

第3部 テストの信頼性と測定概念

4. 数学と国語の信頼性の検証

4.1 数学の信頼性

数学のテストの信頼性については、過去 2 年間の調査研究により検討済みである。本調査研究で使用了問題は全て、昨年度または一昨年度で使用了したものであるため、信頼性が大きく異なっていることは考えにくい。そこで、本報告書では数学に関しては、信頼性係数を算出するに留める。

表 4.1.1 に、数学データに関して分冊ごとに求めたテストの信頼性係数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は 32 であった。受検者数はいずれも 362 名 (2539/8) 前後に割り当てられていることがわかる。また α 係数は、いずれの分冊も 0.9 以上と高い値であり、分冊単独であっても十分な測定精度を持っていることがわかる。

表 4.1.1 分冊ごとの信頼性係数の推定値

分冊	SB01	SB02	SB03	SB04	SB05	SB06	SB07
項目数	32	32	32	32	32	32	32
受検者	368	356	357	356	369	366	367
信頼性係数	0.924	0.918	0.924	0.929	0.921	0.932	0.923

項目セットごとの信頼性係数の推定値は表 4.1.2 に示したとおりである。各項目セットの信頼性係数の推定値はいずれもほぼ 0.7 であった。信頼性係数は、項目数が少ないほど小さくなるため、項目セットごとの信頼係数は分冊ごとの値に比べて低い値であった。

表 4.1.2 項目セットごとの信頼性係数の推定値

項目セット	S01	S02	S03	S04	S05	S06	S07
項目数	8	8	8	8	8	8	8
受検者	1435	1449	1459	1458	1458	1447	1450
信頼性係数	0.789	0.769	0.774	0.746	0.763	0.753	0.733

4.2 国語の信頼性

4.2.1 テストの信頼性

国語データについて、分冊ごとに求めた信頼性係数の推定値であるクロンバックの α を表 4.2.1 に示した。いずれの分冊も項目数は 12 であった。受検者数はいずれも 363 名 (2540/7) 前後に割り当てられていることがわかる。信頼性係数の推定値は、0.444~0.696 であった。数学に比べて、国語のテストは相対的に信頼性係数が低くなることが経験的に知られている。その理由は、一般に国語の問題は 1 つの題材 (文章) に対して学習指導要領の領域及び目標の異なる小問が複数用意されるためである。一方、数学はその度合いが少ない。そのため、国語の大問ごとの信頼性は数学よりも低くなる傾

向にあるといえる。この傾向は本調査の結果にも表れており、分冊ごとの信頼性係数は国語が数学よりも低かった。

表 4.2.1 分冊ごとの信頼性係数の推定値

分冊	KB01	KB02	KB03	KB04	KB05	KB06	KB07
項目数	12	12	12	12	12	12	12
受検者	366	364	363	358	363	362	364
信頼性係数	0.444	0.581	0.627	0.607	0.696	0.580	0.592

表 4.2.2 に項目セットごとの信頼性係数の推定値を示した。分冊ごとの信頼性係数よりもさらに低い値（0.133～0.509）が得られている。数学における分冊と項目セットの信頼性係数の関係と同じ理由で、国語でも、項目数の少ない項目セット単位の信頼性係数が低くなっている。数学の項目セットごとの信頼性係数に比べて国語が相対的に低いのは、分冊におけるのと同じ理由であるが、国語はさらに項目セットごとの項目数が数学よりも少ないことが影響しているといえる。

表 4.2.2 項目セットごとの信頼性係数の推定値

項目セット	K01	K02	K03	K04	K05	K06	K07
項目数	4	4	4	4	4	4	4
受検者	1093	1092	1091	1087	1085	1088	1084
信頼性係数	0.221	0.133	0.430	0.332	0.432	0.509	0.390

昨年度の調査では、分冊ごとの信頼性係数が 0.511～0.760、項目セットごとの信頼性係数が 0.162～0.497 であった。昨年に比べると、本年度の分冊ごとの信頼性係数はやや低い値が得られている。特に、本年度の分冊 1 は信頼性係数が 0.444 と、昨年度の最低値よりもさらに低い値となっていた。一方、本年度の項目セットごとの信頼性係数はほぼ昨年並みといえる。昨年度と共通の問題を使用した項目セットは、本年度の項目セット K03、K06 であったため信頼性係数を比較したところ、項目セット K03（本年度 0.430、昨年度 0.393）、項目セット K06（本年度 0.509、昨年度 0.440）とも、本年度の方が昨年度よりもやや高い値が得られていた。

4.2.2 信頼性の低い分冊の分析

本年度調査では信頼性係数の低い分冊が見られたため、以下では、信頼性係数の高かった分冊との相違点を明らかにすることにより、当該分冊の信頼性係数が低くなった理由を考察する。信頼性係数の低い分冊として分冊 KB01 を、信頼性係数の高い分冊として分冊 KB03、KB05 を取り上げる。

まず、因子分析モデルに基づいて各分冊で測定した能力の一貫性を検討する。分冊ごとに因子分析を行い固有値を求めたところ、いずれも第一因子の固有値が他の因子よりも突出していた。ただし、固有値の減衰状況は分冊 KB01（図 4.2.1a）の方が、他の 2 つの分冊（図 4.2.1b, c）よりもなだらかであった。したがって、因子分析モデルから判断すると、分冊 KB01 は分冊の一次元性が確保できてい

るというよりも、複数の次元を反映した結果であると捉えられ、多様な国語能力を測定していた分冊であったといえる。一方、信頼性係数の高かった分冊 KB03 と KB05 は一次元性が確保されていると捉えられ、国語の中でも比較的一貫した能力を測定する項目からなっていたと解釈できる。

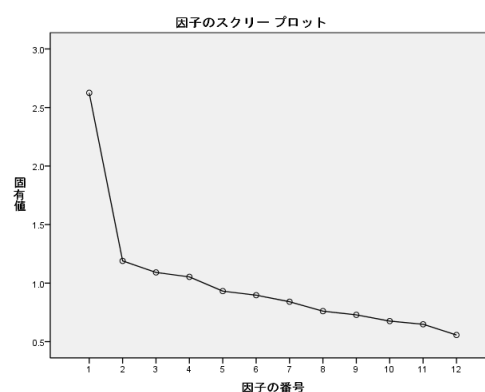
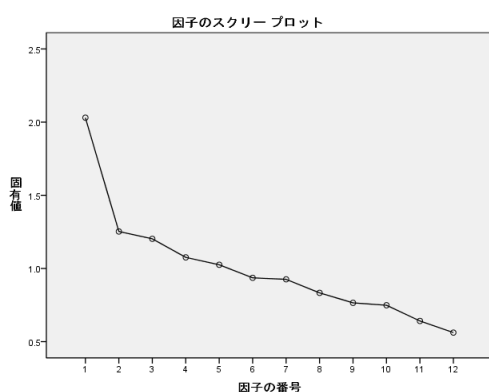


図 4.2.1a 分冊 KB01 のスクリープロット 図 4.2.1b 分冊 KB03 のスクリープロット

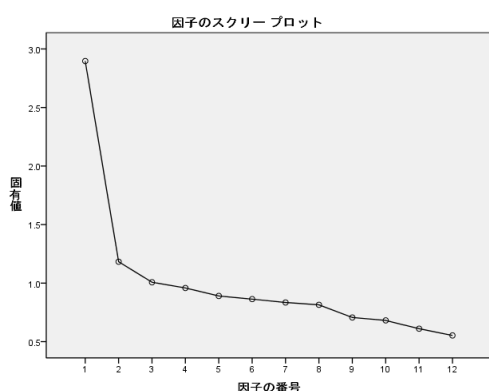


図 4.2.1c 分冊 KB05 のスクリープロット

続いて、個々の項目ごとの識別力を検討するために、点双列相関係数を算出した。点双列相関係数は、その項目とその項目が含まれているテストの総得点との相関を計算したものであり、その項目がテストで測定している特性をどの程度捉えているかを見るものである。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含まない場合とがあり、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する。

図 4.2.2 には、点双列相関係数と正答率との散布図⁴を示した。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、全体的に良好な結果が得られた項目が多

⁴本報告書では、使用した項目の識別に、変数名と項目名の二つを使用している。変数名と項目名及び、それぞれの項目の含まれる項目セットや分冊については、資料編の 2.1.3～2.2 (国語) 及び 1.1.3～1.1.4 (数学) を参照されたい。ここでは、たとえば図中一番下に位置する項目は t12_7 となっているがこれは変数名であり、それに対応する項目名は k07 となる。したがってこれ以降、変数名と項目名の対応をつけやすくするために、t12_7 (k07) のように「変数名 (項目名)」と表記する。本来なら両者は統一して項目名で表記するべきであったが、手法開発と分析を同時並行的に進めなければならなかったため、そこまでの表記の統一ができなかった。なお、分冊は KB、項目は k で区別した。

い。しかし、点双列相関係数が 0.2 を下回った項目が 7 項目見られ、係数の値が小さい順に t12_7(k07), t12_2 (k02), t12_14 (k14), t12_6 (k06), t12_11_11_11 (k11), t12_26 (k26), t12_5 (k05) であった。すなわち、これらの 9 項目はテストで測定している特性をうまく捉えることができなかった項目といえる。このうち、信頼性係数の低かった分冊 KB01 には、t12_2 (k02), t12_5 (k05), t12_6 (k06), t12_7 (k07), t12_14 (k14) の 5 項目が含まれていたのに対し、信頼性係数の高かった分冊 KB01 には t12_11_11_11 (k11), t12_14 (k14) の 2 項目、分冊 KB05 には t12_2 (k02) のみが含まれていた。

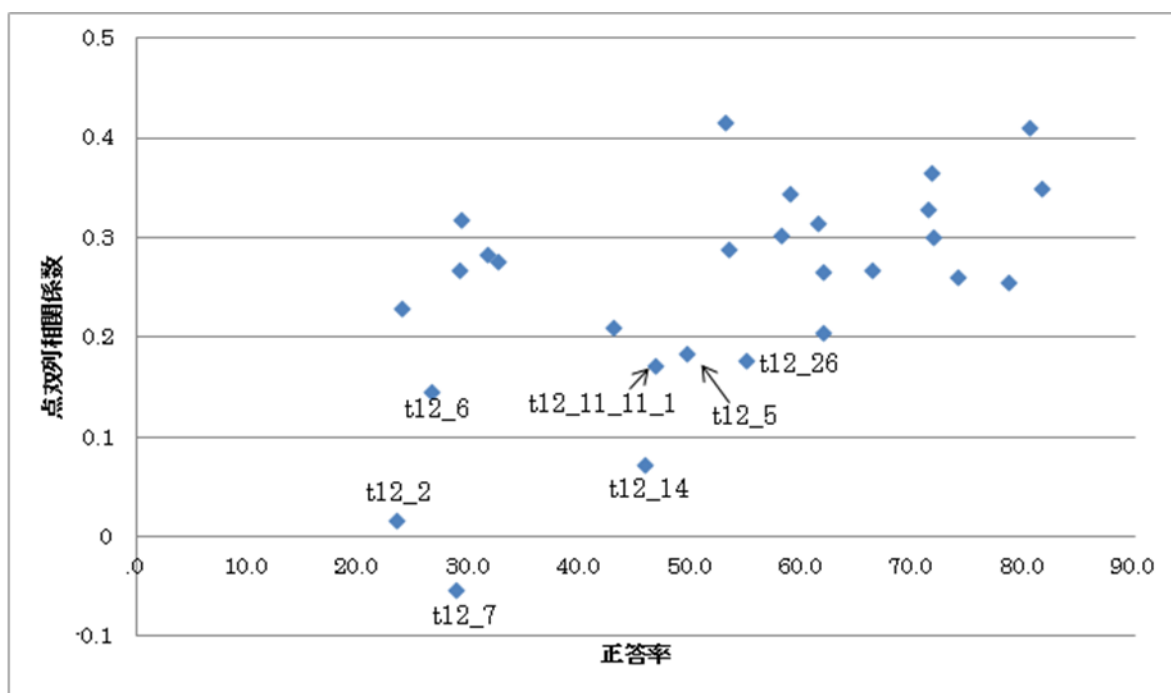


図 4.2.2 国語の正答率と点双列相関係数の散布図

さらに、点双列相関係数と因子負荷との関係を表したのが図 4.2.3 である。データ行列の一次元性が強い場合には、点双列相関係数と因子負荷はほぼ同じ情報を持つため、本調査のデータでも、全ての分冊に関して求めた両者の値をプロットすると直線状に並ぶ。しかし、ここでも分冊 KB01 の t12_2 (k02) と t12_7 (k07) は、点双列相関係数、因子負荷とも著しく低い値となっていることがわかる。一方、分冊 KB03 では t12_14 (k14) が、分冊 KB05 では t12_2 (k02) が点双列相関係数と因子負荷量の低い項目となっていた。

以上の分析結果を踏まえれば、分冊 KB01 の信頼性係数が他の分冊よりも低くなったのは、テストで測定している特性をうまく捉えることのできない項目が複数含まれており、結果として多様な能力を測定するテスト構成になっていたためと結論づけることができる。一方、分冊 KB03 や分冊 KB05 は、比較的一貫した国語能力を測定するテスト構成になっていたために、信頼性係数が高くなったといえる。

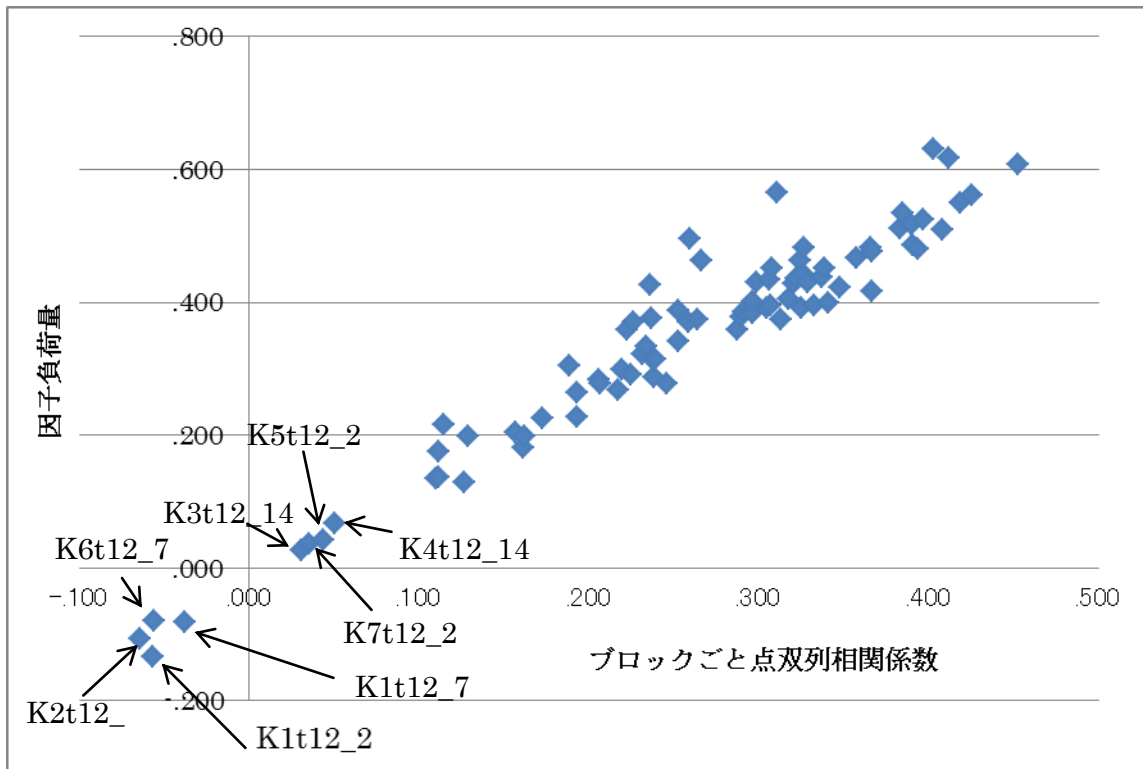


図 4.2.3 点双列相関係数と因子負荷量の散布図

4.2.3 項目分析

上で行った分析から、本年度使用した項目の中に、因子負荷量と点双列相関係数の著しく低い項目が含まれることが明らかになった。具体的には、t12_2 (k02) と t12_7 (k07) がそれに該当する。これらの項目は、テストで測定している特性をうまく捉えることができていなかった。そこで以下では、GP 分析図の結果、及び実際の問題文を分析することにより、これらの項目のどこに問題があったかを検討する。

GP 分析図 (Good-Poor Analysis) は、受検者の学力レベルを分類して、段階ごとに回答パターンがどれくらいの比率でいるかを示したものである。図 4.2.4 は t12_2 (k02) の GP 分析図である。受検者の学力レベルを 5 段階に分類し、レベルの数値が高いほど高い学力を表している。なお、GP 分析図における点双列相関係数は、各項目と学力特性値 (θ) との相関を求めているので、上で求めた点双列相関係数の値とは異なっている。

図 4.2.4 では、いずれの学力レベルでも誤答の選択率の方が正答の選択率よりも高く、正答率も 0.23 と低いことから難しい問題であったことがわかる。また、正答の選択率を表すグラフがレベル 1～4 まではほぼ横ばいで、レベル 5 で上昇していることから、学力の高いグループを識別するのに適した問題であったといえる。

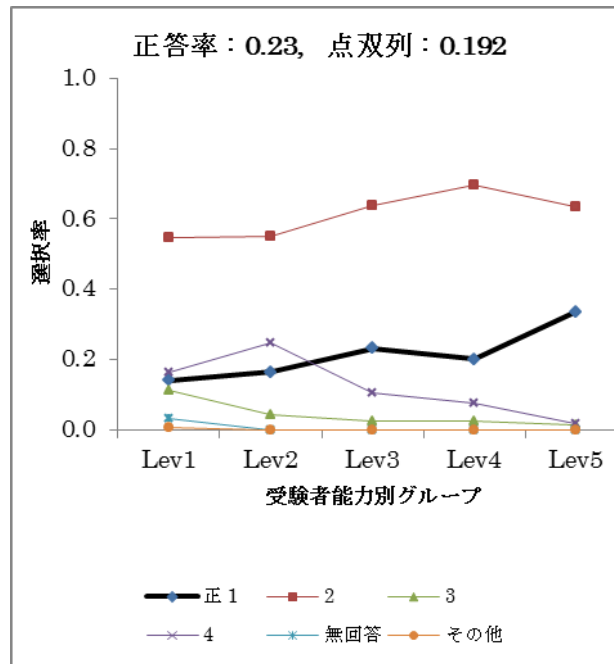


図 4.2.4 t12_2 (k02) の GP 分析図

t12_2 (k02) では特定の誤答の選択率が正答の選択率よりも高かったため、具体的な問題内容を検討する。図 4.2.5 は t12_2 (k02) の問題文であり、物語の登場人物の発言内容を問うものである。前後の文脈から、①の言葉と②の言葉が異なる意味を表すことは推測しやすいものの、それらの内容を考えるには、言葉の意味を吟味する必要がある。ところが題材の文章では、②の言葉の内容は直前に記述されているが、①の言葉の内容はその前後に記述されておらず、かなり後まで読み進めなければならぬ。そのため、②の言葉（「ここ使うなよ」「オレが予約してたんだから、どけよ」）だけを見て答えようとすれば、選択枝 2 を選択しかねない。このように、正答以外に判断に迷うような選択枝が含まれていたために、誤答の選択枝の選択率が高くなったと考えられる。

