

3 テストの設計と開発

3.1 ブロックの配置

数学と国語のいずれにおいても1つのブロックを4つの小問から構成することを基本とした。したがって数学においてはブロック数が16、項目数が64、国語においては基本的にブロック数が12、項目数が48となるが、後で述べる理由により、実際にはそれぞれ13、49となった。項目の詳細配置表については、数学においては資料1.1.3を、国語においては資料2.1.3を参照されたい。ブロック内については、数学では独立した小問から構成、国語においては1つのテキストに関する4つの小問からなるいわゆる大問形式を採用した。

表3.1.1aでは数学の分冊S1～S8の計8冊の分冊に対して、計16個のブロックをB1～B16まで割り当てた様子を示す。分冊S1においてはB3,B5,B7,B9,B11,B12,B15,B16の8ブロック32個の項目（小問）が含まれ、分冊S2においてはB1,B4,B6,B8,B10,B13,B13,B16の同じく8ブロック32個の項目が含まれていることがわかる。また、分冊S1と分冊S2にはB12とB16の2つのブロック、計8項目が共通に含まれていることになる。同様に、分冊S1と分冊S6とではB3,B5,B11,B12,B16の5ブロック20項目が共通に含まれている。逆にブロック単位で見ると、いずれのブロックも全部で8冊ある分冊のうち必ず4冊には含まれるようにデザインされている。

さらに表3.1.1bではそれらのブロックがその冊子のどの位置に配置されているかを示した。情報量としてはこの表から上の表3.1.1aが復元できる関係にあるが、わかりやすさを優先して両方を掲載する。表3.1.1bの表頭にある数値がその分冊において、あるブロックが前から何番目に置かれるのかがしめされている。たとえば分冊1においてブロック11は3番目、ブロック7は5番目の位置に置かれている。

表 3.1.1a 数学の分冊と各ブロックとの対応関係

		ブロック															
		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B16
分冊	S1			○		○		○		○		○	○			○	○
	S2	○			○		○		○		○		○	○			○
	S3	○	○	○		○		○		○				○	○		
	S4		○		○		○		○		○				○	○	
	S5				○		○	○		○			○	○			○
	S6	○		○		○			○		○		○	○			○
	S7	○	○	○		○		○		○				○	○		
	S8		○		○		○		○		○		○			○	○

表 3.1.1b 数学の分冊におけるブロック配置

分冊	位置1	位置2	位置3	位置4	位置5	位置6	位置7	位置8
S1	B3	B5	B11	B12	B7	B9	B15	B16
S2	B4	B6	B12	B13	B1	B8	B10	B16
S3	B5	B7	B13	B14	B1	B2	B3	B9
S4	B6	B8	B14	B15	B2	B4	B10	B11
S5	B7	B9	B15	B16	B4	B6	B12	B13
S6	B1	B8	B10	B16	B3	B5	B11	B12
S7	B1	B2	B3	B9	B5	B7	B13	B14
S8	B2	B4	B10	B11	B6	B8	B14	B15

次に表 3.1.2a には国語の各分冊と各ブロックの対応関係がしめされている。分冊番号 K4 と K7 において K4B と K7B が存在しているが、これはブロック 12 において 4 番目の小問を異なるものにしたためである。小問を後者に差し替えたブロックが入っている分冊の番号に記号「B」が付されている。大問形式における小問入れ替えの影響をこのことによって探る。表 3.1.2b も含めて見方は数学の場合と基本的には同じである。ブロック数が数学に比べて国語の方が少ない分、分冊相互に共通するブロックの数がいずれの組み合わせにおいても 2 つとなるようにしてバランスをとるようにしたが完全ではない。また表 3.1.2c にはブロックごとの問題タイトルを掲載した。後の章でこのタイトルを使った考察がなされる。

表 3.1.2a 国語の分冊と各ブロックの対応関係

		ブロック												
		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B12B
分冊	K1	○							○	○				
	K2		○					○			○			
	K3			○			○					○		
	K4				○	○							○	
	K4B													○
	K5	○				○					○			
	K6		○				○				○			
	K7			○				○						○
	K7B													○
K8				○					○			○		

表 3.1.2b 国語の分冊におけるブロック配置

分冊	位置1	位置2	位置3
K1	B1	B8	B9
K2	B2	B7	B10
K3	B3	B6	B11
K4	B4	B5	B12
K4B	B4	B5	B12B
K5	B1	B5	B10
K6	B2	B6	B9
K7	B3	B7	B12
K7B	B3	B7	B12B
K8	B4	B8	B11

