全国規模の学力調査における
重複テスト分冊法
適用の試み

平成２２年度文部科学省委託研究
「学力調査を活用した専門的課題分析に関する調査研究
研究成果報告書

平成２３年３月３１日

国立大学法人東北大学

はしがき

この報告は，平成 21 年度文部科学省企画公募研究「学力調査を活用した専門的な課題分析に関する調査研究」の『Ａ．全国的な学力調査の調査手法における技術的課題に関する調査研究』に応募し審査会を経て採用された研究，「全国規模の学力調査における重複テスト分冊法適用の試み」の成果をまとめたものである。

重複テスト分冊法のエッセンスは,受検者が,問題構成の異なる分冊とよばれるテスト冊子(Testlet)を受けるものの，その中に含まれている共通な問題項目ブロックを利用して，項目反応理論（Item Response Theory :IRT）を用いた等化（Equating）を行うことにより，全ての分冊を共通の尺度上で比較可能なようにするところにある。すなわち，解いたテスト冊子が異なっても，互いに比較できる得点(尺度値)を計算することができる。この利点を活かせば，1 回の実施あたりに関する受検者の解答負担が軽減できる一方，複数の分冊を生成できるため，測定したい領域の内容を幅広くカバーすることが可能となる。また，学力の経年比較のための指標開発の今後の研究展開の準備として，今回の実施データから IRT 分析等によって得られる項目統計量とともに問題項目をＤＢ化することもあわせて目指した。

重複テスト分冊法自体の目的は，飽くまでも関心下の領域全体に関する各種特性および対象とする母集団の学力分布に関する各種情報の取得にあって，必ずしも参加した児童・生徒個人や所属しているクラスや学校に関する学力状況・理解状況を把握するところにあるものではない。しかし，わが国の義務教育段階におけるテストは「指導のためのツール」としての要請が極めて強い。そのため，ＩＲＴモデルを利用することによって，児童・生徒に関する個別情報をはじめとして，いくつかの学習指導上の判断に資する情報を取り出すことは工夫次第によって可能であることも示した。ただし，この例示は重複テスト分冊法の本来の目的から言えば飽くまでも付随的なものである。

この報告書が，PISA や TIMSS などの国際学力調査ですでに利用されている最新の調査技術の，わが国における全国規模の学力調査への導入の端緒となれば研究代表者として幸甚である。

研究代表者　柴山　直

謝辞

お忙しい年間行事の中にも関わらず，貴重な時間を割いて本調査に参加していただきました新潟市の小学校・中学校，その児童・生徒のみなさま，先生方，ならびに新潟市教育委員会の諸氏に深く感謝申しあげます。

付　記

本報告書執筆中の平成 23 年 3 月 11 日 14 時 46 分､東北地方太平洋沖地震が発生しました。研究環境を支えるライフラインやネットワーク等が寸断される中､懸命の復旧に努められた東北大学関係者の皆様方、さらにはこの状況に鑑み、執筆にあたって多大なご配慮をいただきました文部科学省学力調査室のご関係の方々にはここに記して感謝申し上げます。(T.S.)

事業概要

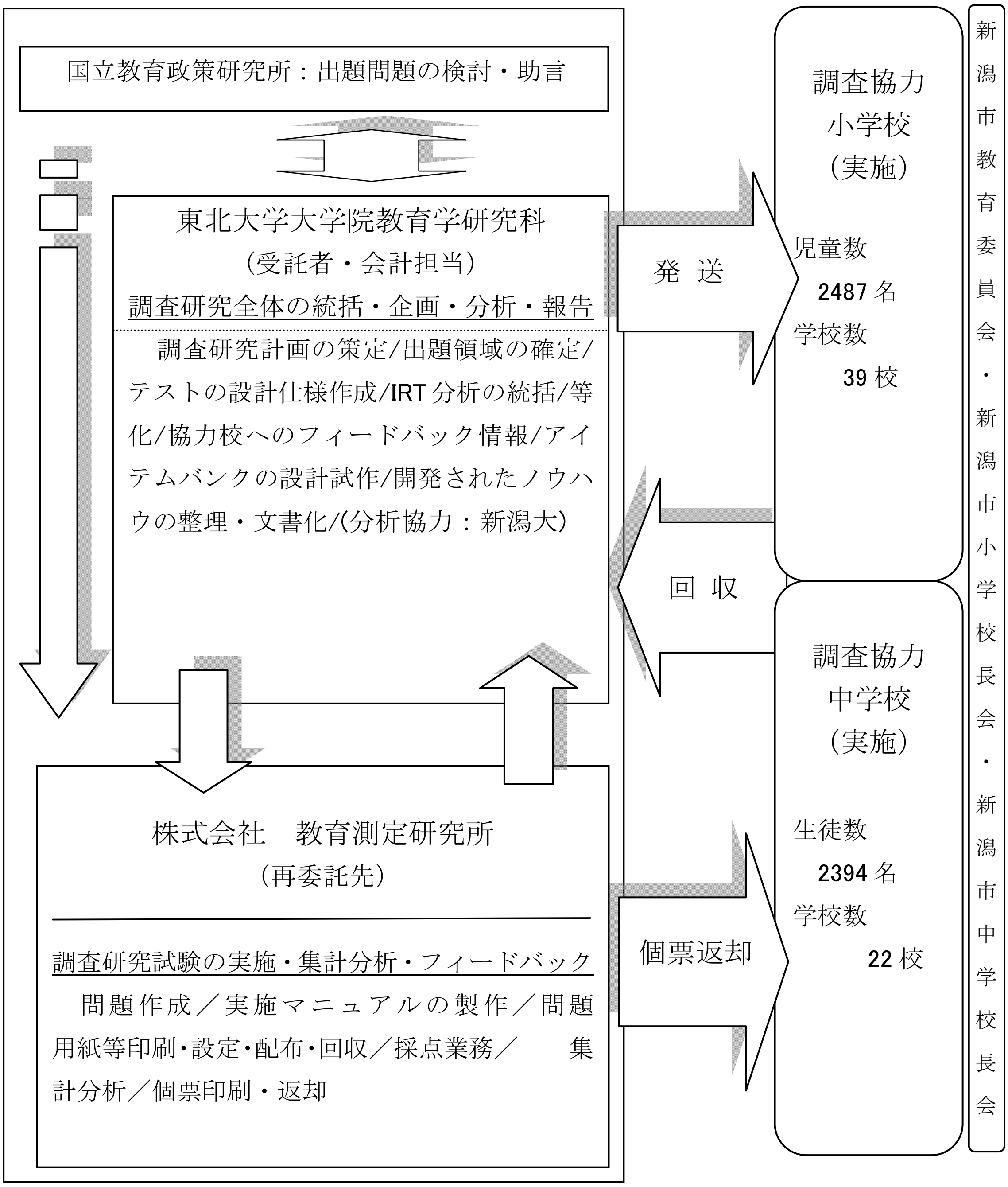
事 業 名	学力調査を活用した専門的な課題分析に関する調査研究
事業内容	全国的学力調査の調査手法における技術的課題に関する調査研究
委託期間	平成２２年９月１日から平成２３年３月３１日
事業者名	国立大学法人東北大学・大学院教育学研究科長・宮腰　英一
事 業 費	６，０１８ 千円

研究組織

研究代表	柴山　直	東北大学大学院教育学研究科　（全体統括/執筆/編集）
研究協力	佐藤　喜一	新潟大学教育・学生支援機構入学センター（第 3 章執筆）
	熊谷　龍一	東北大学大学院教育学研究科（第 7 章執筆）
研究助手	佐藤　誠子	東北大学大学院教育学研究科（第 5 章執筆）
資料整理	蝦名　正司	東北大学大学院教育学研究科
	宮田　佳緒理	東北大学大学院教育学研究科
	中野　友香子	東北大学大学院教育学研究科
	新国　佳祐	東北大学大学院教育学研究科
事務担当	紙屋　雅子	

実施集計	株式会社　教育測定研究所
作題監修	国立教育政策研究所

実施体制


国立教育政策研究所：出題問題の検討・助言
東北大学大学院教育学研究科
（受託者・会計担当）
調査研究全体の統括・企画・分析・報告
調査研究計画の策定/出題領域の確定/テストの設計仕様作成/IRT分析の統括/等化/協力校へのフィードバック情報/アイテムバンクの設計試作/開発されたノウハウの整理・文書化/(分析協力：新潟大)
株式会社　教育測定研究所
（再委託先）
調査研究試験の実施・集計分析・フィードバック
問題作成／実施マニュアルの製作／問題用紙等印刷・設定・配布・回収／採点業務／　集計分析／個票印刷・返却
発 送
回 収
個票返却
調査協力
小学校
（実施）
児童数
2487名
学校数
39校
調査協力
中学校
（実施）
生徒数
2394名
学校数
22校
新潟市教育委員会・新潟市小学校長会・新潟市中学校長会

調査研究計画

【準備】

1. 調査計画の策定
2. 学習指導要領の精査および出題領域の文書化
3. 重複テスト分冊法に基づくテスト・デザインの策定
4. 国立教育政策研究所から助言・指導のもと出題問題(教師アンケート項目を含む)の作成
5. 出題問題の確定
6. 項目配置計画の策定
7. 新潟市教育委員会を通して協力校への依頼と確定
8. 試験実施マニュアルの作成
9. 上記事項に関する教育測定研究所との協議

【実施】

10. 調査用紙等細部点検確認
11. 試験実施マニュアル・問題冊子・解答用紙の印刷
12. 配布準備，搬送
13. 調査実施，回収
14. データ入力作業
15. 採点
16. 上記事項に関する教育測定研究所との協議

【分析・データベース化】

17. 基礎統計量の集計
18. IRT 分析
19. 児童生徒の学習指導に必要な情報(推定正答確率など)のみを分析作業中に返却・報告
20. データ解析作業
21. データ解析結果の整理
22. アイテムバンクの仕様設計
23. アイテムバンクの試作
24. 上記事項に関する教育測定研究所との協議

【文書化】

25. 調査すべき項目事項の数，調査に参加可能な受検者数，および，調査に利用できる時間の三つの現実的な制約条件の下での，重複テスト分冊法による設計ノウハウ
26. ある項目群がある集団に偏って実施されることをさけるために必要な，計画的に受検者を分冊に割り振る手法に関する適用ノウハウ
27. 教師アンケートによる本方法の現場サイドから見た実施運用の実用性・有用性の検証結果
28. データへの項目反応理論の適用による等化および項目特性値（学力値）推定の試み
29. 経年変動をモニターするためのアイテムバンク（問題項目のデータベース）の試作
30. 取得できたノウハウおよび起こりうるトラブルへの対処法の文書化
31. 副次的に取得できた知見及び情報の整理（全国的な学力調査の結果と関連するもの）

実施経過

2010.09.01　　事業開始
株式会社教育測定研究所へ実施および集計作業部分を再委託

2010.09.01～2010.09.30
テストの設計・調査問題の作成・検討・印刷

2010.10.01　　新潟市小学校・中学校への調査協力依頼

2010.10.01～2010.10.08
調査対象学級及び対象人数の確認

2010.10.12～2010.10.16
調査問題の配送

2010.10.18～2010.10.22
調査の実施

2010.10.18～2010.10.29
実施終了校から順次解答済みの調査問題の返却

2010.11.12　　打合せ（個人票に掲載する情報について）
東北大学：柴山直　　於　株式会社教育測定研究所

2010.12.10　　個人票（調査シート）の各学校を通しての返却

2010.12.22　　打合せ（集計結果の中間報告）
東北大学：柴山直　於　株式会社教育測定研究所会議室

2010.12.23　　集計データの整理開始　東北大学：佐藤誠子

2011.02.04　　打合せ（集計結果の確認・データの分析等について）
東北大学：柴山直・熊谷龍一/新潟大学：佐藤善一　於　株式会社教育測定研究所

2011.02.26　　新潟市教育委員会向け個別報告書作成

2011.03.03　　打合せ（データ分析最終打合せ・執筆分担）
新潟大学：佐藤善一・東北大学：熊谷龍一　　於　東北大学大学院教育学研究科

2011.03.04　　新潟市教育委員会へ中間報告
東北大学：柴山直　　於　新潟市教育委員会学校支援課

2011.03.04　　打合せ（集計結果の最終報告）
東北大学：柴山直　　於　株式会社教育測定研究所

（以下，東北地方太平洋沖地震発生により実施中止）

2011.03.24　　報告書原稿入稿

2011.03.25　　最終報告会　於　東北大学大学院教育学研究科

2011.03.30　　報告書納入

目　次

全国規模の学力調査における重複テスト分冊法適用の試み

１．本研究の概要

1.1　本調査研究の意義と目的

本調査研究の最終的な目的は，マトリックスサンプリング(Matrix Sampling)とよばれる大規模な学力調査のための最新の技術を活用するための適用ノウハウの確立を目指すところにある。マトリックスサンプリングは社会調査で採用されるサンプリング調査法を，いわば調査すべき問題項目にも同時にあてはめたものである。そのための理論的なバックボーンとしてはこの半世紀，教育測定学・計量心理学の分野で著しく発展してきた項目反応理論(Item Response Theory : IRT)がある。この IRT モデルを学力の測定モデルとして採用することによって，分冊(testlet)と呼ばれる問題構成の違うテスト冊子をいくつも準備し，それらを異なる受検者集団に実施し，受検者の負担を軽減しながらも，以下のようなメリットを享受することが可能となる。すなわち，

