項目分析の結果から，今回のテストに，難易度が低く識別度の低い項目（≒皆が正解できる問題）が複数含まれていたことはわかる。しかし，こうした問題をテストに含めるべきなのか，それとも除外するべきなのかは，あくまでテストの目的に依存する。すべての子どもに一定の点数をとらせることが目的ならば，こうした問題は残すべきである。他方で，教育政策に活かすことを重視し，成績が正規分布することを目指すのであれば，こうした問題は排除した方が好ましい。計量分析は，あくまでテストに含まれている個々の項目の特徴を明らかにするだけであり，そこから，どの項目を選び，どの項目を削除するのかは，あくまで調査を実施する側が判断しなければならない。

その意味では，自治体学力調査の課題は，因子分析や項目分析による信頼性を高めることももちろんだが，そもそも「調査によって何を測ろうとしているのか」「何のために調査をしようとしているのか」という調査の前提をはっきりさせなければならないという点にある。

なお，誤解のないように指摘しておくが，これは調査の目的や対象の検討という「妥当性」だけを高めればよいという主張ではない。調査の目的や対象の検討という「妥当性」に関する議論と，テストに含まれる各問題の一貫性や安定性の検討という「信頼性」に関する議論は，両立をめざすべきものである。妥当性が高いが信頼性が低い評価方法が信頼できないことと同じように，信頼性が高いが妥当性が検討されていない評価方法は現実を歪めて（極めて限定的に）とらえている可能性がある。たとえば「何を測っているかよくわからないが，子どもの能力を明確に弁別できる指標（≒信頼性は高いが妥当性がない指標）」が開発されたとして，そこにどんな意味があるだろうか。信頼性の高さのみを追究することは，妥当性にかかわる偏りになりうる。こうした偏りというリスクへの思慮深さが求められる。

妥当性との擦り合わせのない評価方法（項目）の決定はもちろんのこと，擦り合わせがおこなわれた場合でも，それは「子どもの学びの一部を切り取ったものであるかもしれない」という前提で運用されるべきである。当然のことながら，学力テストを政策評価・カリキュラム評価・学校評価として活用する場合は，より慎重な態度が求められる[(4)]。

今回は紙幅の都合上，項目反応理論など，近年の新しいテスト理論の技法には，ほとんど触れることができなかった。項目反応理論による分析は，古典的テスト理論よりも高度な知識と技術が求められるが，小学校から中学校までの経時的な算数・数学の学力の変化を検討することができるようになる等，将来的なテスト設計のための豊かな情報を得る事が可能である。すでに，自治体で実施されているテストに対して項目反応理論を用いて分析をしている研究などもあり（e.g., 石井・安永，2011：斉田・柳川，2011），今後，自治体学力調査への更なる応用が期待される。この点については，次回の課題としたい。

＜註＞

(1) 学力に関する因果推論・パネルデータの重要性については，たとえば赤林他編（2016）を参照。

(2) 今回行った因子分析は探索的因子分析と呼ばれるものである。因子分析に用いたデータは，各問題について正解＝１，不正解＝０の二値変数のデータであったため，カテゴリカル因子分析を用いた。因子数については，ポリコリック相関係数行列を用いて，スクリープロットによる固有値の減衰状況を確認し，第１因子に対する寄与率の確認を行った。豊田（2002a）は第１因子の寄与率が 20% を超えていることを１因子構造と見なす基準の１つとして挙げているが，今回の結果は全ての学年のテストにおいて１因子構造からなる構成概念を測定していることを示唆するものであった（小学４年生のテスト結果における第１因子の寄与率＝ 34.3%，小学６年生のテスト結果における第１因子の寄与率＝ 43.8%，中学

校１年生のテスト結果における第１因子の寄与率＝50.1%）。また，平行分析，及びMAP（Minimum Average Partial）基準を用いた判断などの複数の方法で因子数を判断するための情報を得たが，いずれも１因子構造を支持するものであった。

(3) 実際のテストは複数の大問と小問から構成されていたが，自治体や実施年度などが特定されることを避けるために今回は問題番号を全て通し番号とした。

(4) 標準化テストが学校や教師に引き起こした「教育学的に」歪んだ状況については多くの研究蓄積がある（例えば，Apple（2008），Ravitch（2015）など）。

＜参考文献一覧＞

赤林英夫・直井道生・敷島千鶴編著，2016，『学力・心理・家庭環境の経済分析』有斐閣。

Apple, M, W, 2006, *Educating the Right Way* (*2nd Edition*), Routledge（＝大田直子訳，2008，『右派の／正しい教育【第２版】』世織書房）.

石井秀宗・安永和央，2011，「全項目が開示されるテスト文化のもとでの得点分布の経年比較―全国テストと自治体テストのリンキング」『日本テスト学会誌』7（1），pp.24-35.

苅谷剛彦・志水宏吉編，2004，『学力の社会学』岩波書店。

中室牧子，2015，『「学力」の経済学』ディスカバー・トゥエンティワン。

OECD，2014，*PISA 2012 Technical Report*, OECD.

Olson, J.F., Martin, M.O., & Mullis, I.V.S., 2008, *TIMSS 2007 Technical Report*, TIMSS & PIRLS International Study Center.

Ravitch, D, 2013, *Reign of Error*, Knopf Doubleday Publishing（＝末藤美津子訳，2015，『アメリカ 間違いがまかり通っている時代』東信堂）.

斉田智里・柳川浩三，2011，「共通項目デザインによる神奈川県高等学校「県下一斉英語学力テスト」の開発―項目応答理論を用いた等化によるテストの再評価と展望」『日本テスト学会誌』7（1），pp.122-132.

豊田秀樹，2002a，『項目反応理論［事例編］』朝倉書店。

豊田秀樹，2002b，『項目反応理論［入門編］』朝倉書店。

志水宏吉編，2009，『「力のある学校」の探求』大阪大学出版会。