表 3.1.2c 国語のブロックと問題タイトル

ブロック番号	問題タイトル
B1	カレー
B2	水餃子
B3	枕草子
B4	にくきもの
B5	原稿用紙
B6	食政策
B7	インターネット
B8	ナイチンゲール
B9	働かないアリ
B10	よろしかったでしょうか
B11	おはようどこんばんは
B12	土用の丑の日
B12B	土用の丑の日B

なお、PISA 調査等で採用されているアイテム・マトリックス・サンプリング法の場合、ブロックを分冊に配置する際に、さらに厳密なデザインである BIB (Balanced Incomplete Block: 釣合い型不完全ブロック) デザインを利用しているが、おもに実施上の理由で、数学・国語とも本調査研究ではそこまでの工夫は実現できなかった。

3.2 数学

本研究で使用した問題と数学における学習指導要領（平成10年）との関係は資料 1.1.1 から 1.1.2 に掲載したとおりである。また昨年度実施した問題との関係及び各問の変更点は資料 1.1.4 に掲げたとおりである。すなわち昨年度研究での数学の実施問題は 56 項目であったが、表 3.2.1 に示すとおり、そのうち 12 項目をそれぞれの理由で削除し、その代わり、あらたに作問した 20 項目を追加し、本年度実施の 64 項目とした。追加した 20 項目のうち 2 項目は平成 24 年度実施の学指導要領対応のものである。なお、項目は非公開である。これは国際的な学力調査において項目を公開すると難易度等の項目特性が変化してしまい、時系列的な学力の変化を捉えられなくなるなどの弊害があるためである。そのかわり、項目の特性を表す統計的な指標の値を公開することで、第三者による質的なチェックが行うことが可能となる。本調査研究でもその方式を採用した。

削除された 12 項目はおおむね困難度の低いものである一方、追加された 20 個の項目はある程度の難しさを持つように作成したため、結果として、昨年度のテストよりも難しいテスト構成となっていることに注意が必要である。これはいわば学力の時系列変化を追跡する状況を模していることにもなっている。すなわち、長期にわたって学力の時系列変化を捉えるためには、その間に必ず実施されるであろう学習指導要領の改訂にともなう項目の差し替え、あるいは項目内容等の時代にもなう表現などの小規模な手直しなどが必要となるが、それを擬似的に試みていることになる。今回の場合は相対的に易しい項目が減ったため、従来型のテスト得点で比較しようとする、仮に生徒集団に学力の変化がなくても平均点として下がってしまう状況であるが、第 6 章で示すとおり等化(equating)を行うことによってこの問題を解消する。

表 3.2.1 昨年度実施項目のうち削除された項目の特性と削除理由

2010年度 項目番号	識別力	困難度	削除理由	2011年 項目番号
B0-1	1.294	-1.634	昨年正答率が予想より高かったため。	
B0-2	1.065	-1.813	小学校の範囲のため。	
B0-7	1.611	-1.510	記号を使うこと以外は小学校の範囲のため。	
B0-12	1.499	-1.269	他の設問とかぶるため。	
B0-14	1.144	-1.58	想定正答率より高かったため。	
B1-6	0.879	-1.225	既題があったため。	
B2-2	1.614	-1.164	記号を使うこと以外は小学校の範囲のため。	
B2-3	0.641	-0.753	既題があったため。	
B3-3	1.378	-1.597	想定正答率より高かったため。	
B3-6	0.462	-0.645	識別力が悪かったため。	
B4-7	1.288	-1.247	既題があったため。	
B4-8	0.327	1.795	識別力が低かったが分野(単元)を重視しているため残した。	
B5-5	1.156	-1.343	新問を入れる段階で、調整問題となった。	

3.3 国語

3.3.1 本研究におけるリーディング・リテラシーの諸側面の定義

本調査研究におけるリーディング・リテラシーは、PISAにおけるリーディング・リテラシーの定義を参考にしつつ、国語科の学習内容としての学力を測定するために、以下のような4つの側面から独自の定義をおこなった(表3.3.1)。学習指導要領やPISAにおける定義との概念的な関係は第2章3節に述べたとおりである。

表 3.3.1 本研究におけるリーディング・リテラシーの諸側面の定義

<p>1 情報の取り出し</p> <p>文書を読んで、その文書が述べている事柄を正しく取り出すことができることは、国語の力として基本的なものである。述べている事を正しく取り出すだけでなく、曲解しないためにも、書かれていないことを含意していると思わないことや、異なる情報との区別を付けるなど、書かれていることを過不足無く受け取れるかどうかをみる。</p>
<p>2 解釈</p> <p>文書に書かれている情報をよく理解できているかは、文章中にまとまって書かれている情報をそのまま抜き取るだけでなく、例えば2つの事柄がかかかれていればその共通点と差を捉えたり、登場人物の心情を理解したり、著者の主張を捉えたりなど、分散して書かれているものをまとめたり、特徴を捉えたりすることができるかどうかをみる。</p>
<p>3 情報の編集・統合</p> <p>情報を的確に伝えるためには情報を過不足無くまとめコンパクトにしたり、相手にあわせたり、課題にあわせて編集する力が必要になる。求められているものと、現状の捉えた情報からその差を埋めたり、必要な情報のみを切り出したりできるかどうかをみる。</p>
<p>4 判断</p> <p>文書から得た情報を、文書に挙げられていた事例とは異なるものに適用できることは、理解を示す上で大切なことである。別の事例の場合には、その事例を文書の事例をもとに分析し、何をどのように適用するかを示す必要がある。そのためには、得られた情報を再構成することも必要になる。このような力をみる。</p>