問 ——線部①「クラスのみながいる前では言えなくなるはずの言葉」と、——線部②「こういう言葉」はそれぞれどのような言葉ですか。最も適切なものを次の中から一つ選びなさい。

- 1 「言えなくなるはずの言葉」は、相手の実力を認める言葉で、「こういう言葉」は、自分の立場を譲らず主張する言葉。
- 2 「言えなくなるはずの言葉」は、自分のしたことを素直に謝る言葉で、「こういう言葉」は、相手を挑発する言葉。
- 3 「言えなくなるはずの言葉」と「こういう言葉」は同じ意味で、ケンカを続けるために相手を怒らせる言葉。
- 4 「言えなくなるはずの言葉」と「こういう言葉」は同じ意味で、仲直りをするために相手をなだめる言葉。

図 4.2.5 t12_2 (k02) の問題文

続いて、図 4.2.6 に t12_7 (k07) の GP 分析図を示す。この項目も、全ての学力レベルで誤答の選択率が正答の選択率を上回っており、正答率も 0.29 と低いことから、難しい問題であったことがわかる。レベル 1 から 2 へと正答の選択率がやや高まっていることから、学力の低いグループを識別するのに適した問題であったと考えられる。

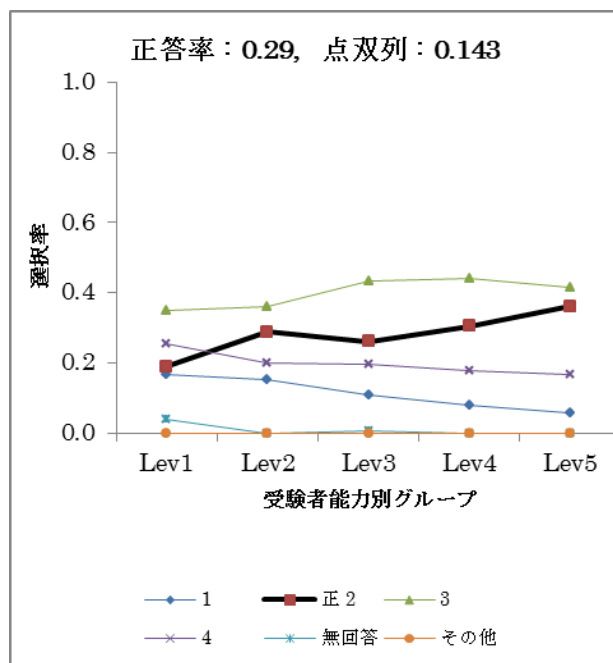


図 4.2.6 t12_7 (k07) の GP 分析図

t12_7 (k07) は、正答以外のどの選択枝も一定程度の比率で選択されていたため、具体的な問題内容を検討する。図 4.2.7 が t12_7 (k07) の問題文であり、題材の文章に関連する対話文を読んで、二人の話し手に共通する考えを答えさせるものである。正答は選択枝 2 であり、B さんの「『おはよう』をいつまで使うかについても、時間以外に何か基準があるんじゃないのかな」という発言が手がかりとなる。一方、選択枝 3 は二人が最終的に至った結論であり、選択枝 4 は題材の文章の記述内容である。つまり、選択枝 3、4 は、問題の答えとしては誤りであるが、内容自体は妥当である。そのため、選択率が高くなったと推察される。

問 A さんと B さんはことばの使い方についてある共通する考え方をしていますが、それはどのような考え方ですか。次の中から適切なものを一つ選びなさい。

- 1 あいさつのことばは、同じような生活をしている人同士であれば同じ時間に使う。
- 2 ことばの使い方について誰にでも分かるような基準があれば、使い方は統一される。
- 3 ことばの使い方には、もともと個人差がある。
- 4 人によって使い方が異なることばもあれば、同じように使われることばもある。

図 4.2.7 t12_7 (k07) の問題文

以上より、因子負荷量と点双列相関係数の低かった t12_2 (k02) , t12_7 (k07) は、判断に迷うような選択枝が含まれていたために、非常に難しい問題となったといえる。いずれの項目も、特定の学力レベルのグループにおける識別力は確認されたものの、学力の高い受検者でも正答よりも誤答の選択率の方が高かったことから、難しすぎる問題であったと言わざるを得ない。これらの項目が項目セットや分冊の信頼性を低下させる一因であったという点からいえば、テストの内的整合性を確保するうえでは適切でない問題といえる。

5. リーディング・リテラシーと国語科における「読解力」の違い

5.1 はじめに

学力や性格、適性など物理的には存在しない「心理量」を測定する際には、その構成概念をいかに定義するかが本質的に重要である、ここでは本年度の国語の中核である「リーディング・リテラシー」に焦点をあて、従来の国語科における「読解力」との違いを検討する。

「リーディング・リテラシー」とは、経済協力開発機構(OECD)が行っている15歳生徒の国際学習到達度調査 PISA(Programme for International Student Assessment)で使用されている用語で、数学的リテラシー(mathematical literacy)・科学的リテラシー(scientific literacy)とともにこの調査の1つの領域を指す。最初は「読解力」と訳されたが、我が国の国語科教育で用いられてきた「読解力」と意味が異なるため、「リーディング・リテラシー」とカタカナ書きで示されることが多くなった。本調査では、リーディング・リテラシーと国語科の問題を分けて考えている。そこで、本章では、リーディング・リテラシーと国語科で扱われている読解力がどのように異なるのか、そのような違いが生まれたのはなぜかを考察することを目的とする。

5.2 リーディング・リテラシーと国語科における読解力の違い

まず、リーディング・リテラシーと国語科における読解力の違いについて、前稿(足立, 2012)では、3点指摘した。まず、その3点について説明する。

1点目は、読まれるテキストについてである。リーディング・リテラシーでは、「物語」「解説」「記述」「議論」「指示」「文書または記録」「ハイパーテキスト」といった、いわゆる文章の形態であられる「連続型テキスト」と、「図・グラフ」「表」「図」「書式」「情報シート」「宣伝・広告」「バウチャー」「証明書」といった文章の形態とは言いがたい「非連続型テキスト」がある。これに対し、国語科では、「文学的文章」(文学的な文章、文学作品などの表現も用いられてきた)と「説明的な文章」(説明的な文章、説明文などの表現も用いられてきた)が二大ジャンルとして扱われてきた。ここで重要なのは、リーディング・リテラシーが取り上げる多様なテキストである。リーディング・リテラシーは、義務教育を修了した若者が「自らの知識と可能性を発達させ、効果的に社会に参加するために」(リーディング・リテラシーの定義の一部)読む能力であるので、グローバル社会の中で若者達が出会う様々な形態のテキストを扱う。このため、リーディング・リテラシーが「私的」「公的」「職業的」「教育的」の4つの状況が設定され、それぞれ21%、25%、25%、29%という割合で設問が作成されている。一方、国語科では、若者が国語科教育の中で学ぶに足る典型的な規範的な文章を対象としてきており、非連続型テキストは対象になっていなかった。加えて、国語科では、状況の設定は行われていない。教科書に載っている文章ということを考えると、「教育的」状況のテキストだけが読むべき文章として扱われていることになる。これは、リーディング・リテラシーでは29%に過ぎない範囲ということになる。さらに、非連続型テキストは、リーディング・リテラシーの約30%を占めるので、これを含めないとすると、国語科が扱ってきたテキストの範囲は、リーディング・リテラシーの約20%という計算になる。加えて、注目したいのは、「説明的文章」の内

容である。「解説」「記述」「議論」「指示」「文書または記録」などの下位分類は、あまり重視されてこなかった。加えて、教科書に掲載される「説明的な文章」は、大抵は教科書用に再編集されたもので、現実の社会にはあまり存在しないような特殊な形になることが多かった。具体的には、見出しが外されたり、図・表が削られたり、問いが1段落目もしくは2段落目に配置されたり、最後の段落に筆者の主張や問題提起が書かれるようになったりということである。つまり、一言で言うと、リーディング・リテラシーは、若者が出会うであろう現実のテキストを対象とし、国語科は、生徒が教科書用に書き換えられた文章を対象としているということになる。

2点目は、リーディング・リテラシーでは、長く記述するような形式の解答欄があるということである。単に書かれたことを自分が理解できればよいということではなくて、自分はこのように理解したということを、しっかりと表明できなければならない。国語科では、長く記述するものは、「読むこと」ではなく「書くこと」としてカリキュラムに位置付けてきた。このため、読む能力をつけるために、書き方を教えるということがあまり行われてこなかったと言ってよい。このことは、我が国では、これまでは比較的等質の家庭環境・文化環境を持った生徒達が教育の対象となってきたことと関係がある。我が国のコミュニケーションの特徴と言える「以心伝心」「言わずもがな」は、等質の文化背景を持つ者同士では有効に働く。しかし、リーディング・リテラシーが前提としているグローバル社会の中で、異質の文化背景を持つ人間が共存していこうとするときには、自分の考えを的確に表明したり、妥協点をすりあわせるために交渉したりということを行っていかなければならない。したがって、自分が読んでいること・考えていることについて、異質な文化背景の相手に分かるように述べる力がリーディング・リテラシーでは、読む力として測定されているのである。

3点目は、リーディング・リテラシーでは、テキストに書かれた内容を理解することだけでなく、そのテキストが自分にとってどのような意味があるかを、自分の先行知識・価値観などとの関係から評価・判断させるような側面「熟考・評価」を調査していることである。一方、国語科では、書かれた内容を理解することが重視されてきた。具体的には、自分の読み方について述べる時はテキストの中に根拠を見いだして述べる必要があるとされ、自分の先行知識や価値観などは根拠とならないとされてきたことである。このことについては後で詳しく述べるが、昭和50年代・60年代に中途半端に導入された読者論と関係があると筆者は考えている。読みというのは、読者の自由であるから、国語科の中で扱えるのはテキストに書かれたものだけであるという考え方である。

以上をまとめると、リーディング・リテラシーは、グローバル化する現実社会に参加するために若者が読まざるを得ない多様なテキストを、異質な文化背景を持つ者とともに読む能力である。一方、我が国の国語科教育で行われてきたのは、比較的等質文化背景を持つ国民として典型的あるいは規範的な文学的な文章や説明的な文章を、文章に書かれたことの内容を正確に読み取る能力だったと言ってよい。なぜ、このような違いが生まれたのかを次に考察していく。

5.3 読解と読書の分立とその歴史

繰り返しになるが、リーディング・リテラシーと国語科の違いをさらに短くまとめるならば、状況への配慮の違いである。リーディング・リテラシーでは、グローバル化する社会に参加する若者の能

力に関心があるため、実際のところ、若者はどのような状況でテキストを読むのかをとらえることが必要であった。しかし、国語科の方は自動的に「教育的」な状況だけを考え、現実生活中で若者がどのような状況で何をどのように読むのか、関心を払ってこなかった。そこで重要なのは、教科書の位置付けである。リーディング・リテラシーは、国際学力調査であるので、ある特定の国の教科書を前提にできないことは当然であるが、逆に言うと、国語科の読解力が上記のような特徴になったのは、教科書という媒体を抜きには考えられない。

我が国では、昭和30年代頃から、読むことの教育について教科書を対象にした「読解」と教科書以外のものを対象にした「読書」の2本立ての体制がとられるようになった。「読解」の方は教室で集団で行うことであり、「読書」の方は家庭や学校図書館で個人で行うこととされた。したがって、読解力と言え、国語科の授業という教育的状況の中で教科書を対象にした「読解」の力ということになる。昭和30年代から平成10年代まで「読書」は学習指導要領の目標には位置付けられているが、内容としては位置付けられず、授業として扱われることも少なかったと言ってよい。「教育的」状況以外の状況（リーディング・リテラシーで言うならば「私的」「公的」「職業的」といった状況）で読まれる教科書以外のテキストは、分類されることも詳しく考えられることもなく、「読書」の方に一括されていた。

昭和40年代・50年代の読書指導論を見てみると、マンガ・写真・印刷メディアなどの書籍の形態をとらないものの読書の可能性が述べられていたり、テレビと読書の関係が論じられていたりする（滑川、1976）といった、いわゆる本ではないものについて論じられている。すなわち、「読書」指導という領域では、当時の児童・生徒が実際に読んでいるものに着目しようとする発想があったと言ってよい。しかし、「読書」は家庭・学校図書館などの授業以外の場面で行うことという意識が強く、授業で教育されることは少なかった。

一方、教科書「読解」の方は、「文学的文章」と「説明的文章」を読ませることに終始していった。端的に言えば、昭和30年代から40年代の国語科教育は、「文学的文章」に登場する人物の気持ちを読み取ったり、行動の理由を理解しようとすることに重きが置かれてきた。これに対し、昭和50年代・60年代には、いわゆる「読者論」の導入があり、児童・生徒が自由に読もうとするのを尊重しようという動きがあった。しかし、この読者論は、国語科での「感じ方は自由」「何でも正解」などと言った教育なき自由論を招き、その反動で、国語科で問うべき、文章のみを根拠とする正確な読み取りへと変貌していった。「説明的文章」の方は、昭和40年代までの知識を読み取る形式から、昭和50年代・60年代は段落構成、筆者の述べ方の工夫、主張との整合性を読むといった、説明的文章の形式面への指導に傾斜がかかっていった。これらは、いずれにせよ、テキスト内容・形式そのものを読み取ることを重視するものである。リーディング・リテラシーの「熟考・評価」のように、自分の先行知識・価値観を根拠に、テキストを評価したり判断したりするようなことは、国語科では行わなかった。

平成20年告示の中学校学習指導要領では、PISAに影響を受けたと言われている。そのうちの1つは、国語科の内容「C読むこと」の枠組みは、「語句の意味の理解」「文章の解釈」「自分の考えの

形成」「読書と情報活用」となったことである。これは、前者3つがリーディング・リテラシーの3側面「1情報の取り出し」「2解釈」「3熟考・評価」に即しており、最後の1つが「読書」ととらえられる。やはり、「読解」と「読書」の分立は、踏襲されていると言える。

5.4 読書観の違い

これまで、主に、教科書を中心として、リーディング・リテラシーと国語科では、どのような内容が含まれているかを述べてきた。最も具体的には、テストで能力として測定されるものの違いであるが、もう少し拡張して考えると、教育の中での「読むこと」の扱われ方の違いということもできるし、OECDと文部科学省という機関の立場の違いとも言える。しかし、もっと根本的な問題としては、これは、読むことに関する考え方、すなわち読書観の違いと言える。乱暴な言い方ではあるが、西欧中心で進められてきたリーディング・リテラシーは西洋の読書観に基づき、我が国の国語科は東洋の読書観を持ってきた。その西洋と東洋の読書観の違いが、結果として、カリキュラムや教材や評価の違いとして現れたとも見ることができる。

ここで言う西洋の読書観とは可視的であり、東洋の読書観は不可視的である。滑川（1979, p.198）は、「読書における東洋的思考」という論考の中で、アメリカ人研究者ハリスと日本人研究者阪本一郎の読書能力のとらえ方を比較している。それによると、阪本は、ハリスの「読書技能的であり、分析的である到達目標に対して『教養的読書』を付加している。おそらく阪本の不満がそうさせたに違いない」と述べている。滑川は、日本では、「読書を通しての『人間形成』を思念する。アメリカ的思考においては、人間形成的な教育目標は、次元の違う問題として別に考えるか、わかりきった当然のこととして、技能的な面を取り上げる。日本では、行動化できるものばかりに限らず、精神的にさえ映るだろう不可視的実在が取り上げられてきた。究極的には西欧文化は『有』を実在の根底にすえ、東洋の文化は『無』を実在の根底にすえているという、東西文化の淵源にまでさかのぼらなければ解明できないのかもしれない。」（滑川, 1979, p.199）と述べている。

西洋の読書観が可視的であれば、それは説明可能なものとしてとらえられる。リーディング・リテラシーの中で、長い記述をさせたり、「熟考・評価」をさせたりするのは、西洋の読書観に基づけば、読者が読んで考えていることは、本人が言葉で説明したり行動で示したりして見せられるということになる。「熟考・評価」はテキスト中に明示されていない生徒の先行知識・価値観を使用して解答させるものであり、その先行知識や価値観は、異質な文化背景を持った相手にあっても説明によって理解可能になるという考え方である。これに対し、東洋の読書観に基づけば、最終的には読者が読書をするということは、他人に示すことができない不可視的なものであるということになる。それを、テストのような形で明示することは不可能であるから、測定できる範囲は、テキストに解答が掲載されている範囲ということになる。「熟考・評価」のような、読者の中にある思考を明示させることには積極的でないことになるのである。

筆者が、欧米の複数の国に滞在して読書指導法（リーディングの教育法）を研究した時に共通して感じたことは、欧米の読書指導法は、自分の読書した事柄（それは読み取った内容だけでなく、自分

が想起した記憶や知識，付随した思考など）を教室内やコミュニティー内のメンバーに対して共有することが多いということであった。それに比べると，我が国の国語科では，読書は個人作業でとらえられることが多く，自分の理解や考えを誰かに伝えるということは少ない。これも，他人の読書が可視的であるととらえるか不可視的ととらえるかの違いであると言えよう。

西洋の教育の場面で，現在主流になってきているのは，読者反応理論である。この理論では，まず，読むという行為を読者がテキストの意味を構成する行為ととらえる。そして，その意味構成を行う際には，テキストの内部だけでなく，読者の経験，読者の思考過程，社会的なコンテキスト，文化といったものが重視される。そして，これらの意味構成は，読者反応として読者から示されうるという考え方である。

5.5 おわりに

以上，やや乱暴でもあるが，リーディング・リテラシーと国語科での読解力はどのように違うのか，また，それはどのようにして異なるものになったのか，国語科教育の歴史を踏まえ，西洋と東洋の読書観の違いというところまで広げてとらえてみた。このようにリーディング・リテラシーと国語科の違いを踏まえたうえでの学力調査は，国語科教育の歴史の上でも有意義なものである。

