８.項目に関する IRT 分析

ここからは項目反応理論にもとづくモデルパラメータ（項目母数や学力特性値）の推定に必要なノウハウを見いだすために，算数と数学のそれぞれの結果を対比させる形で検討をすすめる。その前提条件として，推定に利用した BILOG-ＭＧではいずれのデータにおいても学力特性値（θ）は平均０，分散１の標準正規分布にしたがっていることを仮定している。算数と数学の学力が連続しているものであるなら，本来なら，算数の調査を受検した小学生集団と数学の調査を受検した中学生集団とのあいだで少なくともその学力特性の母集団分布における後者の平均を前者よりも高く設定する必要がある。しかし，今回得られたデータにはそのような情報は含まれていない点に注意する必要がある。なお，教育測定学上，これは垂直等化(vertical equating)の問題として扱われるものであり，本調査研究の目的には含まれないが検討すべき極めて重要な問題である。

8.1 項目母数の推定結果

8.1.1 安定した推定に必要な受検者数

算数データの全体および数学データの全体を利用してもとめた項目母数は，それぞれ資料 13 及び資料 17（または，表 7-1,2 を参照）に掲載した。これらの推定には一つの項目あたり，共通ブロックに含まれる項目には全受検者数 2400 名程度，それ以外の項目では 900 名程度が含まれている。一方，資料 16 及び資料 20 には，それぞれ，算数と数学の各分冊（計 10）のデータのみを利用して推定された識別力母数と困難度母数をグラフと表で示されている。例えば算数テストの共通ブロックに含まれる 15 個の項目の識別力母数と困難度母数に関する分冊データからの推定値の様子は図 8-1 および図 8-2 に示すとおりである。いずれの図も右端が全体での推定値となっている。図をみれば明らかなように，例えば識別力母数においては B0-3，B0-8 が分冊によって大きく推定値の値が異なっていることがわかる。また項目困難度母数に関しては，B0-4 などに顕著な推定結果の差が見られる。

同様に数学データのブロック３に含まれる項目群をみて見ると，図 8-4 と表 8-4 に掲げた項目困難度の母数はどの分冊においても比較的安定した推定結果となっている。その一方で，項目識別力母数の推定結果においては，図 8-3，表 8-3 に示すように B3-3 の違いが激しい。

算数テストならびに数学テストの他の分冊においても上と同じような動きを示す項目が散見される。推定値の違いについては様々な要因が考えられるが，やはり大きなものとしては受検者数の多寡が指摘できるであろう。今回の調査研究の計画を立てるにあたっては，これまでのいわば経験則から，項目一つあたりの受検者数が 1000 名程度以上，逆に受検者一人あたりの項目数が 30 個以上を目処にして，デザインを組み立てている。分冊ごとのデータを使っているため，受検者数としてはいずれも 230 名から 250 名程度であることを考えると，重複テスト分冊法による実施ににあたってのノウハウの一つとして，上の数値は目安となり得るものであると判断できる。

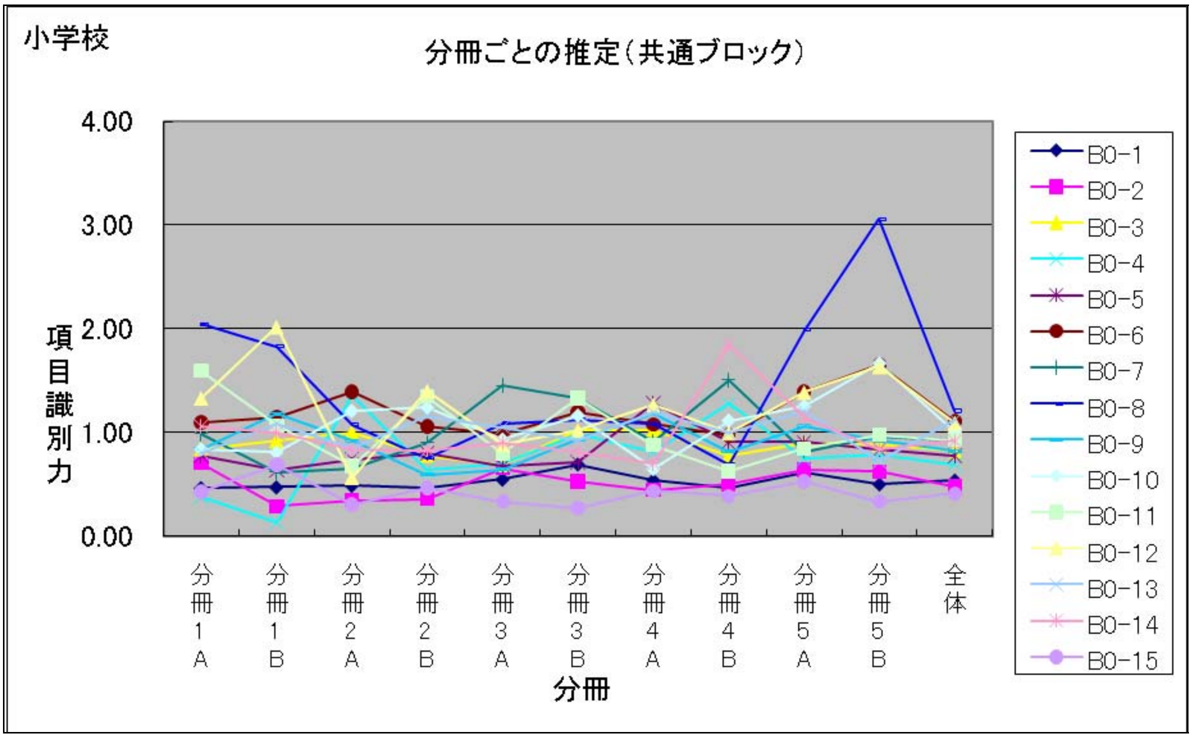

図 8-1 共通ブロックに含まれる項目の分冊ごとの識別力母数の推定値の違い（算数）

表 8-1 共通ブロックに含まれる項目の分冊ごとの識別力母数の推定値（算数）

項目	分冊1A	分冊1B	分冊2A	分冊2B	分冊3A	分冊3B	分冊4A	分冊4B	分冊5A	分冊5B	全体
B0-1	0.46	0.47	0.49	0.46	0.55	0.69	0.54	0.47	0.61	0.50	0.54
B0-2	0.70	0.29	0.34	0.36	0.66	0.53	0.44	0.50	0.64	0.62	0.48
B0-3	0.84	0.92	1.00	0.78	0.69	1.02	1.03	0.78	0.90	0.88	0.82
B0-4	0.37	0.13	1.35	0.64	0.70	0.98	0.81	1.28	0.75	0.78	0.69
B0-5	0.77	0.63	0.74	0.80	0.67	0.71	1.28	0.90	0.91	0.83	0.77
B0-6	1.10	1.15	1.39	1.06	0.98	1.20	1.08	0.97	1.39	1.65	1.11
B0-7	0.99	0.61	0.65	0.90	1.45	1.33	0.91	1.51	0.82	0.96	0.92
B0-8	2.05	1.83	1.08	0.75	1.08	1.13	1.08	0.69	1.99	3.06	1.21
B0-9	0.84	1.18	0.92	0.59	0.64	0.93	1.18	0.82	1.06	0.92	0.82
B0-10	0.84	0.81	1.21	1.24	0.93	1.16	0.65	1.11	1.25	1.66	1.00
B0-11	1.59	1.07	0.68	1.35	0.80	1.33	0.88	0.62	0.84	0.98	0.93
B0-12	1.33	2.02	0.56	1.40	0.89	1.02	1.27	1.00	1.39	1.64	1.09
B0-13	0.80	1.09	0.84	1.16	1.05	0.92	1.20	0.99	1.21	0.72	1.12
B0-14	1.06	1.00	0.84	0.83	0.90	0.82	0.70	1.86	1.14	0.84	0.92
B0-15	0.43	0.69	0.30	0.46	0.33	0.27	0.44	0.39	0.53	0.34	0.41