3.3.2 学習指導要領との関係

本調査研究におけるリーディング・リテラシーの諸側面と新中学校学習指導要領(国語)との対応は表3.3.2aから表3.3.2cに示すとおり関係を想定している。たとえばブロック1を例にとれば、本研究におけるリーディング・リテラシーに対応する学習指導要領の領域は「C読むこと」になっているが、領域「B書くこと」に関しては当然のことながらリーディング・リテラシーの定義には含めず、学習指導要領のみに即した構成をとっていることがわかる。

表 3.3.2a 学習指導要領と本研究における国語の定義との関係（ブロック1～4）

		学習指導要領				本研究におけるリーディング・グリテシーの定義				
ブロック番号	項目番号	変数名	学年	領域	目標	内容	情報の取り出し	解釈	情報の編集・統合	判断
1	1	B1-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	場面の展開や登場人物などの描写に注意して読み、内容の理解に役立てること。		○		
	2	B1-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	場面の展開や登場人物などの描写に注意して読み、内容の理解に役立てること。		○		
	3	B1-3	第2学年	読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章全体と部分との関係、例示や描写の効果、登場人物の言動の意味などを考え、内容の理解に役立てること。			○	
	4	B1-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考えた確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考える。	集めた材料を分類するなどして整理するとともに、段落の役割を考えて文章を構成すること。				
	第2学年		読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章の構成や展開、表現の仕方について、根拠を明確にして自分の考えをまとめること。				○	
2	1	B2-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	場面の展開や登場人物などの描写に注意して読み、内容の理解に役立てること。	○			
	2	B2-2	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考えた確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考える。	伝えたい事実や事柄について、自分の考えや気持ちを根拠を明確にして書くこと。				
			第2学年	読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章全体と部分との関係、例示や描写の効果、登場人物の言動の意味などを考え、内容の理解に役立てること。			○	
	3	B2-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	文章の構成や展開、表現の特徴について、自分の考えをもつこと。				○
4	B2-4	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	場面の展開や登場人物などの描写に注意して読み、内容の理解に役立てること。				○	
3	1	B3-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	本や文章などから必要な情報を集めるための方法を身に付け、目的に応じて必要な情報を読み取ること。		○		
	2	B3-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	場面の展開や登場人物などの描写に注意して読み、内容の理解に役立てること。		○		
	3	B3-3	第2学年	読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章全体と部分との関係、例示や描写の効果、登場人物の言動の意味などを考え、内容の理解に役立てること。				○
	4	B3-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考えた確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考える。	伝えたい事実や事柄について、自分の考えや気持ちを根拠を明確にして書くこと。				
	第2学年		読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章の構成や展開、表現の仕方について、根拠を明確にして自分の考えをまとめること。				○	
4	1	B4-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	文章に表れているもの見方や考え方をとらえ、自分のもの見方や考え方を広げること。		○		
	2	B4-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	文章の構成や展開、表現の特徴について、自分の考えをもつこと。				○
	3	B4-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	場面の展開や登場人物などの描写に注意して読み、内容の理解に役立てること。		○		
	4	B4-4	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してもの見方や考え方を広げようとする態度を育てる。	文章の構成や展開、表現の特徴について、自分の考えをもつこと。				
	第2学年		書くこと	目的や意図に応じ、社会生活にかかわることなどについて、構成を工夫して分かりやすく書く能力を身に付けさせるとともに、文章を書いて考えを広げようとする態度を育てる。	自分の立場及び伝えたい事実や事柄を明確にして、文章の構成を工夫すること。					