A)　調べたい学力に関する全ての領域にわたってその定着度を従来の調査方法と比較してはるかに詳細に調べること。
B)　互いに異なる問題項目の組み合わせからなる分冊であっても，それぞれの分冊から得られた学力を共通の尺度上で表現して比較すること。
C)　過年度の問題と共通する問題を一部含んだ分冊をいくつも準備することにより，高い精度で学力の経年比較の追跡をすること。

などを実現することが期待できるのである。さらには，この方法と抽出調査法とを組み合わせることによって，同一予算規模であっても，教育施策を立案・遂行する上で必須の情報を，現在よりもはるかにきめ細かい観点から系統的に得ることができるようになる。また，学校外での生活習慣や学習に対する意識などの他の外的な情報に関する調査項目を分冊中に含めることによって，学力データと同時にそれらの情報をとることも当然可能であり，その場合においても極めて識別力の高い統計値を得ることができるようになる。

ただし，マトリックスサンプリング自体はかなり複雑な技術を必要とするため，本研究ではまず，上の目的のうち，分冊によって調査するところにまず焦点を当て，A),B)についてのノウハウを得ることを主たる目標とした上で，C)の経年比較や抽出調査法との組み合わせによる母集団分布の特性把握の具体的方法，さらには本調査研究で得られたノウハウの算数・数学以外の他の教科への汎用化については今後の課題として，可能ならば段階的に展開していくことにした。

1.2　学力分布と得点分布との区別

学力の調査において出題問題が異なれば学力の経年比較はできない。問題が異なる通常のテストにおいては，得点の変化が，受検者の学力の変化によるものなのか，問題の困難度の変化によるものかの区別ができないためである。通常のテストで使われているテスト得点にはこの両者が分かちがたく含まれている。そのため，極端な場合には，やさしいテストを受けた学力の低い受検者と難しいテストを受けた学力の高い受検者のテスト得点が実際の学力の高低と逆転してしまう現象も起こり得る。そこで学力を追跡するため,毎年同じ問題を繰り返し使うなどの方策が採られることがある。しかし,

長期間にわたる学力調査では，例えば学習指導要領の改訂等の理由で，問題自体が使えなくなる，問題が公開できない，問題の場面設定が年月とともに古くなる，などの問題点が生じると考えられる。

こうした問題点を克服する有効な手段として，項目反応理論モデル（IRT モデル）がある。IRT モデルは，従来の単なる加算方式によって求められるテスト得点とは違い，受検者の問題項目への正誤反応パターンにもとづいて，受検者の学力そのものを，問題項目の困難度から明確に分離して扱うことができる。その原理を簡単に見ておこう（詳細は第 3 章参照）。次の問題は現在の大学入試センター試験の前身である大学共通第 1 次学力試験のある年の数学試験で出された 24 問のうちのある問題である。受検者数は約 30 万名であり，配点に重みを設けずに正答なら 1 点，誤答なら 0 点とする（このような得点のことを素得点と呼ぶ）と，その得点は 0 点から 24 点の範囲に入る。得点ごとにこの問題に正答した受検者の割合を計算する。このテストで 0 点であったものの中にはこの問題が正答できたものは存在しないからその割合は当然 0 となる。逆に満点の 24 点を取った受験生達は全員この問題に正答しているからその割合は 1 となる。

問題　半径√5/2 の円に内接する二等辺三角形 ABC において，AB=AC=2 とする．また，A を通るこの円の直径を AD とする．このとき，sin∠BAC=キ／クである．

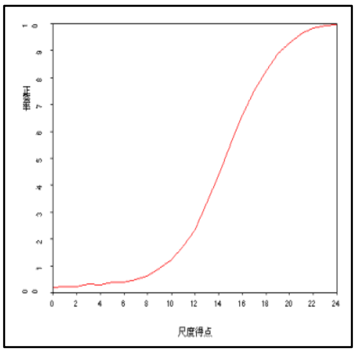


図 1-1 項目特性曲線の例（良い問題）

さらに，横軸に素得点，縦軸に素得点ごとの正答率をプロットし，それらを結べば図 1-1 の様な曲線が得られる。この曲線のことを項目特性曲線(Item Characteristic Curve : ICC)と呼ぶ。当然ながら得点が高い層ほど正答率は高くなる。

次の問題はやはり同じ大学共通一次試験の国語で出題された問題である。この問題の項目特性曲線を描いたものが図 1-2 である。テスト得点が高い層よりも低い層の方で若干正答率が高くなっていることが読み取れる。すなわちこの問題は調べたい国語の学力を識別できていないという意味で悪い項目ということになる。

問題　傍線部「夙に」の意味として最も適当なものを，次の①～⑤のうちから一つ選べ．		
①うまれつき	②朝早くから	③心のそこから
④ずっと以前から	⑤十二分に	

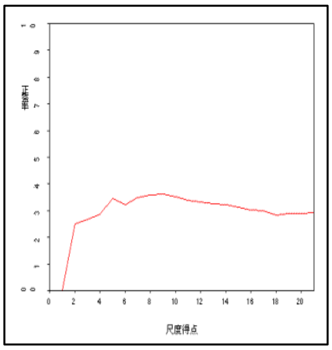


図 1-2 項目特性曲線の例（悪い問題）

ここまでの議論ではテスト得点と学力を同一視してきたが，IRT モデルにおいてはこの二つを明確に区別して扱う。すなわち学力を表す母数として特性値 θ を導入し，ある θ のもとである項目に正答する確率を表すことを考える。例えば，2 母数ロジスティックモデルと呼ばれるものは，次式で記述される。

$$P\left(X_j = 1 \mid \theta\right) = \frac{1}{1 + \exp\left\{-1.7 a_j\left(\theta - b_j\right)\right\}}$$

この数式を使えば，母数の組み合わせによって，図 1-3 にしめすように様々な項目特性曲線をもつ項目を特性値 θ のもとで扱えるようになる。点線で表された項目は実線で表された項目よりも難しく，実線の左側の曲線で表された項目は実線の項目よりやさしく，さらに実線を横切るような曲線で表された項目は実線の項目よりも学力の識別ができていないことを表している。この特徴を利用すれば，もはやわれわれはある特定のテスト冊子のみに依存することなく目指す学力 θ を測定することが可能となるのである。

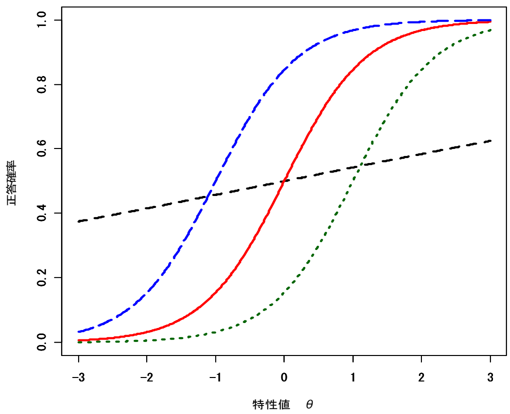


図 1-3 様々な項目特性曲線

さらに IRT モデルを採用することによって，混同されがちなテストの得点分布と測定すべき学力分布とを分けて扱うことができるようになる。これを端的に示したのが次ページの図 1-4a～1-4d である。図 1-4a が全体的に測定精度の高いテストの場合に学力分布がテスト特性曲線と呼ばれるものを通してテストの得点分布に変換されるプロセスを示している。図 1-4b がそれとは逆に，いい加減に作られた

結果，測定精度が低い場合の様子を示している。どちらの図においても真に知りたい学力分布はまったく同じであるが，テストの性質によって手にできるスコアの分布が異なっていることがわかる。前者にくらべ後者は分布の広がり方がかなり縮まっており，児童・生徒の個人差の識別ができていない，したがって精度の悪いテストということになる。

さらに，例えば資格試験や検定試験などのようにある水準以上の学力水準・能力水準であるか否かを判断する場合のテストにおいては，図 1-4 c のような双峰分布（ふた山分布）にテスト得点の分布がなるようなテストを準備する必要がある。真に知りたい能力分布は図 1-4a, b と同じであるにも関わらず，テストの作りによっては，テスト得点の分布がこのように変化する。いわゆる「学力のふた山分布化」現象の議論が，テスト得点の分布をみての議論なのか，学力分布の形を推定しての議論なのかを明確に区別しておかないと判断を間違う可能性をこのことは示しているのである。

最後の図 1-4 d は図 1-4a にくらべて全体的に難しい問題からなるテストを作った場合のテスト得点の分布の様子を表したものである。能力分布がここでも同じであるにも関わらず，テスト得点の分布は得点の低い方に山ができていることがわかる。