第4部 結果と考察

6. リーディング・リテラシーと数学との関係

6.1 リーディング・リテラシーと数学問題との相関

第 5 章で述べたように本調査では国語科における「読解力」とリーディング・リテラシーとを区別している。そのことを前提に、本章では、リーディング・リテラシーの学力特性値（以下、RL θ ）と数学問題から数学の学力特性値（以下、数学 θ ）の関係について検討する。推定値は数学 θ および RL θ は、EAP 法によるものである。いずれの学力特性値も、項目固定法による等化後の項目母数を用いて推定された（詳細は、7 章を参照のこと）。なお、数学 θ の算出には、全 56 項目を使用した。RL θ の算出には、推定誤差が大きくモデルへの適合が悪いと判断された 2 問を除いた 26 項目を使用した。

数学 θ と RL θ の相関係数は 0.547 であり、全体としては中程度の相関関係があった。本調査で測定した数学の学力と、RL の学力との関係を、より詳細に検討するため、RL θ と各数学問題との相関係数および数学 θ と各 RL 問題との相関係数を算出した。

表 6.1 RL θ と数学問題の各項目との相関係数

領域	単元	内容	item	相関係数
数と式	【一次方程式】	等式の性質を基にして、方程式が解けることを知ること。	s23	0.196
		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。	s05	0.334
			s18	0.348
	【式の計算】	目的に応じて、簡単な式を変形すること。	s10	0.221
		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。	s46	0.277
			s30	0.285
		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。	s17	0.339
			s14	0.341
		s09	0.331	
	【正負の数】	正の数と負の数の四則計算をすること。	s21	0.245
		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。	s25	0.280
			s06	0.360
		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。	s49	0.311
	正の数と負の数の必要性和意味を理解すること。	s50	0.311	
	【文字と式】	文字を用いることの必要性和意味を理解すること。	s29	0.385
		簡単な一次式の加法と減法の計算をすること。	s33	0.291
s41			0.308	
s53			0.309	
s45	0.349			
【連立方程式】	二元一次方程式とその解の意味を理解すること。	s34	0.285	
	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。	s01	0.374	
		s13	0.331	
		s37	0.355	

(次ページにつづく)

(前ページのつづき)

領域	単元	内容	item	相関係数
図形	【空間図形】	空間における直線や平面の位置関係を知ること.	s26	0.208
		扇形の弧の長さや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.	s38	0.292
			s47	0.320
		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にとらえたり、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.	s02	0.405
			s11	0.312
	【作図と移動】	平行移動、対称移動及び回転移動について理解し、二つの図形の関係について調べること.	s07	0.242
		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.	s31	0.334
	【図形の合同】	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.	s51	0.216
			s54	0.206
		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.	s42	0.326
		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.	s16	0.314
			s22	0.366
	【平行線と角】	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.	s35	0.261
			s39	0.305
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.		s43	0.346	
関数	【一次関数】	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.	s27	0.254
		二元一次方程式を関数を表す式とみること.	s48	0.249
		一次関数を用いて具体的な事象をとらえ説明すること.	s32	0.374
			s08	0.354
		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.	s24	0.357
	【比例と反比例】	比例、反比例の意味を理解すること.	s12	0.286
		関数関係の意味を理解すること.	s44	0.216
		座標の意味を理解すること.	s56	0.239
		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.	s19	0.309
			s40	0.325
			s03	0.327
		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.	s15	0.318
	資料の活用	【ヒストグラムと代表値】	ヒストグラムや代表値を用いて資料の傾向をとらえ説明すること.	s04
【確率】		確率を用いて不確実な事象をとらえ説明すること.	s20	0.273
		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.	s28	0.304
			s36	0.310

6.2 リーディング・リテラシーと相関の高い数学問題の特徴

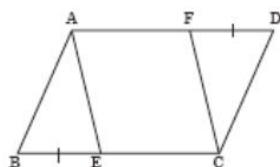
RL0 と数学問題との相関係数は、0.196～0.405 であった(表 6.1)。RL0 との相関係数が比較的高かった項目として、s32, s08, s24, s04, s02, s22, s55, s06, s29, s01, s37 の 11 項目を挙げることができる。この 11 項目は、正の数や図形など、異なる単元で学ぶ内容を扱っているもの、その問題の質において共通する特徴がある。RL0 と相関の高い数学の問題は、その内容的な特徴から、3 種類に分類することができよう。

第一に、数学科で習得した知識や技術を活用する能力を測定する項目である。s32 と s08 は「一次関数」、s22, s55 は「図形」、s06 は「正負の数」、s01 と s37 は「連立方程式」、と単元は異なるものの、これら 7 項目は、「～を用いて、～する」「～を基に、～する」という学習指導要領の内容に対応しており（表 6.1 より）、その単元で習得した知識や技術を活用する能力を測る項目であるといえる。特に s32, s08, s06, s01, s37 は、習得した知識を、具体的な事象や場面と対応づける問題となっている。

第二に、数学科での既習の事象や、複数の事象を関連づけて解答する項目である。s02, s04, s24 の 3 項目は、複数の数学で学ぶ事象や知識を結びつけることが求められる項目である。s02 は、投影図を通じて、それまでに学んできた平面図形の知識を使い、空間図形の特徴を考察する問題である。s04 は、小学校算数科で学ぶ平均値に代表される代表値と、ヒストグラムとを関連づける必要がある。s24 では、関数関係を、式とグラフという複数の表象に対応づけて理解することが求められる。

第三に、数学的知識を適切に用いた、論理的思考力を測る項目である。s22 及び s55 は、「図形」という単元における証明問題であり、例えば、s22 は、図 6.1 のように、二つの図形が合同であることを、基礎的な数学的知識である合同条件を用い、論理的に説明することが求められている（図 6.1）。s29 は、「文字と式」という単元における思考過程を、根拠を明らかにして表現する問題であり、これら 3 項目で測定しているのは、習得した数学的知識や技術を適切に活用し、思考過程を説明する能力といえる。

- (6) 下の図のような平行四辺形 ABCD があります。辺 BC、辺 DA 上に $BE = DF$ となる点 E、点 F をそれぞれとります。このとき、 $AE = CF$ となることを、次のように証明しました。□に入る証明の続きを書きなさい。



証明

△ ABE と △ CDF において、
 仮定から、 $BE = DF$ ……①
 四角形 ABCD は平行四辺形なので向かい合う辺と、向かい合う角はそれぞれ等しいから、

合同な三角形の対応する辺の長さは等しいので、
 $AE = CF$

図 6.1 s22 の問題

以上、見てきたように、RL0 と関連の高かった数学の問題項目は、数学的知識及び技術の活用能力、既習の内容や複数の数学的事象を結びつける能力、数学的知識を用いて説明する能力のいずれかを測定する項目であるという特徴がある。これらの能力は、中学校数学科において重視されている「数学

的活動」と関係しているといえよう。数学的活動とは、「生徒が目的意識をもって主体的に取り組む数学にかかわりのある様々な営み」であり（学習指導要領，2008），特に，中学校数学科では，既習の数学を基にして，数や図形の性質を見出し，発展させる活動，日常生活や社会で数学を利用する活動，数学的な表現を用いて，根拠を明らかにし筋道立てて説明し伝え合う活動を重視している。これは，本調査の結果，RL0 と相関の高かった項目の 3 つの特徴と対応するものといえる。すなわち，数ある数学の項目のうち，RL0 と相関の高かった項目は，「数学的活動」で培われる能力であることが示唆された。

6.3 リーディング・リテラシーと相関の低い数学問題の特徴

6.2 で見てきたように RL0 と相関の高かった項目には，大きく 3 つの特徴があることが示された。それでは，反対に，RL0 と相関の低かった 10 項目（s23, s54, s26, s44, s51, s10, s56, s07, s21, s48）には，どのような特徴があるのだろうか。s23, s54, s44, s51, s56 の 5 項目は，学習指導要領の「～を知ること」「～を理解すること」という内容と対応していることから，各単元において重要な知識の習得度を測る項目と判断できる。s07 と s48, s26 も，例えば s07 なら対称移動，s48 なら「連立方程式の解」，s26 なら「二つの平面が垂直に交わる条件」など，各単元の重要事項の理解度を測る内容となっており，以上の 8 項目は，重要な数学的知識の習得度を測る項目であるといえる。また，s10 は文字を含む与式の変形，s21 は，図 6.2 のような文字を含まない与式の計算を求める問題であり（図 6.2），各領域において習熟すべき，数学の基礎的技術を測る項目である。このように，RL0 と相関の低かった項目は，数学科の各単元の基礎的かつ重要な知識及び技術の習得度を測る項目といえる。

(5) 次の に当てはまる数を求めなさい。

$$(-6^2) \div (-3)^2 = \text{$$

図 6.2 s21 の問題

6.4 リーディング・リテラシーから見る数学問題の回答パターンの特徴

この結果を踏まえると，RL0 の高低によって，数学的活動で培われる能力を測定する項目の解答パターンが異なることが予想される。そこで，RL0 を 5 段階に分類して，段階ごとの数学の各項目の解答パターンの比率を示した図を描き，リーディング・リテラシーと数学的活動に関連する項目の解答パターンとの関係を検討する。これは，資料に添付した GP 分析図の，横軸を数学 0 から，RL0 に変えたものである（GP 分析図の詳細については，本書の 4.2.3 を参照されたい）。

ここでは，紙面の都合から，基礎的な数学的知識や技術の習得度に関連する問題項目として先に挙げた 10 項目の中から，s23 と s21 の 2 項目を取り上げ，それらの基礎的な数学的能力の活用力に関連する問題項目として先に挙げた 11 項目のうち，s06 と s22 の 2 項目を取り上げて，図の比較を行う。

まず，数学の基礎的な技術に関係する s21 の図 6.3 をみると，RL0 が高い者ほど，正答率が高くな

っているものの、誤答の一種である類型2の選択率が、ほぼ横ばいであり、RL0の高さで、この類型2の解答者を上手く説明できていないことが示されている。s23についても同様に、RL0の高さに従い、正答率は上がっているものの、その他の解答の比率はほぼ横ばいになっている(図6.4)。それに対し、基礎的な数学的能力の活用力を問うs06は、RL0の高さに伴い、正答率が高くなり、無答率が下がっていることがわかる(図6.5)。s22は、その傾向がより顕著であり、正答率が右上がりになっているのに対し、無答率が右下がりになっている(図6.6)。

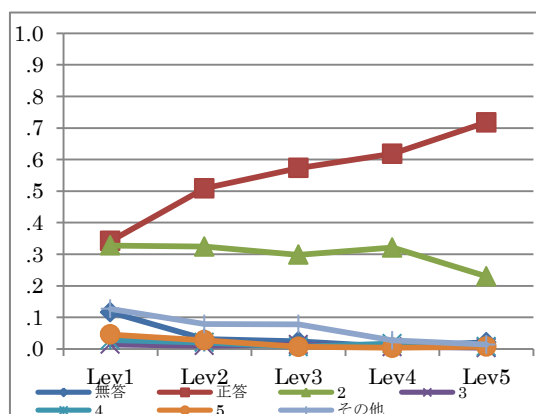


図 6.3 s21

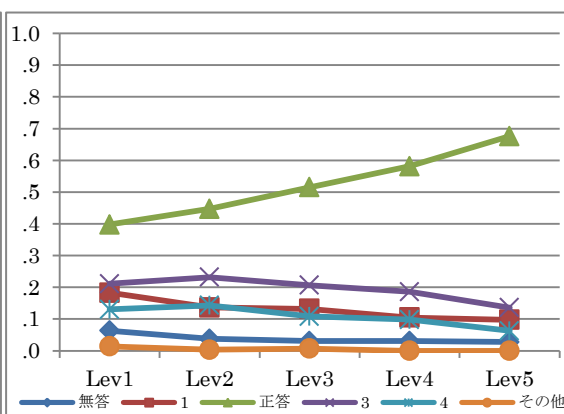


図 6.4 s23

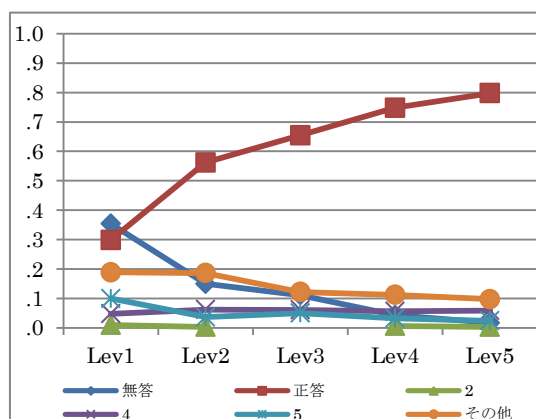


図 6.5 s06

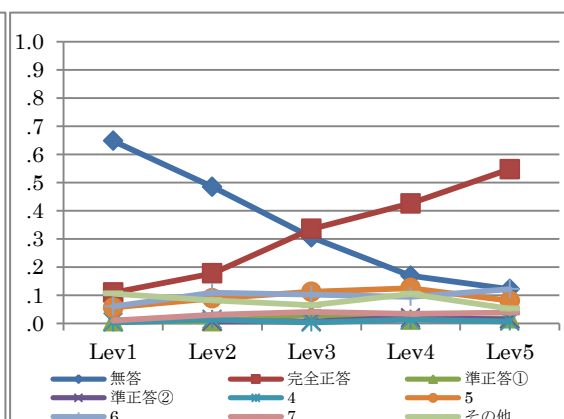


図 6.6 s22

以上の GP 分析図からも、RL0 は、数学の基礎的な知識や技術というよりも、それらを活用する力と関係があることが示された。このように、数学の問題の質によって、リーディング・リテラシーとの関係が異なる結果が得られた背景には、リーディング・リテラシーと、数学的活動で培われる、基礎的な数学的スキルの活用力との共通点があると推察される。

従来の国語科教育と比較したリーディング・リテラシーの特徴の一つとして、図表やグラフ等の非連続型テキストを扱うことがある。数学で扱われるヒストグラムや円グラフ、表などから情報を読み取り、それらを自分の主張に活用することまで、リーディング・リテラシーには含まれる。図表やグラフの意味や利点、それらのよみ方や表現方法は、数学科教育の範疇にある。よって、図表やグラフを、自分の意見や主張に活用することは、数学を活用することと同義と見なすことができよう。さら

に言えば、リーディング・リテラシーと数学的活動では扱うテキストの質（前者は連続型テキストの割合が多く、後者は非連続型テキスト、特に式や図など数学特有の表象が中心である）に違いがあるものの、両者は共に、テキストから得られた情報を正確に読み取り、活用する能力に焦点を当てている。このリーディング・リテラシーと数学的活動で培われる能力との共通点を考慮すると、本調査でRL0と、数学のうち、特に数学的活動と関係する問題項目との間に関係がみられたことは、納得できる結果といえよう。

6.5 考察

本章では、リーディング・リテラシーと、数学の問題項目との関係を考察することで、3つの重要な知見が得られた。第一に、全体として、リーディング・リテラシーと数学能力との間に中程度の相関がみられたことである。従来、国語の能力と数学の能力は、あまり相関がないと考えられてきた。しかし、本調査で測られたリーディング・リテラシーは、一般に想起されやすい読解力とは異なる（伝統的な国語科教育とリーディング・リテラシーとの違いについては、平成23年度報告書の2章及び本報告書の5章を参照のこと）。昨年度の調査では、リーディング・リテラシーのうち、情報を適切に読み取る能力は、数学の文章題と関係し、情報を端的にまとめて記述する能力は、数学の証明問題と関係することが示唆された。本調査では、昨年度よりも厳密にリーディング・リテラシーの定義に則った結果、より信頼性の高いリーディング・リテラシーの問題が作成され（4章を参照のこと）、数学との相関も顕著になった。この結果は、依然としてアウトラインを捉えることが難しいリーディング・リテラシーの実態を明らかにするための有効な知見を提供するものといえよう。

第二に、一口に数学的能力といっても、質的に異なる能力が含まれることが示唆された。本調査では、リーディング・リテラシーとの関係性という観点で、数学の問題項目を概観する中で、数学的能力を、基礎的な数学的事項や計算スキルの習得度と、それらに関連づけたり活用する能力の、二つに大別することができた。後者の数学の活用力は、現行の学習指導要領における中学校数学科の目標に盛り込まれた「数学的活動」と関連し、PISAの数学的リテラシー⁵とも通じる能力であると考えられるため、特に近年重視されている。この数学の活用力を支えるのが、前者の各単元での基礎的かつ重要な数学的知識や技術であることから、両者を、数学的能力の両輪として、バランスよく伸ばしていくことを目指すことが求められる。

最後に、本調査の結果、リーディング・リテラシーは、数学的能力の中でも、特に数学的活動で培われるような力と関係があることが示唆された。従来、国語科教育で養われてきた読解力は、一般に数学的能力と関係がうすいとされてきた。また、本調査で示されたように、数学的能力といっても、数学という領域に固有な知識や技術は、リーディング・リテラシーとの関係性が低い。この結果を踏まえると、リーディング・リテラシーや数学的活動で培われる能力は、国語科や数学科という教科に固有な能力というよりは、領域間を跨いで養われる、多領域に共通した要素を含む能力といえる可能性がある。新学習指導要領では、教科を超えた目標として、言語活動の充実を重視している（文部科

⁵量や空間、確率あるいはその他の数学的概念を含む様々な状況において、生徒が問題を数学的に設定し、定式化し、解決し、それを解釈するように、有効に分析し、推論し、他者に伝達することのできる能力に関係したもの（経済協力開発機構、2010）

学省，2011）．これは，国語科教育で行われるような言葉の美しさを体感するような活動や，基本的な国語力の習得及び熟達に通じる活動に限らず，例えば，日常的事象について数学的に表現するような数学的活動なども含まれている．このような，他者とのコミュニケーションや，考えや主張を表現することなど，言語的活動を通じて，各教科で学んだ知識や技術を活用することは，学習者が，各教科の基礎的な知識及び技術を習得することの意義を見出したり，学習意欲を促したり，そういった基礎的な学力を定着させるような効果が予想され，結果として，全体的な学力の向上へと繋がることが考えられる．本調査の結果は，このような教科共通した，知識の活用能力，すなわち，リテラシーの存在を示唆するものといえる．

7. 等化手続きと経年変化分析

本章では 2010 年，2011 年，2012 年と，これまでの 3 年間で実施された数学のテスト結果，及び，2011 年と 2012 年に実施されたリーディング・リテラシーのテスト結果を，それぞれ（経年）比較することが目的となる。

