

博士論文

数学教育開発における大規模教育調査の二次分析に関する研究

—PISA 調査を事例として—

渡邊耕二

広島大学大学院国際協力研究科

2014年3月

数学教育開発における大規模教育調査の二次分析に関する研究  
—PISA 調査を事例として—

D102093

渡邊耕二

広島大学大学院国際協力研究科博士論文

2014年3月

広島大学大学院国際協力研究科

論文名: 数学教育開発における大規模教育調査の二次分析に関する研究  
—PISA 調査を事例として—  
学位の名称: 博士 (学術)  
学生番号: D102093  
氏名: 渡邊耕二

2014年1月24日

審査委員会

委員長・広島大学大学院国際協力研究科 教授

馬場 真也



広島大学大学院国際協力研究科 教授

池田 秀雄



広島大学大学院国際協力研究科 教授

清水 銀也



広島大学大学院教育学研究科 教授

志崎 和樹



広島大学大学院教育学研究科 教授

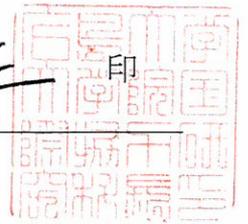
山口 博敏



2014年2月28日

研究科長

藤原 章正



## 謝辞

本研究を遂行し、学位論文をまとめるに当たり、多くのご支援とご指導を賜りました、指導教員である馬場卓也教授に深く感謝を申し上げます。時に応じて、厳しくご指導いただいたこと、励ましてくださったことを通じて、私自身の至らなさを実感することが多くありました。馬場先生には、博士課程前期・後期の5年半もの長い間、私の研究を温かく見守って下さいました。この場を借りて厚く御礼申し上げます。

国際協力研究科の池田秀雄教授には、IDECで蓄積された知見を含めた研究をさらに高めるための方向性や理科教育との関わりなど、国際理数科教育協力に関するご指導を多く頂きました。この場を借りまして、池田先生に心より感謝申し上げます。

国際協力研究科の清水欽也教授には、項目反応理論など二次分析に必要な知識を教えてくださいました。また何気ない会話の中でも、私にとっては先生のお話は刺激的なものでした。清水先生に、心より感謝申し上げます。

教育学研究科の岩崎秀樹教授には、博士課程前期の頃から、全国数学教育学会をはじめ多くの場面でご指導を頂きました。数学教育研究の発展に貢献できるように、日々を積み重ねたい覚悟でございます。岩崎先生のご指導とご厚情に厚く御礼申し上げます。

教育学研究科の山崎博敏教授には、教育社会学研究に関する知識や考え方だけでなく、計量的な分析による研究のABCを惜しみなく教えて頂きました。先生から頂いたご指導を心に深く刻み、今後の研究活動に精進したい所存でございます。山崎先生のご指導とご支援に心より感謝申し上げます。

大学院での研究生生活を共にしてきた馬場研究室の皆さまには、日頃から研究の話を聞いて頂きました。本論文執筆に向けて、精神的な支えになっていたことは言うまでもありません。ともに研究に励んだ院生の皆さまにも、この場を借りて厚く御礼申し上げます。

本論文は、本当に多くの方々に支えられて執筆に至りました。本当にありがとうございました。最後に、私のわがままな進路を見守り、今日まで応援してくれた両親（父・克修、母・君枝）に感謝致します。

2014年2月 渡邊耕二

## 目次

<b>第1章 問題の所在と研究の目的およびその方法</b>	<b>1</b>
1.1. 問題の所在	1
1.1.1. Mathematics for All からみた今日的な数学教育研究の課題	1
1.1.2. 数学教育研究における大規模教育調査の活用に向けた課題	6
1.2. 本研究の目的およびその方法	15
1.2.1. 本研究の目的	15
1.2.2. 本研究における分析の観点と主要な分析手法	15
1.2.3. 使用データとデータ処理および使用する項目反応モデル	21
1.3. 本研究の構成	25
<b>第2章 数学学力とカリキュラムの特徴の関連性</b>	<b>30</b>
2.1. PISA2003 における数学的リテラシー調査のテスト得点の国際比較	30
2.2. 日本に注目した 13 先進国に関する分析	33
2.2.1. 確率・統計を活用する能力の分布の国際比較	34
2.2.2. 項目母数の推定と等化	34
2.2.3. 13 カ国の解答パターンの類似性と特異性	37
2.2.4. 日本の解答パターンの特徴	38
2.2.5. 「不確実性」における日本と平均国との差異の分布	40
2.2.6. 日本における数学学力とカリキュラムの関連性の総合的考察	41
2.2.7. 日本以外の 12 カ国の平均国との差異	47
2.3. 7 途上国に関する分析	50
2.3.1. 項目母数の推定と等化	51
2.3.2. 7 カ国の解答パターンの類似性と特異性	51
2.3.3. 7 カ国の解答パターンの特徴	54
2.4. 本章のまとめ	60
<b>第3章 数学学力と数学に対する情意的側面の関連性</b>	<b>63</b>
3.1. 数学的リテラシー調査のテスト得点と 5 つの情意的側面に関する因子得点の国際比較	63
3.1.1. 使用データとそのデータ処理	63
3.1.2. 数学的リテラシー調査のテスト得点の国際比較	63
3.1.3. 5 つの情意的側面における因子得点の国際比較	66
3.1.4. 階層線形モデルによるテスト得点と 5 つの情意的側面の関連性の国際比較	66
3.2. 13 先進国における「自己効力感」の水準別に注目した分析	77
3.2.1. 異なる自己効力感の水準におけるテスト得点の分布と困難度に注目した解答パターン	78
3.2.2. 日本に注目した分析結果の総合的考察	84
3.3. 7 途上国における「自己効力感」の水準別に注目した分析	86

3.3.1. 異なる自己効力感の水準におけるテスト得点の分布と困難度 に注目した解答パターン	86
3.3.2. 途上国に注目した分析結果の総合的考察	91
3.4. 本章のまとめ	93
<b>第4章 数学学力と言語的側面の関連性</b>	<b>98</b>
4.1. 数学的リテラシーと読解力調査のテスト得点の国際比較	98
4.1.1. 使用データとそのデータ処理	98
4.1.2. 数学的リテラシーと読解力調査のテスト得点とその経年変化 の国際比較	99
4.1.3. 階層線形モデルによる数学的リテラシーと読解力調査のテスト得点 およびテスト言語の関連性の国際比較	103
4.2. 13先進国における読解力の水準別に注目した分析	114
4.2.1. 異なる読解力の水準におけるテスト得点の分布と困難度 に注目した解答パターン	115
4.2.2. 先進国に分析結果の総合的考察	121
4.3. 7途上国における読解力の水準別に注目した分析	122
4.3.1. 異なる読解力の水準におけるテスト得点の分布と困難度 に注目した解答パターン	123
4.3.2. 途上国に注目した分析結果の総合的考察	128
4.4. 本章のまとめ	129
<b>第5章 本研究の総合的考察と今後の課題</b>	<b>134</b>
5.1. 各章で得られた研究成果	134
5.1.1. 第2章の概要	134
5.1.2. 第3章の概要	135
5.1.3. 第4章の概要	136
5.2. 本研究の総合的考察と今後の課題	138
5.2.1. 本研究の総合的考察	138
5.2.2. 今後の課題	141
<b>参考・引用文献</b>	<b>144</b>

## 図・表一覧

### 【第1章 問題の所在と研究の目的およびその方法】

- 表 1.1 先進国と途上国における MFA の枠組み
- 図 1.1 SES と数学的リテラシー得点の散布図
- 図 1.2 Mathematics for All における分析の観点
- 表 1.2 TIMSS と PISA の参加国数の推移
- 表 1.3 SACMEQ、PASEC および LLECE の参加国
- 表 1.4 TIMSS および PISA における得点の上位・下位 10 カ国
- 表 1.5 PISA における経年変化（「数学的リテラシー」変化率の降順）
- 表 1.6 PISA2003 における 13 種類のブックレット（テスト）の構成
- 図 1.3 2 値反応データの例（シミュレーション）
- 図 1.4 識別力の高い項目と低い項目の例（シミュレーション）
- 図 1.5 2 つの集団における正答率と識別力（シミュレーション）
- 図 1.6 2 つの集団における SES とテスト得点の散布図（シミュレーション）
- 図 1.7 データ処理後のデータセットの一例
- 図 1.8 本研究の構成

### 【第2章 数学学力とカリキュラムの特徴の関連性】

- 図 2.1 2 母数モデルの項目母数
- 図 2.2 ラッシュモデルの項目母数
- 図 2.3 2 母数モデルとラッシュモデルによるテスト得点の散布図
- 表 2.1 2 母数モデルによる各国のテスト得点（平均値の降順）
- 表 2.2 包括的アイディア（数学的な内容）の各領域の内容と項目数
- 表 2.3 13 カ国における数学および確率・統計を活用する能力の分布
- 図 2.4 39 カ国の数学および確率・統計を活用する能力の平均値の散布図
- 表 2.4 等化後の項目母数の一部
- 図 2.5 困難度のテンドログラム
- 図 2.6 識別力のテンドログラム
- 表 2.5 平均国との差の標準偏差
- 図 2.7 日本の困難度の分布
- 図 2.8 平均国の困難度の分布
- 図 2.9 日本の識別力の分布
- 図 2.10 平均国の識別力の分布
- 図 2.11 日本の困難度差と識別力差の散布図（80 項目）
- 図 2.12 日本の困難度差と識別量差の散布図（不確実性の 19 項目）
- 表 2.6 「不確実性」における公開項目の困難度差と識別力差
- 図 2.13 「輸出 1」
- 図 2.14 「輸出 2」
- 図 2.15 「いろいろな色のキャンディ」
- 図 2.16 「地震」

図 2.17	「テストの点数」
図 2.18	「大統領の支持率」
図 2.19	オーストラリア
図 2.20	カナダ
図 2.21	ドイツ
図 2.22	フィンランド
図 2.23	フランス
図 2.24	香港
図 2.25	アイルランド
図 2.26	イタリア
図 2.27	韓国
図 2.28	オランダ
図 2.29	ニュージーランド
図 2.30	アメリカ
表 2.7	4つの領域における平均国との距離の平均値（不確実性の降順）
図 2.31	4つの領域における平均国との距離の平均値の折れ線グラフ（不確実性の降順）
表 2.8	等化後の項目母数の一部
図 2.32	困難度のテンドログラム
図 2.33	識別力のテンドログラム
表 2.9	7ヵ国における困難度と識別力の相関係数
表 2.10	平均国との差の標準偏差
図 2.34	平均国における困難度と識別力の散布図
図 2.35	困難度における箱ひげ図
図 2.36	識別力における箱ひげ図
図 2.37	ブラジル
図 2.38	インドネシア
図 2.39	メキシコ
図 2.40	タイ
図 2.41	チュニジア
図 2.42	トルコ
図 2.43	ウルグアイ
表 2.11	4つの領域における平均国との距離の平均値（不確実性の降順）
図 2.44	4つの領域における平均国との距離の平均値の折れ線グラフ（不確実性の降順）
図 2.45	ブラジル
図 2.46	インドネシア
図 2.47	メキシコ
図 2.48	タイ
図 2.49	チュニジア
図 2.50	トルコ
図 2.51	ウルグアイ

### 【第3章 数学学力と数学に対する情意的側面の関連性】

- 図 3.1 84 項目における項目特性曲線
- 表 3.1 ラッシュモデルによる各国のテスト得点（平均値の降順）
- 表 3.2 5つの情意的側面の内容とその因子寄与率、 $\alpha$ 信頼係数および $\omega$ 信頼係数
- 表 3.3 5つの情意的側面における因子得点（興味・関心の平均値の降順）
- 図 3.2 39カ国の5つの情意的側面の平均値の散布図（横軸：興味・関心）
- 表 3.4 国レベルの記述統計量（39カ国）
- 表 3.5 生徒レベルの記述統計量（228032名）
- 表 3.6 階層線形モデルにおける固定効果の分析結果
- 表 3.7 モデル1とモデル4における $u_{0j}$ の大きさ（モデル4の降順）
- 図 3.3 モデル1とモデル4における各国の $u_{0j}$ の大きさ（モデル4の降順）
- 図 3.4 39カ国における $u_{0j}$ の相対的な変化の大きさ
- 表 3.8 モデル4における39カ国の切片と回帰係数（興味・関心の降順）
- 図 3.5 5つの情意的側面におけるランダム効果を含む回帰係数の箱ひげ図
- 図 3.6 階層線形モデルと他の分析方法による結果の散布図
- 表 3.9 4つの分析方法における相関係数
- 表 3.10 13カ国の3つの水準におけるテスト得点の平均値と標準偏差（平均値の降順）
- 図 3.7 39受験者集団における困難度の相関係数の分布
- 表 3.11 日本における自己効力感の水準別の困難度の平均値
- 図 3.8 日本の出題形式における困難度の水準別比較
- 表 3.12 13カ国における水準別比較の国際比較（一の位の0は省略）
- 表 3.13 13カ国における全体平均値の国際比較（一の位の0は省略）
- 図 3.9 「自由記述」、「教育的」、「科学的」、「空間と形」および「複合的選択肢」における困難度の国内水準別比較
- 表 3.14 7カ国の3つの水準におけるテスト得点の平均値と標準偏差（平均値の降順）
- 図 3.10 21受験者集団における相関係数の分布
- 表 3.15 ブラジルにおける自己効力感の水準別の困難度の平均値
- 図 3.11 ブラジルの出題形式における困難度の水準別平均値
- 表 3.16 7カ国における水準別比較の国際比較（一の位の0は省略）
- 表 3.17 7カ国における全体平均値の国際比較（一の位の0は省略）
- 図 3.12 「自由記述」、「選択肢」、「教育的」、「職業的」および「空間と形」における困難度の国内水準別比較
- 図 3.13 質問紙調査における39カ国の多次元性

### 【第4章 数学学力と言語的側面の関連性】

- 表 4.1 3つのPISAにおける項目母数の基本統計量
- 表 4.2 数学的リテラシーにおける3つのPISA間の共通項目と等化係数
- 表 4.3 読解力における3つのPISA間の共通項目と等化係数
- 表 4.4 数学的リテラシーにおけるテスト得点とその経年変化（PISA2003の平均値の降順）
- 表 4.5 読解力におけるテスト得点とその経年変化（PISA2003の平均値の降順）
- 図 4.1 PISA2003における数学的リテラシーと読解力の平均値の散布図

- 図 4.2 PISA2003 と PISA2009 における数学的リテラシーの平均値の散布図
- 図 4.3 PISA2003 と PISA2009 の数学的リテラシーと読解力における各国の平均値の差の効果量  
(数学的リテラシーの効果量の降順)
- 表 4.6 3つの PISA における国レベルと生徒レベルの記述統計量
- 表 4.7 各 PISA における階層線形モデルによる固定効果の分析結果
- 表 4.8 PISA における読解力とテスト言語の回帰係数 (PISA2003 の読解力の降順)
- 図 4.4 読解力の回帰係数の散布図
- 図 4.5 テスト言語の回帰係数の散布図
- 図 4.6 PISA2003 における数学的リテラシーの平均値と読解力の回帰係数の散布図
- 図 4.7 階層線形モデルと多母集団回帰分析および相関係数の散布図
- 表 4.9 13カ国の3つの水準におけるテスト得点の平均値と標準偏差 (平均値の降順)
- 図 4.8 39受験者集団における困難度の相関係数の分布
- 表 4.10 日本における読解力の水準別の困難度の平均値
- 図 4.9 日本の出題形式における困難度の水準別比較
- 表 4.11 13カ国における水準別比較の国際比較 (一の位の0は省略)
- 表 4.12 13カ国における全体平均値の国際比較 (一の位の0は省略)
- 図 4.10 「科学的」、「教育的」、「自由記述」、「空間と形」における困難度の国内水準別比較
- 表 4.13 7カ国の3つの水準におけるテスト得点の平均値と標準偏差 (平均値の降順)
- 図 4.11 39受験者集団における困難度の相関係数の分布
- 表 4.14 ブラジルにおける読解力の水準別の困難度の平均値
- 図 4.12 ブラジルの「科学的」、「自由記述」、「複合的選択肢」、「選択肢」の水準別平均値
- 表 4.15 7カ国における水準別比較の国際比較 (一の位の0は省略)
- 表 4.16 7カ国における全体平均値の国際比較 (一の位の0は省略)
- 図 4.13 「科学的」、「教育的」、「自由記述」、「複合的選択肢」および「選択肢」における困難度の国内水準別比較

## 【第5章 本研究の総括と今後の課題】

- 図 5.1 本研究の総合的考察の略図

## 【付録】

- 付録 1.1 5つの情意的側面に関する質問項目
- 付録 1.2 PISA2003 におけるデータ処理後の数学的リテラシーの受験者数および数学的リテラシーと読解力の共通受験者数
- 付録 3.1 5つの情意的側面における出題項目と因子負荷量
- 付録 3.2 PISA2003 の数学的リテラシー調査における領域・種別
- 付録 4.1 PISA2006 における数学的リテラシーの平均値と回帰係数の散布図
- 付録 4.2 PISA2009 における数学的リテラシーの平均値と回帰係数の散布図
- 付録 4.3 複合的選択肢の項目例 (項目名「花壇」)

## 略語一覧

OECD	経済協力開発機構 (Organization for Economic Cooperation and Development)		
PISA	OECD生徒の学習到達度調査 (Programme for International Student Assessment)		
IEA	国際教育到達度評価学会 (International Association for the Evaluation of Educational Achievement)		
TIMSS	国際数学・理科教育動向調査 (Trend in International Mathematics and Science Study)		
SACMEQ	教育の質測定のための東南部アフリカ連合 (Southern and Eastern Consortium for Monitoring Educational Quality)		
OREALC	ユネスコのラテンアメリカ・カリブ海地域事務所 (Oficina Regional de Educación para América Latina y el Caribe)		
LLECE	ラテンアメリカにおける教育の質評価研究 (Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación)		
PASEC	フランス語圏アフリカ諸国における学校教育システム分析プログラム (Programme d'Analyse des Systèmes Éducatifs des Pays de la Confemen)		
NAEP	全米学力調査 (the National Assessment of Education Progress)		
AUS	オーストラリア	KOR	韓国
AUT	オーストリア	LUX	ルクセンブルク
BEL	ベルギー	LVA	ラトビア
BRA	ブラジル	MAC	マカオ
CAN	カナダ	MEX	メキシコ
CHE	スイス	NLD	オランダ
CZE	チェコ	NOR	ノルウェー
DEU	ドイツ	NZL	ニュージーランド
DNK	デンマーク	POL	ポーランド
ESP	スペイン	PRT	ポルトガル
FIN	フィンランド	RUS	ロシア
FRA	フランス	SVK	スロバキア
GRC	ギリシャ	SWE	スウェーデン
HKG	香港	THA	タイ
HUN	ハンガリー	TUN	チュニジア
IDN	インドネシア	TUR	トルコ
IRL	アイルランド	URY	ウルグアイ
ISL	アイスランド	USA	アメリカ
ITA	イタリア	YUG	旧セルビア・モンテネグロ
JPN	日本		

## 第1章 問題の所在と研究の目的およびその方法

### 1.1. 問題の所在

#### 1.1.1. Mathematics for All からみた今日的な数学教育研究の課題

教育は、基本的人権の一つであると同時に社会経済開発においても重要な役割りを果たしている（黒田、p.1）。だからこそ「Education for All（万人のための教育、以下EFA）」が謳われるように、教育の普及とその質の向上は、国際社会や各国が果たすべき責務となっている。今日においては、開発途上国（以下、途上国）を中心に教育の量的拡大は一定程度の成果がみられるが、子どもの学習成果が十分には向上していないとされる（UNESCO、2008、p.108）。つまり教育の質を子どもの学力<sup>1</sup>から捉えるならば、依然多くの課題が残されている。

現代社会において、子どもの社会・経済・文化的な背景などの社会的出自が、生徒の学力に何らかの影響を及ぼすことが、日本だけに限らず、実証された社会学的な事実として認識されている（荻谷・志水、2004）。つまり国、ジェンダー、民族、社会経済的地位（socio-economic status：以下、SES）あるいは学校など、生徒が属する集団の違いによって、生徒の学習成果に差異が生じているのである。集団間に存在する学習成果の格差、言い換えると格差問題の縮小あるいは解消、そして教育の社会的公正（equity）の実現は、今日の教育が抱える主要な課題の一つである。格差問題に対しては、これまで教育社会学研究が大きな役割りを担ってきたし、現在でもそうである。欧米を中心にした研究をみると、人種や民族の違いあるいは移民に焦点を置き、SESの階層差とともに格差問題が論じられてきた（e.g.、Grace & Jennifer、2003；Jennifer & Bryndl、2007；Kimberly & Yu、1999；Wiggan、2007）。またわが国においては、SESの階層差や学校間にみられる格差問題に焦点を置きながら「学習方略」あるいは「効果的な学校」などをキーワードに国際比較にまで対象を広げて考察されている（e.g.、金子、2004；川口、2009；近藤、2012；須藤、2007）。

文化的あるいは言語的な側面に依存しないとされる数学を扱う数学教育においても、上述のような集団間にみられる格差問題が議論されてきた。その契機は、1984年に開催された第5回数学教育国際会議（International Congress on Mathematics Education：以下、ICME5）における課題研究テーマの一つであった「Mathematics for All（万人のための数学、以下MFA）」であろう。そこでは、数学教育研究で目が向けられつつある途上国、少数民族、ジェンダー、地域、経済状況などを集中的に取り上げ、社会的・文化的な側面から数学教育を捉え直すという新たな観点が提起された（Damerow et al.、1985）。その主な

---

<sup>1</sup> 「学力」という言葉は、定義が難解な用語である（荻谷・志水、2004；市川、2001）といわれるように、様々に議論されている。本章では、学習意欲やコミュニケーション力を含めた広い意味で「学力」という用語を用いるが、第2章以降では、学力の内的規定には注目せず、その他の側面との外的な関係を取り上げるため、テストによって浮かび上がる測定可能な学習成果として「学力」を規定する。

背景には、数学教育の大衆化および数学教育現代化<sup>2</sup>の流れを汲むカリキュラムが数学者を目指すエリートの生徒を主要な対象とし、大多数の生徒への対応が軽視されていたことを挙げられる。ICME5 で提起された MFA は、それまで学校教育を受けることが困難であった生徒や不適当な内容や方法で数学を学んでいた生徒に目を向け、全ての生徒にとって価値ある数学教育を模索する必要性を唱えたのである。その後の ICME においても MFA に関する議論が継続的に取り上げられており（モーモーニェン、2006）、2012 年に韓国で開催された ICME12 では、ジェンダーや教授言語の観点から特に論じられている（e.g., Leder、2012 ; Mamokgethi、2012）。

その一方で Gates & Vistro-Yu (2003) は、表 1.1 に示したように MFA 実現に向けた課題をカリキュラムの内容、数学の教授・学習における文化的な価値、試験などの評価実践および社会的公正という視点から捉え直している。そこでは先進国・途上国に共通する課題として、格差問題の緩和と数学の教授・学習に影響を及ぼす試験や調査の存在が取り上げられている。先進国と途上国を区別する枠組みでは、カリキュラムと文化的価値について触れられている。カリキュラムに関しては、欧米諸国を中心としたエリートの育成を目指すカリキュラムが途上国に広まったことを課題の一つに捉えている。そのため多くの生徒は、数学を学ぶことに不安あるいは拒否感を抱いている（p.34）とも述べ、数学に対する情意的な側面への働きかけを必要としている。また文化的側面については、先進国や途上国における少数民族などの各集団の固有性を重視し民族数学<sup>3</sup>を取り上げている（p.38）。民族数学とは、MFA と同じ ICME5 でブラジル人数学教育研究者 D'Ambrosio によって導入され、第三世界における欧米で発展した数学を基盤とする数学教育に対して批判的な立場から論じる契機を与えた。これを背景に持つ馬場（1998、1999、2001、2002）の一連の研究では、批判的数学教育との相互参照を通して民族数学の数学教育への適用を考察し、民族数学を数学的活動として捉えた動詞型カリキュラムの提案としてカリキュラム構成原理について検討している。これは、途上国の数学教育におけるカリキュラムの基盤を文化的な側面から捉え直し、再構築しようという試みである。また馬場（2005）は、途上国の数学教育全体で重要な役割りを果たす教授言語をはじめとする言語的側面に関する研究の必

表 1.1 先進国と途上国における MFA の枠組み

	カリキュラム	文化的価値	評価実践	社会的公正
先進国	数学のエリート育成のための内容	欧米を中心とした文化的背景	数学の教授・学習に影響を与える試験や調査の存在	ジェンダー、民族間や先進国と途上国などの集団間における格差問題
途上国	先進国のカリキュラムの再生産	多様な文化的背景の軽視		

出典：Gates & Vistro-Yu (2003) をもとに筆者作成

<sup>2</sup> 1957 年に旧ソ連の人工衛星スプートニク 1 号の打ち上げ成功を発端に、世界的に広まったカリキュラム改革運動を指す。

<sup>3</sup> 民族数学とは、文化、地域、民族、職業、年齢などの共通した目的や習慣を持つ集団において実践される数学のことを意味する（D'Ambrosio、2006）。

要性を主張している。そして Gates & Vistro-Yu (2003, p.68) は、MFA が対象とする範囲は、非常に広く扱い切れない部分が生じるとしながらも、その研究を継続することの重要性を述べている。

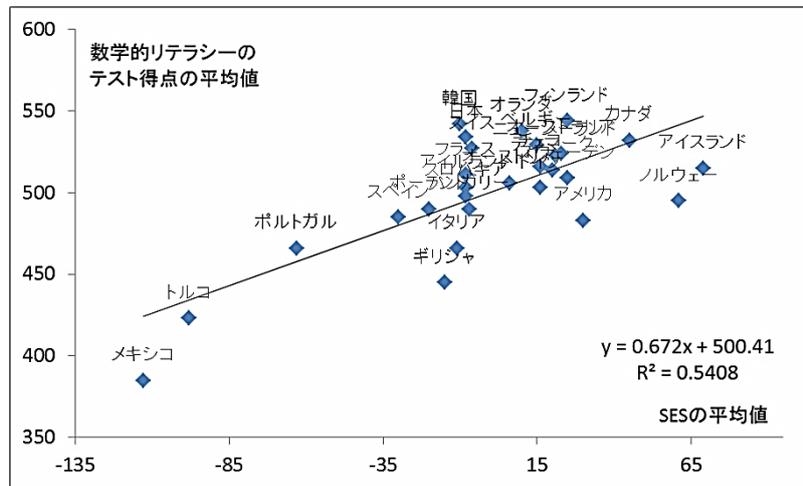
このように ICME5 で提唱された MFA は、数学教育の大衆化および数学教育現代化の際に軽視された社会文化的側面を再考することで、全ての生徒にとって価値ある数学教育の実現という理念を有している。今日では、これらの 4 つの視点に留まらず、それらから派生する情意的側面や教授言語といった観点に具体化されながら、更なる議論が求められている。そこで本研究では、数学教育研究における教育調査の活用を考察していく上で、社会的・文化的な側面から数学教育を捉え直すという ICME5 の MFA の理念を拠り所に、上述のいくつかの観点から MFA 実現に向けた議論を進めることとする。

MFA 実現に向けて、現状を計量的に把握できる研究成果に目を向けると、経済協力開発機構 (Organization for Economic Cooperation and Development : 以下、OECD) が行う生徒の学習到達度調査 (Programme for International Student Assessment : 以下、PISA) や国際教育到達度評価学会 (International Association for the Evaluation of Educational Achievement : 以下、IEA) の国際数学・理科教育動向調査 (Trend in International Mathematics and Science Study : 以下、TIMSS) といった教育に関する調査 (以下、教育調査) では、生徒の学力に影響を及ぼす要因を国際比較を通じて調べている (e.g., OECD, 2004, 2007 ; IEA, 2005, 2008a, 2008b)。例えば OECD (2004) では、PISA が規定する数学を活用する能力である「数学的リテラシー<sup>4</sup>」に対して、学習への態度、ジェンダー、SES および学校の違いなどの影響を国際比較によって検討している。図 1.1 は、OECD 加盟国における PISA で計算された SES と数学的リテラシー得点の平均値の散布図である。この図から分かるように SES と生徒の学力には、一定の相関関係が認められる。このような分析結果は、自国の特徴を国際的な視点から捉えるのに役立つ。

アメリカにおいては、コールマン・レポート (Coleman, 1966) を皮切りにスパニック系、アフリカ系、原住民系アメリカ人などの民族やジェンダーあるいは学校の違いにおいて、生徒の学力格差が早くから問題視され、1980 年代から数多くの調査が行われてきた。それらの調査結果を受け、1989 年に全米数学教師協会 (National Council of Teachers of Mathematics : 以下、NCTM) は、「学校数学のためのカリキュラムとその評価基準 (Curriculum and Evaluation Standards for School Mathematics)」を公表している (NCTM, 1989)。その後も 1995 年と 2000 年に同様な評価基準が作成され、社会的公正の実現を数学教育の目標として位置付けている。そのため、数学教育における社会的公正に関する実証的な研究が多くみられる (e.g., Jemes et al., 2008 ; Lucille, 1997 ; Michael et al., 2007 ; Thomas et al., 2008 ; William, 1997 ; Xin & Jesse, 2007)。これらの数

<sup>4</sup> PISA の「数学的リテラシー」は、「数学が現実で果たす役割を見つけ、理解し、現在及び将来の個人の生活、職業生活、友人や家族や親族の社会生活、建設的で関心を持った思慮深い市民としての生活において確実な根拠に基づき判断を行い、数学に携わる能力」と規定されている (国立教育政策研究所、2004)。

図 1.1 SES と数学的リテラシー得点の散布図



出典：OECD（2004、p.358）をもとに筆者作成

学教育研究では、テストで測定した生徒の学力に対して、SES、学校、ジェンダー、民族、言語能力（language proficiency）の影響あるいは実施された数学教育改善プログラムの効果といった観点で計量的な検証が行われている。例えば William（1997）では、民族間にみられる学力格差は縮小傾向にあるが、ジェンダーにおいてはほとんど変化がないと報告している。また Thomas et al.（2008）は、新しく導入したあるプログラムの効果がオープンエンドな問題や問題解決（problem solving）の学習において確認され、生徒の学習意欲に対しても肯定的な影響を及ぼしたとしている。

その一方でわが国では2007年から、義務教育の機会均等とその水準の維持向上の観点から、日本国内の児童・生徒の学力と学習状況を把握する取り組みが進められている（国立教育政策研究所、2012a、2012b）が、アメリカをはじめとする国際的な動向と比べると、MFA 実現に繋がる計量的な数学教育研究は、国立教育政策研究所が公刊する報告書を読み解く方法に留まる傾向にあり、これまで十分に議論されてきたとは言い難い。それは、先進国の中で類をみないほど経済格差が少なかったことと単一民族・言語であり宗教による差別もほとんど存在しないという社会文化的な特徴が関係していると思われる。ところが清水（2012）は、今日的なわが国の数学教育研究について、

「今後は教育社会的な関心に基づく学力の外的要因（例えば、保護者の収入の違い）が学力に及ぼす影響や、教育測定（テスト）理論に基づく標準化された調査結果分析の報告など、教室における数学科の学習指導とはやや離れたところの議論も進められることになるだろう。・・・(省略)・・・教科の立場からは、学習指導要領に指導内容として示された数学的活動について、具体的な問題場面やそこではたらく数学的に考える力に特に焦点化して指導と評価の研究を進める必要がある。また、併せて子ども達の学力や

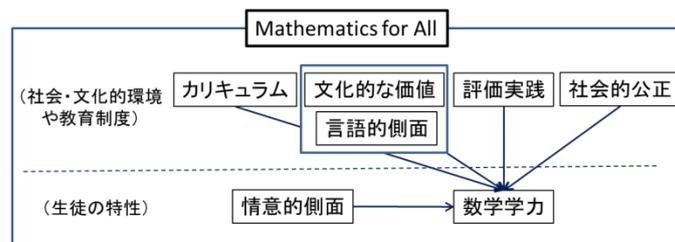
学習状況の実態と学習指導の改善の方向性を探る努力を、一層積み重ねる必要がある。

》

と述べている。他方でこの指摘に関連して近年では、説明責任を背景に学力と効果的な学校に関する研究が、教育社会学研究においていくつか報告されるようになってきた (e.g., 荻谷・志水、2004 ; 川口、2009 ; 須藤、2010 ; 古田、2012 ; 山崎・藤井・水野、2009)。例えば荻谷・志水 (2004) は、生徒の学力に対する SES の階層差が及ぼす影響は拡大傾向にあると指摘している。その結果を受け須藤 (2010) は、学習方略と学力および階層差の関連性について PISA2003 の質問紙と数学的リテラシー調査のデータを活用し、階層差は学習方略にも影響を及ぼすと報告した。これを翻ると、学習方略への働きかけが学力格差の縮小に効果を生むことを主張できる。また川口 (2009) は、ある地方自治体が実施した教育調査のデータを分析し、わが国は学校間の学力格差は小さいが、生徒の学力を大きく伸ばす学校が存在することを示した。これは、学校教育が階層差の影響を軽減させる可能性を示唆している。これらの教育社会学研究は、わが国における格差問題の実態把握とその解消・縮小に向けて、数学教育に関連するデータを用いた取り組みである。これを考慮するならば、MFA 実現に向けた計量的な数学教育研究は、今後増々充実すべき研究領域である。

そこで Gates & Vistro-Yu (2003) や馬場 (2005) によって提示された観点、具体的には、数学教育におけるカリキュラム、文化的な価値、評価実践、社会的公正、情意的側面および言語的側面を手掛かりに、数学教育研究が担うべき研究課題を整理する。ここで言語的側面を文化的な価値の一面として位置付ける (馬場、2005) ならば、図 1.2 のように生徒の学力育成に繋がる 5 つの観点を設定できる。これらの観点について述べれば、これまでは MFA 実現に向けて社会的公正に軸足を置く研究が大半であったといえる。しかし MFA 実現を単なるスローガンに終止させないためには、他の 4 つの観点からも考察が深められるべきであろう。言うまでもなくカリキュラムおよび数学に対する情意的側面への考察は、これまで数学教育研究が主として取り組んできた研究課題である (e.g., 今井、2010 ; 根本、2010)。教授言語の習得度合いなどに代表され得る言語的側面は、途上国を対象とし

図 1.2 Mathematics for All における分析の観点



出典 : Gates & Vistro-Yu (2003) と馬場 (2005) をもとに筆者作成

た数学教育研究 (e.g., 内田、2011 ; 澁谷、2008、2010 ; 中和、2012 ; 馬場、2008) や MFA の議論 (Mamokgethi、2012) で頻繁に取り上げられている。それは日常生活でほとんど使用しない言語、言い換えると第二言語による数学教育においては、不十分な第二言語能力が阻害要因となり、数学の学習が望ましい形で進まないことを容易に想像できる。この観点は、日本では決して馴染み深いものではないが、国際的な視座に立てば、数学教育研究の主要な課題に位置付いている (e.g., Barton & Barton、2005 ; Barwell et al.、2007 ; Jamal & Carol、2001 ; Máire & John、2009 ; Mamokgethi、2005 ; Mamokgethi & Jill、2001)。最後に評価実践 (assessment practice) では、数学の教授・学習が選抜試験や教育調査で高い成績を残すだけに行われてしまうことを案じている (Gates & Vistro-Yu、2003、p.45)。また選抜試験に使用されるテストの妥当性や信頼性がしばしば問題になる。教育調査においては、単純にテスト得点のみに目が向けられ、得られる情報が限定的になってしまうことなども問題視される。しかし選抜試験や教育調査は、学校教育で育成された能力を表現あるいは評価するためのいわば道具または手段であり、それらが生徒の学力を育成するわけではない。したがって学力の育成とそれを評価する活動は、互いに関連することを認めながらも、それらは区別して取り扱うべきと考える。

以上を纏めると、MFA 実現に向けた数学教育研究の課題が明らかになる。それは数学教育における社会的公正だけでなく、①カリキュラム、②情意的側面、③言語的側面の 3 つの視点に立つ研究を蓄積していくことである。そして「生徒の学力」を軸に社会的公正を教育調査のデータを通じて論じる教育社会学研究を考慮するならば、上記の 3 つの視点と「生徒の学力」を教育調査を活用して実証的に関連付ける試みが必要と考える。そこで MFA 実現に向けて、提起した 3 つの視点と「生徒の学力」の関連性を計量的に明らかにする取り組みを今日的な数学教育研究の課題として位置付けたい。なお学力に影響を与える要因としての選抜試験などの評価実践は、特に取り上げないこととした。それは本研究自体が生徒の学力を評価する活動に含まれるためである。

### 1.1.2. 数学教育研究における大規模教育調査の活用に向けた課題

理数科科目を含む調査を中心にみると、TIMSS や PISA が全世界的に行われる大規模な教育調査として有名である。TIMSS は、1995 年から 4 年毎に基礎学校第 4 学年 (日本の小学校第 4 学年相当) および第 8 学年 (日本の中学校第 2 学年相当) を対象に実施されており、また PISA は 2000 年から 3 年毎に 15 歳児 (日本の高等学校第 1 学年相当) を対象に行われている。表 1.2 から分かるように TIMSS や PISA に参加する国は、回を追う毎に増加する傾向にあり、国際的な指標を用いて自国の教育を把握する取り組みが普及しつつある。

また全世界的ではなく、地域を限定した教育調査も存在する。例えば、東南部アフリカ地域を対象に行われているジンバブエ教育省と UNESCO が企画した、教育の質測定のための東南部アフリカ連合の調査 (Southern and Eastern Consortium for Monitoring Edu

表 1.2 TIMSS と PISA の参加国数の推移

TIMSS		PISA	
年	国・地域数	年	国・地域数
1995	45	2000	43
1999	38	2003	41
2003	54(4)	2006	57
2007	67(8)	2009	74(3)
2011	78(14)	2012	64

注) ( )内は、州レベルでの参加数

出典：TIMSS および PISA のホームページより筆者作成

ational Quality：以下、SACMEQ) や、フランス語圏アフリカ諸国における学校教育システム分析プログラム (Programme d'Analyse des Systèmes Éducatifs des Pays de la Confemen：以下、PASEC) およびユネスコのラテンアメリカ・カリブ海地域事務所 (Oficina Regional de Educación para América Latina y el Caribe：以下、OREALC) によるラテンアメリカにおける教育の質評価研究 (Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación：以下、LLECE) などが知られている (表 1.3 参照)。これらの調査結果は、国際比較を通じて自国の教育を特徴付けるのに役立つ。

加えて教育調査は国際的なものだけでなく、国内においても実施されている。アメリカ合衆国 (以下、アメリカ) では、全米学力調査 (the National Assessment of Education Progress：以下、NAEP) と呼ばれる全米規模で行われる調査がある。その目的は、特定

表 1.3 SACMEQ、PASEC および LLECE の参加国

SACMEQ (2007)	PASEC (2006)	LLECE (2006)
ボツワナ	ブルキナファソ	アルゼンチン
レソト	カメルーン	ブラジル
ケニア	コンゴ	チリ
マラウイ	マダガスカル	コロンビア
モーリシャス	セネガル	コスタリカ
モザンビーク	チャド	キューバ
ナミビア	ベナン	エクアドル
セイシェル	ガボン	エルサルバドル
南アフリカ	モーリタニア	グアテマラ
スワジランド		メキシコ
タンザニア*		ニカラグア
ウガンダ		パナマ
ザンビア		パラグアイ
ザンジバル**		ペルー
ジンバブエ		ドミニカ共和国
		ウルグアイ
		レオン***

注) \*は、タンザニア本土を意味する。  
 \*\*は、タンザニア・バンジバル島である。  
 \*\*\*は、メキシコにあるヌエボ・レオン州である。  
 ( )内は、調査が実施された年。

出典：SACMEQ と PASEC のホームページおよび LLECE (2009) より筆者作成

の個人についてではなく、アメリカ全体における生徒の学力やその変化の把握にある（NAEP ホームページ参照）。そのために、社会や時代の変化に応じた教育課題に焦点を当てた調査（通称、メイン NAEP）と英単語や四則演算のように、時代の変化に依らず必要とされる基礎的な能力の経年的な変化を捉える調査（通称、トレンド NAEP）および各州で行う調査（通称、州別 NAEP）の3つが行われている。その調査結果は、様々なレベルで活用されており、アメリカの教育政策の改善に大きな影響を与えている。わが国においては、学力低下についての世論の高まりを受け2007年に「全国学力・学習状況調査」として教育調査が再開された。その目的には、児童・生徒の学力や学習状況の実態から教育施策の成果と課題を検証すること、教育に関する継続的な検証改善サイクルを確立すること、学校における教育指導の充実や学習状況の改善を図ることの3つが設定されている（文部科学省ホームページ参照）。調査内容は、国語と算数・数学における主に「知識」に関するA問題と「活用」に関するB問題の2つのテストを用いた学力および質問紙による児童・生徒の生活習慣や学校環境に関する調査である。途上国における国内教育調査については、例えば、アフリカ地域に位置するザンビア共和国（以下、ザンビア）では、ザンビア国家試験委員会（Examination Council of Zambia：以下、ECZ）が1999年から2年から3年毎に継続的に調査を行っている（ECZ、2000、2003、2006、2008、2010）。ここでは英語、数学、理科に加えて現地語による読解力と質問紙調査による生徒の家庭環境について調べられている。その他多くの途上国においても教育調査が盛んに実施されており、教育の実態を探る独自の取り組みが行われている（AEAA、2011；PREAL、2006；UNESCO、2012が詳しい）。

一般にこれらのような教育調査では、数学や理科あるいは読解力に関する学力とともに、学校での学習環境、生徒が置かれている家庭環境、学習に対する姿勢・態度・意欲などについても調査されている。前者は、教育の質向上あるいは保証に対して、生徒の学力に対する評価は不可欠な指標であり、それに対する後者の観点の影響を無視することは難しいと考えられるからである。例えば、表1.4に示したように国際比較というマクロな視点に立つと、先進国や欧米あるいは東アジアの国々は高い成績を修めている（ただし参加国全体で平均値500、標準偏差100で示してある）。しかしながら中東、東南アジア、中米、南米、アフリカ地域の国々は、下位層に位置している。この背景には当然、各国の経済的あるいは教育に対する政策的な側面など様々な要因が考えられるが、一般的にいわれる先進国と途上国<sup>5</sup>の間には、生徒の学力に大きな差異が認められる。加えてPISAやTIMSSをはじめ多くの教育調査では、経年変化を掴むために継続的に調査を行っている。例えば、PISAの「数学的リテラシー」と「科学的リテラシー」調査における経年変化を国立教育政策所（2010）を参考にしながら表1.5に示した。この表からは、極端な変化を示す国は確認されないが、比較的平均値が低い国で変化率が大きい傾向にあるといえるだろう。

<sup>5</sup> 本研究では、OECD 開発援助委員会（Development Assistance Committee：通称、DAC）援助受取国・地域リスト2007年度版（外務省、2008）に挙げられた国々を途上国と規定する。

表 1.4 TIMSS および PISA における得点の上位・下位 10 カ国

TIMSS2007 第8学年 (48)		PISA2006 (57)		PISA2003 (40)			
理科		数学		科学的リテラシー		数学的リテラシー	
国名	平均値	国名	平均値	国名	平均値	国名	平均値
シンガポール	567	台湾	598	フィンランド	563	香港	550
台湾	561	韓国	597	香港	542	フィンランド	544
日本	554	シンガポール	593	カナダ	534	韓国	542
韓国	553	香港	572	台湾	532	オランダ	538
イギリス	542	日本	570	エストニア	531	リヒテンシュタイン	536
ハンガリー	539	ハンガリー	517	日本	531	日本	534
チェコ	539	イギリス	513	ニュージーランド	530	カナダ	532
スロベニア	538	ロシア	512	オーストラリア	527	ベルギー	529
香港	530	アメリカ	508	オランダ	525	マカオ	527
ロシア	530	リトアニア	506	リヒテンシュタイン	522	スイス	527
⋮		⋮		⋮		⋮	
コロンビア	417	アルジェリア	387	モンテネグロ	412	イタリア	466
レバノン	414	コロンビア	380	メキシコ	410	ギリシャ	445
エジプト	408	オマーン	372	インドネシア	393	旧セ*	437
アルジェリア	408	パレスチナ	367	アルゼンチン	391	トルコ	423
パレスチナ	404	ボツワナ	364	ブラジル	390	ウルグアイ	422
サウジアラビア	403	クウェート	354	コロンビア	388	タイ	417
エルサルバドル	387	エルサルバドル	340	チュニジア	386	メキシコ	385
ボツワナ	355	サウジアラビア	329	アゼルバイジャン	382	インドネシア	360
カタール	319	ガーナ	309	カタール	349	チュニジア	359
ガーナ	303	カタール	307	キルギス	322	ブラジル	356

注) ( ) 内は、参加した国および地域数。ただし、国際比較に必要な基準を満たさなかった国の数は除外してある。網掛けは、DAC援助受取国・地域リスト2007年度版にある国(外務省、2008)。  
\*は、旧セルビア・モンテネグロである。

出典：国立教育政策研究所（2004、2007）、IEA（2008a；2008b）と OECD（2004；2007）  
をもとに筆者作成

このように、国際的あるいは各国独自の指標を使って自国の教育を把握する動きが先進国・途上国を問わず拡大している。それに併せて、メイン NAEP やトレンド NAEP あるいはわが国における A 問題や B 問題のように調査内容は多様化し、経年変化などを含め多面的な分析が可能なるようにその実施方法も高度になりつつある。

調査内容が多様化する中で理数科目などの学力を測定する場合、一部コンピュータ上で解答するテスト (Computer-Based Testing) を用いた取り組み (OECD、2010a) もみられるが、現在でも紙ベースのテスト (以下、テスト) を使用するのが一般的である。テストを用いて測定される生徒の学力は、グローバル化によって促進される知識基盤社会に対応するための能力を含む形で捉えられ、その実態を把握する動きが高まっている。言い換えれば、従来からテストが対象としていた「何を知っているか」から「何ができるか」という観点にまで拡張して学力を測定する時代となった (村木、2005)。国際的な教育調査である TIMSS と PISA を例にとると、まず TIMSS では、

≪初等中等教育段階における児童・生徒の算数・数学および理科の教育到達度 (education

表 1.5 PISA における経年変化（「数学的リテラシー」変化率の降順）

国名	数学的リテラシー			科学的リテラシー		
	平均値		変化率	平均値		変化率
	2003	2009		2006	2009	
メキシコ	385	419	8.83%	410	416	1.46%
ブラジル	356	386	8.43%	360	405	12.50%
トルコ	423	445	5.20%	424	454	7.08%
ギリシア	445	466	4.72%	473	470	-0.63%
ポルトガル	466	487	4.51%	474	493	4.01%
イタリア	466	483	3.65%	475	489	2.95%
チュニジア	359	371	3.34%	386	401	3.89%
インドネシア	360	371	3.06%	393	383	-2.54%
ドイツ	503	513	1.99%	516	520	0.78%
スイス	527	534	1.33%	512	517	0.98%
ウルグアイ	422	427	1.18%	428	427	-0.23%
セルビア	437	442	1.14%	436	443	1.61%
ポーランド	490	495	1.02%	498	508	2.01%
香港	550	555	0.91%	542	549	1.29%
アメリカ	483	487	0.83%	489	502	2.66%
韓国	542	546	0.74%	522	538	3.07%
ノルウェー	495	498	0.61%	487	500	2.67%
タイ	417	419	0.48%	421	425	0.95%
ヒリテンシュタイン	536	536	0.00%	522	520	-0.38%
ハンガリー	490	490	0.00%	504	503	-0.20%
ロシア	468	468	0.00%	479	478	-0.21%
スロバキア	498	497	-0.20%	488	490	0.41%
ラトビア	483	482	-0.21%	490	494	0.82%
マカオ	527	525	-0.38%	511	511	0.00%
スペイン	485	483	-0.41%	488	488	0.00%
フィンランド	544	541	-0.55%	563	554	-1.60%
ニュージーランド	523	519	-0.76%	530	532	0.38%
ルクセンブルク	493	489	-0.81%	486	484	-0.41%
日本	534	529	-0.94%	531	539	1.51%
カナダ	532	527	-0.94%	534	529	-0.94%
アイスランド	515	507	-1.55%	491	496	1.02%
オーストラリア	524	514	-1.91%	527	527	0.00%
デンマーク	514	503	-2.14%	496	499	0.60%
オランダ	538	526	-2.23%	525	522	-0.57%
ベルギー	529	515	-2.65%	510	507	-0.59%
フランス	511	497	-2.74%	495	498	0.61%
スウェーデン	509	494	-2.95%	503	495	-1.59%
アイルランド	503	487	-3.18%	508	508	0.00%
チェコ	516	493	-4.46%	513	500	-2.53%

出典：国立教育政策研究所（2010）をもとに筆者作成

al achievement) を国際的な尺度によって測定し、児童・生徒の学習環境条件等の諸要因との関係を参加国間におけるそれらの違いを利用して組織的に研究すること」(国立教育政策研究所、2005、p.2)

と目的が設定されている。これは、学校教育カリキュラムで学習した知識や技能をどの程度習得しているかという視点から学力を捉えて、それに対応する内容のテストが使用され

ている。つまり、以前から対象とされていた学力の側面に焦点を当てていると言えよう。他方で PISA では、「生徒が身に付けるべき総合的な技能は、教科横断的に育成されるもの」という考え方から、その目的を、

「多くの国で義務教育修了段階にある 15 歳児を対象に、読解力、数学的リテラシー、科学的リテラシーの各分野について、将来生活していく上で必要とされる知識や技能をどの程度身に付けているかを、国際的な水準に照らして測り、それを各国の教育政策の改善や見直しに活かすことを目的とする」(国立教育政策研究所、2010、p.2)

と定めている。これは、特定の学校教育カリキュラムに依存しない観点で学力を測定しようという試みであり、今日的な学力の捉え方を基本としている。そのため、TIMSS とは異なる側面を測定するために開発されたテストが使用されている。なお上記の 3 つの分野は、学校教育の科目に対応しているが、その内容は教科横断的な側面を十分に意識したものとされる。

このように TIMSS では、学校教育カリキュラムに対応した知識や技能の習得度合いに主眼があり、他方で PISA では、カリキュラムの枠組みを超え、習得した知識や技能を活用する能力の測定に重点がある。言い換えると TIMSS は、カリキュラムに即した知識や技能の習得度合い (achievement) を測定するためのテスト (以下、アチーブメント・テスト)、PISA では、それらを活用する能力の習得度合い (assessment) を測定するためのテスト (以下、アセスメント・テスト) がそれぞれ開発され、実用されている (日本テスト学会、2010、p.61)。この 2 つのテストの特徴は TIMSS や PISA だけでなく、アメリカにおけるメイン NAEP とトレンド NAEP、わが国における A 問題と B 問題にも当てはまる枠組みである。ちなみに SACMEQ、PASEC および LLECE では、アチーブメント・テストを用いて生徒の学力を測定している (UNESCO、2009)。このように、生徒の学力に対する考え方は大きく 2 つに分かれ、それに合わせてアチーブメント・テストまたはアセスメント・テストが作成され、使用されている。

その一方で、測定される生徒の学力が多様になるにつれ、テストを使った調査の実施方法も高度化しつつある。なぜならば調査時間などの現実的な制約のもとで、いかに効率的に幅広く学力を測定するかが問われるからである。ここでは PISA を具体的に取り上げ、その実施方法を確認したい。対象とする学力を幅広く測定するには、多種多様な出題項目 (以下、項目) を作成し、内容に偏りなく数多くの項目を組み込んだテストが望まれる。ところが各受験者が解答できる項目数には限りがあり、調査時間が長くなることもあまり好ましくない。このような問題点を考慮して、PISA では、「重複テスト分冊法」と呼ばれる実施方法が採用されている。これは、一人の受験者では解答しきれない多数の項目を大勢の受験者に割り当てながら解答してもらうように工夫された方法である。例えば PISA2003 では、数学的リテラシー85項目、科学的リテラシー35項目、読解力28項目、問題解決19

項目の合計 167 項目をいくつかの項目群に分け、それらを組み合わせて 13 種類のブックレット、つまり 13 種類のテストが準備されている (表 1.6 参照)。各項目群における解答時間は 25 問 30 分であり、各受験者は、13 種類のブックレットのうちの一冊を合計 2 時間で解答する。それらを合算することで、全 167 項目に対するデータが収集される。またこの他にも、特別な対応が必要な受験者を想定した項目数が少ない UH ブックレットも作られている。なお、PISA2006 では PISA2003 と同様にそれぞれ 4 つの項目群で 13 種類のブックレットが使用されており、PISA2009 では 20 種類のブックレットが準備された (OECD、2006、2009)。このように、重複テスト分冊法を用いて効率的に測定を行い、テスト得点が

表 1.6 PISA2003 における 13 種類のブックレット (テスト) の構成

ブックレット 番号	項目群1 30分	項目群2 30分	項目群3 30分	項目群4 30分
1	M1	M2	M4	R1
2	M2	M3	M5	R2
3	M3	M4	M6	PS1
4	M4	M5	M7	PS2
5	M5	M6	S1	M1
6	M6	M7	S2	M2
7	M7	S1	R1	M3
8	S1	S2	R2	M4
9	S2	R1	PS1	M5
10	R1	R2	PS2	M6
11	R2	PS1	M1	M7
12	PS1	PS2	M2	S1
13	PS2	M1	M3	S2

注) M: 数学的リテラシー項目 (M1からM7の7項目群)  
 S: 科学的リテラシー項目群 (S1とS2の2項目群)  
 R: 読解力項目群 (R1とR2の2項目群)  
 PS: 問題解決項目群 (PS1とPS2の2項目群)

出典: OECD (2005a) をもとに筆者作成

算出され、国際比較可能な数値が公表されている。ところが、異なる種類のテストを用いるこの方法では、測定対象 (例えば、数学的リテラシー) が各テストで統一され、なおかつその結果が比較可能なのか、という疑問が生じる。そこで表 1.6 をみると、ブックレット間には共通に含まれる項目群を確認できる。例えば、ブックレット 1 とブックレット 2 には、M2 が共通に含まれ、ブックレット 2 とブックレット 3 には、M3 が共通の項目群になっている。このように、一定数の項目を共有させて異なるテストを作成することを「共通項目計画」といい、それらの項目は「共通項目」と呼ばれる (豊田、2002)。この共通項目計画によるテスト間では、項目反応理論<sup>6</sup> (Item Response Theory: 通称、IRT) と呼ばれ

<sup>6</sup> 項目反応理論について、芝 (1991)、大友 (1996)、豊田 (2002a、2002b、2012b)、村木 (2011) および植野・荘島 (2010) などが詳しい。項目反応理論の概要を述べるならば、この理論は、各項目の特性を項目反応モデルを用いて同定し、受験者の能力特性を推定するために開発された

る現代テスト理論を適用すれば、テスト得点を同質のものとして扱うことが可能になる。纏めると PISA では、共通項目を含む重複テスト分冊法によって幅広く測定ができるように工夫され、項目反応理論を適用してテスト得点を算出し、国際比較可能な数値が公表されている。また表 1.5 でみたように、PISA2003 と PISA2006 および PISA2009 間でも共通項目計画によってテストが作成されており、項目反応理論を用いることで経年変化を捉えることができる。翻って考えると、共通項目計画において精度の高い調査を実現するためには、共通項目の役割りは、重要かつ本質的であり、その管理は、経年変化を調べる上で特に注意しなければならない。そのため共通項目は、非公開にするのが一般的であり、PISA では、いくつかの項目のみが公開されている。なおこのような重複テスト分冊法と項目反応理論による方法は、PISA だけでなく、TIMSS や LLECE あるいは NAEP においても採用されている。加えて PISA や TIMSS では、収集したデータをウェブ上で公開しており（PISA ホームページあるいは TIMSS ホームページ参照）、誰でも公開データ (raw data) を入手し、分析できる体制を整えている。そのため OECD や IEA が刊行する報告書に記載された結果だけでなく、様々な視点からの二次的な分析（以下、二次分析）が可能なのである。

他方でわが国の全国学力・学習状況調査では、重複テスト分冊法を用いた項目反応理論によるテスト得点の算出は行われていない。そのため TIMSS や PISA のように経年変化を捉える分析が容易ではないのが現状である。その背景には、わが国独特のテストに対する風潮が影響しているとされる（柴山、2008）。例えば経年変化をみる場合、原理的には、同一のテストを異なる時点で同等の集団に対して行うことで可能になる。ところがわが国においては、実施後に項目が全て公開されテスト対策の指導に使用するという考え方があるため、一度使用したテストを別の時点で実施したとしても適切に経年変化が捉えられるとは限らない。つまり、わが国における全体的な特徴を調べることを目的の一つに据えた調査に対してテスト対策が行われ、テストに対する余剰な学習効果が生じてしまい、精度の高い調査が困難なのである（柳井・石井、2008）。これはテストを実施する目的が、個々の生徒の能力特性を浮かび上がらせること（例えば、選抜や個に応じた指導への診断的評価）に集中しがちであり、集団の特性を明らかにするということにあまり置かれていないとい

---

テスト理論の一つである。項目反応モデルとは、測定の対象となる能力特性を独立変数として持つその項目に正答する確率を表現するための数理モデルである。一般の項目反応モデルには、困難度や識別力などの母数を含んで定義される。母数の困難度は、その値が大きいほど難易度の高い項目を表わす。また識別力は、その値が大きいほど能力特性を識別するのに効果的な項目であることを意味する。項目反応理論の特質すべき点は、①どんな異なったテストを用いても共通の尺度で能力特性を測定できること、②どんな受験者集団においても出題項目の特性（困難度や識別力）を共通に表現できることに集約され、項目反応理論の不変性と呼ばれる（大友、1996）。しかしこの不変性の利点を利用するためには、「等化」という過程を踏む必要がある。なお等化の方法については、様々なものが提案されている。その詳細については、豊田（2012b）が詳しい。なお本研究では、芝（1991）および豊田（2002a、2002b、2012b）を参考にした。また項目反応理論を用いた分析では、統計解析ソフト R に含まれるパッケージ ltm (Dimitris, 2006) を用いて数値を算出した。

うわが国独特の文化的な側面なのかもしれない。そのため、全ての項目を公開するという特徴を考慮しながら、テスト得点など、得られる数値を比較可能とするために様々な研究がされ始めている。例えば石井・安永（2011）は、全国学力・学習状況調査などの公的な大規模調査のテストと地方自治体が独自に行う調査のテストを共通に受験した生徒に着目して対応付け、経年変化をみることを試みている。また重複テスト分冊法と項目反応理論による方法は、国際的にみて標準的なものになりつつあることから、それらをわが国の文脈に合わせる形で実施するための基礎的な研究も進められている（柴山、2012）。

これまで述べたように教育の質向上を目的とし、今日では、国内外で盛んに教育調査が行われ、テストを使って測定する内容は多様化し、その実施方法も高度化している。そのため UNESCO（2000、p.66）が指摘するように、教育調査の需要が拡大するにつれ、テストの開発・実施・利用・管理に至るまでの「テストの専門家」が今後ますます必要になると思われる。特に PISA や TIMSS のような国際的な教育調査では、収集したデータを公開しており、もしこれらのデータを二次分析するならば、そのテスト開発や実施方法から原則的には項目反応理論を適用しなければならない。なぜならば、同一のテストを使用することを前提とした従来の古典的テスト理論（Classic Test Theory：通称、CTT）による分析では、対応できないテストの構成になっているからである。ところがわが国においては、教育測定・統計などテストに関する科目が大学の学部教育でほとんど扱われていない状況にある。この背景について木村（2010）は、

「戦後直後の教員養成カリキュラムに「教育測定・統計」科目が位置付けられていたにも拘らず、その後、中教審や臨教審の各答申の中で、「教育心理学」が、カウンセリングや生徒指導といった児童理解ツールにのみ定義が縮小化されていく中で、教員養成カリキュラムの中から「教育測定・統計」を含んだ教育心理学という概念自体が自然消滅していった」

と述べており、テストに関する知識・技能の習得は、就職後あるいは修士課程在学中になされる場合が多いとされる。特に古典テスト理論や項目反応理論のテスト理論に関して、独学で習得した者は少なくないと報告している。また豊田（2012b）は、

「項目反応理論は、アメリカ合衆国はもとより、ヨーロッパの多くの国でもテスト理論のスタンダードとして不動の地位を築いています。中国や台湾など、アジア諸国の統一試験の運用にも使用されています。産業界にも浸透しています。・・・(省略)・・・しかし、現在、わが国のその他の関連学会には項目反応理論が、あまり普及していません。学会を飛び越えて、産業界にだけ普及してしまう現状は、たいへん奇妙です。同時に、学会の怠慢を示しています。・・・(省略)・・・尺度構成やテスト作成の領域で、わが国の学部生が項目反応理論を学習することが当たり前になる状況を願っています」(p.1、

下線部は筆者の加筆)

と述べ、その重要性を主張している。このようにわが国においては、PISA や TIMSS で公開されたデータを効果的に分析するための方法論が普及していないのが現状である。

実際、教育調査で頻繁に使われる科目の数学に関する数学教育研究に目を向けると、PISA や TIMSS といった大規模調査の結果を平均点などの単純な比較を超えて、生徒の学習状況や教育課程の課題をどう読み解くのかに繋がる多面的な分析が求められている（清水、2006）。ところが長崎・萩原（2004）では、日本の数学教育研究で項目反応理論を用いた研究が欧米と比べて少ないとし、その利用法を事例的に示している。また鈴木・豊田・川端（2008）は、PISA2003 の数学的リテラシー調査の公開データに項目反応理論を適用し、わが国の生徒は日常に近い状況とされた問題で数学を活用することに難があると指摘する中で、項目反応理論を有効に利用した数学教育研究の必要性を説いている。つまり清水（2012）が述べるように、信頼性の高い分析に向けて新たな手法を導入することは、今日的な数学教育研究における課題として認識されるべきものである。

## 1.2. 本研究の目的およびその方法

### 1.2.1. 本研究の目的

前小節で設定した今日的な数学教育研究の課題から本研究では、生徒の数学学力と①カリキュラムの特徴、②生徒の情意的側面および③言語的側面の 3 つの視点の関連性を実証的に明らかにすることを目的とする。そしてその際には、公開された大規模調査の結果を単純に比較するのではなく、項目反応理論を適用しながら、公開データを活用する方法論的な課題にも応えていく。

### 1.2.2. 本研究における分析の観点と主要な分析手法

わが国の数学教育研究において、教育調査の結果を解釈する手段として公刊される報告書にあるテスト得点の分布や正答率を参考にするのが主流であり、収集されたデータが限定的に活用されているのが現状である。言い換えると、生徒の数学学力と設定した 3 つの視点との関連性は、報告書にある分析結果の範囲で考察されており（e.g.、鈴木、2006；瀬沼、2008；清水、2007）、その範囲を超えた研究が今日求められている。これを踏まえ本研究では、教育調査の公開データ（raw data）から出発する二次分析を試みる。

加えて、前小節で述べた MFA 実現に向けた今日的な数学教育研究の課題は、わが国だけでなく、途上国を含めた他の国にも目を向けて論じるべきと考える。なぜならば MFA 実現に向けて Gates & Vistro-Yu（2003）が提起した課題は、先進国だけでなく途上国にも目を向ける必要性に触れているからである（表 1.1 参照）。ゆえに本研究では、わが国をはじめとする先進国だけでなく、途上国に関する考察も同様に取り組んでいく。わが国の途上国に関する理数科教育研究は始まったばかり（馬場、2008）とされるが、生徒の社会的・文