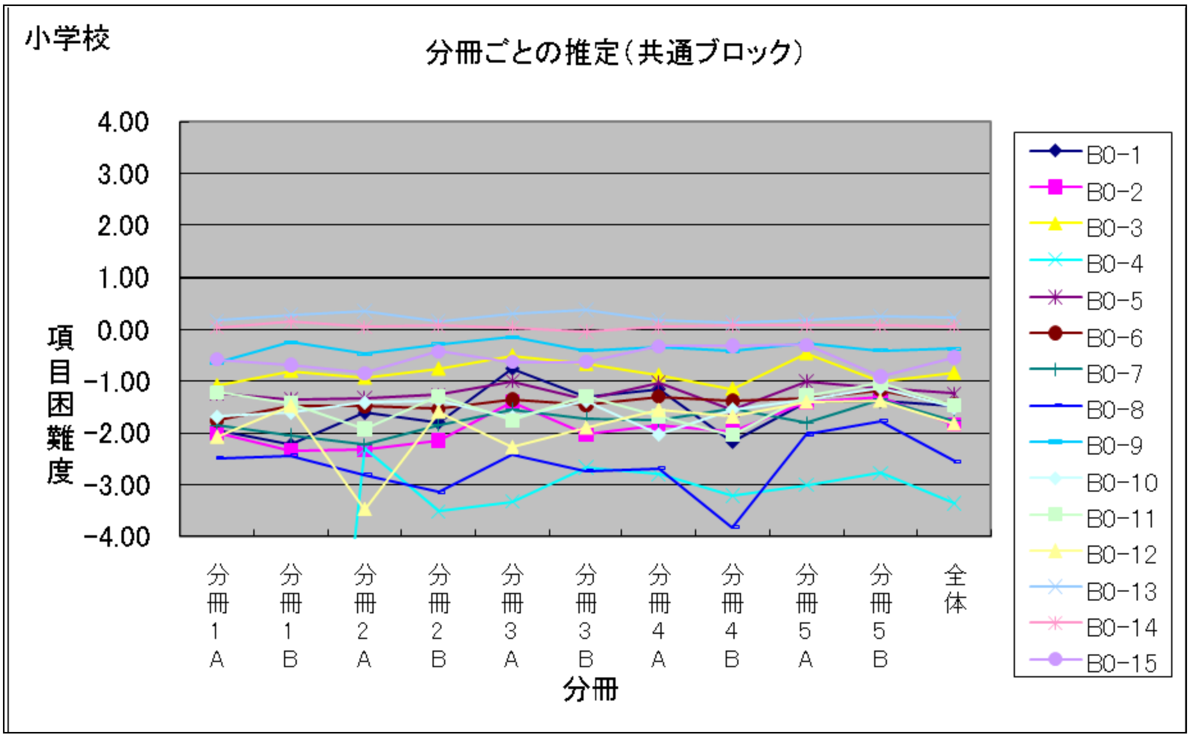

図 8-2　共通ブロックに含まれる項目の分冊ごとの困難度母数の推定値の違い（算数）

表 8-2　共通ブロックに含まれる項目の分冊ごとの困難度母数の推定値（算数）

項目	分冊1A	分冊1B	分冊2A	分冊2B	分冊3A	分冊3B	分冊4A	分冊4B	分冊5A	分冊5B	全体
B0-1	-1.94	-2.22	-1.60	-1.79	-0.75	-1.29	-1.14	-2.16	-1.39	-1.38	-1.48
B0-2	-2.00	-2.33	-2.33	-2.14	-1.40	-2.03	-1.85	-1.96	-1.41	-1.29	-1.84
B0-3	-1.08	-0.79	-0.93	-0.75	-0.50	-0.66	-0.88	-1.14	-0.45	-0.99	-0.83
B0-4	-5.74	-15.46	-2.30	-3.52	-3.33	-2.66	-2.80	-3.21	-3.00	-2.77	-3.36
B0-5	-1.23	-1.35	-1.32	-1.26	-1.00	-1.34	-1.03	-1.56	-1.01	-1.14	-1.24
B0-6	-1.76	-1.48	-1.47	-1.53	-1.35	-1.46	-1.29	-1.38	-1.33	-1.17	-1.45
B0-7	-1.85	-2.04	-2.23	-1.84	-1.54	-1.72	-1.75	-1.53	-1.81	-1.35	-1.77
B0-8	-2.49	-2.43	-2.81	-3.14	-2.41	-2.73	-2.69	-3.82	-2.02	-1.76	-2.55
B0-9	-0.64	-0.24	-0.47	-0.28	-0.14	-0.40	-0.33	-0.41	-0.27	-0.40	-0.37
B0-10	-1.68	-1.60	-1.41	-1.37	-1.69	-1.38	-2.02	-1.54	-1.35	-1.12	-1.51
B0-11	-1.21	-1.43	-1.92	-1.29	-1.76	-1.29	-1.67	-2.04	-1.27	-1.03	-1.47
B0-12	-2.07	-1.48	-3.48	-1.59	-2.28	-1.90	-1.56	-1.68	-1.40	-1.37	-1.82
B0-13	0.18	0.29	0.36	0.17	0.31	0.38	0.19	0.14	0.19	0.26	0.24
B0-14	0.04	0.15	0.06	0.08	0.04	-0.03	0.07	0.09	0.08	0.08	0.07
B0-15	-0.58	-0.68	-0.84	-0.42	-0.62	-0.62	-0.32	-0.31	-0.29	-0.91	-0.53

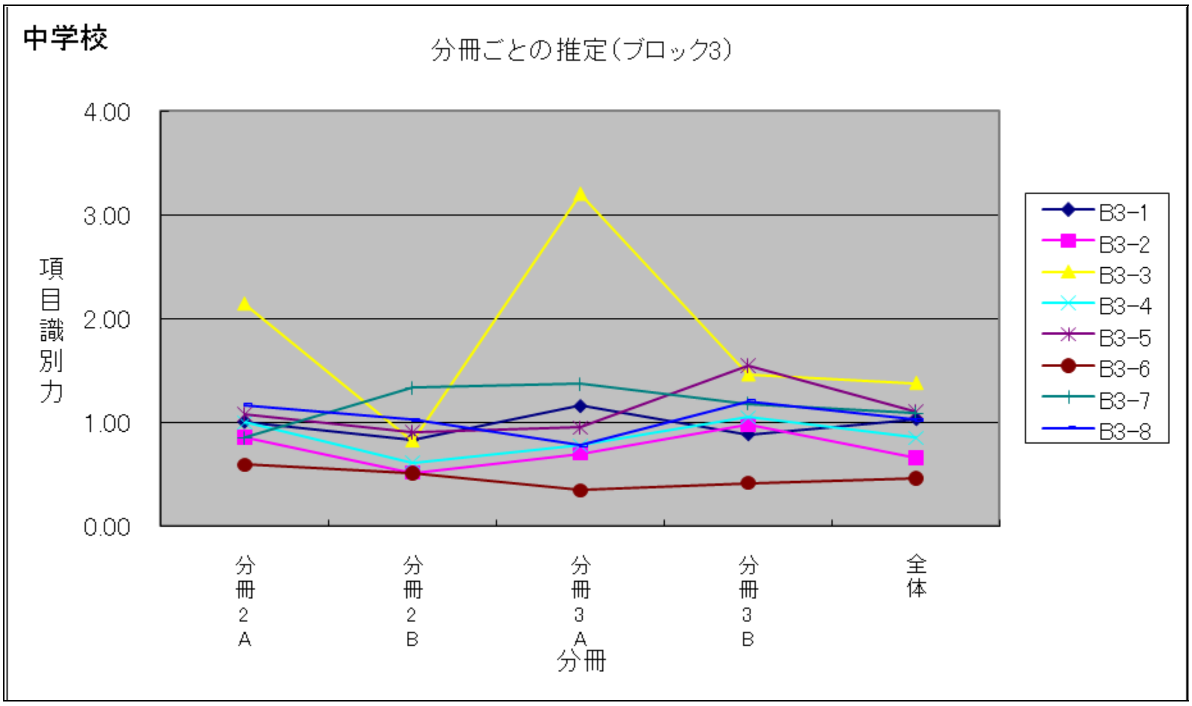

図 8-3 共通ブロックに含まれる項目の分冊ごとの識別力母数の推定値の違い（数学）

表 8-3　共通ブロックに含まれる項目の分冊ごとの識別力母数の推定値（数学）

項目	分冊2A	分冊2B	分冊3A	分冊3B	全体
B3-1	1.00	0.84	1.16	0.88	1.03
B3-2	0.86	0.51	0.70	0.98	0.66
B3-3	2.14	0.82	3.21	1.46	1.38
B3-4	1.00	0.61	0.78	1.05	0.85
B3-5	1.08	0.91	0.95	1.55	1.10
B3-6	0.59	0.51	0.35	0.42	0.46
B3-7	0.85	1.33	1.37	1.18	1.09
B3-8	1.17	1.03	0.78	1.20	1.03

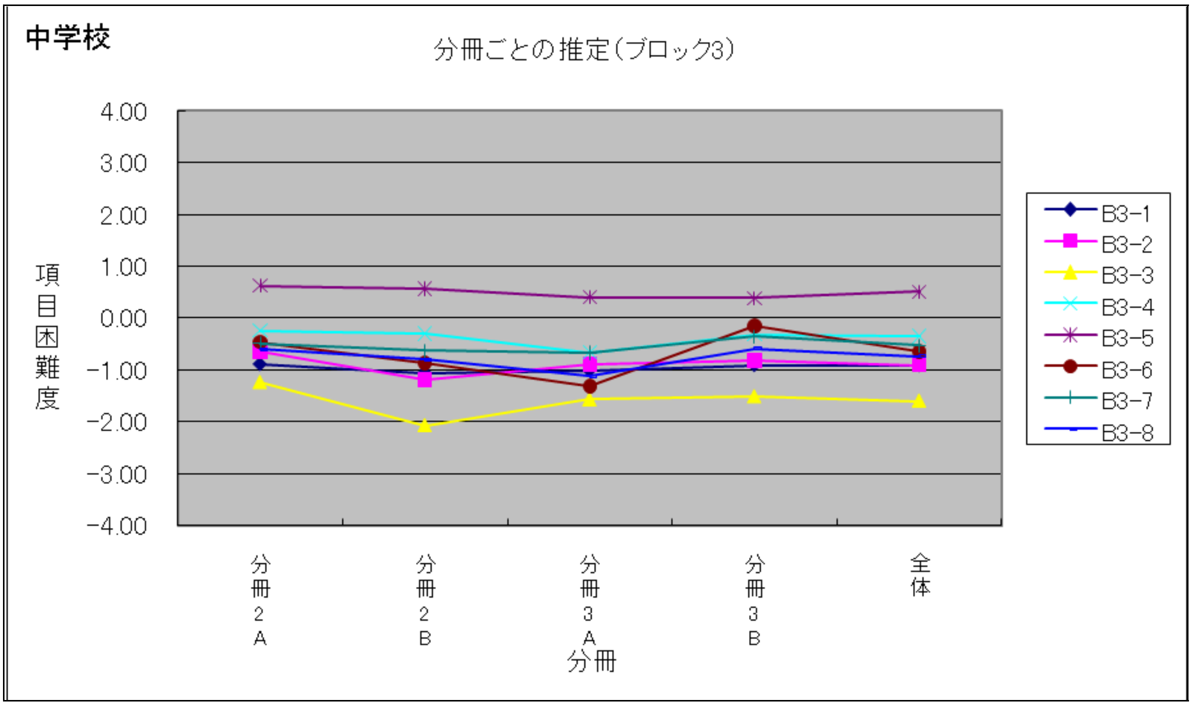

図 8-4　共通ブロックに含まれる項目の分冊ごとの困難度母数の推定値の違い（数学）

表 8-4　共通ブロックに含まれる項目の分冊ごとの困難度母数の推定値（数学）

項目	分冊2A	分冊2B	分冊3A	分冊3B	全体
B3-1	-0.89	-1.06	-1.01	-0.91	-0.92
B3-2	-0.65	-1.19	-0.90	-0.81	-0.91
B3-3	-1.23	-2.07	-1.57	-1.51	-1.60
B3-4	-0.24	-0.30	-0.66	-0.31	-0.35
B3-5	0.63	0.57	0.40	0.39	0.51
B3-6	-0.46	-0.87	-1.31	-0.14	-0.64
B3-7	-0.48	-0.63	-0.67	-0.35	-0.53
B3-8	-0.60	-0.80	-1.13	-0.59	-0.74

8.1.2　項目困難度からの検討

図 8-5 には算数テストと数学テストにおける項目困難度の推定値が示されている。比較しやすいようにほぼ同じスケールで全ての項目を困難度順に並べた。図中左が算数，右が数学のものである。先に述べたように算数の受検者についても数学の受検者についてもその学力特性の母集団はいずれも標準正規分布に従うと仮定しているので，困難度自体の値はほぼ同程度のところに位置している。ここで重要なのは，その分布の仕方が，数学に比べ算数の方が相対的に広い範囲に散らばっていることである。すなわち算数の方が学力特性の広い範囲にわたって相対的に精度のよい測定ができること意味している。調査を設計するにあたってこの性質は重要な意味をもっている。そのことは後述するテスト情報量からの考察のところで再度取り上げて検討する。