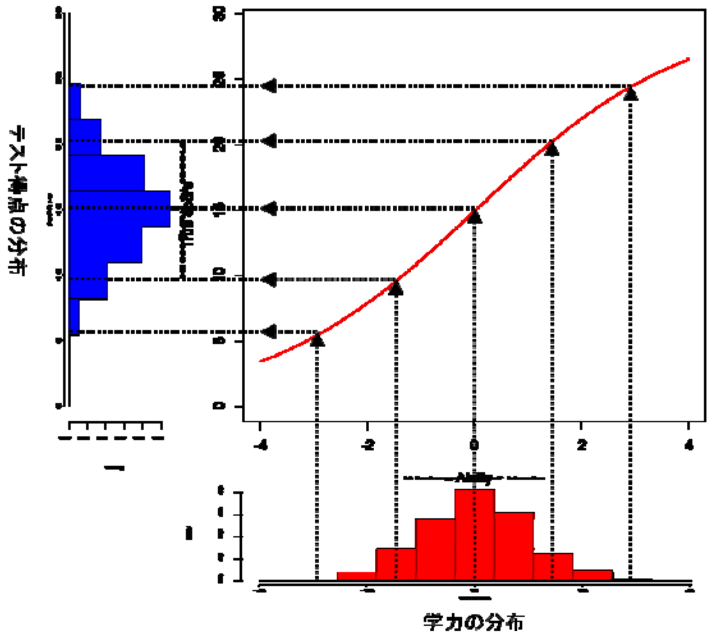


図 1-4a 学力分布と得点分布（信頼性が高いテスト）

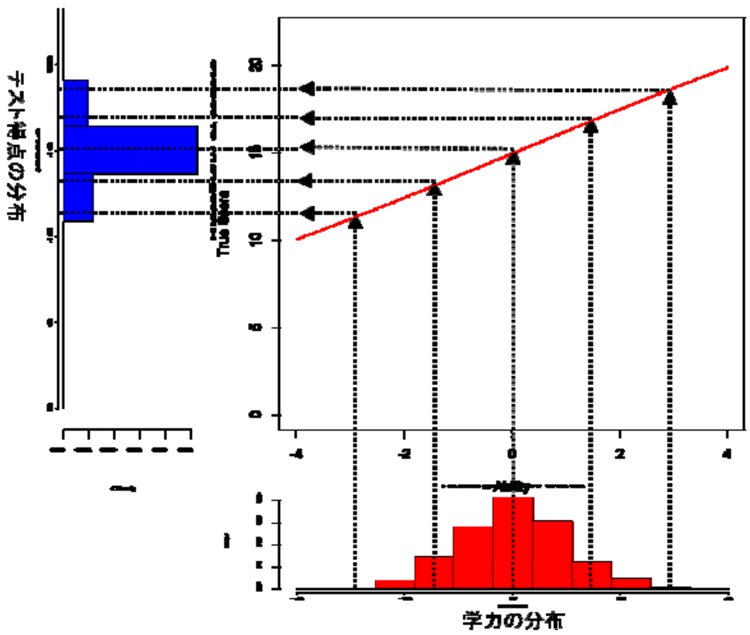


図 1-4 b 学力分布と得点分布（信頼性が低いテスト）

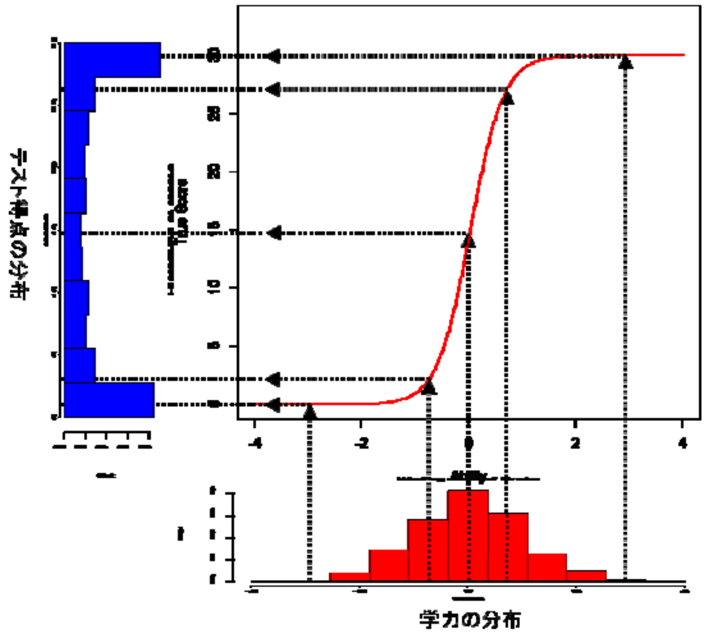


図 1-4ｃ　学力分布と得点分布（平均的な難しさの問題を多くした場合）

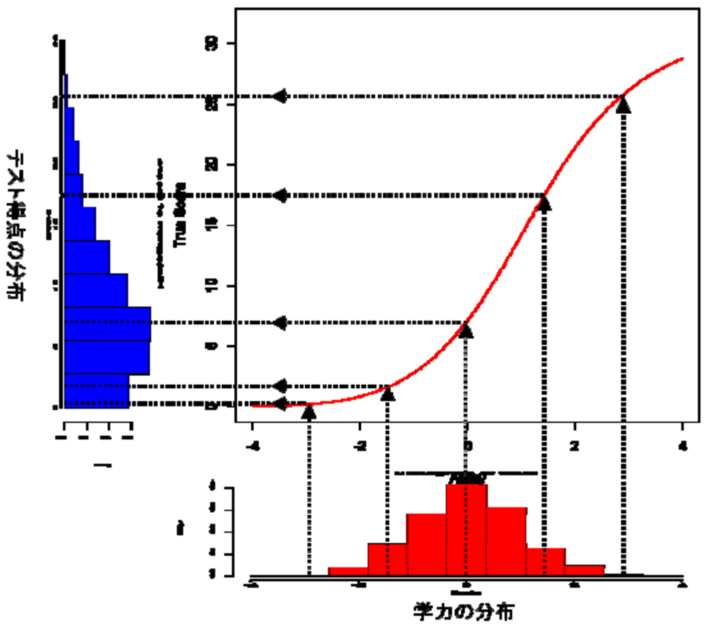


図 1-4ｄ　学力分布と得点分布（難しい問題を多くした場合）

1.3　重複テスト分冊法

このように学力調査において,真に知りたい学力分布とそこからテストによって生み出されてくるテスト得点の分布とを区別して考えることが，技術論的には必須の前提条件となる。このアイディアを応用すれば，ある条件下では例えば問題構成が年ごとに変わっても，学力分布の経年変化をトレースすることが可能となる。また，同様に，一斉に同一冊子を全ての児童・生徒に実施しなくても，互いに問題構成の異なるテスト冊子（分冊）を幾通りか準備することで，手分けして児童・生徒にそれを受けてもらえば，全体としての学力分布を推定することが可能となる。本研究で試みた重複テスト分冊法はまさにこの考え方が基本となっている。

具体的には,様々な理由で問題の入れ替えがあっても,数学なら数学の学力そのものをとらえ続けることが可能となる。このことによって得られるメリットとしては,

1）　長期間にわたる調査で起こりうる問題の自然漏洩に対して，問題の入れ替えによる対処が可能
2）　使われなくなった問題を公開することにより，事後的にではあるが，問題の妥当性・信頼性の確認が可能。
3）　最新の教育思潮を反映した問題を追加することが可能。
4）　事前テストが，対象集団とは同質だが別の受検者からなる集団に実施できれば，等化によって当該問題の難易度などをあらかじめ統計的に調整でき，本調査での利用後直ちに問題を公表することが可能。

などが指摘できる。

さらに,この原理を応用することによって,互いに異なる問題項目からなる複数のテスト冊子(分冊)を別々の受検者に実施しても，学力を分冊相互で比較可能なように数値化することもできる。IRT モデルを利用することによって，受検者の学力と分冊を構成している問題項目の困難度とを分離して扱えるからである。年間数回実施される TOEFL などで，いずれの回を受検してもそこで得られる結果が受検した回に関わらず同値とされ，そのスコアが選抜や採用で使われるのも同じ原理によっている。

このような IRT モデルのもつ種々の利点をさらに展開応用したのが,この「全国規模の学力調査における重複テスト分冊法適用の試み」である。重複テスト分冊法の基本的な考え方は，調査対象の受検者が，それぞれ異なるテスト冊子を受けるものの，分冊とよばれる各テスト冊子の中に含まれている互いに共通な問題項目ブロックを利用して，項目反応理論（IRT）を用いた等化(equating）を行うことにより，全ての分冊を共通の尺度上で比較可能なようにするところにある。言い換えれば，解いたテスト冊子が異なっても，互いに比較できる得点(尺度値)を計算することができる。その性質を利用すれば多種の分冊を生成でき，測定したい領域の内容を，1 回の実施であっても幅広くカバーすることが可能となる。

重複テスト分冊法のそもそもの発想は,この方法が OECD の児童生徒の学習到達度調査(PISA)や IEA の国際数学理科教育動向調査（TIMSS）で採用されていることからも明らかなように，行政として必要なのは個々の児童・生徒の学力よりはむしろグループ全体としての学力を幅広くとらえるところにあ

る。その際，いわゆる全数調査ではなく対象とする母集団からの児童・生徒の抽出による調査をするための標本調査法の採用がすぐに念頭に浮かぶ。しかし学力調査の場合においてはもう一つ重要な側面がある。それは調査すべき学習内容の範囲もまた広範囲に及んでいるという点である。ところが児童生徒一人が一科目あたり答えられる問題数というのは実際上限りがある。本来ならば調査すべき学習内容を全て網羅しかつそれぞれにおいて多角的多面的に丁寧な調査をしようとするとそのために必要な問題数は莫大なものになる。このように本来準備すべき問題の側面から見た母集団のことをいま仮に項目母集団と呼ぶとすると，一人の生徒にそれを全て解いてもらうのは事実上不可能である。
これは従来の学力調査の実施方式をめぐってほとんど看過されてきた点であるが,児童生徒が抽出されたものか全数を対象にしたものかに関わらず調査の目的からいえばたいへん本質的な問題である。

この問題を解決するために開発されたのが項目マトリックスサンプリング(Item-Matrix Sampling または単にマトリックスサンプリング(Matrix Sampling))と呼ばれる方法である。要するにこの方法は児童・生徒からも問題（項目）からもそれぞれ標本抽出を行い，その両者を組み合わせて集団としての学力調査(Group-score assessments)を効率良く行おうというものである。しかし，単に別々の問題からなる分冊を別々の成員からなる集団に実施するだけでは不十分である。そのことを単純化して示したのが下の図である。

	問題ブロック		
受検者集団	A	B	C
a	○		
b		○	
c			○

表 1-1a　受検者集団と問題ブロックの配置
（共通ブロックがない場合）

	問題ブロック		
受検者集団	A	B	C
a	○	○	
b		○	○
c	○		○

表 1-1 b　受検者集団と問題ブロックの配置
（共通ブロックがある場合）

表 1-1a では受検者集団はそれぞれ互いに異なる問題を解くことになり，その結果は比較できない。しかし表 1-1b では受検者集団間で共通の問題が存在することになる。そこで先に述べた IRT モデルを介して共通部分の情報を使うことによって，例えば受検者集団 a の問題ブロック C における正答率を推定することができるようになる。また，受検者集団 a に実施される問題ブロック A と B をあわせたもの，同様に受検者集団 b に実施される問題ブロック B と C を合わせたたものを分冊(testlet)と呼ぶことにする。
さらにここに実験計画法の考え方を導入することによって,統計的には同質でありながら互いに成員の異なる受検者集団にこれらの分冊を実施すれば，どの問題に対しても受検者集団の影響は同等になると考えることができる。そのようなデザインの組み方で代表的なものが釣合型不完備ブロックデザイン(Balanced Incomplete Block Design)である。ただし本調査研究ではまず，共通した（重複した）ブロックの情報から学力調査に必要な情報をどのようにして取り出すかということを主目的とし，もし遂行可能なら次の研究段階において厳密な意味での BIB デザイン使用に関するノウハウの開発を行うこととする。また，この意味で本調査研究ではマトリックスサンプリングという専門用語ではなくて，共通ブロックを使うことを強調した重複テスト分冊法という用語を使うことにした。

1.4　テスト・デザイン

新潟市の小中学校の協力を得て行った本調査研究の実際のテスト・デザインは下の図のようになる。この図では，例えば分冊１には共通ブロック，ブロック１，ブロック２が含まれている。同様にその下の分冊２には共通ブロック，ブロック２，ブロック３が含まれている。いずれの分冊も設問数としては 32 ということになるが，分冊１を受ける児童生徒と分冊２を受ける児童生徒が共通して解く問題は共通ブロックとブロック２のあわせて 24 個で，残りの８設問は互いに異なる問題となる。このようにして分冊相互に共通な項目を含めながら，５つの分冊を準備して領域全体をカバーするのが重複テスト分冊法の考え方である。なお，下図の設問数は説明のためのもので，実際には若干数増減する。また解答方式は記入式とする。さらに，実施問題の内容的妥当性等の検討については，全国学力・学習状況調査や教育課程実施状況調査の問題作成の実績がある国立教育政策研究所の助言を得た。

分冊番号	設問のブロック番号（設問数）					
分冊１		ブロック１ (8)	ブロック２ (8)			
分冊２			ブロック２ (8)	ブロック３ (8)		
分冊３	共通ブロック (16)			ブロック３ (8)	ブロック４ (8)	
分冊４					ブロック４ (8)	ブロック５ (8)
分冊５		ブロック１ (8)				ブロック５ (8)

図 1-3　今回のテストデザインの概念図

またブロックの順序効果を確認するために実際の分冊では同じブロックを持ちながらも,共通ブロックをはさんでブロックの順序を入れ替えたものを準備した。例えば分冊１においては（ブロック１・共通ブロック・ブロック２）とした分冊 1a と（ブロック２・共通ブロック・ブロック１）とした分冊１b とを準備した。順序効果を見るためには厳密には問題ごとに位置を調整しなければならないが，そこまでの精密な検討は別の機会に譲ることとした。なお，ブロックの位置を表す用語としてパートを採用した。すなわちパート１は一番最初，パート２は真ん中，パート３は最後をあらわしている。パート番号を採用することで位置情報が今回の様に３と限定されていなくてもさらに位置情報が増加した場合でも表記が可能となることも念頭においている。

さらに，全てに分冊に共通ブロックを含めた理由は，異なるブロックから推定された学力値の妥当性を共通ブロックの正答数との相関から確認するためである。すなわち、あるブロックのみを利用して求められた等化後の尺度値が高ければ当然共通ブロックの正答数得点も理論的には高くなるはずである。このことが確認できれば，実際のテストにおいては全ての分冊に共通ブロックを含める必要はな

くなり，その分，さらに実施できる問題数を増やせることを意味する。

表 1-2 問題冊子ブロック構造

分冊番号	パート 1	パート 2	パート 3
1a	ブロック 1	共通ブロック	ブロック 2
1b	ブロック 2		ブロック 1
2a	ブロック 2		ブロック 3
2b	ブロック 3		ブロック 2
3a	ブロック 3		ブロック 4
3b	ブロック 4		ブロック 3
4a	ブロック 4		ブロック 5
4b	ブロック 5		ブロック 4
5a	ブロック 5		ブロック 1
5b	ブロック 1		ブロック 5

1.5　個別指導のための推定正答確率の付加的導入

さらに，今回の調査では，項目反応理論モデルを利用して，ある生徒が受けなかった残りのブロック，例えば分冊１を受けた生徒を例にとれば，その生徒がもしブロック３～５の問題を受けたとしたらどの程度の確率でその問題に正しく答えられるかいう推定正答確率も計算する。その生徒は実際には 32 問しか解いていないが，残りの 32 問についての情報も得られるということになり，その生徒の数学に関する学力の全体像を把握できようなる。この情報は個票にして協力校を通して児童生徒に返却した。

2．テストの設計と開発

2.1　テスト設計の方針

テストを設計するに当たって，問題冊子を３パートに分け，それぞれのパートを先行パートとしてのブロック，共通ブロック，後行としてのブロックと重複テスト分冊法の設計が決まった。この方針を最大限活かすために，以下の方針をとった。

（ア）　共通ブロックには，各設問は算数・数学の全領域をまんべんなく均等に含むものとする。ただし，設問をする単位の学習項目は約４０あり，その全てができるわけではない。

（イ）　それゆえ，共通ブロックの学習項目の中から各設問の難易度に関しては平均的なものとする

（ウ）　個別ブロックは，それぞれの個別ブロック毎に各設問は全領域がまんべんなく含まれるようにする。ただし，設問をする単位の学習項目は約４０あり，可能な限り共通ブロックで選ばれなかった項目を優先する。

（エ）　個別ブロックの各設問の難易度は，ブロック全体としては平均的になるようにするが，それぞれは難易度高から低へ幅広いものとする

（オ）　共通ブロックと５つの個別ブロックにより学習指導要領の学習項目の小学校なら第４学年，第５学年，中学校なら第１学年と第２学年を網羅することとする

これらのことにより，基準となる共通ブロックにおいてテストを受けるグループのレベルの差を把握できるようにした。かつ，５つの個別ブロックで学習指導要領上の全学習項目を調査することとした。また，小学校では共通ブロック１５問，個別ブロック９問で，中学校では共通ブロック１６問，個別ブロック８問で構成することとした。

共通ブロックと個別ブロックの組み合わせにおいて，留意点がある。特定の学習項目において複数の設問を選ぶ場合には，ブロックの組み合わせにおいて重複して出題しないようにすることが制約としてある。また，いわゆるＡ型問題のみで構成し，Ｂ型問題は扱わないこととした。問題は経年変化をとらえるための指標開発のために次年度も利用する可能性があるため，非公開とし本報告書には掲載していない。その代わり後述するように問題の質の良し悪しを判断できるように項目分析や IRT 分析の結果，得られた統計指標を報告する。ただし，出題内容については表 1a, 表 1b のブロック出題表を参照されたい。