化的背景あるいは学力水準がわが国と大きく異なる途上国を理解する上で、日本といった国と対比させながら実態を探る意義は、必ずしも小さくないと考える。先進国と途上国と区別して、それぞれの固有あるいは共通する特徴を掴むことは、新たな研究課題の提示に繋がると期待できよう。ゆえに本研究では、途上国を含めた他の国にも目を向けながら各国の特徴を明らかにする方法、つまり国際比較という分析の立場を取る。

そこで途上国に目を向けると、一般に、これらの国の国際的な教育調査における成績は低いことが知られている (e.g., OECD, 2004, 2007, 2010b; IEA, 2005, 2008a, 2008b)。このことは、途上国における教育の質の低さを表わす最も目にする根拠であり、わが国と大きく異なる点の一つである。そのため途上国の教育の質向上に向けて、生徒の学力向上は、数学教育研究が取り組む主要な研究課題となっている。ところがそれに向けて、教育調査の結果から明確な見解は必ずしも得られていない。なぜならば内田 (2009) および馬場・内田 (2008) によると、途上国でも特にアフリカ諸国のように極度に低い結果を得た場合、一般に知られる低い学力という特徴のみが浮かび上がり、生徒の学力向上に繋がる有意義な示唆を得ることが難しいとされるからである。それを受け内田 (2009) は、平易な問題を用いるなど工夫した調査の必要性から、SACMEQ で最も低い結果となったザンビアにおいて、基礎的な計算に特化したテストを用いて調査を行っている。これらの研究の根底にある主張は、途上国の低い学力のみを強調してしまうテストの調査ではなく、その国あるいは地域に見合ういわばオーダーメイドともいえるテストを用いた調査が必要ということである。とはいえ、国際的な教育調査から有益な情報が得られにくいとする見解は、TIMSS や SACMEQ の報告書にあるテスト得点の分布や出題項目の正答率を単純に比較した分析から得られている。

国際比較を可能とする教育調査では、特定の国や地域に特化した内容を含むテストは使用されない。なぜならば信頼性の高い調査結果を得るには、ある国や地域に著しく有利または不利に働くテストは好ましくないためである。つまり教育調査では、オーダーメイドのテストではなく、現状を客観的に捉える標準テストを用いて、国や地域の特徴の全体像を浮かび上がらすことに大きな目的がある。言い換えると、教育調査が担うべき主要な役割りに学力水準の高低を提示することが含まれるのである。テスト得点の分布や正答率は、それらを具体的に示す最も知られた指標として現在使われている。これを踏まえると、テスト得点や正答率に焦点を当てる分析では、おのずと得られる情報が学力水準といった側面に限られてしまう。したがって教育調査において、内田 (2009) や馬場・内田 (2008) が指摘した「途上国の低い学力という特徴が、有意義な示唆を得にくくする」のではなく、テスト得点の分布や正答率を重視する方法にその要因があると考えられる。途上国が参加している国際的な教育調査、例えば TIMSS や PISA あるいは SACMEQ などでは、収集したデータを公開している。公開されるデータは、一般にサンプルサイズが大きく、一個人が収集することは大変難しい貴重なものである。わが国の途上国における理数科教育研究は始まったばかりであるが、生徒の社会・文化的背景あるいは学力がわが国と大きく異なる途

上国のより深い理解に向けて、国際比較が可能な公開データを活用する意義は大きい。纏めると、国内あるいは他の国、特に途上国の教育に関する数学教育研究において、教育調査の公開データを活用した実証的な取り組みが今後求められるだろう。そしてそれに向けて、テスト得点の分布や正答率といった学力水準を反映する指標に注目するだけでなく、新たな分析の観点を導入する必要があると考える。

一般に教育調査の公開データは、無回答、部分点や欠損値を区別できる形で公開されている。しかしそのデータのままでは分析が困難なため、図 1.3 のように、正答を 1 誤答を 0 などとする 2 値反応データ (binary response data) に変換し、それを分析対象とすることが多い。この場合正答率は、各項目に正答した受験者数のその項目を解答した全受験者数に対する比率となる。また 2 値反応データからテスト得点を算出する方法はいくつかあるが、古典的テスト理論の枠組みでは、各受験者の項目得点の総和、つまり正答の数が各受験者のテスト得点となる。このようにテスト得点は、項目得点を統合して得られる指標であるため、各々の項目が持つ特性がテスト得点の分布に大きく関わってくる。

そこでまず、出題項目の難易度として最も使用される指標の正答率について考える。正答率を用いる分析には、少なくとも 2 つの立場がある。1 つ目は、公開問題など具体的に出題内容が分かる場合、内容を吟味しながらその項目に対する生徒の学力水準を把握することである。2 つ目は、具体的な出題内容が明らかでなくとも、図 1.3 のように各項目が数学的な領域などで特徴付いていれば、領域ごとに生徒の傾向が掴むことができる。纏めると前者は、特定の項目に対する考察を深めることを目的とし、後者は、全ての項目を通じて浮かび上がるその受験者集団が有する解答のパターン (以下、解答パターン<sup>7</sup>) を把握する

図 1.3 2 値反応データの例 (シュミレーション)

受験者	出題項目																							
	代	代	代	代	代	幾	幾	幾	幾	幾	関	関	関	関	関	確	確	確	確	確	...			
1	0	0	0	1	0	0	0	1	1	0	1	1	0	1	1	0	0	1	0	0	1	0	1	
2	1	1	0	1	0	0	0	1	0	1	1	0	1	1	1	1	0	0	1	0	1	0	0	
3	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	1	1	0	1	1	0	1	1	
4	0	1	0	1	0	0	0	0	1	1	1	0	1	0	1	1	0	0	0	1	0	0	1	
5	1	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	
6	0	0	1	1	1	1	1	0	1	1	0	1	0	1	0	0	0	0	1	1	0	1	1	
7	1	1	0	0	1	0	1	1	1	1	0	1	1	1	1	0	1	0	0	1	0	0	1	
8	1	0	0	0	1	1	1	1	0	0	1	1	1	0	0	0	1	0	0	1	0	0	1	...
9	0	0	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	1	0	1	0	
10	0	1	1	1	0	1	1	1	1	0	0	1	1	1	1	0	1	0	0	1	0	0	1	
11	0	1	1	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	1	
12	1	1	1	0	1	1	1	1	0	1	1	1	1	0	0	1	1	1	0	0	1	0	0	
13	0	1	1	1	1	1	1	1	0	0	1	1	0	0	0	1	1	1	0	0	0	0	0	
14	0	1	1	1	1	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	
15	0	1	1	0	0	1	1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	
⋮																								

注) 代は代数、幾は幾何、関は関数、確は確率・統計を意味する。

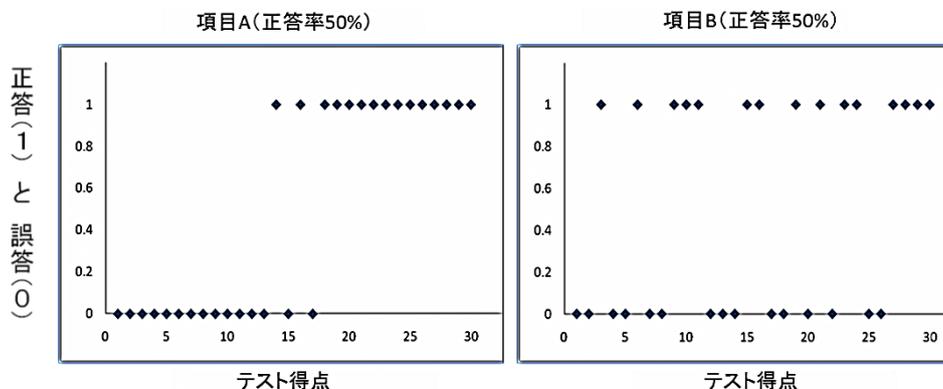
<sup>7</sup> 解答パターンとは、図 1.3 のようなテストを用いて収集されるデータにみられる規則性のことを指す。

ことに重きがある。多くの教育調査では、その実施方法とも関連して一部の項目のみが公開されるため、正答率は、解答パターンを掴むために使われることが多い。

他方でテストを構成する項目は、正答率といった難易度だけでなく、「識別力」と呼ばれる指標によっても特徴付けられている。識別力とは、テストの信頼性に繋がる重要な概念で、「各項目がテスト全体で測定する能力特性をどれだけ反映するか」を示すものである（Steven & Howard、2009、p.5；日本テスト学会、2010、p.10）。例えば図 1.4 のように、正答率が 50%の項目を 2 つ用意した。項目 A では、テスト得点の高い受験者ほど正答を得ているが、項目 B では、テスト得点の高低に拘らず、正答または誤答となっている。このように正答率が同じであっても、テスト得点の高低をうまく反映する項目とそうでない項目が存在することがある。この場合、項目 A は項目 B よりも識別力が高いという。特に識別力が低い項目は、テスト得点との関連性が低いことを意味するため、測定に対してあまり好ましくないと判断され得る。識別力は、正答率のような項目の難易度に関するものではなく、受験者の能力特性を見分けるのにその項目がどれだけ役立つかに関する情報を与えてくれる。なお古典的テスト理論では、各項目とテスト得点の相関係数を識別力の指標とすることが多い。これを踏まえると解答パターンを明らかにする分析において、従来の正答率に重点を置く方法（例えば、国立教育政策研究所、2004、2007、2010 を参照）に加え、識別力にも注目することで各項目の特性をより多面的に分析できると考える。

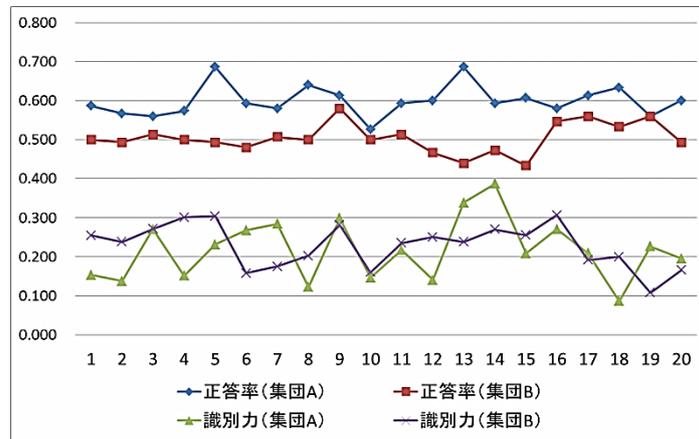
しかしながら国際比較のように、ある特定の受験者集団ではなく、いくつかの集団間でそれらを比較するには、注意しなければならないことがある。そこで図 1.5 に、異なるテスト得点の分布を持つ 2 つの受験者集団における正答率と識別力をシミュレーションした結果を示した。このテストは 20 点満点で、集団 A におけるテスト得点の平均値は 11.92 で標準偏差は 2.13 であり、集団 B ではそれぞれ 10.03 と 2.34 である。また受験者数は、各集団 150 名とした。なお t 検定の結果、2 つの平均値は統計的に有意な差であり、効果量が 0.847 を示したため大体標準偏差 0.85 程度の差異となっている。まず全体的な傾向は、集団 A の

図 1.4 識別力の高い項目と低い項目の例（シミュレーション）



出典：筆者作成

図 1.5 2つの集団における正答率と識別力（シミュレーション）



出典：筆者作成

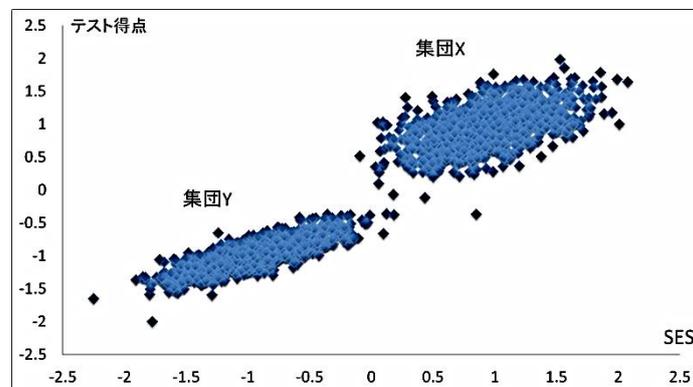
方が集団 B よりも正答率が高くなっている。このことは、両集団のテスト得点の分布の違いからも理解される。集団 A においては、項目 5 と項目 13 は正答率が高く、他と比べて容易な項目といえる。また集団 B では、項目 9 は易しく、項目 15 は難しいことが分かる。ところが、この 2 つの受験者集団において項目 13 に注目すると、集団 A では容易な項目であるが、集団 B においては必ずしもそうではない。また項目 10 と項目 19 は、集団 A では難易度が高いが、集団 B においては易しい項目と判断できる。このようにいくつかの項目の難易度が、集団 A と集団 B の間で相違していることを確認できる。この現象は、出題項目の難易度と受験者集団の能力特性という本来異なるものを同時に説明してしまう正答率が持つ特性によって生じている。言い換えると正答率には、その項目の本来の難易度と受験者集団に依存して決定されるという性質がある（芝、1991）。そのため国際比較のように各国の解答パターンを調べる際には、出題項目と受験者集団の特性を区別することがより客観的な分析に繋がる。このような分析を可能にする手法として、現在では項目反応理論が知られている（大友、1996；豊田、2002）。なお正答率に関する上述の議論は、相関係数を用いた識別力についても同様である。項目反応理論の利用は、解答パターンをより信頼性を確保して明らかにするに向けて、有効な手段の一つと考える。

これまで解答パターンを検討するために、正答率だけでなく、識別力にも目を向けることを述べた。またそれらの数値が有する限界から、項目反応理論を用いることの有効性について触れた。これらは、出題項目に目を向けた分析の観点である。テストを用いる調査では、受験者の総合的な能力特性、つまり総合的な学力の水準を表わすテスト得点に注目する分析も有効である。実際教育社会学研究では、テスト得点を従属変数とし、SES や性別あるいは学校の違いなどを独立変数に置き計量的に分析することが多い。その一方で途上国に関する研究に目を移すと、テスト得点を従属変数に置いた場合、その低さゆえに説明変数で分析しようとしても、一般的に差異が出にくく得られる情報は少ないとされる（馬

場・内田、2008)。つまりテスト得点の分布が低位にある場合には、教育社会学研究で使用されるような方法があまり機能しないことを主張している。そこで図 1.6 にテスト得点を従属変数、SES を独立変数とするシミュレーション結果の散布図を示した。この図では、生徒の SES と学力水準に正の相関関係がみられるという定説を基にしている。図から分かるように、集団 X は、テスト得点と SES とともに集団 Y より高くなっている。ところが相関係数を計算すると、集団 X は 0.515 で集団 Y では 0.810 となった。また回帰分析における決定係数は、それぞれ約 71%と 90%である。このように、独立変数の選び方次第で従属変数を適切に説明し得ることは想像に難くない。実際 PISA2003 の報告書では、「数学的リテラシー」調査で低い結果を得たメキシコにおいて、十分な説明力を持つ分析がされている (OECD、2004、p.200)。ここで取り上げた分析は、従属変数と説明変数の関連の有無あるいはその程度を明らかにすることを目的としており、変数間の相対的な関係性を表わす情報を与えてくれる。つまり学力水準の違いといった情報は失われるが、変数間の関係性という観点で分析が可能なのである。これを翻ると学力水準という情報が一定程度排除されるため、途上国のように低さのみが強調されるとしても、得られる情報は少ないのではなく、むしろそれが直接的に作用しない有益な情報を取り出せる可能性を秘めているのである。ちなみに変数間の関係性を表わす指標としては、相関係数や回帰分析における決定係数などがよく知られている。

本研究では、国際比較の視座から計量的な分析を試みる。つまり国を単位に各受験者が属する集団が明確である入れ子構造を持つデータを分析対象とするため、入れ子構造を持つデータに頻繁に使用される階層線形モデルを主な手法として用いていく。纏めると、これまでのテスト得点の分布や正答率を重視し国際比較する方法を批判的に捉え、項目反応理論による「等化」を利用して解答パターンをより詳細に掴むこと、および階層線形モデルなどを採用し、変数間の相対的な関係を表わす指標に目を向けることの 2 点を留意していく。

図 1.6 2つの集団における SES とテスト得点の散布図 (シミュレーション)



出典：筆者作成

### 1.2.3. 使用データとデータ処理および使用する項目反応モデル

本研究では、生徒の数学学力と①カリキュラムの特徴、②生徒の情意的側面および③言語的側面の3つの視点の関連性を国際比較の視座から明らかにするために、大規模調査の公開データ (raw data) から始まる二次分析を試みる。

TIMSS と PISA といった教育調査では、重複テスト分冊法による数学や理科あるいは読解力に関する学力の測定だけでなく、それらに対する情意的側面について質問紙<sup>8</sup>を用いた調査を行っている。そのため、テストと質問紙を用いて収集された公開データを二次分析することで、生徒が有する学力に対して様々な視点からの考察が可能と考える。TIMSS においては、調査内容が数学と理科に関するものに限定されている (e.g., IEA, 2005, 2008a, 2008b)。ところが PISA では、理数科に関する学力と情意的側面だけでなく、読解力<sup>9</sup>の調査を継続的に行っている。ゆえに生徒の言語的側面を反映する指標の一つとして読解力を位置付けるならば、本研究で焦点を当てる数学学力と言語的側面の関連性に関する考察を行える。そこで本研究では、PISA の公開データを分析対象に定め、数学学力を「数学的リテラシー」に統一して、カリキュラムの特徴と情意的側面および言語的側面の3つの視点に立つ分析を試みる。

加えて本研究では、日本を含むいくつかの先進国と数学的リテラシーの水準が最下位層に位置する途上国に特に焦点を当てながら国際比較分析を行っていく。先進国としては、国立教育政策研究所が公刊する報告書 (国立教育政策研究所, 2004) や鈴木・豊田・川端 (2008) で取り上げられているオーストラリア、カナダ、フィンランド、フランス、ドイツ、香港、アイルランド、イタリア、日本、韓国、オランダ、ニュージーランド、アメリカの13ヵ国を取り上げる。また途上国については、表1.4に示した PISA2003 の数学的リテラシーに関する結果で下位層に位置しているトルコ、タイ、ウルグアイ、メキシコ、ブラジル、チュニジア、インドネシアの7ヵ国に注目する。

では、第2章以降において中心的な分析対象となるテストを用いて収集されたデータの処理方法について、PISA2003 の数学的リテラシー調査のデータを例に取り上げて説明する。なお情意的側面については、第3章のみで扱うためそのデータ処理の方法は、第3章で述べることにする。

PISA2003 には、41ヵ国<sup>10</sup>が参加している。本研究では、それらの国のうちイギリスと

<sup>8</sup> 例えば PISA2003 では、5つの情意的側面に関する調査が行われている。その具体的な質問項目については、付録1.1を参照されたい。

<sup>9</sup> PISA における読解力は、「自ら目標を達成し、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、書かれたテキストを理解し、利用し、熟考する能力」と規定されている (国立教育政策研究所, 2004, p.16)。

<sup>10</sup> オーストラリア、オーストリア、ベルギー、ブラジル、カナダ、スイス、チェコ、ドイツ、デンマーク、スペイン、フィンランド、フランス、ギリシャ、香港、ハンガリー、インドネシア、アイルランド、アイスランド、イタリア、日本、韓国、リヒテンシュタイン、ルクセンブルク、ラトビア、マカオ、メキシコ、オランダ、ノルウェー、ニュージーランド、ポーランド、ポルトガル、ロシア、スロバキア、スウェーデン、タイ、チュニジア、トルコ、ウルグアイ、アメリカ、

リヒテンシュタインを除いた 39 カ国の生徒を分析対象とする。なぜならばイギリスは、PISA2003 における国際的に定められた実施基準を満たしていないためである (OECD、2005a、p.238)。またリヒテンシュタインは、受験者数が 332 名と特に少ないため分析対象に含めなかった。PISA2003 では、数学的リテラシー調査のために 2 時間で解答する 13 種類のブックレットが作成され、受験者はそのうちの 1 冊に取り組むという方法でデータが収集された。ところがこの 13 種類のブックレットの他に、特別な必要がある場合を想定した UH ブックレットという短い時間で解答できるものも準備されている。このブックレットに解答した受験者は、39 カ国の全受験者 266298 名のうち 1048 名と少ないため、これらの受験者も分析対象に含めないことにした。

PISA2003 の数学的リテラシー調査では、85 個の項目を用いてその測定が行われた。ところが 85 項目のうち、全ての国に対して整合性が低いと判断された項目<sup>11</sup>が実施後の分析から 1 つ指摘されたため、実際には 84 項目による測定結果が公表されている。出題された 84 項目は、自由記述形式や選択肢形式など様々な出題形式で作成されており、部分配点を持つ項目が 7 つ含まれている。しかし本研究では、個々の項目の詳細な分析というよりも、84 項目を通じて各国の数学学力の全体的な特徴を浮かび上がらせることに重きを置く。そのため部分配点を考慮すると、分析だけでなく分析結果の解釈が煩雑になる。そこで部分配点を持つ項目が全体の 1 割に満たないことを踏まえ、OECD が公開しているコードブック (OECD、2005b) を参考にしながら、無回答、Not reach、部分点、欠損値は誤答として扱い、全ての項目において正答を 1 誤答を 0 とする 2 値データを作成した。なお PISA のコードブックに示されている N/A は、受験者が解答したブックレットに含まれない項目を意味する。そのため各項目の受験者数は、全受験者の約 3 分の 1 程度に限定されており (OECD、2005a)、N/A を誤答として扱うことは適切でない<sup>12</sup>と判断した。本研究で使用する統計解析ソフトウェア R に含まれる項目反応理論を実行するためのパッケージ ltm (Dimitris、2006 ; 服部、2011) では、N/A を出題されていない項目としての欠損値 (R における欠損値は NA と規定されている) として処理すれば、項目反応理論による計算を適切に行うことができる。したがって作成したテストデータセットは、図 1.7 に示したような正答 1 と誤答 0 および欠損値 NA を要素に持つ行列となっている。

この図から分かるように、本研究では、265250 名を全受験者とするデータセットを作成し、分析に使用する。なお各国の受験者数<sup>12</sup>については、付録 1.2 を参照されたい。

---

旧セルビア・モンテネグロ、イギリスの 41 カ国・地域。なお本稿では、国立教育政策研究所 (2004) に合わせ、国名の表記については、煩雑さを避けるために、アメリカ合衆国をアメリカと表記するなど、全ての国において略称を用いることとする。また香港およびマカオは国 (中国) としての参加ではないが、ここでも煩雑さを避けるために、以下では「国・地域」という表記はせず「国」とする。

<sup>11</sup> 国際的なレベルで削除された項目は、M434Q01T である (OECD、2005a、p.190)。

<sup>12</sup> テストを用いて収集された読解力のデータについても同様な処理を施した。なお第 4 章では、数学的リテラシーと読解力の共通受験者に焦点を当てるため、付録 1.2 に併せて示したように、各国の受験者数は異なっている。

図 1.7 データ処理後のデータセットの一例

受検者	国名	ブックレット	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	...
1	AUS	11	0	NA	NA	NA	NA									
2	AUS	9	NA	1	0	0	NA									
3	AUS	12	NA	1	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	...
4	AUS	1	1	1	NA	NA	1	0	1	0	1	NA	NA	NA	NA	
⋮																
50000	CAN	12	NA	0	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	NA	
50001	CAN	1	1	0	NA	NA	0	0	1	0	1	NA	NA	NA	NA	
50002	CAN	1	1	1	NA	NA	1	1	1	1	1	NA	NA	NA	NA	...
50003	CAN	10	NA	NA	NA	0										
⋮																
100000	FRA	6	NA	0	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	0	
100001	FRA	4	NA	NA	NA	NA	1	0	1	0	NA	1	1	0	NA	
100002	FRA	12	NA	0	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	NA	...
100003	FRA	7	NA	NA	1	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	
⋮																
150000	JPN	6	NA	1	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	0	
150001	JPN	9	NA	1	1	0	NA									
150002	JPN	7	NA	NA	1	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	...
150003	JPN	13	1	NA	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	
⋮																
200000	NOR	5	1	NA	1	0	0	0								
200001	NOR	4	NA	NA	NA	NA	1	0	1	0	NA	1	1	1	NA	
200002	NOR	6	NA	0	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	1	...
200003	NOR	8	NA	NA	NA	NA	1	1	1	1	NA	NA	NA	NA	NA	
⋮																
250000	URY	2	NA	0	0	0	NA	NA	NA	NA	0	0	0	0	NA	
250001	URY	4	NA	NA	NA	NA	0	0	0	0	NA	0	0	0	NA	
250002	URY	8	NA	NA	NA	NA	0	0	1	0	NA	NA	NA	NA	NA	...
250003	URY	5	0	NA	0	0	0	1								
⋮																
265247	YUG	5	0	NA	1	1	0	0								
265248	YUG	9	NA	0	0	0	NA									
265249	YUG	13	1	NA	1	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	...
265250	YUG	12	NA	0	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	NA	

注) AUSはオーストラリア、CANはカナダ、FRAはフランス、JPNは日本、NORはノルウェー、URYはウルグアイ、YUGは旧セルビア・モンテネグロ

出典：筆者作成

続いて、本研究で使用する 2 つの項目反応モデルを紹介する。テスト得点を算出するために使用される項目反応モデルは、一般に 1 母数モデル、2 母数モデルおよび 3 母数モデルがよく知られている（豊田、2002a、2012）。ちなみに PISA で採用されているモデルは、1 母数モデル（通称にならい、本稿ではラッシュモデルと呼ぶ）である（OECD、2005a、2006、2009）。その背景には、ラッシュモデルは項目母数が一つのみであり、テストを運用する上で取り扱いが容易である（豊田、2012b、p.25）ことが挙げられる。ところがラッシュモデルに含まれる項目母数は、その項目が有する困難度と呼ばれる難易度を表わす指標のみであり、「識別力」に関する母数は含まれていない。第 2 章では、困難度だけでなく識別力にも目を向けて各国の解答パターンを検討するため、困難度と識別力を項目母数に含む 2 母

数モデルも採用してテスト得点を算出する。また第3章と第4章では、困難度に焦点を絞り解答パターンを調べるため、ラッシュモデルを用いて受験者のテスト得点を計算する。なお2母数モデルとラッシュモデルの項目*i*における項目反応モデルは、

$$(1.1) \quad 2 \text{ 母数モデル: } P_i(\theta) = \frac{1}{1 + \exp(-Da_i(\theta - b_i))}$$

$$(1.2) \quad \text{ラッシュモデル: } P_i(\theta) = \frac{1}{1 + \exp(-D(\theta - b_i))}$$

とそれぞれ表現される。ただし $\theta$ は受験者の能力特性値（テスト得点）を表わし、 $p_i$ は項目*i*の正答確率を意味する。また $a_i$ と $b_i$ は、それぞれ識別力と困難度の母数であり、定数*D*は正規計量として扱うため 1.702 と設定する。なお項目反応理論を適用する場合には、使用するテストがテストの一次元性<sup>13</sup>を満たす必要がある。PISA は、データ収集の方法に重複テスト分冊法が使われていること分かるように、項目反応理論による分析を前提としている。そのため、テストの一次元性の条件は本来的に満たされている。

加えて第2章、第3章および第4章では、各国の解答パターンをより妥当性を確保して分析するために、項目母数に「等化」を施し、各国における困難度および識別力を比較可能な数値へと変換する。等化の方法について、例えば豊田（2012b）は、mean-sigma 法、mean-mean 法、Haebara 法、Stocking and Lord 法を紹介している。本研究では、芝（1991）でも取り上げられている mean-sigma 法を採用し、等化を行うこととする<sup>14</sup>。ここでは、その方法を簡単に述べておく。項目反応理論における等化は、能力特性値 $\theta$ を

$$(1.3) \quad \theta = k\theta + l$$

と線形変換を施しても、正答確率 $p_i$ は変化しないという性質を利用して行われる。線形変換（1.3）にある*k*と*l*は等化係数と呼ばれ、任意に設定できる。今、*j*受検者集団（*j*=1、2、…、*m*）の*i*番目の項目（*i*=1、2、…、*n*）の困難度を $b_{ij}$ とおき、等化の基準となる受験者集団における困難度を便宜上 $b_{i1}$ とする。2母数モデル（1.1）の場合 mean-sigma 法によれば、等化係数*k*と*l*は、

$$(1.4) \quad k = \frac{(b_{i1} \text{の標準偏差})}{(b_{ij} \text{の標準偏差})}, \\ l = (b_{i1} \text{の平均値}) - k(b_{ij} \text{の平均値})$$

と定義される。そして、これらを用いて各受験者集団の識別力 $a_{ij}$ と困難度 $b_{ij}$ は、

$$(1.5) \quad a_{ij}^* = \frac{a_{ij}}{k}, \quad b_{ij}^* = kb_{ij} + l$$

と基準となる受験者集団に等化される。なおラッシュモデル（1.2）では、識別力 $a_{ij}$ は全て

<sup>13</sup> テストの一次元性を満たすとは、使用するテストがただ一つの能力特性を測定することを意味している。その確認方法は、豊田（2002a、2002b）が詳しい。

<sup>14</sup> その詳細は、豊田（2012b、p.116）や服部（2011）を参照されたい。

の受験者集団で1.0に固定されており、困難度のみ等に等化を施すことになる。ゆえにラッシュモデルの場合、各受験者集団の困難度 $b_{ij}$ は、

$$(1.6) \quad {}^*b_{ij} = b_j + l_i$$

と基準となる受験者集団に等化される。本研究では、第2章において式(1.5)を用いた等化を行い、第3章と第4章で式(1.6)による等化を施すことになる。

最後に OECD (2005a, 2006, 2009) をみると、PISA における数学的リテラシーと質問紙および読解力に関する調査では、データ収集の方法、出題内容やその翻訳あるいはデータ処理の方法などについて国際基準が設定され厳密に吟味されている<sup>15</sup>。そのため、PISA の公開データを使用して行う国際比較分析に対して、一定の信頼性と妥当性が確保されていると判断できよう。とはいえ、個々の質問事項の内容や翻訳における表現などを特定の国の文脈と照らし合わせながら細かく検討すれば、PISA が提供するデータに対する批判的な見解は成り立って然るべきである。しかしある特定の国・地域において成り立つ見解に傾倒したのでは、参加国全体で共通の指標を構成して初めて浮かび上がる国際的な傾向を見逃す可能性は否定できない。本研究はある特定の国・地域ではなく、PISA が設定する国際比較可能なデータあるいは指標を通して、各国・地域の特徴を可能な限り浮かび上がらすことに主眼があることを強調しておきたい。

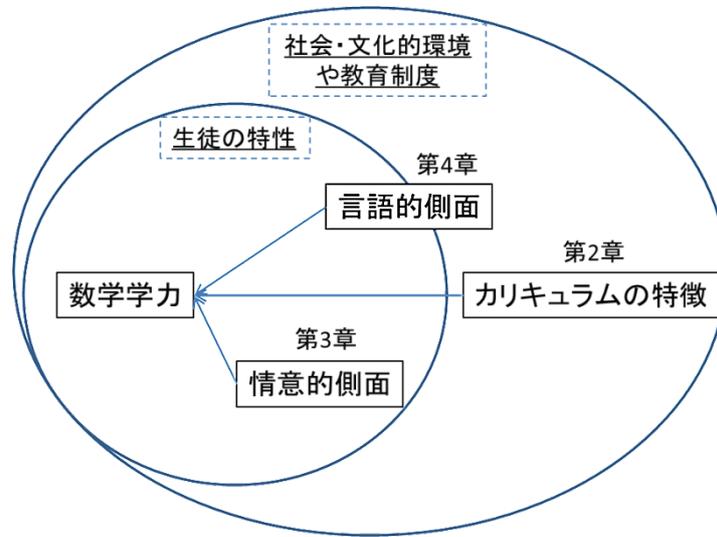
### 1.3 本研究の構成

第2章以降では、本研究の主題に対する計量的な分析を進める。それに向けてすでに述べたが、PISA や TIMSS といった国際比較が可能な教育調査の公開データ (raw data) を活用し、わが国および他の国の特徴を浮かび上がらせる数学教育研究は、これまで十分に蓄積されているとは言い難い。そのため各章では、まず国際比較を通じて日本を中心とした先進国の特徴を明らかにするとともに、途上国の特徴を捉えるための分析枠組みを事例的に示していく。そして、その分析枠組みを用いて途上国の特徴を浮かび上がらせるという手順を踏むことにする。なお、第2章以降で扱う研究課題の関係を示したものが図1.8である。

第2章では、2003年に実施されたPISA (以下、PISA2003) で収集された数学的リテラシー調査のデータを用いて、生徒の数学学力とカリキュラムの関連性について分析する。PISA2003のデータを使用する理由は、2006年と2009年に行われたPISA (以下、それぞれPISA2006とPISA2009) に比べて数学的リテラシーを測定するために作成された項目数が特に多く、出題項目の内容に触れた分析が可能だからである。まずその調査結果をみると、わが国のテスト全体における成績は、40参加国の中で最上位層に位置している。ところが「不確実性」と呼ばれる確率・統計に関する数学的な内容においては、その成績が

<sup>15</sup> その詳細は、PISA のテクニカル・レポート (OECD, 2005a, 2006, 2009) を参照されたい。

図 1.8 本研究の構成



出典：筆者作成

最上位層ではなく第2位層に留まっている（国立教育政策研究所、2004）。この背景について瀬沼（2006）は、2003年当時わが国の中学校数学科に確率・統計、特に統計に関する内容がほとんど含まれていなかったことを指摘している。もしそうであるならば、確率・統計を学習していないことに起因する特徴を掴む必要がある。つまりカリキュラムに含まれる数学的な内容の有無に着目しながら、数学学力とのカリキュラムの関連性を調べるのが求められる。そこで本章では、カリキュラムに含まれる数学的な内容、特にその有無に焦点を合わせながら分析を進めていく。

第3章では、第2章と同様に PISA2003 で収集されたデータの二次分析を通じて、生徒の数学学力と数学に対する情意的側面の関連性について検討する。なぜならば PISA2003 では、数学的リテラシー調査に加え、質問紙を用いて数学に対する情意的側面の調査が行われているためである。なお PISA2006 と PISA2009 では、数学に対する情意的側面に関する調査は行われていない。その具体的な調査内容は、数学に対する①興味・関心（内的な動機付け）、②道具的動機付け（外的な動機付け）、③自己効力感（数学への自信）、④自己概念（数学的な能力への信念）と⑤不安（数学に対する不安）の5つである。わが国のこれらの調査結果は、一貫して国際平均よりも低くなっており（国立教育政策研究所、2004）、今日では、生徒の数学に対する情意的側面へのアプローチが実践的な課題の一つに位置付いている。そして、情意的側面と数学学力との相互作用や因果関係を明らかにすること、およびその研究手法を確立することが数学教育研究に求められている（今井、2010、p.323；齋藤、2010、p.372）。そこで本章では、生徒の情意的側面を PISA が測定した5つの側面に代表させて、数学学力との関連性を国際比較していく。

第4章では、生徒の数学学力と言語的側面の関連性を分析していく。わが国では、この

研究課題が表立って数学教育研究で取り上げられることは少ない。その背景には、単一民族・言語であるというわが国の特徴が大きいと思われる。ところが国際的な視座に立てば、第二言語で数学を学習することは、数学教育研究の重要な課題として位置付いている。なぜならば言語は、思考、理解、コミュニケーションに関わり、数学の学習に重要な役割りを果たすからである。例えば Máire & John (2009) は、アイルランドを事例として挙げ、ゲール語能力が英語で数学を学習するのに大きく関係し、学年が高くなるにつれ、英語能力の役割りが大きくなると報告している。Jamal & Carol (2001) は、アメリカを事例にとして英語学習者 (English language learners) が数学のテストで大きくつまづく要因を、文章問題に焦点を当て、その長さや表現を変えながら調べている。それに加え、英語学習者の SES は決して高くないことに触れ、数学と言語的側面の関連性に関するさらなる研究の必要性を述べている。なお PISA では、2000 年以降継続して、数学的リテラシーと読解力および質問紙調査で使用される言語 (以下、テスト言語<sup>16</sup>) を家で話すかどうか、つまりその国における主要な教授言語の日常生活での使用に関する調査を行っている。そのためテスト言語の使用率は、言語的側面を反映させる指標として扱えると考えられる。したがって本章では、PISA2003 と PISA2006 および PISA2009 における数学的リテラシーと読解力の成績の経年変化を掴むとともに、読解力とテスト言語と言語的側面を代表させながら、数学学力との関連性を国際比較していく。そして第2章と第3章と同様に、出題項目数が他と比べて多い PISA2003 のデータを活用することで、出題項目に焦点を当てた分析に取り組み、数学学力と読解力の関連性をより詳細に調べることにする。

第5章においては、各章で得られた成果を整理し、本研究を総合的に考察する。そしてそれを踏まえ、今後の課題について整理する。

---

<sup>16</sup> PISA では、その国の主要な教授言語によってテストや質問紙が作成され、それを用いてデータ収集されている (OECD、2005a、2006、2009)。

付録 1.1 5つの情意的側面に関する質問項目

	因子負荷量	共通性
<b>興味・関心(内的な動機付け)</b>		
・ 数学についての本を読むのが好きである	0.765	0.585
・ 数学の授業が楽しみである	0.835	0.697
・ 数学を勉強しているのは楽しいからである	0.877	0.769
・ 数学で学ぶ内容に興味がある	0.781	0.610
<b>道具的動機付け(外的な動機付け)</b>		
・ 将来就きたい仕事に役立ちそうだから、数学は頑張る価値がある	0.805	0.648
・ 将来の仕事の可能性を広げてくれるから、数学は学びがいがある	0.829	0.687
・ 自分にとって数学が重要な科目なのは、これから勉強したいことに必要だからである	0.795	0.632
・ これから数学でたくさんのことを学んで、仕事に就くときに役立てたい	0.779	0.607
<b>自己効力感(数学への自信)</b>		
・ 列車の時刻表をみて、ある場所から別の場所までどのくらい時間がかかるか計算する	0.577	0.333
・ あるテレビが30%引きになったとして、それが元の値段よりいくら安くなったかを計算する	0.664	0.441
・ 床にタイルを張るには、何平方メートル分のタイルが必要かを計算する	0.716	0.513
・ 新聞に掲載されたグラフを理解する	0.579	0.335
・ $3x+5=17$ という方程式を解く	0.593	0.352
・ 縮尺10000分の1の地図上にある、2点間の距離を計算する	0.628	0.394
・ $2(x+3)=(x+3)(x-3)$ という方程式を解く	0.572	0.327
・ 自動車のガソリンの燃費を計算する	0.549	0.301
<b>自己概念(数学的な能力への信念)</b>		
・ 数学はまったく得意ではない	0.750	0.563
・ 数学では良い成績をとっている	0.764	0.584
・ 数学はすぐわかる	0.839	0.704
・ 数学は得意科目の一つだといつも思う	0.777	0.604
・ 数学の授業ではどんな難しい問題でも理解できる	0.731	0.534
<b>不安(数学に対する不安)</b>		
・ 数学の授業についていけないのではないかとよく心配する	0.698	0.487
・ 数学の宿題をやるとなると、とても気が重くなる	0.743	0.552
・ 数学の問題をやっているといらいらする	0.745	0.555
・ 数学の問題を解くとき、手も足も出ないと感じる	0.659	0.434
・ 数学でひどい成績をとるのではないかと心配になる	0.574	0.329

出典：国立教育政策研究所（2004）を基に筆者作成

付録 1.2 PISA2003 におけるデータ処理後の数学的リテラシーの受験者数および数学的リテラシーと読解力の共通受験者数

国名	受験者数	
	数学的リテラシー	読解力
1 オーストラリア	12551	6615
2 オーストリア	4568	2380
3 ベルギー	8549	4166
4 ブラジル	4452	2318
5 カナダ	27953	14135
6 スイス	8420	4229
7 チェコ	6192	3281
8 ドイツ	4552	2174
9 デンマーク	4218	2171
10 スペイン	10791	5813
11 フィンランド	5796	3110
12 フランス	4300	2211
13 ギリシャ	4627	2447
14 香港	4478	2336
15 ハンガリー	4371	2272
16 インドネシア	10761	5598
17 アイルランド	3880	2078
18 アイスランド	3350	1749
19 イタリア	11639	5905
20 日本	4707	2383
21 韓国	5444	2926
22 ルクセンブルク	3923	2012
23 ラトビア	4627	2418
24 マカオ	1250	
25 メキシコ	29983	15694
26 オランダ	3908	2003
27 ノルウェー	4064	2092
28 ニューージーランド	4511	2383
29 ポーランド	4383	2357
30 ポルトガル	4608	2406
31 ロシア	5974	3157
32 スロバキア	7288	3873
33 スウェーデン	4624	2313
34 タイ	5236	2783
35 チュニジア	4721	2445
36 トルコ	4855	2614
37 ウルグアイ	5835	3054
38 アメリカ	5456	2844
39 旧セルビア・モンテネグロ	4405	2327
合計	265250	137072

注) 空欄は、本研究で使用しないことデータを意味する。

出典：筆者作成

## 第2章 数学学力とカリキュラムの特徴の関連性

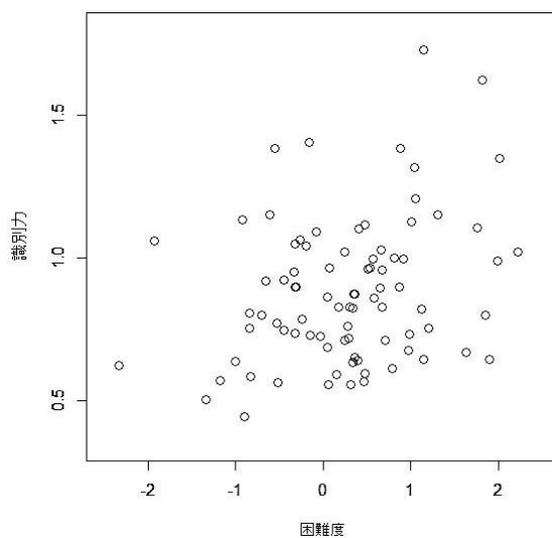
本章では、PISA2003 で収集されたデータの二次分析を通じて、生徒の数学学力とカリキュラムの特徴の関連性を日本を中心に検討し、そこで構築した分析枠組みを用いて、途上国における解答パターンを浮かび上がらせる。

### 2.1. PISA2003 における数学的リテラシー調査のテスト得点の国際比較

本小節では、PISA2003 に参加した 39 カ国の数学的リテラシー調査のテスト得点を国際比較し、先進国と途上国においてどの程度の数学学力の差異があるのかを確認する。本章では、テスト得点を算出するために使用する項目反応モデルとして、2 母数モデル (1.1) を採用する。しかし PISA の報告書 (e.g., OECD, 2004, 2007, 2010b) では、ラッシュモデル (1.2) による分析結果が報告されているため、このモデルでもテスト得点を算出し、2 つのモデル間におけるテスト得点の関係性を確認しておく。

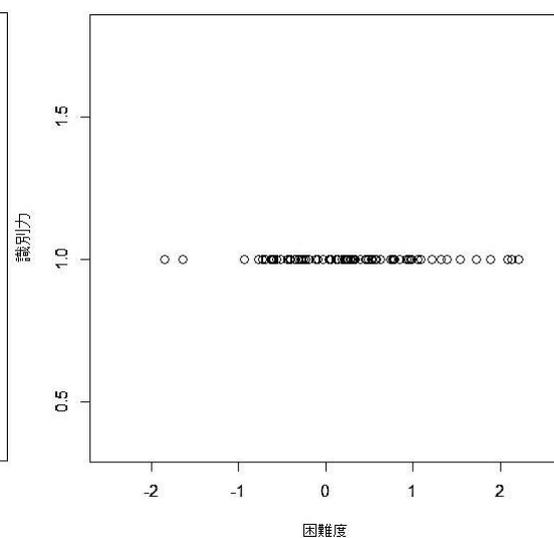
まず、図 1.7 で示したデータセットからテスト得点を算出する準備として、両モデルに含まれる項目母数に極端な値を示すものがないかを確認する。なお 2 母数モデルでは困難度と識別力について検討するが、ラッシュモデルでは困難度のみ注目する。そこでラッシュモデルと 2 母数モデルによる各項目の母数を図 2.1 と図 2.2 に散布図で示した。図 2.2 から分かるように、ラッシュモデルにおける各項目の識別力は 1.0 に固定されている。まずラッシュモデルによる各項目の困難度をみると、その最大値と最小値は、それぞれ 2.214 と -1.842 となった。次に図 2.1 に関して、2 母数モデルによる困難度と識別力の項目母数をみると、困難度の最大値と最小値はそれぞれ 2.221 と -2.330 となり、識別力については、

図 2.1 2 母数モデルの項目母数



出典：筆者作成

図 2.2 ラッシュモデルの項目母数

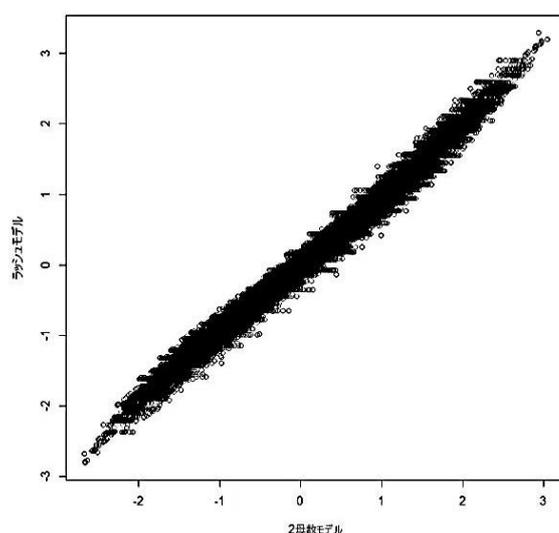


出典：筆者作成

最大値と最小値がそれぞれ 1.729 と 0.444 を得た。これらの値の評価について、芝 (1991) および豊田 (2002b) を参考にすれば、困難度の絶対値は 4.0 以下、識別力では 0.2 から 2.0 の間が望ましいとしている。この基準に従うならば、ラッシュモデルあるいは 2 母数モデルのどちらを用いてもテスト得点を算出できると考える。

そこで、数学的リテラシー調査のテスト得点をラッシュモデルと 2 母数モデルで求め、その散布図を図 2.3 に示した。この 2 つのモデルにおけるテスト得点の相関係数は 0.994 であり、どちらのモデルを採用しても支障がないことが分かる。後述するが本章では、困難度だけでなく、識別力にも注目した分析を行う。そのためここでは、2 母数モデルによって算出したテスト得点についてみていく。

図 2.3 2 母数モデルとラッシュモデルによるテスト得点の散布図



出典：筆者作成

39 カ国における数学的リテラシー調査のテスト得点の分布を平均値や標準偏差といった統計量を用いて表 2.1 に要約した。なお受験者全体で平均値が 0.0、標準偏差が 1.0 になるように調節してある。この表から分かるように、上位層には香港、オランダ、フィンランド、韓国や日本といった先進国が位置している。その一方で、下位層にはウルグアイ、メキシコ、ブラジル、チュニジアやインドネシアといった途上国が位置することを見て取れる。そこで、国別の違いが受験者のテスト得点をどれだけ説明するかを表わす級内相関係数 (interclass correlation coefficient、通称、ICC) を計算したところ、その値は 0.201 となった。このことは、265250 名の全受験者をわずかに 39 カ国に分けた場合、生徒のテスト得点のばらつきが約 20.1%説明されることを意味する。また最も高い平均値を持つ香港と逆に低いインドネシアを比較すると、各国の標準偏差が 1.0 より小さいことから、その差異 1.635 は、決して小さいとはいえないだろう。纏めるならば、PISA で測定される数学的リ

テラシー、つまり本研究で規定する数学学力の差異は、先進国と途上国という枠組みで決して小さくないことが分かる。

そこで本章では、日本を含むいくつかの先進国と数学的リテラシーの水準が下位層に位置する国の途上国に焦点を絞って国際比較分析を試みる。先進国としては、国立教育政策研究所が公刊する報告書（国立教育政策研究所、2004）や鈴川・豊田・川端（2008）で取り上げられているオーストラリア、カナダ、フィンランド、フランス、ドイツ、香港、アイルランド、イタリア、日本、韓国、オランダ、ニュージーランド、アメリカの13カ国に焦点を当てていく。また途上国については、表1.4と同様に表2.1でも下位層に位置してい

表2.1 2母数モデルによる各国のテスト得点（平均値の降順）

国名	平均値	標準誤差	中央値	標準偏差	標準誤差	最大値	最小値	受検者数
香港	0.640	0.014	0.742	0.933	0.010	3.042	-2.663	4478
オランダ	0.548	0.014	0.563	0.876	0.010	2.968	-2.103	3908
フィンランド	0.531	0.011	0.546	0.845	0.008	2.930	-2.403	5796
韓国	0.506	0.012	0.538	0.905	0.009	2.936	-2.531	5444
ベルギー	0.492	0.011	0.554	0.978	0.007	2.970	-2.550	8549
チェコ	0.480	0.012	0.524	0.944	0.008	3.042	-2.550	6192
日本	0.468	0.014	0.556	0.966	0.010	3.042	-2.663	4707
ニュージーランド	0.368	0.014	0.407	0.952	0.010	2.970	-2.578	4511
マカオ	0.361	0.026	0.414	0.903	0.018	2.796	-2.403	1250
オーストラリア	0.349	0.008	0.388	0.950	0.006	3.042	-2.663	12551
スイス	0.340	0.010	0.393	0.934	0.007	2.970	-2.660	8420
カナダ	0.332	0.005	0.360	0.884	0.004	2.970	-2.655	27953
フランス	0.285	0.014	0.351	0.906	0.010	2.752	-2.663	4300
デンマーク	0.282	0.014	0.341	0.912	0.010	3.042	-2.578	4218
アイスランド	0.279	0.016	0.341	0.905	0.011	2.936	-2.518	3350
オーストリア	0.269	0.013	0.286	0.890	0.009	2.970	-2.403	4568
ドイツ	0.253	0.014	0.300	0.935	0.010	2.970	-2.578	4552
スウェーデン	0.223	0.014	0.234	0.934	0.010	2.968	-2.663	4624
スロバキア	0.206	0.011	0.208	0.914	0.008	2.970	-2.498	7288
アイルランド	0.182	0.014	0.209	0.849	0.010	2.796	-2.498	3880
イタリア	0.119	0.009	0.161	0.926	0.006	2.970	-2.498	11639
ハンガリー	0.117	0.014	0.122	0.908	0.010	2.970	-2.431	4371
ノルウェー	0.113	0.014	0.132	0.923	0.010	2.970	-2.550	4064
スペイン	0.104	0.008	0.144	0.875	0.006	2.970	-2.660	10791
ルクセンブルク	0.097	0.014	0.114	0.896	0.010	2.624	-2.518	3923
ポーランド	0.054	0.014	0.057	0.914	0.010	2.796	-2.663	4383
ラトビア	0.032	0.013	0.056	0.886	0.009	2.901	-2.518	4627
アメリカ	-0.026	0.012	-0.012	0.906	0.009	3.042	-2.660	5456
ロシア	-0.073	0.012	-0.072	0.926	0.008	2.901	-2.550	5974
ポルトガル	-0.157	0.013	-0.176	0.883	0.009	2.437	-2.518	4608
ギリシャ	-0.348	0.013	-0.347	0.910	0.009	2.608	-2.578	4627
旧セ	-0.388	0.013	-0.414	0.846	0.009	2.674	-2.663	4405
トルコ	-0.478	0.014	-0.552	0.941	0.010	2.970	-2.663	4855
タイ	-0.525	0.012	-0.601	0.847	0.008	2.803	-2.663	5236
ウルグアイ	-0.536	0.012	-0.595	0.941	0.009	2.832	-2.547	5835
メキシコ	-0.651	0.004	-0.663	0.773	0.003	2.729	-2.663	29983
ブラジル	-0.980	0.012	-1.100	0.826	0.009	2.284	-2.660	4452
チュニジア	-0.990	0.011	-1.068	0.742	0.008	2.284	-2.657	4721
インドネシア	-0.995	0.007	-1.081	0.720	0.005	2.090	-2.663	10761
全体	0.000	0.002	0.002	1.000	0.001	3.042	-2.663	265250

注) 旧セは、旧セルビア・モンテネグロ

るトルコ、タイ、ウルグアイ、メキシコ、ブラジル、チュニジア、インドネシアの7カ国に注目しながら分析を進めることとする。

## 2.2. 日本に注目した13先進国に関する国際比較

前小節では、PISA2003に参加した39カ国の数学的リテラシー調査のテスト得点を国際比較した。その結果、先進国と途上国の間に看過できない数学学力の差異が存在することが明らかになった。本小節では、日本に焦点を当てながら先進国における数学学力とカリキュラムの特徴の関連性、特にカリキュラムに含まれる数学的な内容との関連性について検討していく。それと同時に、焦点を当てる7途上国の特徴を浮かび上がらすための分析枠組みを構築することを試みる。

PISA2003における日本のテスト得点の成績は、参加国の中でも最上位層にあることを確認した。その一方で、「不確実性」と呼ばれる確率・統計に関する領域では、その成績が国際的にみて第2位層に留まっている（国立教育政策研究所、2004）。この結果から日本の生徒は、この領域に関連する知識や技能を活用する能力が国際的にみて低いことを推察できるだろう。この主な要因について瀬沼（2006）は、2003年当時に中学校数学に確率・統計、特に統計に関する内容がほとんど含まれていなかったことを挙げている。もしそうであるならば、確率・統計を学習していないことに起因する特徴を可能な限り掴む必要があると考える。つまり、確率・統計に関する内容がほとんど扱われていなかったという日本のカリキュラム上の特徴に着目し、それが生徒の数学学力とどのように関連するのかを本小節では明らかにする。

一般に、PISAのような国際的な教育調査から各国の特徴を浮かび上がらせる場合には、テスト得点の平均値や出題項目の正答率を国際比較することが多い（e.g.、国立教育政策研究所、2004；鈴木、2006）。2.2小節で述べたようにこれらは、各国の生徒が有する数学学力の水準を特に反映する指標である。したがってこれらの値を比較するだけでは、各国の生徒が有する解答パターンの特徴を把握するには十分ではない。なぜならばテスト得点の平均値や正答率は、ある特定の受験者集団に依存して定められる性質があるため（芝、1991）、各項目が持つ難易度などの特性を客観的に比較した結果になりにくいからである。ゆえに本小節では、より客観的に解答パターンを捉えるために、受験者とテストおよび出題項目の特性を区別できる項目反応理論を用いる。

まず、本小節で焦点を当てる13カ国におけるテスト得点の分布を「不確実性」に限定して確認する。次に、出題項目の特性を困難度だけでなく識別力にも目を向けて国際比較し、日本の生徒が有する解答パターンの全体像を浮かび上がらせる。最後に、浮かび上がった解答パターンを確率・統計に関する領域に絞り、その特徴をカリキュラムと関連付けながら検討する。なお本章では、PISA2003の数学的リテラシー調査がテスト全体で測定する能力を「数学を活用する能力」とし、特にその能力の中で確率・統計に限定する場合には「確率・統計を活用する能力」と呼ぶことで区別する。加えてPISA2003に項目反応理論を適

用した研究としては、鈴川・豊田・川端（2008）が挙げられる。この研究では、ラッシュモデル（1.2）に含まれる困難度に注目して、解答パターンを国際比較し、日本の生徒は日常場面に近い状況で特に数学を活用する能力が国際的にみて低いことを指摘している。ところがテストを構成する項目は、困難度だけでなく識別力によっても特徴付けられる。そのため本章では、困難度だけでなく識別力も母数に含む2母数モデル（1.1）を用いて、日本の生徒が確率・統計を活用する能力を中心に分析していく。

### 2.2.1. 確率・統計を活用する能力の分布の国際比較

表 2.2 に PISA2003 の数学的リテラシー調査における「包括的アイデア」と呼ばれる数学的な内容の4つの評価枠組みとその項目数を示した。不確実性に含まれる項目は19項目であることから、確率・統計を活用する能力は、その19項目から成るテストを用いてテスト得点を算出する。

まず、確率・統計を活用する能力を表わすテスト得点を39カ国の全受験者に対して算出し、その後焦点を当てる13カ国におけるその分布を検討する。それに向けて、19項目で構成されるテストに極端な困難度および識別力を示す項目がないかを確認する。困難度における最大値と最小値は、それぞれ2.340と-1.432であった。また識別力においては、最大値が1.497となり最小値は0.499であった。芝（1991）と豊田（2002b）の基準でこれらの値を評価するならば、19項目全てを使用してテスト得点を算出することができる。そこで、全受験者における確率・統計を活用する能力の特性値の平均値を0.0、標準偏差を1.0に調節し、焦点を当てる13カ国におけるその分布を表2.3に示した。加えて39カ国における数学および確率・統計を活用する能力の平均値の散布図を図2.4に提示した。これらの結果から平均値の順位に多少の違いはみられるものの、この2つのテスト得点の分布が大きく異なる国はなく、一定の相関関係にあることが分かる。

表 2.2 包括的アイデア（数学的な内容）の各領域の内容と項目数

包括的アイデア	内容	項目数
空間と形	空間的、幾何的な現象や関係を扱う	21
不確実性	確率的、統計的な現象や関係を扱う	20
変化と関係	変数間の関数的な関係と依存関係とともに変化の数学的な関係を扱う	21
量	数量的な関係、数量的なパターン、数量的な現象を主に扱う	19

出典：国立教育政策研究所（2004）をもとに筆者作成

### 2.2.2. 項目母数の推定と等化

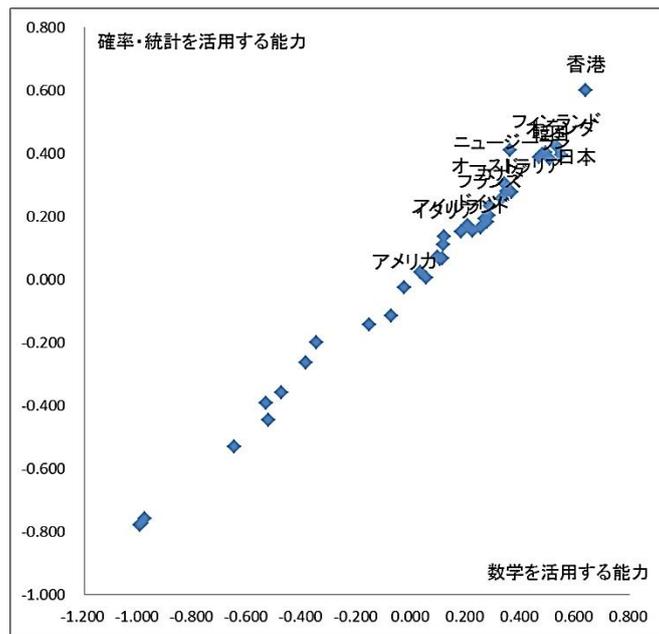
ここでは2母数モデル（1.1）が持つ母数である各項目の困難度と識別力を13カ国ごとに推定し、それらを比較していく。なぜならばこの2つの母数を国際比較することは、これまでみてきた各国の数学学力の水準を比較だけでなく、その内実を表わす各国の解答パターンを浮かび上がらすことに繋がるからである。しかしながら各国で推定される困難度

表 2.3 13 カ国における数学および確率・統計を活用する能力の分布

国名	数学を活用する能力				確率・統計を活用する能力				受検者数
	平均値	標準誤差	標準偏差	標準誤差	平均値	標準誤差	標準偏差	標準誤差	
オーストラリア	0.349	0.008	0.950	0.006	0.280	0.009	1.022	0.006	12551
カナダ	0.332	0.005	0.884	0.004	0.256	0.006	0.963	0.004	27953
ドイツ	0.253	0.014	0.935	0.010	0.163	0.014	0.959	0.010	4552
フィンランド	0.531	0.011	0.845	0.008	0.421	0.012	0.918	0.009	5796
フランス	0.285	0.014	0.906	0.010	0.232	0.015	0.978	0.011	4300
香港	0.640	0.014	0.933	0.010	0.599	0.015	1.003	0.011	4478
アイルランド	0.182	0.014	0.849	0.010	0.150	0.015	0.918	0.010	3880
イタリア	0.119	0.009	0.926	0.006	0.134	0.009	0.981	0.006	11639
日本	0.468	0.014	0.966	0.010	0.387	0.015	1.029	0.011	4707
韓国	0.506	0.012	0.905	0.009	0.382	0.014	1.010	0.010	5444
オランダ	0.548	0.014	0.876	0.010	0.395	0.015	0.950	0.011	3908
ニュージーランド	0.368	0.014	0.952	0.010	0.277	0.015	1.014	0.011	4511
アメリカ	-0.026	0.012	0.906	0.009	-0.027	0.013	0.964	0.009	5456

出典：筆者作成

図 2.4 39 カ国の数学および確率・統計を活用する能力の平均値の散布図



出典：筆者作成

と識別力は、「等化」を施すことによって、比較可能な数値に変換される。

では、項目母数の等化の前に、13 カ国で解答パターンを国際比較するための準備を行う。本研究では、等化の精度をより高めるために、13 カ国間で分析対象とする項目を統一しておく。数学的リテラシー調査では、84 項目で構成されるテストが使用されている。しかし OECD の分析の結果、13 カ国の中で韓国、イタリアおよびカナダで整合性が低いとされたものが 1 項目ずつ指摘されている (OECD、2005a、p.190)。そのため、この 3 つの項目は

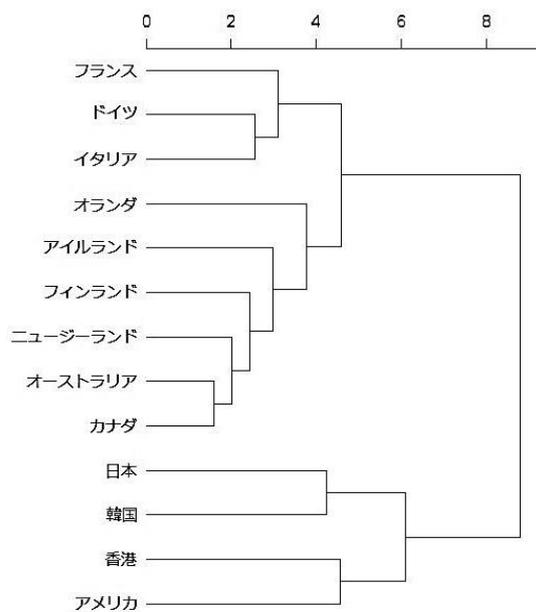


### 2.2.3. 13カ国の解答パターンの類似性と特異性

表 2.4 に示した各国の 80 項目の困難度と識別力を用いて、13 カ国における解答パターンの類似度を調べる。そのために、類似度または相違度をもとにいくつかのクラスターを作り分類する方法として知られる階層的クラスター分析<sup>18</sup>を適用する。なおこの分析のために、各国で困難度と識別力を標準化した。困難度においては、等化の段階で平均値と標準偏差が同一の値に揃えられるため特に必要ないが便宜上行った。

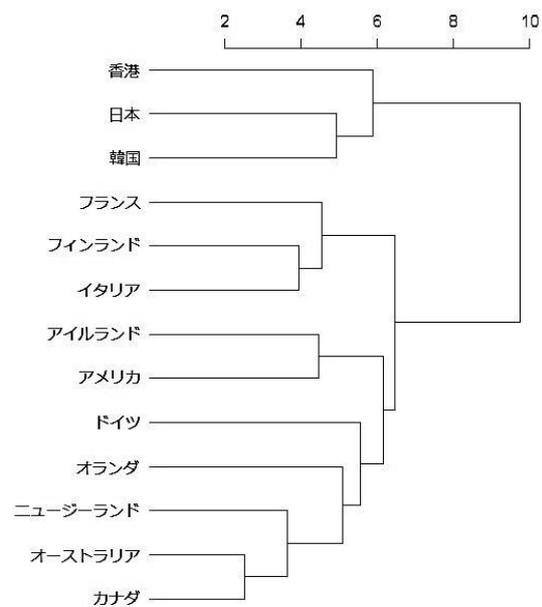
図 2.5 と図 2.6 に困難度と識別力におけるクラスター分析の結果であるテンドログラムをそれぞれ示した。この 2 つの分析結果において、それぞれ距離 7.0 付近を目安に 2 つの組に分けると、日本と韓国および香港は困難度と識別力ともに同じグループに属することが分かる。このことから困難度におけるアメリカを例外とすれば、解答パターンの類似度は、アジア諸国と欧米諸国に分類されるという結果を得た。