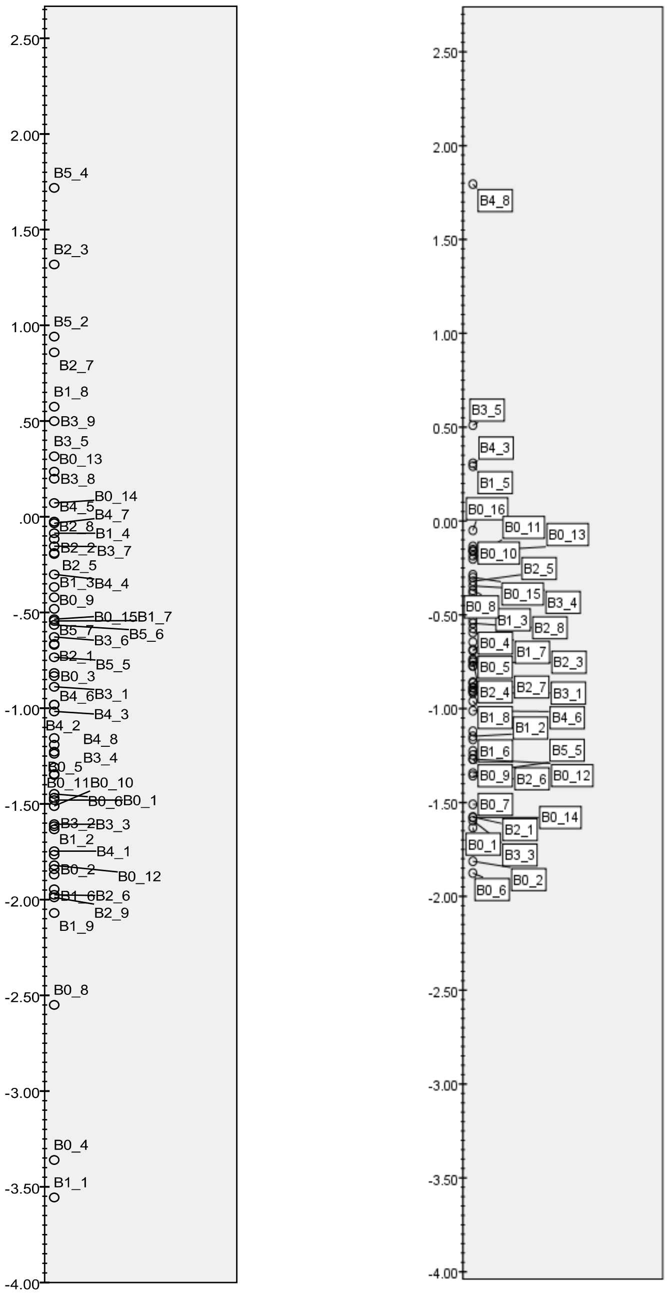

図 8-5　　算数テストと数学テストの項目困難度の推定値の比較

8.1.3　項目識別力からの検討

図 8-6 には算数と数学における項目識別力母数の推定値が示されている。図 8-5 と同様に比較しやすいようにほぼ同じスケールで全ての項目を識別力順にならべた。図中左が算数，右が数学であることも同様である。学力特性についての母集団分布の仮定も同じく標準正規分布である。相対的に数学の方が算数に比べて高めの推定値が得られていることがわかる。算数に比べて数学の方が、学力差が広がることは一般に良く指摘されることであるが，その広がりが推定上は学力特性の分布を両者で同じものに固定した分，項目識別力でとらえられていると見なすことも可能である。これらの結果も次のテスト情報量からの考察の際に改めて取り上げることにする。

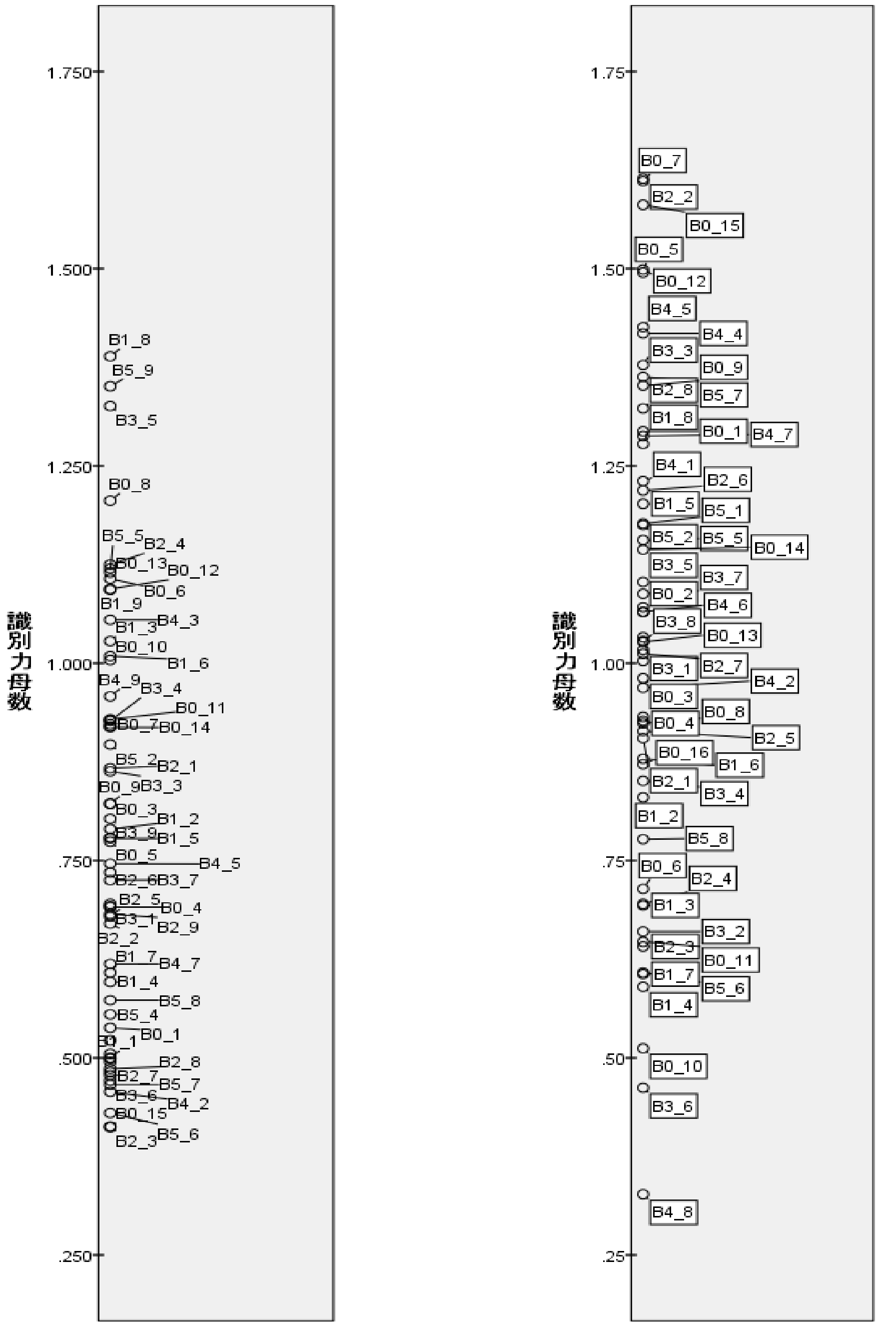

図 8-6　算数テストと数学テストの項目識別力の推定値の比較

8.2 テスト情報量からの検討

数理統計学においてある推定量が一様最小不偏推定量であるかどうかを確かめるために有効な定理としてクラメール・ラオの不等式がある。これを応用して項目反応理論モデルではテストの測定精度を評価するための関数として（3-4）式で定義されるテスト情報関数を用い，ある θ の値に対して定まるこの関数の値をテスト情報量と呼ぶ。さらにクラメールラオの不等式により，その平方根の逆数によって推定量 θ の標準誤差を見積もることが可能となる。

そのことを表したのが，図 8-7 と図 8-8 である。二つのグラフとも全ての項目を使って学力特性値の範囲が－４から＋４に対応するテスト情報量を描いている。実際には各受検者は全ての項目を受検しているわけではないが，ここではあたかも全ての項目から構成されるテストとして扱っている。これは古典的テスト理論の範疇では不可能で有り，項目反応理論でなければできない考察である。

一見して明らかなように二つのテストともピークが θ の尺度上（横軸）では 0 と-1 のあいだにきている。先に述べたようにこれは項目困難度の推定値が低め，したがって，いずれの項目も易しめのものが多いことに対応している。またテストの測定精度の観点からさらに重要なのはこのあたりの学力特性値を一番精度良く測ることができているということである。逆にその近辺から離れるにしたがって精度は極端に落ちていることがわかる。さらにピークの高さは数学テストの方が算数テストに比べて高くなっていることも読み取れる。

同じことを標準誤差の観点から見たのが次頁の図 8-9 と図 8-10 である。やはり０と－１のあいだで推定誤差が最小になっているが，両端にいくにつれてその値は大きくなっていることがわかる。またその大きさも算数に比べて数学の方が大きいことも明らかである。このことは数学では０から-1 近辺の学力特性値の推定精度が高い分，逆に両端で低くなってしまうことを意味しているのである。このことは調査の目的に応じて学力範囲を広く全般的に精度良く測定するのか，それともある範囲に絞ってその部分の測定精度を上げるかの判断が必要なことを意味している。おそらく全国的な学力調査に当てはめて考えれば前者のストラテジーにしたがってテストの設計をした方がその目的に適合した調査が作成できるであろう。