表 3.3.2b 学習指導要領と本研究における国語の定義との関係（ブロック 5～8）

		学習指導要領				本研究におけるリーディング・グリテシーの定義				
ブロック番号	項目番号	変数名	学年	領域	目標	内容	情報の取り出し	解釈	情報の編集・統合	判断
5	1	B5-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的な部分と付加的な部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。			○	
	2	B5-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文脈の中における語句の意味を的確にとらえ、理解すること。			○	
	3	B5-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	(オ)文章に表れているものの見方や考え方をとらえ、自分のものの見方や考え方を広げること。			○	
	4		B5-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	伝えたい事実や事柄について、自分の考えや気持ちを根拠を明確にして書くこと。			
第1学年				読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章に表れているものの見方や考え方について、知識や体験と関連付けて自分の考えをもつこと。			○	
6	1	B6-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	本や文章などから必要な情報を集めるための方法を身に付け、目的に応じて必要な情報を読み取ること。	○			
	2	B6-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の構成や展開、表現の特徴について、自分の考えをもつこと。			○	
	3	B6-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的な部分と付加的な部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。	○			
	4		B6-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	集めた材料を分類するなどして整理するとともに、段落の役割を考へて文章を構成すること。			
第1学年				読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の構成や展開、表現の特徴について、自分の考えをもつこと。			○	
7	1	B7-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的な部分と付加的な部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。	○			
	2	B7-2	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	集めた材料を分類するなどして整理するとともに、段落の役割を考へて文章を構成すること。				
			第2学年	読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	抽象的な概念を表す語句や心情を表す語句などに注意して読むこと。			○	
	3	B7-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的な部分と付加的な部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。			○	
4	B7-4	第2学年	読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章の構成や展開、表現の仕方について、根拠を明確にして自分の考えをまとめること。			○		
8	1	B8-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的な部分と付加的な部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。			○	
	2	B8-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	本や文章などから必要な情報を集めるための方法を身に付け、目的に応じて必要な情報を読み取ること。	○			
	3	B8-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的な部分と付加的な部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。			○	
	4	B8-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	伝えたい事実や事柄について、自分の考えや気持ちを根拠を明確にして書くこと。				
第2学年			読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章の構成や展開、表現の仕方について、根拠を明確にして自分の考えをまとめること。			○		

表 3.3.2c 学習指導要領と本研究における国語の定義との関係（ブロック 9～12）

ブロック番号	項目番号	変数名	学習指導要領			本研究におけるリーディング・グリテシーの定義				
			学年	領域	目標	内容	情報の取り出し	解釈	情報の編集・統合	判断
9	1	B9-1	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	文脈の中における語句の意味を的確にとらえ、理解すること。	○			
	2	B9-2	第2学年	読むこと	目的や意図に応じ、文章の内容や表現の仕方に注意して読む能力、広い範囲から情報を集め効果的に活用する能力を身に付けさせるとともに、読書を生活に役立てようとする態度を育てる。	文章全体と部分との関係、例示や描写の効果、登場人物の言動の意味などを考へ、内容の理解に役立てること。		○		
	3	B9-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的部分と付加的部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。			○	
	4	B9-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	集めた材料を分類するなどして整理するとともに、段落の役割を考へて文章を構成すること。				
	第2学年		読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章全体と部分との関係、例示や描写の効果、登場人物の言動の意味などを考へ、内容の理解に役立てること。				○	
10	1	B10-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	本や文章などから必要な情報を集めるための方法を身に付け、目的に応じて必要な情報を読み取ること。	○			
	2	B10-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	本や文章などから必要な情報を集めるための方法を身に付け、目的に応じて必要な情報を読み取ること。	○			
	3	B10-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的部分と付加的部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。			○	
	4	B10-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	集めた材料を分類するなどして整理するとともに、段落の役割を考へて文章を構成すること。				
	第2学年		読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	多様な方法で選んだ本や文章などから適切な情報を得て、自分の考えをまとめること。				○	
11	1	B11-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	本や文章などから必要な情報を集めるための方法を身に付け、目的に応じて必要な情報を読み取ること。	○			
	2	B11-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的部分と付加的部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。		○		
	3	B11-3	第2学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	多様な方法で選んだ本や文章などから適切な情報を得て、自分の考えをまとめること。			○	
	4	B11-4	第1学年	読むこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	伝えたい事実や事柄について、自分の考えや気持ちを明確にして書くこと。				
	第2学年		読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的部分と付加的部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。				○	
12	1	B12-1	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の中心的部分と付加的部分、事実と意見などを読み分け、目的や必要に応じて要約したり要旨をとらえたりすること。	○			
	2	B12-2	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	文章の構成や展開、表現の特徴について、自分の考えをもつこと。			○	
	3	B12-3	第1学年	読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	本や文章などから必要な情報を集めるための方法を身に付け、目的に応じて必要な情報を読み取ること。		○		
	4	B12-4	第1学年	書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	集めた材料を分類するなどして整理するとともに、段落の役割を考へて文章を構成すること。				
	第2学年		書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	事実や事柄、意見や心情が相手に効果的に伝わるように、説明や具体例を加えたり、描写を工夫したりして書くこと。					
12B		B12B-4	第1学年	書くこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	多様な方法で選んだ本や文章などから適切な情報を得て、自分の考えをまとめること。				○
	第2学年		書くこと	目的や意図に応じ、日常生活にかかわることなどについて、構成を考へて的確に書く能力を身に付けさせるとともに、進んで文章を書いて考えをまとめようとする態度を考へる。	事実や事柄、意見や心情が相手に効果的に伝わるように、説明や具体例を加えたり、描写を工夫したりして書くこと。					
	第2学年		読むこと	目的や意図に応じ、様々な本や文章などを読み、内容や要旨を的確にとらえる能力を身に付けるさせるとともに、読書を通してものの見方や考え方を広げようとする態度を育てる。	多様な方法で選んだ本や文章などから適切な情報を得て、自分の考えをまとめること。				○	