図 2.5 困難度のテンドログラム



出典：筆者作成

図 2.6 識別力のテンドログラム



出典：筆者作成

続いて、分析対象である 13 カ国全体の中で各国の解答パターンがどのくらい特異であるかを検討する。なぜならば解答パターンが類似しているとしても、それらの国の解答パターンがどれだけ全体的な傾向から離れているかを把握できないからである。そこで、それぞれの国が持つ各項目の困難度および識別力の値を使って、各項目の困難度と識別力の平均値を計算する。例えば、13 カ国における項目 1 の困難度と識別力の値から、項目 1 の困難度と識別力の平均値をそれぞれ算出する。このように各項目で求めた困難度と識別力の

<sup>18</sup> 本研究では、ウォード法を採用した。

平均値を各項目が持つ特性値として持つ国を構想すれば、この国は全ての項目で平均的な困難度と識別力を持つと仮定できる<sup>19</sup>。つまりこの平均国は、13 ヶ国における平均的な解答パターンを有すると想定でき、この国と比較することによって、各国がどのくらい特異な解答パターンを有するかを把握できる。

平均国との差異に着目した分析を行うために、13 ヶ国が各々持つ各項目の困難度と識別力から平均国のそれらの値を引き算し、それらの標準偏差を各国で計算する。例えば、日本における各項目の困難度と識別力の値から平均国のそれらの値を引き、それによって得られた各項目の新しい困難度と識別力に対して標準偏差を算出する。これを各国で行う。この値はある種の距離を表わすため、もし各国が平均的な解答パターンを持つならば、困難度と識別力は誤差の許す範囲で平均国に近くなるはずである。つまりこの値が大きいほど平均国との差異を表し、特異な解答パターンを持つ国と考えられる。この分析結果を表 2.5 に示した。この表から分かるように、困難度および識別力ともにアジア諸国における値が大きくなっている。ゆえに特異な解答パターンを持つ国は、アジア諸国であるという傾向を得た。このように類似性と特異性に着目した分析から、分析対象とした 13 ヶ国におけるアジア諸国は、解答パターンに類似性が認められるとともに、その実態は平均的な解答パターンとは異なるという実態が浮かび上がった。

表 2.5 平均国との差の標準偏差

困難度		識別力	
国名	標準偏差	国名	標準偏差
韓国	0.465	香港	0.551
日本	0.437	韓国	0.495
アメリカ	0.402	日本	0.483
香港	0.376	ドイツ	0.480
オランダ	0.320	アメリカ	0.477
フランス	0.286	オランダ	0.443
アイルランド	0.281	フィンランド	0.386
イタリア	0.267	フランス	0.385
ニュージーランド	0.256	アイルランド	0.359
ドイツ	0.220	ニュージーランド	0.356
フィンランド	0.197	イタリア	0.350
カナダ	0.178	オーストラリア	0.241
オーストラリア	0.169	カナダ	0.231

出典：筆者作成

#### 2.2.4. 日本の解答パターンの特徴

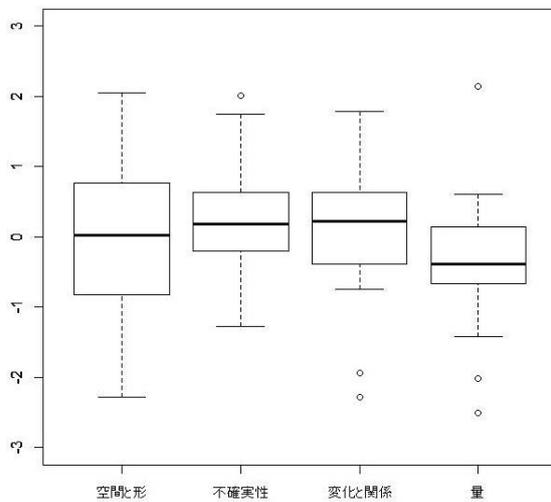
日本をはじめアジア地域にある韓国や香港の解答パターンは、国際的にみて類似性と特異性を併せ持つことが浮かび上がった。それでは、その特徴を PISA2003 の評価枠組みの「包括的アイディア」と呼ばれる 4 つの数学的な領域、具体的には「空間と形」、「不確実性」、「変化と関係」、「量」に注目しながらみていく。なお 13 ヶ国全てと比較することは本

<sup>19</sup> その詳細は、鈴木・豊田・川端（2008）を参照せよ。

質的ではないので、平均的な解答パターンを持つ平均国と比較する。

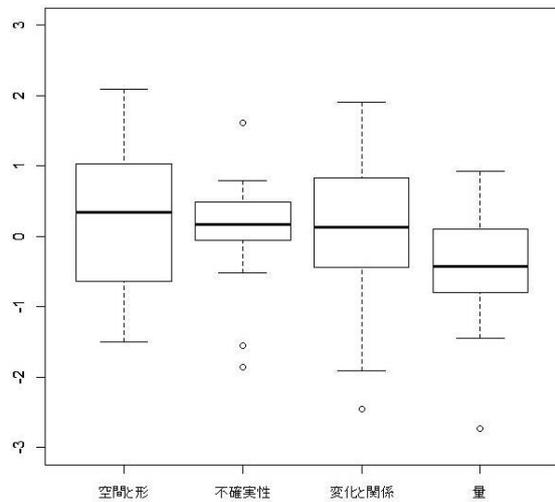
まず、日本と平均国における困難度の分布を図 2.7 および図 2.8 のように、「包括的アイディア」の 4 つの領域ごとに箱ひげ図に要約した。両国において各領域の中央値が大幅に異なる状況にはなっていない。ところがレンジに目を向けると、日本の「不確実性」の幅は明らかに平均国のものより大きい。これは、わが国の「不確実性」における困難度のばらつきが多様であることを意味する。次に、識別力の分布に目を向ける。困難度と同様に、両国における分布を図 2.9 と図 2.10 のように示した。中央値の位置は、日本と平均国ともに各領域で大差ないが、レンジにおいては「不確実性」および「変化と関係」で大きくな

図 2.7 日本の困難度の分布



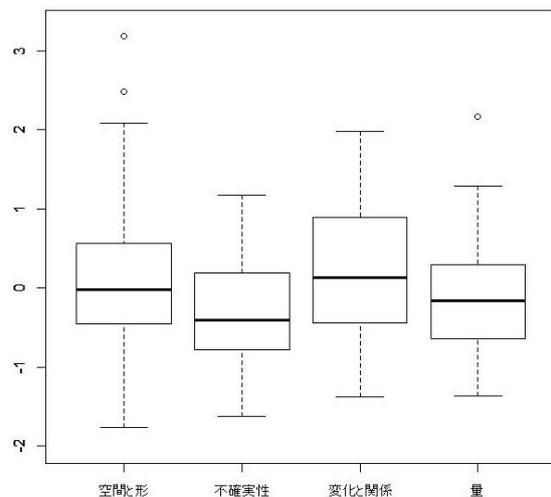
出典：筆者作成

図 2.8 平均国の困難度の分布



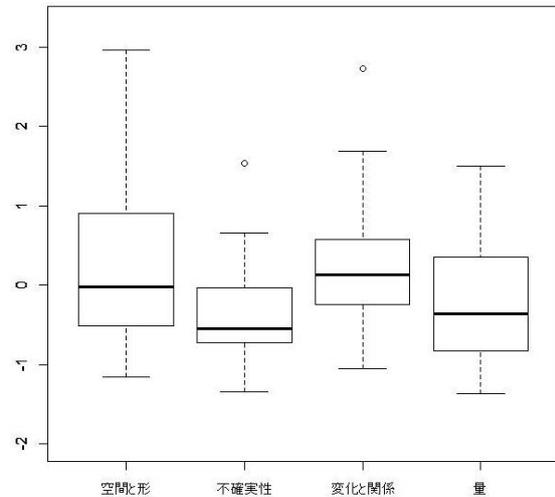
出典：筆者作成

図 2.9 日本の識別力の分布



出典：筆者作成

図 2.10 平均国の識別力の分布



出典：筆者作成

る傾向を見て取れる。

このように、出題項目の困難度と識別力の分布を「包括的アイデア」における4つの領域に着目して調べたところ、「不確実性」における困難度と識別力のばらつきは大きく、特に困難度では顕著であることが分かった。つまりわが国の解答パターンの特異性は、主に「不確実性」に依存するところが大きいといえ、特に困難度においてその傾向が際立っていることを確認した。

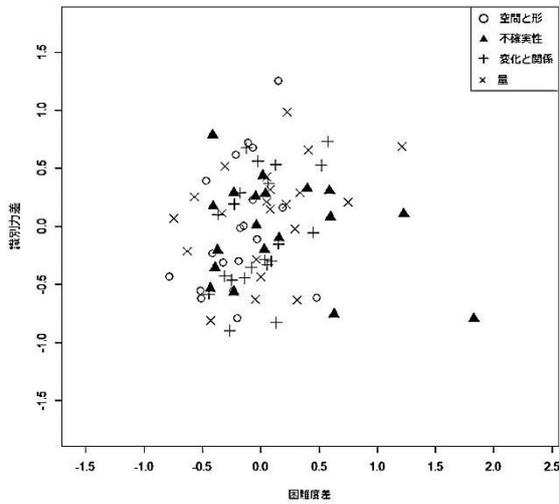
### 2.2.5. 「不確実性」における日本と平均国との差異の分布

これまでの分析から「不確実性」におけるわが国の生徒が有する解答パターンは、13カ国の平均的な解答パターンと比べて特異であることが分かった。特に困難度のばらつきは大きく、その難易度は低いものから高いものまで多様である。また識別力においても困難度ほどではないが、同様な傾向にあることが分かった。

では「不確実性」において、どのような内容の項目が平均的な解答パターンから特異なのであろうか。これを調べるために、日本における各項目の困難度と識別力の値から平均国のそれらの値を引き算し、それによって得られる値を利用する。以下、困難度における引き算の結果を困難度差とし、識別力におけるそれを識別力差と呼ぶ。この定義から正に大きい困難度差ほど、わが国が平均的な解答パターンよりも高い難易度を持つ項目を表し、負の値を示す場合にはその逆である。また識別力の本来の意味から識別力差が正に大きい場合は、日本ではその項目が持つ受験者を識別する機能が高く、負の場合には低いことを意味する。つまり日本において効果的に機能しやすいまたは機能しにくい項目を示すことになる。困難度差と識別力差を用いれば、各項目を平均的な解答パターンとの差異という視点から特徴付けることができ、日本の解答パターンを個々の項目から分析するのに役立つと考える。

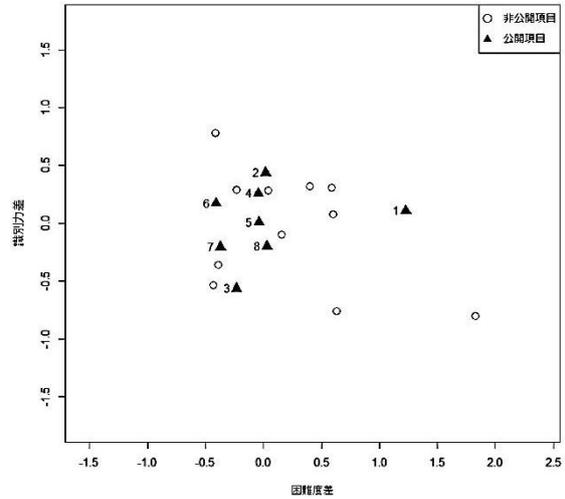
困難度差と識別力差を各項目で算出し、それを図2.11と図2.12のように散布図で表現した。図2.11は、全ての項目を「包括的アイデア」の4つの領域で分類して示しており、図2.12では、「不確実性」に属する項目で公開項目と非公開項目を示している。なお図2.12の公開問題に振った番号は、表2.6の番号欄の数字に対応している。図2.11から分かるように、「不確実性」における分布には、他の領域と比べて極端に正に大きい困難度差を持つ項目を見て取れる。ところが、多くの項目の困難度差は $-0.5$ から $0.5$ の範囲に収まっていることから、日本の生徒は、ある特定の内容において高い困難性を有していると考えられる。また識別力差においては、困難度差のような外れ値的な項目は見当たらず、一様なばらつき具合である。つまり日本の生徒が有する確率・統計を活用する能力は、13カ国における平均的な解答パターンから一様に特異なのではなく、ある特定の内容に高い困難性を抱えているという様相を読み取れる。加えて「不確実性」に属する公開問題の分布は図2.12に示したように、極端な困難度差を持つ項目を含んでおり、識別力差については適度なばらつきにある。このことから公開問題を参考にすれば、「不確実性」における分布をある程

図 2.11 日本の困難度差と識別力差の散布図 (80 項目)



出典：筆者作成

図 2.12 日本の困難度差と識別力差の散布図 (不確実性の 19 項目)



出典：筆者作成

表 2.6 「不確実性」における公開項目の困難度差と識別力差

番号	項目番号	ユニット	困難度差	識別力差	内容
1	M438Q01	輸出1	1.229	0.107	グラフを正しく読む
2	M438Q02	輸出2	0.019	0.438	グラフにある情報を用いての割合
3	M467Q01	いろいろな色のキャンディ	-0.231	-0.564	確率の算出
4	M468Q01T	理科のテスト	-0.043	0.259	平均値の算出
5	M505Q01	ゴミ	-0.038	0.008	棒グラフの適応条件について
6	M509Q01	地震	-0.408	0.172	確率の実現的な意味について
7	M513Q01	テストの点数	-0.370	-0.206	テスト得点の分布の比較
8	M702Q01	大統領の支持率	0.031	-0.199	世論調査の標本抽出について

出典：筆者作成

度項目の内容という視点から考察できると考える。

### 2.2.6. 日本における数学学力とカリキュラムの関連性の総合的考察

まず、クラスター分析と標準偏差を用いて得られた 13 カ国の類似性および特異性の結果 (図 2.5、図 2.6 と表 2.5 参照) のように、13 カ国の類似性および特異性がアジア諸国と欧米諸国に分かれて顕在化した理由について考える。PISA の出題項目は、学校教育の中で習得される一定の知識や技能を超えた範囲まで含めて評価するために、特定のカリキュラムに依らない教科横断的な観点から作成されている (国立教育政策研究所、2004)。また国際比較を意図した調査であるため、データ収集の方法や出題項目の内容およびその翻訳については、国際基準を設定し厳密に吟味されている。そのため、各国における受験者集団や各項目の信頼性および妥当性に著しい欠点があるとは思えない。出題内容は横断的であり受験者や各項目に際立った不備がないため、これまで浮かび上がった解答パターンの差異

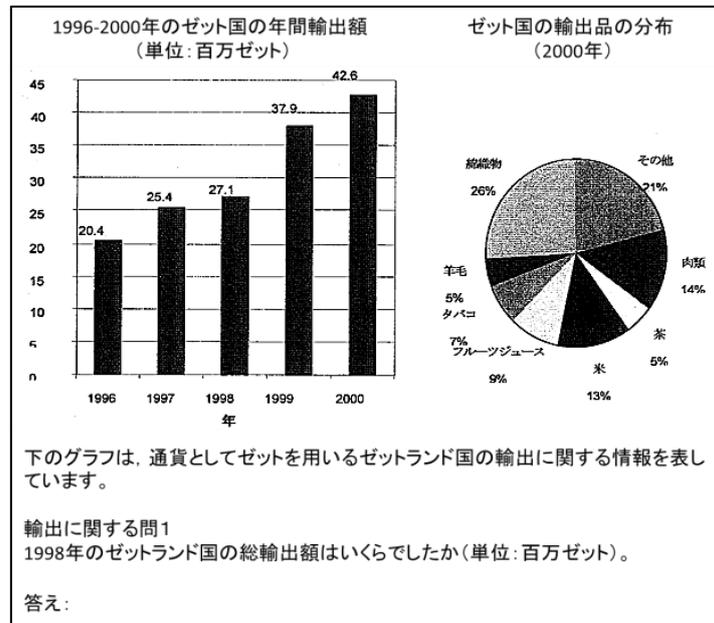
は、各国の数学を活用する能力の特徴を反映するものとする。これを数学教育の目標という立場から眺めると、日本や韓国および台湾では、数学的思考や問題解決能力および創造性などの陶冶的目標に限定されるが、欧米諸国では陶冶的目標よりもむしろ市民社会の形成や国造りに必要な人材といった社会的目標が充実しているといわれている（岩崎・長崎・モーモーニェン、2004）。このことから、日本や韓国あるいは香港の教育目標の基盤にある学力観は、欧米諸国のものとは基本的に異なっていると解釈できる。教育目標に違いがみられればその方法は異とするものになり、生徒が習得する知識や技能の実態も相違するだろう。このように考えると、日本や韓国あるいは香港においては測定内容にいささかそぐわない側面があるとも解釈できよう。このような測定内容の不一致は、異なる識別力をもたらすとされる（田崎、2008）。クラスター分析の結果において、困難度よりも識別力でアジア諸国と欧米諸国が明確にグループ分けされたことは、識別力に注目した分析から新たな示唆を得ることが期待される。

次に、PISAの目的と出題内容およびそれらの国際的な調査としての妥当性と信頼性を考慮すれば、平均国を基準として算出した「不確実性」に属する各項目の困難度差と識別力差は、本小節で焦点を当てている日本の生徒が持つ確率・統計を活用する能力の特徴を示唆すると考える。困難度差に注目するならば、日本と平均国との間で難易度が異なる項目を特定できる。また識別力差に目を向けるならば、日本の生徒が持つ確率・統計を活用する能力を測定するために、どの項目が効果的に機能しているかまたは機能していないかを検討できる。

では「不確実性」に属する公開項目を参考にしながら、困難度差および識別力差を考察する。まず、困難度差からみていく。「不確実性」に属する項目の中には、困難度差が正に極端に大きいものがいくつかあり、その中で公開されているものは、図2.13に示した「輸出1」である。この項目は、与えられた2つのグラフが持つ情報を把握し、質問に合うグラフから指定されたデータを読み取る内容となっている。この種の情報を読み取り理解することは、確率・統計を活用する能力の本質的な要素である（国立教育政策研究所、2004、p.91）とされるが、わが国の生徒は高い困難性を抱えていることが分かった。ところが、全く同じグラフを用いて解答する「輸出2（図2.14）」の困難度差は、ほぼ0.0という結果となっている。この項目で正答を得るには、「輸出1」と同じように2つのグラフが持つ情報を整理した上で、適切な数値を読み取り必要な計算を行わなければならない。「輸出1」では、グラフを正確に読み取るだけが正答を得る過程であったのに対して、「輸出2」においてはグラフを正確に読み取った後に数学的な計算を行うという過程が加わっている。したがって、「輸出1」の困難度差は正に大きく「輸出2」ではほぼ0.0であることから、日本の生徒は、グラフを正確に読み取る段階では高い困難性を有するが、読み取った後に行う数学的な計算は難無くこなすという実態が浮かび上がった。逆に平均国においては、グラフを読み取ることよりも数学的な計算に比較的高い困難性を有するとも解釈できる。またその他の公開問題の中には、数学的な計算を必要としないが困難度差は負の値を示してい

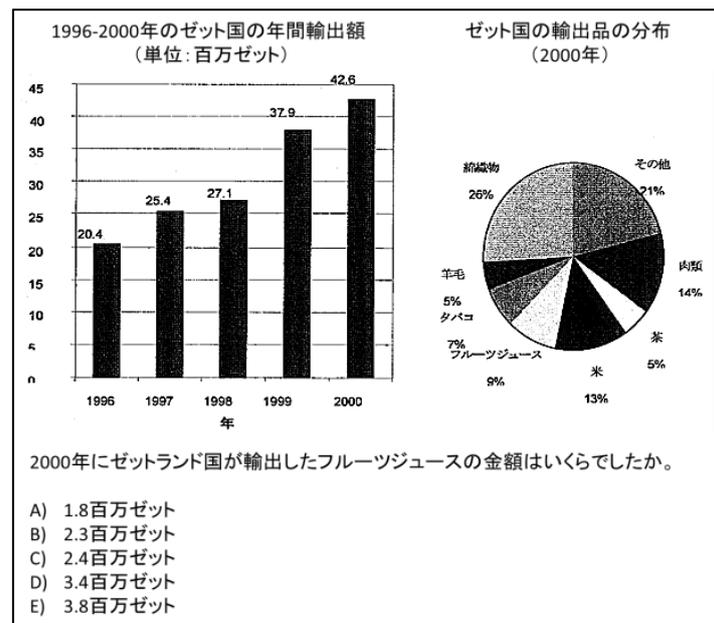
るものがある。このことから、グラフを正しく読むことに対する生徒が持つ困難性は本分析以上に大きい可能性を指摘できるだろう。

図 2.13 「輸出 1」



出典：国立教育政策研究所（2004）

図 2.14 「輸出 2」

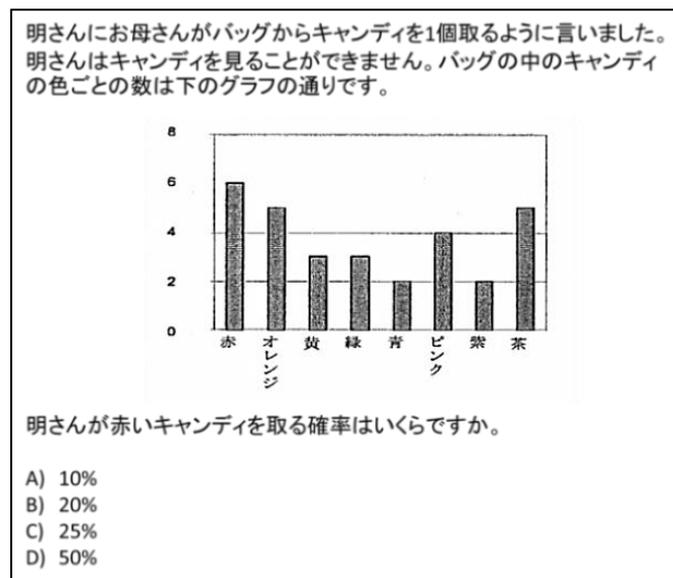


出典：国立教育政策研究所（2004）

続いて、識別力差についてみていく。公開問題の中で識別力差が負の値を示した項目は、表 2.6 から分かるように「テストの点数」、「大統領の支持率」および「いろいろな色のキャンディ」の 3 つである。これらの項目は日本の文脈において、受験者の能力特性の測定に対して効果的に機能しない傾向にあることを示唆している。

では、この 3 つの項目の内容を確認しながらその特徴を調べることにする。まず最も負に大きな値を示したのは、図 2.15 に示した「いろいろな色のキャンディ」である。これは、グラフにある値を用いて指定された事象の確率を求める内容である。グラフから数値を読み取りそれに必要な計算を行う点で類似する「輸出 1」および「輸出 2」では、識別力差については正の値を示し、わが国において識別力の低下は確認されていない。また確率に関する項目として図 2.16 に示した「地震」が公開問題となっている。この項目は確率を求めるものではなく、与えられた確率に対して妥当性の高い解釈を問う内容となっている。このような確率の意味理解を測る項目においても、識別力差は負の値を示していない。つまり「いろいろな色のキャンディ」におけるわが国の識別力の低下を引き起こす要因は、グラフを読み取ることや確率概念の把握に関する側面ではなく、具体的に確率を求める段階に隠れていると思われる。次に、「テストの点数」と「大統領の支持率」の内容を図 2.17 と図 2.18 に示した。前者の項目は、与えられたデータに外れ値が含まれるが、分布の本質的な部分を読み取り説明する内容である。外れ値は、平均値など日常でよく目にする代表的な統計量に影響を及ぼすことが知られている。そのため外れ値に対する知識や対処法は、本来のデータの実態を把握する上で重要な役割を果たす。他方で「大統領の支持率」は、

図 2.15 「いろいろな色のキャンディ」



出典：国立教育政策研究所（2004）

図 2.16 「地震」

地震と地震の頻度についてドキュメンタリー番組が放送されました。番組では地震を予知できるかどうかについても議論が交わされました。

ある地質学者は次のように言いました。「今後20年以内にゼットランド市で地震が起きる確率は3分の2だ」

この地質学者の言葉の意味を一番よく反映しているのは次のどれですか。

A)  $\frac{2}{3} \times 20 = 13.3$ 。だから、いまから13年から14年の間にゼットランド市では地震が起きる。

B)  $\frac{2}{3}$ は $\frac{1}{2}$ より大きい。だから、今後20年の間にゼットランド市ではいつか必ず地震が起きる。

C) 今後20年の間にゼットランド市で地震が起きる確率は、地震が起きない確率より大きい。

D) 地震がいつ起きるかはだれも確信できないので、何が起きるかを予言することはできない。

出典：国立教育政策研究所（2004）

図 2.17 「テストの点数」

下のグラフは、二つの班AとBの理科のテスト結果を示しています。A班の平均点は62.0、B班の平均点は64.5です。50点以下をとった生徒は合格になります。

**理科**

点数	A班	B班
0-9	1	0
10-19	0	0
20-29	0	0
30-39	0	0
40-49	2	2
50-59	4	1
60-69	2	5
70-79	2	3
80-89	1	1
90-100	0	0

■ A班 □ B班

先生はこのグラフを見て、今回のテストでは、B班のほうがA班より良かったと言いました。

A班の生徒たちは先生の意見に納得できません。A班の生徒たちは、B班のほうが必ずしも良かったとは言えないということを先生に納得させようとしています。

グラフを使い、A班の生徒が主張できる数学的な理由を一つ挙げてください。

出典：国立教育政策研究所（2004）

示された支持率に信頼性があるかを標本抽出の観点から問うものである。標本抽出の方法に妥当性がなければ、そこから得られる代表値に信頼性は生まれない。つまりこの2つの項目が測定している内容は、確率・統計を活用する能力において不可欠な要素を含んでい

図 2.18 「大統領の支持率」

ゼットランド国で、次回の選挙での大統領の支持率を調べるために世論調査が行われました。四つの新聞社が別々に全国的な世論調査を実施しました。四つの新聞の世論調査の結果は次の通りでした。

新聞1： 36.5%（投票権をもつ市民500人を無作為に選び、1月6日に世論調査を実施した）。

新聞2： 41.0%（投票権をもつ市民500人を無作為に選び、1月20日に世論調査を実施した）。

新聞3： 39.0%（投票権をもつ市民1000人を無作為に選び、1月20日に世論調査を実施した）。

新聞4： 44.5%（「電話投票」に応じた1000人の読者に1月20日に世論調査を実施した）

選挙が1月25日に行われるとすると、大統領の支持率を最も正確に予測しているのはどの新聞の結果ですか。答の説明として理由を二つ挙げてください。

出典：国立教育政策研究所（2004）

る。しかしながらわが国においては、これらの項目は確率・統計を活用する能力を測定する文脈で、効果的に機能しない傾向にあることが浮かび上がった。

そこで 2003 年当時の日本の中学校数学科を振り返ると、2000 年の学習指導要領改訂の影響で、統計に関する内容を直接扱うことはほとんどなく、また確率については第 2 学年における数量関係の中で、樹形図や同様に確からしい事象の起こり得る確率の概念および単純な事象の確率を求めることを取り上げるに留まっている（深澤・竹内・二宮、2007）。このことから確率・統計に関する領域は、ほとんど取り扱われない状況にあったといえる。

困難度差に注目した分析では、生徒は特にグラフを正確に読み取る段階に高い困難性を有すると述べた。統計的なグラフについては、日本では、小学校の算数で取り上げられる学習内容<sup>20</sup>であり、生徒にとって「輸出 1」のような質問は初めて目にするものではない。しかし中学校数学科では扱われなくなるため、このようなグラフに触れる機会は目立って減るものと予想される。つまり日本の生徒は、グラフから何かを読み取ることは広く認識

<sup>20</sup> 確率・統計教育に関する主な内容については、以下の表を参照いただきたい。

	数と計算	量と測定	図形	数量関係
第1学年	加法、減法、個数の処理、数直線	長さの測定、身近な単位	(該当項目なし)	(該当項目なし)
第2学年	乗法、事柄の分類性、表やグラフによる表現	時刻	直線、三角形、四角形	(該当項目なし)
第3学年	除法、そろばん	単位や計器の選択、時間	正方形、長方形、直角三角形、直角	表やグラフを用いた資料の表現、棒グラフ
第4学年	小数、分数、概数	面積、角度	二等辺三角形、正三角形、円	折れ線グラフ、グラフの特徴や傾向把握、
第5学年	偶数、奇数、記数法	平面図形の面積	平行、垂直、平行四辺形、台形、ひ	百分率(%)、歩合、円グラフ、帯グラフ
第6学年	約数、倍数	体積、速さ	立体図形	比、比例、平均

出典：深澤・竹内・二宮（2007）

しているといえるが、その習得が不確かであると考えられる。したがって困難度差が導く分析結果は、生徒には広く認識されている一方で、習得があいまいな内容を含む項目に対する生徒の解答パターンを反映したものと考えられる。

他方で識別力差における分析では、外れ値を含むデータの本質的な部分を読み取ることや、示された統計量の信頼性を保証する標本抽出の方法についての項目で、日本ではその項目の測定に対する機能が国際的にみて低いという結果を得た。中学校数学科では、統計に関する内容は直接扱うことはないため、外れ値や標本抽出について学習する機会は非常に少ない。これは小学校の算数においても同様である。ところが多くの欧米諸国では、データ収集やその分類などデータを中心に据えた統計教育が充実していると報告されている（深澤・竹内・二宮、2007）。つまり本研究で取り上げた「テストの点数」や「大統領の支持率」に含まれる統計的な要素は、それらの国では広く認識されていると考えられるが、日本では、それらが欧米諸国と同様に一般的であるとは言い難い。ゆえに識別力差に着目して得られた結果は、日本の数学教育の文脈であまり認識されていない概念や考え方を含む項目に対する生徒の解答パターンが顕在化したものと考えられる。このように本研究では、識別力に注目したことで、出題内容に対する習得の度合いだけでなく、それへの認識の度合いという新しい観点で生徒の解答パターンを分析できたといえる。

### 2.2.7. 日本以外の12ヵ国の平均国との差異

本小節では、平均国との比較を通じて日本の特徴を検討し、その結果の一つとして、表2.5に標準偏差を用いた各国の特異性を表わす指標を示した。ここでは、各国個別の考察は行わないが、標準偏差の違いをより理解するために、各国の困難度差と識別力差の散布図（図2.19から図2.30参照）を示しておく。

例えば韓国をみると、表2.5にある困難度と識別力の標準偏差は、それぞれ0.465と0.495と大きな値を示した。そこで図2.27に目を移すと韓国においても、不確実性における困難度差と識別力差にばらつきが目立つことを読み取れる。他方で標準偏差が小さいオーストラリアやカナダをみると、当然ながら困難度差と識別力差のばらつきは小さいことが分かる（図2.19と図2.20参照）。

これまで、標準偏差を指標として困難度と識別力を別々に検討してきた。ところがこれらの図のように平面上で考えるならば、困難度差と識別力差を合わせて平均国との差異を調べることができる。なぜならば、各図の原点は平均国の解答パターンを意味するため、各点と原点との距離は、困難度差と識別力差の2つの視点からみた全体的な差異の指標になるからである。そこで各項目と原点の距離を各国で求め、それらが属する「包括的アイデア」の4つの領域ごとに平均値を算出した。表2.7にその分析結果を示し、それを図2.31に折れ線グラフで表現した。なおここでは、外れ値とも取れる項目を含めて全体像を把握するために平均値を採用している。図2.31において、韓国は「不確実性」における値が突出している。このことは、図2.27から視覚的に捉えても読み取れる。ゆえに表2.7お

図 2.19 オーストラリア

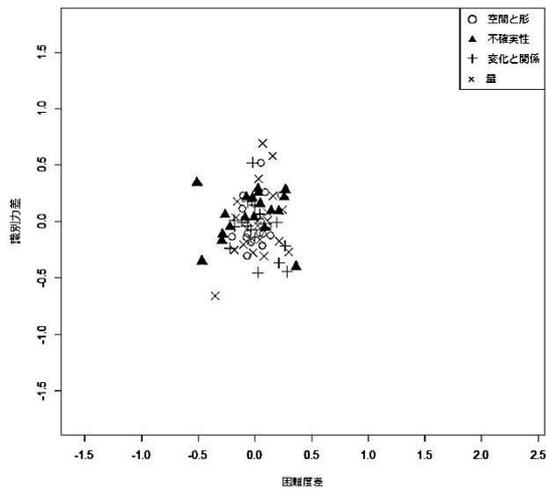


図 2.20 カナダ

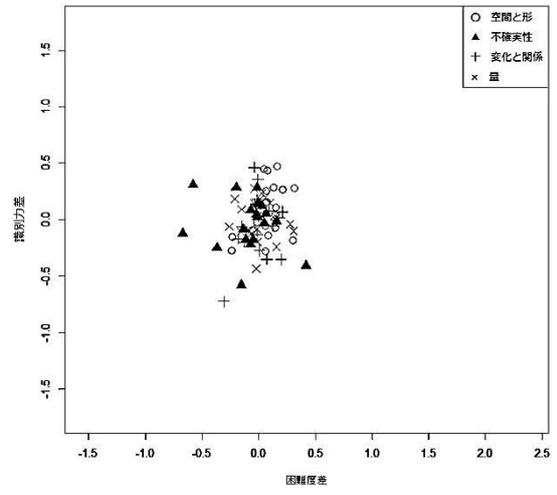


図 2.21 ドイツ

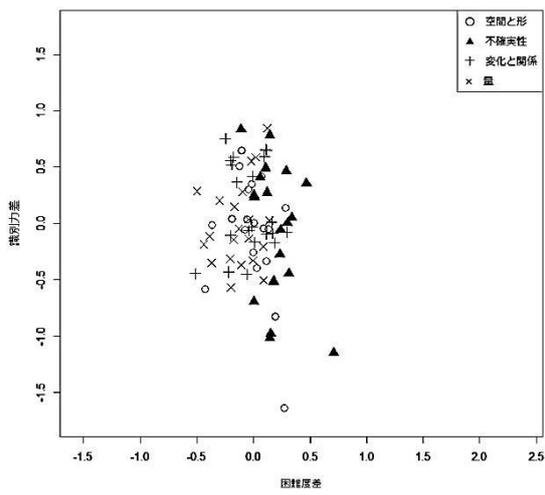


図 2.22 フィンランド

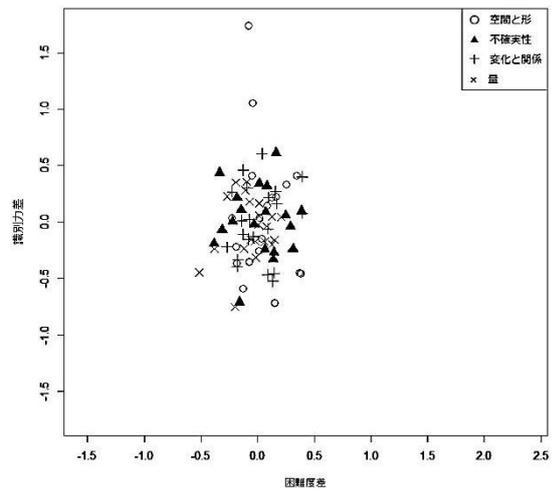


図 2.23 フランス

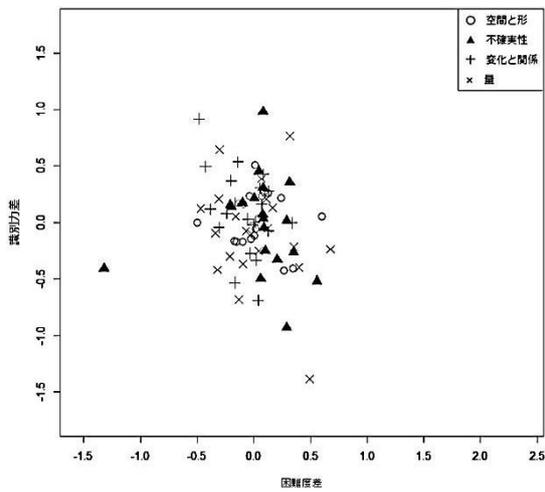


図 2.24 香港

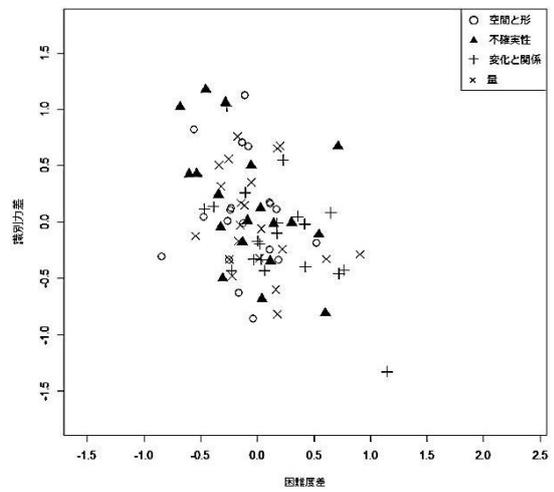


図 2.25 アイルランド

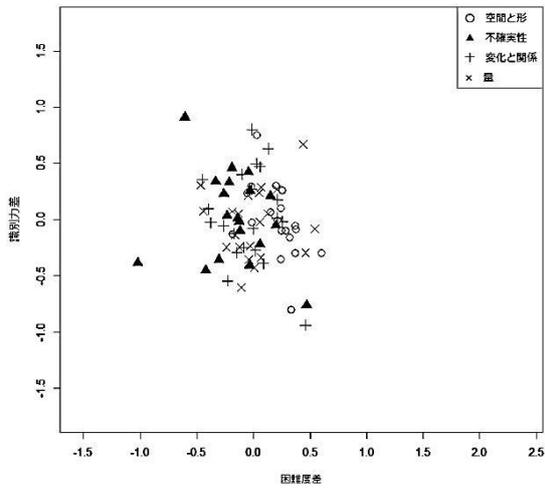


図 2.26 イタリア

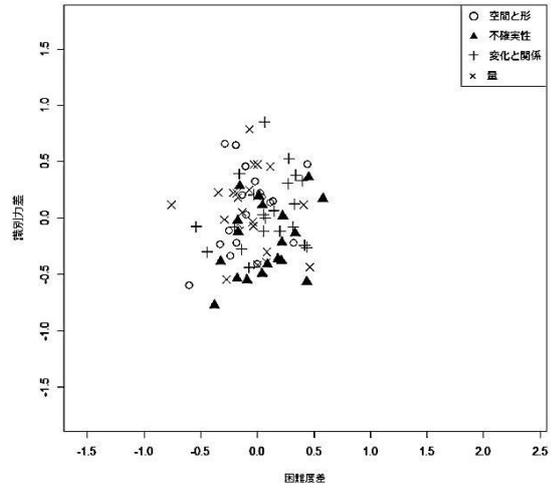


図 2.27 韓国

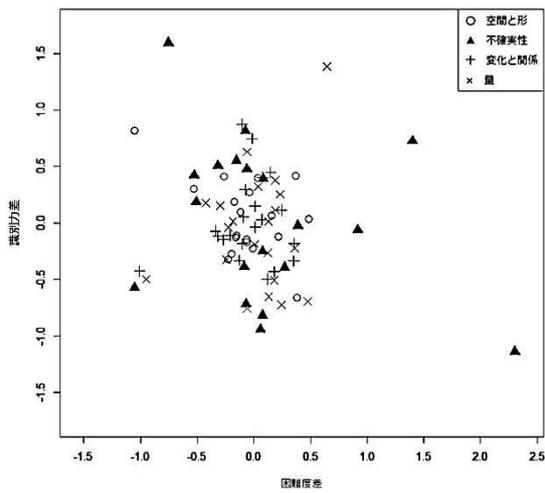


図 2.28 オランダ

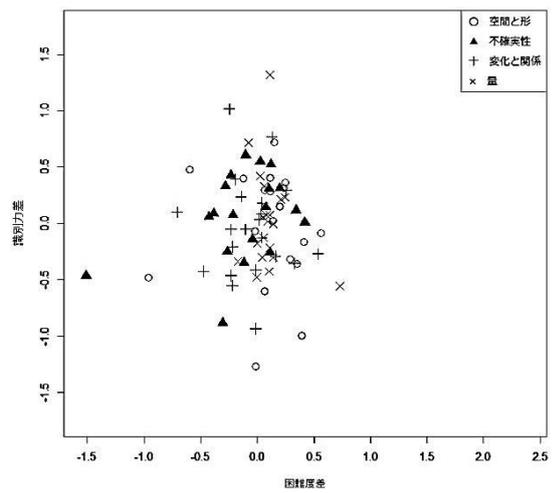


図 2.29 ニュージーランド

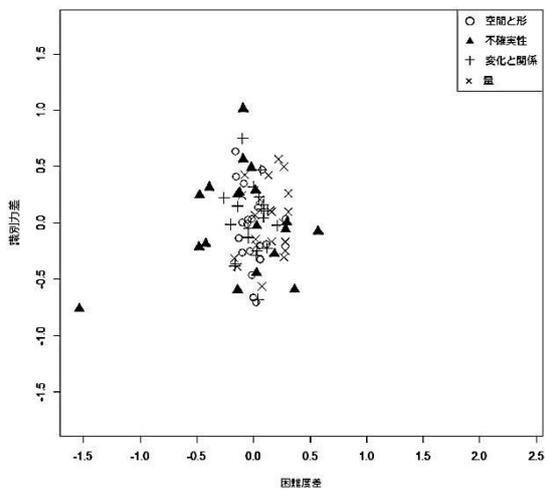


図 2.30 アメリカ

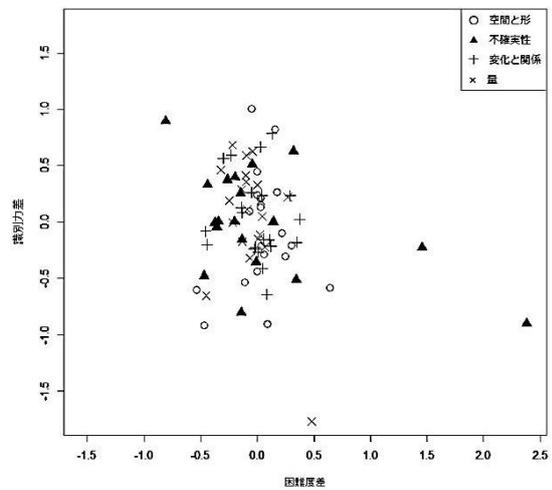
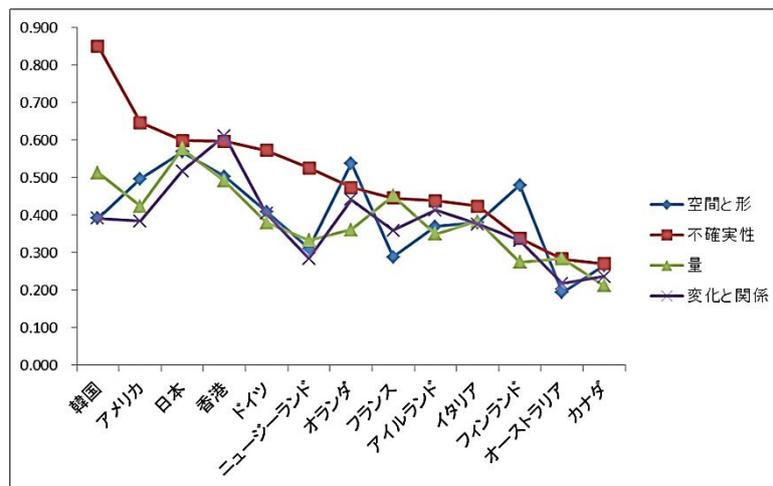


表 2.7 4つの領域における平均国との距離の平均値（不確実性の降順）

国名	空間と形	不確実性	量	変化と関係	最大値	最小値	レンジ
韓国	0.391	0.850	0.513	0.391	0.850	0.391	0.459
アメリカ	0.496	0.647	0.424	0.384	0.647	0.384	0.262
日本	0.568	0.598	0.576	0.517	0.598	0.517	0.081
香港	0.503	0.596	0.491	0.610	0.610	0.491	0.120
ドイツ	0.409	0.572	0.380	0.403	0.572	0.380	0.193
ニュージーランド	0.311	0.525	0.333	0.284	0.525	0.284	0.241
オランダ	0.537	0.473	0.361	0.441	0.537	0.361	0.177
フランス	0.288	0.445	0.452	0.359	0.452	0.288	0.164
アイルランド	0.370	0.438	0.350	0.413	0.438	0.350	0.088
イタリア	0.380	0.424	0.384	0.376	0.424	0.376	0.048
フィンランド	0.479	0.338	0.275	0.332	0.479	0.275	0.205
オーストラリア	0.194	0.284	0.284	0.217	0.284	0.194	0.090
カナダ	0.266	0.271	0.212	0.236	0.271	0.212	0.059
最大値	0.568	0.850	0.576	0.610			
最小値	0.194	0.271	0.212	0.217			
レンジ	0.374	0.579	0.364	0.393			

出典：筆者作成

図 2.31 4つの領域における平均国との距離の平均値の折れ線グラフ（不確実性の降順）



出典：筆者作成

よび図 2.31 は、各国の困難度差と識別力差の実態を一定程度反映する分析結果と考える。したがって図 2.31 から浮かび上がる傾向の一つとして、フィンランドを例外とすれば、「不確実性」において平均国との差異が現れやすい、つまり「不確実性」に各国の解答パターンが顕在化しやすいと解釈できるだろう。また図 2.22 をみるとフィンランドにおいて、識別力が特に高い「空間と形」に属する項目を 2 つ確認できる。ゆえにフィンランドでは、他の国ではあまり認識されていないような空間的・幾何的な現象や関係を扱う内容を数学教育で扱っている可能性を否定はできないだろう。フィンランドのみに拘らず他の国においても、特徴的な機能を有すると考えられる項目が浮かび上がりつつある。そのため、そ

れらを詳細な検討が今後の課題として位置付くと考える。とはいえ全体的な傾向として、すでに述べたように日本だけでなく、「不確実性」と呼ばれる確率・統計に関連する内容に各国固有の特徴が反映されやすいことを主張しておきたい。

### 2.3. 7 途上国に関する国際比較

これまで日本においては、2003年当時の中学校数学科で確率・統計に関する内容がほとんど含まれていなかったというカリキュラム上の特徴を踏まえ、それが生徒の数学学力にどのように関連するかを困難度および識別力を用いた解答パターンの分析から調べてきた。

本小節では、表 2.1 で示したように最下位層に数学的リテラシー調査の成績が位置付く 7 ヶ国、トルコ、タイ、ウルグアイ、メキシコ、ブラジル、チュニジア、インドネシアに焦点を当てて、これらの国の解答パターンを浮かび上がらせる。なお日本を中心とした分析では、困難度と識別力を用いて解答パターンを捉え、それを平均国と比較することで特徴を明らかにした。ここでは、前小節で構築した分析枠組み、具体的には 7 ヶ国における平均国を設定し、それと比較することによって、各国の解答パターンを検討する。

#### 2.3.1. 項目母数の推定と等化

項目母数を各国で算出し、それを用いて 7 ヶ国における平均国を設定する。それに向けて、解答パターンを国際比較するための準備を行う。本小節においても等化の精度を高めるために、7 ヶ国間で分析対象とする項目を統一する。数学的リテラシー調査では、84 項目が使用されたことはすでに述べたが、ここで取り上げる 7 ヶ国では、ウルグアイに整合性が低いとされた項目が 2 つ指摘されている (OECD、2005a、p.190)。そのため、この 2 つの項目は分析対象に含めなかった。また、7 ヶ国それぞれで 82 項目の困難度と識別力を 2 母数モデルを使って求めたところ、芝 (1991) と豊田 (2002b) の基準を満たさないものがいくつかみられた<sup>21</sup>。本小節では、一定の比較の信頼性を確保するために、これらの項目も分析対象には含めず、最終的に残った 71 項目を用いて分析を進めていく。

続いて、各国で得られた 71 項目の困難度と識別力にブラジルのそれに等化を施す。その方法は前小節と同様に、mean-sigma 法の式 (1.5) を採用する。なお等化後の困難度と識別力は、71 項目×7 ヶ国の合計 497 と大きいため、その一部を表 2.8 に示した。

#### 2.3.2. 7 ヶ国の解答パターンの類似性と特異性

等化後の各項目の困難度と識別力を用いて、7 ヶ国における解答パターンの類似度を調べる。その分析結果として、図 2.32 と図 2.33 に困難度と識別力のテンドログラムを提示した。図 2.5 と図 2.6 ではアメリカを例外とすれば、困難度と識別力ともにアジア諸国と欧米諸国と分類された。ところが本分析からは、地域性といった特徴を捉えることは難しい。そこ

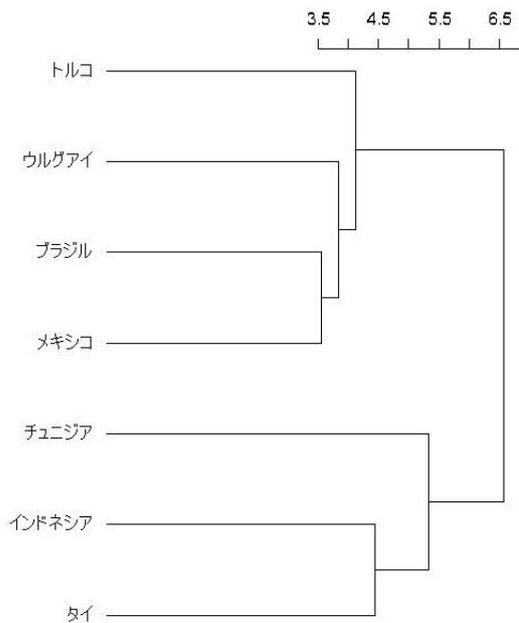
<sup>21</sup> 除外した項目は、M155Q03T、M179Q01T、M266Q01T、M305Q01、M408Q01T、M411Q02、M421Q02T、M446Q02、M462Q01T、M467Q01、M710Q01 の 11 項目である。

表 2.8 等化後の項目母数の一部

項目	ブラジル				インドネシア				メキシコ				...
	困難度	標準誤差	識別力	標準誤差	困難度	標準誤差	識別力	標準誤差	困難度	標準誤差	識別力	標準誤差	
1	0.059	0.058	0.644	0.088	0.167	0.049	0.541	0.053	0.339	0.032	0.491	0.030	
2	1.875	0.113	0.865	0.134	1.885	0.097	0.938	0.096	1.554	0.032	0.905	0.045	
3	1.568	0.067	1.344	0.192	1.125	0.050	1.041	0.087	1.931	0.050	0.755	0.043	
4	3.066	0.248	1.455	0.457	2.741	0.205	1.533	0.332	2.845	0.088	1.653	0.173	
5	1.322	0.105	0.579	0.087	0.712	0.044	0.838	0.066	0.453	0.020	0.861	0.038	
6	2.569	0.212	0.688	0.132	2.061	0.136	0.653	0.072	2.211	0.062	0.741	0.044	
7	0.456	0.048	0.912	0.110	0.570	0.031	1.228	0.090	0.130	0.018	1.182	0.050	
8	1.609	0.081	1.107	0.157	1.311	0.056	1.087	0.092	1.504	0.029	1.011	0.048	
9	0.579	0.042	1.183	0.143	0.757	0.039	1.057	0.083	0.994	0.019	1.102	0.048	
10	0.569	0.048	0.986	0.117	1.414	0.081	0.729	0.068	0.989	0.026	0.753	0.036	
11	0.869	0.048	1.282	0.156	1.746	0.088	0.964	0.096	1.422	0.032	0.838	0.041	
12	1.906	0.098	1.201	0.192	2.122	0.135	0.750	0.085	2.792	0.103	0.780	0.059	
13	0.714	0.050	1.035	0.125	0.863	0.041	1.175	0.094	0.597	0.019	0.975	0.042	
14	1.327	0.060	1.415	0.198	1.294	0.063	1.040	0.094	1.411	0.030	0.995	0.047	
15	0.955	0.071	0.772	0.102	1.044	0.072	0.656	0.062	1.298	0.035	0.708	0.036	
16	2.159	0.156	0.709	0.120	2.345	0.171	0.628	0.077	2.780	0.106	0.620	0.046	
17	1.659	0.147	0.567	0.099	0.814	0.095	0.388	0.048	1.664	0.071	0.429	0.031	
18	-0.794	0.056	1.178	0.169	-0.731	0.058	0.559	0.055	-1.117	0.047	1.047	0.063	
19	0.407	0.039	1.349	0.166	0.459	0.040	0.816	0.067	0.585	0.021	0.813	0.038	
20	1.411	0.061	1.648	0.258	2.584	0.219	0.637	0.089	2.281	0.072	0.808	0.052	
...													
平均値	1.322		0.954		1.322		0.844		1.322		0.833		
標準偏差	0.855		0.351		0.855		0.333		0.855		0.261		
最大値	3.704		1.809		2.741		1.555		2.947		1.653		
最小値	-0.816		0.321		-1.199		0.264		-1.117		0.382		

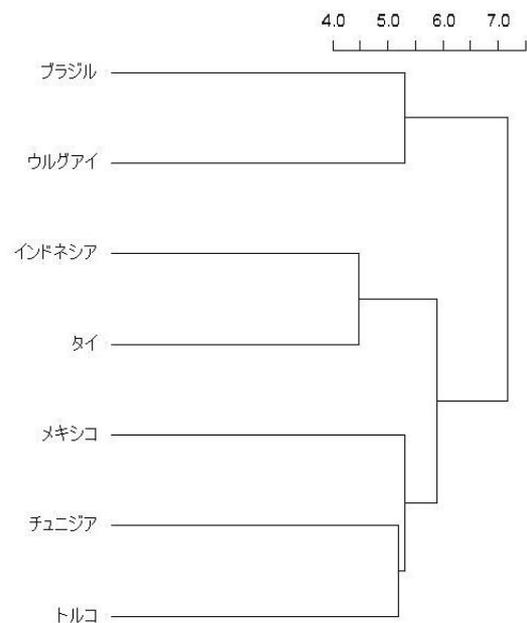
出典：筆者作成

図 2.32 困難度のテンドログラム



出典：筆者作成

図 2.33 識別力のテンドログラム



出典：筆者作成

表 2.9 7カ国における困難度と識別力の相関係数

(下三角行列：困難度、上三角行列：識別力)

国名		BRA	IDN	MEX	THA	TUN	TUR	URY
ブラジル	BRA	-	0.737	0.770	0.698	0.787	0.771	0.798
インドネシア	IDN	0.808	-	0.806	0.858	0.834	0.771	0.692
メキシコ	MEX	0.910	0.858	-	0.817	0.800	0.801	0.747
タイ	THA	0.833	0.859	0.872	-	0.764	0.788	0.725
チュニジア	TUN	0.801	0.847	0.812	0.777	-	0.807	0.760
トルコ	TUR	0.896	0.817	0.884	0.861	0.789	-	0.795
ウルグアイ	URY	0.906	0.811	0.891	0.832	0.759	0.880	-

出典：筆者作成

で、表 2.9 のように 7カ国における 71 項目の困難度と識別力の相関係数を計算した。なおこの表では、困難度の相関係数を下三角行列に、識別力の相関係数を上三角行列に示している。困難度における相関係数の平均値は 0.843、最大値は 0.910 であり最小値が 0.759 となっている。また識別力については、平均値が 0.777、最大値は 0.858 で最小値が 0.692 となった。この結果を単純に考えるならば、相関係数の値が小さい傾向にある識別力に各国の特徴がより反映されやすいと思われる。

続いて、分析対象の 7カ国において各国の解答パターンがどの程度の特異性を有するかをみていく。そのために前小節と同様の手順を踏み、7カ国における平均国を設定する。そして、7カ国が各々持つ各項目の困難度と識別力から平均国のそれらの値を引き算し、それらの標準偏差を各国で計算する。もし得られた標準偏差が小さいならば、その国は 7カ国の平均的な解答パターンを持ち、逆に大きい場合には、特異な解答パターンにあると考えられる。その分析結果を示したものが表 2.10 である。この表から分かるように、識別力における各国の値は、最大値がウルグアイの 0.483 で最小値がチュニジアの 0.407 となっている。困難度における最大値と最小値と比べるとその差異は小さく、各国とも識別力に特異な傾向を有することが示唆される。この分析結果を同様の手順を踏んだ先進国に関する結果である表 2.5 と比べると、困難度と識別力における国の順番にあまり一貫性がみられないことも分かる。

表 2.10 平均国との差の標準偏差

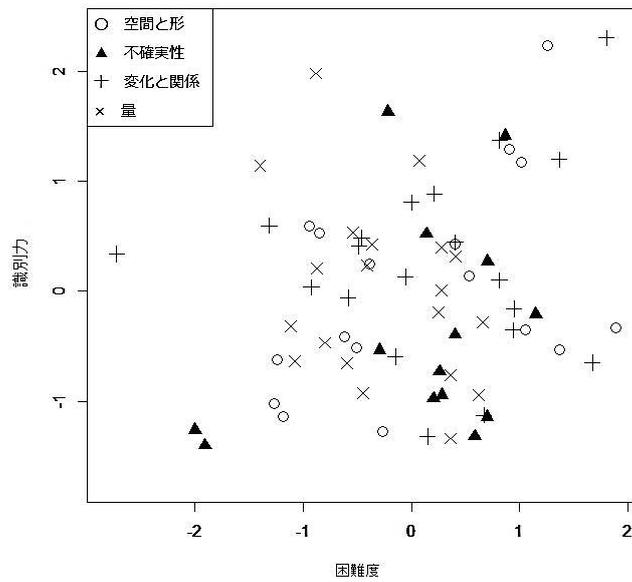
困難度		識別力	
国名	標準偏差	国名	標準偏差
チュニジア	0.461	ウルグアイ	0.483
インドネシア	0.389	ブラジル	0.470
タイ	0.376	タイ	0.442
ウルグアイ	0.358	インドネシア	0.426
トルコ	0.339	トルコ	0.414
ブラジル	0.328	メキシコ	0.411
メキシコ	0.294	チュニジア	0.407

出典：筆者作成

2.3.3. 7カ国の解答パターンの特徴

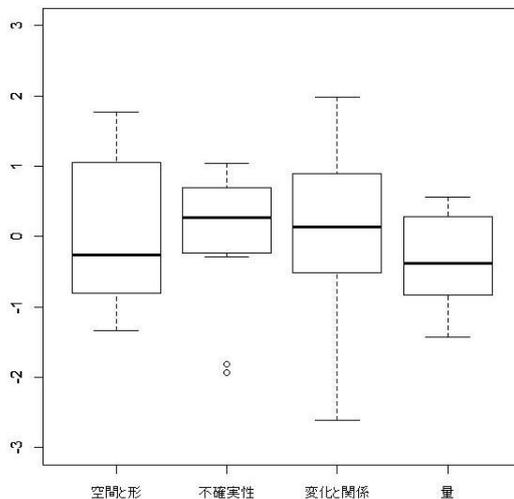
まず7カ国の解答パターンを調べる前に、平均国の解答パターンを確認しておく。平均国における71項目の困難度と識別力の散布図を図2.34に示した。また「包括的アイデア」にある4つの数学的な領域ごとの困難度と識別力の分布を箱ひげ図を用いて、図2.35と図2.36にそれぞれ要約した。例えば図2.35から「不確実性」の困難度は、中央値とレンジから全体的に高いことが分かる。他方で不確実性に属する項目の識別力は、いくらか低

図 2.34 平均国における困難度と識別力の散布図



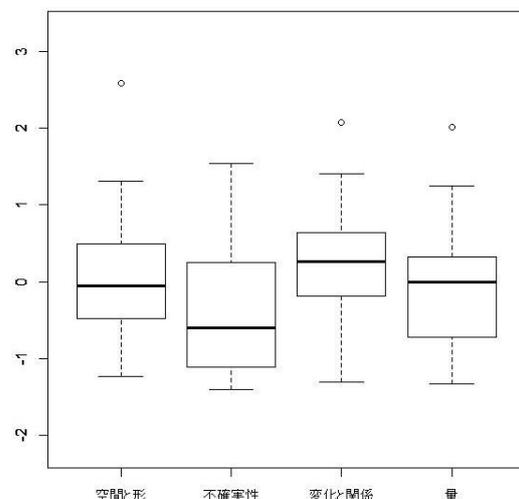
出典：筆者作成

図 2.35 困難度における箱ひげ図



出典：筆者作成

図 2.36 識別力における箱ひげ図



出典：筆者作成

い傾向にあることを読み取れる（図 2.36）。前小節で取り上げた 13 先進国における平均国においても、不確実性に属する項目は、比較的の高い困難度を持ち識別力はいくらか低めという傾向（図 2.8 と図 2.10 参照）を考えると、確率・統計に関連する内容は、平均的にみて途上国においても特徴的な解答パターンが反映されやすいのかもしれない。

では 7 カ国における平均的な解答パターンを持つ国を平均国とし、それとの比較を通じて、各国の解答パターンを浮かび上がらせる。ここでも前小節と同様に、平均国における各項目の困難度と識別力と各国のそれらを引き算し、困難度におけるその結果を困難度差とし、識別力では識別力差と呼ぶことにする。この定義から正の大きい困難度差ほど、平均的な解答パターンより高い難易度を持つ項目を表わし、負の値を示す場合はその逆である。また識別力の意味から、識別力差が正に大きい場合は、その項目が持つ受験者を識別する機能が低い項目であり、負の場合には高いことを意味する。つまりその国において効果的に機能しやすいまたは機能しにくい項目を示すことになる。困難度差と識別力差を用いることで、各項目は平均的な解答パターンとの差異という視点から特徴付けられる。そこで、図 2.37 から図 2.43 に 7 カ国における困難度差と識別力差の散布図を示した。

表 2.10 で示したように、標準偏差を用いて困難度と識別力を別々に検討したところ、識別力よりも困難度に平均国との違いが顕在化しやすいという結果を得た。さらに原点が平均国の解答パターンを意味することを利用して、各項目と原点との距離を用いて、困難度差と識別力差の 2 つの視点からみた全体的な差異をみていく。各項目と原点との距離を各国で求め、それらが属する「包括的アイディア」の 4 つの領域の平均値を計算する。その分析結果を表 2.11 に示し、それを図 2.44 に折れ線グラフに要約した。なおここでは、外れ値とも取れる項目を含めて全体像を把握するために平均値を採用した。

図 2.44 において、チュニジアは「不確実性」と「変化と関係」における値が大きくなっている。このことは、図 2.41 から視覚的に確認できる。また「空間と形」が突出している