2.2　学習指導要領との関係

PISA などに見られるように学力調査に先立っては調査すべき学力の定義が必須であることは言を俟たない。今回の調査研究においては新しい調査方法のノウハウの確立が主たる目的であるため，何をもって学力とすべきかの議論はなされていない。そのため，学力の定義の代用として学習指導要領を仮に使用した。結果として，市販の標準学力テストや教育課程実施状況調査と類似する調査研究になっているが，本研究は飽くまでも重複テスト分冊法のノウハウを確立するところにあることを強調すると同時に，本来的には全国的な学力調査において調査すべき「学力」は方法論より先に決定すべき問題であることをここでは指摘するにとどめる。

以上のことを前提に，試験時間が限られる一般的な調査と異なり，設定した当該の対象領域の全てを設問にすることで，学習指導要領の全体を網羅するようにした。設計では，算数・数学の領域，内容，項目と３段階で設計を行っており，最細分化されている項目から１設問以上を抽出した。

また，形式上大問構成をとっているが，実質的には単問で構成した。これはテスト設計上の観点に起因するものである。すなわち，大問構成の場合には，そもそもの指導要領上の項目が複数入っており，その組み合わせがどのように影響して調査結果がかわるのか等の点は今回の調査研究ではカバーしていない。

なお，開発した項目は全て国立教育政策研究所の監修を経て，何回かブラッシュアップした上で実施した。

表 2-1a ブロック出題表（小学校）

学年	領域		単元		内容		共通	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ
第4学年	A	数と計算	(1)	整数の表し方	ア	億，兆の単位					○	
			(2)	概数と四捨五入	ア	概数が用いられる場合	○					
					イ	四捨五入						○
			(3)	整数の除法	ア	除法の計算の仕方						
					イ	除法の計算を用いること			○			
					ウ	被除数，除数，商及び余りの間の関係	○			○		
					エ	除法について成り立つ性質			○			
			(4)	小数の仕組みとその計算	ア	小数の意味と表し方			○			
					イ	小数の仕組みと数の相対的な大きさ				○		
					ウ	小数の加法，減法	○					○
			(5)	分数の意味とその表わし方	ア	分数の意味と表し方	○	○				
					イ	単位分数					○	
	B	量と測定	(1)	面積の単位と測定	ア	面積の単位と測定の意味						
					イ	面積の単位（cm2）			○			
					ウ	正方形，長方形の面積の求め方	○		○			
			(2)	角の大きさ	ア	回転の大きさ	○				○	
					イ	角の大きさの単位	○				○	
	C	図形	(1)	二等辺三角形，正三角形などの	ア	二等辺三角形，正三角形				○		○
					イ	角						
					ウ	円，球			○	○		
	D	数量関係	(1)	伴って変わる二つの数量	ア	対応や関係を調べること						
					イ	変化の様子と折れ線グラフ	○					
			(2)	数量の関係を表す式	ア	四則の混合した式や（ ）を用いた式			○			○
					イ	公式						
			(3)	資料の分類整理	ア	二つの観点から分類整理すること		○			○	
					イ	資料の落ちや重なりについて調べること						
					ウ	折れ線グラフの読み方とかき方	○					
第5学年	A	数と計算	(1)	整数の性質	ア	偶数，奇数					○	
			(2)	整数，小数の記数法	ア	10倍，100倍，1/10，1/100などの大きさ	○					
			(3)	小数の乗法，除法	ア	小数×整数，小数÷整数の意味と計算		○		○		
					イ	小数の乗法，除法の意味						○
					ウ	小数の乗法，除法の計算，余りの大きさ	○			○	○	
			(4)	分数の意味と加法，減法	ア	大きさの等しい分数						○
					イ	整数，小数と分数の関係				○		
					ウ	除法の結果と分数		○				
					移措	同分母分数の加法，減法	○					
	B	量と測定	(1)	図形の面積	ア	三角形，平行四辺形の面積の求め方	○	○	○		○	○
					移措	ひし形，台形の面積の求め方				○		
					イ	円の面積		○				
	C	図形	(1)	平行四辺形，ひし形，台形	ア	直線の平行や垂直の関係		○				
					イ	平行四辺形，ひし形，台形	○	○			○	
					移措	多角形，正多角形			○			
					移措	図形の合同						
					ウ	図形の性質			○		○	
					エ	円周率	○					
					移措	立方体，直方体						○
					移措	直線や平面の平行や垂直						
	D	数量関係	(1)	四則に関して成り立つ性質	ア	交換法則，結合法則，分配法則						○
			(2)	百分率			○	○		○		
			(3)	円グラフや帯グラフ								
			(4)	数量の関係を表す式					○	○		

表 2-1b　ブロック出題表（中学）

学年	領域	単元	内容	共通	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ
第1学年	A数と式	【正負の数】具体的な場面を通して正の数と負の数について理解し，その四則計算ができるようにするとともに，正の数と負の数を用いて表現し考察することができるようにする	正の数と負の数の必要性と意味を理解すること	○			○		
			小学校で学習した数の四則計算と関連付けて，正の数と負の数の四則計算の意味を理解すること	○		○			
			正の数と負の数の四則計算をすること	○					
			具体的な場面で正の数と負の数を用いて表したり処理したりすること		○		○		
		【文字と式】文字を用いて数量の関係や法則などを式に表現したり式の意味を読み取ったりする能力を培うとともに，文字を用いた式の計算ができるようにする	文字を用いることの必要性と意味を理解すること						○
			文字を用いた式における乗法と除法の表し方を知ること	○					
			簡単な一次式の加法と減法の計算をすること	○		○	○	○	
			数量の関係や法則などを文字を用いた式に表すことができることを理解し，式を用いて表したり読み取ったりすること					○	
		【一次方程式】方程式について理解し，一元一次方程式を用いて考察することができるようにする	方程式の必要性と意味及び方程式の中の文字や解の意味を理解すること	○					
			等式の性質を基にして，方程式が解けることを知ること		○				
			簡単な一元一次方程式を解くこと及びそれを具体的な場面で活用すること			○			○
	B図形	【平面図形】観察，操作や実験などの活動を通して，見通しをもって作図したり図形の関係について調べたりして平面図形についての理解を深めるとともに，論理的に考察し表現する能力を培う	線対称，点対称の意味を理解するとともに，対称性に着目して平面図形についての直観的な見方や考え方を深めること						○
			角の二等分線，線分の垂直二等分線，垂線などの基本的な作図の方法を理解し，それを具体的な場面で活用すること	○					
		【空間図形】観察，操作や実験などの活動を通して，空間図形についての理解を深めるとともに，図形の計量についての能力を伸ばす	空間における直線や平面の位置関係を知ること	○					
			空間図形を直線や平面図形の運動によって構成されるものととらえたり，空間図形を平面上に表現して平面上の表現から空間図形の性質を読み取ったりすること			○		○	
			扇形の弧の長さと面積並びに基本的な柱体，錐(すい)体及び球の表面積と体積を求めること				○		
	C数量関係	【比例と反比例】具体的な事象の中から二つの数量を取り出し，それらの変化や対応を調べることを通して，比例，反比例の関係についての理解を深めるとともに，関数関係を見いだし表現し考察する能力を培う	関数関係の意味を理解すること						
			比例，反比例の意味を理解すること	○					
			座標の意味を理解すること		○				
			比例，反比例を表，式，グラフなどで表し，それらの特徴を理解すること					○	○
			比例，反比例を用いて具体的な事象をとらえ説明すること				○		
第2学年	A数と式	【式の計算】具体的な事象の中に数量の関係を見いだし，それを文字を用いて式に表現したり式の意味を読み取ったりする能力を養うとともに，文字を用いた式の四則計算ができるようにする	簡単な整式の加法，減法及び単項式の乗法，除法の計算をすること	○	○				
			文字を用いた式で数量及び数量の関係をとらえ説明できることを理解すること				○	○	
			目的に応じて，簡単な式を変形すること		○	○			
		【連立方程式】連立二元一次方程式について理解し，それを用いて考察することができるようにする	二元一次方程式とその解の意味を理解すること						○
			連立二元一次方程式の必要性と意味及びその解の意味を理解すること						
			簡単な連立二元一次方程式を解くこと及びそれを具体的な場面で活用すること	○				○	○
	B図形	【平行と合同】観察，操作や実験などの活動を通して，基本的な平面図形の性質を見いだし，平行線の性質を基にしてそれらを確かめることができるようにする	平行線や角の性質を理解し，それに基づいて図形の性質を確かめ説明すること			○			
			平行線の性質や三角形の角についての性質を基にして，多角形の角についての性質が見いだせることを知ること		○				
		【図形と証明】図形の合同について理解し図形についての見方を深めるとともに，図形の性質を三角形の合同条件などを基にして確かめ，論理的に考察し表現する能力を養う	平面図形の合同の意味及び三角形の合同条件について理解すること	○				○	
			証明の必要性と意味及びその方法について理解すること			○	○		○
			三角形の合同条件などを基にして三角形や平行四辺形の基本的な性質を論理的に確かめたり，図形の性質の証明を読んで新たな性質を見いだしたりすること		○				
	C数量関係	【一次関数】具体的な事象の中から二つの数量を取り出し，それらの変化や対応を調べることを通して，一次関数について理解するとともに，関数関係を見いだし表現し考察する能力を養う	事象の中には一次関数としてとらえられるものがあることを知ること	○					
			一次関数について，表，式，グラフを相互に関連付けて理解すること	○			○		
			二元一次方程式を関数を表す式とみること					○	○
			一次関数を用いて具体的な事象をとらえ説明すること	○		○			
		【確率】不確定な事象についての観察や実験などの活動を通して，確率について理解し，それを用いて考察し表現することができるようにする	確率の必要性と意味を理解し，簡単な場合について確率を求めること	○	○				
			確率を用いて不確定な事象をとらえ説明すること						

3. 尺度化のための理論的基礎

3.1　項目反応理論

テスト得点の項目依存性と項目困難度の標本依存性は，長い間，古典的テスト理論（Classical Test Theory, CTT）において解決を期待されてきた問題であった。100 点満点のテストにおける同じ 70 点という得点であっても，むずかしい項目から構成されるテストの 70 点とやさしい項目から構成されるテストの 70 点では表現される学力レベルが異なる。あるいは，正答率が同じ 20%の項目であっても，高い学力レベルの受検者集団における正答率 20%と低い学力レベルの受検者集団における 20%では表現される項目のむずかしさは異なる。このように，テスト得点で表される学力レベルはテストを構成する項目に依存するし，正答率で表される項目困難度はテストを受検した受検者集団に依存する。

項目反応理論（Item Response Theory, IRT）は，CTT が抱えてきたそれらの問題に対し，項目反応モデルという一つの解決策を与えた。IRT モデルは，一人の受検者が一つの項目に応答する際の正答確率をパラメトリックな関数としてモデル化したものである。IRT モデルには，受検者の学力を表す受検者母数とむずかしさなどの項目の特性を表す項目母数が含まれている。テストの結果から，項目に依存しない受検者母数と受検者集団に依存しない項目母数を推定することができる。このような IRT モデルの性質により，異なるテストを受検した受検者どうしの学力を同一尺度上で比較することが可能となる。その際，どんなテスト間でも学力を比較できるわけではなく，本調査研究のように，事前に使用目的に沿うようにテストを綿密に設計しておく必要がある。

すでに欧米では，IRT は学力調査をはじめ資格試験や適性試験などに広く利用されている。近年，日本でも IRT によるテスト運用が注目されつつあり，IRT によるテスト運用を利点の一つにあげる試験も多くなってきた。IRT によって運用されているテストの例として，NAEP（National Assessment of Educational Progress），LSAT（Law School Admission Test），PISA（Programme for International Student），TOEFL（Test of English as a Foreign Language），情報処理技術者試験，（医療系大学間）共用試験（医学系 CBT）などがあげられる。

参考までに，IRT 関連の書籍をいくつか紹介しておく．日本語でのIRTの入門書として豊田（2002），池田（1994），芝（1991）があげられる．言語テストの観点からIRT を平易に解説している本として大友（1996）があげられる．様々なIRT モデルの紹介や母数の推定ついては豊田（2005）が詳しい．洋書でのIRT の入門書としてEmbretson and Reise（2000）があげられる．Baker and Kim（2004）は，母数の推定法に関して非常に詳しい本である．IRT モデルについては，多くのモデルがvan der Linden and Hambleton（1996）に網羅されている．

3.2　2 母数ロジスティックモデル

本調査研究では，IRT モデルの代表格である 2 母数ロジスティックモデルを分析に用いる。2 母数ロジスティックモデルにおいて，尺度値 θ をもつ受検者が項目 j に正答する確率 $P_j(\theta)$ は，

$$P_j(\theta) = \frac{\exp[Da_j(\theta - b_j)]}{1+\exp[Da_j(\theta - b_j)]} \tag{3-1}$$

と表される。ここで，θ は潜在特性，能力母数あるいは被験者母数などと呼ばれ，テストで測定しようとしている何らかの構成概念に対応する。本調査研究の場合，学力テストを扱うことから，θ を受検者母数と呼ぶことにし，θ の具体的な値を尺度値と呼ぶことにする。また，本調査研究において測定しようとしている構成概念は各教科の学力である。θ の定義域は $-\infty < \theta < \infty$ であり，その値が大きいほどテストで測定しようとしている受検者の学力が高いことを示す。b_j は項目 j の項目困難度母数と呼ばれ，項目 j のむずかしさを表す。b_j の定義域は $-\infty < b_j < \infty$ であり，その値が大きいほど項目がむずかしいことを示す。a_j は項目 j の項目識別力母数と呼ばれ，$\theta = b_j$ 付近における受検者の尺度値のちがいがどのくらい敏感に正答確率に反映するかを表す。通常は $a_j > 0$ が仮定され，その値が大きいほど $\theta = b_j$ 付近の尺度値をもつ受検者をはっきりと識別できることを示す。D は尺度要素と呼ばれる定数であり，一般には式(3-1) で表わされるロジスティック曲線を正規累積曲線に近似させるために D=1.7 が用いられる。