数学のテスト問題に関して，2010 年に実施された数学問題と，2011 年に実施された数学問題は，等化済みである（昨年度報告書の 6 章を参照のこと）。2012 年に実施された数学のテストでは，既に等化された 2010 年に実施された数学問題と，2011 年に実施された数学問題から選出しているため，等化を行わずに，これまでの調査結果から得られた項目パラメタを使うことで，経年比較を行うことができる。ただし，2011 年と 2012 年のリーディング・リテラシーのテスト結果を（経年）比較するためには，2011 年に実施されたリーディング・リテラシー問題と 2012 年に実施したリーディング・リテラシー問題を等化する必要がある。そこで，次節では，リーディング・リテラシー問題の等化を行う。

7.1 リーディング・リテラシー問題の等化結果の比較

ここでは，2011 年に実施されたリーディング・リテラシー問題と，2012 年に実施されたリーディング・リテラシー問題を等化することが目的となる。昨年度の調査での数学問題の等化に倣い（昨年度報告書の 6 章を参照のこと），本研究では，「共通項目デザイン」を，等化のデザインとして採用し，そのデザインに従い，テストデータを収集した。

本研究では，複数ある等化方法のうち，（1）同時尺度調整法，（2）項目固定法，（3）共通項目情報を用いた mean&sigma 法，（5）熊谷・野口（2012）による推定母集団分布情報を用いた mean&sigma 法の 4 つの方法を採用した（各方法の詳細については，昨年度報告書 6 章 2 節を参照されたい）。以上の 4 つの方法を用いて，2011 年度テストデータと 2012 年度テストデータを等化することで得られた「等化後」の識別力を表 7.1 に，困難度母数を表 7.2 に示す。各表には，本研究で採用した全方法の平均値も最終例に付してある。なお，リーディング・リテラシー問題の全 28 項目を用いて，等化を行ったところ，項目母数の挙動のおかしい項目が 2 項目あった（t12_2 (k02) は識別力母数 0.088，困難度母数 6.010 であり，t12_7 (k07) は識別力母数が 0.088，困難度母数が 5.980 であった）。そこで，表 7.1 及び表 7.2 には，この 2 項目を外して等化した項目母数を載せている。また，計算には全て EasyEstimation（熊谷，2009）を使用した。

表 7.1 と表 7.2 から，方法間・項目間で，項目母数にバラつきがあることがわかる。この背景には，2011 年度のリーディング・リテラシー問題と，2012 年度のリーディング・リテラシー問題の共通問題の少なさがある。共通問題は 8 項目（2 ブロック×各 4 問）であり，項目母数の挙動のおかしかった 2 問を外した 26 問で等化したことを考えても，全体の 3 割程度しか占めていない。また，この共通問題は 8 項目とも，古典を題材にした問題であり，材料となるテキストの種類が偏っていたという問題もある。さらに，本調査のデザインで示されているように（詳細は 3 章），分冊によって，共通問題の

数が 0 から 8 項目まで、バラついていた。このように、共通問題が少なく、かつ、質的に偏っていたことから、等化後の項目母数が安定しなかったと推察される。ただし、(1) と (5) のように、計算手法が異なるにもかかわらず、等化後の項目母数の値が近い項目もあり、適切に等化が行われたことが示唆される。

本研究では、今後のさらなるテストデータの追加などの発展可能性や、計算作業の容易さ、等化後項目母数が平均値に近いことなどを考慮し、昨年度の数学の等化に採用した方法と同じ、(2) 項目固定法によるものを、等化の方法として採用することとする。

表 7.1 等化後項目の項目母数の比較 (識別力母数)

		変数名	受検者数	(1) 同時 尺度調整法	(2) 項目 固定法	(3) 共通項目 (M&S)	(5) 推定 母集団	平均値
K01	1	t12_1	1093	.541	.466	.417	.542	.491
	3	t12_3	1093	.338	.299	.259	.338	.309
	4	t12_4	1093	.873	.786	.674	.876	.802
K02	1	t12_5	1092	.275	.243	.213	.278	.252
	2	t12_6	1092	.271	.242	.209	.272	.249
	4	t12_8	1092	.624	.563	.480	.624	.573
K03	1	t12_9_11_9	1091	.922	.948	.711	.948	.882
	2	t12_10_11_10	1091	.490	.460	.392	.460	.450
	3	t12_11_11_11	1091	.263	.185	.229	.185	.215
	4	t12_12_11_12	1091	.577	.490	.476	.490	.508
K04	1	t12_13	1087	.665	.588	.511	.665	.607
	2	t12_14	1087	.097	.085	.075	.098	.089
	3	t12_15	1087	.401	.346	.311	.404	.366
	4	t12_16	1087	.740	.635	.572	.745	.673
K05	1	t12_17	1085	.524	.479	.402	.523	.482
	2	t12_18	1085	.524	.457	.402	.524	.477
	3	t12_19	1085	.505	.450	.384	.500	.460
	4	t12_20	1085	.584	.512	.452	.588	.534
K06	1	t12_21_11_13	1088	.993	.779	.883	.779	.858
	2	t12_22_11_14	1088	.563	.569	.430	.569	.532
	3	t12_23_11_15	1088	.548	.743	.366	.743	.600
	4	t12_24_11_16	1088	.911	.805	.752	.805	.818
K07	1	t12_25	1084	.493	.426	.379	.493	.448
	2	t12_26	1084	.271	.237	.208	.271	.247
	3	t12_27	1084	.899	.759	.694	.903	.813
	4	t12_28	1084	.445	.405	.343	.446	.410

表 7.2 等化後項目の項目母数の比較 (困難度母数)

		変数名	受検者数	(1) 同時 尺度調整法	(2) 項目 固定法	(3) 共通項目 (M&S)	(5) 推定 母集団	平均値
K01	1	t12_1	1093	-1.687	-1.978	-2.070	-1.673	-1.852
	3	t12_3	1093	.495	.527	.764	.505	.573
	4	t12_4	1093	.806	.877	1.164	.812	.915
K02	1	t12_5	1092	.012	-.009	.134	.021	.039
	2	t12_6	1092	2.278	2.532	3.073	2.279	2.541
	4	t12_8	1092	1.014	1.109	1.436	1.021	1.145
K03	1	t12_9_11_9	1091	-.912	-1.081	-.970	-1.081	-1.011
	2	t12_10_11_10	1091	-1.010	-1.264	-1.056	-1.264	-1.148
	3	t12_11_11_11	1091	.272	.263	.475	.263	.318
	4	t12_12_11_12	1091	.982	1.354	1.221	1.354	1.228
K04	1	t12_13	1087	-.401	-.470	-.403	-.392	-.417
	2	t12_14	1087	1.018	1.144	1.434	1.019	1.154
	3	t12_15	1087	-.779	-.921	-.888	-.765	-.838
	4	t12_16	1087	-.936	-1.100	-1.093	-.922	-1.013
K05	1	t12_17	1085	.987	1.063	1.405	.997	1.113
	2	t12_18	1085	-.456	-.542	-.472	-.445	-.479
	3	t12_19	1085	-.655	-.758	-.736	-.648	-.699
	4	t12_20	1085	-.190	-.237	-.126	-.179	-.183
K06	1	t12_21_11_13	1088	-1.246	-1.464	-1.374	-1.464	-1.387
	2	t12_22_11_14	1088	-1.107	-.990	-1.391	-.990	-1.120
	3	t12_23_11_15	1088	-.660	-.628	-.773	-.628	-.672
	4	t12_24_11_16	1088	-.095	-.073	-.015	-.073	-.064
K07	1	t12_25	1084	-1.431	-1.667	-1.740	-1.419	-1.564
	2	t12_26	1084	-.450	-.538	-.466	-.441	-.474
	3	t12_27	1084	-1.360	-1.605	-1.644	-1.345	-1.488
	4	t12_28	1084	1.720	1.873	2.351	1.725	1.917

7.2 学力特性値の年度間比較

従来、テストの得点は、各問題項目の正答数、もしくは、その正答数に配点による重みづけを行い、表現されてきた。しかしながら、昨年度の報告書で示されたように、この正答数や、配点による重みづけを行った正答数で表現されるテスト得点を経年比較に使う場合、その得点差に、項目の困難度を反映してしまう危険性がある。すなわち、年度間での正答数の変化が、実際の学力差そのものではなく、その年度の調査に使用された各問題項目の難度の違いを示してしまっている恐れがある。

そこで、ここでは、数学については、2010～2013年度の3年分のデータを基に、リーディング・リテラシーについては、2011年度及び2012年度の2年分のデータを基に、年度間の正答率の比較及び、IRTによる等化を行ったもとの、学力特性値の比較を行う。

7.2.1 数学

2010年度及び2011年度、2012年度のテスト得点を（配点をすべて1として）算出し、比較してみる。分冊ごとに項目数が異なるため、テスト得点を受検した項目数で割った「正答率」で表現する。年度ごとの各受検者の正答率について、ヒストグラムで表示したものが図7.2.1.1である。通常ヒストグラムは棒グラフで表現されるが、年度間の比較を容易にするために、推定密度で表現したものを表示している。また、箱ひげ図は図7.2.1.2に、年度ごとの正答率の平均値及び標準偏差を表7.2.1.1に示す。

ヒストグラムから分かるように、2010年度に比べ、2011年度及び2012年度の受検者の正答率は大きく左に偏っている、正答率の平均値も、低下していることがわかる。箱ひげ図からも、2010年度と、2011年度及び2012年度の正答率のずれは明確である。具体的には2010年度から2011年度にかけて正答率が大幅に低下し、2011年度から2012年度にかけてはわずかな低下に留まっている。しかしこの変化の要因がそもそも学力の変化なのか問題の入れ替えによるテストの難易度の変化のためなのかは、正答率からだけでは原理的に特定できないことには注意が必要である。

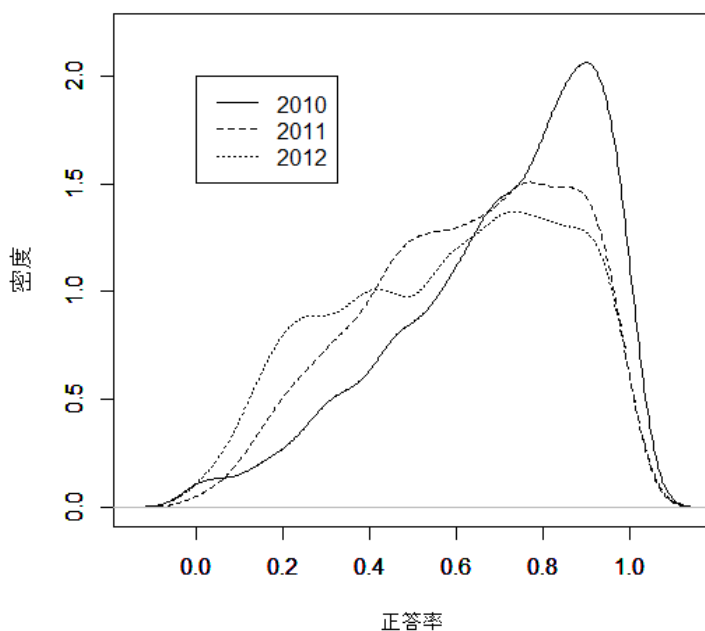


図 7.2.1.1 数学の正答率のヒストグラム（推定密度関数）

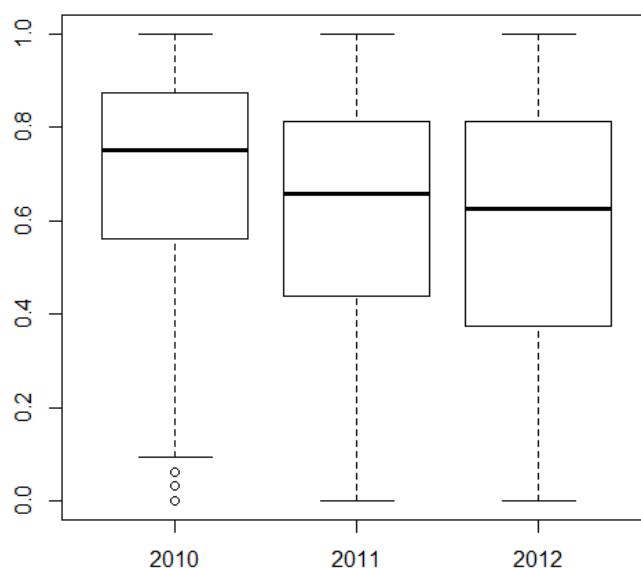


図 7.2.1.2 数学の正答率の箱ひげ図

表 7.2.1.1 数学の正答率の平均・標準偏差

	2010 年度	2011 年度	2012 年度
平均	0.696	0.624	0.588
標準偏差	0.234	0.235	0.258

次に、IRT による等化を行ったもとの学力特性値の比較をしてみる。図 7.2.1.1, 7.2.1.2 と同様に、各年度の学力特性値のヒストグラムと箱ひげ図を示したのが図 7.2.1.3, 7.2.1.4 である。学力特性値の推定には最尤推定法 (MLE) を採用し、全問正答・誤答の受検者は計算から除外している。また、学力特性値の平均・標準偏差を表 7.2.1.2 に示す。

図 7.2.1.3, 7.2.1.4, 表 7.2.1.2 から分かるように、2012 年度の受検者の特性値が、2010 年度及び 2011 年度の受検者の特性値と比べ、わずかに低下している。自治体ごとの比較はこの調査研究の目的ではないので詳細な結果は本報告書に掲載しないが、自治体ごとの分布でも同様の現象が両年度の参加のいずれの自治体でも起こっている。このことと、昨年度の実施時期が 11 月中旬、今年度の実施時期が 10 月中旬であったことを考え合わせると、実施時期が本年度約 1 ヶ月程度昨年度よりも早かったことによるものと考えるのが妥当である。このことは、逆に、中学 3 年生のこの時期の学力の伸びを問題の入れ替えがあっても学力特性値を使うことによって、とらえることができていることを意味する。くり返しになるが問題が入れ替わった場合、正答率からだけではこのような判断は不可能である。

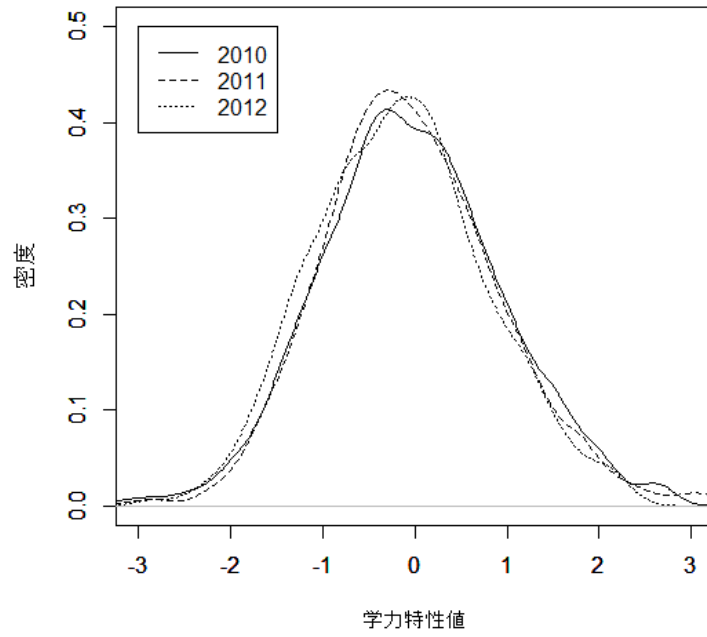


図 7.2.1.3 数学の学力特性値 (MLE) のヒストグラム (推定密度関数)

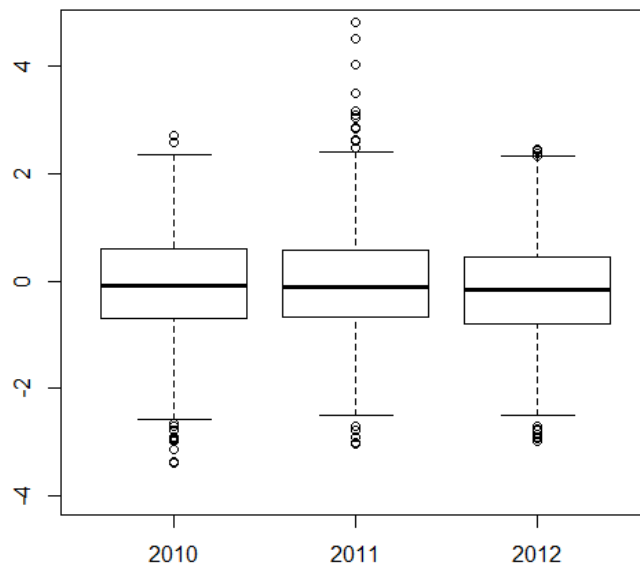


図 7.2.1.4 数学の学力特性値 (MLE) の箱ひげ図

表 7.2.1.2 数学の学力特性値 (MLE) の平均・標準偏差

	2010 年度	2011 年度	2012 年度
平均	-0.033	-0.026	-0.153
標準偏差	0.985	0.978	0.931

最後に、学力特性値の推定方法を EAP (expected a posterior) 法で行ったものの比較を、図 7.2.1.5, 7.2.1.6, 表 7.2.1.3 に示す。EAP 法では、学力特性値の事前分布を仮定することにより、全問正答・誤答の受検者に対しても、推定値を割り当てることができる。EAP 法による結果は、MLE 法による結果とほぼ同様のものとなっていることがわかる。

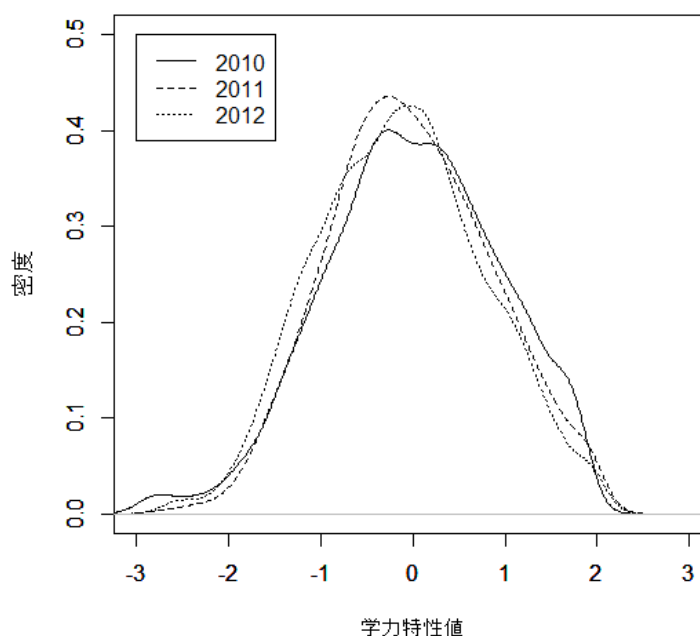


図 7.2.1.5 学力特性値 (EAP) のヒストグラム (推定密度関数)