図 2.37 ブラジル

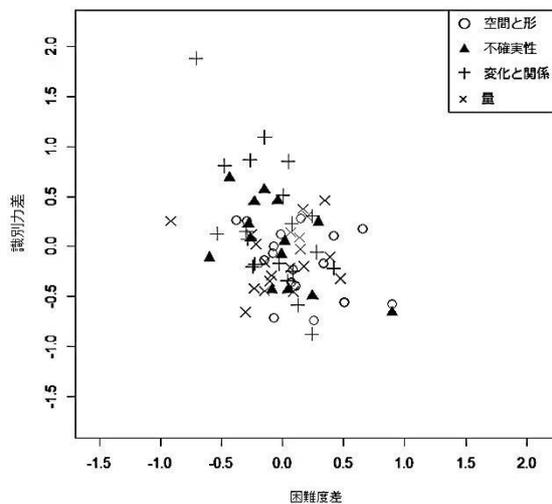


図 2.38 インドネシア

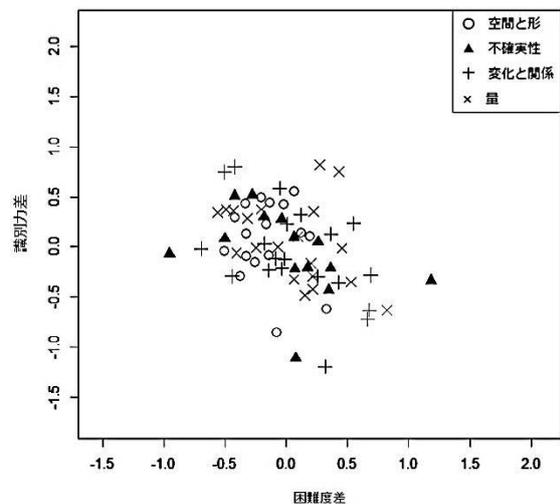


図 2.39 メキシコ

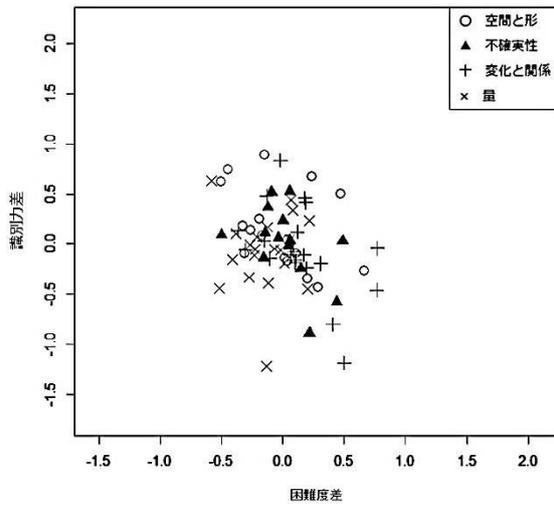


図 2.40 タイ

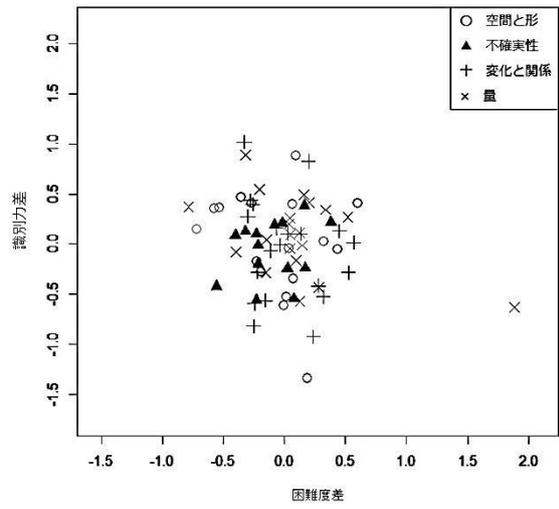


図 2.41 チュニジア

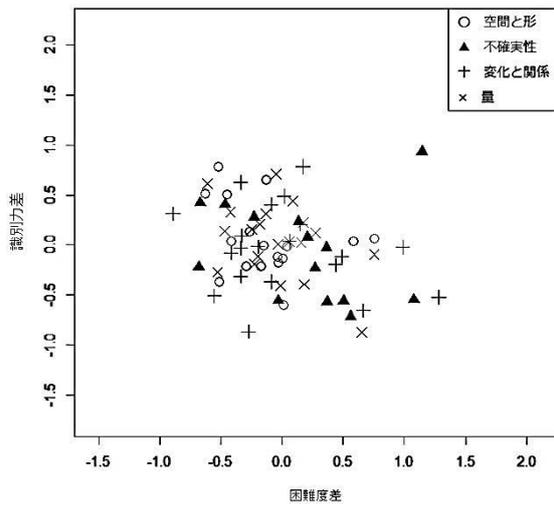


図 2.42 トルコ

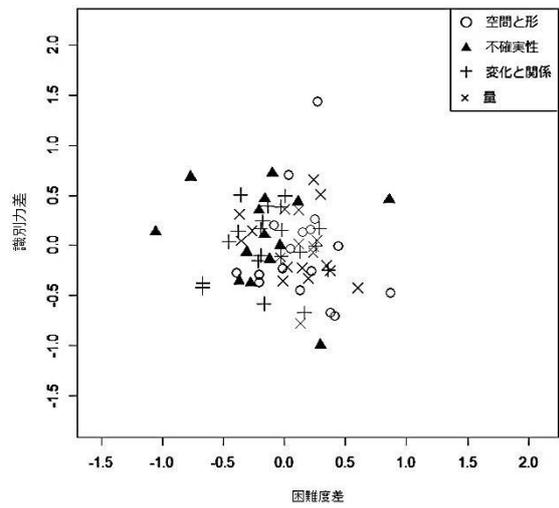


図 2.43 ウルグアイ

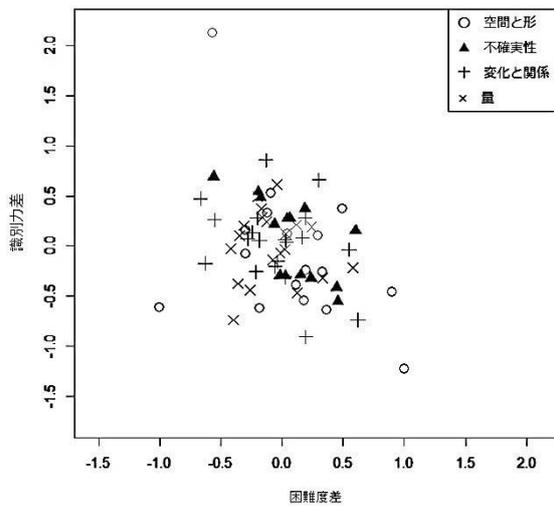
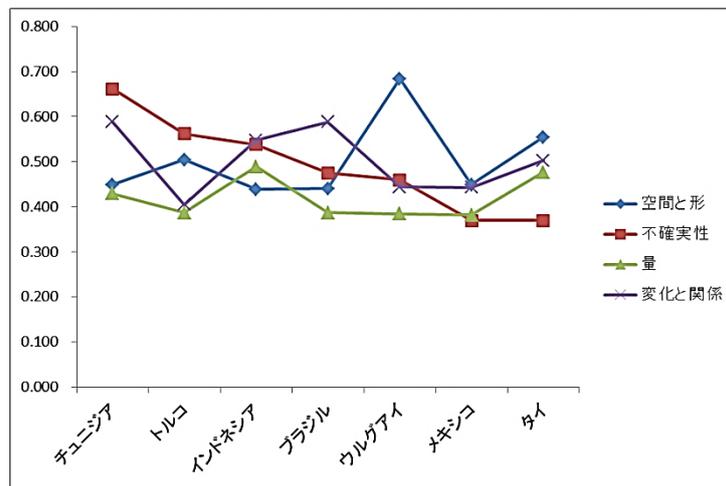


表 2.11 4つの領域における平均国との距離の平均値（不確実性の降順）

国名	空間と形	不確実性	量	変化と関係	最大値	最小値	レンジ
チュニジア	0.450	0.662	0.430	0.589	0.662	0.430	0.232
トルコ	0.504	0.562	0.387	0.405	0.562	0.387	0.175
インドネシア	0.439	0.538	0.489	0.547	0.547	0.439	0.108
ブラジル	0.440	0.475	0.387	0.588	0.588	0.387	0.202
ウルグアイ	0.684	0.459	0.384	0.444	0.684	0.384	0.300
メキシコ	0.449	0.370	0.382	0.443	0.449	0.370	0.079
タイ	0.555	0.369	0.477	0.503	0.555	0.369	0.185
最大値	0.684	0.662	0.489	0.589			
最小値	0.439	0.369	0.382	0.405			
レンジ	0.245	0.293	0.107	0.184			

出典：筆者作成

図 2.44 4つの領域における平均国との距離の平均値の折れ線グラフ（不確実性の降順）



出典：筆者作成

ウルグアイにおいても同様である（図 2.43 参照）。ゆえに表 2.11 および図 2.44 は、各国の困難度差と識別力差のばらつき具合を一定程度反映するものと考えられる。同様な手順を踏んで得た結果である表 2.7 と図 2.31 では、フィンランドを除けば「不確実性」に各国の解答パターンが顕在化しやすいことが浮かび上がった。ところが 7 途上国に対して得た結果の表 2.11 と図 2.44 をみると、各国の解答パターンがある特定の数学的な領域に表面化しやすいという全体的な傾向は確認されない。つまり 13 先進国のようにその国固有の解答パターンがある特定の数学的な領域に現れやすいのではなく、むしろ数学的な内容の枠組みに依存しない形で顕在化するという 7 途上国における全体的な傾向が浮かび上がった。

とはいえ 4 つの数学的な領域の枠組みを外すと分析が煩雑になるため、ここでは「不確実性」に限定しながら各国の解答パターンを調べることにする。図 2.45 から図 2.51 に「不確実性」における公開問題と非公開問題を区別し、各国の困難度差と識別力差の散布図を

作成した。なお提示した7つの図に振った番号は、表2.12の番号欄の数字に対応している。

前小節では13先進国の中で日本は、「輸出1(図2.13参照)」に高い困難性を抱えており、「大統領の支持率(図2.18参照)」では低い識別力を有することを示した。前者の項目は、与えられた2つのグラフが持つ情報を把握し、質問に合うグラフから指定されたデータを読み取る内容である。他方で後者の項目は、示された支持率に信頼性があるかを標本抽出の観点から問う内容であり、その方法に妥当性がなければ、そこから得られる代表値に信頼性は生まれない。したがってこの2つの項目には、確率・統計を活用する能力において不可欠な要素が含まれている。

そこで本小節で取り上げる7カ国に目を向けると、トルコで「輸出1」は比較的に高い困難性を有し、逆にインドネシアでは比較的に易しい内容と考えられる。他方で「大統領の支持率」は、トルコにおいて高い識別力を有し、逆にインドネシアでは低い結果となった。

図 2.45 ブラジル

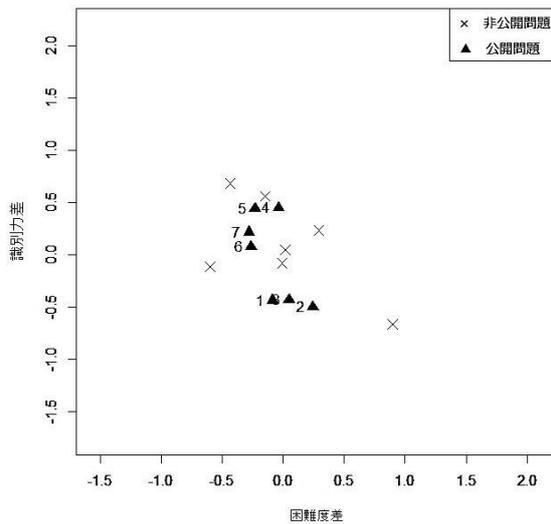


図 2.46 インドネシア

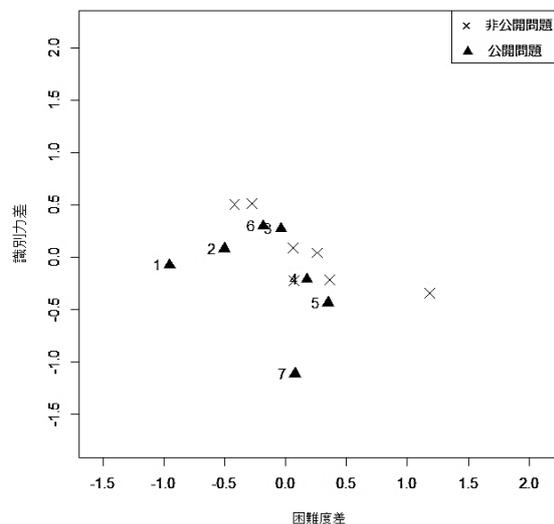


図 2.47 メキシコ

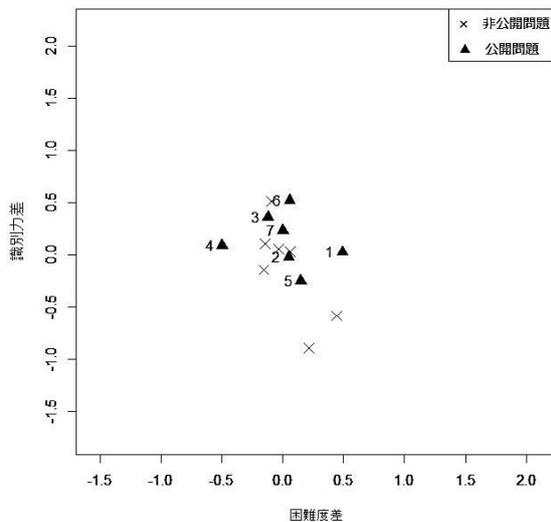


図 2.48 タイ

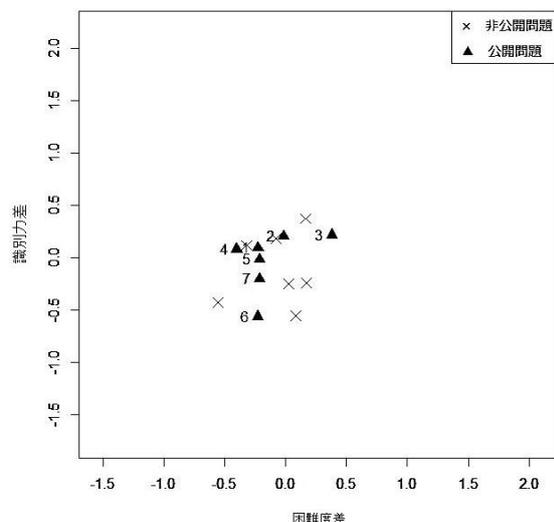


図 2.49 チュニジア

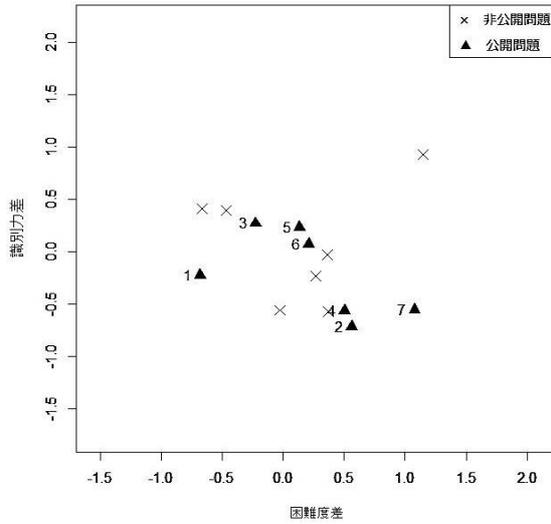


図 2.50 トルコ

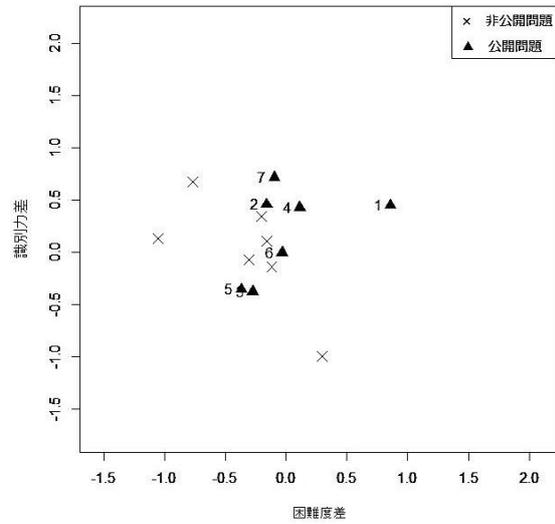


図 2.51 ウルグアイ

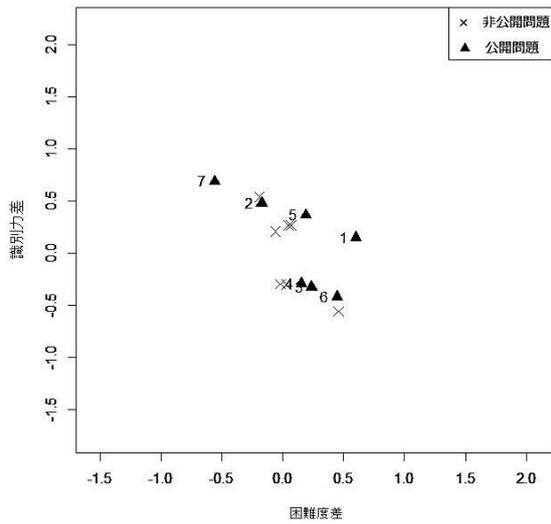


表 2.12 不確実性における 7つの公開問題

番号	項目番号	ユニット	内容
1	M438Q01	輸出1	グラフを正しく読む
2	M438Q02	輸出2	グラフにある情報を用いての割合
3	M468Q01T	理科のテスト	平均値の算出
4	M505Q01	ゴミ	棒グラフの適応条件について
5	M509Q01	地震	確率の実現的な意味について
6	M513Q01	テストの点数	テスト得点の分布の比較
7	M702Q01	大統領の支持率	世論調査の標本抽出について

出典：国立教育政策研究所（2004）をもとに筆者作成

つまり、高い難易度を有する項目と数学的リテラシーを測定するために効果的な項目が各国で異なるのである。この分析結果の背景には、日本の文脈で考察したように、困難度差

はその国の生徒には広く認識されている一方でその習得があいまいな内容を含む項目を反映し、識別力差は出題内容に対する習得の度合いだけでなく、それへの認識の度合いという側面で生徒の解答パターンが顕在化した可能性は必ずしも否定できないだろう。

#### 2.4. 本章のまとめ

本章では、PISA2003の数学的リテラシー調査のデータを活用して、生徒の数学学力とカリキュラムの関連性について調べてきた。

まず、PISA2003に参加した39カ国のうち、日本をはじめとする先進国と途上国の間には、その学力水準に大きな差がみられることを確認した。次に13先進国を分析対象にしながら、日本の生徒が有する解答パターンを困難度と識別力に着目して浮かび上がらせ、2003年当時に中学校数学科において確率・統計に関する内容が扱われていなかったという日本のカリキュラム上の特徴を活かし、数学学力とカリキュラムの関連性を考察した。そしてそこで構築した分析枠組みを用いて、数学的リテラシー調査のテスト得点が最下位層に位置する7途上国の解答パターンを分析した。

日本を中心とした先進国に関する分析では、次のことが明らかになった。まず、テスト得点を算出することで、日本をはじめ韓国と香港は、数学を活用する能力とともに確率・統計を活用する能力も国際的にみて高い水準にあることを確認した。次に、各国の解答パターンを国際比較したことで、日本と韓国および香港は、困難度と識別力の2つの観点から類似した解答パターンを持つと同時に、13カ国における平均的な解答パターンからみて特異であることが浮き彫りとなった。そして日本の特徴としては、「不確実性」と呼ばれる確率・統計に関する領域における困難度でより特異性が認められた。つまり日本の特異性は、4つの数学的な領域で同程度にみられるのではなく、特に「不確実性」に集中しているといえる。最後に、日本の平均国より困難度が高い項目および識別力が低い項目を同定し、その内容を公開問題により確認することで、生徒が抱えている困難性の段階や測定の機能が低下した項目の特徴を調べた。加えてその要因を2003年当時のカリキュラムの特徴を絡めることで、出題内容に対する習得の度合いだけでなく、それへの認識の度合いという観点から論じた。

以上を整理すると、日本の生徒が持つ数学、特に確率・統計を活用する能力は、国際的に高い水準にある一方で、その内実である解答パターンは国際的にみて特異であるという実態が明らかになった。そしてその背景には、日本のカリキュラムの特徴が一定程度反映されていると捉えた。具体的には、日本に焦点を当て平均国と比較することで、困難度差と識別力差は、生徒には広く認識されているが習得があいまいな内容を含む項目、あるいは生徒にほとんど認識されていない概念や考え方を含む項目を反映するという見解を得た。

なお欧米諸国でみられるデータを中心に据えた統計教育で重視される内容として、データ収集やデータの分布に関する理解と活用が挙げられる(渡辺、2007)。そして、日本においてもそのような内容を取り扱う必要性が指摘されている(青山、2007; 深澤、2007)。本

分析の結果では、「大統領の支持率 (図 2.18)」や「テストの得点 (図 2.17)」といったデータ収集やデータの分布に関する項目の測定に対する機能は、国際的にみて低いという結果を得た。つまり本章の分析結果は、習得の度合いだけでなく、数学教育で扱う統計的な学習内容の認識を高める対策の必要性を生徒の解答パターンという視座から確認したと主張できるだろう。

とはいえ「いろいろな色のキャンディ (図 2.15)」のように、確率を求める内容にも拘らず、識別力差が負の値を示した項目がある。言い換えれば、日本に根差した側面を持つ内容ではあるが、測定の機能が低下した項目への示唆がまだ得られていない。この項目は、グラフの値を正しく読み取った上で確率を求める、という統計的な部分と確率的な部分が合わさった構造をしている。そのため、分析の視点もある程度構造的にする必要があると考える。この分析に対する視点の構築は今後の課題に位置付けたい。さらに困難度差と識別力差を平面上で考えた場合に原点は、平均国の解答パターンを表わすため、それを利用して平均国との全体的な差異を各国で調べた。その結果、日本だけでなく他の国においても「不確実性」にその国固有の解答パターンが浮かび上がりやすいことが分かった。このことは、それぞれ国で確率・統計領域の数学教育が他の領域と比べて特徴的に行われていると読み取ることも可能だろう。

続いて、日本の解答パターンを浮かび上がらせた方法、具体的には平均国との比較を通じた分析を7途上国に対しても適用し、それらの国の解答パターンを検討した。その結果、日本を含め取り上げた13先進国では、「不確実性」において固有の解答パターンが顕在化しやすいことが浮かび上がった。ところが数学的リテラシー調査のテスト得点の分布が下位層にある途上国においては、数学的な領域にあまり依存しない形で解答パターンの固有性が表面化するという全体的な傾向が浮かび上がった。そして「不確実性」に属する公開問題を手掛かりに、トルコとインドネシアを取り上げて、いくらか解答パターンを検討した。まずトルコとインドネシアでは、数学的リテラシー調査のテスト得点に一定の差異が認められる(表 2.1 参照)。なおこの2つの国における差異の効果量<sup>22</sup>は、0.617である。ところが「輸出1」のように、質問に合うグラフから指定されたデータを読み取るという単純な内容の項目において、テスト得点の平均値が高いトルコの方にインドネシアより高い困難性が認められた。そこで日本の解答パターンとカリキュラムの特徴の関連性を指摘したように、ここで得られた結果を各国のカリキュラムと照らし合わせながら読み解く取り組みが今後の大きな課題に位置付けたい。

最後に本章では、「包括的アイディア」の中でも特に「不確実性」に着目して解答パター

<sup>22</sup> ここでは、2つの集団における平均値が  $\mu_1$  と  $\mu_2$  で、標準偏差を  $\sigma_1$  と  $\sigma_2$  とするとき、

$$(2.1) \quad \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

と定義される効果量を採用している。

ンの分析を進めた。ここで構築した分析枠組みは、不確実性のみに限らず、各出題項目を特徴付ける評価枠組み、例えば「状況・文脈」、「能力」あるいは出題形式に焦点を合わせて適用することも可能である。分析対象とする国や出題項目の評価枠組みに合わせて、鈴木・豊田・川端（2008）が述べるように、項目反応理論を活用する解答パターンを明らかにする分析が今後さらに蓄積されることを期待したい。

## 第3章 数学学力と数学に対する情意的側面の関連性

第2章と同様に PISA2003 で収集されたデータの二次分析を通じて、生徒の数学学力と数学に対する情意的側面の関連性を日本を中心に検討し、さらにそこで構築した分析枠組みを用いて、途上国に関する分析を試みる。

### 3.1. 数学的リテラシー調査のテスト得点と 5 つの情意的側面に関する因子得点の国際比較

本小節では、PISA2003 に参加した 39 カ国の数学的リテラシー調査のテスト得点と 5 つの数学に対する情意的側面、具体的には①興味・関心（内的な動機付け）、②道具的動機付け（外的な動機付け）、③自己効力感（数学への自信）、④自己概念（数学的な能力への信念）、⑤不安（数学に対する不安）を因子分析を用いて尺度を構成し、その国際比較を通じて先進国と途上国にどのような差異が認められるのかを検討する。そして、それらの関連性を階層線形モデルを用いて調べていく。

#### 3.1.1. 使用データとそのデータ処理

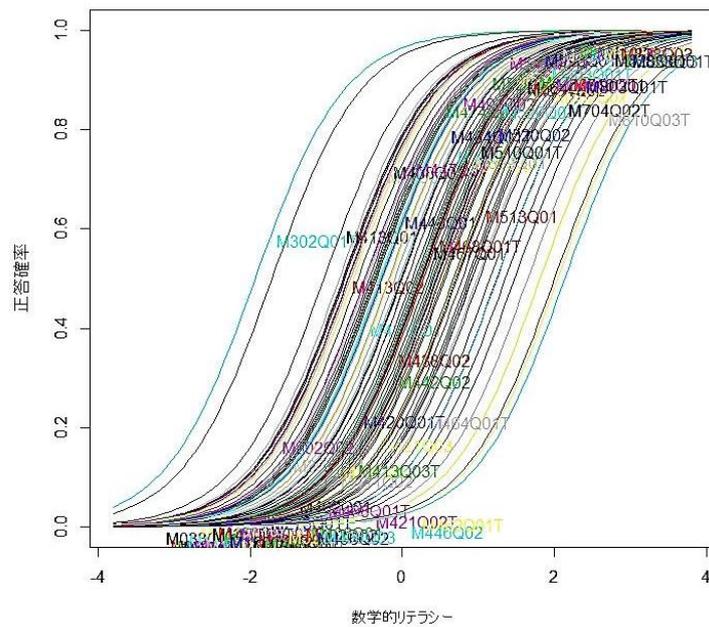
数学的リテラシー調査におけるデータは、図 1.7 で示したように、全ての項目において正答を 1 誤答を 0 とする 2 値データへと処理している。続いて、情意的側面に関する質問紙調査では、「全くその通りだ」、「その通りだ」、「その通りではない」および「全くその通りではない」のような 4 つの選択肢の中から該当するものを選ぶという方式が採られている。とはいえ質問紙調査のデータにおいても、無回答や欠損値が含まれているため、コードブック（OECD、2005b）をみながら、情意的側面の測定のために作成された 26 項目全てに有効回答を持つ生徒に限定した。したがって本章で分析対象となる生徒は、39 カ国の全受験者の約 85%に相当する 228038 名となった。なお全 26 項目のうち 6 項目が否定的に回答する逆転項目となっている。そのため、全ての項目が肯定的な意味を持つようにデータ処理を行った。具体的には、自己概念に関する 1 項目、不安に関する全 5 項目に処理を施した。このようなデータ処理から不安は、全体を通じて不安の無さを表わす尺度となっている。

#### 3.1.2. 数学的リテラシー調査のテスト得点の国際比較

本章では対象とする生徒数が前章とは異なるため、数学的リテラシー調査のテスト得点の分布を各国で改めて確認する。PISA に関する報告書（e.g., OECD、2005a、2006、2009）にある調査結果は、項目反応理論を用いて算出されていた。そのため本章でも、項目反応理論を採用することにする。なお用いる項目反応モデルは、ラッシュモデル（1.2）とした。前章では 2 母数モデルを採用したが、本章では、PISA で使用されているモデルに合わせることで、その扱い方が容易になるためである（豊田、2012b、p.25）。

まずテスト得点を算出する前に、項目母数である困難度に極端な値を示すものがないかを確認する。そこで、各項目の項目特性曲線を図 3.1 に示した。得られた 84 項目の困難度の最大値と最小値を調べると、それぞれ 2.141 と -1.932 を得た。したがって芝 (1991) および豊田 (2002b) の基準を参考にすれば、84 項目全てを用いてテスト得点を算出できると考える。

図 3.1 84 項目における項目特性曲線



出典：筆者作成

ラッシュモデルを用いて算出した各国のテスト得点の分布を平均値と標準偏差を用いて表 3.1 のように要約した。なお受験者全体で平均値 0.0、標準偏差 1.0 に調節してある。この表から分かるように、上位層には香港、オランダ、ベルギー、フィンランド、韓国や日本といった先進国が位置している。その一方で、下位層にはウルグアイ、メキシコ、ブラジル、チュニジアやインドネシアといった途上国が位置することを見て取れる。そこで、国別の違いが受験者のテスト得点をどれだけ説明するかを表わす級内相関係数を計算したところ、その値は 0.185 となった。このことは、228038 名の受験者をわずか 39 カ国に分けた場合、生徒のテスト得点のばらつきが約 18.5%説明されることを意味する。また最も高い平均値を持つ香港と逆に低いインドネシアを比較すると、各国の標準偏差が 1.0 より小さいことから、その差異 1.584 は、決して小さいとはいえないだろう。纏めると、PISA2003 で測定された数学的リテラシー、つまり本研究で規定する数学学力の差異は、先進国と途上国の間で十分に大きいと主張できよう。

そこで本章においても、日本を含むいくつかの先進国と数学的リテラシーの水準が下位

層に位置する途上国に焦点を絞って国際比較分析を行う。対象とする国は前章と同様に、先進国では、オーストラリア、カナダ、フィンランド、フランス、ドイツ、香港、アイルランド、イタリア、日本、韓国、オランダ、ニュージーランド、アメリカの13カ国であり、途上国については、トルコ、タイ、ウルグアイ、メキシコ、ブラジル、チュニジア、インドネシアの7カ国とする。

表 3.1 ラッシュモデルによる各国のテスト得点（平均値の降順）

国名	平均値	標準誤差	中央値	標準偏差	標準誤差	最大値	最小値	受検者数
香港	0.583	0.014	0.687	0.954	0.010	3.282	-2.782	4354
オランダ	0.516	0.015	0.536	0.895	0.011	3.282	-2.080	3532
ベルギー	0.492	0.011	0.551	0.974	0.008	3.175	-2.749	7480
フィンランド	0.491	0.012	0.494	0.861	0.008	3.106	-2.513	5464
チェコ	0.459	0.013	0.486	0.954	0.009	3.200	-2.749	5752
韓国	0.457	0.013	0.467	0.914	0.009	3.282	-2.724	5293
日本	0.413	0.015	0.486	0.975	0.010	3.282	-2.782	4464
ニュージーランド	0.365	0.015	0.385	0.962	0.011	3.175	-2.739	4098
カナダ	0.360	0.006	0.376	0.890	0.005	3.282	-2.622	19463
オーストラリア	0.329	0.009	0.365	0.952	0.006	3.200	-2.782	11543
スイス	0.315	0.011	0.365	0.940	0.008	3.282	-2.918	7218
マカオ	0.289	0.027	0.330	0.927	0.019	2.976	-2.513	1178
デンマーク	0.278	0.016	0.307	0.919	0.011	3.200	-2.369	3414
ドイツ	0.272	0.015	0.302	0.924	0.010	3.282	-2.739	4001
フランス	0.269	0.015	0.307	0.900	0.010	2.810	-2.513	3842
アイスランド	0.242	0.017	0.299	0.926	0.012	3.282	-2.664	3094
オーストリア	0.239	0.014	0.234	0.890	0.010	3.175	-2.513	4128
スウェーデン	0.211	0.015	0.200	0.948	0.011	3.145	-2.513	3968
アイルランド	0.169	0.014	0.189	0.849	0.010	2.976	-2.513	3503
スロバキア	0.169	0.011	0.166	0.912	0.008	3.282	-2.574	6689
ノルウェー	0.097	0.016	0.103	0.939	0.011	3.175	-2.749	3504
ルクセンブルク	0.075	0.015	0.091	0.902	0.011	2.762	-2.622	3421
イタリア	0.074	0.009	0.103	0.934	0.006	3.175	-2.621	10864
スペイン	0.071	0.009	0.097	0.877	0.006	3.175	-2.739	9695
ハンガリー	0.064	0.014	0.061	0.916	0.010	3.282	-2.508	4125
ポーランド	0.019	0.014	0.028	0.918	0.010	2.976	-2.782	4117
ラトビア	-0.018	0.014	0.023	0.887	0.010	3.094	-2.664	4305
アメリカ	-0.051	0.013	-0.066	0.902	0.009	3.200	-2.918	5014
ロシア	-0.106	0.013	-0.095	0.944	0.009	3.094	-2.749	5341
ポルトガル	-0.194	0.014	-0.214	0.879	0.010	2.511	-2.622	4161
旧セ	-0.358	0.014	-0.399	0.831	0.010	2.762	-2.782	3588
ギリシャ	-0.360	0.014	-0.356	0.911	0.010	2.511	-2.749	4037
トルコ	-0.422	0.015	-0.485	0.948	0.011	3.175	-2.918	3879
ウルグアイ	-0.473	0.014	-0.485	0.961	0.010	2.810	-2.742	4435
タイ	-0.573	0.012	-0.628	0.860	0.009	2.810	-2.782	4934
メキシコ	-0.663	0.005	-0.655	0.782	0.003	2.762	-2.912	25158
ブラジル	-0.934	0.015	-1.003	0.858	0.010	2.442	-2.918	3351
チュニジア	-0.956	0.015	-1.001	0.791	0.011	2.442	-2.915	2778
インドネシア	-1.001	0.008	-1.040	0.734	0.006	1.870	-2.918	8853
全体	0.000	0.002	0.004	1.000	0.001	3.282	-2.918	228038

注) 旧セは、旧セルビア・モンテネグロ

出典：筆者作成

## 3.1.3. 5つの情意的側面における因子得点の国際比較

続いて、①興味・関心、②道具的動機付け、③自己効力感、④自己概念、⑤不安の5つの情意的側面ごとに1因子の因子分析を行い、その際に得られる因子得点を用いて情意的側面に関する国際比較分析を行う。その得られた結果の一つとして、表3.2に各内容の定義、項目数、因子寄与率、 $\alpha$ 信頼係数および $\omega$ 信頼係数を提示した。なお具体的な質問事項に対する因子負荷量と共通性については、本章の付録3.1を参照されたい。また、表3.3に各国の因子得点の平均値と標準偏差を示した。この表にある斜体の数値は、39カ国で平均値を比較した際の順位である。例えば日本をみるならば、5つの情意的側面の水準は39カ国の中で最下位層にあるといえる。特に不安以外では、その順位が最下位となっている。そこで図3.2に横軸を「興味・関心」として、39カ国における5つの情意的側面の平均値の散布図を作成した。なお、余白に興味・関心と自己効力感の散布図を反映させ、そこにOECDが使用している国ラベルを記した。日本は、5つの側面で一貫して国際的にみて低い水準にあることが分かる。ところが、チュニジアやインドネシアあるいはタイといった国では、興味・関心の水準が高くても、必ずしも他の側面が同様に高い水準とは限らないことも見て取れる。つまり国際的にみて、日本のように全ての側面で一貫した水準にある国は少なく、多くの場合は、側面ごとに異なる水準にあることが分かった。特に本研究で注目する7途上国では、例えば数学に対する興味・関心が高くとも、自己効力感や不安における水準は高くないという実態が浮かび上がった。

表3.2 5つの情意的側面の内容とその因子寄与率、 $\alpha$ 信頼係数および $\omega$ 信頼係数

内容	定義	項目数	因子寄与率	$\alpha$ 信頼係数	$\omega$ 信頼係数
興味・関心	1つの教科としての数学への興味・関心および数学を学ぶことの楽しみについて	4	0.665	0.887	0.888
道具的動機付け	良い仕事の見込みなどの外的な報酬によって、学ぶことがどの程度促進されるかについて	4	0.644	0.877	0.878
自己効力感	困難を克服して、学習場면을効果的に扱う自分の能力をどの程度信じているかについて	8	0.375	0.825	0.826
自己概念	自分自身の数学の能力に関する信念について	5	0.598	0.879	0.881
不安	数学に取り組んでいるときに、無力であるか、あるいは情緒的な圧力をどの程度感じているかについて	5	0.472	0.811	0.816

出典：筆者作成

## 3.1.4. 階層線形モデルによるテスト得点と5つの情意的側面の関連性の国際比較

数学的リテラシー調査のテスト得点には、先進国と途上国の間で大きな差異が存在する。また5つの情意的側面に関して、その水準が高いあるいは低いとして表れる側面も異なることも浮かび上がった。このように得られた結果を眺めると、測定された数学学力と情意的側面に一見不合理とも解釈できる現状がある。つまり国際的な教育調査において、数学学力と数学に対する情意的側面にある関連性が必ずしも明確でないのである。

表 3.3 5つの情意的側面における因子得点（興味・関心の平均値の降順）

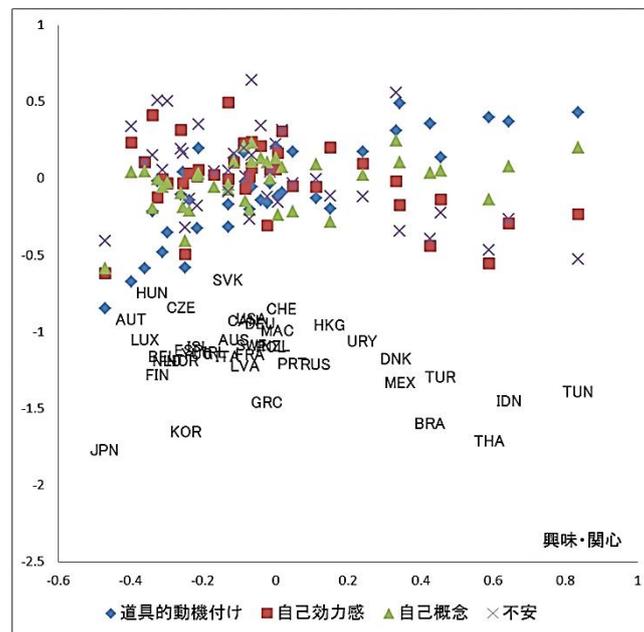
国名	興味・関心		道具的動機付け		自己効力感		自己概念		不安						
	平均値	SD	平均値	SD	平均値	SD	平均値	SD	平均値	SD					
チュニジア	0.835	1	1.052	0.430	2	1.016	-0.236	33	1.085	0.202	4	1.174	-0.529	39	1.029
インドネシア	0.644	2	0.674	0.372	4	0.650	-0.294	34	0.698	0.077	12	0.662	-0.268	33	0.711
タイ	0.590	3	0.709	0.398	3	0.655	-0.554	38	0.949	-0.137	29	0.699	-0.469	38	0.784
トルコ	0.455	4	1.088	0.138	12	0.981	-0.137	31	1.091	0.048	14	1.024	-0.224	31	1.093
ブラジル	0.426	5	0.891	0.356	5	0.870	-0.442	36	0.952	0.037	17	0.948	-0.393	36	0.918
メキシコ	0.343	6	0.821	0.488	1	0.737	-0.175	32	0.911	0.103	10	0.876	-0.345	35	0.851
デンマーク	0.332	7	1.016	0.312	6	0.867	-0.020	24	1.022	0.247	1	1.068	0.561	2	1.024
ウルグアイ	0.240	8	1.057	0.172	9	0.992	0.094	13	0.966	0.021	19	1.052	-0.118	27	1.001
香港	0.150	9	0.962	-0.197	29	0.889	0.199	9	1.006	-0.284	37	0.931	-0.113	25	0.941
ロシア	0.112	10	0.830	-0.127	24	0.930	-0.057	28	0.895	0.090	11	0.810	-0.007	22	0.850
ポルトガル	0.047	11	0.879	0.172	10	0.925	-0.050	27	0.896	-0.216	35	0.970	-0.032	23	0.897
スイス	0.018	12	1.068	-0.091	21	1.057	0.306	4	0.977	0.075	13	1.082	0.317	8	1.049
マカオ	0.006	13	0.908	-0.113	23	0.854	0.164	10	0.902	-0.239	36	0.964	-0.157	29	1.024
ニュージーランド	-0.001	14	0.992	0.207	7	0.911	0.061	15	0.994	0.132	6	0.942	0.223	9	0.906
ポーランド	-0.015	15	0.937	-0.031	19	0.829	0.059	16	0.955	-0.002	21	0.956	0.042	20	0.973
ギリシャ	-0.024	16	1.046	-0.160	27	1.043	-0.305	35	1.009	0.105	9	0.937	-0.117	26	1.007
ドイツ	-0.040	17	1.165	-0.144	26	1.019	0.209	8	0.982	0.134	5	1.162	0.341	6	1.146
スウェーデン	-0.067	18	1.026	-0.055	20	0.960	0.067	14	1.051	0.109	7	1.032	0.639	1	0.931
アメリカ	-0.067	19	1.065	0.091	14	0.970	0.239	5	0.991	0.232	2	1.059	0.151	15	1.082
フランス	-0.072	20	1.004	-0.201	30	1.081	0.008	21	1.013	-0.208	33	1.068	-0.264	32	1.005
ラトビア	-0.084	21	0.774	-0.023	18	0.854	-0.068	29	0.870	-0.148	30	0.862	0.012	21	0.844
カナダ	-0.087	22	1.042	0.170	11	0.993	0.229	7	1.018	0.212	3	1.108	0.197	10	1.087
オーストラリア	-0.116	23	0.971	0.134	13	0.958	0.104	12	0.999	0.109	8	0.945	0.161	13	0.904
イタリア	-0.131	24	0.975	-0.315	32	0.990	-0.004	22	0.893	-0.040	24	1.055	-0.089	24	0.960
スロバキア	-0.131	25	0.859	-0.168	28	0.881	0.493	1	0.895	-0.075	27	0.889	0.050	18	0.882
アイルランド	-0.169	26	0.971	0.016	16	0.964	0.023	19	0.997	-0.055	25	0.963	0.047	19	0.964
アイスランド	-0.214	27	1.083	0.198	8	1.014	0.057	17	1.130	0.031	18	1.185	0.352	5	1.067
旧セ	-0.216	28	0.940	-0.326	33	1.021	0.012	20	0.939	0.018	20	0.907	-0.180	30	1.012
スペイン	-0.238	29	0.984	-0.140	25	1.068	0.033	18	0.931	-0.212	34	1.031	-0.131	28	0.956
韓国	-0.250	30	1.014	-0.582	36	1.062	-0.495	37	1.061	-0.409	38	0.933	-0.323	34	0.900
ノルウェー	-0.256	31	1.093	0.040	15	1.053	-0.033	26	1.137	-0.189	31	1.149	0.165	12	1.107
チェコ	-0.261	32	0.850	-0.106	22	0.869	0.316	3	0.945	-0.099	28	0.960	0.190	11	0.910
オランダ	-0.300	33	0.913	-0.355	34	0.945	-0.032	25	0.969	-0.029	23	1.027	0.504	4	0.839
ベルギー	-0.313	34	0.990	-0.481	35	1.081	-0.005	23	1.008	-0.056	26	1.017	0.056	17	1.012
フィンランド	-0.327	35	0.943	-0.021	17	0.919	-0.126	30	1.039	-0.005	22	1.068	0.506	3	0.886
ハンガリー	-0.340	36	0.899	-0.214	31	0.913	0.411	2	0.902	-0.192	32	0.895	0.151	14	0.924
ルクセンブルク	-0.361	37	1.114	-0.586	37	1.243	0.105	11	1.055	0.046	15	1.146	0.104	16	1.173
オーストリア	-0.400	38	1.056	-0.675	38	1.114	0.234	6	1.016	0.041	16	1.113	0.339	7	1.147
日本	-0.471	39	1.022	-0.849	39	1.154	-0.618	39	1.148	-0.588	39	0.962	-0.409	37	1.080
全体	0.000		1.000	0.000		1.000	0.000		1.000	0.000		1.000	0.000		1.000

注) 旧セは、旧セルビア・モンテネグロ

出典：筆者作成

PISA や TIMSS のように全世界的な規模で行われる調査の情意的側面に関する結果を国際比較するならば、受験者が質問項目に回答する際に想定する判断基準の違い、つまり田崎（2008、p.63）が指摘する、異なる文化圏にみられる価値判断の違いを考慮すべきと考える。例えば、「数学が好き」という概念はどの文化圏にも当てはまるだろう。ところがその判断基準が各文化圏で異なるならば、全体として妥当性の低い調査結果が生じる可能性を否定できない。ゆえに、情意的側面に関して得られた結果を直接比較するだけでなく、

図 3.2 39 カ国の 5 つの情意的側面の平均値の散布図（横軸：興味・関心）



出典：筆者作成

何らかの工夫<sup>23</sup>を施して国際比較を試みる必要がある。

そこで本小節では、それらの値を直接比較するだけでなく、評価の基準を統一して算出される数学的リテラシー調査のテスト得点を介して国際比較するなど工夫し、数学学力と情意的側面の関連性を調べていく。そのためにテスト得点を目的変数とし、5つの情意的側面と性別を説明変数に持つ階層線形モデル<sup>24</sup>による方法を試みる。なお本研究では、国際比較を意図しているため、228032名の受験者が39カ国のどれかの国に属するという入れ子構造 (nested structure) を活かし、国レベルと生徒レベルの2つの階層に焦点を絞ることにする。そこで、表 3.4 と表 3.5 に本分析で使用する変数とその国レベルと生徒レベルにお

表 3.4 国レベルの記述統計量 (39 カ国)

変数名	最小値	最大値	平均値	SD	
数学的リテラシー	$y$	-1.001	0.583	0.031	0.427
興味・関心	$x_1$	-0.471	0.835	-0.019	0.305
道具的動機付け	$x_2$	-0.849	0.488	-0.058	0.313
自己効力感	$x_3$	-0.618	0.493	-0.006	0.250
自己概念	$x_4$	-0.588	0.247	-0.029	0.178
不安	$x_5$	-0.529	0.639	0.024	0.293

出典：筆者作成

<sup>23</sup> 吉野・林・山岡 (2010) によれば、個々の質問項目を個別に比較する場合には、質問の聞き方や多少の表現で無視できない差が生じることがあるが、多数の質問に対する全体のパターンを眺めると、多少の表現の違いの効果をほとんど無視でき、安定した結果を得やすいと述べている。

<sup>24</sup> 本分析では、Rに含まれる階層線形モデルを実行するためのパッケージ `mlmRev` を使用した。

表 3.5 生徒レベルの記述統計量 (228032 名)

変数名		最小値	最大値	平均値	SD
数学的リテラシー	$y$	-2.918	3.282	0.000	1.000
興味・関心	$x_1$	-1.776	2.218	0.000	1.000
道具的動機付け	$x_2$	-2.802	1.483	0.000	1.000
自己効力感	$x_3$	-3.475	1.757	0.000	1.000
自己概念	$x_4$	-2.005	2.145	0.000	1.000
不安	$x_5$	-2.534	2.123	0.000	1.000
性別(男子ダミー)	$x_6$	0	1	0.491	0.500

出典：筆者作成

ける基本統計量を示した。ただし性別は、男子を 1 とするダミー変数である。

本分析では 4 つのモデルを構成し、各国の特徴を浮かび上がらせていく。まず、説明変数を含まない以下のモデル 1 を考える。

$$\text{モデル 1: } y_{ij} = (\gamma_{00} + u_{0j}) + \varepsilon_{ij}$$

モデル 1 は一般にヌルモデルと呼ばれ、数学的リテラシー調査のテスト得点  $y_{ij}$  が受験者全体の平均値を表わす固定効果  $\gamma_{00}$  とその国固有のランダム効果  $u_{0j}$  および生徒レベルの誤差項  $\varepsilon_{ij}$  の 3 つから成ると仮定している。ただし  $y_{ij}$  は、 $j$  国 ( $j=1,2,\dots,39$ ) における  $i$  番目の生徒のテスト得点である。このモデルを通じて、国の違いによってテスト得点の差異がどの程度説明されるのかを把握でき、さらに説明変数を投入したモデルの適合度を評価する基準を設定できる。

次に、モデル 1 で示される国別によるテスト得点の差異に対して、国レベルで 5 つの情意的側面がどの程度影響しているかを調べるために、以下のようなモデルを考える。

$$\text{モデル 2: } y_{ij} = (\gamma_{00} + u_{0j}) + \gamma_{01}(x_{1\bullet j} - x_{1\bullet\bullet}) + \gamma_{02}(x_{2\bullet j} - x_{2\bullet\bullet}) + \gamma_{03}(x_{3\bullet j} - x_{3\bullet\bullet}) + \gamma_{04}(x_{4\bullet j} - x_{4\bullet\bullet}) + \gamma_{05}(x_{5\bullet j} - x_{5\bullet\bullet}) + \varepsilon_{ij}$$

モデル 2 は、モデル 1 に国レベルにおける 5 つの情意的側面を説明変数として投入している。ただし  $x_{k\bullet j}$  ( $k=1,2,3,4,5$ ) は、表 3.4 にある変数  $x_k$  の  $j$  国における平均値をそれぞれ表し、 $x_{k\bullet\bullet}$  は、 $x_{k\bullet j}$  の平均値を意味する。このモデルによって、国レベルの各変数をその全体平均値に中心化させ、国別における差異を浮かび上がらせる。なおランダム効果については、モデル 1 と同様に  $u_{0j}$  のみとなっている。

続いて、モデル 2 に生徒レベルにおける 5 つの情意的側面と性別の説明変数を投入したモデルを考える。生徒レベルの変数を加えることで、国内におけるテスト得点のばらつき具合を検討できる。そのモデルは、以下の 2 つである。

$$\begin{aligned} \text{モデル 3: } y_{ij} &= (\gamma_{00} + u_{0j}) + \gamma_{01}(x_{1\bullet j} - x_{1\bullet\bullet}) + \gamma_{02}(x_{2\bullet j} - x_{2\bullet\bullet}) + \gamma_{03}(x_{3\bullet j} - x_{3\bullet\bullet}) \\ &+ \gamma_{04}(x_{4\bullet j} - x_{4\bullet\bullet}) + \gamma_{05}(x_{5\bullet j} - x_{5\bullet\bullet}) + \gamma_{10}(x_{1ij} - x_{1\bullet j}) + \gamma_{20}(x_{2ij} - x_{2\bullet j}) \\ &+ \gamma_{30}(x_{3ij} - x_{3\bullet j}) + \gamma_{40}(x_{4ij} - x_{4\bullet j}) + \gamma_{50}(x_{5ij} - x_{5\bullet j}) + \gamma_{60}(x_{6ij} - x_{6\bullet j}) + \varepsilon_{ij} \end{aligned}$$

モデル 4 :

$$\begin{aligned}
 y_{ij} = & (\gamma_{00} + u_{0j}) + \gamma_{01}(x_{1\cdot j} - x_{1\cdot\cdot}) + \gamma_{02}(x_{2\cdot j} - x_{2\cdot\cdot}) + \gamma_{03}(x_{3\cdot j} - x_{3\cdot\cdot}) + \gamma_{04}(x_{4\cdot j} - x_{4\cdot\cdot}) \\
 & + \gamma_{05}(x_{5\cdot j} - x_{5\cdot\cdot}) + (\gamma_{10} + u_{1j})(x_{1ij} - x_{1\cdot j}) + (\gamma_{20} + u_{2j})(x_{2ij} - x_{2\cdot j}) + (\gamma_{30} + u_{3j})(x_{3ij} - x_{3\cdot j}) \\
 & + (\gamma_{40} + u_{4j})(x_{4ij} - x_{4\cdot j}) + (\gamma_{50} + u_{5j})(x_{5ij} - x_{5\cdot j}) + \gamma_{60}(x_{6ij} - x_{6\cdot j}) + \varepsilon_{ij}
 \end{aligned}$$

それぞれのモデルでは、生徒レベルの変数  $x_{kij}$  を各国の平均値  $x_{k\cdot j}$  に中心化している。ただしモデル3では、ランダム効果が  $u_{0j}$  に限定されているが、モデル4においては、生徒レベルの変数の回帰係数にもランダム効果  $u_{kj}$  がそれぞれ含まれている。つまりこの2つのモデルには、ランダム切片モデルとランダム切片・係数モデルという違いがある。本小節では、4つのモデルを用いてテスト得点と5つの情意的側面の関連性を検討していく。

本分析の結果である表3.6をみていく。まず、モデル1の国別分散と生徒間分散に注目する。これらを用いて計算される級内相関係数を確認すると0.185であった。このことは、228038名の受験者を国別でわずか39グループに分けた場合、生徒のテスト得点のばらつきが約18.5%説明されることを意味する。つまり本分析のように、国の違いに着目した階層線形モデルを用いる方法に一定の妥当性が認められる。次に、モデル1とモデル2における国別分散の値に目を向ける。それらはそれぞれ0.182と0.052であり、モデル2の方がモデル1よりもかなり小さく、その減少率は71.43%にも及ぶ。モデル1とモデル2の違いは、国レベルにおける5つの情意的側面の説明変数を持つかどうかであった。このことは5つの情意的側面を投入することによって、国別のテスト得点の差異が約7割も説明されることを意味する。つまり国レベルでテスト得点を見るならば、テスト得点に与える情意的側面の影響は決して小さくないと判断される。続いて、モデル2に個人レベルの変数を投入したモデル3をみると、モデル2の生徒間分散0.800よりもモデル3の値0.623の

表3.6 階層線形モデルにおける固定効果の分析結果

		モデル1		モデル2		モデル3		モデル4	
		固定効果	S.E	固定効果	S.E	固定効果	S.E	固定効果	S.E
切片	$\gamma_{00}$	0.031	0.068	0.031	0.037	0.031	0.037	0.031	0.052
国	興味・関心 $\gamma_{01}$			-0.201	0.261	-0.201	0.261	-0.418	0.191
レ	道具的動機付け $\gamma_{02}$			-0.235	0.207	-0.235	0.207	0.140	0.151
ベ	自己効力感 $\gamma_{03}$			0.222	0.187	0.222	0.187	-0.112	0.138
ル	自己概念 $\gamma_{04}$			-0.996	0.313 *	-0.996	0.314 *	0.088	0.231
	不安 $\gamma_{05}$			0.873	0.207 **	0.873	0.207 **	0.518	0.152 *
生	興味・関心 $\gamma_{10}$					-0.112	0.003 **	-0.100	0.012 **
徒	道具的動機付け $\gamma_{20}$					-0.016	0.002 **	-0.021	0.007 **
レ	自己効力感 $\gamma_{30}$					0.351	0.002 **	0.351	0.013 **
ベ	自己概念 $\gamma_{40}$					0.118	0.003 **	0.114	0.015 **
ル	不安 $\gamma_{50}$					0.110	0.002 **	0.101	0.009 **
	性別(男子) $\gamma_{60}$					-0.040	0.003 **	-0.036	0.003 **
	国別分散	0.182		0.052		0.052		0.130	
	生徒間分散	0.800		0.800		0.623		0.610	
	逸脱度	596619		596566		539593		535037	
	AIC	596629		596594		539695		535168	

注) \*: p<0.05 \*\*: p<0.01  
S.Eは標準誤差

出典：筆者作成

方が小さい。また、その減少率は22.13%である。これは生徒レベルの変数を投入したことで、生徒レベルのテスト得点のばらつきが約2割説明されることを意味する。しかしこの値は、国レベルにおける71.43%に比べるとずっと小さい。これは国別による差異に対して生徒の違い、つまり個人差を説明することがより複雑であると解釈でき、生徒レベルの考察に向けて他の視点を取り入れる必要がある。最後に、生徒レベルの変数にランダム効果を加えたモデル4をみると、モデル3の国別分散0.052よりもモデル4の値0.130の方が大きいことが分かる。これはランダム効果の投入によって、各国の差異をより捉えられていると考える。

これまで4つのモデルの国別分散と生徒間分散に目を向けてきた。では次に、構成した4つのモデルとデータの適合度を確認する。その当てはまりの良さを表す指標が、表3.6にある逸脱度とAICである。これらは、値が小さいほど適合度が高いモデルと判断される。そこで表3.6をみると、モデル4の逸脱度とAICの値がともに最も小さくなっている。したがって本分析では、4つのモデルのうちモデル4がデータと最も適合していると判断する。

階層線形モデルでは、モデル1とモデル4にある切片を比較することで、説明変数を投入した際に、目的変数である数学的リテラシー調査のテスト得点がどのように変化するかを調べることができる。4つのモデルに含まれる $u_{0j}$  ( $j=1,2,\dots,39$ )は、各国固有の特徴を反映するランダム効果であり、39カ国の平均的な値からのずれを意味する。ゆえにモデル1における $u_{0j}$ は、単純に各国のテスト得点の平均値を表わし、説明変数を投入した他のモデルにおいて $u_{0j}$ は、各変数の効果を統制した上で得られる新たなテスト得点と理解され得る値を算出できる<sup>25</sup>。ちなみに本分析では、生徒レベルの変数を各国の平均値に中心化させているため、各説明変数の級内相関係数が0.0に変換される。つまり説明変数が有する国別の差異が失われるため、説明変数の投入前と投入後、つまりモデル1とモデル4の間で大きな変化が生じる場合には、5つの情意的側面が与えるテスト得点に対する影響が国際的にみて大きい国と解釈できよう。

では、モデル1とモデル4に含まれる $u_{0j}$  ( $j=1,2,\dots,39$ )に目を向ける。表3.7にモデル1とモデル4における $u_{0j}$ を各国のテスト得点の平均値とともに提示した。なお斜体の値は、39カ国における $u_{0j}$ を標準化得点である。また差異の欄には、モデル1からモデル4の標準化得点を引き算した結果を示し、図3.3にモデル1とモデル4における各国の $u_{0j}$ の大きさを棒グラフに表した。この図をみると、モデル1ではその順位が上位層にあるフィンランドは、5つの情意的側面を統制した後のモデル4では大きく順位が後退することが分かる。統制後、つまり情意的側面の国別の差異が失われた場合にフィンランドは、大きく順位が下がることから、数学に対する情意的側面が国際的にみて数学学力に大きく寄与していると解釈できよう。またその逆の傾向を持つ場合には、39カ国の中で情意的側面が必ずしも大きな寄与を与えるとは限らない国と考えられる。そこで「差異」に示した値を参考に、2つのモデル間における各国の変動をみていく。差異の値が最も小さい香港を基準と

<sup>25</sup> 詳細は、宮崎（2006）や山崎・藤井・水野（2009）および川口（2009）およびを参照せよ。

表 3.7 モデル1とモデル4における $u_{0j}$ の大きさ (モデル4の降順)

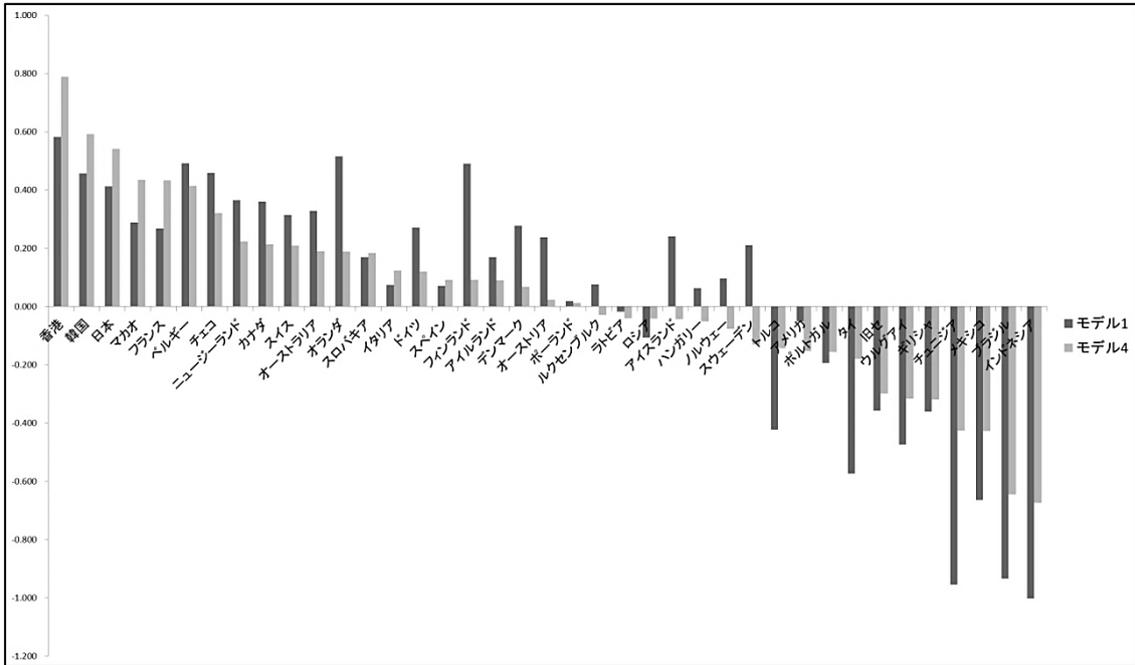
国名	平均値	モデル1	モデル4	差異		
香港	0.583	0.583	1.292	0.788	2.383	-1.090
韓国	0.457	0.457	0.996	0.592	1.766	-0.769
日本	0.413	0.413	0.895	0.541	1.603	-0.708
マカオ	0.289	0.288	0.603	0.434	1.269	-0.666
フランス	0.269	0.269	0.557	0.433	1.264	-0.707
ベルギー	0.492	0.493	1.081	0.414	1.204	-0.123
チェコ	0.459	0.459	1.002	0.321	0.913	0.089
ニュージーランド	0.365	0.365	0.782	0.224	0.608	0.174
カナダ	0.360	0.360	0.771	0.215	0.578	0.193
スイス	0.315	0.315	0.664	0.209	0.559	0.105
オーストラリア	0.329	0.329	0.697	0.191	0.502	0.195
オランダ	0.516	0.516	1.135	0.189	0.496	0.639
スロバキア	0.169	0.169	0.323	0.185	0.483	-0.160
イタリア	0.074	0.074	0.101	0.124	0.292	-0.190
ドイツ	0.272	0.272	0.564	0.120	0.281	0.283
スペイン	0.071	0.071	0.095	0.092	0.193	-0.098
フィンランド	0.491	0.491	1.077	0.091	0.190	0.886
アイルランド	0.169	0.169	0.324	0.090	0.185	0.139
デンマーク	0.278	0.278	0.579	0.069	0.118	0.460
オーストリア	0.239	0.239	0.486	0.024	-0.023	0.508
ポーランド	0.019	0.019	-0.028	0.012	-0.058	0.031
ルクセンブルク	0.075	0.075	0.104	-0.029	-0.188	0.292
ラトビア	-0.018	-0.018	-0.114	-0.040	-0.224	0.109
ロシア	-0.106	-0.106	-0.320	-0.042	-0.228	-0.092
アイスランド	0.242	0.242	0.493	-0.043	-0.233	0.726
ハンガリー	0.064	0.064	0.077	-0.051	-0.257	0.335
ノルウェー	0.097	0.097	0.154	-0.077	-0.339	0.493
スウェーデン	0.211	0.211	0.421	-0.131	-0.509	0.930
トルコ	-0.422	-0.422	-1.061	-0.144	-0.549	-0.511
アメリカ	-0.051	-0.051	-0.193	-0.153	-0.579	0.386
ポルトガル	-0.194	-0.194	-0.526	-0.156	-0.588	0.062
タイ	-0.573	-0.573	-1.415	-0.179	-0.662	-0.753
旧セ	-0.358	-0.358	-0.910	-0.299	-1.038	0.128
ウルグアイ	-0.473	-0.473	-1.180	-0.316	-1.090	-0.089
ギリシャ	-0.360	-0.360	-0.916	-0.319	-1.100	0.184
チュニジア	-0.956	-0.955	-2.309	-0.425	-1.436	-0.873
メキシコ	-0.663	-0.664	-1.626	-0.428	-1.443	-0.183
ブラジル	-0.934	-0.933	-2.258	-0.644	-2.124	-0.133
インドネシア	-1.001	-1.002	-2.417	-0.674	-2.218	-0.199