式(3-1)において，受検者母数 θ を横軸に，項目 j への正答確率 $P_j(\theta)$ を縦軸にとったグラフで表される曲線を項目特性曲線（item characteristic curve, ICC）と呼ぶ。図 3-1 を見ると，2 母数ロジスティックモデルの ICC の特徴を知ることができる。まず，$\theta = b_j$ のとき正答確率がちょうど 0.5 になることがわかる．これは，式(3-1)の右辺に $\theta = b_j$ を代入したとき $P_j(\theta) = 0.5$ となることに対応する．また，項目 1 と項目 2 を比較すると，a_j の値を一定にしたまま b_j の値を変化させると ICC が平行移動することがわかる．さらに，項目 1 と項目 3 を比較すると，a_j の値が大きいほうが $\theta = b_j$ 付近の ICC の傾きが急峻になることがわかる．

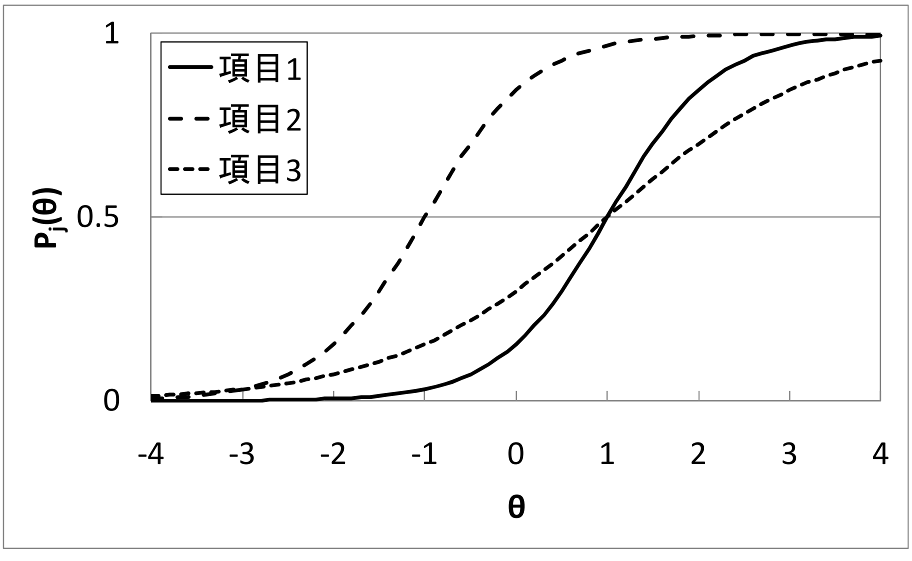


図 3-1 2 母数ロジスティックモデルの ICC の例

（注）項目 1 では a_1=1，b_1=1，項目 2 では a_2=1，b_2=−1，項目 3 では a_3=0.5，b_3=1 である．

IRT モデルは，いくつかの前提条件の上にモデルが構成されている。特に重要な前提条件は，局所独立の仮定と 1 次元性の仮定という概念である。局所独立の仮定とは，一つのテストにおいて，ある尺度水準をもつ受検者がある項目に正答する確率は他の項目に正答する確率の影響を受けないという仮定である。確率論的には，ある受検者が各項目に正答するのは互いに独立な事象であるということを意味する。1 次元性の仮定とは，一つのテストを構成する項目はただ一つの構成概念を測定するものでなければならないという仮定である。なお，2 母数ロジスティックモデルのように，受検者母数が 1 次元(1 変量)の IRT モデルでは,局所独立の仮定と 1 次元性の仮定とは同値である(Lord & Novick, 1968)。実際には，モデルを適用する際に，テストの 1 次元性だけを何らかの方法でチェックするのが一般的である。

テストの 1 次元性は，様々な方法で確認することができる（Hattie, 1985）。よく用いられる方法として，各項目間の四分相関係数行列における固有値のスクリープロットから判断する方法がある。例えば，テスト結果から図 3-2 のようなスクリープロットが得られたとする。図を見ると，第 1 固有値が突出しているとともに第 2 以下の固有値に格段の差を生じていることがわかる。このような傾向が確認できるとき，当該テストにおいて 1 次元性の仮定は満たされていると判断できる。これ以外の客観的な基準としては，各項目間の四分相関係数行列の第 1 固有値の分散説明率が 20%以上あることが Reckase（1979）により推奨されている。

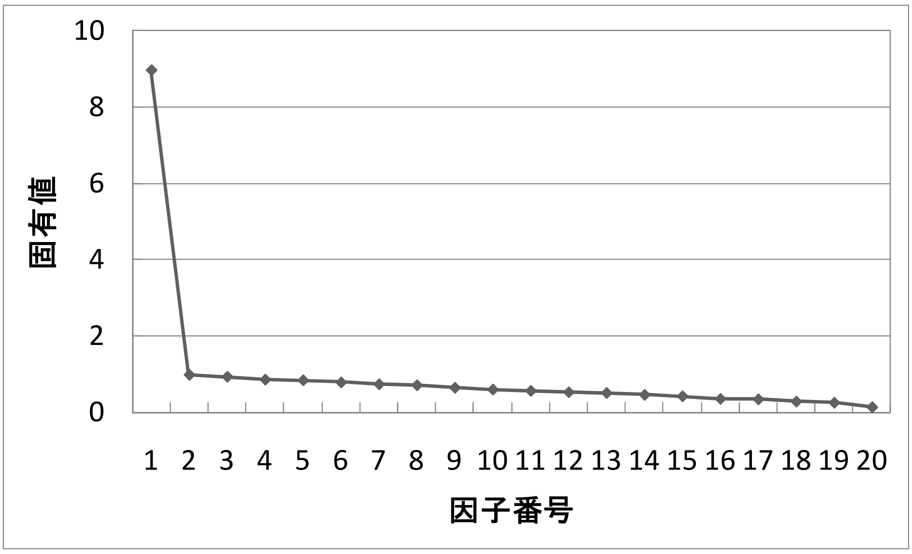


図 3-2 スクリープロットの例

3.3 項目情報関数とテスト情報関数

CTT では，信頼性係数を用いて受検者集団全体についての平均的な測定精度を見積もることができる。IRT では一歩進んで，尺度値 θ に応じた項目ごとの測定精度を表す指標を定義できる。その指標を項目情報関数（item information function）と呼び，

$$I_j(\theta) = \frac{P_j'^2(\theta)}{P_j(\theta)Q_j(\theta)} \tag{3-2}$$

と表せる。ここで，$P_j(\theta)$は尺度値θをもつ受検者が項目 j に正答する確率であり，2 母数ロジスティックモデルを利用したときには式(3-1)で表される。$Q_j(\theta) = 1 - P_j(\theta)$は誤答確率である。また，$P_j'(\theta)$は$\theta$についての$P_j(\theta)$の 1 次導関数である。2 母数ロジスティックモデルを利用する場合は，

$$I_j(\theta) = D^2 a_j^2 P_j(\theta)Q_j(\theta) \tag{3-3}$$

と計算できる。

式(3-2)や式(3-3)で表される項目情報量は加算可能である。テストを構成する項目の項目情報量を全て加算すると，そのテスト全体の情報量となる。項目情報関数の単純和をテスト情報関数（test information function）と呼び，

$$I(\theta) = \sum_{j=1}^{n} I_j(\theta) = \sum_{j=1}^{n} \frac{P_j'^2(\theta)}{P_j(\theta)Q_j(\theta)} \tag{3-4}$$

と表せる。もちろん，2 母数ロジスティックモデルを利用した場合は式(3-3)の単純和となる。

図 3-3 に，前節の図 3-1 で利用した 3 項目の項目情報関数を示す。図 3-4 に，その 3 項目からなるテストを作成した場合のテスト情報関数を示す。図 3-3 を見ると，項目情報関数は尺度値と項目困難度が等しい$\theta = b_j$で最大値をとることや，項目識別力a_jが大きいほど鋭いピークをもつことがわかる。図 3-4 を見ると，テスト情報量はテストを構成する項目の項目情報量の単純和になっている様子がうかがえる。

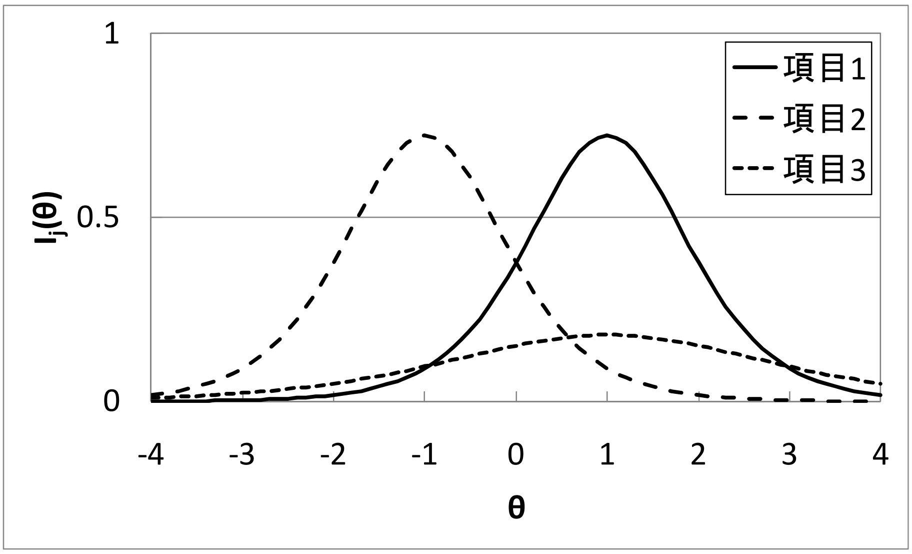


図 3-3 項目情報関数の例

（注）項目 1 では a_1=1，b_1=1，項目 2 では a_2=1，b_2=−1，項目 3 では a_3=0.5，b_3=1 である．

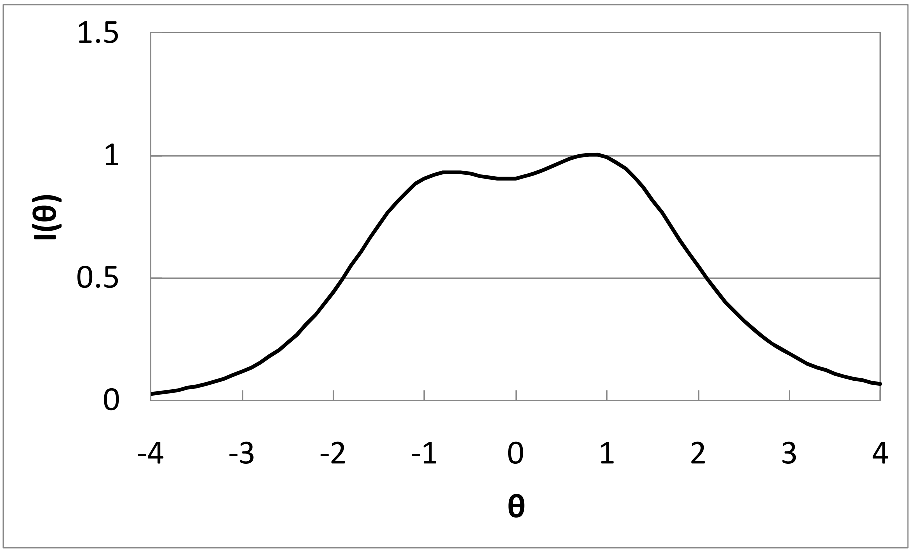


図 3-4 テスト情報関数の例

（注）テスト項目は，図 3-1 および図 3-3 と同じ。

式(3-4)のテスト情報量は，一般的な統計学の分野における Fisher 情報量と同一のものであり，テスト全体の測定精度と密接な関係がある。テスト情報量を用いると，次節で概説する θ の最尤推定値の標準誤差（ばらつき）を$1/\sqrt{I(\theta)}$で見積もることができる（図 3-5）。すなわち，テスト情報関数をみれば，テストがどの付近の尺度値をどのくらい正確に測定できるのかがわかる。テスト情報量の大きい尺度値レベルが尺度値をより正確に測定できる部分であり，テスト情報量の小さい尺度値レベルが尺度値の測定精度が低くなる部分である。例えば，図 3-4 を見ると，当該テストはその受検者集団において平均的な尺度値レベルをもつ受検者の尺度値を他の尺度値より正確に測定できたことが読み取れる。このような尺度値ごとの測定精度は，CTT では得られない情報であり，CTT に対する IRT の利点の一つになっている。