3.3.3 記述式項目の採点基準

いずれのブロックにも記述式の項目が含まれるようにデザインされているが、その採点基準に関しては、たとえばブロック 1「カレー」における項目番号 3（表 3.3.3.1）やブロック 10「よろしかったでしょうか」（表 3.3.3.2）の場合には以下に示す基準を設け、これに基づき採点者が採点した。採点基準の信頼性を担保するために、採点基準を設けるに当たって複数の作題担当者と採点担当者の協議によった。また、実施後サンプル採点をした後、基準の若干の修正を行った。なお、12 個のブロックのうち、第 3 者による検討に供するため 3 つのブロックを公開する（資料 3.1 から 3.3 参照）。

表 3.3.3.1 採点基準の例（国語ブロック 1：項目番号 3：変数名 B1-3）

国語 01 シズコさん（カレー） 問三
<p>条件</p> <p>A) 嫁に来る条件について触れながら解答している。（姑に、嫁に来るのに料理学校も行っていないのか、と云われたこと。）</p> <p>B) 姑の料理について触れながら解答している。（姑の料理が豚肉とキャベツ入りのすき焼きとてんぷらだけだったこと。）</p> <p>C) 私に基本的な家庭料理が身につけている理由について触れながら解答している。</p> <p>D) 四十五字以上六十字以内で書いている。</p>
<p>正答</p> <p>類型 1 条件を全て満たしているもの。</p> <p>類型 2 AとBとCは満たしているが、Dを満たしていないもの。</p> <p>類型 3 Dを満たし、かつ、AとBとCのうち二つのみを満たしているもの。</p> <p>類型 4 Dを満たし、かつ、AとBとCのうち一つのみを満たしているもの。</p>
<p>誤答</p> <p>類型 5 母が優れた料理人である事を知ったので、やったぜ、と思った、という説明がされているもの。</p> <p>類型 9 上記以外の解答。</p> <p>類型 0 無回答。</p>

表 3.3.3.2 採点基準の例 (国語ブロック 10 項目番号 4 : 変数名 B10-4)

国語 10 「よろしかったでしょうか」はよろしくないか 問四
<p>条件</p> <p>A) 本文において「控えめな敬語」は「よろしいでしょうか」ではなく「よろしかったでしょうか」である、ということについて正しく理解して解答している。</p> <p>B) 図のデータから必要な根拠を正しく抜き出して解答している。（「変な言い方だ」と思っている人が多い。「ていねいだ」と思っている人が若い人に少ない。など）</p>
<p>正答</p> <p>類型 1 条件を全て満たしているもの。</p> <p>類型 2 Aについて触れずに、Bを満たしているもの。</p>
<p>誤答</p> <p>類型 3 データの説明について不十分に解答しているもの。（「変な言い方」や「ていねい」に言及せず、）世代や人によって感じ方が変わるから、などと解答しているもの。</p> <p>類型 4 相手を侮辱している（不快にしている）ような印象を与える、という解答をしているもの。</p> <p>類型 9 上記以外の解答。</p> <p>類型 0 無回答。</p>

さらに採点結果の数値化のルールであるが、ブロック 1 の場合には正答とされる類型が 4 通りあり、そのうち類型 1 が必要な条件がもっとも多く、類型 4 になるにしたがって条件の数は減少する。そして類型 5 からは誤答とされ、そのうち無回答の場合は類型 0 と判定された。まず単純な数値化としては正答とされる類型 1 から 5 までに数値 1 を付与し、それ以外に数値 0 を与えるケースが考えられる。これを「2 値データ」と呼ぶ。次に 2 値データの変形で無回答を誤答とは区別したケースが考えられる。このデータは無回答行動の構造を探るときにのみ使用することとした。さらに、正答である類型 1 から 4 までに、条件の厳しい類型ほど高い学力が必要であると考え、類型 1 には数値 4、類型 2 には数値 3、類型 3 には数値 2、類型 4 には数値 1 を付与し、それ以外は数値 0 を与えた。このケースに対応するデータを「多値データ」と称することにする。したがって多値データにおいては記述式以外の部分では 2 値データも含まれていることになる。ブロック 10 も同様である。表 3.3.3.3 にブロック 1 と 10 における数値化の変換ルールを示した。他のブロックにおける記述式問題の類型の 2 値化及び多値化の変換ルールは資料 2.1.4 を参照されたい。