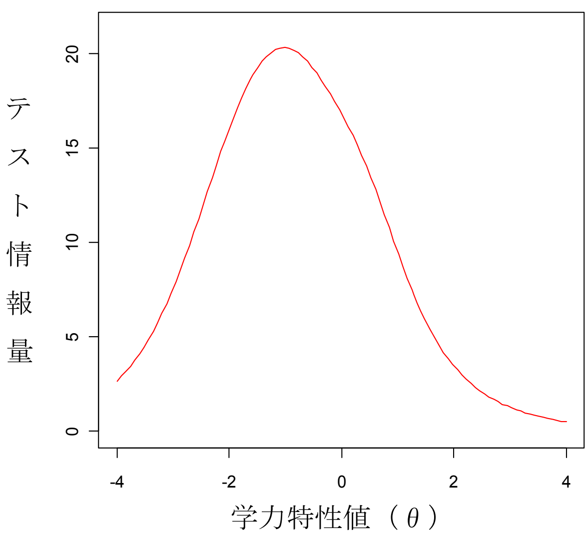

図 8-7　算数のテスト情報量

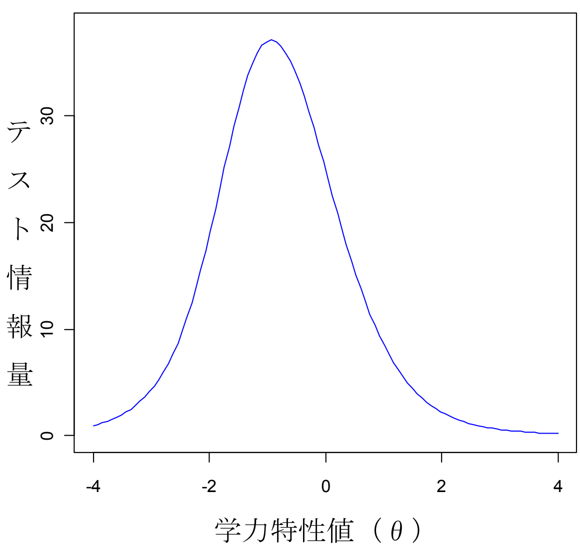

図 8-8　数学のテスト情報量

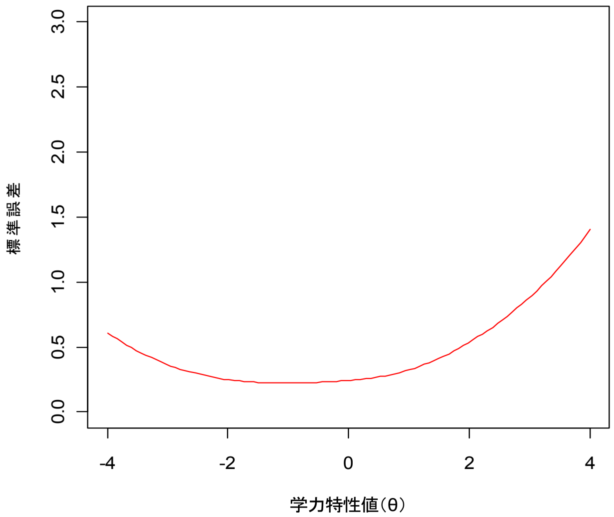

図 8-9　算数の標準誤差

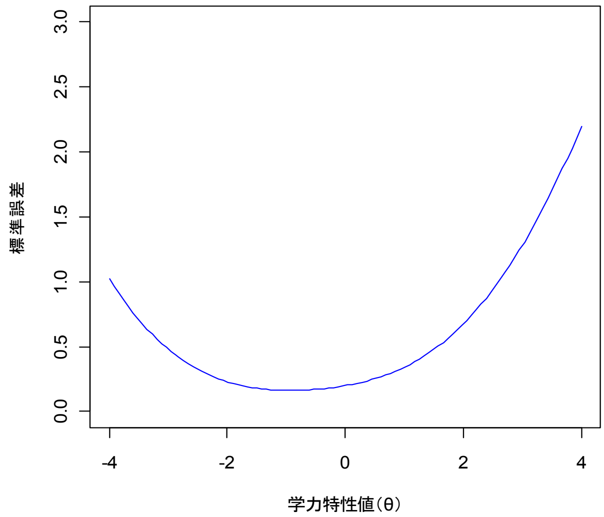

図 8-10　数学の標準誤差

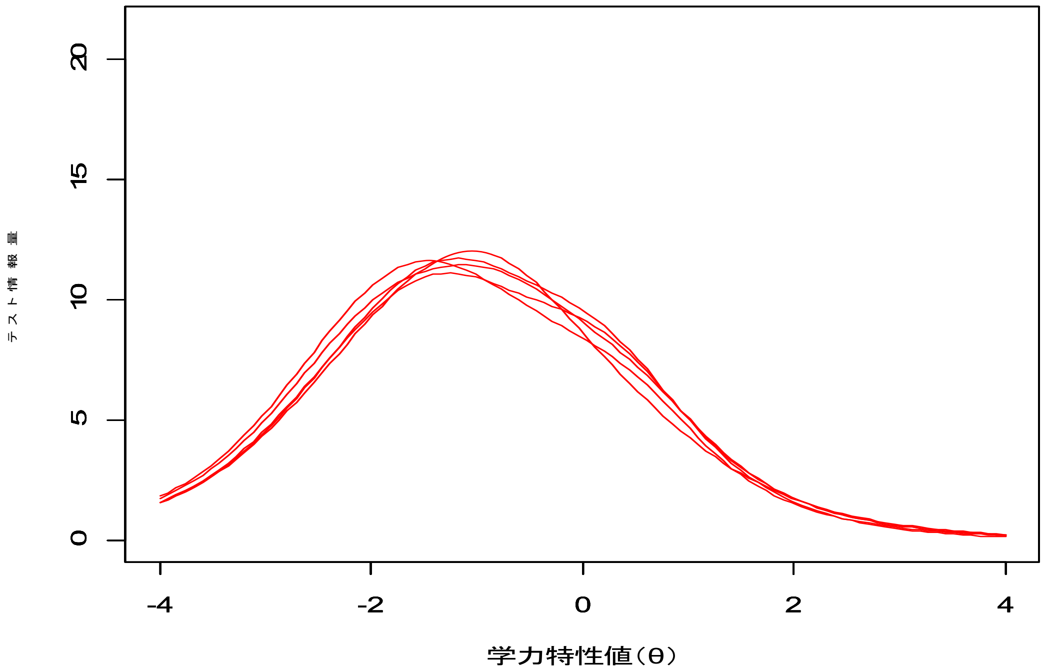

図 8-11　分冊ごとのテスト情報量（算数）

図 8-11，図 8-12 は分冊ごとに描いた算数と数学のテスト情報量，また，図 8-13，図 8-14 は標準誤差である。分冊ごとにみても同様の結論が導けることがわかる。またここで特筆すべきはいずれの分冊もほぼ同様の曲線を描いていることであろう。項目の方から見ても項目サンプリングが意図したとおりうまく機能したことの一つの証左となるであろう。項目数が半分になっている分，図 8-7 や 8-8

とと比べて情報量が落ちていることも図から読み取れる。標準誤差についての議論も同様である。

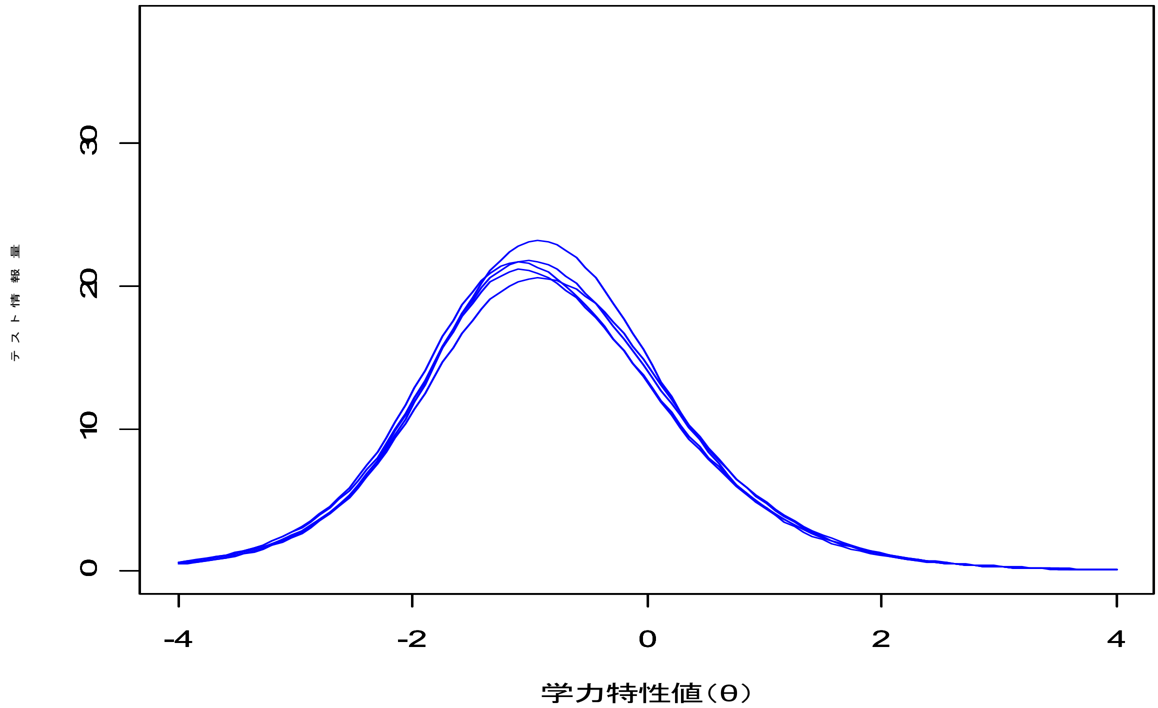

図 8-12　分冊ごとのテスト情報量

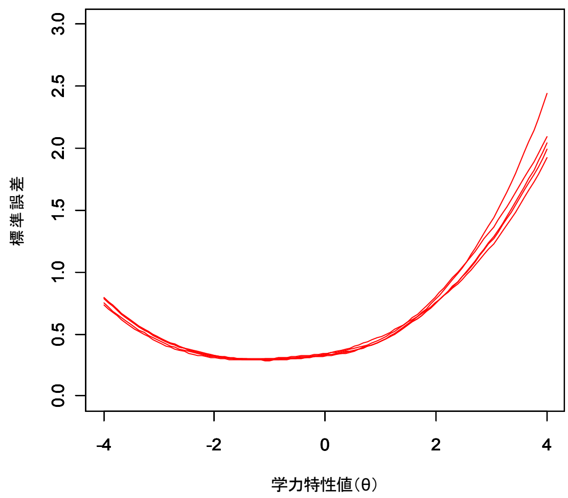

図 8-13　分冊ごとの標準誤差
(算数)

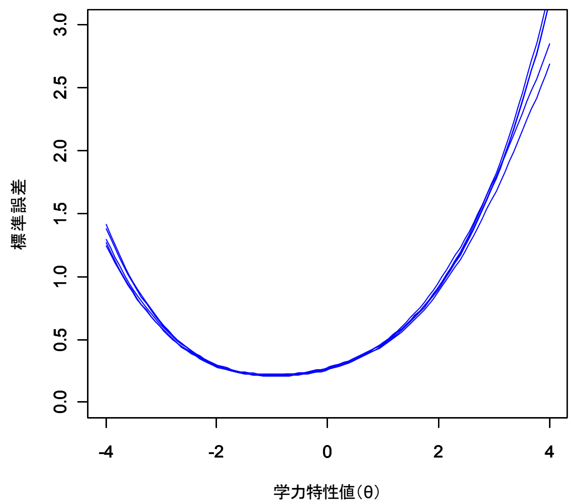

図 8-14　分冊ごとの標準誤差
(数学)