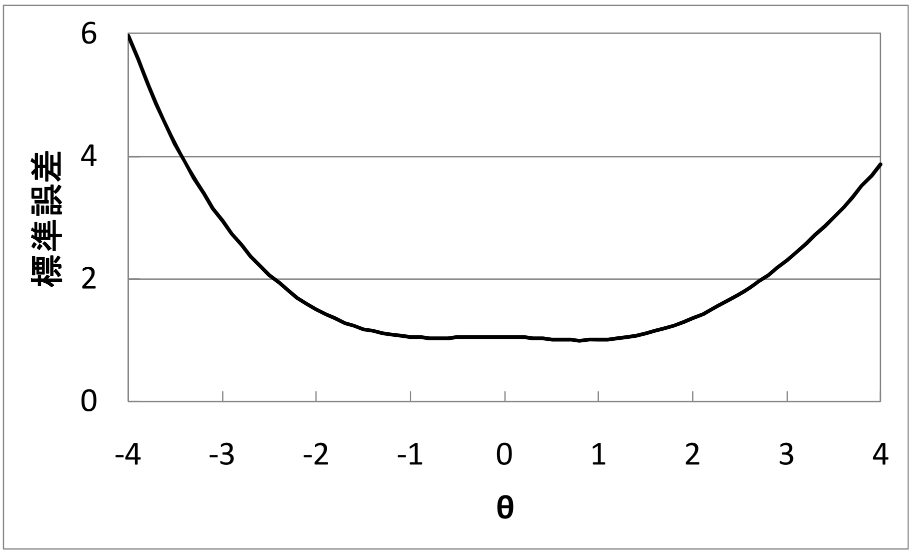


図 3-5 標準誤差の例

（注）テスト項目は，図 3-1 および図 3-3 と同じ。

3.4　IRT 分析の方法

多肢選択式のテストなどでは，テスト結果を 2 値データ（1: 正答，0: 誤答）で表現できる。例えば，10 人の受検者が 5 項目からなるテストを受検したとすると，図 3-6 に示すような 10 行 5 列の行列が得られる。i 行 j 列の要素は，受検者 i が項目 j に正答したか誤答したかを表している。このような 2 値の行列データは，項目反応データあるいは項目反応パターンなどと呼ばれる。本調査研究では，n 項目からなるテスト X を m 人の受検者が受検したときの項目反応データを X = {x_{ij}: i = 1, 2, …, m; j = 1, 2, …, n}と記述することにする。

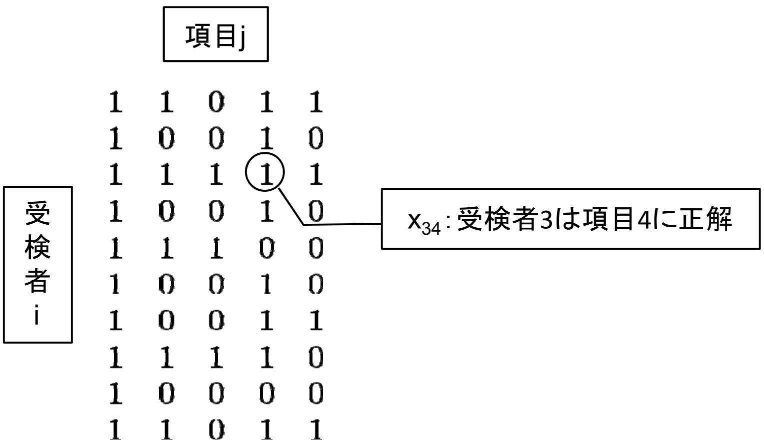


図 3-6 項目反応データの例

テストを IRT 分析するには，テストの結果として得られた項目反応データから IRT モデルに含まれ

る母数を推定する必要がある。その際，項目母数と受検者母数は別々に二段階で推定するのが現在の主流である。本調査研究においても，項目母数を周辺最尤推定法（marginal maximum likelihood estimation，MMLE）によって推定したあと，得られた項目母数の推定値を利用して受検者母数を最尤推定法（maximum likelihood estimation，MLE）によって推定するという手順を踏む。IRTモデルの母数の推定は非常に煩雑であるため，母数を推定するための専用ソフトウェアがいくつか開発されている．本調査研究では，IRT モデルの母数推定に世界中で広く利用されている BILOG-MG（Zimowski，Muraki，Mislevy，& Bock，2003）を主として利用する．

以下の節では，項目母数が所与のときに受検者母数を最尤推定する方法と項目母数を周辺最尤推定する方法について順に概説する。

3.5　受検者母数の推定

BILOG-MGでは，受検者母数の推定法として，最尤推定法（maximum likelihood estimation，MLE），最大事後（maximum a posteriori，MAP）推定法，期待事後（expected a posteriori，EAP）推定法の三つが提供されている。本節では，項目母数が所与のときに受検者母数を最尤推定する方法を中心に概説する。

受検者 i が n 項目からなるテストを受検したとき，その解答パターンが $x_i = (x_{i1}, x_{i2}, \cdots, x_{in})$となる確率は，局所独立の仮定から，

$$L(x_i|\theta_i) = \prod_{j=1}^{n} P_j(\theta_i)^{x_{ij}} Q_j(\theta_i)^{1-x_{ij}} \tag{3-5}$$

と表せる。ここで，$P_j(\theta_i)$は尺度値θ_iをもつ受検者 i が項目 j に正答する確率であり，2 母数ロジスティックモデルを利用したときには式(3-1)で表される。また，$Q_j(\theta_i) = 1 - P_j(\theta_i)$は誤答確率である。

本来，式(3-5)は，受検者母数θ_iが所与のときに解答パターン x_iが得られる確率を表している。しかし，実際に尺度値を推定するには，逆に解答パターン x_iから受検者母数θ_iを推定しなければならない。そこで，変数間の関係を逆にとらえ，式(3-5)を解答パターン x_iが所与のときに受検者母数θ_iが取り得る値のもっともらしさを表す関数と考える。このとき，式(3-5)は受検者母数の尤度関数と呼ばれる。尤度関数を最大化する独立変数の組を求めることを最尤推定するといい，このとき得られる推定値を最尤推定値という。

式(3-5)の両辺の対数をとっても最大値をとるθ_iは変化しないので，計算を容易にするために式(3-5)の両辺の（自然）対数をとれば，

$$\log L(x_i|\theta_i) = \sum_{j-1}^{n}[x_{ij}\log P_j(\theta_i) + (1 - x_{ij})\log Q_j(\theta_i)] \tag{3-6}$$

と対数尤度関数が計算できる。この対数尤度関数を最大化するためには，式(3-6)を θ_i に関して 1 階偏微分した式を 0 とおいた尤度方程式を θ_i に関して解けばよい。式(3-6)の 1 次偏導関数は，

$$\begin{aligned} \frac{\partial \log L(x_i|\theta_i)}{\partial\theta_i} &= \frac{\partial}{\partial\theta_i}\sum_{j=1}^{n}[x_{ij}\log P_j(\theta_i) + (1 - x_{ij})\log Q_j(\theta_i)] \\ &= \sum_{j-1}^{n}\frac{x_{ij}}{P_j(\theta_i)}\frac{\partial P_j(\theta_i)}{\partial\theta_i} + \sum_{j-1}^{n}\frac{1 - x_{ij}}{Q_j(\theta_i)}\frac{\partial Q_j(\theta_i)}{\partial\theta_i} \end{aligned} \tag{3-7}$$

と計算できる。式(3-7)の $P_j(\theta_i)$ が 2 母数ロジスティックモデルであれば，

$$\frac{\partial P_j(\theta_i)}{\partial\theta_i} = Da_jP_j(\theta_i)Q_j(\theta_i) \tag{3-8}$$

と 1 次偏導関数が計算できる。このシンプルな偏導関数の表現が IRT モデルとしてロジスティック関数を用いる利点の一つである。

実際に，式(3-7)を 0 とおいた尤度方程式は，非線形方程式となるために解析的には解くことができない。コンピュータなどを用いて数値的に解かざるを得ず，その際に Newton-Raphson 法や Fisher のスコアリング法がよく利用される。Newton-Raphson 法では，反復計算を繰り返して θ_i の推定値を計算する。t+1 回目の反復計算における更新式は，

$$[\hat{\theta}_i]_{t+1} = [\hat{\theta}_i]_t - [\frac{\partial^2 \log L(x_i|\theta_i)}{\partial\theta_i^2}]_t{}^{1}[\frac{\partial L(x_i|\theta_i)}{\partial\theta_i}]_t \tag{3-9}$$

で与えられる。初期値としては，n 項目からなるテストにおいて受検者 i の正答数得点を X とするとき，$[\theta_i]_0 = \ln[X/(n - X)]$にとるとよい（ln は自然対数）。2 母数ロジスティックモデルを利用する場合，式(3-9)の 2 次偏導関数は，

$$\frac{\partial^2 \log L(x_i|\theta_i)}{\partial\theta_i^2} = \sum_{j=1}^{n}D^2a_j^2P_j(\theta_i)Q_j(\theta_i) \tag{3-10}$$

とシンプルな式となる。式(3-10)の詳しい導出過程については，豊田（2005）に詳しい。また，Fisher のスコアリング法では，式（3-10）の代わりに Fisher の情報関数を式(3-9)の右辺第 2 項に当てはめて反復計算する。Fisher 情報関数は，式(3-10)の期待値にマイナスの符号をつけたものである。2 母数ロジスティックモデルの場合，式(3-10)と Fisher 情報関数は一致するので，見かけ上，

Newton-Raphson 法と Fisher のスコアリング法は区別がつかない。なお，Newton-Raphson 法や Fisher のスコアリング法については，Kendall and Stuart (1979)などが参考になる。

受検者母数の最尤推定値の漸近的な標準誤差は，式(3-4)のテスト情報量を用いて，

$$SE_{\theta_i} = [I(\theta_i)]^{-\frac{1}{2}} \tag{3-11}$$

と計算できる。2 母数ロジスティックモデルの場合は，

$$SE_{\theta_i} = [D^2 a_j^2 P_j(\theta_i) Q_j(\theta_i)]^{-\frac{1}{2}} \tag{3-12}$$

とシンプルな式となる。

なお，テストの項目数 n が大きくなるほど，受検者母数の最尤推定値は受験者の真の尺度値に近づき，その標準誤差も式(3-11)の値に近づく。この漸近的性質は，項目数が 20 以上の場合に実用的な意味で利用可能であることが知られている（豊田，2002）。

BILOG-MG では，受検者母数の最尤推定値を求めるのに Fisher のスコアリング法が採用されている。デフォルトの設定では，事前にコマンドファイルに設定した反復回数の最大値に達するか，全ての θ_i（i = 1, 2, …, m）において t 回目と t+1 回目の反復における推定値の差が 0.01 未満になったときに反復計算は終了するようになっている。推定結果を利用する際には，BILOG-MG の出力結果をよく見て推定値が収束したものであることを確認する必要がある。

最尤推定法は，解答パターン x_i によっては推定プロセスが収束しない場合がある。また，受検者が全問正解あるいは全問不正解の場合は推定値を求めることができない。それに対し，MAP 推定法と EAP 推定法ではどんな解答パターンでも対応する推定値を求めることができる。この点では，MAP 推定法と EAP 推定法のほうが最尤推定法より優れているようにみえる。しかし，最尤推定値が発散するのは，解答パターンに何らかの検討すべき課題が含まれている可能性があるとも考えられる。MAP 推定法や EAP 推定法を用いると，問題のある解答パターンが見過ごされてしまう危険性は否定できない。本調査研究では，各推定法の長所と短所を総合的に考慮し，受検者母数の推定法として最尤推定法を利用することにした。

3.6 項目母数の推定

BILOG-MGにおける項目母数の推定には，ベイズ統計を応用した周辺最大事後推定法（marginal maximum a posteriori estimation, MMAPE）が利用可能である。その際，項目母数の推定値はMAP（maximum a posteriori）推定値として求められる。本調査研究では，BILOG-MGの”SCORE”コマンドのオプショ

ンに”NOTPRIOR”と”NOSPRIOR”を設定し，項目困難度母数と項目識別力母数に事前分布を設定しない。したがって，本調査研究における項目母数の推定法は，周辺最尤推定法（marginal maximum likelihood estimation, MMLE）とみなすことができる。本節では，テスト結果の項目反応データから，項目母数を周辺最尤推定する方法について概説する。

いま，m 人の受検者が n 項目からなるテストを受験し，図 3-6 に示すような項目反応データ $X = \{x_{ij}: i = 1, 2, \cdots, m; j = 1, 2, \cdots, n\}$ が得られたとする。また，2 母数ロジスティックモデルなどの IRT モデルにより，尺度値 θ_i をもつ受検者 i が項目 j に正答する確率が $P_j(\theta_i)$ で与えられているとする。このとき，項目反応データ X が得られる確率は，

$$P(X|\theta,\xi) = \prod_{i=1}^{m}\prod_{j=1}^{n} P_j(\theta_i)^{x_{ij}} Q_j(\theta_i)^{1-x_{ij}} \tag{3-13}$$

と表せる。ここで，$\theta = (\theta_1, \theta_2, \cdots, \theta_m)$ は受検者母数ベクトルである。$\xi = (\xi_1, \xi_2, \cdots, \xi_n)$ は項目母数ベクトルであり，2 母数ロジスティックモデルを利用する場合には $\xi_j = (b_j, a_j)$ である。また，$Q_j(\theta_i) = 1 - P_j(\theta_i)$ は誤答確率である。

式(3-13)を θ と ξ の同時尤度関数とみなし，その同時尤度関数を最大にするような (θ, ξ) を最尤推定する方法は，同時最尤推定法（joint maximum likelihood estimation, JMLE）と呼ばれる。統計的推測における母数の推定値として，データのサンプル数を増やすにつれて推定値の誤差分散は小さくなり推定値が真の値に近づいていくような一致推定量が求められるのが通常である。Neyman and Scott（1948）は，構造母数（項目母数）が付随母数（受検者母数）と同時に推定されるときには，構造母数（項目母数）は一致推定量にならないことを示した。そのため，JMLE には，受検者の数を増やしても項目母数が一致推定量にならないという好ましくない性質がある（Andersen, 1972）。このような理由から，現在ではJMLEはほとんど用いられていない。

