

6. テストの信頼性および項目分析

ここでは古典的テスト理論の観点から、各分冊のテストとしての信頼性の検討ならびに統計学的にみた項目の良否についての分析である項目分析を行う。これらの分析は古典的テスト理論の枠組みで発展してきた方法のため、分冊ごとに行う必要がある。

6.1 算数データの分析

6.1.1 テストの信頼性

表 6-1 に各分冊ごとに求めたテストの信頼性係数の推定値であるクロンバックの α を示した。いずれの分冊も項目数としては 33、受検者数としては 230 名から 260 名程度であり、 α 係数としては安定した値が計算できるサイズである。その値は 0.864 から 0.897 を示しており、いずれの分冊においても十分な精度の測定結果が得られていると判断できる。ブロックの前後を入れ替えたことによる信頼性係数への影響も特に見いだすことはできない。

表 6-1 分冊ごとの信頼性係数の推定値（ブロックの順序の別あり）

分冊	1a	2a	3a	4a	5a
項目数	33	33	33	33	33
受検者数	260	260	258	248	234
信頼性係数	0.871	0.864	0.880	0.887	0.897
分冊	1b	2b	3b	4b	5b
項目数	33	33	33	33	33
受検者数	258	257	252	240	220
信頼性係数	0.865	0.880	0.883	0.877	0.896

推定はクロンバックの α による。以下同様

また、表 6-2 はブロックの入れ替えを無視した場合の分冊ごとの信頼性係数の推定値である。表 6-1 と同様の結論が得られる。このことから、実際に重複テスト分冊法を適用する際にはブロックの順序効果に対する配慮は特に必要ないといえるであろう。

表 6-2 分冊ごとの信頼性係数の推定値（ブロックの順序の別なし）

分冊	1	2	3	4	5
項目数	33	33	33	33	33
受検者数	518	517	510	488	454
信頼性係数	0.868	0.872	0.882	0.882	0.896

なお、参考までに各ブロックをひとつのテストとみなして信頼性係数を求めると表 6-3 のようになる。項目数が 15 の共通ブロックでも 0.771 など、明らかに分冊ごとに求めた信頼性係数の値よりも小さい。このことからわかるようにテストの信頼性は受検者数ではなくて項目数にかなりの程度依存している。信頼性係数と項目数との関係はスピアマン・ブラウンの公式（一般式）として知られているものである。これらのことから実際の調査にあたっては測定精度を担保するためには、ブロックご

とに含まれる項目数は10程度以下であっても、分冊自体には30個程度の項目数が含まれるようにデザインを組む必要があることが重要なノウハウとして指摘できる。

表 6-3 ブロックごとの信頼性形数の推定値

ブロック	共通	1	2	3	4	5
項目数	15	9	9	9	9	9
受検者数	2487	972	1035	1027	998	942
信頼性係数	0.771	0.692	0.638	0.712	0.696	0.686

さらに分冊ごとに因子分析を行い固有値をもとめたところ、いずれも第1因子の固有値の値が他の因子のものに対して突出して高く、因子分析モデルから判断しても、各分冊の一次元性は確保できていると考えられる。図 6-1 は分冊 1 に関するスクリープロットを示す。他の分冊においてもほとんど同様の結果が得られている。

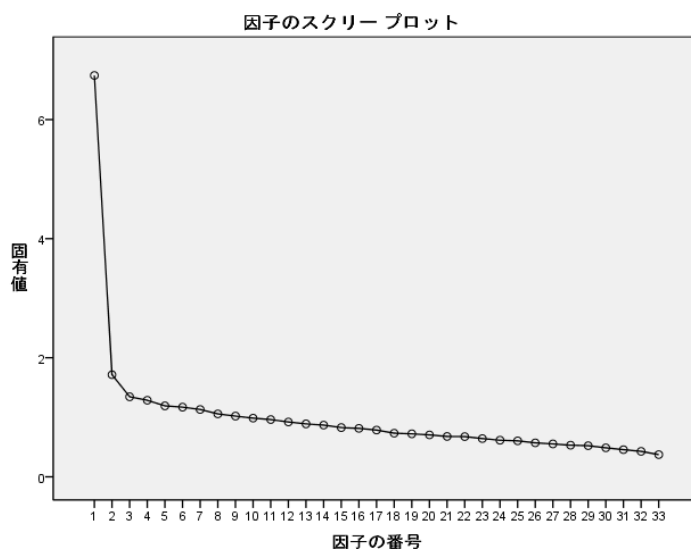


図 6-1 算数データにおける分冊 1 のスクリープロット

6.1.2 項目分析

テストの信頼性の検討によって各分冊ごとにテスト全体としては望ましい精度の測定が実現できていることは担保された。次に考察すべきはそのテストを構成している個々の項目がそれぞれ期待通りの機能を果たしているかどうかの確認である。検討すべき統計量が莫大になるのでこの考察に必要な数値は全て資料編を参照されたい。まず、資料 1 には算数データで使ったテストの各項目のブロック番号、変数名、学年、領域、単元、内容の対照表が示されている。次に資料 2 では各項目の受検者数、正答率、標準偏差、点双列相関係数が記載されている。このうち正答率はその項目を受けた全受検者中何割が正答したかの比率であるので、この値が高いほどその項目は易しかったことになる。また標準偏差はその項目の散らばりを表している。ただし、もとのデータが(1,0)データの場合、正答率を p 、誤答率を $q(=1-p)$ とすれば、その標準偏差は pq の平方根で表現できる。また $p=0.5$ のときが標準偏差が最大になることも簡単な計算で導ける。その様子を示したのが図 6-2 である。このような性質から個人の学力などを識別する上では極端に高い正答率(易しすぎる)や極端に低い正答率(難し過ぎる)

の問題はのぞましくないとされることもある。

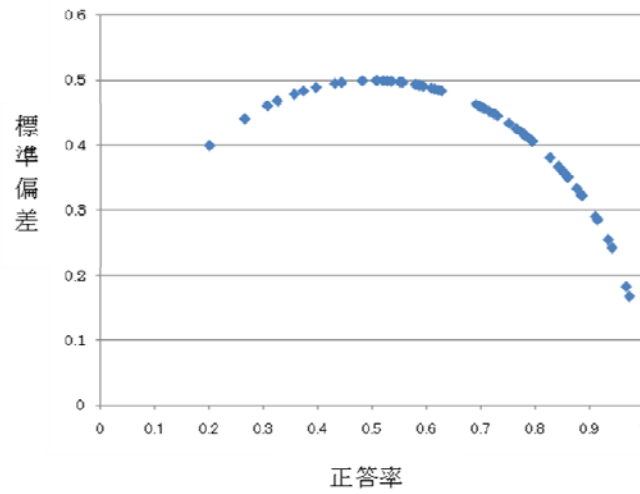


図 6-2 二値データにおける正答率と標準偏差の関係

次に点双列相関係数であるが、これはその項目とその項目が含まれているテストの得点との相関を計算したものである。その意味で項目-テスト得点(Item-Test)間相関、あるいは IT 相関と呼ばれることもある。この指標はいわばその項目がそのテストで測定している特性をどの程度とらえているかを見るものであり、識別力とよばれる統計量のひとつでもある。なお、点双列相関を計算する際のテスト得点の中に、当該項目の正誤情報を含む場合と、それを除外する場合があり、前者の方がその値は高くなる。本研究では後者の得点を利用している。この点双列相関係数と上で求めた正答率との散布図を描いたものが図 6-3 である。