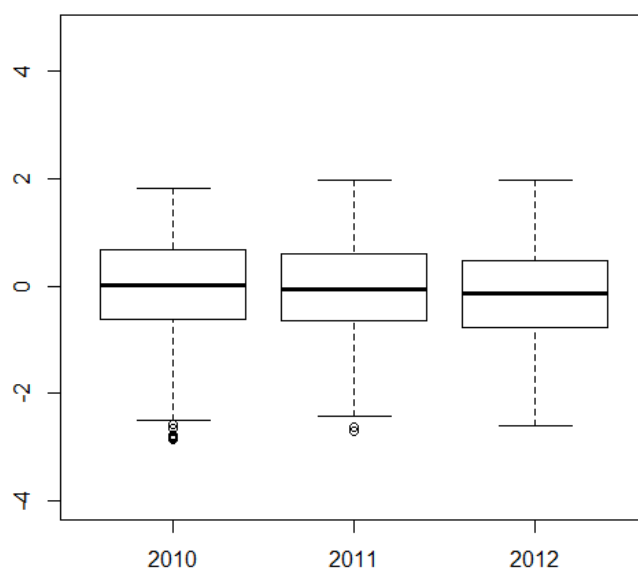


図 7.2.1.6 学力特性値 (EAP) の箱ひげ図

表 7.2.1.3 学力特性値 (EAP) の平均・標準偏差

	2010 年度	2011 年度	2012 年度
平均	-0.000	-0.019	-0.126
標準偏差	0.945	0.879	0.900

7.2.2 リーディング・リテラシー

数学と同様に、リーディング・リテラシーについても、2011 年度及び 2012 年度のデータを使い、正答率及び等化を行ったもとの学力特性値の比較を行った。まず、年度ごとの各受検者の正答率について、ヒストグラムで表示したものを図 7.2.2.7、箱ひげ図で示したものを図 7.2.2.8、平均値及び標準偏差を表 7.2.2.4 に示した。

2012 年度のヒストグラムが、2011 年度と比べて左に偏っていることから分かるように、2011 年度よりも、2012 年度の受検者の正答率が 4.3%ほど低下していることがわかる。箱ひげ図のずれからも、通常であれば、学力が低下したと捉えてしまう結果が得られている。

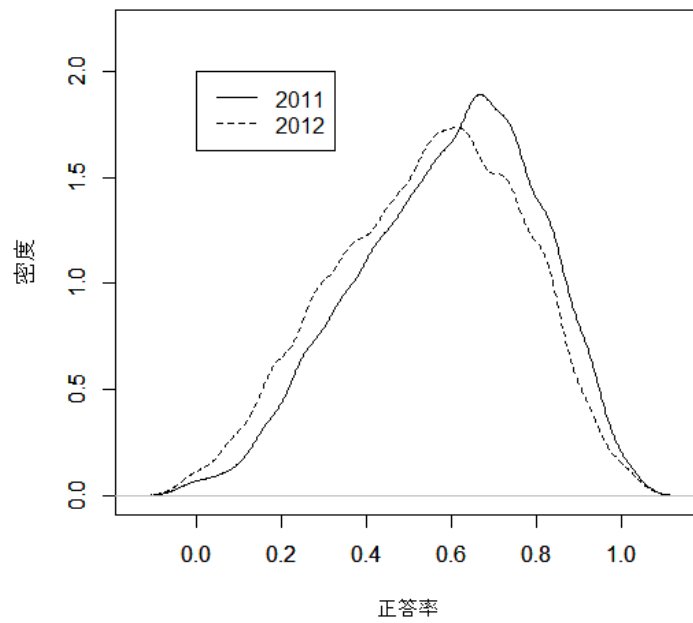


図 7.2.2.7 リーディング・リテラシーの正答率のヒストグラム（推定密度関数）

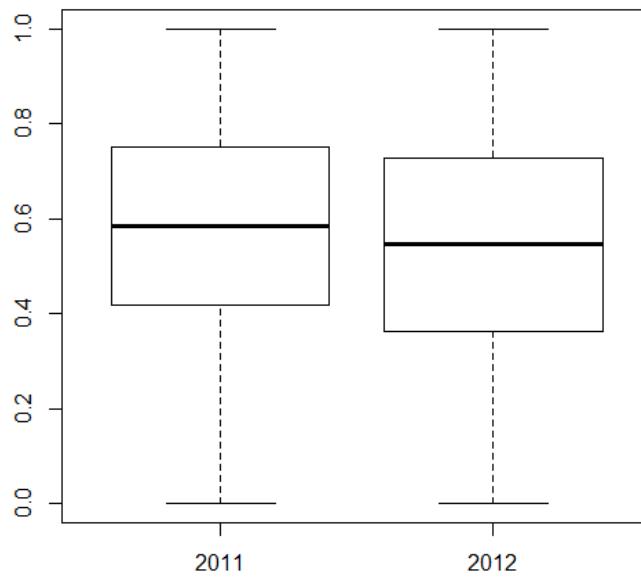


図 7.2.2.8 リーディング・リテラシーの正答率の箱ひげ図

表 7.2.2.4 正答率の平均・標準偏差

	2011 年度	2012 年度
平均	0.590	0.543
標準偏差	0.209	0.217

次に、IRT による等化を行ったもとの学力特性値を比較する。まずは全問正答・誤答の受検者を計算から除外し、最尤推定法（MLE）を採用して、推定した学力特性値のヒストグラムを図 7.2.2.9、箱ひげ図を図 7.2.2.10、平均値及び標準偏差を表 7.2.2.5 に示す。

図 7.2.2.9、図 7.2.2.10、表 7.2.2.5 から分かれるとおり、年度間での受検者の学力特性値には大きな差がみられない。昨年度と今年度の実施時期の違いのため数学と同様の現象がおきているため、若干、平均は今年度のものが小さくなっているが顕著ではない。これは数学にくらべて国語の信頼性が相対的に低いため、数学ほどはその傾向が明確にはならなかったためと考えられる。このように、昨年度の報告書および、先述した数学の結果と同様に、リーディング・リテラシーのデータでも、学力特性値を利用することによって、正答率にからだけでは取得することの不可能な学力分布に関する情報が得られるのである。

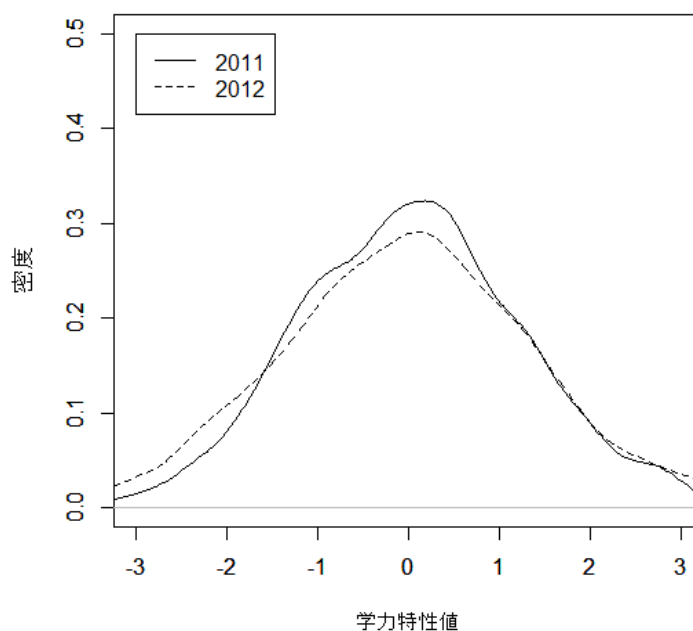


図 7.2.2.9 学力特性値（MLE）のヒストグラム（推定密度関数）

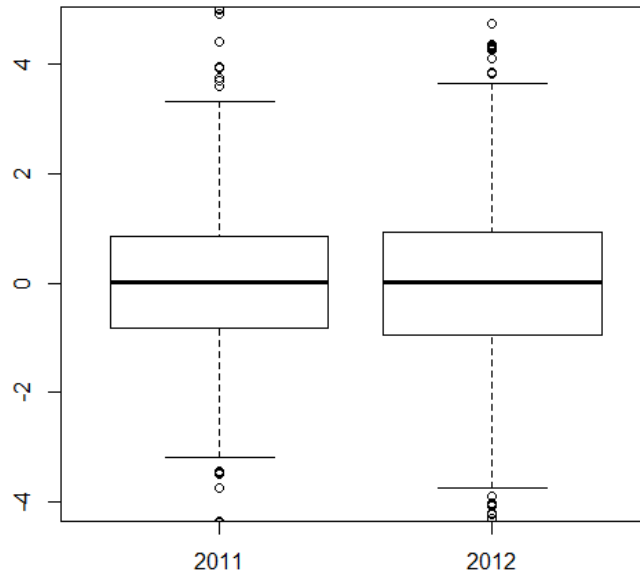


図 7.2.2.10 学力特性値 (MLE) の箱ひげ図

表 7.2.2.5 学力特性値 (MLE) の平均・標準偏差

	2011 年度	2012 年度
平均	0.074	0.018
標準偏差	1.303	1.503

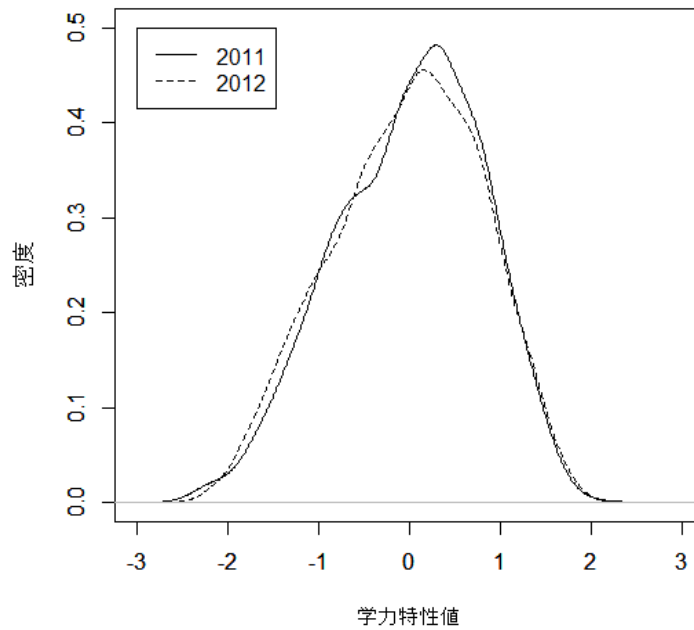


図 7.2.2.11 学力特性値 (EAP) のヒストグラム (推定密度関数)

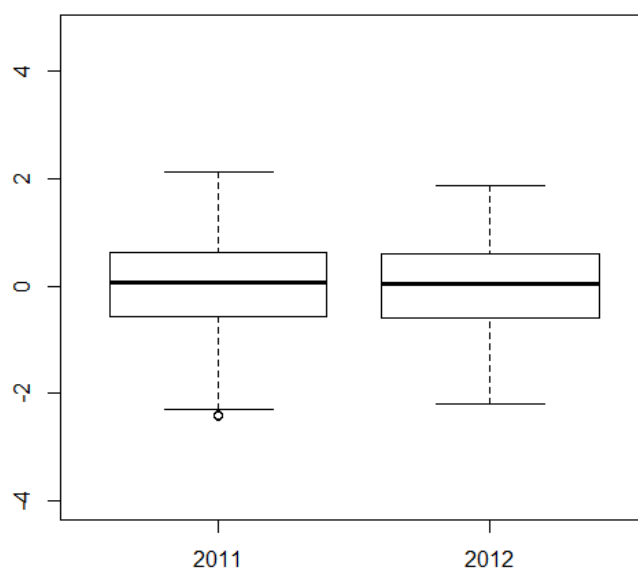


図 7.2.2.12 学力特性値（EAP）の箱ひげ図

表 7.2.2.6 学力特性値（EAP）の平均・標準偏差

	2011 年度	2012 年度
平均	0.000	-0.030
標準偏差	0.809	0.826

7.3 BIBD 導入の効果

ここでは数学をとりあげ、BIBD 導入の効果について確認しておく。第 2 章で述べたように平成 23 年度数学の分冊デザインは BIBD ではない。そこで BIBD である平成 24 年度との比較のために、分冊ごとの平均と標準偏差を対比したのが表 7.3、またそれを箱ひげ図で示したのが図 7.3.1 及び図 7.3.2 である。

分冊ごとの特性値の平均に関しても、標準偏差に関しても、平成 24 年度の方が平成 23 年度よりも散らばりが小さいことがわかる。このことは分冊ごとに割り当てられた生徒群の学力分布が平成 24 年度の方がたがいに等質なことを意味している。実際、平成 24 年度のデータを使って一元配置の分散分析を行っても生徒数が 2539 名と自由度が非常に大きいにもかかわらず、分冊間の学力特性値の平均に統計的な有意差はなかった($p=0.22$)。分冊デザインとして BIBD を導入した効果といえるであろう。

表 7.3 平成 23 年度および平成 24 年度における分冊ごとの統計量（数学）

平成23年度

分冊	S1	S2	S3	S4	S5	S6	S7	S8
人数	256	257	252	251	243	238	233	256
平均	-.048	.035	.065	-.108	-.017	.053	.129	-.109
SD	.892	.997	.941	.885	.971	.886	1.000	.877

平成24年度

分冊	S1	S2	S3	S4	S5	S6	S7
人数	368	356	357	356	369	366	367
平均	-.158	-.045	-.047	-.161	-.180	-.143	-.145
SD	.885	.869	.918	.916	.886	.941	.881

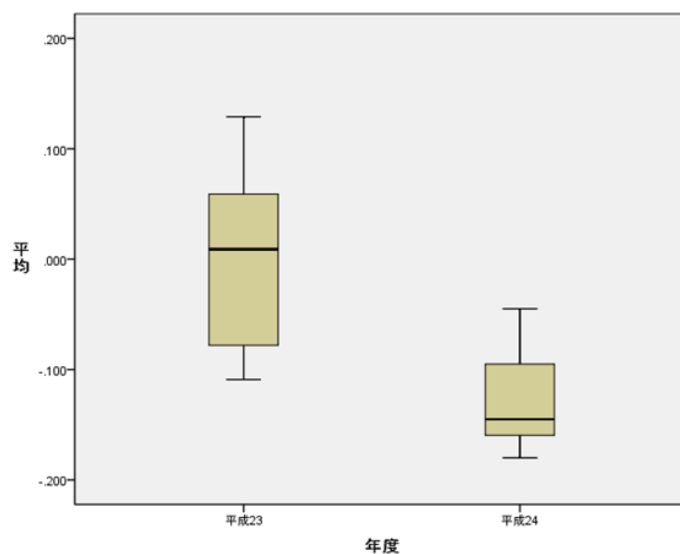


図 7.3.1 平成 23 年度および平成 24 年度の分冊ごとの平均

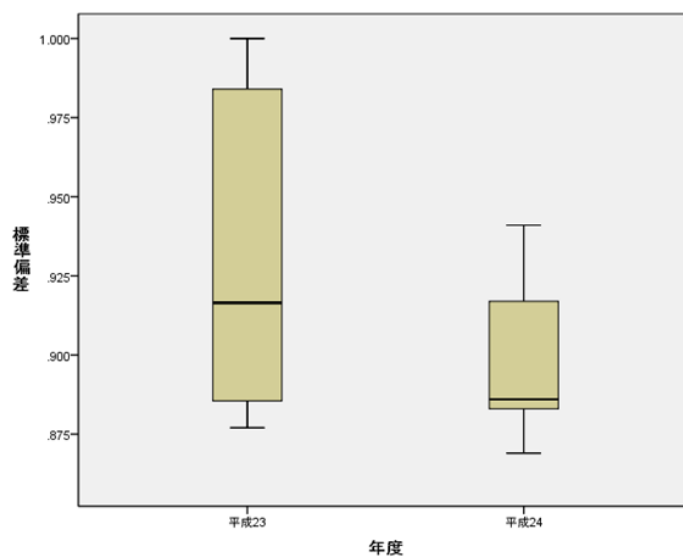


図 7.3.1 平成 23 年度および平成 24 年度の分冊ごとの標準偏差

8. 生徒質問紙の結果

生徒質問紙の詳細については 資料 5.2.5 を参照されたい。ここでは個別の項目に関する集計結果を報告する。

q1 は 13 項目で構成されており、生徒の日々の生活や勉強の様子について尋ねたものである。このうち、q1-1 から q1-3 は主に家族、家庭のことについて、q1-4 から q1-6 では娯楽について、q1-7 から q1-13 では勉強についての項目となっている。

q1-1 によると、4 人、5 人で住んでいる生徒だけで半数以上を占めている (図 8.1)。q1-2 によると、2 人きょうだいの生徒が半数以上であり、3 人きょうだいを含めると 8 割を超える (図 8.2)。このことから、両親と生徒自身、きょうだいの核家族の形態で住んでいる生徒が多いと考えられる。また、q1-3 より、家に車が 2 台あるという生徒が半数以上であり、免許所有者が 1 人 1 台の車を所有していると考えられる (図 8.3)。

q1-4 によると、ほとんどの生徒は家にテレビがある (図 8.4)。q1-5 からは、9 割以上の生徒がゲーム機を所有しており、その中でも特に 6 台以上のゲーム機を所有する生徒の割合が高いことがわかる (図 8.5)。q1-6 では、旅行に行く機会がない生徒は 2 割に満たないほどであり、年に 1、2 回程度旅行に行く生徒と夏休みなどの休みの度に旅行に行くという生徒の合計が 7 割近くであることがわかる (図 8.6)。このことから、多くの家庭では屋内の娯楽だけではなく、遠出する機会を作っていることがわかる。

q1-7 によると、平日に 1 時間以上勉強する生徒は 8 割を超えており、その中でも 2 時間勉強するという生徒の割合が最も高い (図 8.7)。q1-8 では、本を読む習慣がある生徒が 9 割以上であることがわかる (図 8.8)。q1-9 からは、習い事をしている生徒は 8 割を超えており、その中でも特に習い事を 1 つしている生徒が多いことが読み取れる (図 8.9)。q1-10 によると、7 割の生徒が書斎や本棚といった本専用のスペースを確保しており、読書の習慣のあるほとんどの生徒は本を置く場所を決めていることがわかる (図 8.10)。q1-11 によると、大学までの進学を考えている生徒が半数近くおり、次いで高校までの進学を考えている生徒が多い (図 8.11)。q1-12 によると、勉強を教えてくれるのは学習塾と答えた生徒が最も多い (図 8.12)。このことから、生徒がしている習い事は学習塾が多いということが読み取れる。また、学習塾の次に多い回答は友人であり、親や兄弟、姉妹といった家族は 3 番目と 4 番目に多い回答となっている。q1-13 では、学校以外での勉強は学習塾で行っていると答えた生徒が最も多かった (図 8.13)。

q2 は 12 項目で構成されており、数学や国語を勉強する理由を 5 件法で尋ねたものである (具体的には「資料 5.2.5 協力依頼状：生徒質問紙調査」のⅡが、その内容となる)。

q2-1 によると、立派な大人になりたいと考える生徒は 6 割ほどであった (図 8.14)。q2-2 では、問題を解くことがおもしろいと答えた生徒はそう思わない生徒よりも少ない (図 8.15)。また、どちらでもないと答えた生徒が多かった。q2-3 によると、勉強しないと何となく不安だと答えた生徒は 6 割を超えている (図 8.16)。q2-4 では、頭を使うことが好きだと思える生徒は少ないことが読み取れる (図 8.17)。q2-5 によると、よい高校や大学に入りたいと考える生徒は 6 割を超えており、高校や大学への進学を希望している生徒の多くは勉強をしてよい学校に入学したいと考えていることがわかる (図