8.3 予測分布からの検討

図 8-15 は算数の分冊 3 における実際に得られた得点分布である。それに対して図 8-16 は項目母数の推定値を使って理論的に求めた予測分布である。予測分布は同じ得点を生み出す全ての(1,0)パターンについて(3-14)式を足しあわせることでもとめている。また θ については標準正規乱数を 3000 個発生させた。実際の得点分布の形状と予測分布の形状はかなりの程度類似していることがわかる。これは他の分冊でも同様である。ここまでの考察からも明らかなように項目の困難度が総じて低いために分布としてもかなり左に裾の長い，ピークが右に偏っているものが得られている。すなわち低い困難度の項目でテストを構成しているために得点分布で言えば 10 点から 20 点あたりの学力，特性値で言えば-1 から 0 あたりの学力を良く識別している一方，いわゆる平均的な学力から上の部分の識別が芳しくない。特に右端の部分の識別はできていないことがここからも伺える。

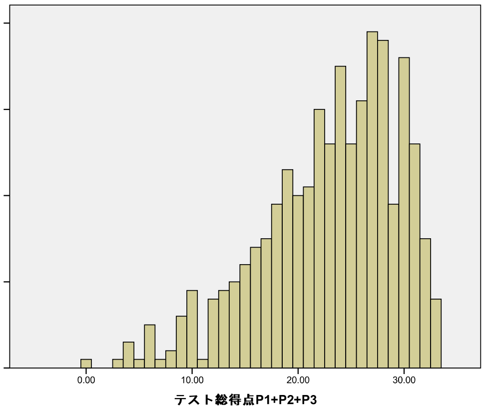

図 8-15　算数の分冊 3 における得点分布

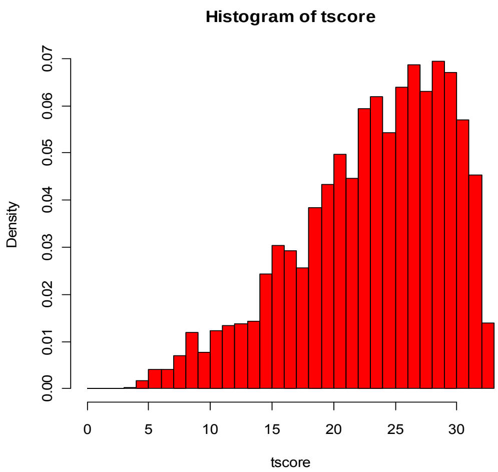

図 8-16　算数の分冊 3 に関する予測分布

この事情はたとえ全ての項目を一人の児童生徒に受けてもらっても同じであることが予想できる。それが図 8-17 である。実際には全ての項目を受けた児童は存在しないが，項目反応理論を利用すればそのような状況もシミュレートが可能である。図 8-17 をみても明らかなようにやはり分布のピークは

右側にきていて学力の高い層の識別が弱いことがわかる。

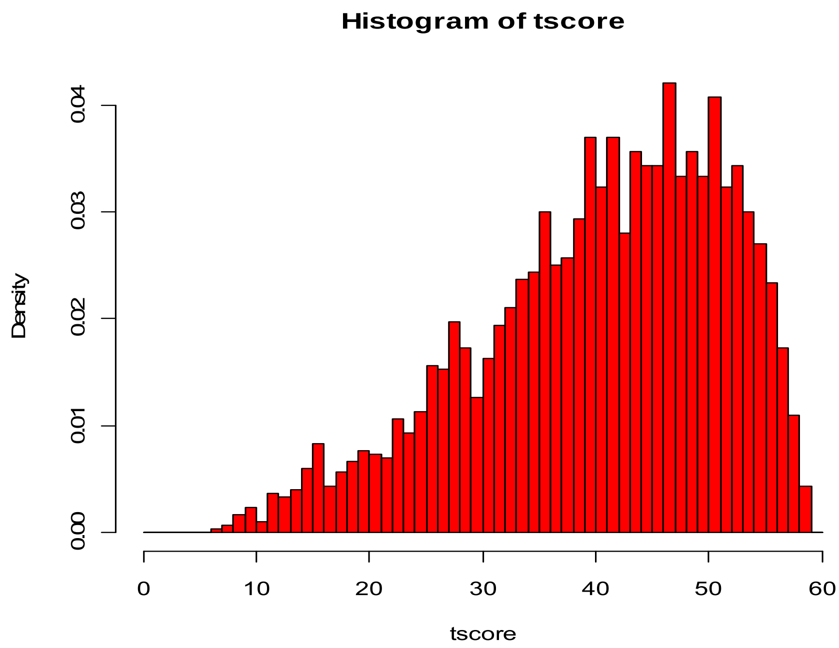

図 8-17　算数の全項目を受検したと想定した場合の得点分布の予測

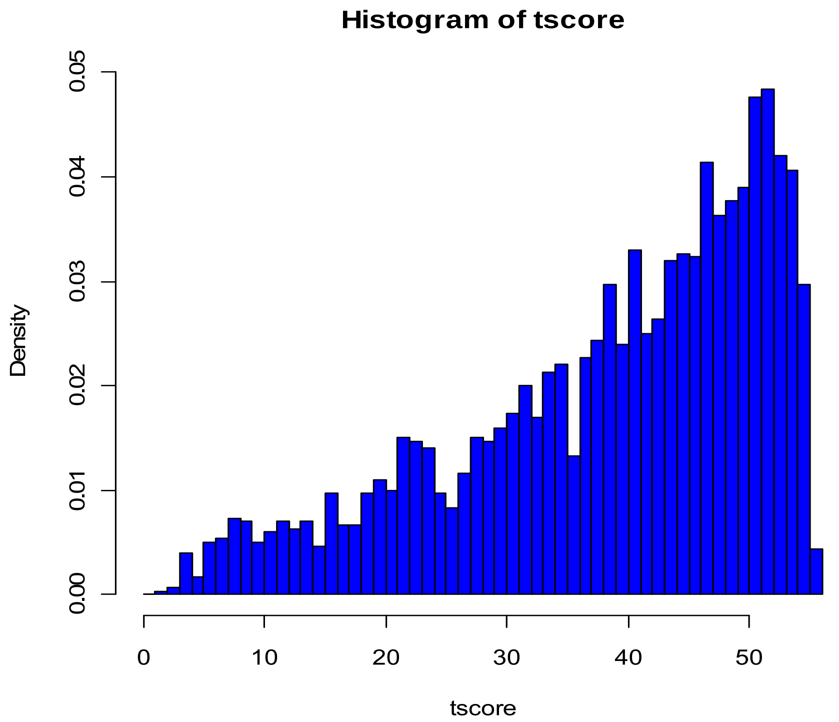

図 8-18　数学の全項目を受検したと想定した場合の得点分布の予測

さらに同様の予測を数学の全項目に関する得点分布について行ったのが図 8-18 である。算数におけるよりも学力の高い層での識別が悪くなっていることがわかる。逆に学力が中以下の層で分布の裾が長くなり識別ができてていることも明らかである。これはテスト情報量の観点から考察したのと同じく，易しい項目によってテストの大半を構成していることが主たる原因である。繰りかえしになるが，どのような目的で調査を実施するのかによって目的とするテストの性質がきまる。そのような検討をおこなう際にも項目反応理論を用いることによって目的に相応しいテストが設計できるのである。

8.4 学力特性値からの検討

ここではＩＲＴ分析の結果得られた学力特性値（θ）の性質について考察する。まず，表 8-5 および表 8-6 には θ と各分冊で計算された得点との相関が示されている。いずれも 0.9 以上の高い値となっていることがわかる。数学に比べて算数の方が全体的にやや高い値となっているのは先に述べたよううに項目困難度の散らばりが算数の方が広いことがおそらく影響しているのであろう。ここで強調すべき点は，分冊ごとの得点は本質的に比較できないが，ＩＲＴモデルを介して得られた θ についてはいずれの分冊からのものであっても同一尺度の上で表現されているため比較可能であるという点である。[2]さらに，相関係数の値からも判断できるように，θ のもつ情報の中には得点のもつ情報自体もほとんどが含まれており，得点を使ってできることは θ を利用することによってもできるのである。

表 8-5　学力特性値と分冊ごとの得点との相関（算数）

		得点				
		分冊1	分冊２	分冊３	分冊４	分冊５
学力推定値	相関係数	0.934	0.95	0.956	0.948	0.953
	N	518	517	510	488	454

表 8-6　学力特性値と分冊ごとの得点との相関（数学）

		得点				
		分冊1	分冊2	分冊3	分冊4	分冊5
学力推定値	相関係数	0.904	0.896	0.945	0.937	0.893
	N	496	491	483	479	445

それでは θ を求めるのに必要な項目数はどの程度であろうか。図 8-19 および図 8-20 には，共通ブロックの項目反応パターンとそれ以外のブロックの項目反応パターンとから得られた二種類の学力特性値のあいだの散布図を描いたものである。満点と零点の受検者を除いている理由については後述する。その結果は，算数においては相関係数にして 0.72，数学においては 0.77 となっている。算数に

[2]なお、今回の場合は特別で、重複テスト分冊法のもつ利点により、受検者集団がいずれの分冊にも等質になるようにランダムに配置されている。その結果、いわゆる等化デザインからいえば、等価集団デザイン(Equivalent-Group Design)となっているため、例えば分冊ごとの得点の平均と標準偏差を調整することで分冊間の得点比較は可能になる。その作業は本調査研究の目的ではないため実施していないが、理論的には実現できる。

おいては共通ブロックに含まれる項目数は 15，それ以外のブロックに含まれる項目数は 9，数学においても同様に前者が 16，後者が 9 である。すなわちこの程度の項目数では θ の推定結果が安定しないことをこれらの結果は示している。ここでも再び，受検者の学力特性値を精度良く求めるための項目数としては 30 項目程度という経験則としての目安が得られるのである。