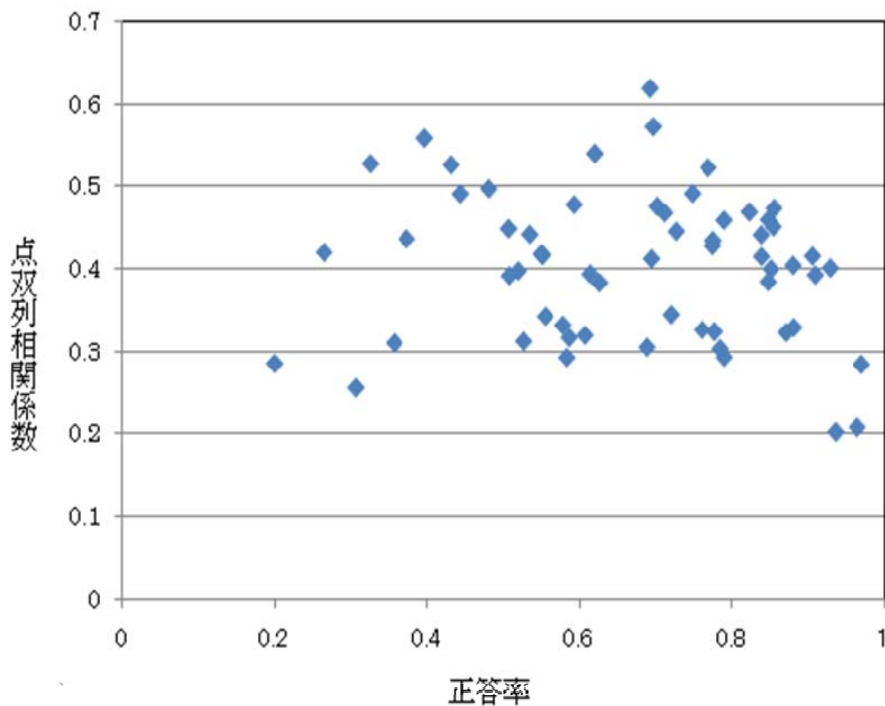


図 6-3 正答率と点双列相関係数の散布図

一般的に点双列相関係数の値は0.2以上が望ましいとされることが多いが、その基準で見ると算数の場合、全ての項目で0.2を超えており、全体的にみて良い問題が作成できたといつてよいであろう。ただ、あえて検討すべき項目をあげれば、点双列相関係数の値が、0.208, 0.202であった項目B0-4と項目B1-1の2つがある。これらの項目は正答率でも0.9を超えておりやさしすぎる問題であった可能性が高い。もっとも、正答率が0.930であっても点双列相関係数の値が0.401となっているB1-9のような項目も存在するため、このような統計的な側面からだけではなく教科教育の知見もあわせた両方からの検討が必要なことは言を俟たない。

さらにブロックの位置による正答率の差をみたのが図6-4である。共通ブロックに含まれる項目もあわせて、分冊Aにおける正答率から分冊Bにおける正答率を引くことによって差を計算し、それを縦軸にとり、項目には通し番号をふってそれを横軸にとった。図を見る限り特に分冊Aに対して位置的には後ろにあることによる系統的な正答率の差の存在はない。実際、共通ブロックのみ、あるいは非共通ブロックのみで、それぞれ、正答率の平均差の検定(対応のあるt検定)をおこなっても統計的有意差($p=0.01$)は見いだせなかった(共通ブロックにおいては自由度73, p 値0.696, 非共通ブロックにおいては自由度88, p 値0.929)。このことは信頼性係数の推定のところで得られたのと同様、実際に重複テスト分冊法を適用する際にはブロックの順序の困難度への影響に関する配慮は特に必要ないといえるであろう。

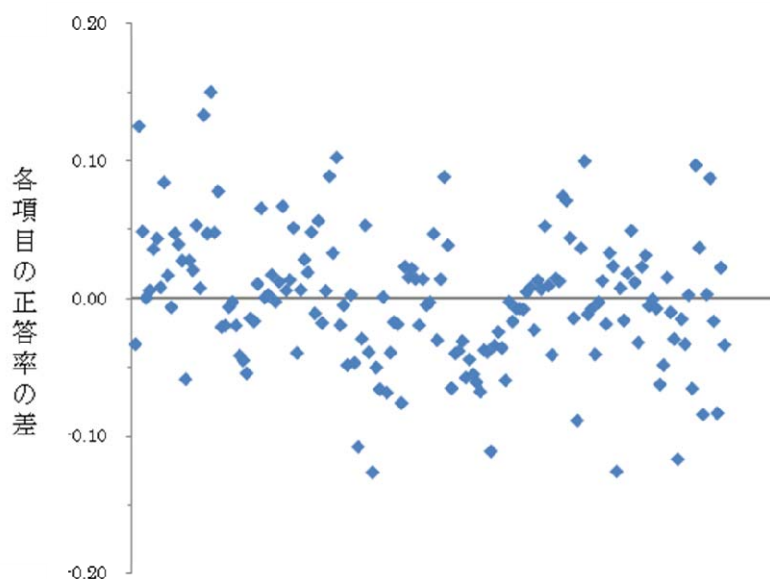


図6-4 各項目の正答率の分冊Aと分冊Bにおける差

次に点双列相関係数と因子負荷との関係から項目の性質を検討しておく。資料5に掲載されている因子負荷の値は、先に述べたように分冊ごとのデータにもとづき因子分析の結果得られたものである。因子分析モデルとしては主因子モデルを使っている。データ行列の一次元性が強い(矩形行列における最大特異値が他の特異値に比して著しく大きい)場合には点双列相関係数と因子負荷はほぼ同じ情報をもつ。このデータの場合も図6-5に示すように、全ての分冊に関してもとめた両者の値をプロットすると直線上にならぶ。ここで分冊1における項目B0-4が外れ値となっていることには注意してお

く必要がある。おそらく正答率が 0.969 と高く個人差の識別ができていないことが最大の原因と考えられるが、ここでも教科教育の知見をあわせて考察する必要がある。

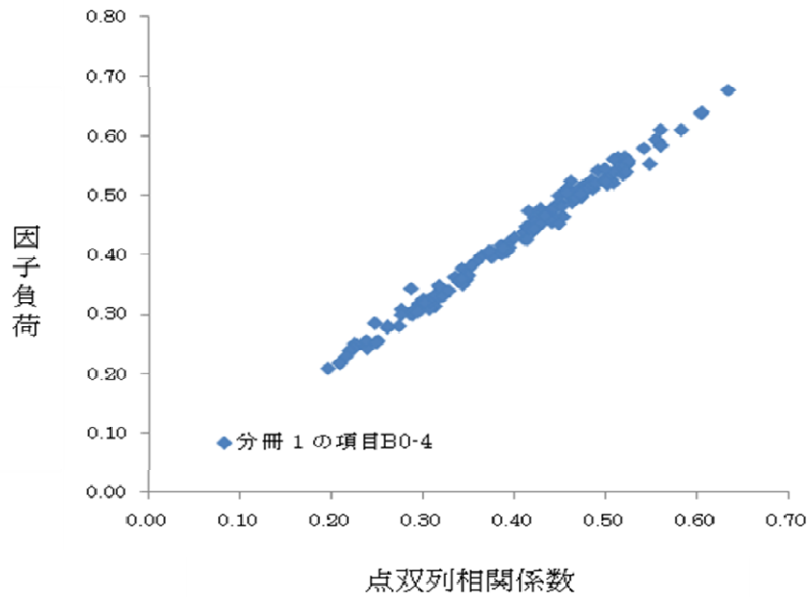


図 6-5 点双列相関係数と因子負荷量の散布図

上のような考察をさらに詳細に行うものが資料 6 に示す項目分析である。項目分析に使われる図の見方について B0-3 を例にとって説明する。各項目の上段には図 6-6 に示すような回答パターンごとの正答率を表したヒストグラムが描かれている。この例の場合正答以外に誤答の主要なパターンが 7 種類、無答、およびそれ以外の回答をあわせたものの計 10 種類のパターンに分類してヒストグラムが描かれている。項目によっては誤答のパターンがこれよりも増減するため合計が 10 種類とは限らないことには注意が必要である。