表 3.3.3.3 記述式項目の解答類型の数値化のルール（例）

ブロック	項目	類型	二値	多値
1	3	1	1	4
		2	1	3
		3	1	2
		4	1	1
		5	0	0
		9	0	0
		0	0	0
10	4	1	1	2
		2	1	1
		3	0	0
		4	0	0
		9	0	0
		0	0	0

4 テストの実施手続

4.1 実施の基本方針

時系列による学力水準の比較のための調査研究を数学において実施する。あわせてリーディング・リテラシーの測定を中心とした国語の調査を試行し、同一学年における複数科目の実施に関するノウハウ及びその結果を学力の相関構造の分析に活かすためのノウハウを開発する。なお、本調査研究のための学力測定技術としては、すでに PISA や PIAAC などの国際調査で採用されている項目反応理論によるものを基盤とする。

4.2 実施内容

対象学年 中学3年生

対象人数 1954名

実施方式 宮城県内の5つの自治体の学校(20校)

実施時期 平成23年11月10日から11月29日の期間

実施時間 上記の期間内の2授業時間(数学・国語各1時間)

(説明や配布、回収の時間5分を含めて1授業時間50分)

出題教科 数学・国語

出題範囲 現行の学習指導要領の内容構成にあわせた。国語についてはPISA型のリーディング・リテラシー問題を中心とした。

問題内容 数学については国立教育政策研究所の助言を受けた。

分冊数 8分冊

解答数 中学校ともに一人当たり30問程度 問題冊子は回収・非公開

解答方式 記入式

個票返却 平成24年1月初旬 各学校へ郵送 (国語の採点に時間がかかるため)

学校の負担 実施にかかる部分のみ。採点等はすべて研究組織側で行った。

4.3 返却した個票の主な内容

- 1) 解答した分冊に含まれている問題についての正誤情報
- 2) 割り当てられなかった分冊に含まれている問題については推定正答確率の値
- 3) すべての問題に関して学習指導要領の項目事項を記載

具体的には「資料4.1.3 実施概要：資料2：調査結果シート」を参照のこと。

4.4 教員向け説明内容

重複テスト分冊法の基本的な考え方を説明した文書を作成し送付した。「資料4.1.2 実施概要：資料1：重複テスト分冊法の基本的な考え方」を参照のこと。

5 数学と国語の信頼性

5.1 数学の信頼性

昨年度の調査研究により、数学のテストの信頼性については検討済みである。本調査研究で使用した問題は、昨年度の問題をベースに作成されたものであり、信頼性が大きく異なっていることは考えにくい。そこで、本報告書では数学に関しては信頼性係数の推定値を算出するに留めることとする。

表 5.1.1 に数学データに関して分冊ごとに求めたテストの信頼性係数の推定値であるクロンバックの α を示した。いずれの分冊も項目数は 32 であった。受検者数はいずれの分冊もほぼ見込み通り 244 名(=1954 名/8 分冊)前後に割り振られていることがわかる。また α 係数は、いずれの分冊もほぼ 0.9 と高い値であり、分冊単独であっても十分な測定精度を持っていることがわかる。問題を非公開にすることで精度のよい項目を再び利用できるためこのようなことが可能となる。

表 5.1.1 分冊ごとの信頼性係数の推定値

分冊	1	2	3	4	5	6	7	8
項目数	32	32	32	32	32	32	32	32
受検者	256	257	252	251	243	238	233	256
信頼性係数	0.906	0.898	0.903	0.907	0.918	0.903	0.905	0.904

また、ブロックごとの信頼性係数は表 5.1.2 に示すとおりである。各ブロックの信頼性係数の推定値は、0.374~0.657 であった。分冊ごとの値に比べると低い値となっているのは、各ブロックの項目数が 4 つと少ないためである。

表 5.1.2 ブロックごとの信頼性係数の推定値

ブロック	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
項目数	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
受検者	980	964	979	979	979	979	984	974	984	974	973	994	985	964	978	994
信頼性係数	0.602	0.537	0.520	0.417	0.503	0.640	0.572	0.374	0.603	0.598	0.657	0.645	0.391	0.605	0.588	0.577