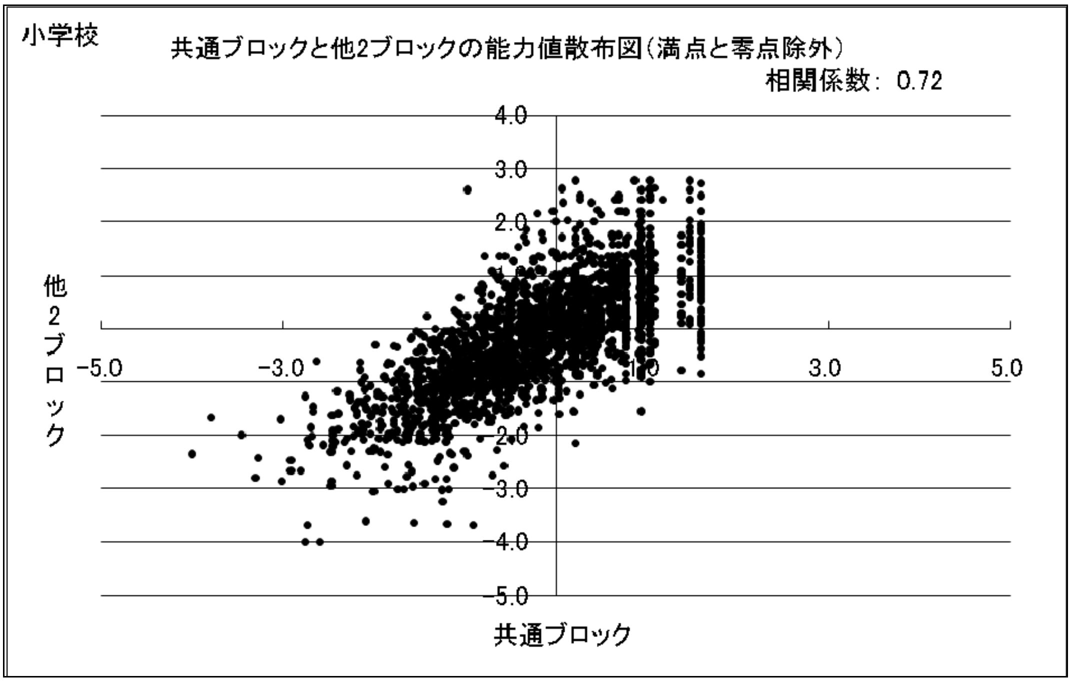

図 8-19　共通ブロックとそれ以外のブロックから得られた θ の相関（算数）

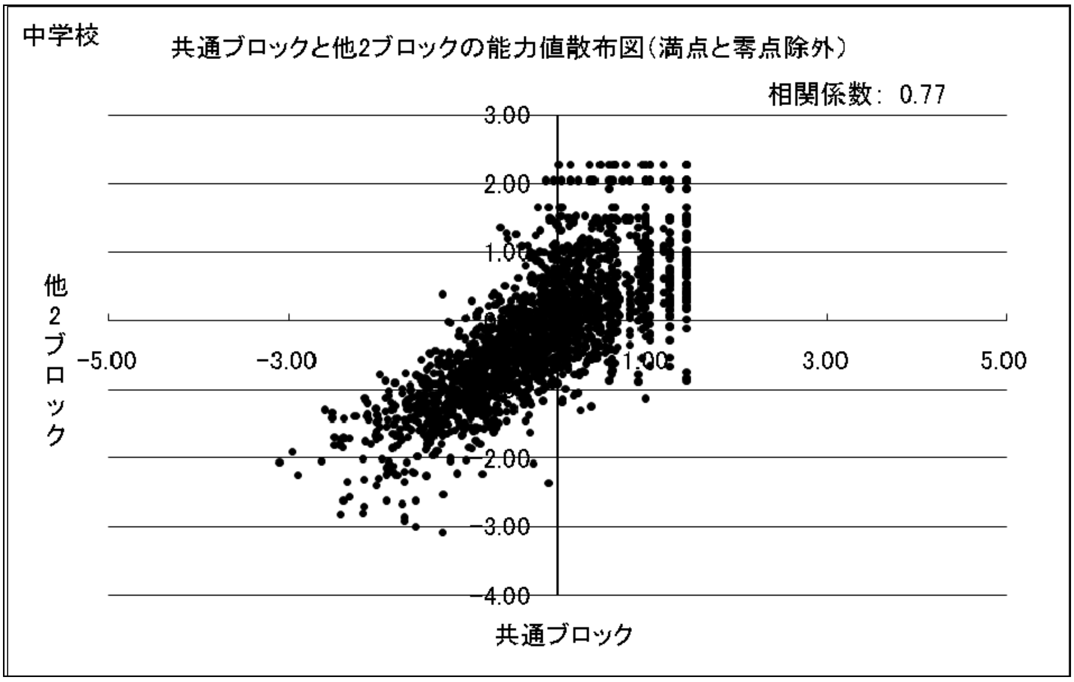

図 8-20　共通ブロックとそれ以外のブロックから得られたθの相関（数学）

次に同じ得点でも分冊が違えば学力特性値が異なることを確認しておこう。図 8-21 は算数の全ての分冊から得られた得点を横軸に，それに対応する学力推定値を縦軸にとった箱ひげ図である。得点の方はいずれの分冊も項目数が等しく 33 項目あるので，その範囲は 0 から 33 の整数値をとる。一方，学力の推定値の方は利用した BILOG-MG のアルゴリズム上-4 から＋４の連続量となっている。この二つの性質が相まって図中のように，ある得点に対しての学力特性値の散らばりが生じる。それぞれの散らばりの中に描かれている細長い長方形の箱の範囲にその得点をとった受検者の半分の人たちの学力特性値があらばっている。さらに箱の中に短い横棒が引かれているが，これはその得点をとった人たちの学力特性値の中央値の位置を示している。0 点から満点まで，中央値の位置は単調増加を示している。

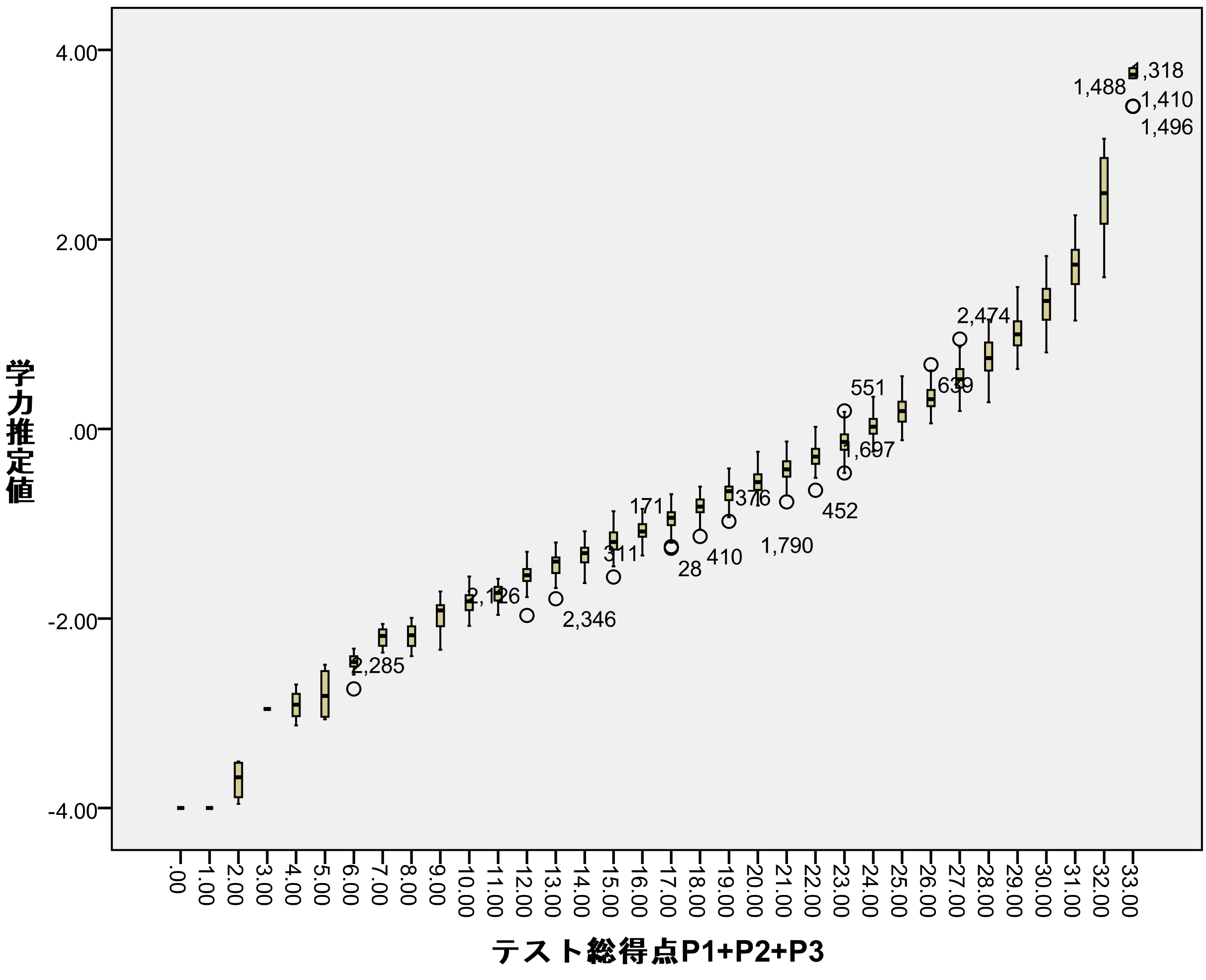

図 8-21　得点と学力推定値の散布図（算数）

しかしながら，隣り合う得点における学力特性値のちらばりを詳細に検討すると，例えば 25 点と 26

点における学力特性値の散らばりが重なっていることがわかる。他の得点同士での比較でも同様の現象が起こっている。これがＩＲＴモデルを介して得られたθの二番目の利点である。そのことを検討するために，いま，算数における得点24，25，26をとった受検者を抜き出し，さらに各得点別に分冊ごとの学力特性値の箱ひげ図を描いたのが，図8-22である。