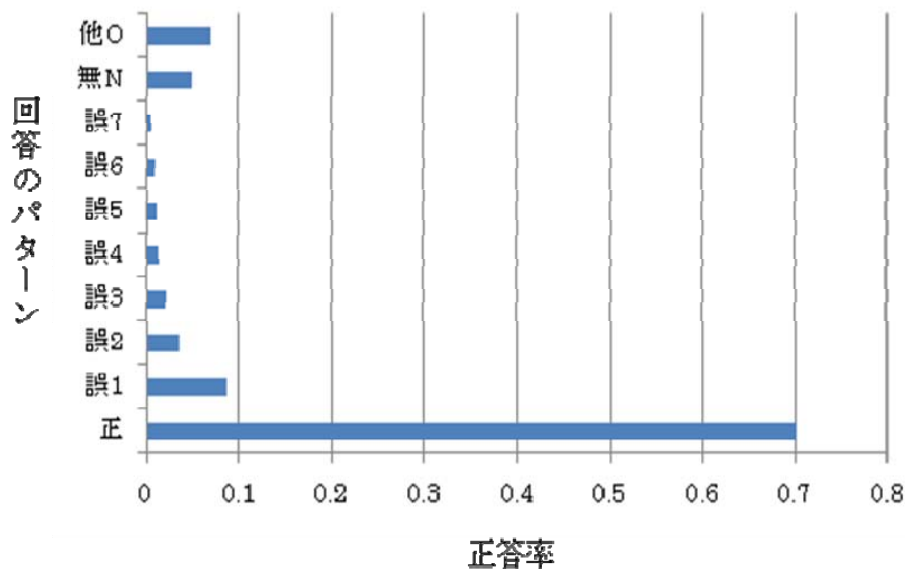


図 6-6 項目 B0-3 における回答パターンごとの正答率

次に中段には受検者の学力レベルを5段階に分類して段階ごとにそれぞれの回答パターンがどのくらいの比率でいるかを示したG P分析図(Good-Poor Analysis)が描かれている。学力レベルが上がるにつれて正答率も上昇していることがわかる。逆に誤答した比率は学力レベルの上昇とともに下がっていることもわかる。なお、これらのG P分析図における点双列相関係数は各項目と後述する学力特性値(θ)との相関をもとめているので、上で求めた点双列相関係数の値とは異なっている。

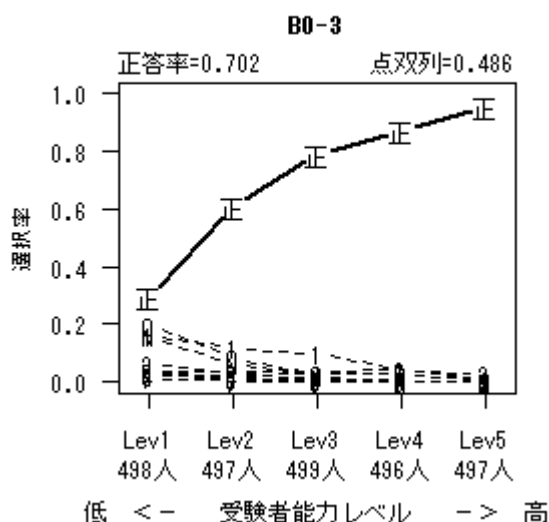


図 6-7 GP 分析図

さらに下段には表 6-4 に示すような上記のグラフを描くための正答率の数値データが埋め込まれている。したがって各項目の良否を検討する上で本質的な情報をわかりやすくとらえているのはG P分析図ということになる。

表 6-4 回答パターンごとの全体および学力レベルごとの正答率

解答	全体	Lev1	Lev2	Lev3	Lev4	Lev5
正	0.776	0.432	0.714	0.828	0.933	0.974
誤1	0.100	0.227	0.147	0.082	0.028	0.014
誤2	0.094	0.239	0.123	0.072	0.032	0.006
誤3	0.010	0.024	0.008	0.010	0.004	0.006
誤4	0.008	0.028	0.006	0.004	0.000	0.000
誤5	0.002	0.010	0.000	0.000	0.000	0.000
誤6	0.001	0.004	0.000	0.000	0.000	0.000
無N	0.005	0.024	0.002	0.000	0.000	0.000
他O	0.004	0.012	0.000	0.004	0.002	0.000

例えば項目 B0-8 は正答率が 0.971 と極めて易しい、したがって多くの児童が正答している項目であるが、そのG P分析図は図 6-8 に示すようになる。当然のことながら学力が低い層の識別に役立つ項目ということになる。ただし図に示すレベル2以上の学力層ではほとんど全員が正答しているためこれらの層での学力の識別にはあまり機能していない項目であることがわかる。

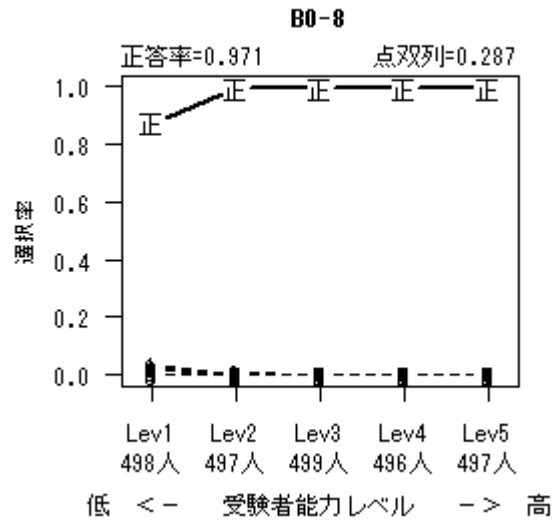


図 6-8 正答率が高い項目のG P分析図の例

また、項目 B5-4 のG P分析図を描くと図 6-9 のようになる。正答の比率は確かに学力層が上がるにつれて上昇しているが、その一方で、誤答のパターン 1 の間違いをした児童の割合もレベル 4 まで同様に上昇していることに加え、その割合はレベル 5 においても正答できた児童の割合よりも高い。このような項目の場合、学力があってもなんらかの要因で誤答してしまう可能性が否定できない。この項目の場合は問題の内容よりもむしろ問題を読む際にいわゆるケアレスミス誘発しやすい文字構成になっていたと考えることができる。

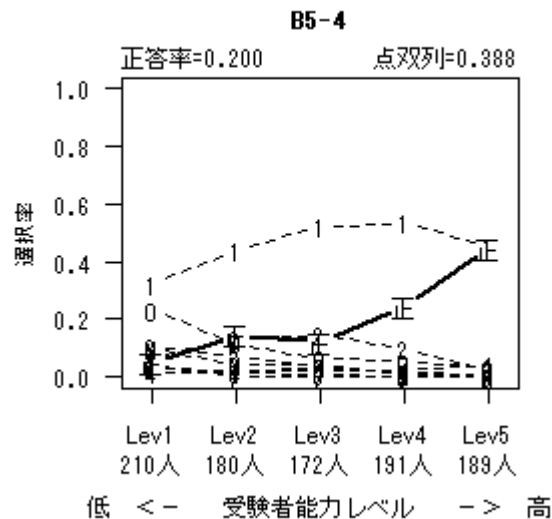


図 6-9 ケアレスミスを誘発しやすい項目のG P分析図の例

一方、第 5 章で考察したように、正答率が比較的低くても点双列相関係数が高いため指導の参考になる項目の例として、項目 B3-8 「数量の関係を表す式」のG P分析図は下図のようになる。これまで見てきたG P分析図の各レベルにおける正答率を結んだ曲線の傾きが急であることがその特徴として指