注) 旧セは、旧セルビア・モンテネグロ

出典：筆者作成

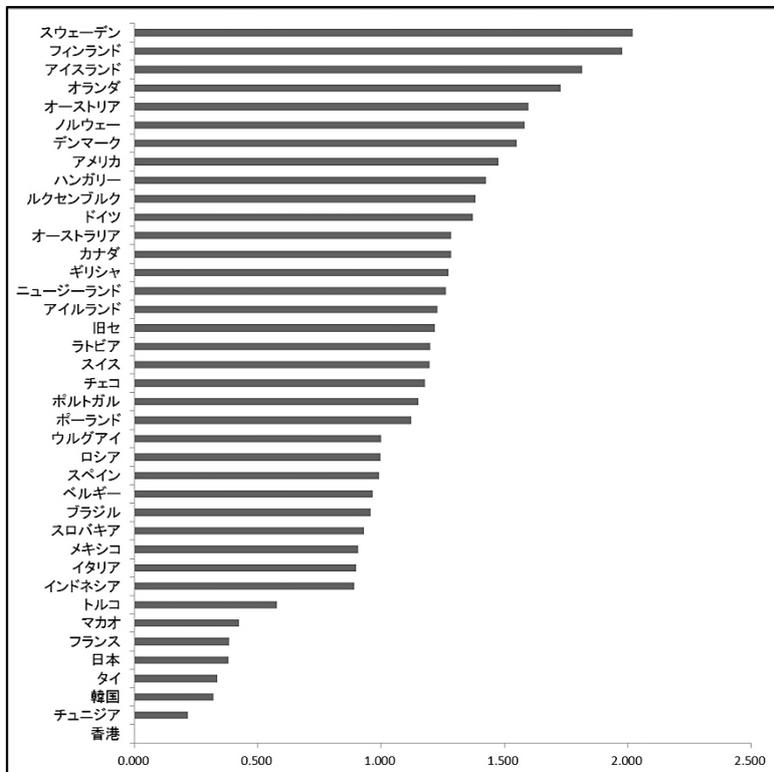
して、各国の5つの情意的側面の統制前と統制後の $u_{0j}$ の変化を図3.4に示した。その結果、スウェーデンやフィンランドあるいはアイスランドといった欧米諸国は、全体的に情意的側面が数学学力に寄与する傾向にあり、韓国や香港もしくは日本のアジア諸国や欧米諸国の中でもフランスは、国際的にみても逆の傾向を有することが分かった。またアジア諸国ほどではないにしても、テスト得点の水準が下位層に位置する途上国、具体的にはウルグアイ、ブラジル、メキシコ、インドネシア、トルコ、タイおよびチュニジアにおいても数学に対する情意的側面が国際的にみても数学学力に大きく寄与しない傾向にあるといえよう。

図 3.3 モデル1 とモデル4 における各国の  $u_{0j}$  の大きさ (モデル4 の降順)



出典：筆者作成

図 3.4 39 カ国における  $u_{0j}$  の相対的な変化の大きさ



出典：筆者作成

続いて切片だけでなく、モデル4を用いた分析結果の回帰係数に着目する(表3.6参照)。国レベルでは不安のみが統計的に有意であり、生徒レベルにおいては6つの変数で有意という結果を得た。まず「不安」はそのデータ処理から、不安の無さを表わす尺度であった。したがって不安の無さが国別のテスト得点に与える効果は、国レベルの分析結果0.518から生徒レベルの値0.101を引いた0.417となる。このように全体として正の効果を示すため、国レベルでは、数学に対する不安が少ない国ほど高いテスト得点を修める傾向にある。

続いて生徒レベルに目を移すと、自己効力感の値が0.351と正に最も大きい、また他の情意的側面の絶対値と比較しても特に大きな値となっている。ゆえに生徒レベルにおいては、「自己効力感」がテスト得点に大きな正の効果を与える、つまり正の方向に強い関連性を持つと解釈できる。なお自己概念と不安も正の値を示していることから、自己効力感に加えて自己概念と不安の無さが高い生徒ほど、平均的に高いテスト得点を修める傾向を読み取れる。他方で興味・関心と道具的動機付けについては、負の値となっていることから、それらが高い生徒であっても必ずしも高い成績を修めるとは限らないことも把握できる。

他方でモデル4には、その国固有の特徴を反映するランダム効果が6種類含まれていた。そこで各国の特徴をさらに調べるために、表3.8にランダム効果を含む切片と5つの回帰係数を提示する。なお表3.3と同様に、その国の39ヵ国中の順位を斜体の数値で表した。得られた回帰係数を比較することで、各国国内における5つの情意的側面の効果の違いが浮かび上がる。例えば日本をみると、自己効力感が0.468と国際的にみても大きな正の値を示し、他の情意的側面の値は0.0に近くなっている。ゆえに日本国内では、「自己効力感」がテスト得点で評価した数学学力に対して、正の方向に最も強い関連性を持つとともに、その程度は国際的に決して小さくないといえるだろう。

さて39ヵ国の全体的な傾向を調べるために、表3.8の記述統計量に目を向ける。まず平均値をみると、それらは表3.6のモデル4における生徒レベルの固定効果と一致することが分かる。これは、5つの情意的側面に関するランダム効果の平均値が0.0になるように推定されるためである。そこで39ヵ国における回帰係数の分布を調べるために、図3.5のように箱ひげ図を作成した。この図から分かるように「自己効力感」は、他と比べて突出した分布となっている。また表3.8と図3.5をみると、自己効力感のみ最小値が0.209と正の値を示している。したがって「自己効力感」は、39ヵ国全てに共通して、テスト得点に対し正の効果の有ることが浮かび上がった。

続いて、テスト得点への効果の大きさを把握するために、回帰係数の絶対値を各国国内で比較する。図3.5から読み取れるように、殆どの国で自己効力感の効果が最も大きい。しかしフィンランド、ノルウェー、メキシコ、ブラジルの4ヵ国において、自己効力感よりも大きい効果を持つ他の情意的側面の存在を確かめた。具体的に前者2ヵ国は、自己概念に最も大きい正の効果があり、後者2ヵ国は、興味・関心において負に大きな効果を有している。とはいえこの4ヵ国においても自己効力感の効果は、それらに続く2番目の大きさであり、その効果は国内で必ずしも小さいとはいえないだろう。したがって、ランダム効果を含む回帰係数を比較した結果、「自己効力感」は39ヵ国に例外なく、テスト得点で評価した数学学力に対し一定の大きさの正の効果の有する、つまり正の方向に強い関連性

を持つという国際的な傾向が明らかになった。

なおここでは、階層線形モデルを用いて回帰係数に含まれるランダム効果に関連性を代表させて、分析を進めてきた。ところが階層線形モデルは、近年教育研究においてその利

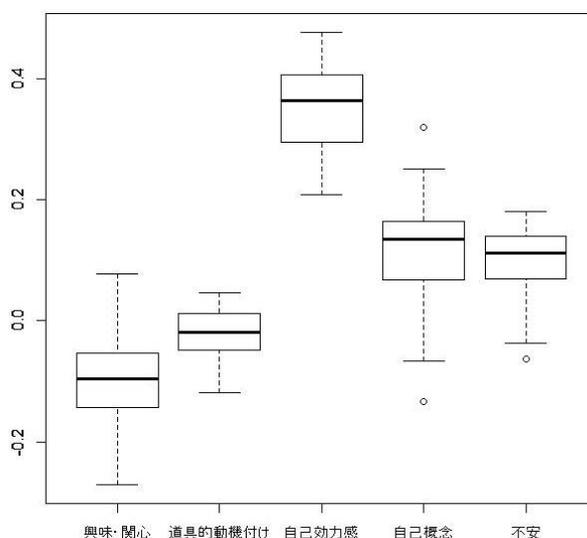
表 3.8 モデル 4 における 39 カ国の切片と回帰係数（興味・関心の降順）

国名	切片	興味・関心	道具的動機 付け	自己効力感	自己概念	不安
日本	0.541 3	0.077 1	-0.004 15	0.468 3	-0.067 38	-0.036 38
マカオ	0.434 4	0.005 2	-0.074 35	0.343 23	0.090 25	0.059 32
ベルギー	0.414 6	0.003 3	-0.013 18	0.364 20	0.007 34	0.123 16
香港	0.788 1	-0.007 4	-0.019 22	0.461 4	0.076 28	0.031 36
韓国	0.592 2	-0.012 5	0.029 4	0.398 12	0.163 12	-0.063 39
チェコ	0.321 7	-0.024 6	-0.082 37	0.477 1	0.072 29	0.137 14
フランス	0.433 5	-0.034 7	-0.029 25	0.383 14	0.102 23	0.055 33
ルクセンブルク	-0.029 22	-0.042 8	-0.052 31	0.374 16	-0.020 36	0.141 10
オランダ	0.189 12	-0.046 9	-0.070 34	0.423 8	-0.006 35	0.068 31
ハンガリー	-0.051 26	-0.050 10	-0.029 26	0.460 5	0.048 33	0.114 19
ドイツ	0.120 15	-0.056 11	-0.091 38	0.468 2	-0.045 37	0.152 6
インドネシア	-0.674 39	-0.067 12	-0.046 29	0.251 34	-0.133 39	0.127 15
ノルウェー	-0.077 27	-0.069 13	0.026 5	0.235 36	0.251 2	0.080 27
ギリシャ	-0.319 35	-0.072 14	-0.019 21	0.241 35	0.188 7	0.111 21
スイス	0.209 10	-0.080 15	-0.118 39	0.457 6	0.053 32	0.109 22
スロバキア	0.185 13	-0.083 16	-0.055 32	0.452 7	0.166 9	0.107 23
オーストリア	0.024 20	-0.086 17	-0.081 36	0.369 17	0.128 21	0.070 29
アイルランド	0.090 18	-0.088 18	-0.042 28	0.367 18	0.084 27	0.104 24
スペイン	0.092 16	-0.092 19	0.046 1	0.301 29	0.146 18	0.095 25
イタリア	0.124 14	-0.095 20	-0.011 17	0.415 9	0.063 31	0.153 4
スウェーデン	-0.131 28	-0.100 21	0.012 10	0.341 24	0.156 13	0.118 17
アイスランド	-0.043 25	-0.111 22	0.005 12	0.254 33	0.232 3	0.028 37
オーストラリア	0.191 11	-0.117 23	-0.024 23	0.412 10	0.154 16	0.075 28
ロシア	-0.042 24	-0.119 24	0.013 9	0.356 21	0.087 26	0.170 2
トルコ	-0.144 29	-0.129 25	-0.051 30	0.376 15	0.096 24	0.140 12
カナダ	0.215 9	-0.134 26	0.012 11	0.317 26	0.196 5	0.069 30
ウルグアイ	-0.316 34	-0.136 27	-0.037 27	0.301 28	0.139 19	0.149 8
フィンランド	0.091 17	-0.139 28	0.016 8	0.225 38	0.319 1	0.033 35
タイ	-0.179 32	-0.139 29	-0.003 14	0.288 30	0.064 30	0.084 26
デンマーク	0.069 19	-0.147 30	-0.017 19	0.307 27	0.219 4	0.116 18
旧セ	-0.299 33	-0.163 31	-0.024 24	0.285 31	0.135 20	0.147 9
ポーランド	0.012 21	-0.167 32	0.023 6	0.352 22	0.168 8	0.169 3
ニュージーランド	0.224 8	-0.177 33	-0.008 16	0.388 13	0.155 15	0.152 5
ラトビア	-0.040 23	-0.178 34	0.037 3	0.322 25	0.192 6	0.151 7
ポルトガル	-0.156 31	-0.178 35	0.023 7	0.402 11	0.148 17	0.112 20
チュニジア	-0.425 36	-0.181 36	0.043 2	0.233 37	0.156 14	0.039 34
アメリカ	-0.153 30	-0.183 37	0.000 13	0.365 19	0.127 22	0.139 13
メキシコ	-0.428 37	-0.212 38	-0.019 20	0.209 39	0.165 10	0.140 11
ブラジル	-0.644 38	-0.271 39	-0.069 33	0.258 32	0.164 11	0.180 1
平均値	0.031	-0.100	-0.021	0.351	0.114	0.101
標準偏差	0.318	0.070	0.040	0.078	0.090	0.055
中央値	0.024	-0.095	-0.019	0.364	0.135	0.112
最大値	0.788	0.077	0.046	0.477	0.319	0.180
最小値	-0.674	-0.271	-0.118	0.209	-0.133	-0.063
レンジ	1.462	0.349	0.165	0.269	0.453	0.244

注) 旧セは、旧セルビア・モンテネグロ

出典：筆者作成

図 3.5 5つの情意的側面におけるランダム効果を含む回帰係数の箱ひげ図



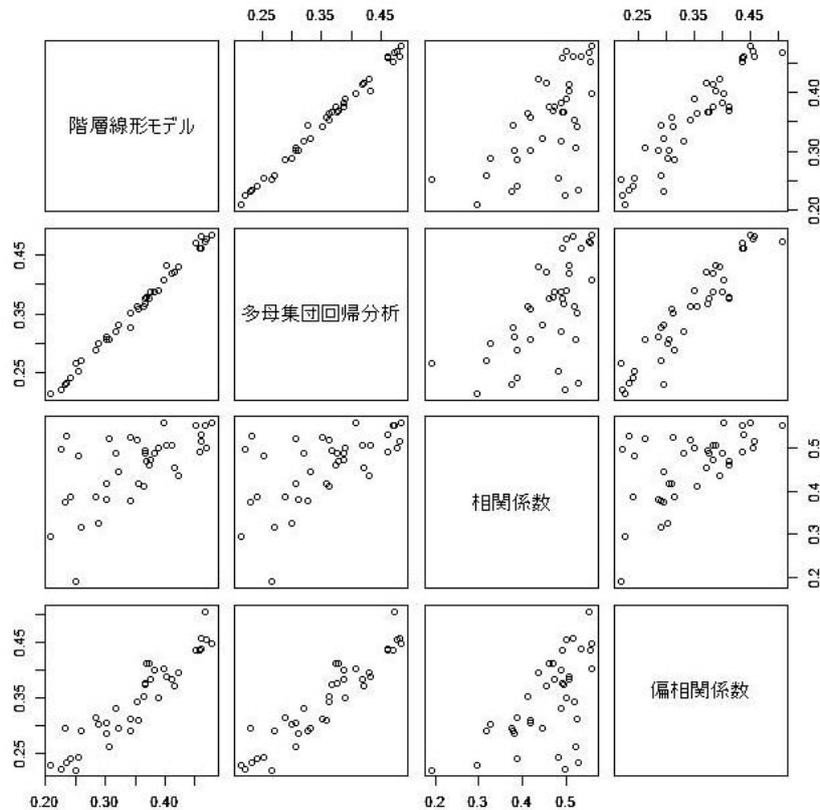
出典：筆者作成

用が推奨されてきた新しいデータ解析の手法である（宮崎、2007）。本分析のようにランダム効果に注目した場合、各国国内でテスト得点と強く正の方向に関連性が強いのは「自己効力感」という結果を得た。そこで、数学的リテラシー調査のテスト得点と5つの情意的側面の関連性を階層線形モデルだけでなく、テスト得点を目的変数、5つの情意的側面を説明変数に持つ多母集団回帰分析<sup>26</sup>およびテスト得点と5つの情意的側面の間の相関係数と偏相関係数を併せて試みた。多母集団回帰分析からは、階層線形モデルとほぼ同様な結果を得た。具体的には、フィンランドとノルウェー以外の国では、自己効力感における回帰係数が最も大きいというものである。また相関係数と偏回帰係数についても、ほとんどの国でテスト得点と自己効力感の間の値が最も大きいという結果となった。

階層線形モデル、多母集団回帰分析、相関係数および偏相関係数によるテスト得点を自己効力感の関連性に各国の分析結果を図3.6のように散布図で提示した。また表3.9に4つの分析方法で得られた結果間の相関係数を示した。階層線形モデルは、生徒が属するグループの違いを考慮して、各国の特徴を反映するランダム効果を算出する。そのため分析に含まれる設定が他の分析方法よりも幾分か複雑である。そこで階層線形モデルによる分析結果を基準とするならば、多母集団回帰分析を採用して分析を試みても、階層線形モデルと同様な結果を得ることができると判断できよう。また階層線形モデルや多母集団回帰分析を採用しなくとも、偏相関係数を確認することでも、図3.6の散布図と表3.9の相関係数から一定程度の信頼性を有する結果を得ることが可能と考える。

<sup>26</sup> 多母集団回帰分析については、豊田（1998）を参照されたい。また本分析では、Rに含まれる多母集団比較に対する共分散構造分析のパッケージ lavaan を使用した。

図 3.6 階層線形モデルと他の分析方法による結果の散布図



出典：筆者作成

表 3.9 5つの分析方法における相関係数

	階層線形 モデル	多母集団 回帰分析	相関係数	偏相関係数
階層線形モデル	1.000			
多母集団回帰分析	0.996	1.000		
相関係数	0.618	0.601	1.000	
偏相関係数	0.937	0.938	0.592	1.000

出典：筆者作成

### 3.2. 13 先進国における「自己効力感」の水準別に注目した分析

これまでの分析で、「自己効力感」は 39 カ国において例外なく、テスト得点で評価した数学学力に対して正の方向に強い関連性を有するという国際的な傾向が明らかになった。そこで数学学力と情意的側面の関連性をさらに調べるために、自己効力感に着目し、その水準の違いが数学的リテラシー調査のテスト得点や出題項目の難易度とどのような関係にあるかをいくつかの先進国に注目しながら検討する。

### 3.2.1. 異なる自己効力感の水準におけるテスト得点の分布と困難度に着目した解答パターン

本小節では、前章と同様の先進国、具体的にはオーストラリア、カナダ、フィンランド、フランス、ドイツ、香港、アイルランド、イタリア、日本、韓国、オランダ、ニュージーランド、アメリカの13カ国に焦点を当てて分析を試みる。そこで、1因子の因子分析によって算出した自己効力感における総合的な指標を降順に均等三分割し、上位層、中位層および下位層の3つの水準を設定する。表3.10に設定した3つの水準における13カ国のテスト得点の分布を平均値と標準偏差を用いて要約した。イタリアとアメリカに関しては、他の国の中位層や下位層が一つ高い水準の平均値を上回る場合もみられるが、テスト得点は、概ね3つの水準で階層的と考えられる。

次に、自己効力感の水準別でテストを構成する出題項目の難易度にどのような差異が生じるのかを検討する。一般に出題項目の難易度を表す指標には、正答率が使われる場合が多い。ところが正答率は、その項目そのものが持つ固有の難易度を表すのではなく、受験者集団にも依存して定められるという欠点を持つ(芝, 1991)。そこで難易度の比較をより妥当性を確保して行うために、テスト得点を算出する際に用いたラッシュモデル(1.2)に含まれる項目の難易度に関するパラメータである困難度に着目する。得られる各項目の困難度は、受験者集団に依らず算出されるもので、測定の尺度を等化することで比較可能な数値に変換される。なお困難度はその値が大きいほど、高い難易度の項目と判断される。

PISA2003の数学的リテラシー調査では、全部で84項目がその測定のために使用された。しかしここでは13カ国の3つの水準、つまり39の受験者集団を対象とした比較を行うために、集団ごとに算出される困難度に極端な値を示すものがないかを確認する必要がある。また本分析では、等化の精度をより高めるために、39の受験者集団間で分析対象とする項目を統一することにする。それに向けてまずOECD(2005a, p.190)によれば、韓国とイタリアおよびカナダで整合性が低いとされた項目が1つずつ指摘されている。また39受験者集団それぞれで81項目の困難度をラッシュモデルを用いて求めたところ、芝(1991)と豊田(2002b)の基準を満たさない項目が2つみつかった。本分析では、一定の比較の信頼性を確保するために、これらの項目も分析対象とはせず、最終的に残った79項目<sup>27</sup>を対象に分析を進めた。なお等化については、式(1.6)を用いる方法を採用し、日本の下位層集団を基準となる受験者集団とする。この方法によれば、各受験者集団における困難度の平均値が日本の下位層の値に揃えられる。そのため、もし各集団が同程度の困難度を持つならば、それらは誤差の許す範囲で近い値になるはずである。また等化後の得られる困難度は、79項目×39受験者集団の合計3081と膨大であるため、その提示は割愛する。

13カ国の3つの水準における困難度のパターンの全体像を掴むために、39×39の相関行列を作成し、国内の3つの水準間および他国の3つの水準との相関係数を確認した。国内

<sup>27</sup> M124Q03T、M155Q01、M603Q02、M704Q01T、M810Q03Tの5つを除いた79項目である。

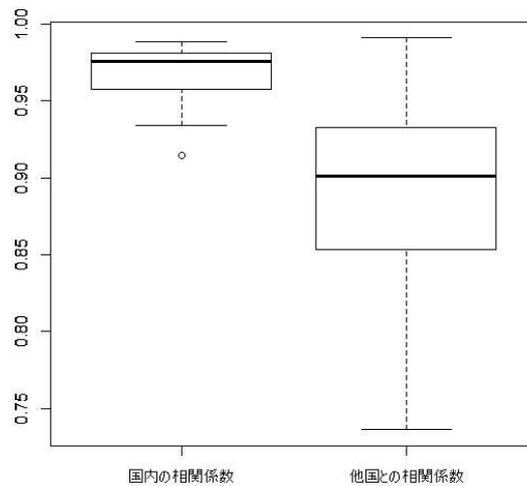
表 3.10 13 カ国の3つの水準におけるテスト得点の平均値と標準偏差（平均値の降順）

国別の自己効力感水準	平均値	標準誤差	標準偏差	標準誤差	受験者数	
香港	上位	1.126	0.019	0.730	0.014	1451
日本	上位	1.047	0.020	0.783	0.014	1488
韓国	上位	1.043	0.018	0.763	0.013	1764
フィンランド	上位	0.990	0.018	0.778	0.013	1821
オランダ	上位	0.968	0.024	0.809	0.017	1177
ニュージーランド	上位	0.915	0.024	0.883	0.017	1366
オーストラリア	上位	0.890	0.014	0.847	0.010	3847
カナダ	上位	0.870	0.010	0.805	0.007	6487
フランス	上位	0.788	0.023	0.811	0.016	1280
ドイツ	上位	0.779	0.023	0.825	0.016	1333
アイルランド	上位	0.656	0.022	0.759	0.016	1167
香港	中位	0.651	0.022	0.826	0.015	1452
イタリア	上位	0.571	0.015	0.879	0.010	3621
オランダ	中位	0.486	0.024	0.830	0.017	1178
アメリカ	上位	0.485	0.021	0.874	0.015	1671
韓国	中位	0.478	0.018	0.774	0.013	1765
フィンランド	中位	0.477	0.018	0.754	0.012	1822
日本	中位	0.439	0.021	0.829	0.015	1488
カナダ	中位	0.339	0.010	0.803	0.007	6488
ニュージーランド	中位	0.322	0.024	0.876	0.017	1366
オーストラリア	中位	0.313	0.014	0.855	0.010	3848
ドイツ	中位	0.298	0.023	0.827	0.016	1334
フランス	中位	0.245	0.023	0.816	0.016	1281
アイルランド	中位	0.140	0.023	0.778	0.016	1168
オランダ	下位	0.091	0.024	0.819	0.017	1177
イタリア	中位	0.075	0.014	0.838	0.010	3622
香港	下位	0.016	0.024	0.921	0.017	1451
フィンランド	下位	0.012	0.017	0.729	0.012	1821
アメリカ	中位	-0.110	0.020	0.808	0.014	1672
カナダ	下位	-0.128	0.009	0.742	0.007	6488
韓国	下位	-0.152	0.019	0.803	0.014	1764
ニュージーランド	下位	-0.166	0.021	0.783	0.015	1366
日本	下位	-0.196	0.023	0.881	0.016	1488
フランス	下位	-0.218	0.021	0.762	0.015	1281
オーストラリア	下位	-0.227	0.013	0.802	0.009	3848
ドイツ	下位	-0.283	0.022	0.799	0.015	1334
アイルランド	下位	-0.304	0.021	0.707	0.015	1168
イタリア	下位	-0.416	0.013	0.809	0.010	3621
アメリカ	下位	-0.524	0.018	0.743	0.013	1671

出典：筆者作成

における 39 個の相関係数と他国との 702 個の相関係数を区別し、その分布を図 3.7 のように箱ひげ図に要約した。国内をみると、他国との相関係数よりも中央値が極端に大きく、レンジも相当に小さい。つまり国内の 3 つの水準における相関係数は、13 カ国とも他国との間にある値よりも大きいといえるだろう。実際、国内の相関係数で最も高い値を示したのは、ドイツの上位層と中位層の間の 0.988 であり、逆に最も低かったのは、韓国の上位層と下位層の 0.915 となった。国内においては、全ての相関係数が 0.900 以上を示すが、他国との相関係数に目を向けると、中央値が 0.901 であることから約半数が 0.900 を超え

図 3.7 39 受験者集団における困難度の相関係数の分布



出典：筆者作成

ることが分かる。なお他国との相関係数の最大値と最小値は、それぞれカナダとニュージーランドの下位層間の 0.991 とイタリアの下位層と韓国の上位層間の 0.736 である。この分析結果のように 39 受験者集団の特徴を調べたところ、各国とも 3 つの水準間でテスト得点の分布が相違するにも拘らず、困難度に注目した解答パターンは、国内の 3 つの水準間でより類似する傾向が浮かび上がった。

他方で数学的リテラシー調査の出題項目は、「包括的アイデア（数学的な内容）」、「数学用いられる状況・文脈」、「能力（数学的な過程）」の 3 つの評価枠組みに関する領域によって特徴付けられており、出題形式も自由記述形式や選択肢形式など様々である。そこで PISA が定める 3 つの評価枠組みの領域（以下、領域）と出題形式の種類別（以下、種別）を用いて、39 の受験者集団が有する困難度に注目した解答パターンを分析する。

まず本分析の流れを明確にするために、日本における分析結果を表 3.11<sup>28</sup>に提示する。この表にある「水準別平均値」には、各領域と種別における 3 つの水準の平均値を示し、その最大値から最小値を引いた値を「水準別比較」は意味している。また「全体平均値」には、3 つの水準ごとではなく 3 つの水準を合わせた全体の平均値を表わしている。この定義から、もし水準別比較の値が 0.0 に近いならば、それは自己効力感の 3 つの水準における難易度の差異が小さい領域または種別であること意味し、大きい場合にはその逆と解釈され得る。つまり水準別比較は、自己効力感の 3 つの水準にある差異の大きさを表わす指標として設定できる。表 3.11 の水準別比較をみると、最も小さい値は「空間と形」の 0.015 であり、最も大きいものは「自由記述」の 0.233 となっている。また「複合的選択肢」においても、0.174 と比較的大きい値となった。そこで、種別ごとの困難度を図 3.8 のように折れ線グラフで表した。水準別比較が最も大きかった「自由記述」は、水準が上位層である

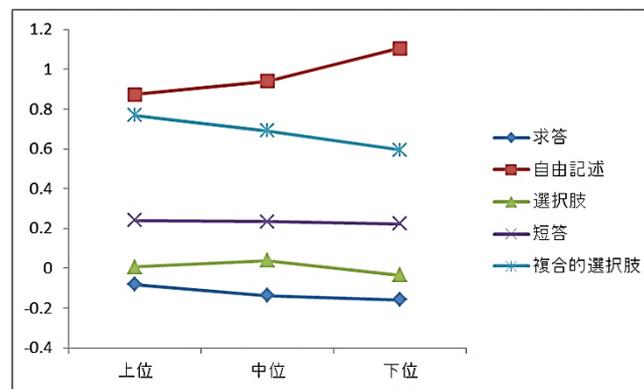
<sup>28</sup> 各領域と種別の規定は、付録 3.2 を参照いただきたい。

表 3.11 日本における自己効力感の水準別の困難度の平均値

領域と種別 ( )内は項目数	水準別平均値			水準別 比較	全体平 均値
	上位	中位	下位		
【包括的アイデア】				*	**
空間と形 (20)	0.370	0.384	0.369	0.015	0.374
不確実性 (19)	0.555	0.580	0.540	0.040	0.558
量 (21)	0.072	0.074	0.043	0.030	0.063
変化と関係 (19)	0.418	0.377	0.466	0.089	0.420
【状況・文脈】					
科学的 (15)	0.781	0.745	0.758	0.036	0.761
教育的 (15)	0.266	0.258	0.247	0.019	0.257
公共的 (27)	0.290	0.304	0.363	0.072	0.319
私的 (16)	0.084	0.093	0.013	0.079	0.063
職業的 (5)	0.026	0.032	-0.038	0.070	0.007
【能力】					
関連付け (35)	0.501	0.464	0.418	0.084	0.461
再現 (26)	-0.242	-0.234	-0.211	0.032	-0.229
熟考 (18)	0.897	0.958	1.015	0.117	0.957
【出題形式】					
求答 (13)	-0.081	-0.137	-0.159	0.077	-0.126
自由記述 (17)	0.873	0.941	1.106	0.233	0.973
選択肢 (17)	0.007	0.039	-0.034	0.073	0.004
短答 (21)	0.241	0.234	0.224	0.017	0.233
複合的選択肢 (11)	0.768	0.692	0.594	0.174	0.684

出典：筆者作成

図 3.8 日本の出題形式における困難度の水準別比較



出典：筆者作成

ほど難易度が低くなり、「複合的選択肢」は逆に高くなることから「自由記述」では、高い水準ほど正答を得る確率が高くなると考えられ、「複合的選択肢」ではその逆の傾向を指摘できるだろう。ここで注目すべきことは、水準が下位層になるほど「自由記述」の難易度が際立つという特徴である。全体平均値については、この値が高いほど難易度が高い領域または種別となるため、0.973と値が最も大きい「自由記述」は、最も難しいと判断される。ゆえに「自由記述」は、自己効力感の水準が低くなるほど難易度が目

立って高くなり、全体的にも難しい種別といえるだろう。

これと同様な分析を他の12カ国にも行った。それを経て得られる各国の「水準別比較」を表3.12<sup>29</sup>に示した。また、表3.13に各領域と種別ごとの各国の「全体平均値」を記し、この表にある「13カ国間比較」の欄に13カ国中の最大値から最小値を引いた結果を提示した。この13カ国間比較は、その算出方法から自己効力感の水準別ではなく、各領域あるいは種別における国別の差異の大きさを表わす指標として扱える。

まず、表3.12で示した国内水準別比較に目を向けると、日本において最も自己効力感の水準間で差異の大きかった「自由記述」は、香港を例外とすれば、他の国においても同様に最も高い値であることが分かる。また「熟考」も各国で比較的に高い値と読み取ることできるだろう。他方で表3.13の「自由記述」における13カ国間比較をみると、平均値や中央値と同程度であるため、必ずしも特に高い種別とは言い切れない。つまり「自由記述」は、全体的な国別の差異は必ずしも大きくはないが、国内における「自己効力感」の水準別では看過できない差異が認められる種別という国際的な傾向を主張できよう。続い

表3.12 13カ国における水準別比較の国際比較（一の位の0は省略）

領域と種別 ( )内は項目数	AUS	CAN	DEU	FIN	FRA	HKG	IRL	ITA	JPN	KOR	NLD	NZL	USA
【包括的アイデア】									*				
空間と形 (20)	.043	.070	.088	.023	.072	.060	.005	.030	.015	.075	.070	.001	.075
不確実性 (19)	.015	.039	.011	.067	.023	.014	.034	.055	.040	.033	.016	.036	.043
量 (21)	.072	.081	.125	.066	.107	.032	.020	.062	.030	.087	.056	.083	.118
変化と関係 (19)	.043	.032	.050	.091	.027	.042	.048	.060	.089	.040	.005	.072	.091
【状況・文脈】													
科学的 (15)	.032	.034	.019	.049	.070	.108	.046	.015	.036	.040	.045	.020	.021
教育的 (15)	.057	.010	.022	.034	.053	.008	.019	.045	.019	.047	.049	.045	.037
公共的 (27)	.087	.075	.066	.102	.036	.088	.056	.066	.072	.066	.049	.088	.077
私的 (16)	.110	.116	.163	.144	.115	.055	.060	.113	.079	.136	.053	.150	.105
職業的 (5)	.039	.048	.115	.036	.037	.112	.066	.046	.070	.074	.059	.035	.065
【能力】													
関連付け (35)	.039	.016	.005	.013	.018	.054	.034	.010	.084	.059	.024	.029	.032
再現 (26)	.029	.062	.031	.074	.098	.022	.023	.037	.032	.017	.028	.051	.048
熟考 (18)	.119	.121	.054	.131	.114	.124	.034	.074	.117	.140	.088	.130	.079
【出題形式】													
求答 (13)	.237	.147	.060	.156	.132	.193	.119	.108	.077	.144	.189	.184	.067
自由記述 (17)	.336	.241	.281	.308	.267	.173	.215	.267	.233	.236	.226	.285	.251
選択肢 (17)	.086	.076	.128	.099	.042	.069	.063	.104	.073	.035	.020	.063	.104
短答 (21)	.006	.035	.013	.043	.065	.014	.045	.045	.017	.029	.044	.038	.077
複合的選択肢 (11)	.141	.070	.180	.102	.069	.119	.181	.084	.174	.189	.063	.088	.047
平均値	.088	.075	.083	.090	.079	.076	.063	.072	.074	.085	.064	.082	.079
中央値	.057	.070	.060	.074	.069	.060	.046	.060	.072	.066	.049	.063	.075
最大値	.336	.241	.281	.308	.267	.193	.215	.267	.233	.236	.226	.285	.251
最小値	.006	.010	.005	.013	.018	.008	.005	.010	.015	.017	.005	.001	.021
レンジ	.331	.231	.276	.295	.249	.185	.210	.257	.218	.219	.221	.284	.230

出典：筆者作成

<sup>29</sup> OECD が使用している国ラベルを用いて、国名を記載した。

表 3.13 13 カ国における全体平均値の国際比較（一の位の0は省略）

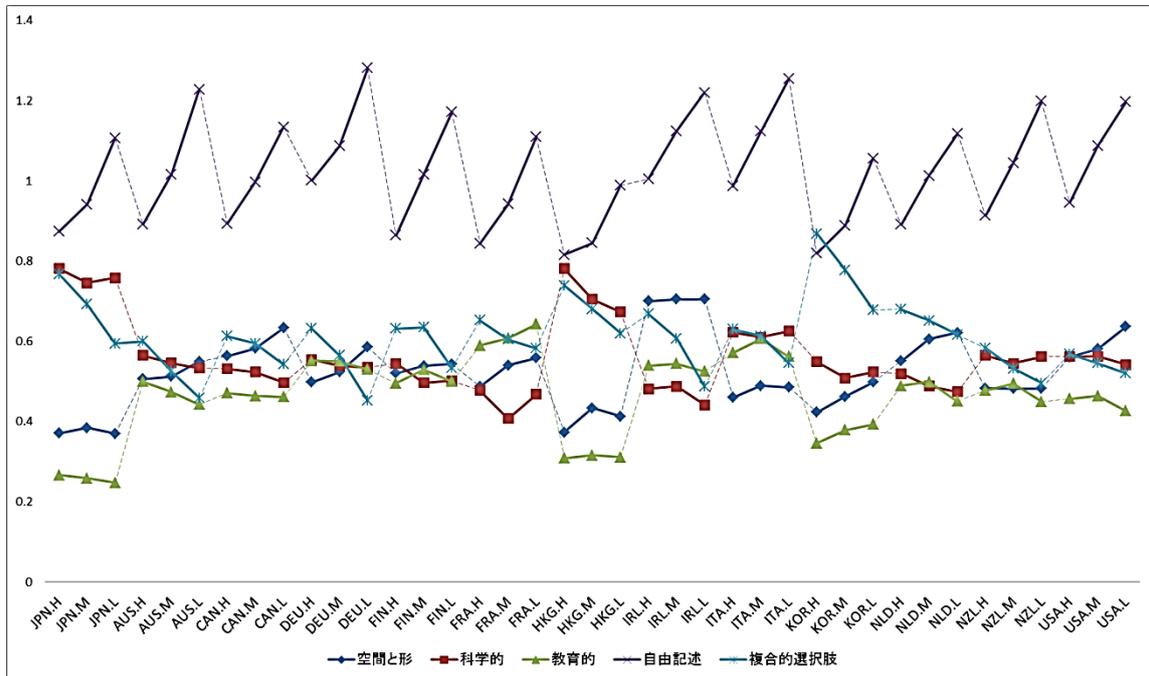
領域と種別	AUS	CAN	DEU	FIN	FRA	HKG	IRL	ITA	JPN	KOR	NLD	NZL	USA	13カ国間比較
【包括的アイデア】									**					
空間と形	.522	.593	.535	.534	.528	.406	.703	.478	.374	.461	.592	.482	.593	.329
不確実性	.429	.387	.590	.495	.522	.402	.330	.502	.558	.503	.357	.434	.412	.260
量	.078	.030	-.079	-.049	.036	.026	.003	-.053	.063	.075	.079	.117	.030	.196
変化と関係	.377	.398	.377	.440	.324	.584	.368	.495	.420	.371	.376	.372	.373	.260
【状況・文脈】														
科学的	.548	.517	.543	.514	.451	.719	.470	.619	.761	.527	.494	.557	.555	.311
教育的	.472	.465	.544	.508	.613	.311	.536	.580	.257	.372	.479	.473	.449	.356
公共的	.278	.303	.268	.263	.236	.217	.278	.231	.319	.354	.296	.271	.353	.137
私的	.058	.065	.034	.100	.137	.137	.092	.055	.063	.115	.054	.069	.013	.124
職業的	.174	.129	.153	.106	.132	.302	.103	.029	.007	.195	.183	.175	.095	.295
【能力】														
関連付け	.443	.461	.465	.455	.436	.460	.458	.450	.461	.482	.482	.444	.452	.046
再現	-.195	-.227	-.266	-.258	-.223	-.195	-.209	-.217	-.229	-.236	-.314	-.218	-.197	.118
熟考	.943	.955	1.003	1.010	.996	.909	.933	.961	.957	.927	1.039	.973	.927	.129
【出題形式】														
求答	-.098	-.026	-.133	-.042	-.091	.169	-.065	-.052	-.126	.016	-.053	-.129	-.011	.302
自由記述	1.045	1.008	1.123	1.017	.965	.883	1.117	1.121	.973	.921	1.007	1.052	1.076	.240
選択肢	.133	.137	.138	.125	.136	.002	.118	.093	.004	-.083	.062	.118	.026	.220
短答	.135	.089	.079	.092	.149	.128	.038	.042	.233	.210	.133	.156	.134	.195
複合的選択肢	.528	.583	.549	.600	.613	.680	.588	.596	.684	.775	.649	.537	.544	.247
平均値	.345	.345	.348	.348	.351	.361	.345	.349	.340	.352	.348	.346	.343	.222
中央値	.377	.387	.377	.440	.324	.311	.330	.450	.319	.371	.357	.372	.373	.240
最大値	1.045	1.008	1.123	1.017	.996	.909	1.117	1.121	.973	.927	1.039	1.052	1.076	.356
最小値	-.195	-.227	-.266	-.258	-.223	-.195	-.209	-.217	-.229	-.236	-.314	-.218	-.197	.046
レンジ	1.241	1.235	1.389	1.275	1.219	1.105	1.325	1.338	1.203	1.162	1.352	1.270	1.273	.311

出典：筆者作成

て、表 3.13 にある 13 カ国間比較に注目すると、「教育的」、「科学的」、「空間と形」および「求答」においてその値が高い結果を得た。その一方でこれらの領域と種別の表 3.12 における国内水準別比較は、各国の平均値や中央値を参考にすると、決して大きな差異があるとは読み取れない。特に「教育的」において、全体的に差異が現れやすい傾向にあるといえる。つまり上述の 4 つの領域と種別では、「教育的」がより明示的であるが、国内の自己効力感の水準別よりも国別における差異が反映されやすいと判断できるだろう。

これまでの分析結果を纏めると、困難度に着目した解答パターンが顕在化する国際的な傾向として、①国内における自己効力感の水準別に解答パターンが顕在化しやすいのは「自由記述」であること、②自己効力感の水準別ではなく国別でそれが現れやすい場合があり、それは「教育的」、「科学的」あるいは「空間と形」という領域であることが浮かび上がった。そこで図 3.9 に、「自由記述」、「教育的」、「科学的」、「空間と形」に加えて「複合的選択肢」における 13 カ国の 3 つの水準の平均値を折れ線グラフで示した。なおこの図では OECD が使用している国ラベルを用いて国名を示し、国名と共に記した H と M および L は、それぞれ上位層、中位層、下位層を意味する。①に関して「自由記述」は、どの国においても自己効力感の水準が低くなるほど、目立って高い困難度の平均値、つまり難易度

図 3.9 「自由記述」、「教育的」、「科学的」、「空間と形」および「複合的選択肢」における困難度の国内水準別比較



出典：筆者作成

が高くなるという特徴を把握できる。また②について国別の方に差異が認められる「科学的」をみると、日本や香港で難易度が高い領域であることを読み取れる（図 3.9 参照）。ところがこの 2 カ国において「教育的」は、逆に易しい領域といえよう。このように、自己効力感の水準に依らない形で顕在化する国際的な傾向も同時に浮かび上がってきた。

### 3.2.2. 日本に注目した分析結果の総合的考察

日本に焦点を絞りながら、得られた分析結果から数学学力と数学に対する情意的側面の関連性について考察する。PISA2003 で調査された 5 つの情意的側面ごとに総合的な指標を算出したところ、日本は、全ての内容において 39 カ国で最も低い水準にあることを確かめた（表 3.3）。その一方で、数学的リテラシー調査のテスト得点の水準は国際的にみて高いことも確認した（表 3.1）。これらは、相反する結果として捉えることができるだろう。ところが情意的側面に関する結果の背景には、数学に対して同程度に肯定的であったとしても、その評価が異なって顕在化する可能性、つまり異なる文化圏にみられる価値判断の違いが考えられる（田崎、2008、p.63）。そのため、これらの値を直接比較するだけでなく、各国で評価の基準を統一して算出されるテスト得点を介して国際比較するなどの工夫が必要になる。実際本章では、日本の情意的側面の水準は国際的にみて低いけれども、数学的リテラシー調査のテスト得点、つまり本研究における数学学力との関連性は、全体的にス

ウェーデンやフィンランドといった欧米諸国と比べると決して大きくない(図 3.4)が、「自己効力感」という側面では、むしろ国際的にみて強いことを示した(表 3.8)。そして、自己効力感の水準の違いが出題項目の難易度の観点から最も反映されやすいのは、「自由記述」であることを明らかにした(表 3.12)。

他方で国際比較の視座に立つと、自己効力感がテスト得点に対して正に強い関連性を有することは、分析対象とした 39 カ国に共通する傾向の一つである(表 3.8 と図 3.5)。また自己効力感の水準の違いが「自由記述」に反映されるという特徴も、日本だけでなく、13 先進国に共通するものである(図 3.9)。したがって、自己効力感という情意的側面とテスト得点で評価した数学学力の関連性、および自己効力感の水準の違いによる自由記述形式問題の難易度の変化について、一定の普遍性を示す結果を得たことになる。

これらの分析結果を日本の文脈で考えると、今日議論されている「数学嫌い」に集約されるような生徒の情意的側面の改善に向けて、一つの示唆を投げかけている。それは、数学学力との関連を意識ながら情意的側面の改善を試みるならば、数学に対する興味・関心というよりも自己効力感、言い換えると「数学に対して自信をつける」という観点から取り組むべきである、と主張できることである。ちなみに日本の「自由記述」における無答率は、2003 年以降実施された数学的リテラシー調査で一貫して国際平均よりも高いことが知られている(国立教育政策研究所、2010;室岡、2006)。本分析では、正答部分のみに焦点を当てているため、誤答の内実に触れることはできない。しかし「自由記述」の難易度が他の国と比べて目立って高くない(表 3.13)ことから、日本の生徒が解答した場合の正答率は高いと思われる。ゆえに高い無答率の背景には、自己効力感の低さが関係していると推測される。もし無答率の高さを改善しようとするならば、自己効力感の向上に焦点を置く取り組みは効果的であると示唆される。

加えて本小節では、日本を含めた 13 カ国に注目して分析を進めてきた。この 13 カ国における自己効力感の 3 つの水準別に数学的リテラシー調査のテスト得点の平均値と標準偏差をみると、その水準は同程度であるとは言い難い(表 3.10)。それにも拘らず、39 の受験者集団で困難度に注目した解答パターンを調べたところ、国内の水準別における類似性の方が国別よりも強く(図 3.7)、平均的にも国別の方に解答パターンの違いが顕在化しやすいことが浮かび上がった(表 3.12 と表 3.13)。つまり、テスト得点の分布が大きく異なる受験者集団よりも、その水準がより同程度と思われる国別の方に解答パターンの差異が現れやすいことになる。このことから自己効力感の水準別だけでなく、日本の全体的な特徴を掴む副次的な分析が可能であろう。鈴川・豊田・川端(2008)が指摘したように、本分析では、日本の生徒は数学の授業で典型的にみられるような疑似現実的な内容が含まれる「教育的」な文脈の問題を国際的にみて難無くこなす一方で、より抽象的とされる「科学的」な文脈では、他の国と比べて難易度が高いことが明らかになった(図 3.9)。「科学的」については、技術的な過程、理論的な場面と明らかに数学的な問題についての理解(国立教育政策研究所、2004)と規定されるように、数学教育の中でその育成が期待される中心

的なものの一つであろう。ところ日本の生徒は、そのような文脈で数学を活用していく能力が同程度のテスト得点の水準にある国と比べて弱いことを指摘できる。なお図 3.9 から分かるように、日本と同様の傾向が香港においても確認されよう。

### 3.3. 7 途上国における「自己効力感」の水準別に注目した分析

分析の結果、「自己効力感」は 39 ヶ国において例外なく、テスト得点で評価した数学学力に対して正の方向に強い関連性を有するという国際的な傾向が明らかになった。そこで数学学力と情意的側面の関連性をさらに調べるために、「自己効力感」に着目し、その水準の違いが数学的リテラシー調査のテスト得点や出題項目の難易度とどのような関係にあるかをいくつかの途上国を対象を絞りながら検討する。なお階層線形モデルを用いて、数学学力と 5 つの情意的側面の関連性について調べた結果、例えば日本では、自己効力感に目立った正の効果が認められ、他の情意的側面の効果は国際的にみて決して大きくないことを指摘した。その一方でメキシコやブラジルといった途上国では、自己効力感だけでなく「自己概念」や「不安」においても一定の正の効果を示しており、「興味・関心」とテスト得点の関連性についても日本とは異なる実態を読み取れる（表 3.7）。このように得られた分析結果を個々にみるならば、各国固有の特徴をさらに詳しく捉える分析が十分に可能と思われる。しかし本章では、「自己効力感」が数学学力に対して正の方向に強い関連性を有するという国際的な傾向に着目し、先進国と低い学力と称され得る途上国に共通する特徴の把握という立場から分析を進めていく。

#### 3.3.1. 異なる自己効力感の水準におけるテスト得点の分布と困難度に注目した解答パターン

本小節では、前章と同様の途上国、具体的にはトルコ、ウルグアイ、タイ、メキシコ、ブラジル、チュニジア、インドネシアの 7 ヶ国に焦点を当てながら数学学力と数学に対する情意的側面の関連性を検討する。そこで、1 因子の因子分析によって求めた自己効力感の総合的な指標となる因子得点を降順に均等三分割し、前小節と同様に上位層、中位層、下位層の 3 つの水準を設定する。表 3.14 に設定した 3 つの水準における 7 ヶ国のテスト得点の分布を平均値と標準偏差を用いて要約した。チュニジアとブラジルおよびインドネシアに関しては、他の国の中位層や下位層が一つ高い水準の平均値を上回る場合もみられるが、テスト得点は、概ね 3 つの水準で階層的と考えられる。

次に、自己効力感の水準の違いによってテストを構成する出題項目の難易度にどのような差異が生じるのかを検討する。本小節でも前小節と同様に、より妥当性を確保して比較を行うために、テスト得点を算出するときに用いたラッシュモデル (1.2) に含まれる項目の難易度に関するパラメータである困難度に注目する。得られる各項目の困難度は、受験者集団に依らず算出されるもので、測定の尺度を等化することで比較可能な数値に変換される。

表 3.14 7カ国の3つの水準におけるテスト得点の平均値と標準偏差（平均値の降順）

国別の自己効力感水準	平均値	標準誤差	標準偏差	標準誤差	受験者数	
トルコ	上位	0.146	0.029	1.026	0.020	1293
ウルグアイ	上位	-0.049	0.027	1.020	0.019	1478
タイ	上位	-0.289	0.024	0.960	0.017	1644
メキシコ	上位	-0.388	0.009	0.842	0.007	8386
ウルグアイ	中位	-0.491	0.023	0.879	0.016	1479
トルコ	中位	-0.550	0.023	0.813	0.016	1293
チュニジア	上位	-0.584	0.029	0.868	0.020	926
タイ	中位	-0.586	0.020	0.825	0.014	1645
ブラジル	上位	-0.651	0.030	0.986	0.021	1117
メキシコ	中位	-0.703	0.008	0.720	0.006	8386
ウルグアイ	下位	-0.849	0.020	0.768	0.014	1478
インドネシア	上位	-0.860	0.015	0.816	0.011	2951
トルコ	下位	-0.889	0.019	0.695	0.014	1293
タイ	下位	-0.892	0.017	0.674	0.012	1645
メキシコ	下位	-0.913	0.007	0.657	0.005	8386
チュニジア	中位	-0.977	0.023	0.705	0.016	926
ブラジル	中位	-0.978	0.023	0.780	0.017	1117
インドネシア	中位	-1.027	0.013	0.719	0.009	2951
インドネシア	下位	-1.154	0.011	0.624	0.008	2951
ブラジル	下位	-1.237	0.019	0.647	0.014	1117
チュニジア	下位	-1.286	0.020	0.601	0.014	926

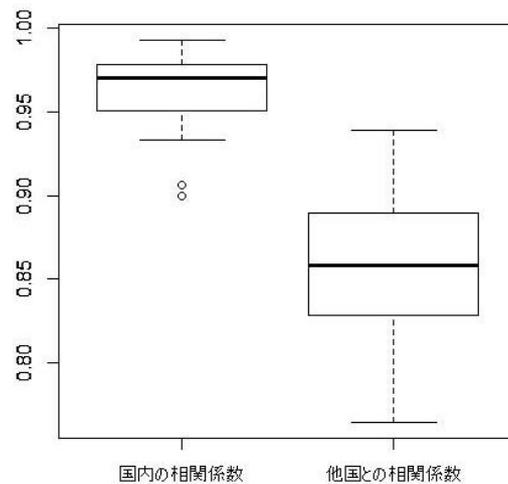
出典：筆者作成

PISA2003の数学的リテラシー調査では、全部で84項目が調査項目として使用された。しかしここでは、7カ国の3つの水準、つまり21の受験者集団を対象とするため、集団ごとに算出される困難度に極端な値を示すものがないかを確認する必要がある。それに向けOECD（2005a）によれば、ウルグアイにおいて整合性の低い項目が2つ指摘されている。21の受験者集団で項目数を揃えて比較を行うために、この2つの項目は分析対象に含めなかった。続いて、各集団における82項目の困難度を算出したところ、いくつかの集団で困難度の値の絶対値が4.0を超えていた。このような芝（1991）と豊田（2002b）の基準を満たさなかった項目も対象としないこととし、本研究では、最終的に残った76項目<sup>30</sup>を対象に分析を進める。なお等化については、式（1.6）を用いる方法を採用する。この方法によれば、各集団で得られる76項目の困難度の平均値が同一の値に揃えられる。そのため、もし各集団が同程度の困難度を持つならば、それらは誤差の許す範囲で近い値になるはずである。なお等化後に得られる困難度は、76項目×21受験者集団の合計1596と大きな数になるため、その提示は割愛する。また本分析では、ブラジルの下位層に等化を施した。

7カ国の3つの水準における困難度のパターンの全体像を掴むために、21×21の相関行列を作成し、国内の3つの水準間および他国の3つの水準との相関係数を確認した。国内における21の相関係数と他国との189の相関係数を区別し、その分布を図3.10のように

<sup>30</sup> M124Q03T、M155Q03T、M442Q02、M446Q02、M462Q01T、M513Q0、M800Q01、M810Q03Tの8つを除いた76項目である。

図 3.10 21 の受験者集団における相関係数の分布



出典：筆者作成

箱ひげ図に要約した。国内をみると、他国との相関係数よりも中央値が大きく、レンジは小さくなっている。つまり国内の3つの水準における相関係数は、7カ国とも他国との間にある値よりも大きいといえるだろう。実際、国内の相関係数で最も高い値を示したのは、メキシコの下位層と中位層の間の0.993であり、逆に最も低かったのはチュニジアの下位層と上位層の0.900であった。国内においては、全ての相関係数が0.900以上を示すが、他国との相関係数に目を向けると、それを超えるものは189の相関係数のうち29個で全体の15.34%ほどであった。したがって21受験者集団の特徴を調べた結果、各国とも3つの水準間でテスト得点の分布が相違するにも拘らず、困難度のパターンは、国内の3つの水準間でより類似する傾向にあるといえよう。

数学的リテラシー調査の出題項目は、「包括的アイディア（数学的な内容）」、「数学が用いられる状況・文脈」、「能力（数学的な過程）」の3つの評価枠組みに関する領域によって特徴付けられており、出題形式も自由記述形式や選択肢形式など様々であった。そこでPISAが定める3つの評価枠組みの領域（以下、領域）と出題形式の種類別（以下、種別）を用いて、21の受験者集団が有する困難度に注目した解答パターンを分析する。

例えば、ブラジルにおける分析結果を表3.15に示す。この表にある「水準別平均値」には、前小節と同様に各領域と種別における3つの水準の平均値を示し、その最大値から最小値を引いた結果を「水準別比較」に記載した。また同様に「全体平均値」は、3つの水準ごとではなく、3つの水準を合わせた全体の平均値を表わしている。このような規定から、もし水準別比較の値が0.0に近いならば、それは自己効力感の3つの水準間で困難度の差異が小さい領域または種別であり、逆に大きい場合にはその逆と判断できる。つまり水準別比較は、自己効力感の3つの水準にある差異の大きさを表す指標となる。水準別比較をみると、最も小さい値は「関連付け」の0.004であり、最も大きいものは「自由記述」の0.284

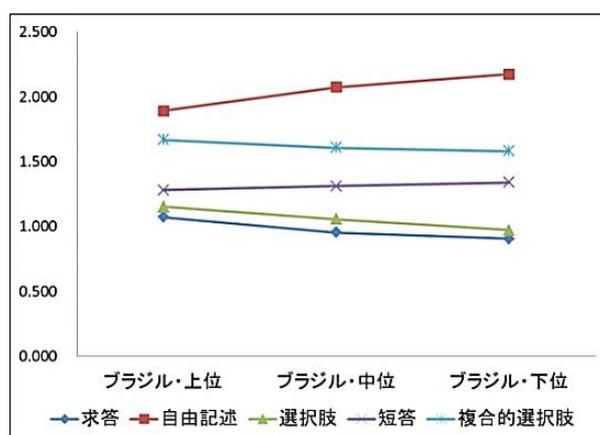
表 3.15 ブラジルにおける自己効力感の水準別の困難度の平均値

領域と種別 ( )内は項目数	水準別平均値			水準別 比較	全体平 均値
	上位	中位	下位		
【包括的アイデア】				*	**
空間と形 (19)	1.578	1.631	1.619	0.053	1.609
不確実性 (19)	1.589	1.562	1.512	0.077	1.554
量 (21)	1.148	1.126	1.189	0.063	1.154
変化と関係 (18)	1.313	1.309	1.305	0.008	1.309
【状況・文脈】					
科学的 (15)	1.390	1.344	1.384	0.046	1.373
教育的 (14)	1.692	1.619	1.586	0.105	1.632
公共的 (27)	1.402	1.471	1.496	0.094	1.457
私的 (16)	1.161	1.124	1.102	0.059	1.129
職業的 (4)	1.446	1.553	1.441	0.113	1.480
【能力】					
関連付け (37)	1.571	1.573	1.576	0.004	1.573
再現 (25)	1.010	1.014	1.020	0.010	1.015
熟考 (14)	1.669	1.658	1.640	0.029	1.656
【出題形式】					
求答 (12)	1.072	0.953	0.907	0.166	0.978
自由記述 (16)	1.891	2.073	2.175	0.284	2.046
選択肢 (16)	1.152	1.058	0.972	0.179	1.061
短答 (21)	1.280	1.311	1.340	0.060	1.310
複合的選択肢 (11)	1.667	1.608	1.580	0.087	1.618

出典：筆者作成

となっている。また、「選択肢」と「求答」においても比較的大きい値を示している。そこで図 3.11 に、種別ごとの困難度の平均値を折れ線グラフで表した。水準別比較が最も大きい「自由記述」は、水準が上位層になるほど困難度が小さくなり、他の種別はその逆の傾向をわずかながら示すことが分かる。ここで注目すべきことは、水準が下位層になるほど「自由記述」の困難度が高くなるという特徴である。全体平均値については、この値が

図 3.11 ブラジルの出題形式における困難度の水準別平均値



出典：筆者作成

高いほど難易度が高い領域あるいは種別となるため、2.046 と値が最も大きい「自由記述」は、最も難しいと判断できる。ゆえに「自由記述」は、自己効力感の水準が低くなるほど難易度が高くなり、全体的にも難しい種別といえよう。

これと同様な分析を他の6か国にも行い、各国の「水準別比較」を表3.16に提示した。また表3.17に各領域と種別ごとの各国の「全体平均値」を記し、この表にある「7カ国間比較」の欄に7カ国中の最大値から最小値を引いた結果を示した。この7カ国間比較は、その算出方法から自己効力感の水準別ではなく、各領域あるいは種別における国別の差異を表わす指標と扱える。

まず表3.16で示した国内水準別比較をみると、ブラジルで自己効力感の水準別で最も差異の大きかった「自由記述」は、チュニジアを例外とすれば、他の国においても同様に高い値という結果となっている。とはいえチュニジアにおいても、平均値や中央値を参考にすると、自由記述の値は必ずしも低くはない。ゆえに自己効力感の3つの水準間で困難度の差異が顕在化しやすいのは、種別の「自由記述」であることが浮かび上がった。他方で表3.17の7カ国間比較をみると、その平均値と中央値を参考にすれば、「自由記述」における国別での差異は小さくはないが、国内水準別比較と比べると、特に目立つ特徴がある

表 3.16 7カ国における水準別比較の国際比較（一の位の0は省略）

領域と種別 ( )内は項目数	IND	MEX	THA	TUN	TUR	URY	BRA
【包括的アイデア】							*
空間と形 (19)	.018	.016	.018	.050	.083	.052	.053
不確実性 (19)	.066	.026	.073	.187	.048	.024	.077
量 (21)	.024	.029	.030	.099	.036	.038	.063
変化と関係 (18)	.031	.013	.043	.060	.054	.043	.008
【状況・文脈】							
科学的 (15)	.031	.022	.084	.040	.062	.060	.046
教育的 (14)	.026	.043	.037	.049	.073	.031	.105
公共的 (27)	.017	.027	.037	.020	.127	.079	.094
私的 (16)	.075	.035	.039	.068	.083	.085	.059
職業的 (4)	.104	.083	.121	.151	.113	.017	.113
【能力】							
関連付け (37)	.032	.009	.027	.041	.034	.011	.004
再現 (25)	.023	.013	.048	.059	.022	.017	.010
熟考 (14)	.043	.045	.122	.185	.085	.029	.029
【出題形式】							
求答 (12)	.030	.091	.110	.066	.184	.067	.166
自由記述 (16)	.128	.143	.186	.181	.259	.202	.284
選択肢 (16)	.106	.102	.117	.254	.115	.082	.179
短答 (21)	.024	.051	.085	.125	.067	.033	.060
複合的選択肢 (11)	.063	.057	.153	.096	.118	.095	.087
平均値	.049	.047	.078	.102	.092	.057	.085
中央値	.031	.035	.073	.068	.083	.043	.063
最大値	.128	.143	.186	.254	.259	.202	.284
最小値	.017	.009	.018	.020	.022	.011	.004
レンジ	.110	.134	.168	.234	.237	.191	.280

出典：筆者作成

表 3.17 7カ国における全体平均値の国際比較（一の位の0は省略）

領域と種別 ( )内は項目数	IND	MEX	THA	TUN	TUR	URY	BRA	7カ国間 比較
【包括的アイデア】							**	
空間と形 (19)	1.406	1.514	1.424	1.461	1.622	1.574	1.609	.216
不確実性 (19)	1.587	1.609	1.610	1.651	1.495	1.643	1.554	.156
量 (21)	1.190	1.080	1.193	1.153	1.256	1.123	1.154	.176
変化と関係 (18)	1.449	1.434	1.402	1.366	1.245	1.288	1.309	.204
【状況・文脈】								
科学的 (15)	1.508	1.408	1.403	1.436	1.313	1.385	1.373	.195
教育的 (14)	1.500	1.518	1.430	1.546	1.548	1.702	1.632	.272
公共的 (27)	1.381	1.494	1.447	1.422	1.489	1.397	1.457	.113
私的 (16)	1.209	1.177	1.270	1.252	1.211	1.190	1.129	.141
職業的 (4)	1.625	1.304	1.579	1.287	1.450	1.350	1.480	.338
【能力】								
関連付け (37)	1.584	1.559	1.562	1.602	1.615	1.558	1.573	.057
再現 (25)	.982	.993	.990	1.018	.918	.986	1.015	.100
熟考 (14)	1.685	1.734	1.731	1.575	1.718	1.749	1.656	.173
【出題形式】								
求答 (12)	1.015	1.019	.844	1.003	1.064	1.104	.978	.259
自由記述 (16)	1.978	1.959	1.958	1.896	2.004	1.813	2.046	.233
選択肢 (16)	1.101	1.225	1.228	1.305	1.137	1.328	1.061	.267
短答 (21)	1.271	1.157	1.285	1.221	1.235	1.199	1.310	.154
複合的選択肢 (11)	1.692	1.755	1.697	1.624	1.617	1.644	1.618	.138
平均値	1.422	1.408	1.415	1.401	1.408	1.414	1.409	.188
中央値	1.449	1.434	1.424	1.422	1.450	1.385	1.457	.176
最大値	1.978	1.959	1.958	1.896	2.004	1.813	2.046	.338
最小値	.982	.993	.844	1.003	.918	.986	.978	.057
レンジ	.995	.966	1.113	.894	1.086	.827	1.069	.280