Bock and Lieberman（1970）は，項目母数を一致推定量として推定可能な周辺最尤推定法（marginal maximum likelihood estimation, MMLE）を提案した。MMLE は，個々の被験者の尺度値を必要とせず，被験者数を増やせば一致推定量として項目母数を推定することができる。JMLE で一致推定量が得られないのは，受検者母数と項目母数が同時に推定されるからである。MMLE では，受検者母数に正規分布などの分布を仮定し，式(3-13)から局外母数となる受検者母数を積分消去して得られる項目母数の周辺尤度関数を最大化する。

受検者 i が n 項目からなるテストを受検したとき，その解答パターンが $x_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ となる確率は，局所独立の仮定から，

$$P(x_i|\theta_i,\xi) = \prod_{j=1}^{n} P_j(\theta_i)^{x_{ij}} Q_j(\theta_i)^{1-x_{ij}} \tag{3-14}$$

と表せる。ただし，$\xi = (\xi_1, \xi_2, \cdots, \xi_n)$は項目母数ベクトルであり，2母数ロジスティックモデルを利用する場合には$\xi_j = (b_j, a_j)$である。$P_j(\theta_i)$は受検者 i が項目 j に正答する確率であり，$Q_j(\theta_i) = 1 - P_j(\theta_i)$は誤答確率である。

MMLE では，受検者集団の尺度値は母集団分布$g(\theta | \tau)$からのランダムサンプルであると仮定する。ここで，τは平均や分散など分布の形を決める母数ベクトルである。局外母数を取り除くため，$g(\theta | \tau)$を用いて式(3-14)から受検者母数を積分消去すると，

$$P(x_i|\xi,\tau) = \int_{-\infty}^{\infty} P(x_i|\theta_i,\xi)g(\theta_i|\tau)d\theta_i \tag{3-15}$$

という解答パターンx_iの周辺確率が得られる。簡便のため，これ以降は式(3-15)を$P(x_i)$と略記すれば，全ての解答パターン（項目反応データ）Xが得られる確率は，

$$L = \prod_{i=1}^{m} P(x_i) \tag{3-16}$$

と表せる。MMLEでは，式(3-16)を項目母数についての周辺尤度関数と呼び，その尤度関数が最大値をとるときの項目母数を推定値として計算する。

計算を容易にするために式(3-16)の両辺の（自然）対数をとれば，

$$\log L = \sum_{i-1}^{m} \log P(x_i) \tag{3-17}$$

と周辺対数尤度関数が計算できる。この周辺対数尤度関数を最大化するためには，式(3-17)を項目母数に関して1階偏微分した式を0とおいた尤度方程式を項目母数に関して解けばよい。尤度方程式の導出はかなり繁雑になるため，ここでは結果のみを記しておく。詳しい導出過程については，Baker and Kim（2004）や豊田（2005）を参照していただきたい。2母数ロジスティックモデルを用いた場合，a_jとb_jに関する尤度方程式は，

$$\frac{\partial}{\partial a_j}(\log L) = \sum_{i=1}^{m} \int_{\infty}^{\infty} [x_{ij} - P_j(\theta_i)](\theta_i - b_j)[P(\theta_i|x_i,\xi,\tau)]d\theta_i = 0 \tag{3-18}$$

$$\frac{\partial}{\partial b_j}(\log L) = -a_j \sum_{i-1}^{m} \int_{-\infty}^{\infty} [x_{ij} - P_j(\theta_i)][(\theta_i|x_i,\xi,\tau)]d\theta_i = 0 \tag{3-19}$$

となる。

推定値を求めるためには，式(3-18)と式(3-19)を連立方程式としてa_jとb_jについて数値的に解けばよい．その際，Newton-Raphson法やFisherのスコアリング法を用いればよいものの，式内の積分の計算と2次偏微分の具体的な計算が困難といった問題がある。それらの問題は，区分求積法と数値微分を利用して解決可能である。

区分求積法を利用すれば，積分に含まれる連続型の分布を離散的な分布で近似して関数の評価を簡単にすることが可能である。ここでは，θ軸上に求積点X_h(h = 1, 2, …, H)をとり，母集団分布$g(\theta|\tau)$を近似する。母集団分布として標準正規分布を利用する場合は，$-4 \leqq \theta \leqq 4$を対象として等間隔に求積点を区切ってもよいし，さらに効率的な方法としてHermite-Gauss求積法を用いてもよい。なお，$A(X_h)$はX_hの近傍における関数$g(\theta|\tau)$の値である。このとき，積分を含んでいた式(3-18)と式(3-19)はそれぞれ，

$$a_j: \sum_{h=1}^{H}\sum_{i-1}^{m}[x_{ij} - P_j(X_h)](X_h - b_j)[P(X_h|x_i,\xi,\tau)] = 0 \tag{3-20}$$

$$b_j: -a_j\sum_{h=1}^{H}\sum_{i=1}^{m}[x_{ij} - P_j(X_h)][P(X_h|x_i,\xi,\tau)] = 0 \tag{3-21}$$

と近似できる。また，$P(X_h|x_i, \xi, \tau)$は，式(3-14)とベイズの定理より，

$$P(X_h|x_i,\xi,\tau) = \frac{\prod_{j=1}^{n} P_j(X_h)^{x_{ij}} Q_j(X_h)^{1-x_{ij}} A(X_h)}{\sum_{h'=1}^{H}\prod_{j=1}^{n} P_j(X_{h'})^{x_{ij}} Q_j(X_{h'})^{1-x_{ij}} A(X_{h'})} \tag{3-22}$$

と書き換えられる。

Newton-Raphson法やFisherのスコアリング法に必要な式(3-18)と式(3-19)の2次偏導関数は，数値微分のテクニックを利用することで近似可能である。例えば，

$$\frac{\partial^2}{\partial a_j \partial b_j}(\log L) \approx \frac{\frac{\partial \log L(b_j+\Delta b_j)}{\partial a_j} - \frac{\partial \log L(b_j)}{\partial a_j}}{\Delta b_j} \tag{3-23}$$

と計算できる。Δa_jやΔb_jとしては，0.001程度の微小な値を用いればよい。

以上で，式(3-18)と式(3-19)を数値的に解く準備が整った。尤度方程式をNewton-Raphson法で解く場合，反復計算のための更新式は次のとおりである。

$$\lambda_{t+1} = \lambda_t - H_t^{-1} g_t \tag{3-24}$$

ここで，λ_tはサイズ2n×1の項目母数ベクトルである。また，g_tはサイズ2n×1の1次偏導関数ベクトル，H_tはサイズ2n×2nの2次偏導関数行列（Hessian行列）である。Fisherのスコアリング法で解く場合は，Hessian行列の代わりにFisherの情報関数行列を利用すればよい。Fisher情報関数行列は，Hessian行列の期待値にマイナスの符号をつけた行列として計算できる。なお，初期値としては，$[a_j]_0 = \rho_{jX}/\sqrt{(1-\rho_{jX}^2)}$ および$[b_j]0 = -z_j/\rho_{jX} = -\Phi^{-1}(p_j)/\rho_{jX}$が利用できる。ここで，$\rho_{jX}$はテスト得点Xと項目jとの双列相関係数（biserial correlation coefficient），z_jは標準正規分布上において下側面積が項目正答率p_jに等しくなるときの横軸の値である。

Newton-Raphson法にせよ，Fisherのスコアリング法にせよ，式(3-24)における逆行列の計算にもっとも大きな計算負荷がかかる。Bock and Lieberman（1970）の当時，まだコンピュータが今ほど発達していなかった時代においては，MMLEは項目数が少ないときにしか利用できない推定法であった。Bock and Lieberman（1970）はJMLEの一致推定量に関する問題を解決したものの，テストデータ解析の実用における問題の解決は，Bock and Aitkin（1981）のEMアルゴリズムを用いたMMLEの開発まで待たなければならなかった。

Bock and Aitkin（1981）は，EMアルゴリズムを用いて項目母数を推定する方法を開発した。EMアルゴリズムの統計学的研究は，すでにDempster，Laird，and Rubin（1977）においてなされていた。EMアルゴリズムは，潜在変数が含まれる確率モデルの母数を最尤推定法に基づいて推定する手法の一つである。EMアルゴリズムでは，Eステップ（expectation step）とMステップ（maximization step）を交互に繰り返すことで計算が進められる。Eステップでは期待対数尤度を求め，Mステップではその対数尤度を最大化するような項目母数の値が計算される。Bock and Lieberman（1970）のMMLEと異なる部分は，尤度方程式を解く前に式(3-25)と式(3-26)で表される期待度数を求めるステップを挿入すること，および尤度方程式が期待度数を用いて書き換えられていることである。

[Eステップ]

項目母数の仮の値（反復1回目は初期値，それ以降は前回のMステップで求められた値）を与え，求積点 X_h (h = 1, 2, …, H)において項目jに解答する期待人数f_{jh}とその中での期待正答者数r_{jh}を計算する。

$$f_{jh} = \sum_{i=1}^{m} P(X_h|x_i,\xi,\tau) = \sum_{i=1}^{m}[\frac{L_i(X_h)A(X_h)}{\sum_{h'=1}^{H} L_i(X_{h'})A(X_{h'})}] \tag{3-25}$$

$$r_{jh} = \sum_{i=1}^{m} x_{ij}P(X_h|x_i,\xi,\tau) = \sum_{i=1}^{m}[\frac{x_{ij}L_i(X_h)A(X_h)}{\sum_{h'=1}^{H} L_i(X_{h'})A(X_{h'})}] \tag{3-26}$$

[Mステップ]

Eステップで計算したf_{jh}とr_{jh}の値を式(3-27)と式(3-28)に代入し，Newton-Raphson法やFisherのスコアリング法を用いて尤度方程式を数値的に解く。収束基準を満たす場合は終了し，満たさない場合はEステップに戻る。

$$a_j:\ \sum_{h=1}^{H}(X_h - b_j)[r_{jh} - f_{jh}P_j(X_h)] = 0 \tag{3-27}$$

$$b_j:\ -a_j\sum_{h=1}^{H}[r_{jh} - f_{jh}P_j(X_h)] = 0 \tag{3-28}$$

なお，推定された項目母数の標準誤差は，推定が終了した時点でのFisher情報関数行列の逆行列における対角要素の平方根として求められる。Fisher情報関数行列は，式(3-24)で登場したHessian行列の期待値にマイナスの符号をつけた行列-E[H]として計算できる。

Bock and Aitkin (1981) のEMアルゴリズムを用いた周辺最尤推定法 (marginal maximum likelihood estimation with EM algorithm, MMLE-EM) は，本調査研究で利用するBILOG-MGをはじめ，PARSCALE, TESTFACT, EasyEstimation (熊谷, 2009) などのIRT専用ソフトウェアで利用できる。BILOG-MGのデフォルトの設定では，尺度値の母集団分布g(θ|τ)には標準正規分布を仮定している。また，区分求積法における求積点の数Hについては，受検者集団の尺度値が単一母集団からのサンプルとみなせる場合は10，多母集団からのサンプルとみなせる場合は20を利用している。また，Mステップの反復計算は，事前にコマンドファイルに設定した反復回数の最大値に達するか，全ての項目母数においてt回目とt+1回目の反復における推定値の差が0.01未満になったときに終了するようになっている。受検者母数の推定の場合と同様に，推定結果を利用する際には，BILOG-MGの出力結果をよく見て推定値が収束したものであることを確認する必要がある。

4.　実施手続

4.1 実施の基本方針

全国的な学力調査実施において参加児童・生徒への実施負担を軽減するため，PISA 等の国際的な学力調査ではすでに採用されている新しい調査方法である「重複テスト分冊法」の日本における適用ノウハウを確立することを目的とした実施内容とする。

あわせて参加協力校へ学習指導に役立つ情報を返却するとともに，平成 21 年度全国学力調査抽出校と希望利用校との集団比較を行うことも副次的な目的とする。

4.2　実施内容

対象学年　新潟市の小学６年生・中学３年生
対象人数　小学校 6 年生：2487 名　　中学 3 年生：2394 名
協力校数　小学校　39 校　　中学校　22 校
実施方式　学校単位
抽出校は新潟市教育委員会の指定による
実施時期　平成 22 年 10 月 18 日～22 日
実施時間　上記の期間内の 1 時限
（説明や配布，回収の時間 5 分を含めて，小学校 40 分，中学校 45 分）
出題教科　算数・数学
出題範囲　現行の学習指導要領の内容構成にあわせ，
小学校：第 4 学年と第 5 学年で学習する 4 領域
Ａ数と計算
Ｂ量と測定
Ｃ図形
Ｄ数量関係
中学校：第 1 学年と第 2 学年で学習する 4 領域
Ａ数と式
Ｂ図形
Ｃ数量関係
Ｄ資料の活用の
問題内容　国立教育政策研究所のチェック済み
分冊数　　基本的には 5 分冊,ただし分冊の順序効果を見るためにブロックを入れ替えたものを分冊ごとに準備したため実質は 10 分冊であった（表 1-2 を参照）。また児童・生徒がどの分冊を解くかはランダムになるよう割り当てた。ブロック毎の項目数については表 3-1 のとおり。