摘できる。すなわち測定したい数学の学力を鋭敏にとらえている項目と判断できるのである。

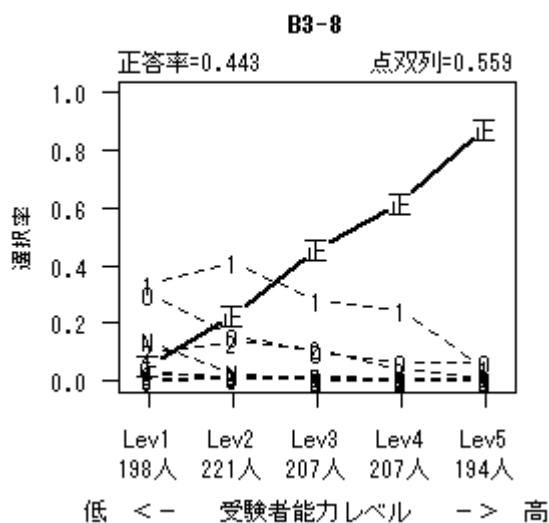


図 6-10 点双列相関の高い項目のG P分析図の例

算数データにおけるその他の項目のG P分析図から判断して著しく不適切なものはない。しかしながら比較的易しい項目が多いため、項目数の割にテスト全体としての測定精度が高くなく、できればもう少し難易度の高い項目に入れ替えるなどして、精度を上げる工夫が実際の調査では必要となるであろう。

6.2 数学データの分析

6.2.1 テストの信頼性

表 6-5 に数学データに関して各分冊ごとに求めたテストの信頼性係数の推定値であるクロンバックの α を示した。いずれの分冊も項目数としては 32、受検者数としては 220 名から 250 名程度であり、 α 係数としては安定した値が計算できるサイズである。その値は 0.898 から 0.927 を示しており、いずれの分冊においても十分な精度の測定結果が得られていると判断できる。経験的に算数のテストに対して数学のテストはこのように高い信頼性が得られことが多い。その理由としてはいくつか考えられるが、受検者の学力分布の散布度が学年の上昇にしたがって広がっていくことなどが直接的な要因と指摘できる。算数データと同様ブロックの前後を入れ替えたことによる信頼性係数への影響も特に見いだすことはできない。

表 6-5 分冊ごとの信頼性係数の推定値（ブロックの順序の別あり）

分冊	1a	2a	3a	4a	5a
項目数	32	32	32	32	32
受検者数	247	246	244	242	226
信頼性係数	0.916	0.921	0.927	0.925	0.898
分冊	1b	2b	3b	4b	5b
項目数	32	32	32	32	32
受検者数	249	245	239	237	219
信頼性係数	0.915	0.923	0.919	0.918	0.930

また、表 6-6 はブロックの入れ替えを無視した場合の分冊ごとの信頼性係数の推定値である。表 6-3 と同様の結論が得られる。このことから、算数においてと同様、数学においても実際に重複テスト分冊法を適用する際にはブロックの順序効果に対する配慮は特に必要ないといえるであろう。

表 6-6 分冊ごとの信頼性係数の推定値（ブロックの順序の別なし）

分冊	1	2	3	4	5
項目数	32	32	32	32	32
受検者数	496	491	483	479	445
信頼性係数	0.917	0.922	0.923	0.922	0.916

ブロックごとの信頼性に関しても表 6-7 に示すとおりである。数学は算数にくらべて相対的に信頼性の高いテストを作成しやすいといわれているが、それでも最低 20 項目ぐらひは必要といえるであろう。調べるべき領域の広さを考えるとたとえ重複テスト分冊法を採用したとしても、各分冊あたり算数と同様 30 項目程度は必要と見込まれる。

表 6-7 ブロックごとの信頼性係数の推定値

ブロック	共通	1	2	3	4	5
項目数	16	8	8	8	8	8
受検者数	2394	941	987	974	962	924
信頼性係数	0.852	0.721	0.763	0.728	0.766	0.778

さらに分冊ごとに因子分析を行い固有値をもとめたところ、算数データと同じように、いずれも第 1 因子の固有値の値が他の因子のものに対して突出して高く、因子分析モデルから判断しても、各分冊の一次元性は確保できていると考えられる。図 6-11 は数学の分冊 1 に関するスクリープロットを示す。他の分冊においてもほとんど同様の結果が得られている。

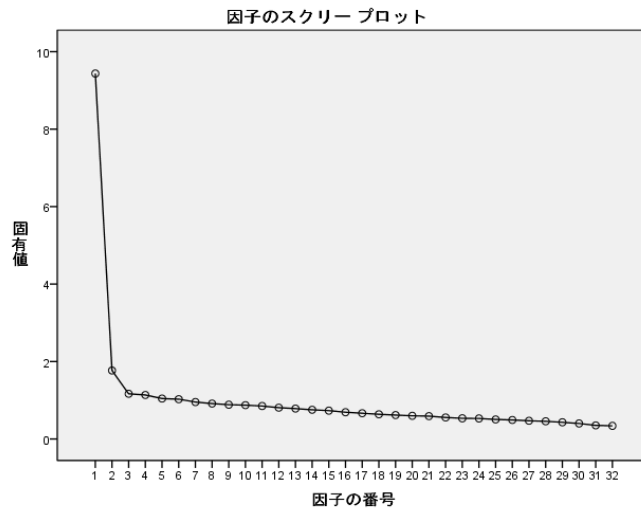


図 6-11 数学データにおける分冊 1 のスクリープロット

図 6-12 には正答率と点双列相関係数の散布図が描かれている。いずれの指標から見てもほとんどの項目において十分良好な結果がえられており、先にみた分冊ごとの高い信頼性係数の値を生み出すことにむすびついていことがわかる。ただし、項目 B4-8 については点双列相関係数の値が 0.2 をしたまわっており検討が必要である。この項目についてはすでに第 5 章でも指摘されているものである。

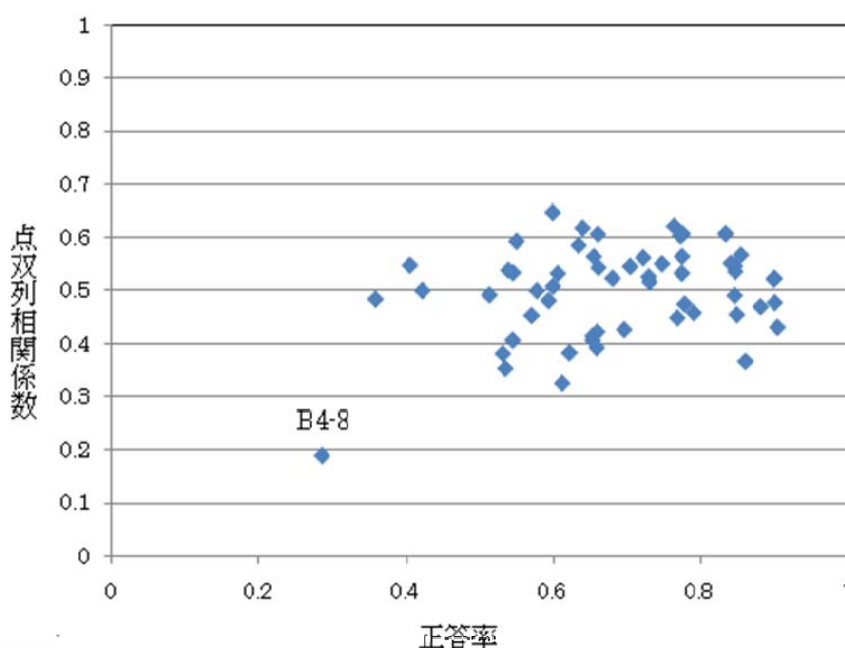


図 6-12 正答率と点双列相関係数の散布図

次にブロックの位置による正答率への影響を調べるために、共通ブロックと非共通ブロックのそれぞれについて、算数と同様、対応のある t 検定をおこなった結果、1%水準で統計的有意差は見いだせなかった。(共通ブロックにおいては自由度 78, p 値 0.040, 非共通ブロックにおいては自由度 78, p 値 0.047)。しかし p 値の大きさが 0.05 以下であるため慎重を期するために、各項目の正答率の差を図示した検討をおこなった。