出典：筆者作成

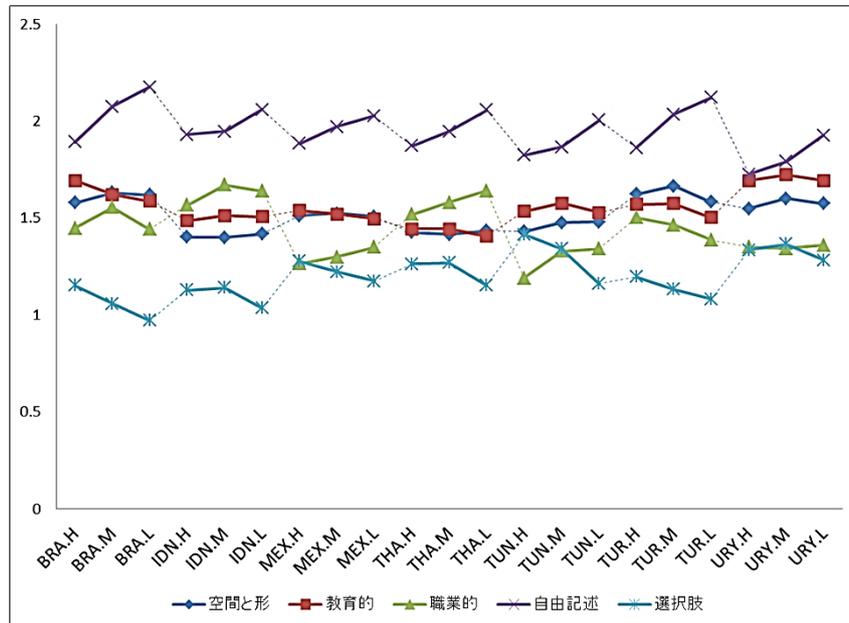
とはいえないだろう。むしろ国別の差異は「自由記述」よりも、「教育的」や「職業的」あるいは「選択肢」といった領域あるいは種別に反映されやすいと捉えるべきと考える。

これまで得られた結果を纏めると、困難度に注目した解答パターンが顕在化する傾向として、①国内における自己効力感の水準別に解答パターンが顕在化しやすいのは、「自由記述」であること、②自己効力感の水準別ではなく国別それが現れやすい場合があり、それは「教育的」や「職業的」あるいは「選択肢」といった領域・種別にみられることが浮かび上がった。そこで図 3.12 に、「自由記述」、「選択肢」、「教育的」、「職業的」および「空間と形」における 7カ国の 3つの水準の平均値を折れ線グラフで提示した。この図から分かるように「自由記述」は、どの国においても自己効力感の水準が低くなると、その難易度が高くなるという国際的な傾向を見て取れる。

### 3.3.2. 途上国に注目した分析結果の総合的考察

5つの情意的側面ごとに総合的な指標を算出したところ、本小節で注目する 7カ国は、興味・関心や道具的動機付けで国際的にみて高い水準にあるが、自己効力感や不安では、逆に低い傾向にあることが分かった（表 3.3）。なおわが国は、全てにおいて国際的に低い水

図 3.12 「自由記述」、「選択肢」、「教育的」、「職業的」および「空間と形」における困難度の国内水準別比較



出典：筆者作成

準という特徴を持つ。つまり取り上げた7ヵ国では、日本のような5つの情意的側面の水準に一貫性はみられず、それらが現れやすい側面は国際的にみて異なるという特徴を有する。ところが質問紙調査の国際比較には、田崎（2008）が指摘する評価の基準が異なって顕在化する可能性を考慮しなければならないだろう。そこでそれらの値を直接比較するだけでなく、各国で評価の基準を統一して算出されるテスト得点との関連性という視座に立ち、階層線形モデルによる分析を試みた。その結果、数学的リテラシー調査のテスト得点を国別の視点で5つの情意的側面から捉え直すことに一定の意味があることを確認した(表 3.6)。そしてデータとの適合が最も良かったモデル4を用いることで、「自己効力感」は、テスト得点で評価した学力に対して、正の方向に強い関連性を有する情意的側面という先進国と途上国に共通する特徴が浮かび上がった(表 3.8 および図 3.5)。

纏めると、本研究で取り上げた7ヵ国は、日本のように国際的にみて一貫して低い水準の情意的側面というような特徴はみられず、その内実は一見矛盾とも読み取れるものであった。しかしながら数学的リテラシー調査のテスト得点、つまり数学学力との関連性という視座に立つことによって、先進国だけでなく、途上国においても「自己効力感」は、数学学力に対して正の方向に強い関連性を有することが浮かび上がった。

そこで本分析では、先進国と低い学力と称され得る途上国に共通する特徴の把握という視座に立ち、7途上国において、「自己効力感」の水準の違いがテスト得点や出題項目の難易度とどのように関連するのかを調べた。まず、各国国内で自己効力感の水準別に3つの

受験者集団に分けた場合、テスト得点の分布はその水準別で階層的であった（表 3.14）。ところがテスト得点の分布が大きく異なるとしても、困難度に注目した解答パターンは、国内の3つの水準間で類似することが分かった（図 3.10）。次に解答パターンは、国内の水準間でより類似する状況の下で、各領域と種別の難易度を調べたところ、その差異は「自由記述」という種別で顕在化しやすいことが浮かび上がった。これは、前小節で得た先進国に対する結果と同様なものである。このように程度の大きさは同様であるとは言えないが、自己効力感の水準別に関して途上国と先進国に共通する特徴を掴むことができたと考える。それは、「自己効力感」の水準別が最も反映されやすいのは、出題形式の「自由記述」という特徴である（表 3.16 および図 3.12）。

これまでの分析結果を纏めると、途上国における数学学力と数学に対する情意的側面の関連性について、先進国と共通する2つの国際的な傾向が浮かび上がる。それはテスト得点で評価した数学学力に対して、①「自己効力感」は正の方向に強い関連性を有する情意的側面であること、②その水準の違いは「自由記述」という種別で顕在化されやすいという特徴である。これらは程度の差こそあれ、前小節で明らかにされた13の主要先進国に対して得られた結果が途上国にもみられることを意味する。したがって、数学学力の向上を意識しながら情意的側面の改善を試みるならば、途上国においても特に「自己効力感の育成」という視点の可能性を示唆するものである。特に本研究で示したように途上国は、PISAといった国際的な学力調査における数学の成績は低く、情意的側面も日本とは異なる様相を呈していた。しかしながら本分析の結果から、そのような国においても自己効力感の向上に軸足を置く取り組みの有効性を主張できると考える。加えて自由記述形式の出題項目を正答するには、答えを導く考え方や求め方などを説明する必要がある。日本においては、それに対する無答率の高さがしばしば問題視される（国立教育政策研究所、2010；室岡、2006）。もしその改善を試みるならば、自己効力感に焦点を当てることに一定の効果を期待できよう。本研究では、正答部分のみに焦点を当てたため、誤答の内実に触れることはできないが、成績下位層に位置付きがちな途上国においても同様な効果を促す可能性は必ずしも否定できないと考える。

### 3.4. 本章のまとめ

本章では、PISA2003の数学的リテラシー調査と質問紙調査の分析を通じて、生徒の数学学力と数学に対する情意的側面の関連性について調べてきた。

その結果、スウェーデンやフィンランドあるいはアイスランドといった欧米諸国では、特に情意的側面が数学学力に寄与する傾向にある（図 3.3 と図 3.4）が、そのような状況の中でも特に「自己効力感」は、39カ国に例外なく数学学力と強い関連性を有することが明らかになった。そしてその水準の違いは、程度の大きさが同様であるとはいえないが先進国と途上国とともに「自由記述」形式の出題項目で顕在化しやすいという特徴を併せて掴んだ（表 3.12、表 3.16、図 3.9 および図 3.12）。これを踏まえて数学学力の向上を意識する

ならば、その上位層と階層、つまり先進国と途上国を問わず、「自己効力感の向上」に焦点を絞る情意的側面に対する取り組みの有効性が示唆される。なお、伊藤・神藤（2003）および松沼（2004）によれば、自己効力感の向上には学習方略、特に自己調整学習方略<sup>31</sup>が関係するとされる。PISA では、生徒の学習方略についての調査を継続して実施しており、自己効力感の向上に繋がる学習方略に着目した分析が可能である。自己効力感と学習方略の関連性を明らかにする分析は、今後の課題としたい。

加えて本分析では、数学学力の水準が国際的にみて低い途上国に焦点を当てた検討も試みた。日本の途上国に関する数学教育研究は端緒についたばかり（馬場、2008）とされるが、馬場・内田（2008）および内田（2009）によれば、国際的な学力調査でその成績が特に低い途上国について、その低さのみが浮かび上がり、教育の質向上に繋がる有益な情報を得ることは難しいとされてきた。ところが本章では、関連性と困難度に注目した解答パターンという新たな視座に立つことによって、程度の違いは否めないとしても、先進国で見られる特徴が途上国でも確認されることを浮き彫りにした。文化・社会的な背景や生徒の学力水準など、多くの点で日本と異なる途上国のより深い理解に向けて、日本をはじめとする他の国との共通点を明らかにした本分析の意義は大きいと考える。

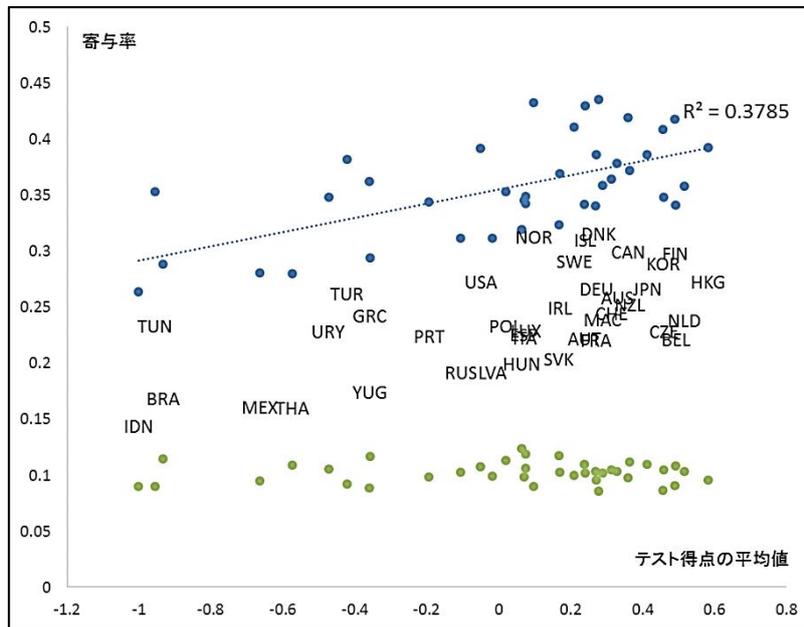
続いて、自己効力感の水準別に依らない副次的に得られた分析結果について触れる。鈴木・豊田・川端（2008）が指摘したように、本分析では、日本の生徒は数学の授業で典型的にみられるような疑似現実的な内容が含まれる「教育的」な文脈の問題を国際的にみて難無くこなす一方で、より抽象的とされる「科学的」な文脈では、他の国と比べて難易度が高いことを浮き彫りにした（図 3.9）。「科学的」については、技術的な過程、理論的な場面と明らかに数学的な問題についての理解（国立教育政策研究所、2004）と規定されるように、数学教育の中でその育成が期待される中心的なものの一つであろう。ところ日本の生徒は、そのような文脈で数学を活用していく能力が同程度のテスト得点の水準にある国と比べて弱いことを指摘しておく。

最後に本分析のように、先進国と途上国に共通する特徴を捉えることだけでなく、相違する点をより詳しく調べることも意義ある取り組みと考える。例えば図 3.9 と図 3.12 を比べるならば、先進国の方により明確に自己効力感の水準別の差異が現れている。本章では、PISA で設定されている 5 つの情意的側面を利用して分析を進めた。このような下位尺度を用いる方法もあるが、情意的側面に関する 26 項目における因子構造を調べることもできる。例えば 26 項目の次元性を確認したところ、テスト得点の平均値が高い国ほど、26 項目に高い次元性が認められる傾向が浮かび上がった。その結果を示したのが図 3.13 である。この図では、横軸が各国のテスト得点の平均値であり、縦軸がその国の次元性に関する寄与率を表わしている。なお、余白に各国の平均値と次元性を表わす指標となる相関係数行列の最大固有値の寄与率の散布図を反映させ、OECD が使用する国ラベルを記した。この図から分かるように、数学学力の水準が高い国ほど、高い情意的側面の次元性を持つ

<sup>31</sup> 学習者自身によってなされる認知および行動を調節する意志的制御を指す。

とも理解できる。つまり先進国では、自己効力感に傾倒する形で次元性が高まっていると思われる。これを翻るならば、次元性が低い傾向にある途上国に対して、下位尺度の一つである「自己効力感」のみに焦点を当てるのではなく、より情意的側面の水準別の差異が浮かび上がりやすい尺度を新たに構成して再検討するなどの工夫が必要かもしれない。先進国と途上国の差異をさらに浮かび上がらせるために、質問紙調査の結果をより詳細に調べることを今後の大きな課題として位置付けたい。

図 3.13 質問紙調査における 39 カ国の多次元性



出典：筆者作成

## 付録 3.1 5つの情意的側面における出題項目と因子負荷量

	因子負荷量	共通性
興味・関心(内的な動機付け)		
・ 数学についての本を読むのが好きである	0.765	0.585
・ 数学の授業が楽しみである	0.835	0.697
・ 数学を勉強しているのは楽しいからである	0.877	0.769
・ 数学で学ぶ内容に興味がある	0.781	0.610
道具的動機付け(外的な動機付け)	因子負荷量	共通性
・ 将来就きたい仕事に役立ちそうだから、数学は頑張る価値がある	0.805	0.648
・ 将来の仕事の可能性を広げてくれるから、数学は学びがいがあ	0.829	0.687
・ 自分にとって数学が重要な科目なのは、これから勉強したいことに必要だからである	0.795	0.632
・ これから数学でたくさんのことを学んで、仕事に就くときに役立てたい	0.779	0.607
自己効力感(数学への自信)	因子負荷量	共通性
・ 列車の時刻表をみて、ある場所から別の場所までどのくらい時間がかかるか計算する	0.577	0.333
・ あるテレビが30%引きになったとして、それが元の値段よりいくら安くなったかを計算する	0.664	0.441
・ 床にタイルを張るには、何平方メートル分のタイルが必要かを計算する	0.716	0.513
・ 新聞に掲載されたグラフを理解する	0.579	0.335
・ $3x+5=17$ という方程式を解く	0.593	0.352
・ 縮尺10000分の1の地図上にある、2点間の距離を計算する	0.628	0.394
・ $2(x+3)=(x+3)(x-3)$ という方程式を解く	0.572	0.327
・ 自動車のガソリンの燃費を計算する	0.549	0.301
自己概念(数学的な能力への信念)	因子負荷量	共通性
・ 数学はまったく得意ではない	0.750	0.563
・ 数学では良い成績をとっている	0.764	0.584
・ 数学はすぐわかる	0.839	0.704
・ 数学は得意科目の一つだといつも思う	0.777	0.604
・ 数学の授業ではどんな難しい問題でも理解できる	0.731	0.534
不安(数学に対する不安)	因子負荷量	共通性
・ 数学の授業についていけないのではないかとよく心配する	0.698	0.487
・ 数学の宿題をやるとなると、とても気が重くなる	0.743	0.552
・ 数学の問題をやっているといらいらする	0.745	0.555
・ 数学の問題を解くとき、手も足も出ないと感じる	0.659	0.434
・ 数学でひどい成績をとるのではないかと心配になる	0.574	0.329

出典：筆者作成

付録 3.2 PISA2003 の数学的リテラシーにおける領域・種別

領域・種別	内容
<b>【包括的アイデア】</b>	
空間と形	空間的、幾何的な現象や関係を扱う内容
不確実性	確率的、統計的な現象や関係を扱う内容
量	数量的な関係、数量的なパターン、数量的な現象を扱う内容
変化と関係	変数間の関数的な関係と依存関係とともに変化の数学的な関係を扱う内容
<b>【状況・文脈】</b>	
科学的	より抽象的な文脈で、技術的な過程、理論的な場面、明らかに数学的な問題についての理解に関連する文脈
教育的	生徒の学校生活に現れるような文脈
公共的	生徒が生活する地域社会における文脈
私的	生徒の日々の活動に直接関係する文脈
職業的	職業の場面に現れるような文脈
<b>【能力】</b>	
関連付け	再現の上に位置付くもので、やや見慣れた場面、または見慣れた場面から拡張され発展された場面において、手順がそれほど決まりきってはいない問題を解く能力
再現	比較的良好に見慣れた、練習された知識の再現を主に要する問題を解く能力
熟考	関連付けのさらに上に位置付くもので、洞察、反省的思考、関連する数学を見付け出す創造性、解を生み出すために関連する知識を結びつける能力
<b>【出題形式】</b>	
求答	答えが問題のある部分に含まれる問題
自由記述	答えの求め方や考え方を説明する問題
選択肢	選択肢の問題が1問だけの問題
短答	答えを求めるのに計算が必要な問題
複合的選択肢	選択肢の問題が連続して複数問ある問題

出典：国立教育政策研究所（2004）をもとに筆者作成

## 第4章 数学学力と言語的側面の関連性

本章では、生徒の数学学力と言語的側面の関連性を明らかにする。PISA2003、PISA2006 および PISA2009 で収集されたデータを活用し、読解力とテスト言語の使用状況と言語的側面を代表させながら、数学学力との関連性を国際比較していく。そして第2章と第3章と同様に、数学的リテラシー調査の項目数が他と比べて多い PISA2003 に着目し、困難度に注目した解答パターンを検討し、数学学力と読解力の関連性をさらに調べていく。

### 4.1. 数学的リテラシーと読解力調査のテスト得点の国際比較

本小節では、PISA2003 と PISA2006 および PISA2009 における数学的リテラシーと読解力調査のテスト得点の経年変化を掴むとともに、それらの関連性を調べていく。

#### 4.1.1. 使用データとそのデータ処理

PISA では、2000 年から継続して数学的リテラシーと読解力に関する調査が行われている。そのため、言語的側面を読解力などに代表させるならば、それらの水準や関連性に関する経年変化を掴むことができると思われる。そこで経年変化をみるために、PISA2003、PISA2006 および PISA2009 に参加した国を分析対象とする。

まず、分析対象とした受験者について説明する。PISA では表 1.6 で示したように、重複テスト分冊法によって数学的リテラシー、読解力および科学的リテラシーに関する学力が測定された。例えば表 1.6 をみると、ブックレット 5 とブックレット 6 の受験者は、読解力に関するテストに解答していないことが分かる。そこで本分析では、数学的リテラシーと読解力調査の共通受験者に対象を限定する。しかし分析対象を 2 つの調査の共通受験者に絞るならば、いくつかの出題項目に対する受験者数が極端に少なくなる場合がある。PISA では、全ての項目において全受験者の約 3 割程度が解答するように設定されていることを踏まえ、受験者数が全体の 3 割を大きく下回るような項目は、分析対象から除外することにした。この過程を踏むと、数学的リテラシーや読解力調査のテスト得点を持たない受験者が生じてしまう。そこで、そのような生徒も分析対象に含めないことにした。なお数学的リテラシーと読解力調査で収集されたデータは、図 1.7 で示したように、全ての項目において正答を 1 誤答を 0 とする 2 値データへと処理してある。

また PISA では、その国の主要な教授言語でテストや質問紙が作成されており、テスト言語をどの程度日常生活で使用するかという調査が継続的に行われている。つまりテスト言語の使用率は、言語的側面を反映させる指標として扱えると判断できるため、その調査結果も活用する。テスト言語の使用に関する調査の質問事項は、「どの言語を家で使いますか」というものである。その回答方法は、PISA2003 では「テスト言語」、「その他の公用語」、「その他の現地語」、「その他」の 4 つから該当するものを選ぶという選択肢形式である。また PISA2006 では、「テスト言語」、「その他の公用語」、「その他」の 3 つの中から選び、

PISA2009では、「テスト言語」または「その他」の2つから選ぶという方法が取られている。そこでテスト言語のデータ処理は、PISA2009の調査内容に合わせる形で、3つのPISAそれぞれで、テスト言語の使用を1それ以外の選択肢は0として扱い、ここでも2値データを作成した。また当然ながらこの調査においても、無回答や欠損値が含まれる。そのためこれらの回答を有する受験者は、分析対象には含めないことにした。ゆえに本分析の対象は、数学的リテラシーと読解力およびテスト言語の共通受験者かつテスト言語の使用に関する項目に完全回答を有する生徒である。

このような処理を行うことによって、PISA2003とPISA2006およびPISA2009における数学的リテラシーと読解力の共通受験者は、全体のそれぞれ44.49%、44.50%、59.16%程度に留まってしまう。この受験者数は、一見分析に対して不十分とも思われる。ところが3つのPISAでは、それぞれに主要な調査内容が設定され、主な調査内容ではないものは表1.6から分かるように、全受験者でなく一部の生徒を対象に実施されている。上記の受験者数は、主要な調査内容でない他の調査内容の受験者数のそれぞれ約85%、100%、約88%を占める。他の調査内容の分析結果も主要な内容と併せてPISAの各報告書(e.g., OECD、2004、2007、2010b)で公開されていること踏まえれば、本分析のように、分析対象を共通受験者でありテスト言語の使用について完全回答を持つ生徒に限定しても、一定の信頼性を失わず国際比較分析が行えると考えられる。とはいえ、各国で一定の受験者数を確保する必要もあるため、データ処理によって受験者数が1000名を下回る国は、本分析では分析対象外とした。したがってPISA2003とPISA2006およびPISA2009では、それぞれ38カ国<sup>32</sup>の117447名、55カ国<sup>33</sup>の174418名、72カ国<sup>34</sup>の303431名の受験者を対象に分析を進めていく。

#### 4.1.2. 数学的リテラシーと読解力調査のテスト得点とその経年変化の国際比較

本章では、対象とする生徒数が一般に公開されている報告書と異なるため、数学的リテラシーと読解力調査のテスト得点を算出し、その分布を確認する。前章と同様に本章においても、ラッシュモデル(1.2)を採用する。なぜならば、PISAではラッシュモデルが使用されていることに加え、豊田(2012b, p.25)が述べるようにその扱い方が容易であるためである。

まずテスト得点を算出する前に、項目母数である困難度に極端な値を示すものがないか

<sup>32</sup> 第2章と第3章で対象とした39カ国からマカオを除いた38カ国。

<sup>33</sup> PISA2003の38カ国の中でPISA2006に含まれていない旧セルビア・モンテネグロと読解力調査で不備が指摘され比較の対象にならなかったアメリカを除き、アルゼンチン、アゼルバイジャン、ブルガリア、チリ、コロンビア、エストニア、イギリス、クロアチア、イスラエル、ヨルダン、カザフスタン、リトアニア、マカオ、モンテネグロ、カタール、ルーマニア、セルビア、スロバニア、台湾の19カ国を加えた55カ国。

<sup>34</sup> PISA2006の55カ国に、アルバニア、アラブ首長国連邦、コスタリカ、グルジア、キルギスタン、モルドバ、マルタ、モーリシャス、マレーシア、パナマ、ペルー、上海、タミルナドゥインド、ミランダベネズエラ、シンガポール、トリニダードトバコ、アメリカを加えた72カ国。

を確認する。そこで表 4.1 に、各 PISA における数学的リテラシーと読解力の各項目の困難度の基本統計量を示した。最大値と最小値をみると、どの PISA においても芝 (1991) と豊田 (2002b) の基準を満たすことが分かる。したがって、各 PISA における数学的リテラシーと読解力のテスト得点を項目数の欄に示した数の項目を用いて算出する。

表 4.1 3つの PISA における項目母数の基本統計量

基本統計量	PISA2003		PISA2006		PISA2009	
	数学的リテラシー	読解力	数学的リテラシー	読解力	数学的リテラシー	読解力
項目数	72 (84)	28 (28)	48 (48)	28 (28)	35 (35)	70 (131)
平均値	0.321	0.019	0.320	0.104	0.427	-0.120
標準偏差	0.804	0.815	0.828	0.828	0.772	0.895
最大値	2.236	2.212	2.182	2.399	2.148	2.297
最小値	-1.761	-1.437	-2.052	-1.403	-1.306	-2.763
レンジ	3.997	3.649	4.234	3.802	3.454	5.060

注) ()内は、本来の項目数。

出典：筆者作成

続いて、各 PISA における数学的リテラシーと読解力調査のテスト得点を算出する。また PISA では、3つの調査に共通項目が設定されており、それらを利用して等化を施すことによって、尺度を統一して国際比較を行うことができる。そこで本分析においても、共通項目を利用して等化を施し、経年変化を含めてテスト得点を国際比較していく。

等化を行うため、共通項目を利用して等化係数を算出しなければならない。本分析ではラッシュモデル (1.2) を用いるため、等化係数  $l$  のみを求めればよい。式 (1.4) から分かるように等化係数  $l$  は、等化の基準となる受験者集団における共通項目の困難度の平均値から等化を施す集団のそれを引き算することによって得られる。そこで表 4.2 と表 4.3 に、それぞれ数学的リテラシーと読解力における共通項目とそれから得られる等化係数  $l$  を示した。なお本分析では、PISA2006 と PISA2009 を PISA2003 に等化した。

得られた等化係数と式 (1.3) から、3つの PISA における各受験者の数学的リテラシーと読解力調査のテスト得点を PISA2003 の尺度に等化する。等化後の各受験者のテスト得点に関する結果として、表 4.4 と表 4.5 のように数学的リテラシーと読解力の各国の分布を平均点と標準偏差を用いて要約した。なおこれらの表では、結果の提示が煩雑になるのを防ぐために PISA2003 に参加した 38 カ国に焦点を絞り、それらの国における PISA2006 と PISA2009 の分析結果を示した。なお斜体の数値は、3つの PISA における平均値を比較した際の順位である。また図 4.1 に、PISA2003 における参加国の数学的リテラシーと読解力調査の平均値の散布図を提示した。これらの表と図において、第 2 章と第 3 章で取り上げた 13 先進国と 7 途上国に目を向けると、前者の国は各 PISA の数学的リテラシーと読解力調査ともに上位層に位置し、後者の国は下位層に位置する傾向を読み取れる。ちなみ図 4.1 における相関係数は 0.965 であり、PISA2006 と PISA2009 における同様の散布図に関

表 4.2 数学的リテラシーにおける3つのPISA間の共通項目と等化係数

共通項目	PISA2003		PISA2006		共通項目	PISA2003		PISA2009	
	困難度	S.E	困難度	S.E		困難度	S.E	困難度	S.E
M033Q01	-0.793	0.009	-0.732	0.007	M033Q01	-0.793	0.009	-0.582	0.004
M034Q01T	0.418	0.008	0.344	0.007	M034Q01T	0.418	0.008	0.591	0.004
M192Q01T	0.589	0.008	0.609	0.007	M192Q01T	0.589	0.008	0.616	0.004
M273Q01T	0.073	0.008	0.123	0.007	M273Q01T	0.073	0.008	0.155	0.005
M302Q01T	-1.761	0.012	-2.052	0.011	M406Q01	1.108	0.009	1.035	0.005
M302Q02	-0.692	0.008	-0.777	0.007	M406Q02	1.540	0.010	1.499	0.006
M302Q03	0.870	0.009	0.889	0.008	M408Q01T	0.547	0.008	0.629	0.005
M305Q01	-0.352	0.008	-0.081	0.007	M411Q01	0.428	0.008	0.377	0.004
M406Q01	1.108	0.009	1.015	0.008	M411Q02	0.512	0.008	0.357	0.004
M406Q02	1.540	0.010	1.489	0.009	M423Q01	-0.941	0.009	-0.827	0.005
M408Q01T	0.547	0.008	0.553	0.007	M446Q01	-0.308	0.008	-0.207	0.005
M411Q01	0.428	0.008	0.142	0.007	M446Q02	2.236	0.014	2.147	0.008
M411Q02	0.512	0.008	0.230	0.007	M462Q01T	2.215	0.014	2.148	0.007
M421Q01	-0.204	0.008	-0.149	0.007	M464Q01T	0.951	0.009	1.182	0.006
M421Q02T	1.369	0.010	1.408	0.009	M474Q01	-0.432	0.008	-0.495	0.004
M421Q03	0.572	0.008	0.565	0.007	M559Q01	-0.046	0.008	-0.193	0.005
M423Q01	-0.941	0.009	-0.762	0.007	M564Q01	0.104	0.008	0.290	0.004
M446Q01	-0.308	0.008	-0.178	0.007	M564Q02	0.213	0.008	0.332	0.004
M446Q02	2.236	0.014	2.182	0.012	M571Q01	0.416	0.008	0.377	0.004
M462Q01T	2.215	0.014	1.965	0.011	M603Q01T	0.265	0.008	0.425	0.004
M464Q01T	0.951	0.009	1.147	0.008	M603Q02T	0.573	0.009	0.751	0.005
M474Q01	-0.432	0.008	-0.712	0.007	M800Q01	-1.554	0.011	-1.306	0.006
M559Q01	-0.046	0.008	-0.117	0.007	M803Q01T	1.037	0.009	1.127	0.005
M564Q01	0.104	0.008	0.326	0.007	M828Q01	0.523	0.008	0.840	0.005
M564Q02	0.213	0.008	0.338	0.007	M828Q02	0.014	0.008	0.079	0.005
M571Q01	0.416	0.008	0.369	0.007	M828Q03	0.701	0.009	0.930	0.005
M598Q01	-0.060	0.008	-0.293	0.007					
M603Q01T	0.265	0.008	0.406	0.007					
M603Q02T	0.573	0.009	0.716	0.007					
M710Q01	0.935	0.009	0.760	0.007					
M800Q01	-1.554	0.011	-1.130	0.008					
M803Q01T	1.037	0.009	0.894	0.008					
M810Q01T	-0.576	0.008	-0.231	0.007					
M810Q02T	-0.727	0.008	-0.473	0.007					
M810Q03T	1.614	0.011	1.804	0.010					
M828Q01	0.523	0.008	0.829	0.007					
M828Q02	0.014	0.008	0.207	0.007					
M828Q03	0.701	0.009	1.002	0.008					
M833Q01T	0.726	0.008	0.844	0.008					
平均値	0.310	0.009	0.345	0.008	平均値	0.400	0.009	0.472	0.005
等化係数		-0.035			等化係数		-0.073		

注) S.Eは、標準誤差を意味する。

出典：筆者作成

する相関係数は、それぞれ 0.934 と 0.951 となった。加えて平均値の経年変化を掴むために、図 4.2 に PISA2003 と PISA2009 における 38 カ国の平均値の散布図を示した。なおこの図に対する相関係数は、0.979 である。また PISA2003 と PISA2006 および PISA2006 と PISA2009 との間の数学的リテラシー調査の平均値の散布図に対する相関係数は、それぞれ 0.987 と 0.982 である。これを踏まえると、数学的リテラシーと読解力の上位層あるいは下位層の国は3つのPISA間で類似し、先進国と途上国の間には、依然大きな学力水準

表 4.3 読解力における3つのPISA間の共通項目と等化係数

共通項目	PISA2003		PISA2009		共通項目	PISA2003		PISA2009	
	困難度	S.E	困難度	S.E		困難度	S.E	困難度	S.E
R055Q01	-0.963	0.008	-0.849	0.007	R055Q01	-0.963	0.008	-1.207	0.007
R055Q02	0.261	0.007	0.390	0.006	R055Q02	0.261	0.007	0.029	0.006
R055Q03	-0.062	0.007	0.042	0.006	R055Q03	-0.062	0.007	-0.319	0.006
R055Q05	-0.503	0.007	-0.398	0.006	R055Q05	-0.503	0.007	-0.814	0.006
R067Q01	-1.437	0.009	-1.403	0.007	R067Q01	-1.437	0.009	-1.016	0.007
R067Q04	0.447	0.007	0.463	0.006	R067Q04	0.447	0.007	0.603	0.006
R067Q05	-0.157	0.007	-0.150	0.006	R067Q05	-0.157	0.007	0.065	0.006
R102Q04A	0.987	0.008	1.003	0.007	R102Q04A	0.987	0.008	1.282	0.007
R102Q05	0.478	0.007	0.504	0.006	R102Q05	0.478	0.007	0.741	0.006
R102Q07	-1.019	0.008	-1.017	0.007	R102Q07	-1.019	0.008	-0.675	0.006
R104Q01	-0.811	0.008	-0.692	0.006	R104Q01	-0.811	0.008	-0.968	0.006
R104Q02	0.735	0.008	0.818	0.006	R104Q02	0.735	0.008	0.460	0.006
R104Q05	2.212	0.012	2.399	0.011	R104Q05	2.212	0.012	2.297	0.013
R111Q01	-0.302	0.007	-0.185	0.006	R111Q01	-0.302	0.007	-0.531	0.006
R111Q02B	1.545	0.009	1.600	0.008	R111Q02B	1.545	0.009	1.214	0.007
R111Q06B	0.670	0.008	0.840	0.006	R111Q06B	0.670	0.008	0.508	0.006
R219Q01T	-0.418	0.007	-0.218	0.006	R219Q02	-0.924	0.008	-0.537	0.006
R219Q01E	-0.024	0.007	0.190	0.006	R220Q01	0.544	0.007	0.908	0.006
R219Q02	-0.924	0.008	-0.717	0.006	R220Q02B	-0.093	0.007	0.152	0.006
R220Q01	0.544	0.007	0.580	0.006	R220Q04	-0.123	0.007	0.234	0.006
R220Q02B	-0.093	0.007	-0.161	0.006	R220Q05	-0.822	0.008	-0.447	0.006
R220Q04	-0.123	0.007	-0.121	0.006	R220Q06	-0.230	0.007	0.092	0.006
R220Q05	-0.822	0.008	-0.899	0.007	R227Q01	0.113	0.007	-0.143	0.006
R220Q06	-0.230	0.007	-0.283	0.006	R227Q02T	0.896	0.008	0.716	0.006
R227Q01	0.113	0.007	0.265	0.006	R227Q03	0.022	0.007	-0.233	0.006
R227Q02T	0.896	0.008	1.033	0.007	R227Q06	-0.484	0.007	-0.712	0.006
R227Q03	0.022	0.007	0.122	0.006					
R227Q06	-0.484	0.007	-0.254	0.006					
平均値	0.019	0.008	0.104	0.006	平均値	0.038	0.008	0.065	0.006
等化係数	-0.085				等化係数	-0.028			

注) S.Eは、標準誤差を意味する。

出典：筆者作成

の差異が存在するといえよう。

他方で得られた分析結果において、3つのPISAにおいて各国の平均値を比較することは当然可能であるが、各PISA間の比較、つまり経年変化を掴むことが項目反応理論による等化によって行える。例えば表4.4をみると日本は、PISA2006で数学的リテラシーの得点が国際的にみて低下したが、フィンランドでは、その逆の傾向にあると読み取れる。読解力では、日本はPISA2003とPISA2009の間に比較的大きな向上が認められる。そこで式(2.1)を用いて、PISA2003とPISA2009の数学的リテラシーと読解力調査における各国の平均値の差の効果を計算し、図4.3のように棒グラフで表現した。数学で最も大きな効果量と認められるギリシャの値は、0.367である。これは、PISA2003とPISA2009における平均的な標準偏差の約37%分の差異があることを意味する。他方で読解力で最も大きな効果量は、1.071を持つメキシコである。つまりメキシコは、PISA2003とPISA2009の間に1標準偏差に近い学力水準の向上がみられたと解釈される。図4.3をみると全体的に、数学的リテラシー、つまり数学学力よりも読解力により経年変化が現れやすいといえよう。つまり

表 4.4 数学的リテラシーにおけるテスト得点とその経年変化(PISA2003 の平均値の降順)

国名	平均値			標準偏差			受験者数		
	2003 (38)	2006 (57)	2009 (72)	2003	2006	2009	2003	2006	2009
香港	0.531 <i>1</i>	0.502 <i>2</i>	0.559 <i>3</i>	0.773	0.687	0.721	1991	2111	2905
オランダ	0.493 <i>2</i>	0.433 <i>5</i>	0.456 <i>6</i>	0.706	0.674	0.670	1722	2150	2725
ベルギー	0.474 <i>3</i>	0.411 <i>7</i>	0.385 <i>10</i>	0.814	0.731	0.744	3564	3699	4739
フィンランド	0.454 <i>4</i>	0.501 <i>3</i>	0.445 <i>7</i>	0.693	0.639	0.657	2656	2169	3530
韓国	0.444 <i>5</i>	0.465 <i>4</i>	0.505 <i>5</i>	0.749	0.696	0.673	2518	2383	3018
チェコ	0.420 <i>6</i>	0.423 <i>6</i>	0.317 <i>14</i>	0.766	0.759	0.734	2813	2646	3479
日本	0.402 <i>7</i>	0.330 <i>10</i>	0.425 <i>8</i>	0.790	0.691	0.712	2039	2647	3635
スイス	0.325 <i>8</i>	0.368 <i>8</i>	0.422 <i>9</i>	0.777	0.717	0.717	3635	5336	6764
ニュージーランド	0.318 <i>9</i>	0.348 <i>9</i>	0.337 <i>13</i>	0.784	0.717	0.721	2040	2111	2835
オーストラリア	0.313 <i>10</i>	0.267 <i>13</i>	0.257 <i>17</i>	0.777	0.704	0.719	5662	6382	8586
ドイツ	0.301 <i>11</i>	0.237 <i>17</i>	0.339 <i>12</i>	0.744	0.725	0.724	1860	1961	2588
カナダ	0.298 <i>12</i>	0.280 <i>12</i>	0.311 <i>15</i>	0.720	0.678	0.680	12075	10029	13642
オーストリア	0.269 <i>13</i>	0.252 <i>14</i>	0.240 <i>19</i>	0.730	0.722	0.713	2033	2192	3673
フランス	0.258 <i>14</i>	0.137 <i>22</i>	0.205 <i>21</i>	0.767	0.714	0.726	1896	2116	2532
アイスランド	0.254 <i>15</i>	0.217 <i>18</i>	0.245 <i>18</i>	0.743	0.697	0.690	1504	1705	2154
デンマーク	0.234 <i>16</i>	0.242 <i>16</i>	0.133 <i>29</i>	0.745	0.654	0.669	1856	1974	3359
スウェーデン	0.211 <i>17</i>	0.192 <i>19</i>	0.176 <i>23</i>	0.761	0.695	0.704	1989	1985	2658
スロバキア	0.171 <i>18</i>	0.119 <i>26</i>	0.163 <i>26</i>	0.759	0.698	0.693	3318	2123	2782
アイルランド	0.160 <i>19</i>	0.182 <i>21</i>	0.123 <i>30</i>	0.690	0.667	0.658	1776	2076	2370
ルクセンブルク	0.106 <i>20</i>	0.137 <i>23</i>	0.224 <i>20</i>	0.730	0.704	0.706	1746	1863	2515
イタリア	0.105 <i>21</i>	0.036 <i>32</i>	0.171 <i>24</i>	0.766	0.714	0.685	5062	8732	16853
ハンガリー	0.103 <i>22</i>	0.111 <i>27</i>	0.154 <i>27</i>	0.755	0.692	0.684	1955	2036	2792
スペイン	0.101 <i>23</i>	0.184 <i>20</i>	0.139 <i>28</i>	0.725	0.674	0.704	4979	8814	15712
ノルウェー	0.092 <i>24</i>	0.098 <i>28</i>	0.185 <i>22</i>	0.757	0.699	0.670	1796	2080	2825
ポーランド	0.039 <i>25</i>	0.133 <i>24</i>	0.168 <i>25</i>	0.739	0.679	0.680	2015	2498	2992
ラトビア	0.036 <i>26</i>	0.077 <i>30</i>	0.097 <i>32</i>	0.725	0.649	0.624	2082	2167	2677
アメリカ	0.002 <i>27</i>		0.079 <i>34</i>	0.707		0.689	2420		3176
ロシア	-0.087 <i>28</i>	-0.008 <i>33</i>	-0.037 <i>39</i>	0.782	0.724	0.656	2703	2664	3231
ポルトガル	-0.098 <i>29</i>	-0.031 <i>34</i>	0.092 <i>33</i>	0.710	0.701	0.700	2076	2320	3782
ギリシャ	-0.290 <i>30</i>	-0.119 <i>37</i>	-0.026 <i>38</i>	0.753	0.704	0.675	2106	2138	2965
旧セ	-0.309 <i>31</i>			0.677			2012		
トルコ	-0.345 <i>32</i>	-0.330 <i>41</i>	-0.203 <i>43</i>	0.758	0.709	0.683	2253	2274	3049
タイ	-0.383 <i>33</i>	-0.353 <i>42</i>	-0.320 <i>50</i>	0.694	0.675	0.650	2384	2812	3724
ウルグアイ	-0.449 <i>34</i>	-0.266 <i>40</i>	-0.298 <i>46</i>	0.782	0.745	0.678	2624	2131	3577
メキシコ	-0.518 <i>35</i>	-0.379 <i>44</i>	-0.332 <i>51</i>	0.642	0.631	0.598	13434	14219	22984
インドネシア	-0.752 <i>36</i>	-0.630 <i>50</i>	-0.649 <i>67</i>	0.597	0.596	0.551	4783	4865	3147
チュニジア	-0.764 <i>37</i>	-0.731 <i>53</i>	-0.690 <i>69</i>	0.611	0.689	0.569	2091	1981	3043
ブラジル	-0.794 <i>38</i>	-0.705 <i>52</i>	-0.604 <i>64</i>	0.670	0.683	0.576	1979	4268	11934
全受験者	0.019	-0.022	0.019	0.821	0.781	0.821	117447	174418	303431

注) 旧セは、旧セルビア・モンテネグロ。空欄は、その年にデータがないことを意味する。

( )内は、各PISAにおける分析対象国の数を表わす。

出典：筆者作成

数学学力は、読解力に比べてより緩やかな動きの中で変化することも捉えられる。

#### 4.1.3. 階層線形モデルによる数学的リテラシーと読解力調査のテスト得点およびテスト言語の関連性の国際比較

先進国と途上国では、数学的リテラシーと読解力の学力において大きな差が依然みられる。また経年変化をみたところ数学学力は、読解力に比べ緩やかな動きを持ちながら変化

表 4.5 読解力におけるテスト得点とその経年変化 (PISA2003 の平均値の降順)

国名	平均値			標準偏差			受験者数					
	2003 (38)	2006 (57)	2009 (72)	2003	2006	2009	2003	2006	2009			
フィンランド	0.437	1	0.479	2	0.400	4	0.659	0.648	0.671	2656	2169	3530
韓国	0.429	2	0.543	1	0.483	2	0.674	0.686	0.597	2518	2383	3018
オランダ	0.358	3	0.331	5	0.358	8	0.673	0.693	0.641	1722	2150	2725
ニュージーランド	0.351	4	0.363	4	0.367	7	0.769	0.758	0.738	2040	2111	2835
ベルギー	0.336	5	0.304	6	0.340	9	0.775	0.759	0.727	3564	3699	4739
オーストラリア	0.332	6	0.240	9	0.283	11	0.768	0.748	0.735	5662	6382	8586
香港	0.317	7	0.432	3	0.475	3	0.662	0.636	0.640	1991	2111	2905
アイルランド	0.308	8	0.304	7	0.234	17	0.692	0.718	0.698	1776	2076	2370
チェコ	0.273	9	0.273	8	0.231	18	0.736	0.804	0.715	2813	2646	3479
スウェーデン	0.261	10	0.239	11	0.211	23	0.754	0.746	0.769	1989	1985	2658
カナダ	0.257	11	0.239	10	0.294	10	0.713	0.750	0.690	12075	10029	13642
ドイツ	0.250	12	0.230	12	0.276	12	0.749	0.754	0.711	1860	1961	2588
日本	0.243	13	0.202	15	0.374	6	0.787	0.787	0.747	2039	2647	3635
オーストリア	0.221	14	0.170	17	0.091	35	0.760	0.750	0.734	2033	2192	3673
フランス	0.206	15	0.120	21	0.226	20	0.738	0.785	0.762	1896	2116	2532
スイス	0.200	16	0.174	16	0.223	22	0.729	0.723	0.693	3635	5336	6764
デンマーク	0.173	17	0.148	18	0.133	32	0.728	0.720	0.682	1856	1974	3359
イタリア	0.168	18	0.054	27	0.202	24	0.763	0.776	0.711	5062	8732	16853
アイスランド	0.142	19	0.104	22	0.239	16	0.756	0.731	0.721	1504	1705	2154
アメリカ	0.135	20			0.200	25	0.759		0.715	2420		3176
ノルウェー	0.125	21	0.069	26	0.228	19	0.793	0.771	0.715	1796	2080	2825
ポーランド	0.098	22	0.220	13	0.269	13	0.757	0.773	0.691	2015	2498	2992
ハンガリー	0.059	23	0.078	24	0.223	21	0.732	0.711	0.674	1955	2036	2792
ラトビア	0.049	24	0.029	28	0.173	28	0.750	0.715	0.647	2082	2167	2677
スペイン	0.046	25	0.026	29	0.103	34	0.746	0.670	0.706	4979	8814	15712
ルクセンブルク	0.031	26	0.077	25	0.156	30	0.779	0.753	0.741	1746	1863	2515
スロバキア	0.017	27	-0.045	33	0.080	36	0.740	0.776	0.700	3318	2123	2782
ポルトガル	-0.033	28	0.010	30	0.122	33	0.735	0.758	0.685	2076	2320	3782
ギリシャ	-0.183	29	-0.122	35	0.169	29	0.816	0.773	0.751	2106	2138	2965
ロシア	-0.266	30	-0.239	39	-0.009	39	0.753	0.767	0.711	2703	2664	3231
トルコ	-0.272	31	-0.215	37	-0.022	40	0.742	0.709	0.681	2253	2274	3049
メキシコ	-0.464	32	-0.365	41	-0.250	49	0.709	0.728	0.649	13434	14219	22984
タイ	-0.468	33	-0.435	42	-0.255	50	0.678	0.692	0.670	2384	2812	3724
ウルグアイ	-0.485	34	-0.356	40	-0.314	53	0.887	0.889	0.766	2624	2131	3577
旧セ	-0.486	35					0.642			2012		
ブラジル	-0.631	36	-0.628	50	-0.446	59	0.766	0.743	0.723	1979	4268	11934
インドネシア	-0.691	37	-0.670	51	-0.458	61	0.618	0.610	0.609	4783	4865	3147
チュニジア	-0.759	38	-0.698	52	-0.471	62	0.681	0.700	0.710	2091	1981	3043
全受験者	-0.001		-0.081		-0.033		0.810	0.816	0.778	117447	174418	303431

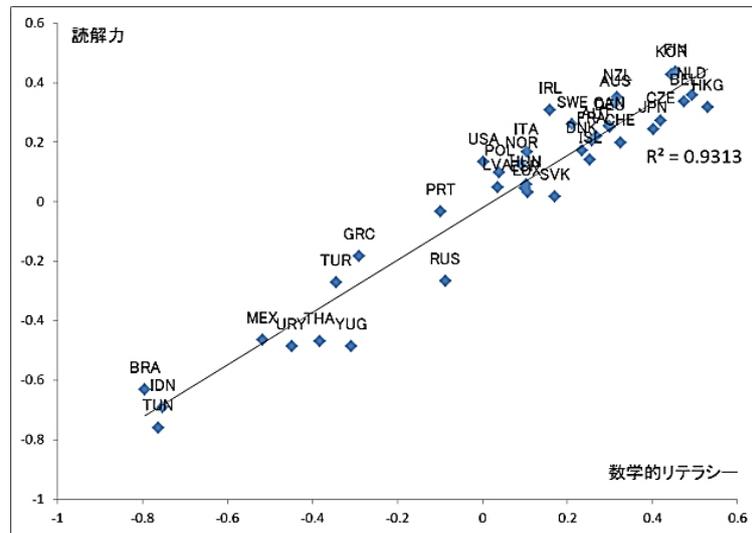
注) 旧セは、旧セルビア・モンテネグロ。空欄は、その年にデータがないことを意味する。

( )内は、各PISAにおける分析対象国の数を表わす。

出典：筆者作成

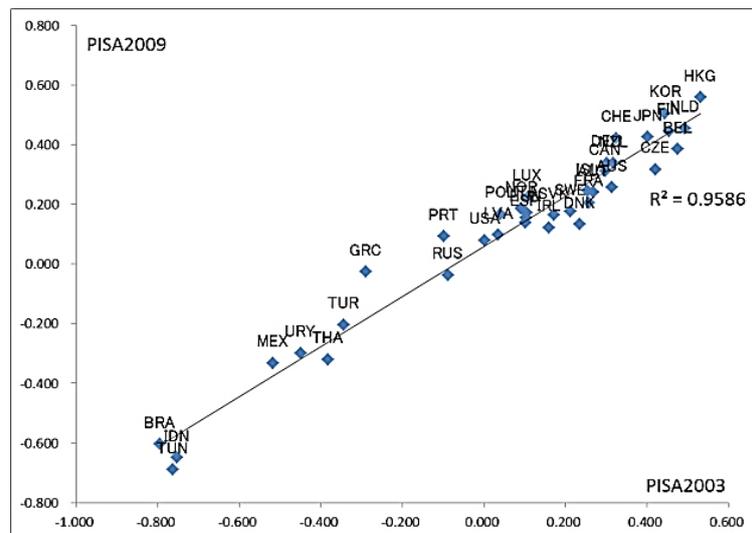
することも分かった。しかし各国の数学学力や読解力の水準を表わす指標（例えば、平均値など）を用いてそれらの関連性をみたとしても、各国国内でそれらがどの程度関連するかまでは把握できない。例えば、数学学力と読解力の水準が国際的に低いとしても、その国の国内でそれらの関連性がどの程度なのかを掴むことはできない。つまり学力水準の違いを取り除きながら、各国国内における関連性を調べる必要がある。そこで、言語的側面を読解力のテスト得点とテスト言語の使用という変数に代表させながら、数学的リテラシ

図 4.1 PISA2003 における数学的リテラシーと読解力の平均値の散布図



出典：筆者作成

図 4.2 PISA2003 と PISA2009 における数学的リテラシーの平均値の散布図

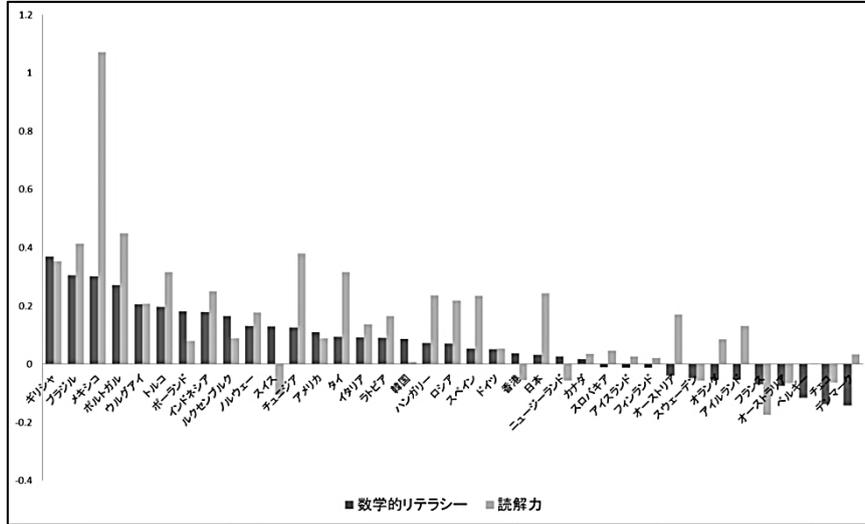


出典：筆者作成

一調査のテスト得点を目的変数に持つ階層線形モデル<sup>35</sup>による分析を試みた。なお本分析では、国際比較を意図しているため、各受験者がいずれかの国に属するという入れ子構造を活かし、国レベルと生徒レベルの2つの階層に焦点を絞ることとする。表 4.6 に本分析で使用する変数とその国レベルと生徒レベルにおける基本統計量を示した。ただし、テスト言語はテスト言語の使用を1とし、性別は男子を1とするダミー変数である。

<sup>35</sup> 本分析では、Rに含まれる階層線形モデルを実行するためのパッケージ `mlmRev` を使用した。

図 4.3 PISA2003 と PISA2009 の数学的リテラシーと読解力における各国の平均値の差の効果量（数学的リテラシーの効果量の降順）



出典：筆者作成

表 4.6 3つの PISA における国レベルと生徒レベルの記述統計量

変数名	国レベル				生徒レベル				
	最小値	最大値	平均値	SD	最小値	最大値	平均値	SD	
<b>【PISA2003】</b>									
数学的リテラシー	$y$	-0.794	0.531	0.056	0.367	-2.131	2.733	0.019	0.821
読解力	$x_1$	-0.759	0.437	0.028	0.335	-1.919	2.128	-0.001	0.810
テスト言語	$x_2$	0.006	0.999	0.864	0.235	0	1	0.868	0.339
性別	$x_3$	0.434	0.588	0.497	0.028	0	1	0.493	0.500
<b>【PISA2006】</b>									
数学的リテラシー	$y$	-0.794	0.531	0.056	0.367	-2.095	2.408	-0.022	0.781
読解力	$x_1$	-0.759	0.437	0.028	0.335	-1.975	2.136	-0.081	0.816
テスト言語	$x_2$	0.011	1.000	0.878	0.199	0	1	0.878	0.328
性別	$x_3$	0.418	0.533	0.494	0.023	0	1	0.492	0.500
<b>【PISA2009】</b>									
数学的リテラシー	$y$	-0.842	0.811	-0.066	0.402	-1.875	2.473	-0.053	0.765
読解力	$x_1$	-0.906	0.585	-0.051	0.360	-2.765	2.297	-0.033	0.778
テスト言語	$x_2$	0.017	0.999	0.849	0.230	0	1	0.873	0.333
性別	$x_3$	0.424	0.535	0.493	0.022	0	1	0.491	0.500

出典：筆者作成

本分析では4つのモデルを構成し、3つのPISAにおける各国の特徴をそれぞれ浮かび上がらせていく。なおここでは、同一の4つのモデルをPISA2003、PISA2006およびPISA2009に適用し、経年変化を視野に入れた分析を試みる。まず、説明変数を含まない以下のモデル1を考える。

$$\text{モデル 1: } y_{ij} = (\gamma_{00} + u_{0j}) + \varepsilon_{ij}$$

モデル1は一般にヌルモデルと呼ばれ、数学的リテラシー調査のテスト得点 $y_{ij}$ が受験者

全体の平均値を表わす固定効果 $\gamma_{00}$ とその国固有のランダム効果 $u_{0j}$ および生徒レベルの誤差項 $\varepsilon_{ij}$ の3つから成ると仮定している。ただし $y_{ij}$ は、 $j$ 国における $i$ 番目の生徒の数学的リテラシー調査のテスト得点である。このモデルを通じて、国の違いによってテスト得点の差異がどの程度説明されるのかを把握でき、さらに説明変数を投入したモデルの適合度を評価する基準を設定できる。

次に、モデル1で示される国別によるテスト得点の差異に対して、国レベルにおける読解力がどの程度影響しているかを調べるために、以下のようなモデルを考える。

$$\text{モデル 2: } y_{ij} = (\gamma_{00} + u_{0j}) + \gamma_{01}(x_{1\bullet j} - x_{1\bullet\bullet}) + \varepsilon_{ij}$$

モデル2は、モデル1に国レベルにおける5つの情意的側面を説明変数として投入している。ただし $x_{1\bullet j}$ は、表4.6にある変数 $x_1$ の $j$ 国における平均値をそれぞれ表し、 $x_{1\bullet\bullet}$ は、 $x_{1\bullet j}$ の平均値を意味する。このモデルによって、国レベルの各変数をその全体平均値に中心化させ、国別における差異を浮かび上がらせる。なおランダム効果については、モデル1と同様に $u_{0j}$ のみとなっている。

続いて、モデル2に国レベルのテスト言語と性別および生徒レベルの3つ説明変数を投入したモデルを考える。生徒レベルの変数を加えることで、国内におけるテスト得点のばらつき具合を検討できる。そのモデルは、以下の2つである。

$$\text{モデル 3: } y_{ij} = (\gamma_{00} + u_{0j}) + \gamma_{01}(x_{1\bullet j} - x_{1\bullet\bullet}) + \gamma_{02}(x_{2\bullet j} - x_{2\bullet\bullet}) + \gamma_{03}(x_{3\bullet j} - x_{3\bullet\bullet}) \\ + \gamma_{10}(x_{1ij} - x_{1\bullet j}) + \gamma_{20}(x_{2ij} - x_{2\bullet j}) + \gamma_{30}(x_{3ij} - x_{3\bullet j}) + \varepsilon_{ij}$$

モデル 4:

$$y_{ij} = (\gamma_{00} + u_{0j}) + \gamma_{01}(x_{1\bullet j} - x_{1\bullet\bullet}) + \gamma_{02}(x_{2\bullet j} - x_{2\bullet\bullet}) + \gamma_{03}(x_{3\bullet j} - x_{3\bullet\bullet}) \\ + (\gamma_{10} + u_{1j})(x_{1ij} - x_{1\bullet j}) + (\gamma_{20} + u_{2j})(x_{2ij} - x_{2\bullet j}) + \gamma_{30}(x_{3ij} - x_{3\bullet j}) + \varepsilon_{ij}$$

それぞれのモデルでは、生徒レベルの変数 $x_{kij}$ を各国の平均値 $x_{k\bullet j}$ に中心化している。ただしモデル3では、ランダム効果が $u_{0j}$ に限定されているが、モデル4においては、生徒レベルの読解力とテスト言語の回帰係数にもランダム効果 $u_{1j}$ と $u_{2j}$ がそれぞれ含まれている。つまりこの2つのモデルには、ランダム切片モデルとランダム切片・係数モデルという違いがある。本小節では、4つのモデルを用いて数学的リテラシー調査のテスト得点と読解力とテスト言語および性別、つまり説明変数は限定的であるが、数学学力と言語的側面の関連性を調べていく。

では、本分析の結果である表4.7をみていく。ここではPISA2003、PISA2006およびPISA2009に関する結果を提示したが、まずはPISA2003における結果を確認し、それを参考にしながらPISA2006とPISA2009の結果を検討する。PISA2003の分析結果にあるモデル1の国別分散と生徒間分散を用いて計算される級内相関係数(interclass correlation coefficient)を求めたところ0.203であった。この値は、117447名の受験者を国別でわずか38グループに分けた場合、生徒の数学的リテラシー調査のテスト得点のばらつきが約20.3%説明されること意味する。つまり本分析のように、国の違いに着目した階層線形モデルを用いる方法に一定の妥当性が認められる。

次に、モデル1とモデル2における国別分散に目を向ける。それらはそれぞれ0.134と

表 4.7 各 PISA における階層線形モデルによる固定効果の分析結果

			モデル1		モデル2		モデル3		モデル4	
			固定効果	S.E	固定効果	S.E	固定効果	S.E	固定効果	S.E
<b>【PISA2003】</b>										
	切片	$\gamma_{00}$	0.056	0.059	0.056	0.016 **	0.056	0.016 **	0.056	0.018 **
国レ	読解力	$\gamma_{01}$			1.058	0.048 **	1.065	0.055 **	0.960	0.049 **
ベル	テスト言語	$\gamma_{02}$					-0.046	0.077	-0.078	0.070
	性別	$\gamma_{03}$					0.253	0.628	-0.219	0.575
生徒	読解力	$\gamma_{10}$					0.532	0.002 **	0.547	0.018 **
レベ	テスト言語	$\gamma_{20}$					0.060	0.007 **	0.088	0.020 **
ル	性別	$\gamma_{30}$					0.162	0.004 **	0.164	0.004 **
	国別分散		0.134		0.009		0.010		0.034	
	生徒間分散		0.528		0.528		0.375		0.367	
	逸脱度		258485		258383		218215		215979	
	AIC		258495		258402		218274		216043	
<b>【PISA2006】</b>										
	切片	$\gamma_{00}$	-0.026	0.053	-0.026	0.019	-0.026	0.018	-0.026	0.018
国レ	読解力	$\gamma_{01}$			0.939	0.049 **	0.908	0.049 **	0.879	0.049 **
ベル	テスト言語	$\gamma_{02}$					-0.116	0.092	-0.116	0.092
	性別	$\gamma_{03}$					2.131	0.829 *	2.053	0.827 *
生徒	読解力	$\gamma_{10}$					0.506	0.002 **	0.508	0.013 **
レベ	テスト言語	$\gamma_{20}$					0.039	0.005 **	0.057	0.015 **
ル	性別	$\gamma_{30}$					0.163	0.003 **	0.165	0.003 **
	国別分散		0.154		0.020		0.018		0.036	
	生徒間分散		0.464		0.464		0.328		0.323	
	逸脱度		361590		361477		301039		298506	
	AIC		361600		361495		301098		298570	
<b>【PISA2009】</b>										
	切片	$\gamma_{00}$	-0.066	0.047	-0.066	0.015 *	-0.066	0.014 **	-0.066	0.015 **
国レ	読解力	$\gamma_{01}$			1.063	0.041 **	1.044	0.040 **	0.917	0.039 **
ベル	テスト言語	$\gamma_{02}$					-0.144	0.067 *	-0.104	0.065
	性別	$\gamma_{03}$					2.012	0.671 **	1.910	0.654 **
生徒	読解力	$\gamma_{10}$					0.571	0.001 **	0.568	0.011 **
レベ	テスト言語	$\gamma_{20}$					0.015	0.003 **	0.036	0.012 **
ル	性別	$\gamma_{30}$					0.191	0.002 **	0.193	0.002 **
	国別分散		0.162		0.016		0.013		0.033	
	生徒間分散		0.440		0.440		0.280		0.276	
	逸脱度		612570		612401		475405		470999	
	AIC		612581		612420		475468		471066	

注) S.Eは、標準誤差。

出典：筆者作成

0.009 であり、モデル 2 の方がモデル 1 よりも相当に小さく、その減少率は 93.11%にも及ぶ。モデル 1 とモデル 2 の違いは、国レベルにおける読解力の変数を持つかどうかであった。このことは読解力を投入することによって、国別の数学的リテラシーの得点の差異が約 9 割も説明されることを意味する。つまり国レベルで数学的リテラシー調査のテスト得点を見るならば、それに与える読解力の影響は決して無視できないことを意味する。続いて、モデル 2 に国レベルのテスト言語と性別および生徒レベルの読解力、テスト言語と性別を加えてモデル 3 をみると、モデル 2 の生徒間分散 0.528 よりもモデル 3 の値 0.375 の方が小さい。また、その減少率は 28.98%である。これは生徒レベルの変数を投入したこと、生徒レベルにおける数学的リテラシー調査のテスト得点のばらつきが 3 割近く説明さ

れることを意味する。しかしこの値は、国レベルの 93.11% と比べると随分小さい。このことは国別に対して生徒の違い、つまり個人差を説明することがより複雑であると解釈でき、生徒のレベルの考察に向けて更なる工夫を施す必要がある。最後に、生徒レベルに変数にランダム効果を加えたモデル 4 をみると、モデル 3 の国別分散 0.010 よりもモデル 4 の値 0.034 の方が大きい。これはランダム効果の投入によって、各国の差異をより捉えられていると考える。続いて、構成した 4 つのモデルとデータの適合度を確認する。その当てはまりの良さを表わす指標が、表 4.7 に示した逸脱度と AIC である。これらは、値が小さいほど適合度が高いモデルと判断される。そこで PISA2003 におけるそれらを確認すると、モデル 4 の値がともに最も小さいことが分かる。ゆえに PISA2003 においては、4 つのモデルのうちモデル 4 がデータと最も適合していると判断できる。

続いて、PISA2006 や PISA2009 における分析結果にも目を向ける。PISA2006 と PISA2009 におけるモデル 1 から算出される級内相関係数は、それぞれ 0.249 と 0.269 である。この結果は、2 つの PISA の受験者数 174418 名と 303431 名をそれぞれ国別でわずか 55 グループと 72 グループに分けた場合に、数学的リテラシー調査のテスト得点のばらつきが約 24.9% と 26.9% 説明されることを意味する。したがって PISA2006 と PISA2009 においても階層線形モデルによる分析に一定の妥当性を主張できよう。続いて、モデル 1 とモデル 2 における国別分散をみると、2 つの調査ともモデル 2 の値が大きく減少しており、その減少率は PISA2006 で 87.14%、PISA2009 で 90.39% にもなる。これは国レベルにおける読解力を投入することで、国別の数学的リテラシーの得点が 9 割近く説明されると理解される。つまりこの 2 つの PISA においても、読解力の影響は決して看過できないと結論できる。加えて逸脱度と AIC に目を向けると、モデル 4 の値が最も小さくなっている。つまり、PISA2003 の結果と同様にモデル 4 による分析が 4 つのモデルのうち最もデータと適合していると判断できる。