さらに本年度のブロック配置計画（資料 1.1.3 参照）では分冊 3 と分冊 7，分冊 4 と分冊 8 でブロックの順次のみが異なり、構成するブロックそのものは同一のデザインを組んでいる。そのため、たとえば、分冊 3 と分冊 7 の受検者群は同じ母集団からの確率標本とみなせることから集団としての学力分布はほぼ等しいと仮定できる。また、構成する項目も同じであるため、項目の順序効果が無視できるならば、分冊 3 と 7 の正答数得点の分布は等しいはずである。分冊 4 と 8 においても同様である。実際、分冊 3 を受けた集団の分冊 3 の正答数得点の平均と標準偏差はそれぞれ 10.73 と 3.93，分冊 7 を受けた集団の分冊 7 の正答数得点の平均と標準偏差は 10.98 と 3.89 でありほとんど等しいことがわかる。分冊 4 と 8 についても、同様に、前者が 8.61 と 4.07，後者が 8.51 と 4.05 となる。自由度が大きいものにも関わらず、独立 2 標本の平均差に関する t 検定を行っても統計的に有意とはならない。したがって、各分冊には重複テスト分冊法の見込み通り、等質な集団を割り当てることができると考えることができる。

5.2 国語

5.2.1 テストの信頼性

表 5.2.1 に国語データに関して分冊ごとに求めたテストの信頼性係数の推定値であるクロンバックの α を示した。いずれの分冊も項目数は 12 であった。受検者数としては、分冊 4, 4B, 7, 7B が、実施のスケジュール調整のミスと、ブロック 12 の改訂を行う必要が生じたための二つの要因が重なり、結果として、同じ分冊であっても 2 種類のバリエーションが生じてしまい、そのブロックが含まれるこれらの分冊については、バリエーションに対応して 60 名から 170 名程度と少なくなってしまった。その他の 6 分冊では 220 名～280 名程度であった。全分冊の α 係数は、0.511 から 0.760 を示していた。

5.1 に記載した数学に比べ、国語のテストは相対的に信頼性係数が低くなるのが経験的に知られている。その理由としては、一般に、国語の問題は、1 つの題材（文章）に対して、学習指導要領の領域及び目標の異なる小問が複数用意される。例えば、小説の一部抜粋を用いて大問を作成して、漢字の読み書きを問う項目や、登場人物の心情を問う項目をいくつか小問として出題することが多い。このように、同じ大問内に質的に異なる内容を問う小問が複数含まれていることから、その度合いが少ない数学に比して国語の大問ごとの信頼性は低くなる傾向にある。

表 5.2.1 分冊ごとの信頼性係数の推定値

分冊	1	2	3	4	4B	5	6	7	7B	8
項目数	12	12	12	12	12	12	12	12	12	12
受検者	274	257	255	170	62	246	244	162	60	224
信頼性係数	0.606	0.511	0.664	0.760	0.683	0.573	0.666	0.562	0.509	0.723

ブロックごとの信頼性係数は表 5.2.2 に示すとおりである。分冊ごとの信頼性係数よりも、さらに低い値（0.162～0.497）が得られている。数学における分冊とブロックの信頼性係数の関係と同じ理由で、国語でも、同様に項目数の少ないブロック単位の信頼性係数が低くなっている。また、数学のブロック単位の信頼性係数に比べても相対的に国語のブロック単位の信頼性係数が低くなっているのも分冊におけるのと同じ理由である。

表 5.2.2 ブロックごとの信頼性係数の推定値

ブロック	1	2	3	4	5	6	7	8	9	10	11	12	12B
項目数	4	4	4	4	4	4	4	4	4	4	4	4	4
受検者	520	501	477	456	478	499	479	498	518	503	479	332	122
信頼性係数	0.231	0.340	0.393	0.514	0.403	0.440	0.162	0.415	0.497	0.357	0.378	0.465	0.431

さらに、分冊ごとに因子分析を行い、固有値を求めたところ、いずれも第一因子の固有値の値が他の因子のものに対して突出して高かったものの、分冊によって固有値の減衰状況が異なっていた。分冊ごとのデータにもとづき主因子解因子分析の結果得られた因子負荷の値を資料 2.5 に掲載した。

公開問題 2 問が入っている分冊 4・分冊 4B（ブロック 4, 5, 12）と分冊 7・分冊 7B（ブロック 3, 7, 12）の、スクリープロットを図 5.2.1～5.2.4 に示す。分冊 4, 分冊 4B は α 係数がそれぞれ 0.760, 0.683 と高いという特徴がある。一方、分冊 7, 分冊 7B は α 係数が 0.559, 0.518 と他の分冊と比べる