まず図 6-13 には共通ブロックにおける各項目の分冊 A での正答率から分冊 B での正答率を減じて差をとりそれを縦軸に、通し番号を振ってそれ横軸にとったものである。分冊 A での正答率がプラスの項目が多いことが伺える。共通ブロックであるので位置の効果はないはずにもかかわらずこの現象が見いだせることには注意が必要である。同様に非共通ブロックの項目についてもその差をプロットしたものが図 6-14 に示されている。図 6-13 に比べて分冊 A の場合の方が正答率が高い傾向が読み取れる。現時点ではこのような傾向が読み取れる原因が特定できないため、今後さらに詳細な分析が必要となろう。

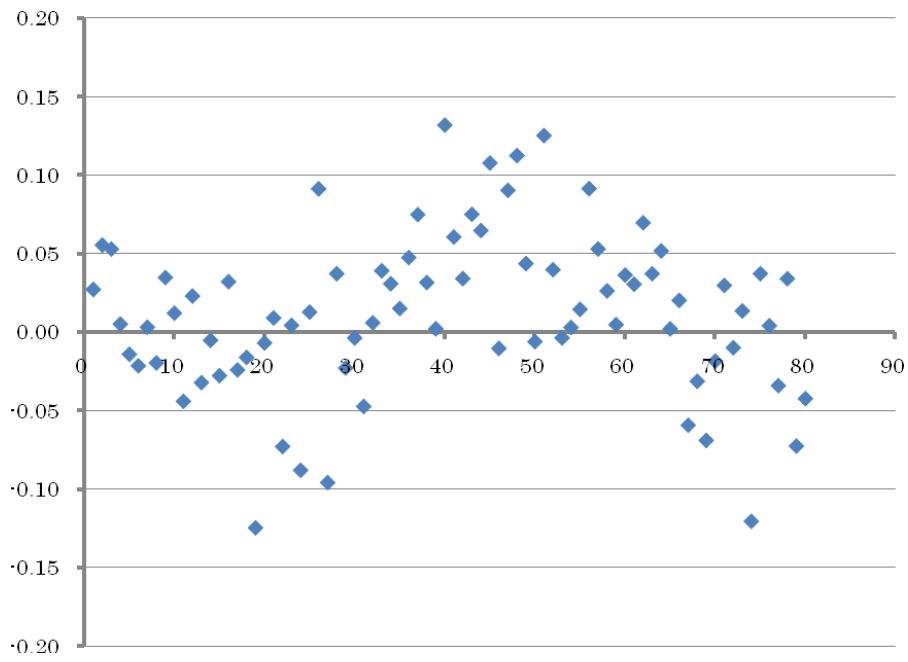


図 6-13 各項目の正答率の分冊Aと分冊Bにおける差（共通ブロック）

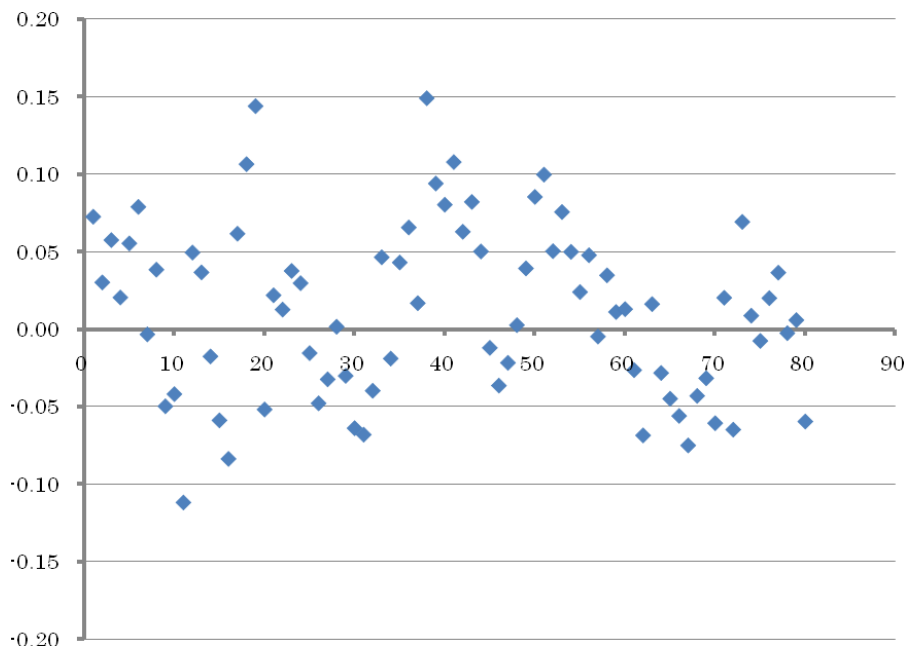


図 6-14 各項目の正答率の分冊Aと分冊Bにおける差（非共通ブロック）

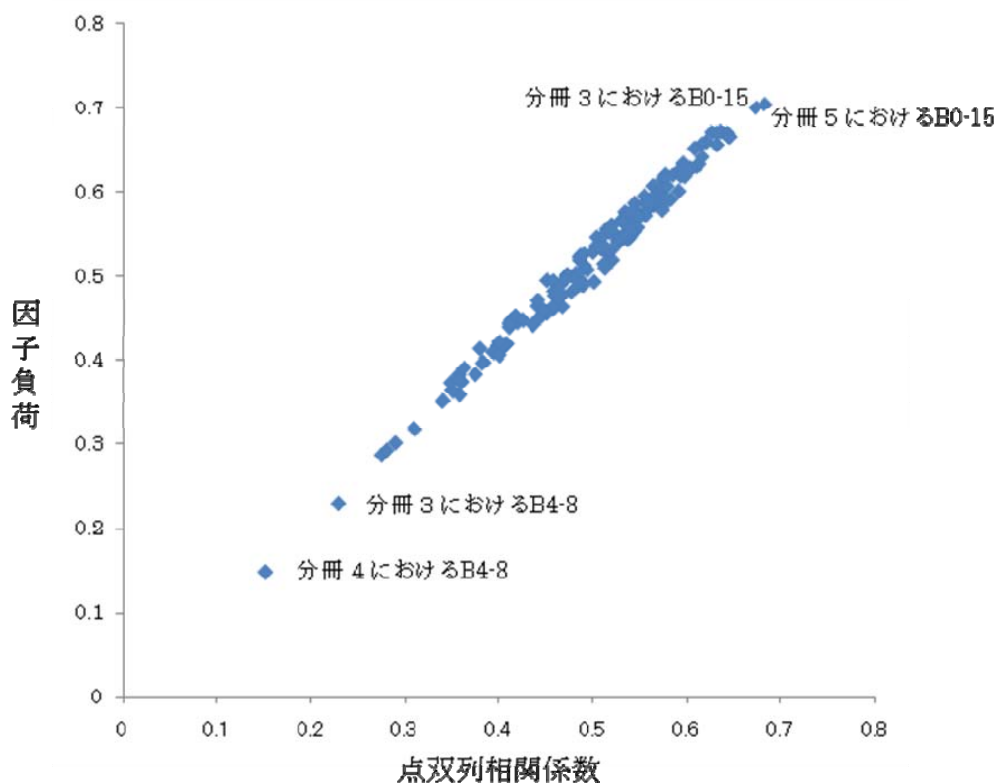


図 6-15 点双列相関係数と因子負荷量の散布図

次に点双列相関係数と因子負荷との関係から項目の性質を検討しておく。資料 11 に掲載されている因子負荷の値は、算数と同じく分冊ごとのデータにもとづき主因子解因子分析の結果得られたものである。このデータの場合も図 6-15 に示すように、全ての分冊に関してもとめた両者の値をプロットすると直線上にならぶ。ここで分冊 3 および分冊 4 における項目 B4-8 がここでも外れ値となっていることがわかる。逆に点双列相関係数の点からも因子負荷の点からも最大の値をしめしたのが共通ブロックにおける項目 B0-15 である。問題としては知識を問うだけの非常にシンプルな内容であるにも関わらずこのような高い値を示すこのようなケースの場合、問うている概念がかなりその領域において本質的なものであることが多い。統計的な指標を参照することでこうした考察も可能となる。