このように PISA2003 における結果と同じように PISA2006 と PISA2009 を読み解くならば、モデル 2 とモデル 3 間の生徒間分散の変動およびモデル 3 とモデル 4 間の国別分散の変動も PISA2003 と同様の傾向を持つ。つまり PISA2003 と PISA2006 および PISA2009 において、言語的側面を読解力とテスト言語に代表させて調べた結果、極めて大きな経年変化は確認されないと考える。一般に PISA のような国際的な調査では、なだらかな動向を持ちながら経年的な変化が観察できるとされる（日本テスト学会、2010、p.62）。本分析では、数学的リテラシーと読解力の関連性という視点から経年変化をみたが、大きな変化は確認できず、全体的にそれらの関連性も緩やかな動きを持つものといえるだろう。

続いて、表 4.7 にあるモデル 4 を用いた分析結果に着目する。まず、PISA2003 における結果に着目する。国レベルでは、読解力だけが統計的に有意であり、生徒レベルにおいては 3 つの変数で有意という結果を得た。そこで読解力が国別の数学的リテラシー調査のテスト得点に与える効果は、国レベルの分析結果 0.960 から生徒レベルの値 0.547 を引いた 0.413 となる。これは、表 4.6 にある PISA2003 の生徒レベルの数学的リテラシーの標準偏

差 0.821 の約半分を占める。このことから、読解力が及ぼす数学学力に対する文脈効果は小さくないことを意味する。つまり国別の数学学力の差異に対して、読解力の視座から考察する必要性を示唆している。

また生徒レベルに目を移すと、読解力の値が最も大きくなっている。また全受験者における傾向として、テスト言語の使用よりも性別の方により大きな効果があることも分かる。この結果は、もし男子でテスト言語を日常で使用し、読解力が全体の平均値より 1 標準偏差ほど高い生徒であれば、数学的リテラシー調査のテスト得点が 0.799<sup>36</sup>ほど平均的に高いことを意味する。生徒レベルの数学的リテラシーの標準偏差が 0.821 であることを考慮すれば、必ずしも小さな効果とはいえないだろう。とはいえ生徒レベルでは、読解力が数学的リテラシーの得点に最も大きな正の効果を与える、つまり正の方向に強い関連性を持つと理解するべきであろう。次に PISA2006 と PISA2009 における結果に目を移すと、国レベルでは、読解力ともに有意となっており文脈効果は、それぞれ 0.371 と 0.349 となった。また生徒レベルでは、3 つの変数ともに有意であり、得られた固定効果の符号は PISA2003 と同様である。ゆえに両調査において、もし男子でテスト言語を使用し、読解力が全体の平均値よりも 1 標準偏差ほど高い生徒であれば、数学的リテラシーの得点が、PISA2006 では 0.730、PISA2009 では 0.797 となる。このことは、それぞれの数学的リテラシーの標準偏差が 0.781 と 0.765 であることを踏まえれば、決して小さくない効果と思われる。特に PISA2009 では、数学的リテラシーの得点の 1 標準偏差以上の効果という結果を得た。

他方で、モデル 4 には、その国固有の特徴を反映するランダム効果が 3 種類含まれていた。そこで各国国内において言語的側面に関する変数、つまり読解力とテスト言語の使用がどの程度数学的リテラシー調査のテスト得点と関連するかを検討する。表 4.8 に、読解力とテスト言語に関する各国のランダム効果を含む回帰係数を提示した。なお、斜体の数字で各 PISA における各国の順位を表わしている。また 3 つの PISA におけるそれらの散布図を図 4.4 と図 4.5 に示した。まずこの 2 つの散布図に目を向けると、読解力における PISA2003 と PISA2006 の間の相関係数は 0.930 であり、PISA2006 と PISA2009 の間は 0.894、PISA2003 と PISA2009 では 0.880 であった。相関係数の大きさから判断すると、読解力の回帰係数に大きな経年変化はみられないと考える。しかし PISA2003 を基準に考えると、PISA2009 との値の方が PISA2006 より小さいため、若干の経年変化は捉えていとも思われる。他方でテスト言語における相関係数は、PISA2003 と PISA2006 および PISA2009 の間でそれぞれ 0.703 と 0.490 であり、PISA2006 と PISA2009 で 0.568 となった。相関係数が読解力と比べて小さいこと、および時間差が大きい PISA2003 と PISA2009 の間で値が小さいことから、テスト言語の使用に関する状況は、読解力と比較して一定の大きさを持ちながら経年的に変化しているといえるだろう。

これまでの分析で読解力は、全体的に生徒レベルにおいて最も大きな正の効果、つまり

<sup>36</sup> 表 4.7 にある PISA2003 のモデル 4 における生徒レベルの回帰係数の総和  $0.547+0.088+0.164$  で算出される。

表 4.8 PISA における読解力とテスト言語の回帰係数 (PISA2003 の読解力の降順)

国名	PISA2003 (38)		PISA2006 (55)		PISA2009 (72)							
	読解力	テスト言語	読解力	テスト言語	読解力	テスト言語						
香港	0.747	1	0.162	10	0.605	6	0.204	1	0.722	4	0.100	16
ニュージーランド	0.694	2	-0.121	38	0.616	5	-0.067	53	0.645	12	0.080	22
オランダ	0.689	3	0.064	23	0.600	7	0.092	20	0.646	11	0.123	10
韓国	0.670	4	0.044	26	0.622	4	0.083	23	0.690	6	0.063	29
ベルギー	0.660	5	0.020	31	0.626	3	0.014	41	0.693	5	0.056	32
チェコ	0.648	6	0.081	21	0.634	2	0.037	37	0.681	7	-0.021	57
アメリカ	0.646	7	0.020	30					0.619	24	0.107	15
アイルランド	0.635	8	-0.028	35	0.560	18	0.090	21	0.582	36	0.111	12
オーストラリア	0.633	9	0.004	32	0.593	10	-0.065	52	0.663	8	-0.056	65
ドイツ	0.632	10	0.133	14	0.587	11	0.106	15	0.650	10	0.138	4
日本	0.613	11	0.175	7	0.544	21	0.147	8	0.628	18	0.076	23
スイス	0.608	12	0.215	2	0.599	8	0.174	2	0.627	20	0.189	1
カナダ	0.603	13	-0.029	36	0.537	25	-0.059	51	0.618	25	-0.047	62
オーストリア	0.601	14	0.125	15	0.598	9	0.146	9	0.621	23	0.128	8
スロバキア	0.592	15	0.094	16	0.535	27	0.150	6	0.628	19	0.008	48
フランス	0.581	16	0.090	18	0.522	29	0.166	3	0.636	15	0.084	21
ルクセンブルク	0.580	17	0.002	33	0.560	19	0.094	19	0.605	29	0.064	28
アイスランド	0.575	18	0.089	19	0.571	15	0.134	10	0.607	28	0.038	39
デンマーク	0.564	19	0.157	12	0.519	30	0.148	7	0.562	44	0.185	2
フィンランド	0.564	20	0.061	24	0.535	26	0.131	11	0.568	41	-0.047	63
トルコ	0.554	21	0.202	3	0.499	36	0.110	14	0.590	34	0.088	20
ノルウェー	0.553	22	0.087	20	0.512	34	0.101	17	0.571	40	0.137	5
ポーランド	0.547	23	0.041	28	0.515	33	0.038	36	0.604	30	0.016	47
スウェーデン	0.546	24	0.162	9	0.539	23	0.116	12	0.580	38	0.071	24
ハンガリー	0.537	25	0.135	13	0.574	13	0.153	5	0.631	16	0.036	41
イタリア	0.524	26	-0.052	37	0.478	40	-0.048	50	0.580	39	-0.022	58
ポルトガル	0.506	27	-0.021	34	0.537	24	-0.014	45	0.643	13	-0.009	55
スペイン	0.504	28	0.042	27	0.539	22	0.087	22	0.623	22	0.029	43
ラトビア	0.494	29	0.074	22	0.487	39	0.050	33	0.566	42	0.002	52
旧セ	0.486	30	0.175	6								
タイ	0.455	31	0.182	5	0.515	31	0.078	26	0.580	37	0.045	37
ロシア	0.419	32	0.184	4	0.424	45	0.080	25	0.549	46	0.046	36
インドネシア	0.396	33	0.040	29	0.393	48	0.010	42	0.444	65	0.067	25
ギリシャ	0.377	34	0.160	11	0.450	43	0.100	18	0.508	53	0.091	19
ブラジル	0.350	35	0.093	17	0.396	47	-0.001	43	0.449	64	0.049	35
ウルグアイ	0.347	36	0.164	8	0.348	53	0.055	31	0.532	50	0.108	14
メキシコ	0.337	37	0.270	1	0.371	50	0.157	4	0.497	56	0.152	3
チュニジア	0.320	38	0.060	25	0.361	51	-0.018	46	0.407	69	0.018	46
平均値*	0.547		0.088		0.508		0.057		0.568		0.036	
標準偏差*	0.108		0.085		0.094		0.078		0.095		0.079	
中央値*	0.564		0.088		0.527		0.061		0.581		0.046	
最大値*	0.747		0.270		0.673		0.204		0.770		0.189	
最小値*	0.320		-0.121		0.230		-0.194		0.245		-0.205	
レンジ*	0.427		0.391		0.443		0.398		0.525		0.394	

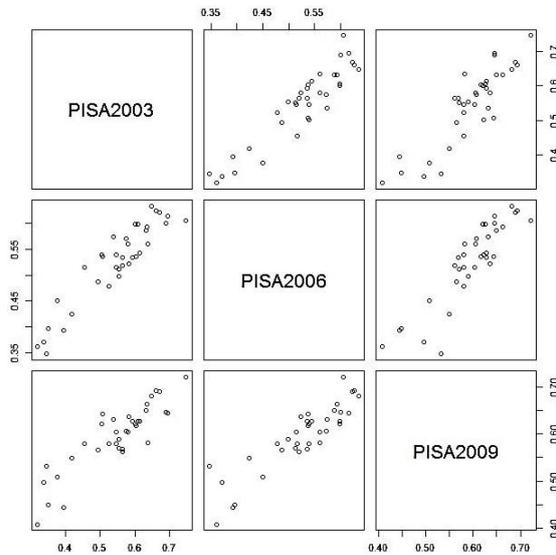
注) 旧セは、旧セルビア・モンテネグロ。空欄は、その年にデータがないことを意味する。

( )内は、各PISAにおける分析対象国の数を表わす。

\*は、各PISAにおける参加国全体で算出している。

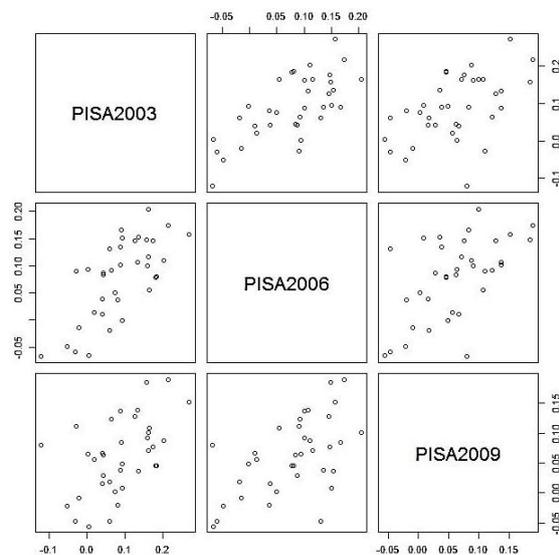
出典：筆者作成

図 4.4 読解力の回帰係数の散布図



出典：筆者作成

図 4.5 テスト言語の回帰係数の散布図



出典：筆者作成

正の方向に強い関連性を持つことが分かった。またランダム効果を含む各国の回帰係数を調べた結果、その経年変化は、3つの PISA の間で特に大きいと必ずしも判断できず、回帰係数も緩やかな動きの中で経年的に変化するものと思われる。また、表 4.8 にある読解力とテスト言語の回帰係数に関する基本統計量にみると、3つの PISA 全てにおいて、読解力の最小値がテスト言語の最大値を上回っている。これは参加国全てにおいて、読解力が有する正方向の関連性がテスト言語より大きいことを意味する。

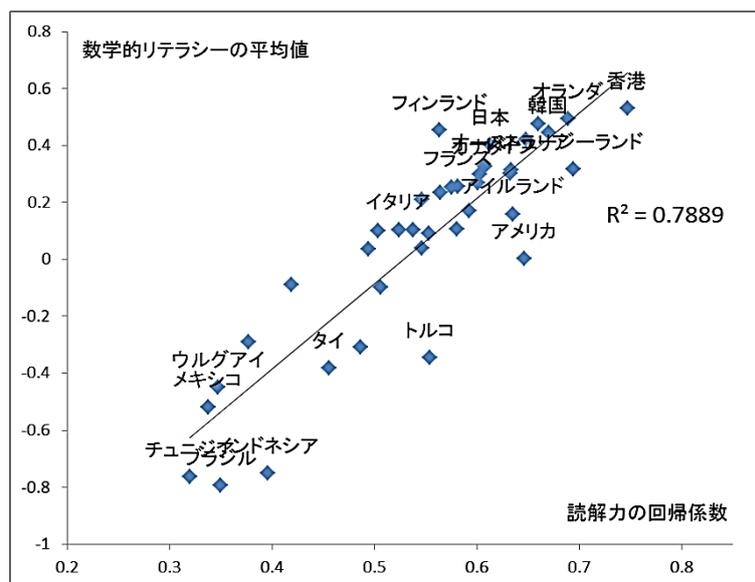
そこで図 4.2 と図 4.4 の散布図とそれらの各相関係数が高いことから、経年変化は必ずしも大きくないと判断できる。そのため、議論が煩雑になるのを防ぐために PISA2003 における各国の読解力の回帰係数に着目しながら各国の特徴を調べることにする。PISA2003 における分析結果をみると、例えば日本は、その値の大きさが 0.613 と全体の 11 番目にあたるため、国内における読解力と数学学力の関連性は、国際的にみて大きい国と捉えるべきと判断できる。この値は、日本国内において読解力が 1 標準偏差ほど高い生徒ならば、平均的に数学的リテラシーの得点が 0.613 程度高くなることを意味する。数学的リテラシーの標準偏差が 0.821 であることを考えれば、その効果の大きさは決して小さくない。したがって日本よりも大きな回帰係数を有する香港やニュージーランドでは、さらに大きな効果を持つことになる。他方で最も値の小さいチュニジアでは、その効果の大きさは上位層にある国の半分程度と読み取ることができる。

この分析結果に対して、第 2 章と第 3 章と同様の 13 先進国と 7 途上国、具体的に先進国は、オーストラリア、カナダ、フィンランド、フランス、ドイツ、香港、アイルランド、イタリア、日本、韓国、オランダ、ニュージーランド、アメリカの 13 カ国であり、途上国は、ブラジル、インドネシア、メキシコ、タイ、トルコ、チュニジア、ウルグアイの 7 カ

国を対象を絞るならば、一定の傾向が浮かび上がる。13先進国に関しては、イタリアの0.524が最も低く全体の26位にある。他方で途上国は、トルコが0.554の21位と目立って高いが他の国は、表4.8から分かるように最下位層に位置している。このように先進国と途上国の枠組みにおいて、国内における読解力の数学的リテラシー、つまり数学学力に対する正方向の関連性は、先進国では高く、逆に途上国では低いという傾向が浮かび上がった（図4.6<sup>37</sup>参照）。例えばメキシコやチュニジアにおける関連性の大きさは、香港の半分にも満たない。ここでは、数学学力と読解力の関連性の大きさに関して、テスト得点の水準の順序性を一定程度保持することに注目したい。また図4.4で示したように、経年変化は小さいことから今日においても数学的リテラシー調査のテスト得点が下位層にある途上国では、国内における読解力の正方向の関連性が先進国と比べて大きくないと思われる。

読解力といった言語的側面が数学の学習を進める上で重要な役割りを果たすことは明らかだろう。本小節では、数学学力の水準が低い国の国内における読解力と数学学力の関連性は国際的にみて決して強くないことを明らかにした。途上国に関する数学教育研究について、馬場（2008）は端緒についたばかりと指摘するが、PISAのような国際的な学力調査からは、その低さのみが強調され、有益な情報が得られにくいとされてきた（馬場・内田、2008；内田、2009）。ところが本分析では、情報を得ることが難しいとされながらも、途上国の特徴を先進国との対比を通じて計量的に浮かび上がらせた。文化・社会的な背景や生徒の学力水準など、多くの点で日本をはじめとする先進国と異なる途上国に理解に向けて、先進国と対比させる形で、更なる検討が必要と考える。

図 4.6 PISA2003 における数学的リテラシーの平均値と読解力の回帰係数の散布図

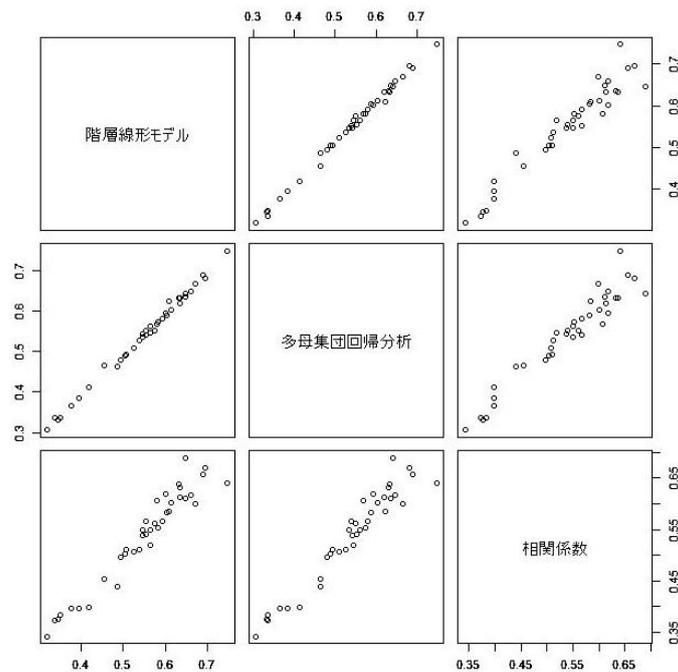


出典：筆者作成

<sup>37</sup> PISA2006 と PISA2009 における同様な散布図は、付録 4.1 と付録 4.2 を参照されたい。

なおここでは、階層線形モデルを用いて国別における読解力の影響の大きさなどを確認しながら、ランダム効果を含む回帰係数に数学学力と読解力あるいはテスト言語の各国の特徴を反映させて調べてきた。ところが、数学的リテラシーと読解力のテスト得点の関連性を調べる方法はいくつかある。そこで、数学的リテラシー調査のテスト得点を目的変数におき、読解力を説明変数に持つ多母集団回帰分析<sup>38</sup>および各国における数学的リテラシーと読解力のテスト得点の相関係数を用いた分析結果と比べてみる。その比較の結果を図4.7に散布図で示した。この図から分かるように、多母集団回帰分析から同様な結果を得ることが可能である。この3の分析方法による結果の相関係数を求めたところ、階層線形モデルと多母集団回帰分析では0.998であり、階層線形モデルと相関係数では0.966であった。また多母集団回帰分析と相関係数では、0.965となった。これらの分析結果から、国内における関連性について検討するならば、相関係数といった比較的扱いやすい指標を参考にすることも可能と思われる。

図4.7 階層線形モデルと多母集団回帰分析および相関係数の散布図



出典：筆者作成

#### 4.2. 13 先進国における読解力の水準別に注目した分析

これまでの分析から読解力は、数学学力に対して正の方向に強い関連性を有することは想像に難くないが、その程度に先進国と途上国の間で大きく異なることが明らかになった。

<sup>38</sup> 多母集団回帰分析については、豊田（1998）を参照されたい。また本分析では、Rに含まれる多母集団比較に対する共分散構造分析のパッケージ lavaan を使用した。

具体的には、先進国において強い関連性が認められ、途上国では逆の傾向を有するという結果である。本小節では、日本に関する分析を軸足に置きながら、先進国が有する特徴を検討する。なお経年変化が大きくないこと踏まえ、本小節では、数学的リテラシー調査における出題項目数が最も多いPISA2003のデータを分析していく。

#### 4.2.1. 異なる読解力の水準におけるテスト得点の分布と困難度に注目した解答パターン

本小節では、第3章と第4章で取り上げた先進国、具体的にはオーストラリア、カナダ、フィンランド、フランス、ドイツ、香港、アイルランド、イタリア、日本、韓国、オランダ、ニュージーランド、アメリカの13カ国に焦点を当てた分析を試みる。そこで、読解力のテスト得点を降順に均等三分割し、読解力の上位層、中位層および下位層の3つの水準を設定する。表4.9に設定した3つの水準ごとの各国の数学的リテラシー調査のテスト得点の分布を平均値と標準偏差を用いて要約した。アメリカとイタリアに関しては、他の国の中位層や下位層が一つ高い水準の平均値を上回る場合もみられるが、テスト得点は、概ね3つの水準で階層的であるといえよう。

次に、読解力の水準別でテストを構成する出題項目の難易度にどのような差異が生じるのか、をラッシュモデル(1.2)に含まれる母数の困難度に注目して検討する。PISA2003の数学的リテラシー調査では、全部で84項目がその測定のために使用された。しかし本分析のデータ処理の段階で、12項目<sup>39</sup>が分析対象から除外されている。また13カ国の3つの水準、つまり39の受験者集団を対象とした比較を行うために、集団ごとに算出される困難度に極端な値を示すものがないかを確かめなければならない。加えて本分析では、等化の精度をより高めるために、39の受験者集団間で分析対象とする項目を統一する。それに向けてまずOECD(2005a, p.190)によれば、残った72項目の中でイタリアおよびカナダで整合性が低いとされた項目が1つずつ指摘されている<sup>40</sup>。また39受験者集団それぞれで70項目の困難度をラッシュモデルを用いて求めたところ、芝(1991)と豊田(2002b)の基準を満たさない項目が4つ見つかった<sup>41</sup>。本分析では、一定の比較の信頼性を確保するために、これらの項目も分析対象とはせず、最終的に残った66項目を対象に分析を進めていく。なお等化については、前章と同様に式(1.6)を用いる方法を採用する。なお本分析では、日本の下位層集団を基準として等化を施す。そのため等化後に得られる各受験者集団の困難度の平均値は、日本の下位層における値に統一される。したがって、もし各集団が同程度の困難度を持つならば、それらは誤差の許す範囲で近い値になるはずである。なお等化後の得られる困難度は、66項目×39受験者集団の合計2574と膨大であるため、その提示は割愛する。

<sup>39</sup> M155Q01、M155Q02T、M155Q03T、M155Q04T、M420Q01T、M442Q02、M447Q01、M468Q01T、M484Q01T、M496Q01T、M496Q02、M509Q01の12項目。

<sup>40</sup> M179Q01TとM603Q02の2項目。

<sup>41</sup> M124Q03T、M446Q02、M505Q01、M810Q03Tの4項目。

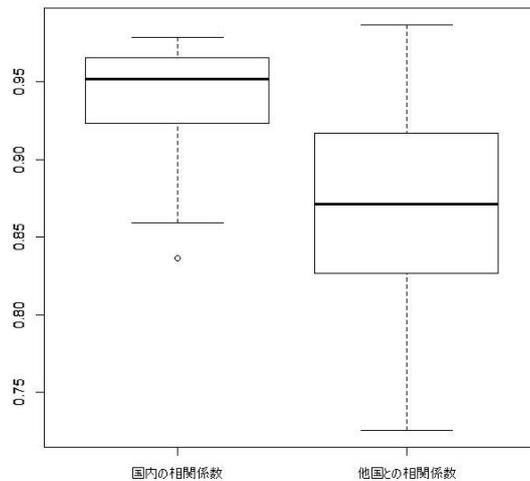
表 4.9 13 カ国の 3 つの水準におけるテスト得点の平均値と標準偏差（平均値の降順）

国別の読解力水準		平均値	標準誤差	標準偏差	標準誤差	受験者数
香港	上位	1.234	0.028	0.718	0.020	663
オランダ	上位	1.195	0.028	0.672	0.020	574
韓国	上位	1.083	0.027	0.785	0.019	839
ニュージーランド	上位	1.071	0.028	0.735	0.020	680
日本	上位	1.058	0.030	0.787	0.021	679
オーストラリア	上位	1.018	0.018	0.782	0.013	1887
フィンランド	上位	1.005	0.026	0.759	0.018	885
ドイツ	上位	0.956	0.029	0.723	0.021	620
カナダ	上位	0.911	0.012	0.748	0.008	4025
フランス	上位	0.858	0.032	0.804	0.023	632
アイルランド	上位	0.750	0.028	0.671	0.019	592
香港	中位	0.725	0.029	0.754	0.021	664
アメリカ	上位	0.674	0.025	0.703	0.018	806
イタリア	上位	0.635	0.020	0.838	0.014	1687
オランダ	中位	0.623	0.029	0.683	0.020	574
韓国	中位	0.596	0.026	0.751	0.018	840
フィンランド	中位	0.580	0.024	0.718	0.017	886
日本	中位	0.574	0.030	0.778	0.021	680
ドイツ	中位	0.437	0.029	0.732	0.021	620
ニュージーランド	中位	0.411	0.028	0.734	0.020	680
オーストラリア	中位	0.383	0.018	0.764	0.012	1888
カナダ	中位	0.355	0.012	0.733	0.008	4025
フランス	中位	0.315	0.030	0.761	0.021	632
アイルランド	中位	0.200	0.027	0.668	0.019	592
イタリア	中位	0.126	0.020	0.807	0.014	1688
フィンランド	下位	0.004	0.025	0.736	0.017	885
アメリカ	中位	-0.054	0.023	0.657	0.016	807
オランダ	下位	-0.087	0.029	0.692	0.020	574
香港	下位	-0.090	0.032	0.828	0.023	664
韓国	下位	-0.126	0.026	0.759	0.019	839
日本	下位	-0.233	0.032	0.840	0.023	680
カナダ	下位	-0.247	0.012	0.736	0.008	4025
フランス	下位	-0.303	0.034	0.846	0.024	632
オーストラリア	下位	-0.328	0.018	0.764	0.012	1887
ドイツ	下位	-0.364	0.029	0.724	0.021	620
ニュージーランド	下位	-0.392	0.029	0.762	0.021	680
アイルランド	下位	-0.438	0.030	0.720	0.021	592
イタリア	下位	-0.450	0.020	0.818	0.014	1687
アメリカ	下位	-0.681	0.022	0.617	0.015	807

出典：筆者作成

13 カ国の 3 つの水準における困難度のパターンの全体像を掴むために、 $39 \times 39$  の相関行列を作成し、国内の 3 つの水準間および他国の 3 つの水準との相関係数を確認した。国内における 39 個の相関係数と他国との 702 個の相関係数を区別し、その分布を図 4.8 のように箱ひげ図で表した。国内をみると、他国との相関係数よりも中央値が極端に大きく、レンジも小さい。つまり国内の 3 つの水準における相関係数は、13 カ国とも他国との間にある値よりも全体的に大きいといえる。実際、国内の相関係数で最も高い値を示したのは、フィンランドの上位層と中位層の間の 0.979 であり、逆に最も低かったのは、アメリカの

図 4.8 39 受験者集団における困難度の相関係数の分布



出典：筆者作成

上位層と下位層の 0.837 となった。国内においては、中央値が 0.952 であり約 8 割の相関係数が 0.900 以上を示すが、他国との相関係数に目を向けると、中央値が 0.871 であることから約半数以上が 0.900 を下回っている。なお他国との相関係数の最大値と最小値は、それぞれオーストラリアとニュージーランドの下位層間の 0.987 とアメリカの上位層と日本の下位層間の 0.725 である。この分析結果のように 39 受験者集団の特徴を調べたところ、各国とも 3 つの水準間で数学的リテラシー調査のテスト得点の分布が相違するにも拘らず、困難度に注目した解答パターンは、国内の 3 つの水準間でより類似する傾向が浮かび上がった。

他方で数学的リテラシー調査の出題項目は、「包括的アイディア（数学的な内容）」、「数学が用いられる状況・文脈」、「能力（数学的な過程）」の 3 つの評価枠組みに関する領域によって特徴付けられており、出題形式も自由記述形式や選択肢形式など様々である。そこで PISA が定める 3 つの評価枠組みの領域（以下、領域）と出題形式の種類別（以下、種別）を用いて、39 の受験者集団が有する困難度に注目した解答パターンを分析する。

まず、日本における分析結果を表 4.10<sup>42</sup>に提示する。この表にある「水準別平均値」には、各領域と種別における 3 つの水準の平均値を示し、その最大値から最小値を引いた値を「水準別比較」に記した。また「全体平均値」には、3 つの水準ごとではなく 3 つの水準を合わせた全体の平均値を提示している。この水準別比較の定義から、もしその値が 0.0 に近いならば、それは読解力の 3 つの水準における難易度の差異が小さい領域または種別と考えられ、大きい場合にはその逆と解釈され得る。つまり水準別比較は、読解力の 3 つの水準にある差異の大きさを表わす指標として扱える。表 4.10 の水準別比較をみると、最も小さい値は「求答」の 0.041 であり、最も大きい値は「自由記述」の 0.288 を得た。そこで、種

<sup>42</sup> 各領域と種別の規定は、付録 3.2 を参照いただきたい。

表 4.10 日本における読解力の水準別の困難度の平均値

領域と種別 ( )内は項目数	水準別平均値			水準別 比較	全体平 均値
	上位	中位	下位		
【包括的アイデア】				*	**
空間と形 (19)	0.607	0.558	0.554	0.052	0.573
不確実性 (15)	0.739	0.735	0.683	0.055	0.719
量 (17)	0.184	0.277	0.259	0.093	0.240
変化と関係 (15)	0.367	0.327	0.404	0.077	0.366
【状況・文脈】					
科学的 (10)	1.019	0.992	1.073	0.081	1.028
教育的 (14)	0.407	0.402	0.312	0.095	0.374
公共的 (23)	0.523	0.553	0.673	0.150	0.583
私的 (15)	0.136	0.134	-0.008	0.144	0.087
職業的 (4)	0.319	0.242	0.192	0.127	0.251
【能力】					
関連付け (30)	0.688	0.642	0.584	0.104	0.638
再現 (24)	-0.082	-0.051	-0.022	0.060	-0.051
熟考 (12)	1.046	1.101	1.185	0.140	1.111
【出題形式】					
求答 (12)	0.090	0.049	0.056	0.041	0.065
自由記述 (13)	0.954	0.990	1.241	0.288	1.062
選択肢 (15)	0.174	0.198	0.110	0.088	0.161
短答 (18)	0.421	0.429	0.413	0.017	0.421
複合的選択肢 (8)	0.945	0.884	0.668	0.276	0.833

出典：筆者作成

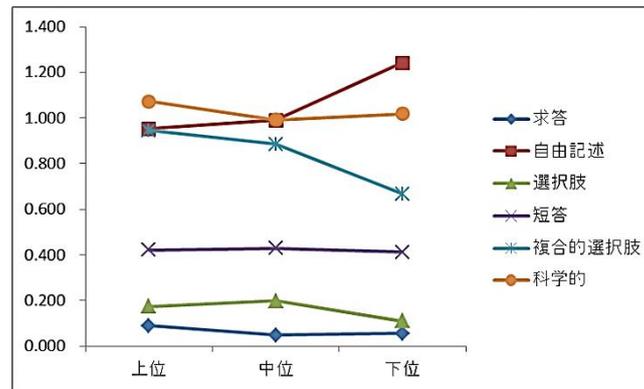
別に加えて「科学的」に対して得られた結果を図 4.9 に折れ線グラフで示した。水準別比較が最も大きかった「自由記述」は、水準が上位層であるほど難易度が低くなり、「複合的選択肢」は逆に高くなる事が分かる。ここで注目したい特徴は、水準が下位層になるほど「自由記述」の難易度が際立つことである。全体平均値に関しては、この値が高いほど難易度が高い領域また種別となるため、その値が高い「熟考」、「科学的」と「自由記述」の3つは特に難易度が高い領域または種別と判断できる。これらを踏まえ「自由記述」について整理すると、読解力の水準が低くなるほど難易度が目立って高くなり、全体的にも難しい種別といえるだろう。

これと同様な分析を他の12カ国でも行う。その分析結果として得られる12カ国の「水準別比較」を表 4.11<sup>43</sup>に提示した。加えて、表 4.12 に各領域と種別ごとの13カ国の「全体平均値」を記し、この表にある「13カ国間比較」の欄に13カ国中の最大値から最小値を引き算した結果を示した。この13カ国間比較は、その定義から読解力の水準別ではなく、各領域あるいは種別における国別の差異の大きさを表わす指標として位置付けられる。

まず表 4.11 に示した国内水準別比較をみると、日本において最も読解力の水準間で差異が大きかった「自由記述」は、他の12カ国においても同様に最も高いことが分かる。他方で、表 4.12 にある「自由記述」の13カ国間比較をみると、その平均値や中央値を参考にすれば、「科学的」や「教育的」ほどではないが、同様に国別の差異が認められる種別と判

<sup>43</sup> OECD が使用している国ラベルを用いて、国名を記載した。表 4.12 についても同様である。

図 4.9 日本の出題形式における困難度の水準別比較



出典：筆者作成

表 4.11 13 カ国における水準別比較の国際比較 (一の位の 0 は省略)

領域と種別 ( )内は項目数	AUS	CAN	DEU	FIN	FRA	HKG	IRL	ITA	JPN	KOR	NLD	NZL	USA
【包括的アイデア】									*				
空間と形 (19)	.022	.006	.088	.040	.068	.045	.059	.043	.052	.081	.075	.053	.028
不確実性 (15)	.049	.097	.051	.058	.042	.057	.011	.119	.055	.104	.069	.041	.084
量 (17)	.062	.031	.022	.036	.172	.020	.029	.112	.093	.021	.048	.053	.045
変化と関係 (15)	.147	.109	.178	.138	.143	.081	.073	.118	.077	.080	.130	.119	.158
【状況・文脈】													
科学的 (10)	.075	.113	.119	.040	.149	.049	.100	.032	.081	.048	.055	.084	.077
教育的 (14)	.178	.143	.228	.129	.055	.132	.080	.102	.095	.065	.141	.173	.149
公共的 (23)	.154	.118	.207	.167	.090	.122	.080	.096	.150	.100	.135	.166	.153
私的 (15)	.098	.116	.166	.104	.136	.082	.100	.051	.144	.089	.101	.161	.151
職業的 (4)	.091	.065	.071	.037	.072	.070	.115	.183	.127	.143	.091	.055	.120
【能力】													
関連付け (30)	.082	.069	.067	.097	.037	.049	.052	.089	.104	.094	.104	.072	.092
再現 (24)	.056	.024	.090	.017	.071	.018	.034	.076	.060	.059	.086	.045	.062
熟考 (12)	.154	.138	.133	.248	.113	.087	.068	.107	.140	.242	.116	.140	.106
【出題形式】													
求答 (12)	.203	.140	.135	.157	.084	.105	.049	.069	.041	.095	.214	.201	.145
自由記述 (13)	.376	.327	.405	.340	.300	.270	.279	.245	.288	.283	.390	.373	.464
選択肢 (15)	.076	.113	.087	.049	.079	.064	.165	.048	.088	.075	.072	.113	.158
短答 (18)	.055	.048	.052	.037	.081	.039	.066	.056	.017	.096	.042	.037	.019
複合的選択肢 (8)	.267	.198	.242	.221	.181	.167	.254	.232	.276	.189	.210	.200	.280
平均値	.126	.109	.138	.113	.110	.086	.095	.105	.111	.110	.122	.123	.135
中央値	.091	.113	.119	.097	.084	.070	.073	.096	.093	.094	.101	.113	.120
最大値	.376	.327	.405	.340	.300	.270	.279	.245	.288	.283	.390	.373	.464
最小値	.022	.006	.022	.017	.037	.018	.011	.032	.017	.021	.042	.037	.019
レンジ	.354	.321	.384	.323	.263	.252	.267	.213	.271	.262	.348	.336	.445

出典：筆者作成

断できる。つまり「自由記述」は、全体的な国別の差異は一定程度認められるが、国内における読解力の水準別では看過できない差が認められる種別という国際的な傾向が浮かび上がった。次に表 4.12 をみると、「科学的」、「教育的」あるいは「職業的」といった「状況・

表 4.12 13 カ国における全体平均値の国際比較（一の位の0は省略）

領域と種別	AUS	CAN	DEU	FIN	FRA	HKG	IRL	ITA	JPN	KOR	NLD	NZL	USA	13カ国間比較
【包括的アイデア】									**					
空間と形	.720	.782	.737	.758	.747	.610	.881	.699	.573	.663	.837	.702	.761	.308
不確実性	.604	.569	.756	.663	.686	.544	.527	.693	.719	.705	.504	.604	.551	.252
量	.213	.149	.070	.064	.149	.163	.115	.087	.240	.187	.191	.249	.180	.185
変化と関係	.326	.354	.314	.387	.280	.580	.310	.404	.366	.325	.301	.307	.363	.300
【状況・文脈】														
科学的	.779	.753	.623	.704	.551	.982	.683	.670	1.028	.679	.657	.753	.706	.478
教育的	.591	.610	.698	.632	.772	.478	.682	.764	.374	.555	.626	.585	.579	.398
公共的	.500	.497	.520	.490	.477	.390	.468	.492	.583	.566	.517	.501	.589	.199
私的	.144	.154	.099	.164	.162	.225	.197	.096	.087	.134	.152	.165	.092	.138
職業的	.379	.355	.442	.404	.381	.596	.287	.271	.251	.413	.429	.375	.287	.345
【能力】														
関連付け	.651	.666	.689	.670	.669	.667	.669	.644	.638	.707	.715	.635	.641	.079
再現	-.071	-.102	-.116	-.127	-.098	-.047	-.074	-.038	-.051	-.095	-.201	-.074	-.042	.163
熟考	1.117	1.144	1.112	1.184	1.126	1.030	1.077	1.069	1.111	1.024	1.217	1.163	1.085	.193
【出題形式】														
求答	.033	.081	.006	.095	.061	.274	.020	.125	.065	.105	.042	.002	.092	.272
自由記述	1.208	1.182	1.280	1.178	1.083	.990	1.266	1.211	1.062	1.073	1.203	1.200	1.181	.290
選択肢	.284	.262	.264	.257	.316	.135	.251	.256	.161	.085	.226	.305	.155	.231
短答	.259	.236	.233	.243	.277	.305	.201	.212	.421	.386	.261	.265	.325	.221
複合的選択肢	.778	.840	.796	.820	.834	.947	.895	.792	.833	.976	.875	.786	.825	.197
平均値	.501	.502	.501	.505	.499	.522	.497	.497	.498	.499	.503	.501	.492	.250
中央値	.500	.497	.520	.490	.477	.544	.468	.492	.421	.555	.504	.501	.551	.231
最大値	1.208	1.182	1.280	1.184	1.126	1.030	1.266	1.211	1.111	1.073	1.217	1.200	1.181	.478
最小値	-.071	-.102	-.116	-.127	-.098	-.047	-.074	-.038	-.051	-.095	-.201	-.074	-.042	.079
レンジ	1.279	1.284	1.396	1.311	1.224	1.077	1.340	1.249	1.162	1.168	1.418	1.274	1.223	.398

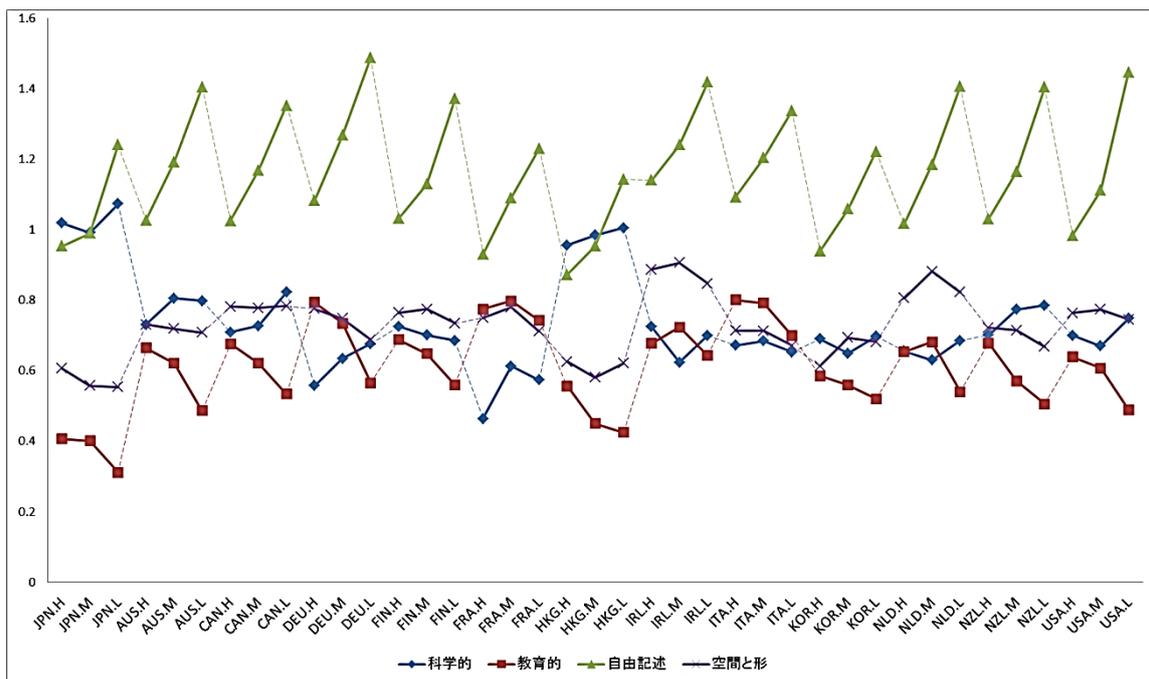
出典：筆者作成

文脈」において 13 カ国間比較の値が高いという結果を得た。ところが、これらの表 4.11 の国内水準別比較は「自由記述」と比較すると特に大きな差異があると捉えることはできない。特に「科学的」は 13 カ国間比較が大きいにも拘らず、国内水準別比較は、小さくなっている。つまり上述の 4 つの領域と種別では、「科学的」がより明示的であるが、国内の読解力の水準別よりも国別における差異が反映されやすいと判断できるだろう。

これまでの分析結果を纏めると、困難度に着目した解答パターンが顕在化する国際的な傾向として、①国内における読解力の水準別に解答パターンが顕在化しやすいのは「自由記述」であること、②読解力の水準別ではなく国別でそれが現れやすい場合があり、それは「科学的」や「教育的」という領域であることが明らかになった。そこで図 4.10 に、「科学的」、「教育的」、「自由記述」および「空間と形」における 13 カ国の 3 つの水準の平均値を折れ線グラフで表現した。なおこの図では、OECD が使用している国ラベルを用いて国名を示し、国名と共に記した H と M および L は、それぞれ上位層、中位層、下位層を意味する。①に関して「自由記述」は、どの国においても読解力の水準が低くなるほど、目立って高い困難度の平均値、つまり高い難易度を有するという特徴を掴むことができる。また②について、国別により大きな差異がみられる「科学的」をみると、日本や香港で高

い難易度の領域であることが分かる(図4.10参照)。しかしこの2カ国において「教育的」は、逆に易しい傾向、特に日本では国際的にみて特に容易に正答を得やすい領域といえよう。このように読解力の水準に依らないで、国別で顕在化する国際的な傾向も同時に浮かび上がった。

図4.10 「科学的」、「教育的」、「自由記述」、「空間と形」における困難度の国内水準別比較



出典：筆者作成

#### 4.2.2. 先進国に関する分析結果の総合的考察

本小節では、先進国に焦点を当てて分析を進めてきた。その結果13先進国に共通する傾向として、①国内における読解力の水準別に困難度に注目した解答パターンが顕在化しやすい領域・種別は「自由記述」であること、読解力の水準別でなく国別でそれが現れやすい場合があり、それは②「科学的」や「教育的」という領域に特にみられることの2点が浮かび上がった。ここで取り上げた13先進国においては、前小節の分析結果からイタリアが若干低い位置にあるが、回帰係数の大きさを考慮すれば、全体的に数学学力に対する読解力の正方向の関連性は国際的にみてと高いといえる(表4.8参照)。ところがそのような状況にも拘らず、読解力の水準の違いは、「自由記述」という種別で顕在化するという国際的な傾向が明らかになった。PISAの数学的リテラシー調査のように文章問題で構成されるテストを用いて生徒の数学学力を測定する場合には、当然ながら一定程度の読解力が求められる。ところが本分析では、「自由記述」において目立った特徴が顕在化した。もし問題

文を読み解くことに困難性を有するならば、自由記述形式に特化して読解力の水準別の違いが現れるとは考えにくい。ゆえに文章問題の題意を読み解くことだけでなく、答えを導いた考え方や求め方あるいは理由を説明するといった「記述する」という場面においても読解力が大きく関係していると解釈できよう。つまり「記述する」ことに対して、読解力を視野に入れながら考察する有効性を国際的な普遍性を確保して主張できよう。前章でも取り上げたが、日本においては、自由記述形式の出題項目に対する無答率の高さがしばしば問題視される（国立教育政策研究所、2010；室岡、2006）。もし高い無答率の改善を試みるならば、読解力の水準別が「自由記述」で顕在化しやすいことから、読解力の視座から検討する意義は小さくないだろう。なぜならば近年の数学教育では、表現力の育成が研究課題の一つに位置付いており（齋藤・藤田、1998；金本、2010）、読み解くこと、つまり問題文を理解し、その上で数学的な考えを表現することには、一定の連続性を仮定できるからである。

加えて表 4.9 をみると、13 ヶ国における読解力の 3 つの水準別に数学的リテラシー調査のテスト得点の平均値と標準偏差から、その分布は同様であるとは言い難かった。ところが 39 の受験者集団で困難度に注目した解答パターンを調べたところ、国内の水準別における類似性の方が国別より強く（図 4.8）、全体的には国別の方に解答パターンの違いが顕在化しやすいことが浮かび上がった（表 4.11 と表 4.12）。つまり、テスト得点の分布が大きく異なる受験者集団よりも、その水準が同程度と考えられる国別の方に解答パターンの差異が現れやすいことになる。このことから読解力の水準別だけでなく、国別に注目した分析を行えるだろう。例えば図 4.10 から日本と香港では、読解力の水準別に依らず、「科学的」の難易度が目立って高い。その一方で「教育的」においては、むしろ低い難易度と解釈できる。つまり鈴川・豊田・川端（2008）が指摘するように、国別に目を向けた場合、日本や香港の生徒は数学の授業で典型的にみられるような疑似現実的な内容が含まれる「教育的」な文脈の問題を国際的にみて難無くこなす一方で、より抽象的とされる「科学的」な文脈では、他の国と比べて高い難易度を有することが明らかになった。「科学的」については、技術的な過程、理論的な場面と明らかに数学的な問題についての理解（国立教育政策研究所、2004）と規定されるように、数学教育の中でその育成が期待される中心的なものの一つであろう。ところが日本に加え香港の生徒は、そのような文脈で数学を活用する能力が同程度の数学学力にある国と比べて弱いことを指摘できる。特に「科学的」という領域では、国内水準別比較が小さく 13 ヶ国間比較が大きいため、読解力の水準別に依存せず、各国の数学教育により起因する固有性が反映されやすい領域とも考えられるだろう。

#### 4.3. 7 途上国における読解力の水準別に注目した分析

階層線形モデルの分析から読解力は、先進国においてより強い関連性が認められ、途上国は全体としてその逆の傾向にあることが浮かび上がった。本小節は、数学的リテラシー調査のテスト得点が下位層にある途上国に焦点を当てて分析を試みる。なお数学学力の経

年変化が大きくないことを踏まえ、本小節でも、PISA2003 のデータを分析対象とする。

#### 4.3.1. 異なる読解力の水準におけるテスト得点の分布と困難度に注目した解答パターン

本小節では、前章で取り上げた途上国、具体的にはトルコ、ウルグアイ、タイ、メキシコ、ブラジル、チュニジア、インドネシアの7カ国に焦点を当てながら数学学力と読解力の関連性を調べていく。そこで、読解力のテスト得点を降順に均等三分割し、読解力の上位層、中位層、下位層の3つの水準を設定する。表4.13に設定した3つの水準ごとの各国の数学的リテラシー調査のテスト得点の分布を平均値と標準偏差を用いて要約した。この表から、インドネシア、チュニジアおよびブラジルの中位層は、タイ、ウルグアイ、トルコとメキシコの下位層よりも低い水準にあることが分かる。ゆえに前者の3カ国と後者の4カ国に分かれる形で階層的であるといえよう。

次に、読解力の水準別でテストを構成する出題項目の難易度にどのような差異が生じるのかを検討する。前小節と同様に、ここでもラッシュモデル(1.2)に含まれる難易度に関する母数の困難度を活用する。その理由は、等化を行うことで正答率とは異なり、比較可能な数値に変換できるためである。

7カ国の3つの水準、つまり21受験者集団を対象として困難度の比較を行う。そこで各集団で極端な値を示す困難度がないかを確認する。それに向けてOECD(2005a)によれば、ウルグアイにおいて整合性の低い項目が2つ指摘されている。21の受験者集団で項目

表4.13 7カ国の3つの水準におけるテスト得点の平均値と標準偏差(平均値の降順)

国別の読解力水準	平均値	標準誤差	標準偏差	標準誤差	受験者数	
トルコ	上位	0.169	0.034	0.939	0.024	751
タイ	上位	-0.034	0.032	0.896	0.022	794
ウルグアイ	上位	-0.097	0.035	1.023	0.024	874
メキシコ	上位	-0.296	0.012	0.802	0.008	4478
トルコ	中位	-0.566	0.027	0.747	0.019	751
タイ	中位	-0.577	0.026	0.737	0.018	795
ブラジル	上位	-0.581	0.035	0.893	0.025	659
インドネシア	上位	-0.589	0.019	0.769	0.014	1594
チュニジア	上位	-0.660	0.030	0.801	0.021	697
ウルグアイ	中位	-0.702	0.029	0.856	0.020	875
メキシコ	中位	-0.708	0.010	0.699	0.007	4478
タイ	下位	-0.861	0.024	0.665	0.017	795
ウルグアイ	下位	-0.914	0.026	0.768	0.018	875
トルコ	下位	-0.933	0.025	0.699	0.018	751
メキシコ	下位	-0.960	0.010	0.692	0.007	4478
インドネシア	中位	-1.004	0.016	0.644	0.011	1595
チュニジア	中位	-1.010	0.025	0.662	0.018	697
ブラジル	中位	-1.108	0.027	0.696	0.019	660
チュニジア	下位	-1.191	0.025	0.660	0.018	697
インドネシア	下位	-1.226	0.015	0.608	0.011	1594
ブラジル	下位	-1.284	0.026	0.673	0.019	660

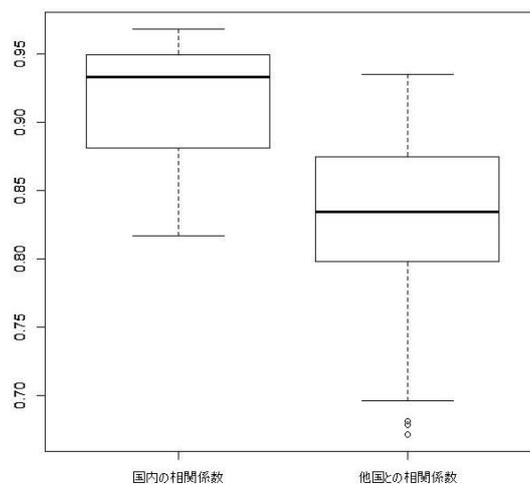
出典：筆者作成

数を揃えるために、この2つの項目は分析の対象外とした。また本分析におけるデータ処理の段階で、12項目が分析対象から除外されている。つまり全84項目のうち、70項目に関する困難度を21受験者集団で芝(1991)と豊田(2002b)の基準を参考に確認した。その結果、6項目<sup>44</sup>が分析には不適切と判断できたため、これらの項目も分析に含めず、最終的に残った64項目を分析対象とした。なお等化については、式(1.6)を用いる方法を採用する。この方法を用いれば、各集団で得られる64項目の困難度の平均値が同一の値に揃えられるため、もし各集団が同程度の困難度を持つならば、それらは誤差の許す範囲で近い値になるはずである。等化後に得られる困難度は、64項目×21受験者集団の合計1344と大きいため、その提示は割愛する。また本分析では、ブラジルの下位層に等化を施した。

7カ国の3つの水準における困難度のパターンの全体像を掴むために、21×21の相関行列を作成し、国内の3つの水準間および他国の3つの水準との相関係数を確認した。国内における21の相関係数と他国との189を区別し、その分布を図4.11に箱ひげ図で要約した。国内をみると、他国との相関係数よりも中央値が大きく、レンジは同程度である。したがって国内の3つの水準における相関係数は、7カ国とも他国との間にある値よりも大きいといえるだろう。実際、国内の相関係数における中央値は0.933であり、他国との相関係数は0.834である。このことから国内では、半数以上が0.900を超える相関係数であるが、他国との相関係数は大半が0.900を下回っている。纏めると21受験者集団の困難度に注目した解答パターンを調べて結果、各国とも3つの水準間でテスト得点の分布が相違するにも拘らず、国内の3つの水準間でより類似する傾向にあるといえよう。

数学的リテラシー調査の出題項目は、「包括的アイディア(数学的な内容)」、「数学が用いられる状況・文脈」、「能力(数学的な過程)」の3つの評価枠組みに関する領域によって

図 4.11 39 受験者集団における困難度の相関係数の分布



出典：筆者作成

<sup>44</sup> M124Q03T、M406Q02、M406Q03、M421Q02T、M446Q02、M462Q01T の6項目。

特徴付けられており、出題形式も自由記述形式や選択肢形式など様々であった。そこで PISA が定める 3 つの評価枠組みの領域（以下、領域）と出題形式の種類別（以下、種別）を用いて、21 の受験者集団が有する困難度に注目した解答パターンを分析する。

例えば、ブラジルにおける分析結果を表 4.14 に示す。この表にある「水準別平均値」には、前小節と同様に各領域と種別における 3 つの水準の平均値を示し、その最大値から最小値を引いた結果を「水準別比較」に記載した。また同様に「全体平均値」は、3 つの水準ごとではなく、3 つの水準を合わせた全体の平均値を表わしている。このような規定から、もし水準別比較の値が 0.0 に近いならば、それは読解力の 3 つの水準間で困難度の差異が小さい領域または種別であり、逆に大きい場合にはその逆と判断できる。つまり水準別比較は、読解力の 3 つの水準にある差異の大きさを表す指標となる。水準別比較をみると、最も小さい値は「科学的」の 0.013 であり、最も大きいものは「複合的選択肢」の 0.386 となっている。また、「自由記述」においても比較的大きい値を示している。そこで図 4.12 に、種別ごとの困難度の平均値を折れ線グラフで表した。全体平均値については、この値が高いほど難易度が高い領域あるいは種別となるため、2.400 と値が最も大きい「自由記述」は、最も難しいと判断できる。

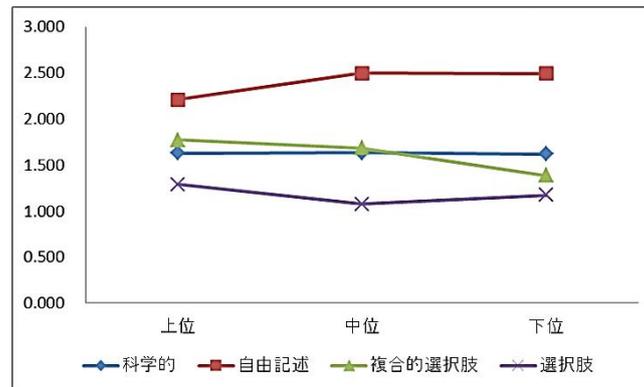
これと同様な分析を他の 6 カ国にも行い、各国の「水準別比較」を示したのが表 4.15 である。また表 4.16 に各領域と種別ごとの各国の「全体平均値」を示し、この表にある「7 カ国間比較」に 7 カ国中の最大値から最小値を引いた結果を提示した。この 7 カ国間比較

表 4.14 ブラジルにおける読解力の水準別の困難度の平均値

領域と種別 ( )内は項目数	水準別平均値			水準別 比較	全体平 均値
	上位	中位	下位		
<b>【包括的アイデア】</b>				*	**
空間と形 (17)	1.751	1.670	1.648	0.102	1.690
不確実性 (15)	1.648	1.559	1.579	0.089	1.595
量 (16)	1.375	1.417	1.405	0.041	1.399
変化と関係 (16)	1.556	1.684	1.701	0.145	1.647
<b>【状況・文脈】</b>					
科学的 (10)	1.631	1.632	1.619	0.013	1.627
教育的 (12)	1.748	1.628	1.625	0.123	1.667
公共的 (23)	1.680	1.697	1.757	0.077	1.711
私的 (15)	1.271	1.317	1.295	0.046	1.294
職業的 (4)	1.603	1.689	1.463	0.226	1.585
<b>【能力】</b>					
関連付け (29)	1.768	1.737	1.638	0.130	1.715
再現 (23)	1.172	1.210	1.281	0.110	1.221
熟考 (12)	1.931	1.931	2.033	0.103	1.965
<b>【出題形式】</b>					
求答 (12)	1.279	1.207	1.287	0.080	1.258
自由記述 (13)	2.210	2.498	2.493	0.288	2.400
選択肢 (14)	1.287	1.077	1.173	0.210	1.179
短答 (18)	1.493	1.530	1.521	0.038	1.515
複合的選択肢 (7)	1.774	1.685	1.388	0.386	1.616

出典：筆者作成

図 4.12 ブラジルの「科学的」、「自由記述」、「複合的選択肢」、「選択肢」の水準別平均値



出典：筆者作成

表 4.15 7カ国における水準別比較の国際比較（一の位の0は省略）

領域と種別 ( )内は項目数	IND	MEX	THA	TUN	TUR	URY	BRA
【包括的アイデア】							*
空間と形 (17)	.025	.032	.043	.053	.153	.069	.102
不確実性 (15)	.152	.087	.113	.128	.114	.005	.089
量 (16)	.088	.071	.050	.088	.110	.057	.041
変化と関係 (16)	.113	.039	.051	.022	.118	.043	.145
【状況・文脈】							
科学的 (10)	.052	.072	.112	.206	.097	.152	.013
教育的 (12)	.061	.053	.107	.071	.114	.143	.123
公共的 (23)	.096	.029	.067	.015	.107	.069	.077
私的 (15)	.088	.073	.011	.113	.010	.145	.046
職業的 (4)	.086	.107	.067	.022	.064	.217	.226
【能力】							
関連付け (29)	.125	.087	.124	.047	.115	.078	.130
再現 (23)	.151	.066	.151	.058	.147	.061	.110
熟考 (12)	.018	.083	.031	.004	.032	.074	.103
【出題形式】							
求答 (12)	.091	.032	.126	.078	.090	.098	.080
自由記述 (13)	.205	.167	.182	.104	.213	.196	.288
選択肢 (14)	.223	.075	.142	.128	.166	.003	.210
短答 (18)	.107	.033	.064	.111	.057	.068	.038
複合的選択肢 (7)	.365	.284	.358	.123	.168	.385	.386
平均値	.120	.082	.106	.081	.110	.110	.130
中央値	.096	.072	.107	.078	.114	.074	.103
最大値	.365	.284	.358	.206	.213	.385	.386
最小値	.018	.029	.011	.004	.010	.003	.013
レンジ	.347	.255	.348	.202	.203	.382	.373

出典：筆者作成

は、その算出方法から読解力の水準別ではなく、各領域あるいは種別における国別の差異を表わす指標となる。

表 4.15 で示した国内水準別比較をみると、ブラジルで読解力の水準別で最も差異の大き

表 4.16 7カ国における全体平均値の国際比較（一の位の0は省略）

領域と種別 ( )内は項目数	IND	MEX	THA	TUN	TUR	URY	BRA	13カ国 間比較
【包括的アイデア】							**	
空間と形 (17)	1.562	1.617	1.539	1.533	1.692	1.683	1.690	.160
不確実性 (15)	1.662	1.703	1.711	1.761	1.611	1.746	1.595	.166
量 (16)	1.319	1.259	1.396	1.403	1.477	1.365	1.399	.217
変化と関係 (16)	1.800	1.763	1.701	1.654	1.551	1.547	1.647	.253
【状況・文脈】								
科学的 (10)	1.825	1.666	1.686	1.633	1.560	1.619	1.627	.265
教育的 (12)	1.513	1.568	1.420	1.618	1.601	1.748	1.667	.328
公共的 (23)	1.579	1.727	1.674	1.658	1.669	1.685	1.711	.147
私的 (15)	1.460	1.340	1.511	1.464	1.480	1.317	1.294	.216
職業的 (4)	1.691	1.522	1.582	1.381	1.493	1.426	1.585	.310
【能力】								
関連付け (29)	1.742	1.722	1.726	1.761	1.787	1.735	1.715	.073
再現 (23)	1.177	1.171	1.175	1.207	1.102	1.172	1.221	.119
熟考 (12)	1.983	2.043	2.026	1.879	2.017	2.011	1.965	.164
【出題形式】								
求答 (12)	1.291	1.254	1.092	1.244	1.288	1.348	1.258	.256
自由記述 (13)	2.335	2.324	2.297	2.233	2.321	2.153	2.400	.247
選択肢 (14)	1.272	1.389	1.439	1.386	1.298	1.504	1.179	.325
短答 (18)	1.423	1.314	1.428	1.418	1.427	1.360	1.515	.201
複合的選択肢 (7)	1.730	1.861	1.795	1.785	1.698	1.668	1.616	.246
平均値	1.610	1.603	1.600	1.589	1.593	1.593	1.593	.217
中央値	1.579	1.617	1.582	1.618	1.560	1.619	1.616	.217
最大値	2.335	2.324	2.297	2.233	2.321	2.153	2.400	.328
最小値	1.177	1.171	1.092	1.207	1.102	1.172	1.179	.073
レンジ	1.158	1.152	1.205	1.026	1.219	.981	1.221	.255