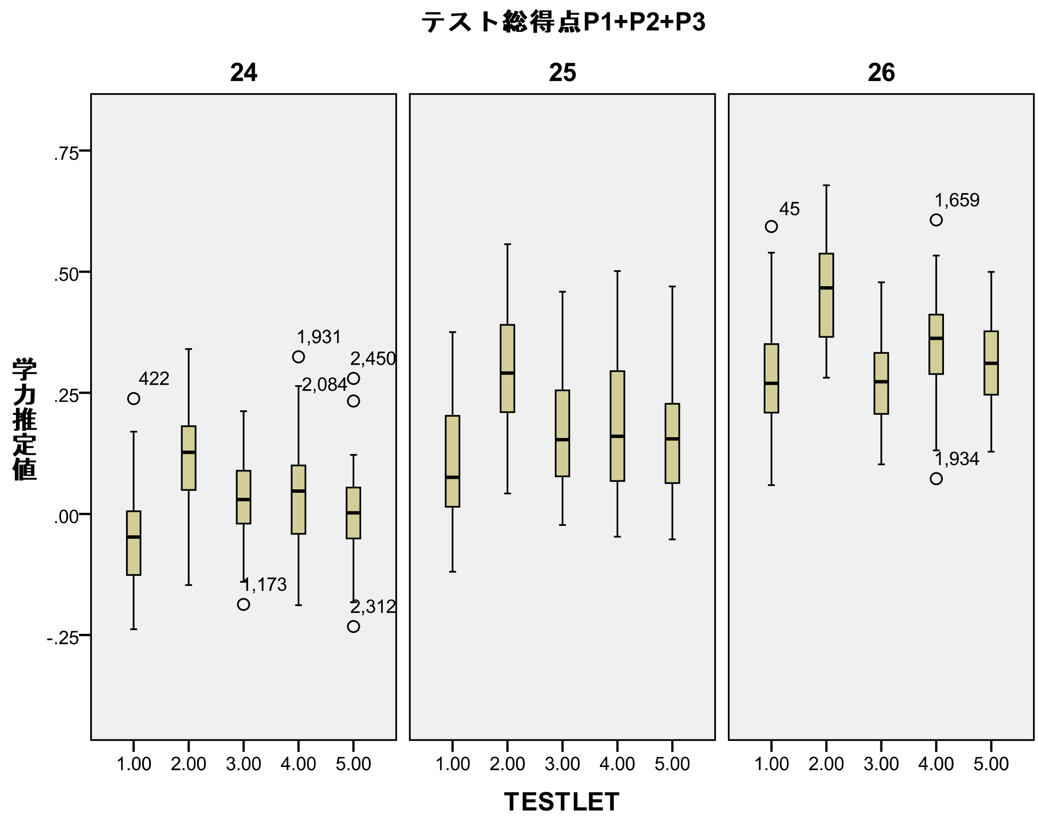

図 8-22　各得点別にプロットした分冊ごとの学力特性値の箱ひげ図（算数）

24 点から 26 点に上昇するにつれて全体の分布も上方向に移動していることがわかる。その一方で得点別に分冊ごとの散らばり具合を見ると例えば分冊２の分布はいずれの得点においても他の分冊の分布よりも上側に位置している。このことは分冊に含まれる項目の構成によってたまたま分冊２のものが平均的にみて若干易しめだったことを示している。このような考察は従来の得点のみの情報からは得られないものである。ここにもＩＲＴモデルを介して θ を求めるメリットがあるといえるであろう。また，数学においてもまったく同じ議論が成り立つので，次頁に二つの図を示すにとどめる。

なお，実用上配慮する必要がある点としては満点及び零点の受検者の学力特性値の扱いである。表 8-7 に示すように，数学の各分冊で満点をとった受検者に対する θ の推定値である。分冊３が 2.510 で最も小さく，分冊５が 3.649 と最も大きい。これは最尤推定法を採用するとその性質上，満点もしくは零点の場合，関数が発散して推定不能となるにもかかわらず，BILOG-MG ではいずれのかの一つの項目への正解に１を与えるのではなく 0.5 を付与することでそれをさけているというアルゴリズム上の工夫が原因で生じている問題である。これは今後実用化に向けて何らかの対応が迫られる問題となろう。

表 8-7　数学における各分冊の満点受検者に対して推定された θ の値

	分冊１			分冊２			分冊３			分冊４			分冊５		
	平均	度数	SD	平均	度数	SD	平均	度数	SD	平均	度数	SD	平均	度数	SD
得点	32	18	.0	32	23	.0	32	13	.0	32	21	.0	32	26	.0
θ	3.647	18	.0	3.626	23	.0	2.510	13	.0	2.351	21	.0	3.649	26	.0

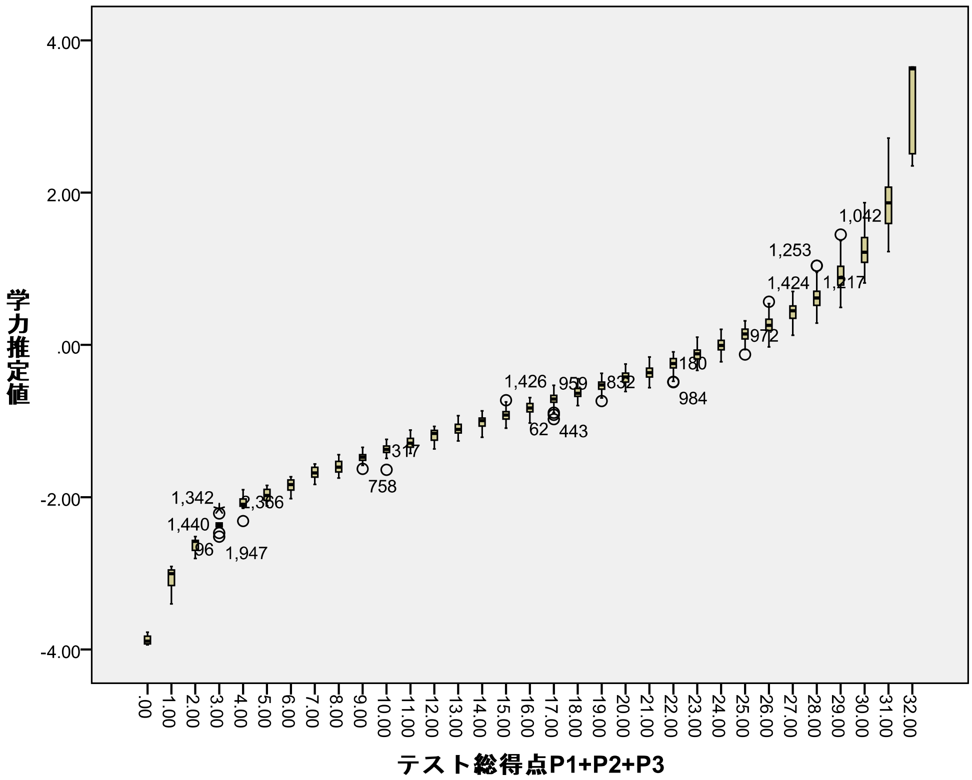

図 8-23　得点と学力推定値の散布図（数学）

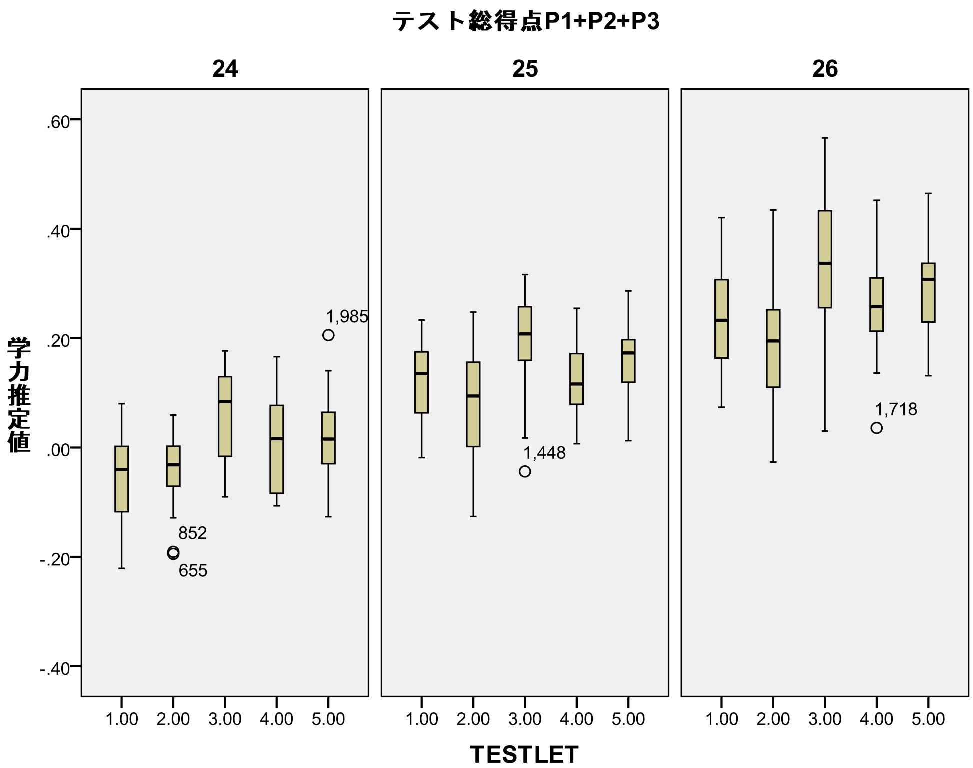

図 8-24　各得点別にプロットした分冊ごとの学力特性値の箱ひげ図（数学）

9. 試作したデータベースの仕様

以下に今回の調査研究で得られたデータをＤＢ化する際に設定した仕様を掲載する。今後の研究調査に必要な情報としては項目に関するものと受検者に関するものとの大別して 2 種類のＤＢが必要となる。また，分冊がどのブロックから構成されているかの対応情報も必須である。なお，まだ，用語に多少の混乱がのこっているが，それぞれのＤＢで補足する。

9.1 項目ＤＢ

用語としては SEQ は通し番号，項目名の a,b はそれぞれ項目識別力母数，項目困難度母数の推定値，ＳＥＴは分冊を意味する。

項目	備考
SEQ	
SEQ_sub	小中別SEQ
教科	1:算数 2:数学
年度	
回次	1:新潟
学種	1:小 2:中
学年	
ブロック	
ブロック№.	
a	
b	
平均正答率	
平均期待正答率	
1a	SET内正答率
1b	SET内正答率
2a	SET内正答率
2b	SET内正答率
3a	SET内正答率
3b	SET内正答率
4a	SET内正答率
4b	SET内正答率
5a	SET内正答率
5b	SET内正答率
学年1	
領域1	
単元1	
内容1	無データあり
学年2	無データあり
領域2	無データあり
単元2	無データあり
内容2	無データあり

9.2 受検者ＤＢ

用語としては項目の「抽出」が「平成 22 年度全国学力・学習状況調査」の抽出校となったものか，いわゆる希望調査校であるかの区別をするものである。全受検者の所属校に割り当てられる。

項目	備考
SEQ	
SEQ_sub	小中別SEQ
教科	1:算数 2:数学
年度	
回次	1:新潟
学種	1:小 2:中
学年	
学校№	
抽出	1:抽出2:その他
クラス	無データあり
性別	1:男 2:女
SET	
θ	
総得点	
P1	パート1得点
P2	パート2得点
P3	パート3得点

9.3　分冊仕様

各分冊に割り当てられたブロックの位置情報。例えば，分冊 3b はブロック４が前，共通ブロック（ブロック 0）が中央，ブロック３が後ろに位置することを示す。

	P1	P2	P3
1a	1	0	2
1b	2	0	1
2a	2	0	3
2b	3	0	2
3a	3	0	4
3b	4	0	3
4a	4	0	5
4b	5	0	4
5a	5	0	1
5b	1	0	5