6.2.2 項目分析

さらに資料 12 の項目分析の結果を利用した項目の良否に関する考察であるが、項目 B4-8 を除いてはほとんどの項目において極めて良好な結果が得られている。そのため、上で単純ではあるが良い問題と判断できた項目 B0-15 と比較するかたちで、考察する。図 6-15 には項目 B0-15 の G P 分析図が掲載されている。学力レベルが上がるにつれて正答率も上昇し、その傾きも急である。さらにレベル 2 とレベル 3 のあいだでその差が一番大きいことから、この概念の理解の程度がこのふたの層で著しく異なると判断できる。こうした知見は実際の学習指導上に活かせるであろう。

一方、項目 B4-8 の G P 分析図は図 6-17 に示すとおりである。学力レベルに応じて正答の割合も上昇しているが、詳細に見ると一番学力の低いレベル 1 とその次のレベル 2 で正答率がそれぞれ 0.207, 0.159 と逆転していることがわかる。さらに誤答 1 のパターンが学力レベルとともに本来なら下から

なければならないのにも関わらず、逆に上昇していることもこの項目の特徴である。おそらく問題文中の「点の全体」という言葉に誘導されて、それが直線になるところまで理解できなかった生徒が多いことがその理由であろう。これも教科教育の観点からさらに考察すれば指導上有益な情報が得られるものと考えることができる。なお問題としては「二元一次方程式 $3x-y=2$ に対応するグラフを選びなさい」と修正すればG P分析図としては改良できるであろう。もっとも「点の集合」がある条件下で直線になることを理解させるのが目的ならこの修正では不十分であることはいうまでもない。

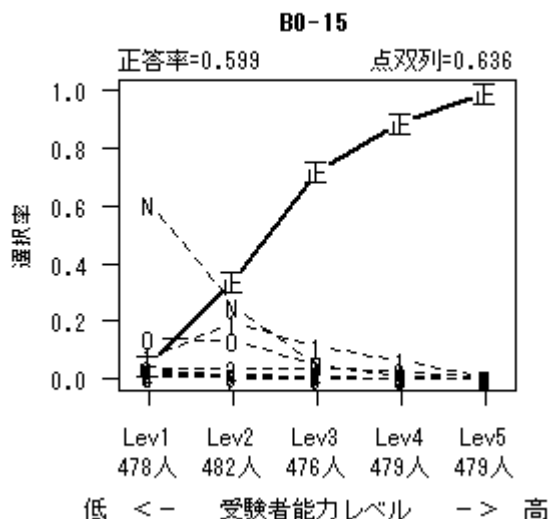


図 6-16 項目 B0-15 のG P分析図

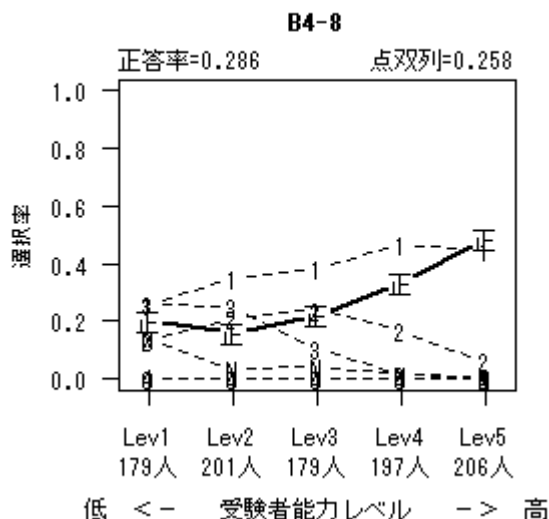


図 6-17 項目 B4-8 のG P分析図

7. IRT分析に関する妥当性の検証

7.1 2母数ロジスティックモデル利用の妥当性の確認

測定モデルとしてIRTモデルを採用し、テストで得られたデータから様々な情報を取り出すためには、まず、そもそも、そのデータとIRTモデルが適合しているかの検討が必要である。さらにIRTモデルにもとづいて、問題項目の難しさ（困難度）などの項目母数と呼ばれるいくつかの指標や特性値と呼ばれる児童生徒の学力を示す指標などを求めるためには、従来のテスト得点を求める場合とは異なり、かなり複雑な推定手続が必要となる。ここでは優れたユーザーインターフェイスを持ち、学術目的に利用する際には無料で使えるEasyEstimation（熊谷，2009）¹を用いて、2母数ロジスティックモデル利用の妥当性の検討ならびにBILOG-MGによって得られた母数の推定値のクロスバリデーションを行う。

7.2 局所独立および一次元性の検証

本報告で分析に用いた2母数ロジスティックモデルでは、3章でも論じたように、a) 局所独立の仮定、および、b) 尺度の一次元性の仮定の2つの前提条件が満たされていなければならない。その際、注意すべきは、2母数ロジスティックモデルのように、測定対象とする構成概念（本調査研究においては算数・数学の学力）が一つである場合には、この二つの仮定は同値となる。さらに実際上はテストに含まれる問題項目が、例えば別の問題を解いて得られた答を使って次の問題を解くといったことがなければ、局所独立の仮定はほぼ充足されていると見なしてよい。したがってここでは得られたデータから一次元性の確認を行うことで、2母数ロジスティックモデル利用の妥当性を検証する。

その方法としては四分相関係数行列における固有値のプロットから判断する方法（スクリーテスト）を適用する。EasyEstimationでは、四分相関係数行列を推定するにあたっては、Olson, U (1979) による2変量正規分布モデルに基づいた最尤推定法を採用している。

小学校版テストおよび中学校版テストにおけるスクリープロット（四分相関係数行列からの固有値をグラフにしたもの）を図7-1および図7-2に示す（固有値の大きさからみて11個目以降はほとんど変化しないため、便宜上プロットする固有値の数は10個とした）。これらの図から、小学校および中学校版においても、ほぼ同様の形状のグラフが得られていることがわかる。スクリーテストでは、第1固有値の値のみがそれ以降の値に比べて著しく大きい場合に、一次元性があると見なす。典型的な形状は図3-2に示された例のような場合であるが、本結果においては、確かに第1固有値の値は、それ以降の値に比べてかなり大きいものの第2、第3固有値の値も、それ以降の固有値に比べてやや大きい値を示している点が異なる。おそらくこれは重複テスト分冊法に起因して四分相関係数行列を求めるためのもとのデータ行列がいわゆる欠測値を含む不完全データ行列となっている影響が出ていると考えられる。その理由についてはさらに経験をつめばより明確になると考えられるが、いずれにせよ第1

¹ <http://irtanalysis.main.jp>

固有値の大きさから判断して、次元性の仮定は十分に満足していると考えられる。したがってこのまま2母数ロジスティックモデルにもとづいた考察をすすめることは妥当であると判断できる。

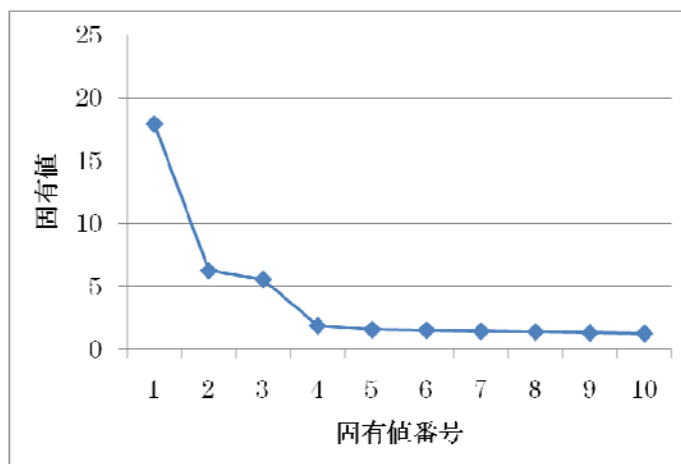


図 7-1 スクリープロット (小学校)