表 4-1 ブロックごとの項目数の内訳

ブロック	小学校	中学校
共通	１５	１６
1	9	8
2	9	8
3	9	8
4	9	8
5	9	8
計	６０	５６

解答数　　　小中学校ともに一人当たり 30 問程度　問題冊子は回収

解答方式　　客観式テスト（手書き解答をコード化した）

個票返却　　平成 22 年 12 月初旬　各学校へ郵送により返却済み

4.3 個票内容(主なもの)

1)　解答した分冊に含まれている問題についての正誤情報

2)　割り当てられなかった分冊に含まれている問題については推定正答確率の値

3)　全ての問題に関して学習指導要領の項目事項を記載

詳細内容は表 4-2 および表 4-3 を参照のこと。実際には児童生徒向けには「個票」ということばではなくて，それぞれ「算数調査結果シート」「数学調査結果シート」という名称を用いた。

4.4 教員向け説明内容

個票を利用する際の注意事項として，教員向けにテスト・デザインの説明の後，以下のような内容の解説文書を配布した。

さらに，今回の調査では，最新のテスト理論を利用して，ある生徒が受けなかった残りのブロック，例えば分冊１を受けた生徒を例にとれば，その生徒がもしブロック３～５の問題を受けたとしたらどの程度の確率でその問題に正しく答えられるかという推定正答確率も計算しています。その生徒は実際には 32 問しか解いていませんが，残りの問題についての情報も得られるということになり，その生徒の数学に関する学力の全体像を把握できるようになっています。

このように，今回の調査は，従来の学力調査のように，児童・生徒が解答する問題は全て同一であるという構成ではありません。そうすると，「解答していない問題があるのだから，その児童・生徒個人の学力は総合的にはわからないのではないか」と思われるかもしれません。しかし，項目反応理論という最新のテスト理論を応用することにより，実際は受けていない問題項目の正答確率を推定することがで

きます。この推定正答確率を見ることによって，出題されていなかった領域の学力（理解度）の概要を把握することが可能になります。

＜平均正答率の見方＞

もし「平均正答率」が 70％の問題すなわち易しい問題で×になっている，という場合，その問題の単元・内容の理解が不十分であることを表しています。反対に，もし「平均正答率」が 30％の非常に難しい問題で○になっているという場合，その問題の単元・内容の理解がよくなされていると解釈することができるでしょう。

＜推定正答確率の意味＞

個人票の「推定正答確率」は，データ分析によって算出された個人の「学力」から，「その個人が，解答していない問題に対してどれくらいの確率で正答できそうか」を推定したものです。したがって，個人ごとに，この「推定正答確率」は異なります。

ただし，このテスト冊子の全ての問題に正解した人あるいは全て不正解だった人の学力は技術的に推定することはできません。

＜推定正答確率の使い方・見方＞

ある領域・内容の「推定正答確率」を見るときは，「推定正答確率が 30％だから，この領域・内容の理解が不十分である」「80％だから十分である」というような絶対的な正答率の見方ではなく，「平均正答率」との比較でみてください。

「推定正答確率」は，正誤データをもとに算出されたその児童・生徒の「学力」によって，「実際には解いていない問題にどれだけの確率で正答できそうか」を推定したものですので，その個人の学力が全体の学力の平均に比べて“高め”であると，「推定正答確率」も実際の「平均正答率」よりも高く出ている傾向があります。

表 4-2 個票（小学生）

算　数　調　査　結　果　シ　ー　ト

学校番号		学校名		組		出席番号		性別	

学年	領　域	単　　元	内　　容	新潟市の平均正答率	解いた問題の正誤	解かなかった問題の推定正答確率
第4学年	数と計算	整数の表し方	億，兆の単位	69%	-	74%
		概数と四捨五入	概数が用いられる場合	78%	×	-
			四捨五入	63%	○	-
		整数の除法	除法の計算を用いること	71%	-	81%
			被除数，除数，商及び余りの間の関係	59%	×	-
				40%	-	43%
			除法について成り立つ性質	87%	-	93%
		小数の仕組みとその計算	小数の意味と表し方	88%	-	94%
			小数の仕組みと数の相対的な大きさ	85%	-	93%
			小数の加法，減法	76%	○	-
				85%	○	-
		分数の意味とその表わし方	分数の意味と表し方	97%	○	-
				62%	○	-
			単位分数	77%	-	90%
	量と測定	面積の単位と測定	面積の単位（cm2）	31%	-	31%
			正方形，長方形の面積の求め方	84%	○	-
				31%	-	31%
		角の大きさ	回転の大きさ、角の大きさの単位	86%	○	-
				51%	-	57%
	図　形	二等辺三角形，正三角形などの図形	二等辺三角形，正三角形	61%	-	66%
				59%	○	-
			円，球	36%	-	37%
				61%	-	66%
	数量関係	伴って変わる二つの数量	変化の様子と折れ線グラフ	86%	○	-
		数量の関係を表す式	四則の混合した式や（　）を用いた式	55%	-	61%
				26%	○	-
		資料の分類整理	二つの観点から分類整理すること	61%	×	-
				78%	-	87%
			折れ線グラフの読み方とかき方	86%	○	-
第5学年	数と計算	整数の性質	偶数，奇数	56%	-	60%
		整数，小数の記数法	10倍，100倍，1/10，1/100などの大きさ	70%	○	-
		小数の乗法，除法	小数×整数，小数÷整数の意味と計算	94%	○	-
				78%	-	83%
			小数の乗法，除法の意味	20%	○	-
			小数の乗法，除法の計算，余りの大きさ	79%	○	-
				70%	-	78%
				79%	-	84%
		分数の意味と加法，減法	大きさの等しい分数	72%	○	-
			整数，小数と分数の関係	79%	-	90%
			除法の結果と分数	84%	○	-
			同分母分数の加法，減法	97%	○	-
	量と測定	図形の面積	三角形，平行四辺形の面積の求め方	48%	×	-
				52%	×	-
				55%	-	61%
				75%	-	86%
				70%	×	-
			ひし形，台形の面積の求め方	54%	-	60%
			円の面積	73%	○	-
	図　形	平行四辺形，ひし形，台形	直線の平行や垂直の関係	93%	○	-
			平行四辺形，ひし形，台形	88%	○	-
				91%	○	-
				91%	○	-
				51%	-	56%
			多角形，正多角形	36%	-	37%
			図形の性質	53%	-	56%
				82%	-	92%
			円周率	58%	○	-
			立方体，直方体	58%	○	-
	数量関係	四則に関して成り立つ性質	交換法則，結合法則，分配法則	69%	○	-
		百分率		43%	○	-
				33%	×	-
				37%	-	39%
		数量の関係を表す式		85%	-	95%
				44%	-	49%

表の見方
1）「平均正答率」はこの調査を受けた人たち全体のなかでその問題に正しく答えられた人たちの割合です。数値が小さいほどむずかしい問題です。
2）○と×はあなたが実際に解いた問題の正解・不正解をあらわしています。
3）解かなかった問題の「推定正答確率」は、あなたが受けたテスト冊子の中にはその問題は含まれていないのですが、仮にこの問題を解いていたらどの程度正しく答えられるかを示したものです。

表 4-3 個票（中学生）

数　学　調　査　結　果　シ　ー　ト

学校番号		学校名		組		出席番号		性別	

学年	領域	単　　元	内　　容	新潟市の平均正答率	解いた問題の正誤	解かなかった問題の推定正答確率
第1学年	数と式	【正負の数】具体的な場面を通して正の数と負の数について理解し，その四則計算ができるようにするとともに，正の数と負の数を用いて表現し考察することができるようにする	正の数と負の数の必要性と意味を理解すること	86%	○	-
				70%	×	-
			小学校で学習した数の四則計算と関連付けて，正の数と負の数の四則計算の意味を理解すること	90%	○	-
				65%	○	-
			正の数と負の数の四則計算をすること	90%	○	-
			具体的な場面で正の数と負の数を用いて表したり処理したりすること	77%	-	92%
				90%	○	-
		【文字と式】文字を用いて数量の関係や法則などを式に表現したり式の意味を読み取ったりする能力を培うとともに，文字を用いた式の計算ができるようにする	文字を用いることの必要性と意味を理解すること	72%	-	94%
			文字を用いた式における乗法と除法の表し方を知ること	90%	○	-
			簡単な一次式の加法と減法の計算をすること	61%	○	-
				85%	○	-
				75%	×	-
				55%	-	83%
			数量の関係や法則などを文字を用いた式に表すことができることを理解し，式を用いて表したり読み取ったりすること	73%	-	92%
		【一次方程式】方程式について理解し，一元一次方程式を用いて考察することができるようにする	方程式の必要性と意味及び方程式の中の文字や解の意味を理解すること	60%	○	-
			等式の性質を基にして，方程式が解けることを知ること	53%	-	69%
			簡単な一元一次方程式を解くこと及びそれを具体的な場面で活用すること	58%	×	-
				63%	-	90%
	図　形	【平面図形】観察，操作や実験などの活動を通して，見通しをもって作図したり図形の関係について調べたりして平面図形についての理解を深めるとともに，論理的に考察し表現する能力を培う	線対称，点対称の意味を理解するとともに，対称性に着目して平面図形についての直観的な見方や考え方を深めること	85%	-	98%
			角の二等分線，線分の垂直二等分線，垂線などの基本的な作図の方法を理解し，それを具体的な場面で活用すること	85%	○	-
		【空間図形】観察，操作や実験などの活動を通して，空間図形についての理解を深めるとともに，図形の計量についての能力を伸ばす	空間における直線や平面の位置関係を知ること	54%	○	-
			空間図形を直線や平面図形の運動によって構成されるものととらえたり，空間図形を平面上に表現して平面上の表現から空間図形の性質を読み取ったりすること	85%	○	-
				77%	-	98%
			扇形の弧の長さと面積並びに基本的な柱体，錐（すい）体及び球の表面積と体積を求めること	36%	○	-
	数量関係	【比例と反比例】具体的な事象の中から二つの数量を取り出し，それらの変化や対応を調べることを通して，比例，反比例の関係についての理解を深めるとともに，関数関係を見いだし表現し考察する能力を培う	比例，反比例の意味を理解すること	86%	○	-
			座標の意味を理解すること	79%	-	94%
			比例，反比例を表，式，グラフなどで表し，それらの特徴を理解すること	84%	-	98%
				64%	-	92%
			比例，反比例を用いて具体的な事象をとらえ説明すること	66%	○	-
第2学年	数と式	【式の計算】具体的な事象の中に数量の関係を見いだし，それを文字を用いて式に表現したり式の意味を読み取ったりする能力を養うとともに，文字を用いた式の四則計算ができるようにする	簡単な整式の加法，減法及び単項式の乗法，除法の計算をすること	68%	○	-
				54%	-	79%
			文字を用いた式で数量及び数量の関係をとらえ説明できることを理解すること	59%	○	-
				42%	-	63%
			目的に応じて，簡単な式を変形すること	65%	-	83%
				66%	○	-
		【連立方程式】連立二元一次方程式について理解し，それを用いて考察することができるようにする	二元一次方程式とその解の意味を理解すること	66%	-	90%
			簡単な連立二元一次方程式を解くこと及びそれを具体的な場面で活用すること	76%	○	-
				78%	-	98%
				78%	-	94%
	図　形	【平行と合同】観察，操作や実験などの活動を通して，基本的な平面図形の性質を見いだし，平行線の性質を基にしてそれらを確かめることができるようにする	平行線や角の性質を理解し，それに基づいて図形の性質を確かめ説明すること	66%	○	-
			平行線の性質や三角形の角についての性質を基にして，多角形の角についての性質が見いだせることを知ること	62%	-	78%
		【図形と証明】図形の合同について理解し図形についての見方を深めるとともに，図形の性質を三角形の合同条件などを基にして確かめ，論理的に考察し表現する能力を養う	平面図形の合同の意味及び三角形の合同条件について理解すること	55%	×	-
				77%	-	95%
			証明の必要性と意味及びその方法について理解すること	73%	○	-
				61%	○	-
				66%	-	81%
			三角形の合同条件などを基にして三角形や平行四辺形の基本的な性質を論理的に確かめたり，図形の性質の証明を読んで新たな性質を見いだしたりすること	78%	-	97%
	数量関係	【一次関数】具体的な事象の中から二つの数量を取り出し，それらの変化や対応を調べることを通して，一次関数について理解するとともに，関数関係を見いだし表現し考察する能力を養う	事象の中には一次関数としてとらえられるものがあることを知ること	55%	○	-
			一次関数について，表，式，グラフを相互に関連付けて理解すること	60%	○	-
				70%	○	-
			二元一次方程式を関数を表す式とみること	29%	-	34%
				57%	-	77%
			一次関数を用いて具体的な事象をとらえ説明すること	88%	○	-
				83%	○	-
		【確率】不確定な事象についての観察や実験などの活動を通して，確率について理解し，それを用いて考察し表現することができるようにする	確率の必要性と意味を理解し，簡単な場合について確率を求めること	51%	○	-
				41%	-	66%

表の見方
1）「平均正答率」はこの調査を受けた人たち全体のなかでその問題に正しく答えられた人たちの割合です。数値が小さいほどむずかしい問題です。
2）○と×はあなたが実際に解いた問題の正解・不正解をあらわしています。
3）解かなかった問題の「推定正答確率」は、あなたが受けたテスト冊子の中にはその問題は含まれていないのですが、仮にこの問題を解いていたらどの程度正しく答えられるかを示したものです。