8.18) . q2-6 では、通知表をよくしたいと答えた生徒が 6 割を超えていることがわかる (図 8.19) . q2-7 によると、大人になって必要だと思う生徒は 6 割を超えている (図 8.20) . q2-8 では、わかることが楽しいと答えた生徒は 4 割を超えているが、どちらでもないと答えた生徒は 3 割と多いことがわかる (図 8.21) . q2-9 では、新しいことを知ることができるかと答えた生徒は 4 割に満たず、どちらでもないと答えた生徒が多かった (図 8.22) . q2-10 によると、自分の力をためすことができると答えた生徒は 3 割ほどであり、どちらでもないと答えた生徒の割合が最も多かった (図 8.23) . q2-11 では、テストでよい点がとりたいと答えた生徒が 7 割を超えていたことがわかる (図 8.24) . q2-12 によると、できるようになるのがおもしろいと答えた生徒は 5 割ほどである (図 8.25) . どちらでもないと答えた生徒が多い。以上のことから、問題を解くことがおもしろい、頭を使うことが好き、わかることが楽しい、新しいことを知ることができる、自分の力をためすことができる、できるようになることがおもしろいといった勉強すること自体に興味を持ち、楽しんでいる生徒が少ないことがわかる。

q3 は 6 項目で構成されており、友人や家族、周囲の人たちとの様子について尋ねたものである。

q3-1 では、大きな声であいさつができる生徒は 8 割ほどであることがわかる (図 8.26) . q3-2 によると、学校に親しい友達がいると答えた生徒は 9 割を超えており、その内親しい友達がたくさんいると答えた生徒は 5 割を超えている (図 8.27) . q3-3 によると、引っ越しなどで会えなくなった友達がいる生徒は 7 割ほどである (図 8.28) . q3-4 では、学校での出来事などについて家族と話をする生徒は 7 割を超えている (図 8.29) . q3-5 によると、家の掃除を手伝う生徒はほぼ 9 割である (図 8.30) . q3-6 では、知らない人に道を尋ねられた際に教えてあげる生徒は 7 割ほどであることがわかる (図 8.31) . 以上のことから、ほとんどの生徒が挨拶や会話によって周囲とうまくコミュニケーションをとることができており、学校において多くの親しい友人を持つことができていると考えられる。

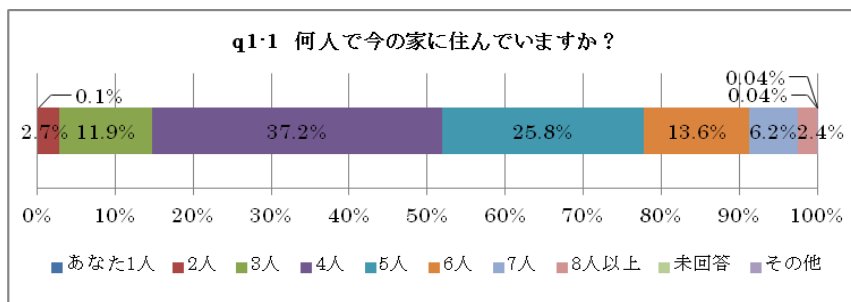


図 8.1 q1-1 の各選択枝の選択率

表 8.1 q1-1 の度数分布表

選択枝	度数
あなた 1 人	3
2 人	69
3 人	302
4 人	946
5 人	655
6 人	345
7 人	157
8 人以上	62
未回答	1
その他	1
合計	2541

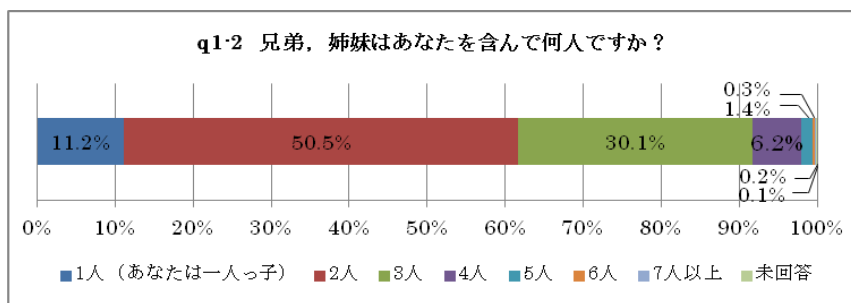


図 8.2 q1-2 の各選択枝の選択率

表 8.2 q1-2 の度数分布表

選択枝	度数
1 人	284
2 人	1284
3 人	764
4 人	157
5 人	36
6 人	8
7 人以上	5
未回答	3
合計	2541

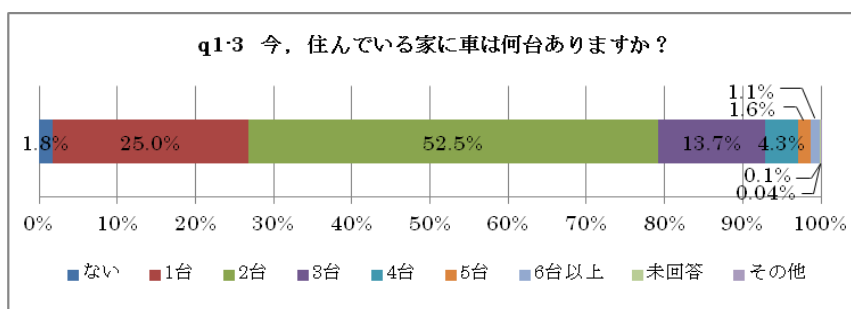


図 8.3 q1-3 の各選択枝の選択率

表 8.3 q1-3 の度数分布表

選択枝	度数
ない	45
1 台	634
2 台	1334
3 台	347
4 台	108
5 台	41
6 台以上	28
未回答	3
その他	1
合計	2541

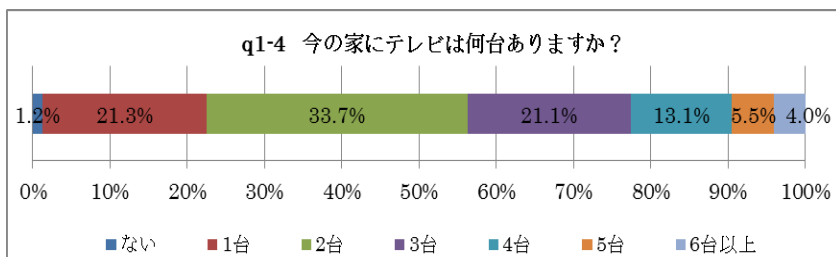


図 8.4 q1-4 の各選択枝の選択率

表 8.4 q1-4 の度数分布表

選択枝	度数
ない	31
1台	542
2台	856
3台	537
4台	334
5台	139
6台以上	102
合計	2541

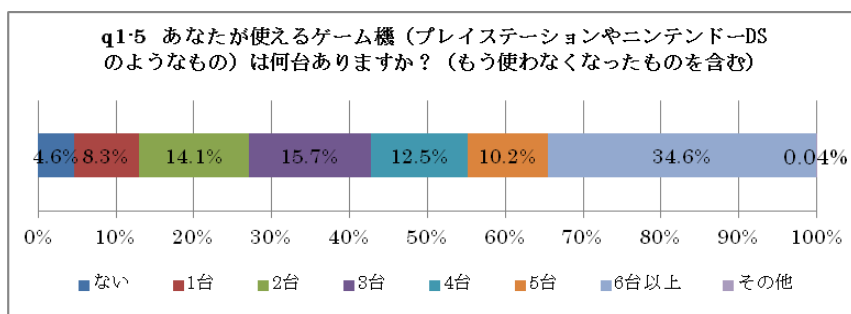


図 8.5 q1-5 の各選択枝の選択率

表 8.5 q1-5 の度数分布表

選択枝	度数
ない	118
1台	210
2台	359
3台	399
4台	317
5台	259
6台以上	878
その他	1
合計	2541

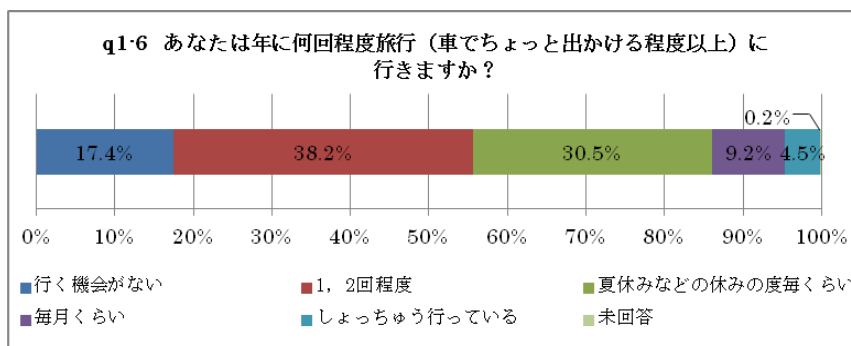


図 8.6 q1-6 の各選択枝の選択率

表 8.6 q1-6 の度数分布表

選択枝	度数
行く機会がない	443
1, 2回程度	970
夏休みなどの休みの度毎くらい	775
毎月くらい	235
しょっちゅう行っている	114
未回答	4
合計	2541

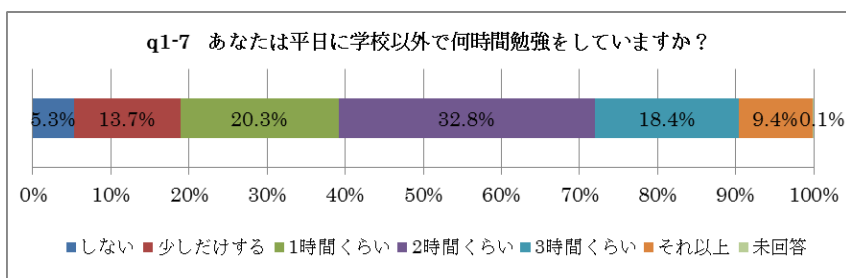


図 8.7 q1-7 の各選択枝の選択率

表 8.7 q1-7 の度数分布表

選択枝	度数
しない	134
少しだけする	347
1時間くらい	516
2時間くらい	833
3時間くらい	468
それ以上	240
未回答	3
合計	2541

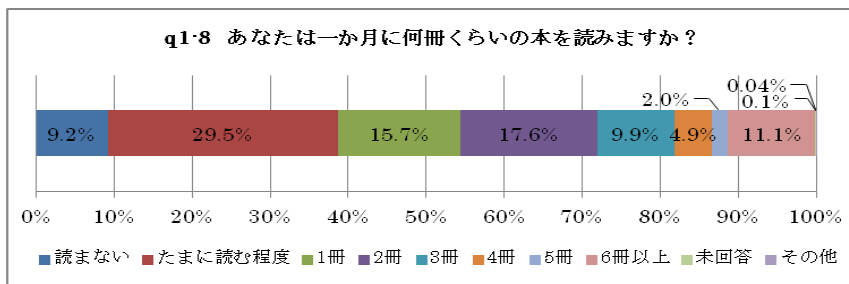


図 8.8 q1-8 の各選択枝の選択率

表 8.8 q1-8 の度数分布表

選択枝	度数
読まない	233
たまに読む程度	749
1冊	399
2冊	446
3冊	251
4冊	124
5冊	52
6冊以上	283
未回答	3
その他	1
合計	2541

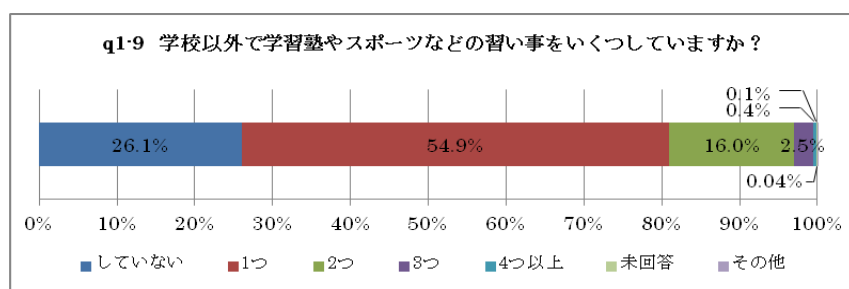


図 8.9 q1-9 の各選択枝の選択率

表 8.9 q1-9 の度数分布表

選択枝	度数
していない	662
1つ	1396
2つ	407
3つ	64
4つ以上	9
未回答	2
その他	1
合計	2541

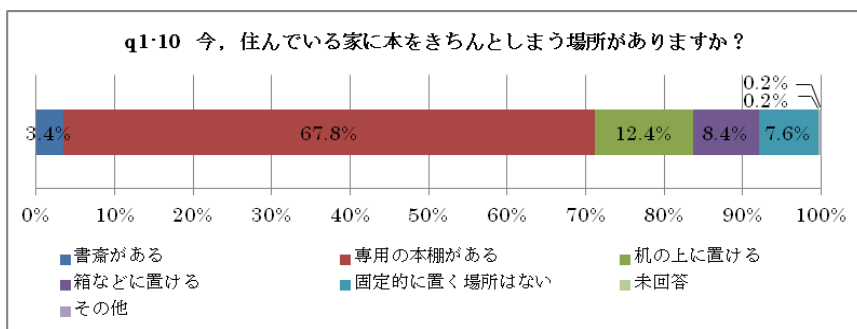


図 8.10 q1-10 の各選択枝の選択率

表 8.10 q1-10 の度数分布表

選択枝	度数
書斎がある	87
専用の本棚がある	1722
机の上に置ける	316
箱などに置ける	213
固定的に置く場所はない	194
未回答	4
その他	5
合計	2541

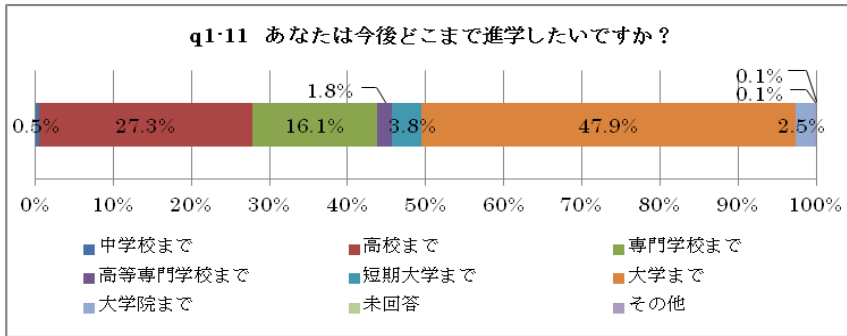


図 8.11 q1-11 の各選択枝の選択率

表 8.11 q1-11 の度数分布表

選択枝	度数
中学校まで	12
高校まで	693
専門学校まで	409
高等専門学校まで	47
短期大学まで	96
大学まで	1216
大学院まで	63
未解答	2
その他	3
合計	2541

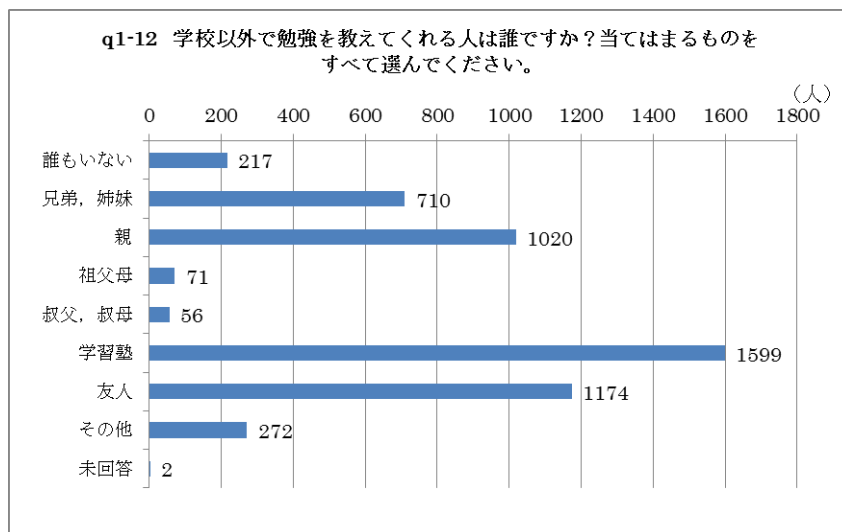


図 8.12 q1-12 の各選択枝の選択率

表 8.12 q1-12 の度数分布表

選択枝	度数
誰もいない	217
兄弟, 姉妹	710
親	1020
祖父母	71
叔父, 叔母	56
学習塾	1599
友人	1174
その他	272
未回答	2
計	5121