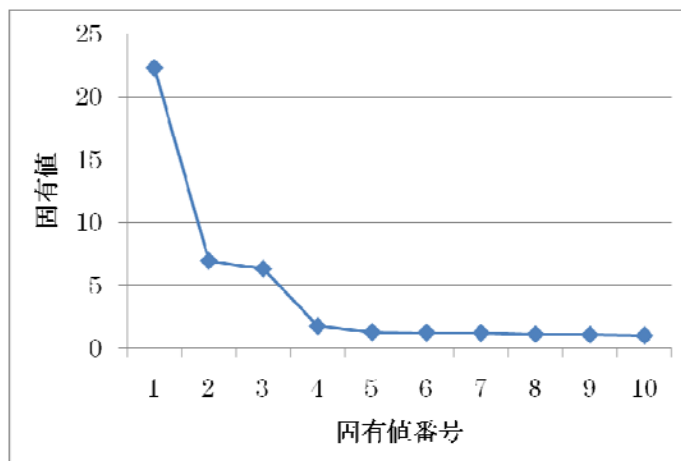


図 7-2 スクリープロット (中学校)

7.3 母数の推定結果の妥当性について

本報告書では、項目反応理論の2母数ロジスティックモデルにもとづき、項目母数および受検者特性値の推定がBILOG-Mを用いて行われている。その推定値についてEasyEstimationによってえられたものと照合する。なお、EasyEstimationにおける計算についてはデフォルト設定での推定を行った。

7.3.1 項目母数の推定値

BILOG-MGで得られた識別力母数、困難度母数、EasyEstimationで得られた識別力母数、困難度母数、および両プログラムでの母数の差の値を、表7-1、表7-2に示す。これらの表からも分かるとおり、BILOG-MGおよびEasyEstimationで得られた項目母数推定値は、識別力母数でBILOG-MGがいずれの推定値も低い値となっているが、絶対値とすれば非常に近い値であり、妥当な推定値が得られていることが分かる。いずれの推定値を使っても実用上は問題ないが、当初の予定通り、項目母数の推定値としてはBILOG-MGのものを利用することとする。

表 7-1 BILOG-MG と EasyEstimation との項目母数比較 (小学校)

項目番号	BILOG-MG		EasyEstimation		両プログラムの差	
	識別力	困難度	識別力	困難度	識別力	困難度
B0-1	0.538	-1.480	0.541	-1.473	-0.003	-0.007
B0-2	0.482	-1.841	0.484	-1.832	-0.002	-0.009
B0-3	0.822	-0.832	0.826	-0.829	-0.004	-0.003
B0-4	0.691	-3.360	0.695	-3.338	-0.004	-0.021
B0-5	0.774	-1.238	0.778	-1.232	-0.004	-0.006
B0-6	1.107	-1.448	1.112	-1.441	-0.005	-0.007
B0-7	0.923	-1.766	0.928	-1.757	-0.005	-0.010
B0-8	1.206	-2.550	1.214	-2.534	-0.007	-0.016
B0-9	0.822	-0.367	0.826	-0.367	-0.004	0.000
B0-10	1.004	-1.511	1.009	-1.503	-0.006	-0.008
B0-11	0.929	-1.468	0.934	-1.460	-0.005	-0.008
B0-12	1.094	-1.821	1.100	-1.811	-0.006	-0.011
B0-13	1.120	0.236	1.124	0.233	-0.004	0.003
B0-14	0.919	0.072	0.923	0.070	-0.004	0.002
B0-15	0.412	-0.535	0.414	-0.533	-0.002	-0.002
B1-1	0.500	-3.556	0.502	-3.536	-0.002	-0.020
B1-2	0.790	-1.617	0.795	-1.608	-0.005	-0.009
B1-3	1.028	-0.421	1.033	-0.420	-0.005	-0.001
B1-4	0.596	-0.086	0.598	-0.087	-0.003	0.001
B1-5	0.778	-0.981	0.781	-0.976	-0.004	-0.005
B1-6	1.009	-1.946	1.015	-1.935	-0.006	-0.011
B1-7	0.608	-0.544	0.610	-0.542	-0.003	-0.002
B1-8	1.389	0.575	1.392	0.570	-0.003	0.004
B1-9	1.093	-2.070	1.099	-2.057	-0.006	-0.013
B2-1	0.867	-0.816	0.870	-0.813	-0.004	-0.003
B2-2	0.670	-0.194	0.673	-0.195	-0.003	0.001
B2-3	0.413	1.318	0.415	1.310	-0.002	0.008
B2-4	1.125	-1.344	1.130	-1.337	-0.006	-0.007
B2-5	0.679	-0.189	0.681	-0.189	-0.003	0.000
B2-6	0.735	-1.975	0.738	-1.965	-0.004	-0.010
B2-7	0.478	0.859	0.480	0.854	-0.002	0.006
B2-8	0.486	-0.116	0.488	-0.117	-0.002	0.001
B2-9	0.682	-1.988	0.686	-1.978	-0.003	-0.010
B3-1	0.695	-0.888	0.698	-0.885	-0.003	-0.003
B3-2	0.522	-1.633	0.525	-1.625	-0.003	-0.008
B3-3	0.863	-1.606	0.867	-1.598	-0.004	-0.008
B3-4	0.927	-1.191	0.931	-1.186	-0.004	-0.005
B3-5	1.326	0.316	1.330	0.313	-0.005	0.004
B3-6	0.471	-0.628	0.474	-0.626	-0.002	-0.002
B3-7	0.725	-0.154	0.728	-0.155	-0.003	0.001
B3-8	0.920	0.199	0.924	0.196	-0.004	0.003
B3-9	0.803	0.500	0.806	0.495	-0.003	0.005
B4-1	0.499	-1.747	0.502	-1.738	-0.002	-0.009
B4-2	0.457	-1.156	0.459	-1.151	-0.002	-0.006
B4-3	1.055	-1.017	1.061	-1.012	-0.006	-0.005
B4-4	0.495	-0.301	0.498	-0.301	-0.002	0.000
B4-5	0.746	-0.025	0.750	-0.027	-0.003	0.002
B4-6	0.928	-0.982	0.933	-0.978	-0.005	-0.005
B4-7	0.619	-0.034	0.622	-0.036	-0.003	0.002
B4-8	0.779	-1.227	0.783	-1.222	-0.004	-0.006
B4-9	0.958	-1.349	0.963	-1.343	-0.005	-0.007
B5-1	0.692	-1.869	0.695	-1.858	-0.004	-0.012
B5-2	0.897	0.941	0.900	0.934	-0.003	0.006
B5-3	0.505	-1.309	0.507	-1.303	-0.002	-0.006
B5-4	0.555	1.718	0.557	1.708	-0.002	0.010
B5-5	1.115	-0.734	1.121	-0.731	-0.006	-0.003
B5-6	0.430	-0.565	0.432	-0.564	-0.002	-0.002
B5-7	0.466	-0.481	0.468	-0.480	-0.002	-0.001
B5-8	0.573	-0.664	0.576	-0.662	-0.003	-0.003
B5-9	1.351	-0.669	1.358	-0.667	-0.007	-0.002

表 7-2 BILOG-MG と EasyEstimation との項目母数比較 (中学校)