10. 教師質問紙の分析

実施校の担当教員におこなったアンケート結果である。選択式の質問と自由記述の部分から構成されている。

10.1 選択式質問

回答比率が 50 パーセントを超えたものには網掛けをした。問 4 での統計用語に関する質問から現場において教育測定学の初歩知識，例えばテストの信頼性など，の不足が見受けられる。今後，エビデンスベースドな教育を進めて行かなければならない時代にあって，教師教育カリキュラムなどの改善が必要と考えさせられる結果である。

問1 今回の調査では、10種類の問題冊子を児童生徒一人ひとりに割り当てて行う方式を取ることで、全体で学習指導要領の広範囲の問題を出題できるように設計されています。こういった調査方式について、どのようにお考えですか？ あてはまるものすべ

	小学校	中学校
(1) 幅広い範囲についての学力・学習状況がつかめるので有益な方法と思う。	32%	28%
(2) 事前に新しい調査方式に関する十分な説明がなされれば実施可能と思う。	34%	43%
(3) 児童生徒がそれぞれ異なる問題を解いているので、結果を返すときの指導に工夫が必要と思う。	74%	69%
(4) 調査用紙の配布方法など従前の方法に比べ多少煩雑さを感じる。	21%	28%
(5) その他	6%	7%

問2 学力調査の結果を平素の学級の指導に役立てるために、必要と思われる情報は次のうちどれですか？ あてはまるものすべてに〇をつけてください。

	小学校	中学校
(1) 平均点（学校全体及び学級別）	57%	45%
(2) 大問別、観点別、内容別の正答率（学校全体及び学級別）	88%	83%
(3) 児童生徒一人ひとりの正誤状況	84%	64%
(4) 学校全体としての経年変化	32%	48%
(5) 児童生徒個人ごとの経年変化	30%	29%
(6)その他	1%	3%

問3 学力調査の方法として、国際調査（例：PISA、TIMSS）のように全体傾向を測るためにあえて個人には成績結果をフィードバックしない方法もあります。このような調査方法について、どのようにお考えですか？ あてはまるものに〇をつけてください。

	小学校	中学校
(1) 目的によってはあってもよいと考える。	40%	38%
(2) 児童生徒指導の観点からは、個人への結果のフィードバックは必ず必要である。	15%	34%
(3) あってもよいと考えるが、参考としては個人の成績表は必要である。	43%	26%
(4) 児童生徒一人ひとりが異なる問題を解いているのであれば、個人成績表は必要ない。	13%	17%
(5)その他	1%	5%

問4 上記のような国際調査の報告では以下のような統計用語が使われています。このうち目にされたことのある用語がございましたら〇をつけてください。

	小学校	中学校
(1) 平均	98%	95%
(2) 分散	70%	74%
(3) 標準偏差	93%	93%
(4) 中央値	60%	72%
(5) 相関係数	48%	52%
(6) 識別力	13%	12%
(7) 通過率	88%	78%
(8) 信頼性係数	11%	7%
(9) 妥当性	49%	41%
(10) 換算点	12%	43%
(11) パーセンタイル順位	9%	10%
(12) 信頼区間	15%	5%
(13) 測定の標準誤差	21%	26%
(14) 回帰式	5%	16%
(15) 等化	2%	12%

10.2　自由回答

（問1）

小学校

：調査が多すぎる。

：有益かどうか、よくわからない。

：配布した種類を行う児童の学力に偏りがある　母数が小さければ10種類が人数は均等でも能力は違う

：初めての方式なので分からない

：効果に疑問を感じる

中学校

：普段の授業日に行うのは45分という長さではむずかしい。

：1つの問題に対する母集団が小さくなるため、データの有効性に疑問を感じる

：検査で突然授業ができなくなるのは困る。

：学組全体の苦手部分の把握という点では疑問が残る．問題回収の意図は？

（問2）

小学校

：6年生の10月では遅い．授業を進めるだけで手一杯．

中学校

：全国や県・市・区の平均点

：母集団の大問別、観点別、内容別の正答率

（問3）

小学校

：放置されて終わる．現場はそれどころではない．やる事が多すぎる．

中学校

：今の学力調査であれ必ずフィードバックをされてきた生徒なので、欲しがるとは思う。
移行するのであれば特に必要はないが、この生徒たちが大人になってから生かされるのであれば、先をみすえた分析が必要だと思う。

：⑵の時期が、検査実施日から間があきすぎると、フィードバックの効果が期待できない。

：フィードバックしないのであれば、やらない方が良い

参考文献

Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *The Journal of the Royal Statistical Society, series B, 34*(1), 42-54.

Baker, F. B., & Kim, S. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York: Marcel Dekker.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 443-459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35*(2), 179-197.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*(1), 1-38.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*(2), 139-164.

池田　央（1994）. 現代テスト理論. 朝倉書店.

Kendall, M. G., & Stuart, A. (1979). *The advanced theory of statistics* (4th ed., Vol. 2). New York: Oxford University Press.

熊谷龍一（2009）. 初学者向けの項目反応理論分析プログラム EasyEstimation シリーズの開発. 日本テスト学会誌, 5(1), 107-118.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica, 16*(1), 1-32.

Olson,U 1979 Maximum likelihood estimation of the polychoric correlation coefficient, Psychometrika 44, 4

大友賢二（1996）. 項目応答理論入門. 大修館書店.

Reckase, M., D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207-320.

芝　祐順（編）（1991）. 項目反応理論―基礎と応用―. 東京大学出版会.

豊田秀樹（2002）. 項目反応理論[入門編]―テストと測定の科学―. 朝倉書店.

豊田秀樹（2005）. 項目反応理論[理論編]―テストの数理―. 朝倉書店.

van der Linden, W. J., & Hambleton, R. K. (Eds.) (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Zimowski, M., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *BILOG-MG: Multiple-group BILOG.* Chicago, IL: Scientific Software International, Inc.

用語・表記一覧

【限界】

・選択枝情報、段階反応、パフォーマンスアセスメントの際の段階分けへ等への対応はしていない。
・学年集団の区別はしていない
・推算値の対応はしていない
・サンプリングの際とは独立な表記システム（一部重なってはいる）。

【注意】
・添え字等の複雑な数式表記はLaTeX表記を援用している。適宜読み替えのこと。

【用語】

受検者:	受験者・被験者は使わない。
項目:	受検者が解く問題の最小単位。「項目」と「問題」の書き分けに注意。
ブロック:	分冊の中の項目群の分割単位。 異なるブロック間に共通の項目はない（排他的）。
分冊:	受検冊子・冊子は使わない。一般名称としての冊子はOK。 異なる分冊に共通に含まれるブロックが存在する。 したがって異なる分冊には共通に含まれる項目が存在する。
共通項目	複数の分冊に含まれている項目のこと。
共通ブロック:	複数の分冊に含まれているブロックのこと。
パート:	全分冊を通してブロックが定置される位置情報。 （本研究においては1=前,2=中,3=後）
学力:	用語として「能力」は使わない・一般総称としては心理学的特性。 「学力の推定値」というような使い方をして、なるべく「学力」とは言い切らない。 IRTモデルにもとづいているので学術的にもこの表現が正確。
学力推定値:	一般名称としては尺度値。θと表記する。
調査結果シート	参加校へ返却した個別結果票のこと。
実施時期:	経年変化を見る際の年度を識別する情報。
集団:	都道府県別などの集団を識別する情報。

【表記法】

基本方針

一般的に使われている表記を優先させる
（たとえば当て推量母数のcなどはそちらの定義を優先する）
なるべくワードで操作しやすい表記にする。

顕在変数はローマ数字
潜在変数はギリシャ文字

定義

全受検者数:	半角小文字の n
全項目数:	半角小文字の m
全ブロック数:	半角小文字の H（P や p はIRTモデルや確率で使用する）
全分冊数:	半角大文字の K（Lは対数尤度のときに使う可能性あり）
全パート数:	半角大文字の O（オー:orderの略：本研究ではO=3で固定）
全実施時期数:	半角小文字の Y（Yearの略：T や t は真値で使う）
集団数:	半角小文字の G（Groupの略：サンプリングの記法とは別体系）

受検者:	半角小文字のi	i=1,…,n
項目:	半角小文字のj	j=1,…,m （通し番号の場合の表記）
ブロック:	半角小文字のh	h=1,…,H
分冊:	半角小文字のk,l(エル)	k,l=1,…,K
共通ブロック:	半角大文字のC	ただし分冊全部に共通する場合のみ
	半角小文字でkl(エル)	ただし分冊kと分冊lに共通する場合
パート:	半角小文字のo（オー）	o=1 or 2 or 3
実施時期:	半角小文字のy	y=10,…,Y（10は2010年度）
集団:	半角小文字のg	g=1,…G

項目反応:	総称および行列は大文字のX ベクトルおよび要素は小文字のx
平均:	¥bar{X},¥bar{x}など
標準偏差:	行列なら半角大文字のS、変数のものなら半角小文字のs

学力の推定値:	総称およびベクトルはθ、添字でベクトル、要素を区別する。
平均:	μ
標準偏差:	行列ならΣ、変数ならσ

項目識別力:	総称およびベクトルはa、添字でベクトル、要素を区別する。
項目困難度:	総称およびベクトルはb、添字でベクトル、要素を区別する。

用例

（注）推定値を意味する¥hatはワードでは煩雑になるので使用しない。

テストX
反応パターン行列X
データ行列X

2011年度のテスト、データ、データ行列	X^{(11)}
都道府県gのデータ	X^{(g)}
都道府県gの2011年度のデータ	X^{(g:11)}
2011年度の学力の平均、標準偏差、分散	¥mu^{(11)}, ¥sigma_{(11)}, ¥sigma_{(11)}^{2}

受検者iの項目jの正誤	x_{ij}=1 or 0
受検者iの項目正答数	x_{i} = 分冊での正答数
受検者iのブロックhから計算された項目正答数	x_{ih}
受検者iの共通項目おける項目正答数	x_{iC}
受検者iの共通項目以外の項目正答数	x_{i[C]}

受検者iの学力の推定値（分冊全部の項目による）	¥theta{i}
受検者iのブロックhから計算された学力の推定値	¥theta_{ih}
受検者iの共通ブロックおける学力の推定値	¥theta_{iC}
受検者iの共通ブロック以外の項目による学力の推定値	¥theta_{i[C]}

分冊kの受検者数	n_{k}
ブロックhの受検者数	n_{h}
分冊kの項目数	m_{k}
ブロックhの項目数	m_{h}

分冊kのデータ	X^{(k)}
分冊kのデータから推定された項目識別力	a^{(k)} , a_{j}^{(k)}
分冊kのデータから推定された項目困難度	b^{(k)} , b_{j}^{(k)}

項目母数はブロックh、共通ブロックCについても同様