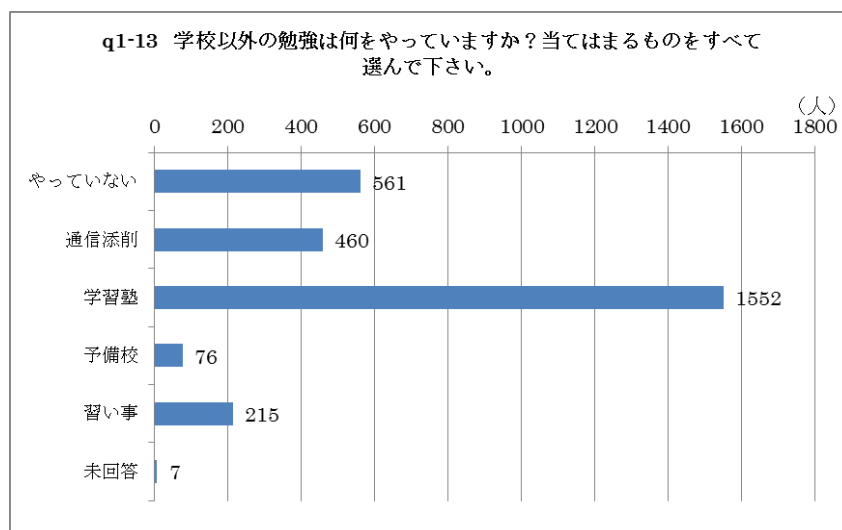


図 8.13 q1-13 の各選択枝の選択率

表 8.13 q1-13 の度数分布表

選択枝	度数
やっていない	561
通信添削	460
学習塾	1552
予備校	76
習い事	215
未回答	7
計	2871

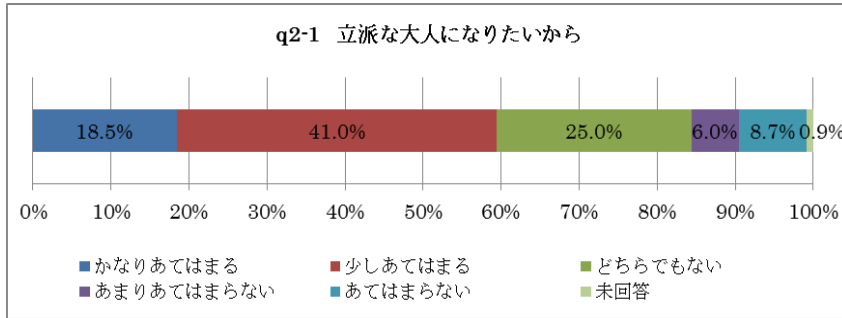


図 8.14 q2-1 の各選択枝の選択率

表 8.14 q2-1 の度数分布表

選択枝	度数
かなりあてはまる	469
少しあてはまる	1042
どちらでもない	635
あまりあてはまらない	152
あてはまらない	221
未回答	22
その他	0
計	2541

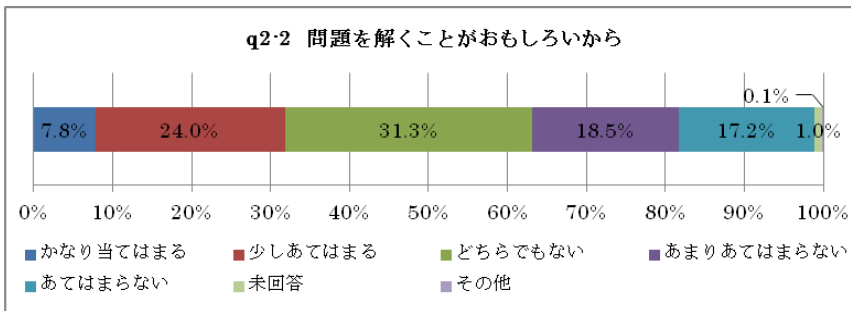


図 8.15 q2-2 の各選択枝の選択率

表 8.15 q2-2 の度数分布表

選択枝	度数
かなりあてはまる	198
少しあてはまる	611
どちらでもない	796
あまりあてはまらない	471
あてはまらない	437
未回答	25
その他	3
計	2541

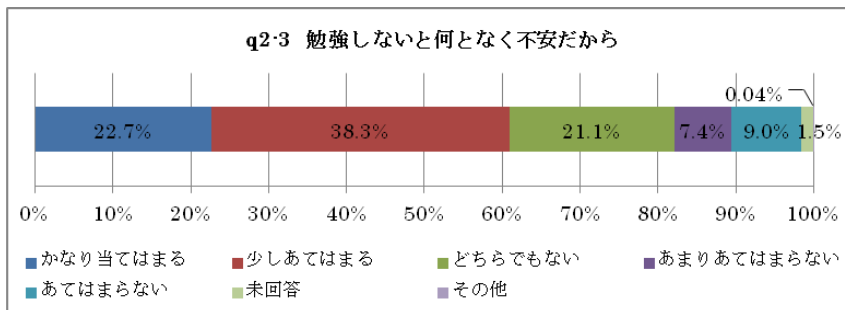


図 8.16 q2-3 の各選択枝の選択率

表 8.16 q2-3 の度数分布表

選択枝	度数
かなりあてはまる	576
少しあてはまる	973
どちらでもない	536
あまりあてはまらない	187
あてはまらない	229
未回答	39
その他	1
計	2541

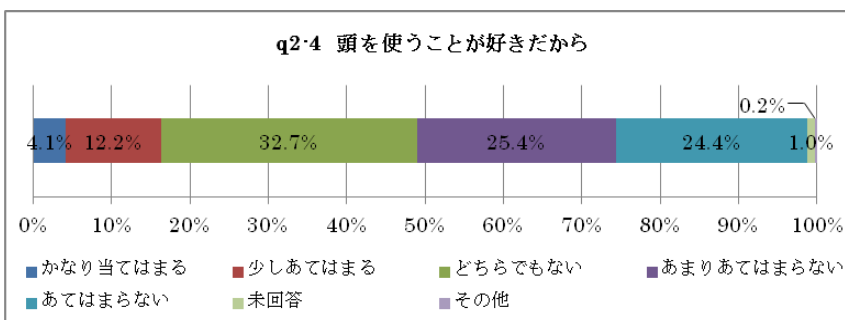


図 8.17 q2-4 の各選択枝の選択率

表 8.17 q2-4 の度数分布表

選択枝	度数
かなりあてはまる	105
少しあてはまる	309
どちらでもない	830
あまりあてはまらない	645
あてはまらない	620
未回答	26
その他	6
計	2541

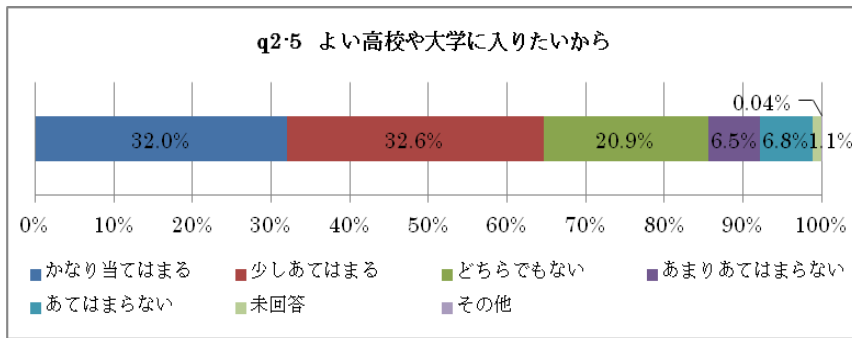


図 8.18 q2-5 の各選択枝の選択率

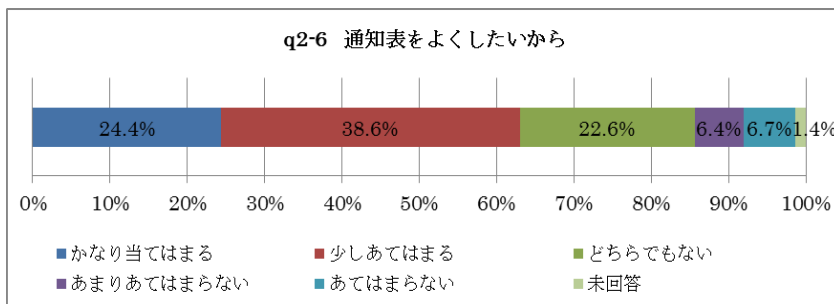


図 8.19 q2-6 の各選択枝の選択率

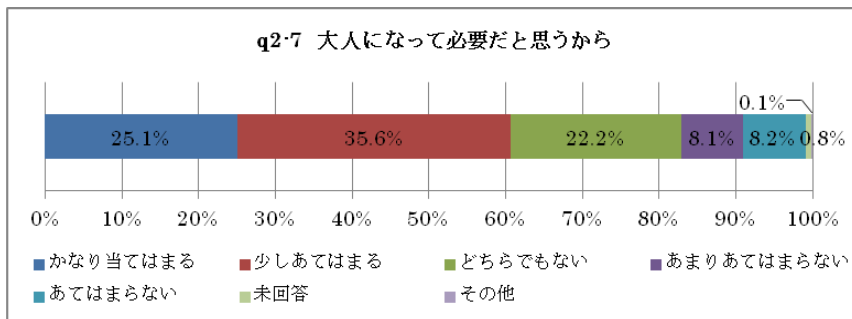


図 8.20 q2-7 の各選択枝の選択率

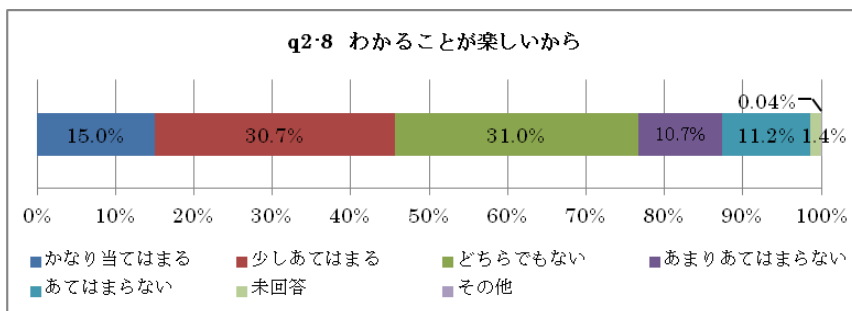


図 8.21 q2-8 の各選択枝の選択率

表 8.18 q2-5 の度数分布表

選択枝	度数
かなり当てはまる	814
少し当てはまる	829
どちらでもない	532
あまり当てはまらない	166
当てはまらない	172
未回答	27
その他	1
計	2541

表 8.19 q2-6 の度数分布表

選択枝	度数
かなり当てはまる	620
少し当てはまる	980
どちらでもない	574
あまり当てはまらない	163
当てはまらない	169
未回答	35
その他	0
計	2541

表 8.20 q2-7 の度数分布表

選択枝	度数
かなり当てはまる	637
少し当てはまる	905
どちらでもない	563
あまり当てはまらない	205
当てはまらない	208
未回答	20
その他	3
計	2541

表 8.21 q2-8 の度数分布表

選択枝	度数
かなり当てはまる	380
少し当てはまる	779
どちらでもない	788
あまり当てはまらない	272
当てはまらない	285
未回答	36
その他	1
計	2541

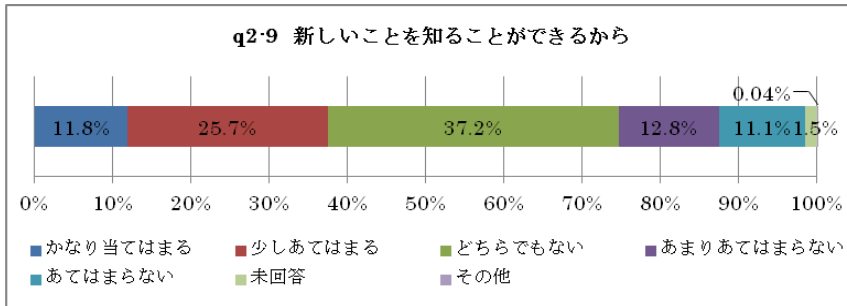


図 8.22 q2-9 の各選択枝の選択率

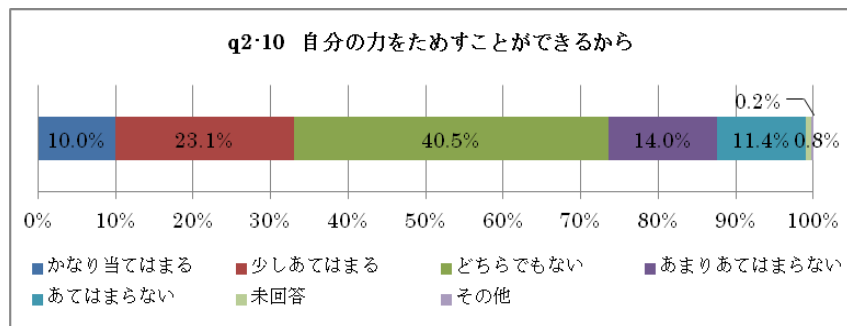


図 8.23 q2-10 の各選択枝の選択率

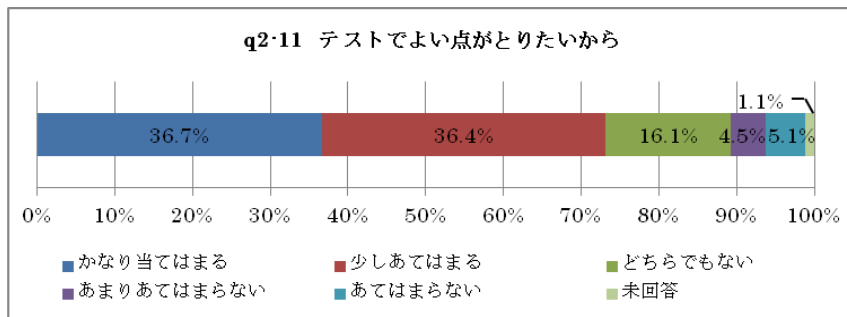


図 8.24 q2-11 の各選択枝の選択率

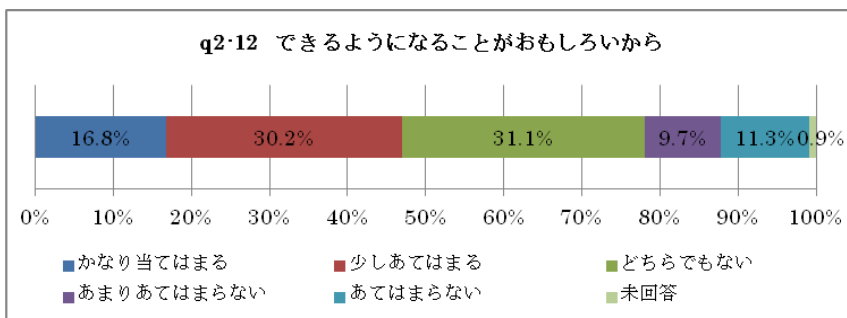


図 8.25 q2-12 の各選択枝の選択率

表 8.22 q2-9 の度数分布表

選択枝	度数
かなり当てはまる	301
少し当てはまる	652
どちらでもない	944
あまり当てはまらない	325
あてはまらない	281
未回答	37
その他	1
計	2541

表 8.23 q2-10 の度数分布表

選択枝	度数
かなり当てはまる	253
少し当てはまる	587
どちらでもない	1030
あまり当てはまらない	355
あてはまらない	290
未回答	20
その他	6
計	2541

表 8.24 q2-11 の度数分布表

選択枝	度数
かなり当てはまる	933
少し当てはまる	926
どちらでもない	408
あまり当てはまらない	115
あてはまらない	130
未回答	29
その他	0
計	2541

表 8.25 q2-12 の度数分布表

選択枝	度数
かなり当てはまる	426
少し当てはまる	768
どちらでもない	789
あまり当てはまらない	247
あてはまらない	287
未回答	24
その他	0
計	2541