項目番号	BILOG-MG		EasyEstimation		両プログラムの差	
	識別力	困難度	識別力	困難度	識別力	困難度
B0-1	1.294	-1.634	1.300	-1.627	-0.006	-0.007
B0-2	1.065	-1.813	1.070	-1.806	-0.005	-0.008
B0-3	0.981	-0.384	0.983	-0.384	-0.003	0.000
B0-4	0.932	-0.685	0.935	-0.684	-0.003	-0.001
B0-5	1.495	-0.864	1.502	-0.862	-0.007	-0.002
B0-6	0.714	-1.876	0.716	-1.869	-0.003	-0.007
B0-7	1.611	-1.510	1.619	-1.504	-0.008	-0.006
B0-8	0.924	-0.370	0.927	-0.370	-0.003	0.000
B0-9	1.352	-1.270	1.359	-1.265	-0.006	-0.005
B0-10	0.512	-0.185	0.513	-0.186	-0.001	0.001
B0-11	0.647	-0.203	0.648	-0.204	-0.002	0.001
B0-12	1.499	-1.269	1.506	-1.264	-0.007	-0.005
B0-13	1.027	-0.160	1.030	-0.161	-0.003	0.001
B0-14	1.144	-1.580	1.149	-1.573	-0.005	-0.007
B0-15	1.581	-0.301	1.586	-0.302	-0.005	0.001
B0-16	0.872	-0.050	0.874	-0.052	-0.002	0.002
B1-1	1.003	-0.152	1.005	-0.153	-0.002	0.001
B1-2	0.830	-1.147	0.833	-1.144	-0.003	-0.004
B1-3	0.693	-0.692	0.695	-0.691	-0.002	-0.001
B1-4	0.590	-0.162	0.592	-0.163	-0.001	0.001
B1-5	1.202	0.292	1.203	0.289	-0.002	0.003
B1-6	0.879	-1.225	0.882	-1.221	-0.003	-0.004
B1-7	0.606	-0.595	0.608	-0.594	-0.002	-0.001
B1-8	1.278	-0.960	1.282	-0.958	-0.004	-0.003
B2-1	0.905	-1.577	0.909	-1.570	-0.004	-0.007
B2-2	1.614	-1.164	1.623	-1.159	-0.008	-0.004
B2-3	0.641	-0.753	0.643	-0.751	-0.002	-0.002
B2-4	0.695	-0.744	0.698	-0.742	-0.003	-0.002
B2-5	0.914	-0.321	0.917	-0.321	-0.003	0.000
B2-6	1.219	-1.357	1.225	-1.351	-0.006	-0.006
B2-7	1.012	-0.885	1.016	-0.883	-0.004	-0.003
B2-8	1.323	-0.545	1.328	-0.544	-0.005	-0.001
B3-1	1.033	-0.917	1.037	-0.914	-0.004	-0.003
B3-2	0.660	-0.908	0.662	-0.906	-0.002	-0.002
B3-3	1.378	-1.597	1.386	-1.590	-0.007	-0.007
B3-4	0.851	-0.346	0.854	-0.346	-0.003	0.000
B3-5	1.103	0.510	1.105	0.507	-0.002	0.003
B3-6	0.462	-0.645	0.463	-0.644	-0.001	-0.001
B3-7	1.088	-0.526	1.092	-0.526	-0.004	-0.001
B3-8	1.028	-0.735	1.032	-0.733	-0.004	-0.002
B4-1	1.231	-0.135	1.234	-0.136	-0.003	0.002
B4-2	0.969	-0.858	0.972	-0.855	-0.004	-0.003
B4-3	1.017	0.308	1.019	0.305	-0.002	0.003
B4-4	1.418	-0.904	1.424	-0.902	-0.006	-0.002
B4-5	1.426	-0.890	1.432	-0.888	-0.005	-0.002
B4-6	1.065	-1.011	1.069	-1.008	-0.004	-0.003
B4-7	1.288	-1.247	1.294	-1.242	-0.006	-0.004
B4-8	0.327	1.795	0.327	1.789	0.000	0.006
B5-1	1.177	-0.457	1.180	-0.457	-0.003	0.000
B5-2	1.175	-0.774	1.179	-0.773	-0.004	-0.002
B5-3	1.071	-0.572	1.074	-0.571	-0.003	-0.001
B5-4	0.927	-1.120	0.930	-1.117	-0.003	-0.003
B5-5	1.156	-1.343	1.161	-1.338	-0.005	-0.005
B5-6	0.608	-0.770	0.610	-0.769	-0.002	-0.001
B5-7	1.363	-0.453	1.367	-0.453	-0.004	0.000
B5-8	0.777	-0.286	0.779	-0.286	-0.002	0.001

7.3.2 児童・生徒の特性値

ここでは児童生徒に関する素得点，BILOG-MG で得られた特性値，EasyEstimation で得られた特性値，および両プログラムでの推定値の差について，小学校，中学校について，それぞれ最初の 20 名分を抜粋して代表例として表 7-3，7-4 に示す。他の児童・生徒についても同様の結果が得られている。

表 7-3 BILOG-MG と EasyEstimation との特性値の比較（小学校）

受検者番号	素得点	BILOG-MG	EasyEstimation	両者の差
S0001	16	-1.109	-1.104	-0.005
S0002	22	-0.399	-0.398	-0.001
S0003	20	-0.501	-0.500	-0.002
S0004	25	0.076	0.074	0.002
S0005	24	-0.029	-0.031	0.002
S0006	18	-0.749	-0.747	-0.003
S0007	24	-0.100	-0.101	0.001
S0008	24	-0.126	-0.127	0.001
S0009	18	-0.906	-0.902	-0.004
S0010	22	-0.262	-0.262	0.001
S0011	11	-1.765	-1.755	-0.010
S0012	30	1.485	1.476	0.009
S0013	31	2.052	2.040	0.012
S0014	32	2.727	2.710	0.017
S0015	20	-0.568	-0.567	-0.001
S0016	13	-1.520	-1.512	-0.008
S0017	22	-0.253	-0.254	0.001
S0018	21	-0.470	-0.469	-0.001
S0019	31	1.467	1.458	0.009
S0020	24	-0.047	-0.048	0.001

表 7-4 BILOG-MG と EasyEstimation との特性値の比較（中学校）

受検者番号	素得点	BILOG-MG	EasyEstimation	両者の差
T0001	26	0.167	0.164	0.003
T0002	27	0.270	0.267	0.003
T0003	30	0.943	0.937	0.007
T0004	17	-0.690	-0.689	-0.002
T0005	20	-0.456	-0.456	0.000
T0006	9	-1.398	-1.392	-0.006
T0007	29	0.662	0.657	0.005
T0008	31	2.062	2.052	0.009
T0009	26	0.399	0.395	0.004
T0010	24	-0.186	-0.187	0.001
T0011	24	0.008	0.006	0.002
T0012	31	1.595	1.586	0.009
T0013	23	-0.302	-0.303	0.001
T0014	23	-0.262	-0.263	0.001
T0015	14	-1.045	-1.042	-0.003
T0016	26	0.272	0.269	0.003
T0017	24	-0.142	-0.144	0.002
T0018	27	0.364	0.360	0.004
T0019	9	-1.506	-1.500	-0.006
T0020	26	0.230	0.227	0.003

上の表から分かるとおり，両者の値は非常に近いものである。数値のわずかな違いは，計算アルゴ

リズムよる違いのほか、前項で確認された項目母数推定値自体のわずかな違いが影響しているものと考えられる。その差は絶対値にすれば非常に近い値であり、再び、ここでも妥当な推定値が得られていることが分かる。項目母数の場合と同様、いずれの推定値を使っても実用上は問題ないが、当初の予定通り、項目母数の推定値としてはBILOG-MGのものを利用することとする。なお、BILOG-MGとEasyEstimationで得られた特性値の間の相関係数は、小学校で.99996、中学校で.9999998となり、ほぼ1と見なしても良いこともその証左となる。

ただ注意すべきは、各表での「両者の差」の絶対値の最大値は、小学校で0.431、中学校で0.015であった点である。小学校でやや大きな値を示しているが、これは素得点が1であった受検者について、BLOG-MGでは-4.000、EasyEstimationでは-4.431と推定されていたことによる。BLOG-MGではプログラムのデフォルトとして、特性値の推定を-4.000から4.000の範囲で行うという制限がかけられており、EasyEstimationではこのような制限をおいていないため、このような違いが生じている。これと同様の問題は、ある分冊で満点(全問正解)であったもの、逆に全問不正解であったものについても生じる可能性が有る。重複テスト分冊法の実用化の際には避けて通れない問題であり、今後の検討が必要な課題の一つである。