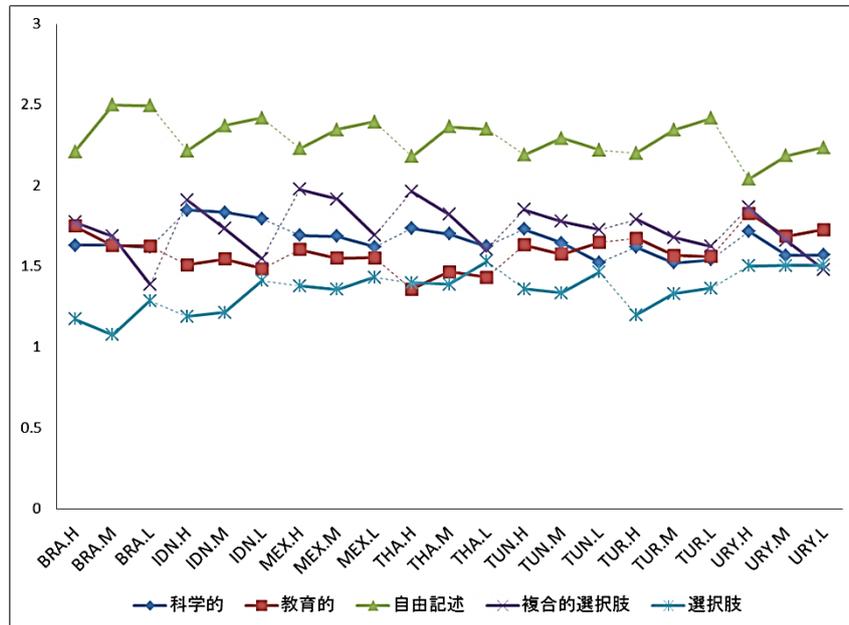
出典：筆者作成

かった「複合的選択肢」は、7カ国の中で5カ国において同様の傾向を有することが分かる。この分析結果は、前小節の13先進国で一貫してみられた特徴を異なる。「複合的選択肢」は、与えられた選択肢の中から選択する問いが連続している項目とされる<sup>45</sup>。本小節で焦点を当てた7途上国においては、全体的に国内水準別比較、つまり読解力の水準の差異が「複合的選択肢」に現れやすいという傾向が浮かび上がった。他方で表4.16にある7カ国間比較をみると、その平均値と中央値を参考にすれば、種別と「教育的」、「科学的」、項目数は少ないが「職業的」では、値が大きいことが分かる。そこで図4.13に「科学的」、「教育的」、「自由記述」、「複合的選択肢」および「選択肢」における7カ国の3つの水準の平均値を折れ線グラフで示した。この図から「自由記述」は、全体的に難易度が高い種別であるが、前小節で取り上げた13先進国とは異なり、読解力の水準別の違いが顕在化しにくい傾向を確認できる。

これまでの分析結果を前小節の先進国に関する結果と対比させながら纏めると、数学学力と読解力の水準が下位層にある途上国では、①「自由記述」は国内の読解力の水準別が必ずしも顕在化しやすい種別ではないこと、②読解力の水準別ではなく国別で差異が現れ

<sup>45</sup> 例えば、付録4.3に示した項目（項目名「花壇」）のような形式である。

図 4.13 「科学的」、「教育的」、「自由記述」、「複合的選択肢」および「選択肢」における困難度の国内水準別比較



出典：筆者作成

やすいのは、「教育的」や「科学的」あるいは出題形式であることである。図 4.13 では、先進国に関する分析結果の図 4.10 と比べると、読解力の水準別と国別の差異が現れにくいことを読み取れるだろう。

#### 4.3.2. 途上国に注目した分析結果の総合的考察

本小節では、途上国に焦点を当てて分析を進めてきた。ここで得た分析結果を 13 先進国に関する結果と照らし合わせると、「自由記述」という種別で読解力の水準別の差異が現れにくく、「複合的選択肢」の方に顕在化しやすいことが分かった。また国別においては、先進国と同様に「教育的」や「科学的」にいくらか表面化することを掴んだ。

では、途上国に関する結果を先進国と対比させながら考察する。13 先進国は、数学学力と読解力の国際的な水準は高く、それらの国内における関連性は国際的にみて強かった。そこで、もし読解力が数学学力の向上に寄与するならば、高い読解力の水準の国ほど、読解力と数学学力の関連性は強いと判断される。そのため先進国の高い数学学力の背景には、それを担保する読解力の存在を否定できない。これを翻って途上国に目を向けると、低い数学学力と称され得る途上国の背景には、数学学力を向上させるだけの読解力が身に付いていないとも解釈できる。また先進国では、文章問題で構成されるテストにおいて自由記述形式で読解力の水準別が目立って顕在化したため、「記述する」ことに読解力が関係するとした。ところが途上国では、与えられた選択肢の中から選択する問いが連続するという

幾分か形式が複雑な「複合的選択肢」に「自由記述」より差異が表面化した。この分析結果と読解力が十分に担保されていないと考えるならば、途上国における特徴としては、「記述する」ことよりもむしろ「問題文を読み解く」段階に困難性を有すると指摘できるだろう。ゆえに「問題文を読み解く」ことは、自由記述形式だけでなく他の出題項目にも関係するため、途上国における読解力の向上をもたらす数学学力全体への貢献は、先進国よりも大きいと考えられる。他方で国別においては、「教育的」や「科学的」で差異が顕在化する傾向にあるといえ、これらの領域では、先進国と同様に読解力の水準の違いを受けにくいともいえよう。

#### 4.4. 本章のまとめ

本章では、PISA2003 と PISA2006 および PISA2009 における数学的リテラシーと読解力調査の分析を通じて、生徒の数学学力と言語的側面、特に読解力との関連性について調べてきた。

その結果、3つの調査における38カ国の数学的リテラシーと読解力調査の平均値をみると、それらの水準の上位層と下位層、つまり先進国と途上国の学力水準には大きな差がみられ、大きな経年変化は確認されなかった(表4.4、表4.5および図4.2)。しかし経年変化の効果量を算出したところ、数学的リテラシーの方が読解力よりも経年変化を反映しにくい傾向を拮んだ(図4.3)。そして、国内における数学的リテラシーと読解力とテスト言語に代表させた言語的側面をみると、級内相関係数やモデル1とモデル2の国別分散の減少率およびモデル2とモデル3の生徒間分散の減少率(表4.7)、加えてランダム効果を含む回帰係数の比較(表4.8、図4.4と図4.5)から、3つのPISAの間に大きな経年変化はあまりみられないと捉えた。

経年変化は必ずしも大きくないという結果から、数学的リテラシー調査の出題項目が最も多いPISA2003に対象を絞ると、図4.1と図4.6で示したように、数学学力と読解力の水準とそれらの国内における関連性は、正の高い相関関係にある。つまり、数学学力の水準と読解力の水準およびそれらの関連性は、連動するという国際的な特徴を把握できる。

またPISA2003のデータを活用して、読解力の水準の違いによって困難度に注目した解答パターンがどのように異なるのかを調べた。その結果、前章では「自己効力感」の水準別は先進国と途上国に程度の違いはあるが、共通して「自由記述」で顕在化しやすいという結果を得た。ところが本章では、読解力の水準別に関して、先進国と途上国で異なる傾向が浮かび上がった。

先進国については、国内における読解力の水準別に困難度に注目した解答パターンが顕在化しやすい領域・種別は「自由記述」であること、読解力の水準別でなく国別でそれが現れやすい場合があり、それは「科学的」や「教育的」という領域に特にみられることの2点が浮かび上がった。PISAの数学的リテラシー調査のように文章問題で構成されるテストを用いて生徒の数学学力を測定する場合には、当然ながら一定程度の読解力が求められる。

ところが本分析では、「自由記述」において目立った特徴が確認された。もし問題文を読み解くことに困難性を有するならば、自由記述形式に特化して読解力の水準別の違いが現れるとは考えにくい。ゆえに文章問題の題意を読み解くことだけでなく、答えを導いた考え方や求め方あるいは理由を説明するといった「記述する」という場面においても読解力が大きく関係していると解釈できる。つまり「記述する」ことに対して、読解力を視野に入れながら考察する有効性を国際的な普遍性を確保して主張できよう。前章でも取り上げたが、日本においては、自由記述形式の出題項目に対する無答率の高さがしばしば問題視される（国立教育政策研究所、2010；室岡、2006）。もし高い無答率の改善を試みるならば、読解力の視座から検討する意義は小さくないだろう。つまり数学教育研究の周辺に読解力を位置付ける意義は否定できず、それは自由記述形式の項目、つまり「記述する」ことに関連した研究課題に成り得ることを示唆している。

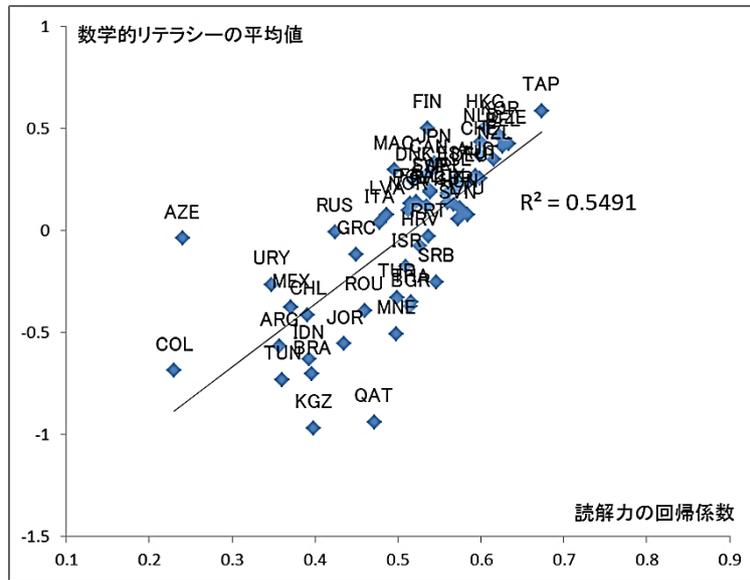
途上国に関しては、先進国とは異なり「自由記述」という種別で読解力の水準別の違いが顕在化しにくく、「複合的選択肢」の方にそれがいくらか現れやすかった。また国別においては、わずかながら先進国と同様に「教育的」や「科学的」に表面化するという傾向を掴んだ。途上国に関する結果を先進国と対比させながら考察すると、13 先進国は、数学学力と読解力の国際的な水準は高く、それらの国内における関連性は国際的にみて強かった。そこで、もし読解力が数学学力の向上に寄与するならば、高い読解力の水準の国ほど、読解力と数学学力の関連性は強いと判断できる。そのため先進国の高い数学学力の背景には、それを担保する読解力の存在を否定できない。これを翻って途上国に目を向けると、低い数学学力と称され得る途上国の背景には、数学学力を向上させるだけの読解力が身に付いていないとも解釈できる。また先進国では、文章問題で構成されるテストにおいて自由記述形式で読解力の水準別が目立って顕在化したため、「記述する」ことに読解力が関係するとした。ところが途上国では、単純な選択肢問題の困難度は大きくない中で、与えられた選択肢の中から選択する問いが連続する「複合的選択肢」により差異が表面化した。この分析結果と読解力が十分に担保されていない可能性を考慮すると、途上国における特徴として、「記述する」ことよりもむしろ「問題文を読み解く」段階に困難性を有することを指摘できるだろう。ゆえに「問題文を読み解く」ことは、自由記述形式だけでなく他の出題項目にも関係するため、途上国における読解力の向上がもたらす数学学力全体への貢献は、先進国よりも大きいと思われる。

本章では、数学学力と読解力の関連性を主に読解力調査のテスト得点を軸に考えてきた。ところが読解力調査もテストを使って測定されているため、読解力における解答パターンを国際比較することも可能である。例えば第3章で構築した分析枠組みを活用し、読解力における解答パターンを明らかにし、それと対比させながら数学的リテラシー調査における解答パターンを考察できよう。つまりテスト得点だけでなく、解答パターンという視点から数学学力と読解力の関連性を明らかにする取り組みが今後の大きな課題として設定できよう。しかしながら取り上げた13 先進国と7 途上国では、テスト得点に軸足を置く場合

できえ、大きく異なる様相を呈している。先進国では、「自由記述」に代表して「記述する」という立場からの検討が効果的と考える。他方で途上国では、そのような具体的な立場を取るよりは、むしろ全体的な特徴の把握を優先する視点を持つべきと思われる。例えば、解答パターンの類似性や特異性を明らかにした上で、それらを特に反映する公開項目などの内容を具体的に検討するなどが考えられる。

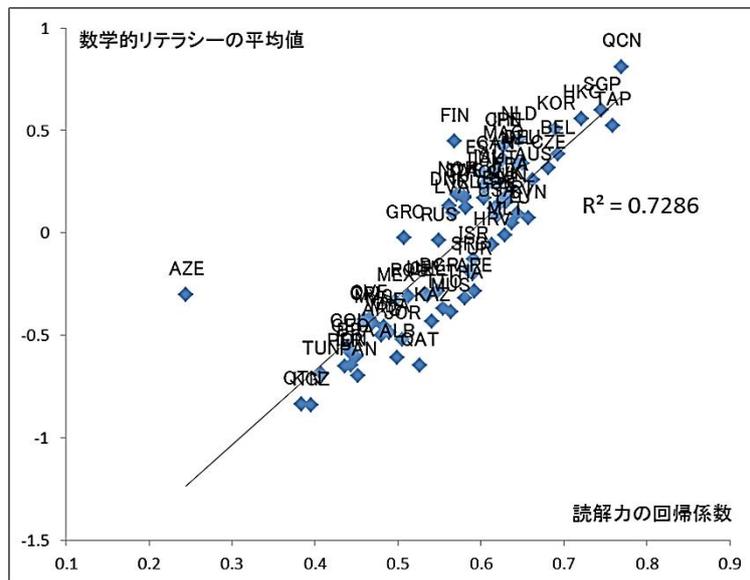
なお本章では、数学学力と読解力に焦点を置きながら言語的側面との関連性について、「自由記述」あるいは「記述する」ことをキーワードに先進国に関する特徴を明らかにしたこと、またそれと踏まえながら途上国固有と捉えるべき特徴を浮かび上がらせた。特に言語的側面から数学学力を捉える必要性は、わが国の途上国に関する数学教育研究でしばしば指摘されてきた (e.g., 馬場、2007 ; 中和、2012)。本研究は、読解力に焦点を当てたため限定的ではあるが、経年変化や数学学力との関連性について日本を含め先進国と対比させながら計量的に程度の違いを示した意義は大きいであろう。本章は、数学学力と言語的側面、特に読解力との関連性に関する研究の第一歩として位置付けたい。

付録 4.1 PISA2006 における数学的リテラシーの平均値と回帰係数の散布図



出典：筆者作成

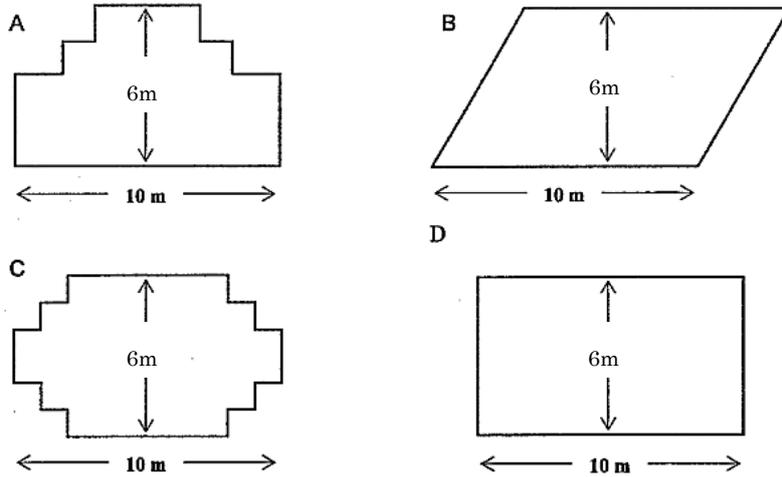
付録 4.2 PISA2009 における数学的リテラシーの平均値と回帰係数の散布図



出典：筆者作成

付録 4.3 複合的選択肢の項目例 (項目名「花壇」)

ある人が、長さが 32 m の木材を使って、花壇の外わくを作りたいと考えています。  
この人は次のようなデザインを考えています。



長さが 32 m の木材で、A~D それぞれのデザインの花壇を、作ることができますか。  
「できる」または「できない」のどちらかを○で囲んでください。

デザインの種類	32メートルの木材で、できるかできないか
デザインA	できる / できない
デザインB	できる / できない
デザインC	できる / できない
デザインD	できる / できない

出典：国立教育政策研究所 (2004、p.117)

## 第5章 本研究の総合的考察と今後の課題

本研究では、生徒の数学学力と①カリキュラムの内容領域の特徴、②生徒の数学に対する情意的側面および③言語的側面の3つの視点の関連性を実証的に明らかにすることを試みた。本章では、得られた研究の成果をまとめ、今後の課題について整理する。

### 5.1. 各章で得られた研究成果

本小節では、第2章から第4章で得られた研究成果の概要を記述する。各章は、PISAが公開するデータの二次分析を通じて得られた数学学力と上述の3つの視点の関連性に関する研究成果にそれぞれ対応する。

#### 5.1.1. 第2章の概要

第2章では、PISA2003の数学的リテラシー調査のデータを活用して、生徒の数学学力とカリキュラムの特徴の関連性を日本を事例として分析し、他の国を検討するための分析枠組みと考察の視点を与えた。

PISA2003に参加した39カ国のうち、日本をはじめとする先進国と途上国の間には、数学学力の水準に大きな差がみられることを確認した。続いて、日本と中心とした先進国に関する分析では、以下の3つを明らかにした。

1つ目は、テスト得点を算出することで、日本をはじめ韓国と香港は、数学を活用する能力とともに確率・統計を活用する能力も国際的にみて高い水準にあることを確認した。

2つ目は、13カ国の解答パターンを国際比較したことで、日本と韓国および香港は、困難度と識別力の2つの観点から類似した解答パターンを持つと同時に、平均的な解答パターンからみて特異であることが浮き彫りとなった。そして日本の特徴としては、「不確実性」と呼ばれる確率・統計に関する領域における困難度でより特異性が認められた。つまり日本の特異性は、4つの数学的な領域で同程度にみられるのではなく、特に「不確実性」に集中しているといえる。

3つ目は、日本の平均国より困難度が高い項目および識別力が低い項目を同定し、その内容を公開問題により確認することで、生徒が抱えている困難性の段階や測定の機能が低下した項目の特徴を調べた。加えてその要因を2003年当時のカリキュラムの特徴と絡めることで、出題内容に対する習得の度合いだけでなく、それへの認識の度合いという観点から論じた。以上を整理すると、日本の生徒が持つ数学、特に確率・統計を活用する能力は、国際的に高い水準にある一方で、その内実である解答パターンは国際的にみて特異であるという実態が明らかになった。そしてその背景には、日本のカリキュラムの特徴が一定程度反映されていると捉えた。特に、日本に焦点を当て平均国と比較することで、困難度差と識別力差は、生徒には広く認識されているが習得があいまいな内容を含む項目、あるいは生徒にほとんど認識されていない概念や考え方を含む項目を反映するという見解を得た。

具体的には、欧米諸国でみられるデータを中心に据えた統計教育で重視される内容として、データ収集やデータの分布に関する理解と活用が挙げられ、そのような内容をわが国においても取り扱う必要性が指摘されている。ところが、そのような項目の日本における識別力は低く測定に対する機能は、国際的にみて必ずしも高くないことを確かめた。つまり本章の分析結果は、習得の度合いだけでなく、数学教育で扱う統計的な学習内容の認識を高める必要性を生徒の解答パターンという視座から確認したといえる。

加えて、困難度差と識別力差を平面上で考えた場合に原点は、平均国の解答パターンを表わすため、それを利用して平均国との全体的な差異を各国で調べた。その結果、日本だけでなく他の国においても「不確実性」にその国固有の解答パターンが浮かび上がりやすいことが分かった。このことは、それぞれ国で確率・統計領域の数学教育が他の領域と比べて固有性を有して行われていると読み取ることも可能だろう。

続いて、日本の解答パターンを浮かび上がらせた方法、具体的には平均国との比較を通じた分析を7途上国に対しても適用し、それらの国の解答パターンを検討した。その結果、日本を含め取り上げた13先進国では、「不確実性」において固有の解答パターンが顕在化しやすいことが浮かび上がった。ところが数学的リテラシー調査のテスト得点の分布が下位層にある途上国においては、数学的な領域にあまり依存しない形で解答パターンの固有性が表面化するという全体的な傾向が浮かび上がった。そして「不確実性」に属する公開問題を手掛かりに、トルコとインドネシアを取り上げて、いくらか解答パターンを検討した。まずトルコとインドネシアでは、数学的リテラシー調査のテスト得点に一定の差異が認められる。ところが、テスト得点の平均値が高いトルコの方にインドネシアより高い難易度である項目を確認した。そこで日本の解答パターンとカリキュラムの特徴の関連性を指摘したように、その解釈の視点を参考にしながら、得られた結果を各国のカリキュラムと関連付ける取り組みを今後の大きな課題として位置付けたい。

### 5.1.2. 第3章の概要

第3章では、PISA2003の数学的リテラシーと質問紙調査の二次分析を通じて、生徒の数学学力と数学に対する情意的側面の関連性を調べた。

まずいくつかの欧米諸国では、情意的側面が特に数学学力に寄与する傾向にあるが、そのような状況の中でも「自己効力感」は、39ヵ国に例外なく数学学力と強い関連性を有することが明らかになった。そして自己効力感の水準の違いは、程度の大きさが同様であるとはいえないが、先進国と途上国とともに「自由記述」形式の出題項目で顕在化しやすいという特徴を明らかにした。これを踏まえて数学学力の向上を意識するならば、先進国と途上国を問わず「自己効力感の向上」に焦点を絞る情意的側面に対する取り組みの有効性が示唆される。なお「自己効力感の向上」には学習方略、特に自己調整学習方略が関係するとされる。PISAでは、生徒の学習方略についての調査を継続して実施しており、自己効力感の向上に繋がる学習方略に着目した分析が可能である。自己効力感と学習方略の関連

性を明らかにする分析は、今後の課題としたい。

本分析では、数学学力の水準が国際的にみて低い途上国に焦点を当てた検討も試みた。その成績が特に低い途上国についても、関連性と困難度に注目した解答パターンという新たな視座に立つことによって、先進国でみられる特徴が途上国でも確認されることを浮き彫りにした。文化・社会的な背景や生徒の学力水準など、多くの点で日本と異なる途上国のより深い理解に向けて、日本をはじめとする他の国との共通点を明らかにした本分析の意義は大きいと考える。

続いて、自己効力感の水準別に依らない副次的に得られた分析結果について触れておく。本章では、日本の生徒は数学の授業で典型的にみられるような疑似現実的な内容が含まれる「教育的」な文脈の問題を国際的にみて難無くこなす一方で、より抽象的とされる「科学的」な文脈では、他の国と比べて難易度が高いことを併せて浮き彫りにした。「科学的」については、技術的な過程、理論的な場面と明らかに数学的な問題についての理解とされ、それは数学教育の中でその育成が期待される中心的なものの一つであろう。ところが日本の生徒は、そのような文脈で数学を活用していく能力が同程度のテスト得点の水準にある国と比べて弱いことを指摘できる。なお、この日本と同様な傾向が香港においても観察された。

最後に、先進国と途上国の間にある相違点をより詳しく調べることも意義ある取り組みと考える。本章では、PISA で設定されている 5 つの情意的側面を利用して分析を進めた。このような下位尺度を用いる方法もあるが、情意的側面に関する 26 項目における因子構造を調べることもできる。例えば 26 項目の次元性を確認したところ、テスト得点の平均値が高い国ほど、26 項目に高い次元性が認められる傾向が浮かび上がった。つまり先進国では、自己効力感に集中する形で高い次元性を持つと思われる。言い換えると、次元性が低い傾向にある途上国に対して、下位尺度の一つである「自己効力感」のみに焦点を当てるのではなく、より情意的側面の水準別の差異が浮かび上がりやすい尺度を新たに構成して再検討するなどの工夫が必要かもしれない。先進国と途上国の差異をさらに浮かび上がらせるために、質問紙調査の結果をより詳細に調べることを今後の大きな課題として位置付けたい。

### 5.1.3. 第4章の概要

第4章では、PISA2003 と PISA2006 および PISA2009 における数学的リテラシーと読解力調査の分析を通じて、生徒の数学学力と言語的側面、特に読解力との関連性を経年変化も含めて検討した。

その結果、3 つの調査における 38 ヶ国の数学的リテラシーと読解力調査の平均値をみると、先進国と途上国の学力水準には依然大きな差がみられる。とはいえ経年変化の効果を算出したところ、数学的リテラシーの方が読解力よりも経年的な変化を反映しにくい傾向を捉えた。そして、読解力とテスト言語に代表させた言語的側面と数学学力の関連性を

階層線形モデルを用いて調べた結果、3つのPISAにおける級内相関係数や国別の読解力の変数を投入後の国別分散の変化率などから、大きな経年変化はみられないと判断した。つまり、数学学力と読解力の水準およびそれらの国内における関連性の2つの視点で大きな経年変化は確認されないと捉えた。

経年変化は必ずしも大きくないという結果から、数学的リテラシー調査の出題項目が最も多いPISA2003に対象を絞って分析を進めた。数学学力と読解力の水準と国内におけるそれらの関連性は、高い正の相関関係を持ち、連動するという国際的な特徴を明らかにした。そしてPISA2003のデータを活用して、読解力の水準の違いによって困難度に注目した解答パターンがどのように異なるのかを検討した。その結果、前章では「自己効力感」の水準別は先進国と途上国に共通する結果が得られたにも拘らず、本章では、読解力の水準別に関して、先進国と途上国で異なる傾向が浮かび上がった。

13 先進国については、国内における読解力の水準別に困難度に注目した解答パターンが顕在化しやすい領域・種別は「自由記述」であること、読解力の水準別でなく国別でそれが現れやすい場合があり、それは「科学的」や「教育的」という領域に特にみられることの2点が浮き彫りになった。PISAの数学的リテラシー調査のように文章問題で構成されるテストを用いて生徒の数学学力を測定する場合には、当然ながら一定程度の読解力が求められる。ところが本分析では、「自由記述」において突出した特徴を確認した。もし問題文を読み解くことに困難性を有するならば、自由記述形式に特化して読解力の水準別の違いが現れるとは考えにくい。ゆえに文章問題の題意を読み解くことだけでなく、答えを導いた考え方や求め方あるいは理由を説明するといった「記述する」という場面においても読解力が大きく関係すると捉えられる。言い換えると「記述する」ことに対して、読解力を視野に入れながら考察する有効性を先進国には仮定できる。つまり数学教育における読み書きという表現力の重要性を主張でき、「記述する」ことに関する研究課題が国際比較の観点から成り立つことを主張できる。

7 途上国に関しては、先進国とは異なり「自由記述」という種別で読解力の水準別の違いは顕在化しにくく、「複合的選択肢」の方にそれがいくらか現れやすかった。また国別においては、わずかながら先進国と同様に「教育的」や「科学的」に表面化するという傾向を掴んだ。途上国に関する結果を先進国と対比させながら考察すると、先進国は、数学学力と読解力の水準は高く、国内におけるそれらの関連性は国際的にみて強かった。そこで、もし読解力が数学学力の向上に寄与するならば、読解力の水準が高い国ほど、数学学力と読解力に強い関連性を有すると判断できる。そのため先進国の高い数学学力の背景に、それを担保する読解力の存在を肯定的に捉えるべきだろう。これを踏まえて低い数学学力と称され得る途上国をみると、数学学力を向上させるだけの読解力が十分身に付いていないと考える。また途上国では、与えられた選択肢の中から選択する問いが連続するという幾分か形式が複雑な「複合的選択肢」に「自由記述」より差異が表面化した。この分析結果と読解力が十分に担保されていない可能性を踏まえると、途上国における特徴として、「記

述する」ことよりもむしろ「問題文を読み解く」段階に困難性を有するといえる。「問題文を読み解く」ことは、自由記述形式だけでなく他の出題項目にも大きく関係するため、途上国における読解力の向上がもたらす数学学力全体への貢献は、先進国よりも大きいと捉えるべきだろう。

最後に本章では、数学学力と読解力の関連性を主に読解力調査のテスト得点を軸に考えてきた。ところが読解力調査もテストを使って測定されているため、読解力における解答パターンを国際比較することも可能である。例えば第2章で構築した分析枠組みを活用し、読解力における解答パターンを明らかにし、それと対比させながら数学的リテラシー調査における解答パターンを考察できよう。テスト得点だけでなく、解答パターンという視点から数学学力と読解力の関連性を明らかにする取り組みが今後の大きな課題として位置付くだろう。

## 5.2. 本研究の総合的考察と今後の課題

前小節では、第2章以降で得られた研究成果の概要を述べた。本研究は、PISAの数学的リテラシー調査のデータを用いて数学学力を規定し、PISAが実施した質問紙調査および読解力を分析対象とし、それらの関連性を論じた。最後に、本研究はPISAで収集されたデータを一貫して使用していることを活かし、得られた研究成果を総合的に検討した場合に得られる研究成果について述べ、今後の課題を整理する。

### 5.2.1. 本研究の総合的考察

各章で得られた研究成果の概要については、前小節で取り上げたが、それらを総合的に検討した場合に得られる成果について考える。

第3章と第4章では、数学的リテラシー調査のテスト得点を目的変数とし、それぞれ5つの情意的側面と読解力を説明変数に含む階層線形モデルによる分析を行った。この2つの章では、それぞれモデル1とモデル2において、国レベルにおける5つの情意的側面と読解力の変数を投入し、国別分散の減少率を計算した。その結果、第3章における減少率は71.43%であったのに対し、第4章ではPISA2003とPISA2006およびPISA2009でそれぞれ93.11%、87.14%、90.39%と約90%前後に至った。このことは国別の数学学力の差異に対して、情意的側面よりも読解力によって説明される部分が多いと解釈できる。とはいえ、情意的側面によって説明される率も小さくないため、情意的側面と読解力への働きかけが国別の数学学力の差異を縮小させる可能性は、決して否定できない。

続いて、それぞれの章で「自己効力感」と「読解力」の水準別にどのような解答パターンの差異が生じるのかを、先進国と途上国を区別して検討した。その結果、第3章と第4章で得られた分析結果に共通点を見い出せる。

まず、先進国に着目して共通点を確認する。それは、①国内における「自己効力感」あるいは「読解力」の水準別と、むしろ②国別でみられる場合の2つに分けられる。①に関

しては、「自由記述」形式の出題項目に特に顕在化するという共通点を指摘できる(表 3.12、表 4.11、図 3.9 および図 4.10 参照)。他方で②については、「科学的」と「教育的」における差異が際立って表面化するという点を主張できる(表 3.13、表 4.12、図 3.9 および図 4.10 参照)。なお日本と香港は、「教育的」における難易度は国際的にみて低く、逆に「科学的」では目立って高いという特徴を持つ。加えて、「科学的」と「教育的」における自己効力感と読解力の国内水準別の差は、各国で比較的小さいことも併せて明らかにした。

①に関する分析結果から、自由記述形式の出題項目について整理すると、この出題形式の難易度は、自己効力感や読解力の水準に左右されると理解できる。ゆえに各国国内において、自由記述形式の項目は、自己効力感や読解力の水準が低い生徒にとって必ずしも適切であるとは限らないといえよう。また本分析の結果は、自由記述形式における課題(例えば日本でいう無答率の高さ)に対して、自己効力感や読解力の視座から取り組む有効性を国際的な傾向の中で位置付けたと考える。

②に関する分析結果に目を移すと、「科学的」と「教育的」、特に「科学的」は、自己効力感と読解力の水準別に依存しないため(表 3.13、表 4.12、図 3.9 および図 4.10 参照)、数学教育の中で普遍的に育成される数学学力が隠れていると考える。この数学教育の固有性は、カリキュラムで規定される学習内容と切り離して考えることはできないだろう。「科学的」あるいは「教育的」に焦点を当て、それらとカリキュラムとの関連性を第2章で示した分析枠組みを活かしながら、より詳細に検討する必要があると考える。

続いて、途上国について検討する。途上国においては、国内の自己効力感と読解力の水準別、および国別の分析結果においても先進国のような際立つ共通点を見出すことは難しい(図 3.12 と図 4.13 参照)。ところが、先進国との対比の中でその特徴を浮かび上がらせることができる。第3章と第4章の要約で述べたように、自己効力感と読解力に関して次の2点を明らかにした。

1 つ目は、「自己効力感」は、先進国と途上国に例外なく数学学力と正方向に強い関連性を有し、その水準の違いは、程度の大きさが同様であるとはいえないが、途上国でも「自由記述」形式の出題項目で顕在化しやすいという特徴である。そのため、途上国においても「自己効力感の向上」に焦点を当てる情意的側面への取り組みが有効であることが示唆される。なお国別に関しては、「教育的」が先進国と共通して表面化しやすいことが明らかになった(表 3.17 参照)。2 つ目は、途上国では先進国と比べると、数学学力の向上を促す読解力が十分に身に付いておらず、それは「記述する」ことよりもむしろ「問題文を読み解く」という段階に困難性を有するという特徴である。そのため、「読解力の向上」が数学学力全体に貢献する可能性は、先進国よりも大きいと肯定的に捉えるべきと考える。ゆえに途上国に関する数学教育研究として、読解力の向上に繋がる考察が増々重視されたいと主張したい。

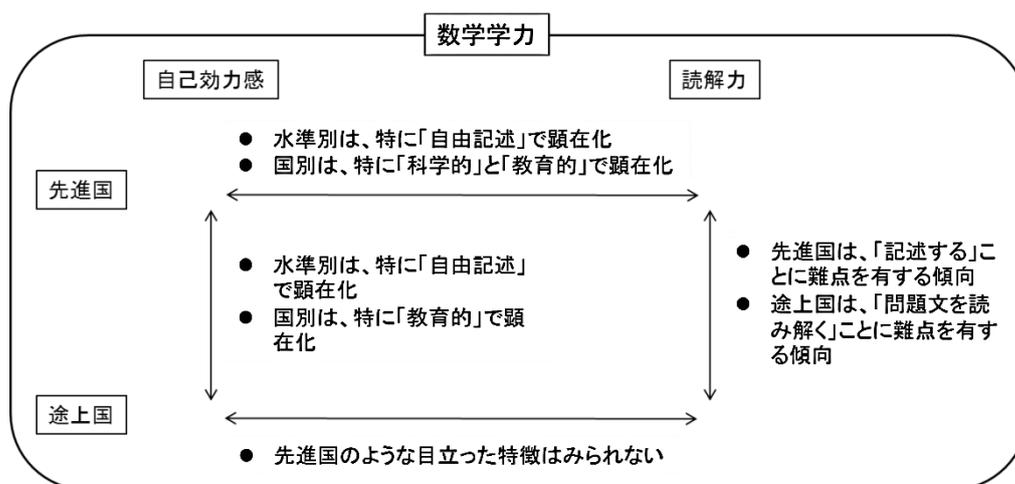
本研究では、先進国の特徴を浮かび上がらせ、それとの対比によって途上国の特徴を一定程度捉えた。先進国と同程度に詳細な情報とはいえないが、先進国と共通あるいは相違

する特徴を計量的に明らかにした本研究の意義は必ずしも小さくない。なお、主に第3章と第4章で得られた結果を総合的に考察して得られた研究成果の略図を図5.1に示した。

加えて、PISAのように国際比較可能な教育調査の公開データを活用する意義と注意点について触れておく。第1章で述べたが、本研究は、PISAが教育調査として高い信頼性と妥当性を持つことを前提としている。個々の出題項目の内容や翻訳などを特定の国の文脈と照らし合わせながら細かく検討すれば、公開されるデータに対する批判的見解は成り立つて然るべきである。ところがある特定の国において成り立つ見解に傾倒したのでは、参加国全体で共通の指標を構成して初めて浮かび上がる国際的な傾向を見逃す可能性は否定できない。本研究は、PISAが果たして有益な教育調査なのかを検証することを意図しない。PISAのデータを通じて各国の実態を浮かび上がらすことを目的とした。つまり、PISAが構築した国際比較可能な尺度や指標を活用し、各国の文脈に依存する特徴を差異の程度に着目して解き明かしたことに意義があるといえよう。

また本研究は、わが国を含む先進国だけでなく、PISAにおいて学力水準が最下位層にある途上国にも目を向けてきた。これまで大規模な教育調査に対して、「途上国の低い学力水準という特徴が、有意義な示唆を得にくくする」という指摘がなされてきた。ところが本研究では、その背景にあるテスト得点の分布や正答率を重視する方法を批判的に捉え、それを乗り越えることを視野に入れながら二次分析を試みた。その結果、項目反応理論の「等化」による困難度あるいは識別力の比較可能な数値への変換、および階層線形モデルによる国内における数学学力と情意的側面および読解力のランダム効果を含む回帰係数への着目を通じて、学力水準の違いという情報を一定程度取り除き、先進国と対比させながら途上国の特徴を考察した。生徒の社会・文化的な背景あるいは学力水準が日本を含めた先進国と大きく異なる途上国をより深く理解する上で、これまで限られた範囲でしか活用され

図5.1 本研究の総合的考察の略図



出典：筆者作成

て来なかった教育調査が有する情報を公開データ (raw data) から分析するという活用方法によって新たな成果を上げた本研究の意義は決して小さくはないと考える。

### 5.2.2. 今後の課題

すでに述べたが本研究は、冒頭で設定した生徒の数学学力と①カリキュラムの特徴、②生徒の数学に対する情意的側面および③言語的側面の3つの視点との関連性を実証的に明らかにすることを試みた。ところが、本研究をその終着点と捉えることは望ましくない。そこで第2章以降の各章で得られる今後の課題を整理し、本研究の更なる発展に向けた準備を行いたい。

第2章では、13先進国と7途上国の解答パターンを検討し、日本に注目してカリキュラムの特徴を数学的な内容の有無に焦点を当てて、解答パターンと関連付けた。そこで日本における解答パターンとカリキュラムの特徴の関連性に関する解釈の視点を参考にしながら、他の国の解答パターンとカリキュラムの特徴を関連付ける取り組みを今後の大きな課題として位置付けたい。先進国では、「不確実性」において各国固有の解答パターンが顕在化しやすい傾向を明らかにした。そのため、確率・統計に関する内容に焦点を合わせながら、数学学力とカリキュラムの関連性を調べるのが効果的と考える。他方で途上国では、数学的な領域にあまり依存しない形で解答パターンの固有性が表面化する傾向を確認した。そのため途上国をみる場合には、特異な解答パターンを示す公開項目に着目して正答を導くために必要な知識や技能を把握しながら、分析結果を慎重に読み解く必要があると考える。

第3章では、自己効力感が数学学力に正方向に強い関連性を持つことを掴んだ。自己効力感の向上には学習方略、特に自己調整学習方略が関係するとされる。PISAでは、生徒の学習方略についての調査を継続して実施しており、自己効力感の向上に繋がる学習方略に着目した分析が可能である。自己効力感と学習方略の関連性を明らかにする分析を今後の課題と位置付けたい。加えて数学学力の水準が高い国ほど、高い情意的側面の次元性を持つことを確認した。つまり先進国では、自己効力感に傾倒する形で次元性が満たされていると思われる。翻ると、次元性が低い傾向にある途上国に対して、下位尺度の一つである「自己効力感」のみに焦点を当てるのではなく、より情意的側面の水準別の差異が浮かび上がる新たな尺度が構成される可能性を否定できない。先進国と途上国にみられる異なる特徴を把握することを目的に、質問紙調査の結果をより詳細に調べる分析を別の今後の課題としたい。

第4章では、数学学力と読解力の関連性を主に読解力調査のテスト得点を軸に考えてきた。ところが読解力調査もテストを使って測定されているため、読解力における解答パターンを第2章で構築した分析枠組みを活用しながら調べるのが可能である。つまりテスト得点だけでなく、解答パターンという視点から数学学力と読解力の関連性を明らかにする取り組みが今後望まれるだろう。数学学力と読解力の関連性を解答パターンの視座から

検討する分析を今後の課題としたい。なおその際には、先進国では「自由記述」に特に注目すべきと考える。ところが途上国については、先進国のように際立った特徴が浮かび上がっていないため、途上国では解答パターンの全体像を掴む視点を持つことが効果的と考える。

第5章の本研究の総合的考察では、先進国において「自己効力感」と「読解力」の水準別に共通する特徴を明らかにした。ゆえに、この2つの関連性にも目を向けて数学学力を再検討するべきと考える。なぜならば、読解力が生徒の情意的側面と深く関わるならば、読解力が数学教育研究においてより焦点を当てるべきと主張できるためである。また途上国における自己効力感と読解力の関連性については、目立った特徴を捉えることに至らなかった。そのため自己効力感のみに注目するのではなく、新たな情意的側面の尺度を構成したり、途上国固有と思われる読解力の特徴を掴むといった分析を通じて、それらの関連性を検討したいと考える。

なお本研究の成果と今後の課題は、PISAで収集されたデータを活用した国際比較分析によって得られたことを述べておく。そのため、分析対象とするデータをPISAだけでなくTIMSSにも目を向けることは、多面的に各国の特徴を捉えるために、今後取り組むべき課題と考える。実際TIMSSでは、PISAと同じように項目反応理論で作成されたテストが作成され、重複テスト分冊法によるデータ収集が行われている。そのため、本研究で構築した分析枠組みを適用し、同様な分析方法による結果を得ることが可能である。またPISAやTIMSSは、表1.2で示したように定期的実施されているため、経年変化を掴む考察も行える。本研究では、PISA2003のデータを中心に国際比較分析を進めたが、PISA2012<sup>46</sup>においても「数学的リテラシー」が主要な調査内容となっている。ゆえに、本研究とPISA2012の二次分析によって得られる成果を照らし合わせることで経年変化を捉えられよう。TIMSSやPISA2012に同様な方法の分析を行い、異なる教育調査で得られる成果を経年変化を含めて総合的に考察することを今後の大きな課題としたい。

また本研究では、途上国として数学的リテラシー調査のテスト得点の最下位層の7カ国を取り上げた。ところがアフリカ諸国など、PISAあるいはTIMSSに参加していない途上国は多い。そのためSACMEQといった地域を限定した調査にも目を向け、途上国に関する分析対象を広げる努力も必要と考える。SACMEQで報告される数学と読解力に関する14カ国・地域(表1.3参照)の学力水準は、項目反応理論によって算出されている。ところが、重複テスト分冊法によるデータ収集は行われていない。なお出題形式は、正答を4つの選択肢から選ぶという選択肢形式が採られている。そのため正答を1誤答を0とするようなデータ処理を行えば、項目反応理論だけでなく、従来の古典テスト理論による分析も行える。つまり古典テスト理論という項目反応理論より一般に馴染みのある方法によっ

<sup>46</sup> 2014年1月現在、PISA2012のテクニカル・レポートは公開されておらず、データ収集後の分析で指摘される整合性の低い項目など、本研究で二次分析に向けて確認した基本的な情報を同様に得ることはできない。

て、解答パターンや数学学力と読解力の関連性などを調べる事が可能である。

最後に本研究のような二次分析は、分析対象とする調査の設計や内容に限定された範囲でしか考察を行えないという制約の下で成り立つ。本研究では、PISA が公開するデータを活用する立場から国際比較を通じて各国の実態を調べた。そのため本研究の成果は、PISA のデータから浮かび上がるものであり、それが本研究の目的に対して唯一無二のものでないことを確認しておく。今後の課題で指摘したように、より洗練された研究成果を得るには、多種多様な教育調査のデータや変数を用いながら、様々な視点から考察を行う必要があるだろう。とはいえ、本研究のような国際比較可能な指標やデータを活用する数学教育研究は、端緒についたばかりである。途上国に関する考察を含め、大規模教育調査の公開データが今後増々数学教育研究に活かされることを望む次第である。

## 参考・引用文献

- Association for Educational Assessment in Africa (AEAA). (2011). *AEAA Newsletter 5th Edition*, Examination Council of Zambia: Lusaka.
- Barton, P. N., & Barton, B. (2005). The Relationship Between English Language and Mathematics Learning for Non-native Speakers, *Teaching & Learning Research Initiative*, UNITEC: New Zealand.
- Barwell, R., Barton, B., & Mamokgethi. S. (2007). Multilingual issues in mathematics education: introduction, *Educational Studies in Mathematics*, Vol.64, pp.113-119.
- Coleman, J. S., (1966). *Equality of Education Opportunity*, Government Printing Office, Wanshington: US.
- Damerow, P., & Westbury. (1985). Conclusions drawn from the experience of the New Mathematics Movement. In P. Damerow, M. E. Dunkley, B. F. Nebres & B. Werry (Eds), *Mathematics for All*, Division of Science and Technical Environmental Education (Science and Technology Education Document Series No.20), UNESCO: Paris.
- Dimitris, R., (2006). ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses, *Journal of Statistics Software*, Vol.17, Issue 5, pp.1-25.
- D'Ambrosio, U. (2006). *Ethnomathematics: Link between Traditions and Modernity*, Sense Publishers: The Netherlands.
- Examination Council of Zambia (ECZ). (2000). *Learning Achievement at the Middle Basic Level 2000*, Ministry of Education, Lusaka: Zambia.
- Examination Council of Zambia (ECZ). (2003). *Learning Achievement at the Middle Basic Level 2003*, Ministry of Education, Lusaka: Zambia.
- Examination Council of Zambia (ECZ). (2006). *Learning Achievement at the Middle Basic Level 2006*, Ministry of Education, Lusaka: Zambia.
- Examination Council of Zambia (ECZ). (2008). *Learning Achievement at the Middle Basic Level 2008*, Ministry of Education, Lusaka: Zambia.
- Examination Council of Zambia (ECZ). (2010). *Learning Achievement at the Middle Basic Level 2010*, Ministry of Education, Lusaka: Zambia.
- Gace, K., & Jennifer, S. T. (2003). Racial and Ethnic Stratification in Educational Achievement and Attainment, *Annu. Rev. Sociol.*, Vol.29, pp.417-442.
- Gates, P., & Vistro-Yu, C. P. (2003). *Second International Handbook of Mathematics Education*. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick & F. K. S. Leung (Eds), *Is Mathematics for All?* (pp.31-73), Kluwer Academic Publishers: UK.
- IEA. (2005). *IEA's TIMSS2003 International Report on Achievement in the Mathematics Cognitive Domains*, IEA and Boston College.

- IEA. (2008a). *TIMSS2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, IEA and Boston College.
- IEA. (2008b). *TIMSS2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, IEA and Boston College.
- Jamal, A., & Carol, L. (2001). The Language Factor in Mathematics Tests, *Applied Measurement in Education*, Vol.14, No.3, pp.219-234.
- James, E. T., Robert, E. R., Barbara, J. R., Óscar, C., Jffrey, S., & Steven, J. O. (2008). The Impact of Middle-Grades Mathematics Curricula and the Classroom Learning Environment on Student Achievement, *Journal for Research in Mathematics Education*, Vol.39. No.3, pp.247-280.
- Jennifer, E. G., & Bryndl, H. M. (2007). Academic Performance of Young Children in Immigrant Families: The Significance of Race, Ethnicity, and National Origins, *International Migration Review*, Vol.41, No.2, pp.371-402.
- Kimberly, G., & Yu, X. (1999). Educational Expectations of Asian American Youths: Determinants and Ethnic Differences, *Sociology of Education*, Vol.72, pp.22-36.
- LLECE. (2009). *Reporte Técnico Segundo Estudio Regional Comparativo y Explicativo*, OREALC/UNESCO: Santiago.
- Leder, G. C. (2012). Mathematics for All? The Case for and against National Testing, *ICME12 Pre-proceedings*, pp.522-542.
- Lucille, C. (1997). *1997 Yearbook of the National Council of Teacher of Mathematics*. In T. Janet (Eds), *Mathematics for All Students: Access, Excellence, and Equity* (pp.1-9), NCTM: USA.
- Mamokgethi, S. (2005). Teaching Mathematics in a Primary Multilingual Classroom, *Journal for Research in Mathematics*, Vol.36, No.5, pp.447-466.
- Mamokgethi, S. (2012). Mathematics Education and Language Diversity: Background, Findings and Future Research Directions, *ICME12 Pre-proceedings*, p.41.
- Mamokgethi, S., & Jill, A. (2001). Between Languages and Discourses: Language Practices in Primary Multilingual Mathematics classroom in South Africa, *Educational Studies in Mathematics*, Vol.43, pp.243-269.
- Michael, R. H., Thomas. R. P., Maeda. Y., Arnold. L. C., Edwin. A., & Jeremy. A. K. (2007). Standards-Based Mathematics Curricula and Secondary Students' Performance on Standardized Achievement Tests, *Journal for Research in Mathematics Education*, Vol.38, No.1, pp.71-101.
- Márie, N. R., & John, O. (2009). The relationship between performance on

- mathematical word problems and language proficiency for students learning through the medium of Irish, *Educational Studies in Mathematics*, Vol.71, pp.43-64.
- National Council of Teacher of Mathematics (NTCM). (1989). *Curriculum and Evaluation Standards for School Mathematics*, Reston: Author.
- OECD. (2004). *Learning for Tomorrow's World: First Results from PISA2003*, OECD: Paris.
- OECD. (2005a). *PISA2003 Technical Report*, OECD: Paris.
- OECD. (2005b). *PISA2003 Data Analysis Manual*, OECD: Paris.
- OECD. (2006). *PISA2006 Technical Report*, OECD: Paris.
- OECD. (2007). *Learning for Tomorrow's World: First Results from PISA2006*, OECD: Paris.
- OECD. (2009). *PISA2009 Technical Report*, OECD: Paris.
- OECD. (2010a). *PISA Computer-Based Assessment of Student Skills in Science*, OECD: Paris.
- OECD. (2010b). *PISA2009 Results: What Students Know and Can Do –Student Performance in Reading, Mathematics and Science (Volume 1)*, OECD: Paris.
- PREAL. (2006). *2006 Cantidad sin Calidad: Un Informe del Progreso Educativo en América Latina*, Inter-American Dialogue: Washington.
- Steven, J. O., & Howard, T. E. (2009). *Differential Item Functioning 2nd ed. (Quantitative Applications in the Social Sciences: 161)*, SAGE Publications, Inc: USA.
- Thomas, R. P., Michael, R. H., Jon, D. D., Maeda, Y., Arnie, C., Edwin, A., Jeremy, A. K., & Ke, W. N. (2008). Standards-Based Mathematics Curricula and Middle-Grades Students' Performance on Standardized Achievement Tests, *Journal for Research in Mathematics Education*, Vol.39, No.2, pp.184-212.
- UNESCO. (2000). *Assessing Learning Achievement*, UNESCO: Paris.
- UNESCO. (2008). *Overcoming Inequality: Why Governance Matters*, UNESCO: Paris
- UNESCO. (2009). *Learning Counts: An Overview of Approaches to Understanding Assessing and Improving the Quality of Education for All*, UNESCO: Bangkok.
- UNESCO. (2012). *Asia-Pacific Education System Review Series: Student Learning Assessment*, UNESCO: Bangkok.
- Wiggan, G. (2007). Race, School Achievement, and Educational Inequality: Toward a Student Based Inquiry Perspective, *Review of Educational Research September*, Vol.77, pp.310-333.
- William, F. T. (1997). Race-Ethnicity, SES, Gender, and Language Proficiency Trends in

- Mathematics Achievement: An Update, *Journal for Research in Mathematics Education*, Vol.28, No.6, pp.652-679.
- Xin, M., & Jesse, L. M. W. (2007). Mathematics Coursework Regulates Growth in Mathematics Achievement, *Journal for Research in Mathematics Education*, Vol.38, No.3, pp.230-257.
- モーモーニエン. (2006). 「数学教育における Equity の視点に関する研究」, 『日本教科教育学会誌』, 第 29 号 第 2 号, pp.39-48.
- 伊藤崇達・神藤隆昭. (2003). 「自己効力感, 不安, 自己調整学習方略, 学習の持続性に関する因果モデルの検証—認知的側面と動機づけの側面の自己調整学習方略に着目して—」, 『日本教育工学会論文誌』, 第 27 卷 第 4 号, pp.377-385.
- 外務省. (2008). 『2007 年版政府開発援助 (ODA) 白書—日本の国際協力—』, 外務省.
- 荻谷剛彦・志水宏吉. (2004). 『学力の社会学—調査が示す学力の変化と学習の課題』, 岩波書店.
- 岩崎秀樹・長崎榮三・モーモーニエン. (2004). 「後期中等段階の数学カリキュラムの国際比較」, 『日本数学教育学会誌』, 第 86 卷 第 9 号, pp.2-12.
- 吉野諒三・林文・山岡和枝. (2010). 『国際比較データの解析—意識調査の実践と活用—』, 朝倉書店.
- 宮崎康夫. (2007). 「階層線形モデル (HLM) の教育研究への応用と分析結果の教育政策への利用の観点」, 『日本テスト学会誌』, 第 3 卷 第 1 号, pp.123-146.
- 近藤博之. (2012). 「社会空間と学力の階層差」, 『教育社会学研究』, 第 90 集, pp.101-121.
- 金子真理子. (2004). 「学力の規定要因—家庭背景と個人の努力は、どう影響するか—」, 荻谷剛彦・志水宏吉編 『学力の社会学』, 岩波書店, pp.153-172.
- 金本良通. (2010). 「表記・表現」, 日本数学教育学会編 『数学教育学研究ハンドブック』, 東洋館出版社, pp.245-252.
- 古田和久. (2012). 「高校生の学校適応と社会文化的背景—学校の階層多様性に着目して—」, 『教育社会学研究』, 第 90 集, pp.123-144.
- 国立教育政策研究所. (2004). 『生きるための知識と技能 2—OECD 生徒の学習到達度調査 PISA2003 年調査国際結果報告書』, ぎょうせい.
- 国立教育政策研究所. (2005). 『TIMSS2003 理科教育の国際比較—国際数学・理科教育動向調査の 2003 年調査報告書』, ぎょうせい.
- 国立教育政策研究所. (2007). 『生きるための知識と技能 3—OECD 生徒の学習到達度調査 PISA2006 年調査国際結果報告書』, ぎょうせい.
- 国立教育政策研究所. (2010). 『生きるための知識と技能 4—OECD 生徒の学習到達度調査 PISA2009 年調査国際結果報告書』, 朝倉書店.
- 国立教育政策研究所. (2012a). 『全国学力・学習状況調査の 4 年間の調査結果から今後の取組が期待される内容のまとめ—児童生徒への学習指導の改善・充実に向けて— (小

- 学校編)』, 教育出版社.
- 国立教育政策研究所. (2012b). 『全国学力・学習状況調査の4年間の調査結果から今後の取組が期待される内容のまとめ—児童生徒への学習指導の改善・充実に向けて—(中学校編)』, 教育出版社.
- 黒田一雄. (2005). 「国際教育協力の潮流」, 黒田一雄・横関祐見子編『国際教育開発論—理論と実践—』, 有斐閣, pp.1-13.
- 今井敏博. (2010). 「情意」, 日本数学教育学会編『数学教育学研究ハンドブック』, 東洋館出版社, pp.318-325.
- 根本博. (2010). 「カリキュラム論」, 日本数学教育学会編『数学教育学研究ハンドブック』, 東洋館出版社, pp.45-62.
- 山崎博敏・藤井宜彰・水野考. (2009). 「学級規模と指導方法が小学生の学力に及ぼす影響: 共分散構造分析とマルチレベルモデル分析の適用」, 『広島大学大学院教育学研究科紀要』, 第三部 第58号, pp.9-16.
- 市川伸一. (2001). 「学力低下論争の構図と『もう一つの学力低下論』」, 中井浩一編『論争・学力崩壊』, 中公新書ラクレ, pp.209-231.
- 室岡和彦. (2006). 「中学・高校数学の改善に向けて—調査結果等から学ぶもの—」, 『日本数学教育学会誌』, 第88巻 第7号, pp.16-22.
- 柴山直. (2008). 「日本のテスト文化について」, 『人事試験研究』, 第208巻, pp.2-13.
- 柴山直. (2012). 『全国規模の学力調査における重複テスト分冊法の展開可能性について』, 平成23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.
- 芝祐順. (1991). 『項目反応理論 基礎と応用』, 東京大学出版会.
- 松沼光泰. (2004). 「テスト不安, 自己効力感, 自己調整学習及びテストパフォーマンスの関連性—小学校4年生と算数のテストを対象として—」, 『教育心理学研究』, 第52巻 第4号, pp.426-436.
- 植野真臣・荘島宏二郎. (2010). 『学習評価の新潮流』, 朝倉書店.
- 深澤弘美・竹内光税・二宮智子. (2007). 「アジア・オセアニア諸国における初等中等統計教育カリキュラムの比較研究」, 『日本統計学会誌』, 第36巻 第7号, pp.279-308.
- 須藤康介. (2007). 「授業方法が学力と学力の階層差に与える影響—新学力観と旧学力観の二項対立を超えて—」, 『教育社会学研究』, 第81集, pp.25-44.
- 須藤康介. (2010). 「学習方略がPISA型学力に与える影響—階層による方略の違いに着目して—」, 『教育社会学研究』, 第86集, pp.139-157.
- 瀬沼花子. (2006). 「学力に関する国際調査PISA, TIMSSにみる日本の小中高生の現状と課題—統計的な内容を中心に—」, 『日本行動計量学会第34回大会発表論文妙録集』, pp.14-15.
- 瀬沼花子. (2008). 「PISAにおける数学的リテラシーについて」, 『科学教育研究』, 第32

- 巻 第 4 号, pp.358-365.
- 清水美憲. (2006). 「OECD・PISA の「数学的リテラシー」論からみた日本の算数・数学教育」, 『日本数学教育学会誌』, 第 88 巻 第 3 号, pp.44-53.
- 清水美憲. (2007). 「OECD/PISA における数学的リテラシー評価問題の特徴」, 教育テスト研究センター第 6 回研究会報告書.
- 清水美憲. (2012). 「評価問題作成における数学的なプロセスへの焦点化—全国学力・学習状況調査 (中学校数学) の動向と課題—」, 『日本数学教育学会誌』, 第 94 巻 第 9 号, pp.30-33.
- 石井秀宗・安永和央. (2011). 「全項目が開示されるテスト文化のもとでの得点分布の経年比較—全国テストと自治体テストのリンキング—」, 『日本テスト学会誌』, 第 7 巻 第 1 号, pp.24-35.
- 川口俊明. (2009). 「マルチレベルモデルを用いた「学校の効果」の分析—「効果的な学校」に社会的不平等の救済はできるのか—」, 『教育社会学研究』, 第 84 集, pp.165-184.
- 村木英治. (2005). 「全米学力調査 (NAEP) 概説—テストデザインと統計手法について—」, 東京大学大学院教育学研究科教育測定・カリキュラム開発 (ベネッセコーポレーション) 講座公開研究会資料.
- 村木英治. (2011). 『項目反応理論』, 朝倉書店.
- 大友賢二. (1996). 『項目反応理論 言語テスト・データの新しい分析法』, 大修館書店.
- 中和渚. (2012). 「本質的学習環境 (SLE) に基づく数学科授業開発研究 (4) —ザンビア共和国で実施した授業開発の視座と教材の再構成の視点—」, 『数学教育学研究』, 第 18 巻 第 1 号, pp.15-22.
- 長崎栄三・萩原康仁. (2004). 「算数到達度の項目反応理論による比較」, 『数学教育論文発表会論文集』, 第 37 号, pp.103-108.
- 田崎勝也. (2008). 『社会科学のための文化比較の方法—等化性と DIF 分析—』, ナカニシヤ出版.
- 渡辺美智子. (2007). 「知識創造社会を支える統計的思考力の育成—アクションに繋がる統計教育への転換—」, 『日本数学教育学会誌』, 第 89 巻 第 7 号, pp.29-38.
- 内田豊海. (2009). 「ザンビア基礎教育における計算能力の診断的評価に関する研究—弁別性と教授的示唆に注目して—」, 『国際教育協力論集』, 第 12 巻 第 2 号, pp.1-12.
- 内田豊海. (2011). 「ザンビア算数科における文章題の診断的評価法開発研究—ニューマン法の限界を乗り越えて—」, 『数学教育学研究』, 第 17 巻 第 2 号, pp.95-102.
- 日本テスト学会. (2010). 『見直そう、テストを支える基本の技術と教育』, 金子書房.
- 馬場卓也. (1998). 「民族数学を基盤とする数学教育の展開 (2) —批判的数学教育と民族数学の接続—」, 『数学教育学研究』, 第 4 巻, pp.29-35.
- 馬場卓也. (1999). 「民族数学を基盤とする数学教育の展開 (3) —数学教育における基礎的活動の動詞による分析—」, 『数学教育学研究』, 第 5 巻, pp.17-25.

- 馬場卓也. (2001). 「民族数学を基盤とする数学教育の展開 (4) —ケニア国初等教育における学習指導要領の動詞による分析—」, 『数学教育学研究』, 第 7 巻, pp.7-17.
- 馬場卓也. (2002). 「民族数学を基盤とする数学教育の展開 (5) —動詞型カリキュラムにおける測定活動の記号論的分析—」, 『数学教育学研究』, 第 8 巻, pp.11-18.
- 馬場卓也. (2005). 「開発途上国における民族数学を基盤としたカリキュラム構成原理の研究」, 『数学教育学論究』, 第 83 巻, pp.17-24.
- 馬場卓也. (2008). 「教育開発研究における教室からの視点—理数科教育の位置付けと課題—」, 澤村信英編『教育開発国際協力研究の展開—EFA (万人のための教育) 達成に向けた実践と課題—』, 明石書店, pp.67-88.
- 馬場卓也・内田豊海. (2008). 「国際比較調査の開発途上国の教育開発に対するインプリケーションに関する考察—ガーナ国 TIMSS レポートを事例として—」, 『国際教育協力論集』, 第 11 号 第 2 号, pp.129-140.
- 服部環. (2011). 『心理・教育のための R によるデータ解析』, 福村出版.
- 豊田秀樹. (1998). 『共分散構造分析 [入門編]—構造方程式モデリング—』, 朝倉書店.
- 豊田秀樹. (2002a). 『項目反応理論 [入門編]—テストと測定の科学—』, 朝倉書店.
- 豊田秀樹. (2002b). 『項目反応理論 [事例編]—新しい心理テストの構成法—』, 朝倉書店.
- 豊田秀樹. (2012b). 『項目反応理論 [入門編] (第 2 版)』, 朝倉書店.
- 木村拓也. (2010). 「日本における「テストの専門家」を巡る人材養成状況の量的把握」, 『日本テスト学会誌』, 第 6 巻 第 1 号, pp.30-49.
- 柳井晴夫・石井秀宗. (2008). 「大規模学力テストと学ぶ力に関する研究をめぐって」, 日本児童研究所編『児童心理学の進歩 Vol.47』, 金子書房.
- 鈴川由美・豊田秀樹・川端一光. (2008). 「わが国の数学教育は数学を日常の中で活用する能力を重視しているか—PISA2003 年調査の DIF による分析—」, 『教育心理学研究』, 第 56 巻 第 2 号, pp.206-217.
- 鈴木康志. (2006). 「PISA2003 年調査・TIMSS2003 年調査の分析」, 『日本数学教育学会誌』, 第 88 巻 第 1 号, pp.23-31.
- 澁谷渚. (2008). 「本質的学習環境 (SLE) に基づく数学科授業開発研究 (1) —ザンビア基礎学校における生徒の活動の分析—」, 『数学教育学研究』, 第 14 巻, pp.187-197.
- 澁谷渚. (2010). 「ザンビアにおける本質的学習環境 (SLE) に基づく数学科授業開発研究 (3) —第 5 学年児童が行った『数の石垣』の学習過程への着目—」, 『数学教育研究』, 第 16 巻 第 2 号, pp.71-79.
- 齋藤昇. (2010). 「評価」, 日本数学教育学会編『数学教育学研究ハンドブック』, 東洋館出版社, pp.396-379.
- 齋藤昇・藤田彰子. (1997). 「数学学習における記述表現力と口述表現力の関係—中学数学 2 年「一次関数」の調査を通じて—」, 『数学教育学研究』, 第 4 巻, pp.197-207.
- 青山和裕. (2007). 「日本の統計教育改善の方向性についての検討」, 『日本統計学会誌』,

第 36 卷 第 2 号, pp.263-277.

深澤弘美. (2007). 「初等・中等統計教育カリキュラムの国際比較研究—ニュージーランドにおける統計教育カリキュラム—」, 『日本数学教育学会誌』, 第 89 卷 第 7 号, pp.39-48.

参考・引用ホームページ

PISA <http://www.oecd.org/pisa/>

TIMSS <http://timss.bc.edu/>

NAEP <http://nces.ed.gov/nationsreportcard/>

文部科学省 <http://www.mext.go.jp/>

SACMEQ <http://www.sacmeq.org/>

PASEC <http://www.confemen.org/>