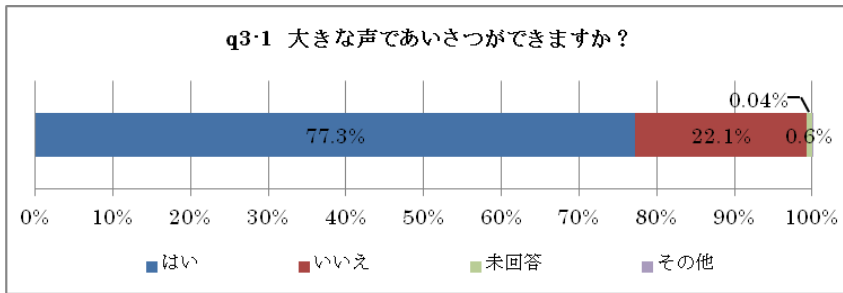


図 8.26 q3-1 の各選択枝の選択率

表 8.26 q3-1 の度数分布表

選択枝	度数
はい	1963
いいえ	561
未回答	16
その他	1
計	2541

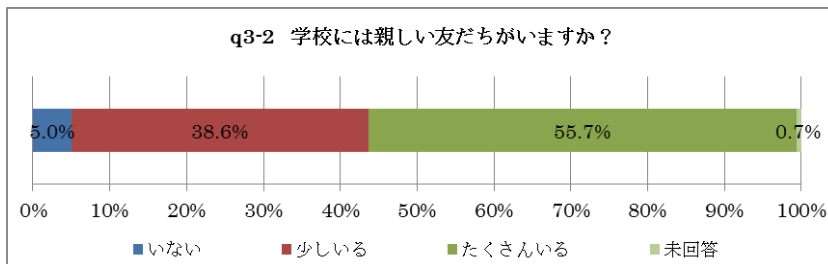


図 8.27 q3-2 の各選択枝の選択率

表 8.27 q3-2 の度数分布表

選択枝	度数
いない	127
少しいる	982
たくさんいる	1415
未回答	17
計	2541

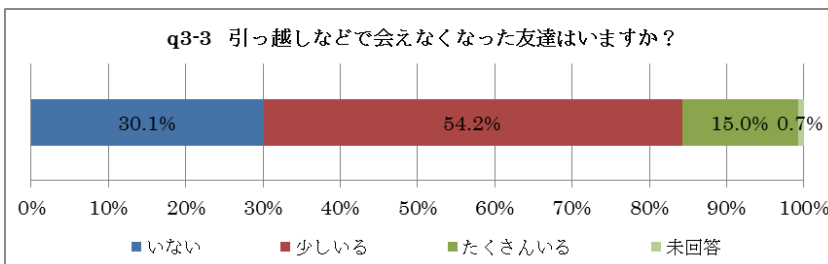


図 8.28 q3-3 の各選択枝の選択率

表 8.28 q3-3 の度数分布表

選択枝	度数
いない	764
少しいる	1376
たくさんいる	382
未回答	19
計	2541

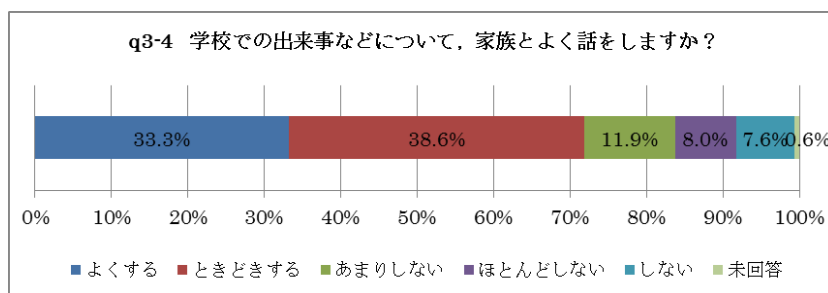


図 8.29 q3-4 の各選択枝の選択率

表 8.29 q3-4 の度数分布表

選択枝	度数
よくする	845
とくどきする	981
あまりしない	302
ほとんどしない	204
しない	193
未回答	16
計	2541

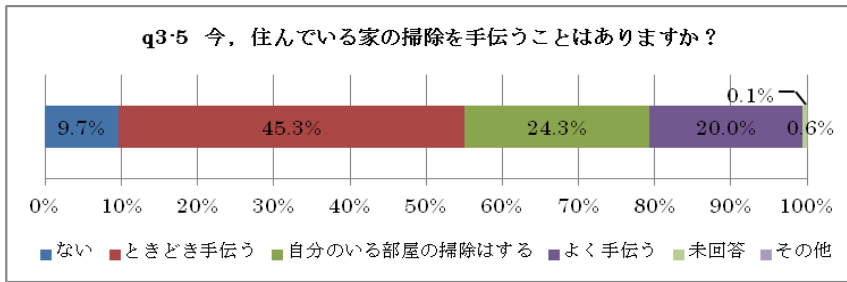


図 8.30 q3-5 の各選択枝の選択率

表 8.30 q3-5 の度数分布表

選択枝	度数
ない	246
ときどき手伝う	1151
自分のいる部屋の掃除はする	618
よく手伝う	509
未回答	15
その他	2
計	2541

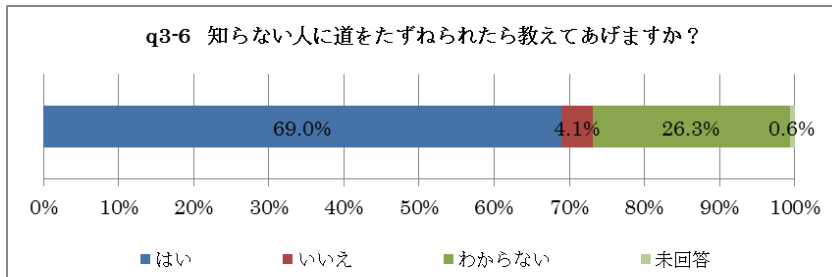


図 8.31 q3-6 の各選択枝の選択率

表 8.31 q3-6 の度数分布表

選択枝	度数
はい	1753
いいえ	104
わからない	668
未回答	16
計	2541

9. 推算値を使ったマルチレベル分析

9.1 推算値とマルチレベル分析の導入

ここでは学力調査から得られた推算値と生徒質問紙から得られた回答データに対して、マルチレベルモデル (multilevel models;たとえば Goldstein(2011),Hox(2010)) を適用し、学力の形成要因を検討することで、推算値とマルチレベル分析の全国学力・学習状況調査への応用可能性を探る。これまでの学力調査データにもとづくさまざまな研究において、ともすれば看過されがちであった問題点としては、1) クラス、学校、地域などの階層構造が必然的に含まれるデータであるにも関わらず、そのことを積極的にモデル化した分析方法が用いられてこなかったこと、2) 学力データ自体にいわゆる測定の標準誤差(standard error of measurement)が含まれているにも関わらず、そのことを考慮せずに学校別平均や標準偏差など推定しているため、が指摘できる。マルチレベル分析は前者の、推算値は後者への、いわば解決策のひとつということになる。

いいかれば、これまで学力調査の分析はさまざまな分野で行われてきてはいるものの、学力調査データのような、いくつかの「階層」がサンプリングされたデータの中に含まれるとき、従来の分析方法、たとえば、分散分析や共分散分析によっては、分析モデル中に多くのパラメータが含まれてしまうことにより、検出力が大幅に落ちてしまい、たとえば、家庭の文化資本が学力に及ぼす影響や学校風土の差など、知りたい効果がうまく抽出できないなどの問題が指摘されている (Luke, D.A., 2004) 。それに対して、マルチレベルモデルを用いることにより、階層の効果を、モデルにおいて変量効果として切片と係数に仮定することで、その効果を扱うことが可能となる。また、推算値を用いることにより、従来の 1 冊子一斉方式と同様に、重複テスト分冊法が抱えていた児童生徒の尺度値から直接的には正確な母集団統計量が推定できていないという本質的な問題も克服できる

以下では、数学の推算値をマルチレベルモデルにおける従属変数、質問紙からえら得たデータを独立変数として、学力の形成要因を検証する。

生徒	推算値	質問項目				
		1	...	j	...	m
生徒 1	PV1_1	X11,	...	X1j,	...	X1m
	PV2_1	X11,	...	X1j,	...	X1m
	PV3_1	X11,	...	X1j,	...	X1m
	PV4_1	X11,	...	X1j,	...	X1m
	PV5_1	X11,	...	X1j,	...	X1m
}						
生徒 i	PV1_i	Xi1,	...	Xij,	...	Xim
	PV2_i	Xi1,	...	Xij,	...	Xim
	PV3_i	Xi1,	...	Xij,	...	Xim
	PV4_i	Xi1,	...	Xij,	...	Xim
	PV5_i	Xi1,	...	Xij,	...	Xim
}						
生徒 n	PV1_n	Xn1,	...	Xnj,	...	Xnm
	PV2_n	Xn1,	...	Xnj,	...	Xnm
	PV3_n	Xn1,	...	Xnj,	...	Xnm
	PV4_n	Xn1,	...	Xnj,	...	Xnm
	PV5_n	Xn1,	...	Xnj,	...	Xnm

図 9.1 推算値を利用する場合のデータ構造

9.2 分析の枠組み

今回は生徒レベルと学校レベルを同時に分析するためにマルチレベルモデルを用いる。その際、ヌルモデルとして

$$\begin{aligned} \text{レベル 1 (生徒レベル)} \quad & \text{数学} PV_{ij} = \beta_{0j} + e_{ij} \quad \text{Var}(e_{ij}) = \sigma^2_{e_{ij}} \\ \text{レベル 2 (学校レベル)} \quad & \beta_{0j} = \beta_0 + u_{0j} \quad \text{Var}(u_{0j}) = \sigma^2_{u_{0j}} \end{aligned}$$

i : 個人 j : 学校

を設定する。ヌルモデルは説明変数は投入せず、切片にのみ変量効果を仮定するモデルである。ここで、級内相関係数 (Intraclass Correlation Coefficient) ,

$$\text{ICC (1)} = \frac{\sigma^2_{u_{0j}}}{\sigma^2_{u_{0j}} + \sigma^2_{e_{ij}}}$$

を用いることで、従属変数における分散のうち、レベル 2 が説明する割合を評価することができる⁶ (Luke, D.A., 2004) .

次に、学力の規定要因として説明変数を投入していくことになるが、

$$\begin{aligned} \text{レベル 1 (生徒レベル)} \quad & \text{数学} PV_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij} \\ \text{レベル 2 (学校レベル)} \quad & \beta_{0j} = \beta_0 + u_{0j} \end{aligned}$$

というように切片にのみ学校間差を仮定するモデルを一般に「ランダム切片モデル」と呼ぶ。言い換えれば、そもそも学力調査を実施した段階で数学の学力にのみ学校間差があるが、それぞれの説明変数からの効果は学校間で一定であることを仮定している。また、この場合、説明変数は1つしか投入していないが、重回帰分析のように説明変数を複数投入することも可能である。

さらに、

$$\begin{aligned} \text{レベル 1 (生徒レベル)} \quad & \text{数学} PV_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij} \\ \text{レベル 2 (学校レベル)} \quad & \beta_{0j} = \beta_0 + u_{0j} \\ \text{レベル 2 (学校レベル)} \quad & \beta_{1j} = \beta_1 + u_{1j} \end{aligned}$$

⁶ マルチレベル分析において、レベル 2 における変量効果を仮定する際、集団レベルの変数を説明変数として、

$$\begin{aligned} \text{レベル 1 (生徒レベル)} \quad & y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij} \\ \text{レベル 2 (学校レベル)} \quad & \beta_{0j} = \beta_0 + \beta_{01} X_j + u_{0j} \end{aligned}$$

などのようにすることも可能である。

特に、集団レベルの変数として個人レベルの得点を集計した平均値を用いる際、

$$\text{ICC(2)} = \frac{k \times \text{ICC(1)}}{1 + (k-1) \times \text{ICC(1)}}$$

k : その集団の平均人数

によって、その信頼性を評価することが知られている (Bliese, 2000) . しかし、今回は集団レベルの変数を用いていないため、ICC (1) のみで評価する。

というように切片，係数ともに学校間差を仮定するモデルは「ランダム切片係数モデル」と呼ぶ。「ランダム切片モデル」では切片にのみ学校間差を仮定していたが，このモデルでは係数にも変量効果を仮定している．言い換えれば，それぞれの説明変数の効果についても学校間差があることを仮定するモデルであると言える．

9.3 分析方法

今回の分析にはオープンソースの統計ソフトウェア R2.15.3 を用い，因子分析には `factanal` 関数，マルチレベル分析には `lme` 関数を使用した．使用するデータは，「平成 24 年度 学力調査 生徒質問紙」（以下，「生徒質問紙」）「平成 24 年度 学力調査 学校質問紙」（自由記述式，以下「学校質問紙」）である．欠測値処理を施し分析対象とされたのは 2,358 名（推算値を 5 つ発生させているため， $N=11,740$ ($2,358 \times 5$))，学校数は 25 となった．その中から学力の形成要因を探るために利用可能な変数を選択し，以下の 4 つを独立変数とした．

まず 1 つ目は，「将来の進学希望」である．「生徒質問紙」において「あなたは今後どこまで進学したいですか」という項目を中学校から大学院までの 7 件法で回答させた．次に，数学と国語の勉強する理由を尋ねた 12 項目（すべて「かなりあてはまる」から「あてはまらない」まで 5 件法）に関して固有値分解，探索的因子分析（最尤法・バリマックス回転）を行い，固有値の減衰状況 (5.80, 1.71, 0.82...) から 2 因子解を採用した．これらの結果は，速水 (1987) の結果と一致し，その後の学習動機分野の研究の進展を踏まえて，2 因子のうち第 1 因子は「頭を使うことがおもしろいから」等，教科学習ではない知的活動に関する項目が高い因子負荷量を示していたことから「知的的好奇心」と命名した．第 2 因子も同様に，「通知表をよくしたいから」等，主に現実の教科学習に関する項目が高い因子負荷量を示していたため「外発的動機づけ」と命名した (表 9.2.1)．そのうえで，それぞれの項目の得点を合計し，独立変数にした．

表 9.2.1 数学と国語の学習動機の因子パターン行列 (最尤法・バリマックス回転)

質問項目	F1	F2	共通性
F1: 「知的的好奇心」			
分かることが楽しいから	0.841	0.241	0.236
新しいことを知ることができるから	0.808	0.213	0.302
できるようになることがおもしろいから	0.798	0.286	0.282
問題を解くことがおもしろいから	0.737	0.168	0.428
自分の力をためすことができるから	0.673	0.354	0.422
頭を使うことが好きだから	0.667	0.187	0.521
F2: 「外発的動機づけ」			
通知表をよくしたいから	0.116	0.788	0.366
テストでよい点がとりたいから	0.252	0.764	0.352
よい高校や大学に入りたいから	0.173	0.684	0.503
勉強しないとなんとなく不安だから	0.232	0.556	0.637
因子寄与率	32.6	22.9	
累積寄与率	32.6	55.5	

最後に、自由記述式の「学校質問紙」における「施設（校舎レベル）の被災状況・使用可能状況は
 いかがでしたか」に関する回答を分類し、授業に東日本大震災の影響があったと判断できるものを「1」、
 そうでなかったものを「0」とし「震災ダミー」というダミー変数を作成した。ただし、今回のマルチ
 レベル分析の中では、それを「生徒個人が学校における授業においてなんらかの影響を受けた」とい
 うように個人レベルの変数として扱う点に注意されたい。また表 9.2.2 はそれぞれの変数の基本統計量
 を示している。

表 9.2.2 使用変数の基本統計量

変数	度数	最小値	最大値	平均値	標準偏差
PV	11740	-3.68	3.65	-0.14	0.89
将来の進学希望	11740	1.00	7.00	4.34	1.83
知的好奇心	11740	6.00	30.00	18.16	5.75
外発的動機付け	11740	4.00	20.00	15.02	3.62
震災ダミー	11740	0.00	1.00	0.33	0.47

9.4 結果

9.4.1 モデル選択

上記の分析枠組みで示したように、ヌルモデルに加え、以下の 2 つのモデルを構築した。モデル 1
 として、

$$\begin{aligned} \text{レベル 1 (生徒レベル)} \quad \text{数学PV}_{ij} = & \beta_{0j} + \beta_1 \text{将来の進学希望}_{ij} \\ & + \beta_2 \text{知的好奇心}_{ij} + \beta_3 \text{外発的動機づけ}_{ij} + \beta_4 \text{震災ダミー}_{ij} + e_{ij} \end{aligned}$$

$$\text{レベル 2 (学校レベル)} \quad \beta_{0j} = \beta_0 + u_{0j}$$

を設定した。それぞれの説明変数からの効果は学校間で等しいが、切片にのみ学校差を認めるモデル
 である。

次にモデル 2 として、

$$\begin{aligned} \text{レベル 1 (生徒レベル)} \quad \text{数学PV}_{ij} = & \beta_{0j} + \beta_1 \text{将来の進学希望}_{ij} \\ & + \beta_2 \text{知的好奇心}_{ij} + \beta_3 \text{外発的動機づけ}_{ij} + \beta_4 \text{震災ダミー}_{ij} + e_{ij} \end{aligned}$$

$$\text{レベル 2 (学校レベル)} \quad \beta_{0j} = \beta_0 + u_{0j}$$

$$\text{レベル 2 (学校レベル)} \quad \beta_{4j} = \beta_4 + u_{4j}$$

を設定した。切片に関する変量効果に加え、説明変数の中でも「震災ダミー」の効果に関してのみ学校間差を仮定したモデルである⁷。

これら 3 つのモデルの間で、尤度比検定を行った結果が表 9.4.1 である。

表 9.4.1 モデル間における尤度比検定の結果

	AIC	BIC	-2対数尤度	p値
ヌルモデル	29926.49	29948.60	29920.50	-
モデル1	26714.21	26765.80	26700.20	0.00
モデル2	26718.10	26784.43	26700.10	0.95

AIC, BIC, -2 対数尤度のいずれもモデル 1 が最も適合度がよいことがわかる。したがって、今回は切片にのみ変量効果を仮定したモデル 1 を採用した。

9.4.2 マルチレベル分析の結果

表 9.4.2 はヌルモデルとモデル 1 のマルチレベル分析結果である。このときヌルモデルにおける ICC は約 5.3% であり、決して高いとは言えない。しかしながら、その絶対的な基準は示されておらず、この値が高い方がマルチレベル分析を行う意義があると捉えられている (Hox, 2010 ; Luke, 2004) 。

各変数を見ていくと、まず「将来の進学希望」は数学の学力に正の効果を持っていた。また、「知的好奇心」「外発的動機づけ」も微弱ながら数学の学力に対して正の効果を持っていた。さらに今回の分析には、東日本大震災が生徒の学力に対してどのような影響を持っていたのかを探索的に検討するために、自由記述をもとに分類した「震災ダミー」を投入していたが、有意な効果は見られなかった。

表 9.4.2 マルチレベル分析の結果

固定効果	ヌルモデル		モデル1	
	推定値	標準誤差	推定値	標準誤差
切片	-0.17***	0.04	-1.73***	0.05
将来の進学希望			0.19***	0.00
知的好奇心			0.02***	0.00
外発的動機付け			0.02***	0.00
震災ダミー			0.12	0.08
変量効果				
$\sigma^2_{u_{oj}}$	0.04		0.03	
$\sigma^2_{e_{ij}}$	0.74		0.57	

***は 0.1%水準で有意

⁷ その他の変数に対しても変量効果を仮定したモデルの間で尤度比検定を行った結果、「震災ダミー」にのみ変量効果を仮定するモデルが最も適合度がよかったため、これをモデル 2 として設定した。

表 9.4.3 には、モデル 1 を採用した後、モデル式中で切片の変量効果として設定した u_{oj} の推定値を一覧にしたものである。ICC (1) は $\sigma^2_{u_{oj}}$ を使用するが、 u_{oj} の実際の推定値を見ることで、そもそもの学校の数学の学力が高いかを知ることができる。今回分析対象となった 25 校の学校のうち、最も数学の学力が高かったのは $u_{oj}=0.42$ の J10 となる。

表 9.4.3 モデル 1 のもとで推定された u_{oj}

学校番号	変量効果	学校番号	変量効果
J01	-0.06	J13	-0.12
J02	0.07	J14	-0.15
J03	-0.18	J15	-0.01
J04	0.02	J16	-0.18
J05	0.22	J17	0.00
J06	0.13	J18	0.07
J07	-0.12	J19	-0.20
J08	-0.22	J20	-0.18
J09	-0.01	J21	-0.01
J10	0.42	J22	0.00
J11	-0.01	J23	0.12
J12	-0.05	J24	0.11
		J25	0.33

9.5 考察

本章では、学力の規定要因を探索的に探り、さらに今後の全国学力・学習状況調査への応用可能性を検討するため、マルチレベルモデルを用いて分析を行った。この分析にあたり、東日本大震災が学校における授業に対して影響を与えたかどうかを評価するダミー変数など 4 変数によって検討した。まず生徒が抱いている「将来の進学希望」、また心理学的尺度の合計点で構成した「知的好奇心」「外発的動機づけ」のそれぞれの変数はすべて数学の学力について正の効果を持っていた。「震災ダミー」については有意な結果が得られなかった。この東日本大震災の影響に関して、今回は自由記述式の「学校質問紙」に関する回答を筆者が分類したものであり、厳密に言えば震災の影響を正確に捉えたものであるとは言い難い。「将来の進学希望」変数などをコントロールすることにより、「震災ダミー」の効果がそれらに吸収されたとも考えられる。この震災に関する精密な検討は、次の調査研究の課題である。

今回、調査モデルとしてアイテム・マトリックス・サンプリング (item-matrix-sampling) を用い、それに加え推算値を導入することを基礎とした。もしこの条件のもとで、全国規模による実施をすれば、全国の様々な学校が抽出されるため本章で述べた ICC (1) の値は十分高く推定されると予想できる。そして本章で試みたように、マルチレベル分析によって各都道府県をはじめ各市町村レベルに関してより精密な調査を行うことができるようになるだろう。