と低い。信頼性係数の高い分冊4のスクリープロット（図5.2.1）では、第一因子が突出しており、因子分析モデルから判断しても、分冊の一次元性は確保できているように見える。このことから、分冊4及び分冊4Bでは、国語の中でも比較的一貫した能力を測定する項目から成っていたといえよう。しかしながら、信頼性係数の低い分冊7Bのスクリープロット（図5.2.4）では、第一因子が高いものの、固有値の減衰状況からも分冊の一次元性が確保できているというよりは、複数の次元を反映した結果であると捉えられ、多様な国語能力を測定していた分冊であったといえよう。

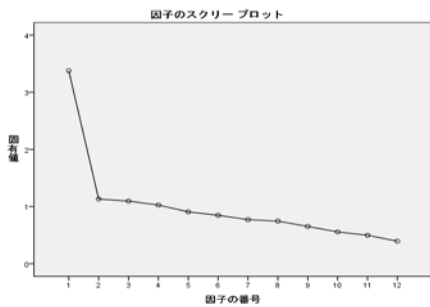


図 5.2.1 分冊4のスクリープロット

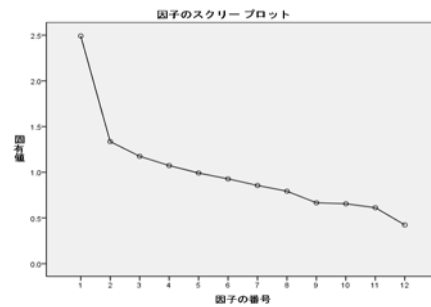


図 5.2.2 分冊4Bのスクリープロット

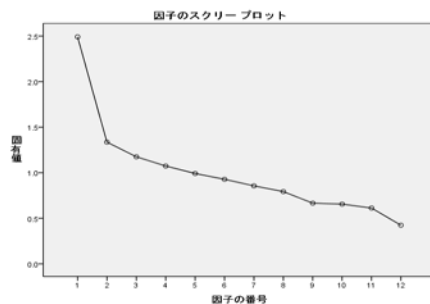


図 5.2.3 分冊7のスクリープロット

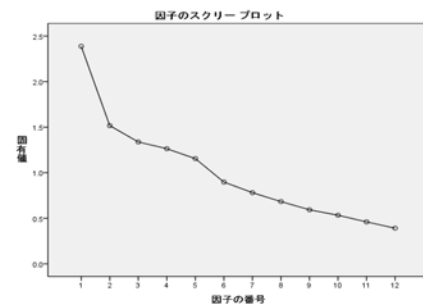


図 5.2.4 分冊7Bのスクリープロット

さらに、因子数を3と固定し、分冊ごとの因子分析（主因子法、プロマックス回転）を行った。その結果得られた因子間相関係数と信頼性係数の相関を見る。因子間相関係数は、因子同士の相関を計算したものであり、いわば分冊内で測定している能力のまとまっている度合いを示す指標の1つであるといえよう。各分冊につき因子間相関係数が3値ずつ計算されたため、その平均を表5.2.3に示した。値は0.07～0.595と異なっている。

平均の低かった分冊に注目してみると、例えば分冊4B, 5, 7などが挙げられる。分冊の4Bや7は、因子間相関係数の最小値と最大値の差が小さいことより、そもそも因子間があまり相関していなかったため、平均が低かったことが伺える。すなわち、この二冊は、それだけ様々な側面を測定している分冊といえよう。それに対し、分冊5は因子間相関係数の最小値が-0.310と中程度の負の相関を示す値である一方で、最大値は0.660と高い正の相関を示す値が得られている。つまり、分冊5は2つの類似する側面に関する項目と、それらとは方向性の異なる側面に関する項目から成っていたことが考えられる。

表 5.2.3 分冊ごとの因子間相関係数の平均

分冊	ブロック	N	因子間相関係数		
			min	MAX	平均
1	1,8,9	274	0.383	0.503	0.445
2	2,7,10	257	0.026	0.546	0.274
3	3,6,11	255	0.442	0.538	0.487
4	4,5,12	170	0.205	0.606	0.385
4B	4,5,12B	62	-0.035	0.275	0.142
5	1,5,10	246	-0.310	0.660	0.079
6	2,6,9	244	0.432	0.563	0.504
7	3,7,12	162	0.109	0.296	0.227
7B	3,7,12B	60	0.260	0.371	0.309
8	4,8,11	224	0.569	0.617	0.595

図 5.2.5 には、分冊ごとの信頼性係数と分冊ごとの因子間相関係数の平均との散布図を示した。両者ともに、分冊内での信頼性を示す指標であることから、右上がりの線形関係にある。図 5.2.5 より、左下の部分に位置にする分冊 2, 5, 7, 7B は信頼性係数と因子相関係数の平均値の両方が低く、幅広い国語能力を測定する項目から構成されていたことがわかる。逆に右上に位置する分冊 3, 4, 6, 8 は、その分冊内の項目の測定する国語能力が比較的一貫していたと解釈できる。

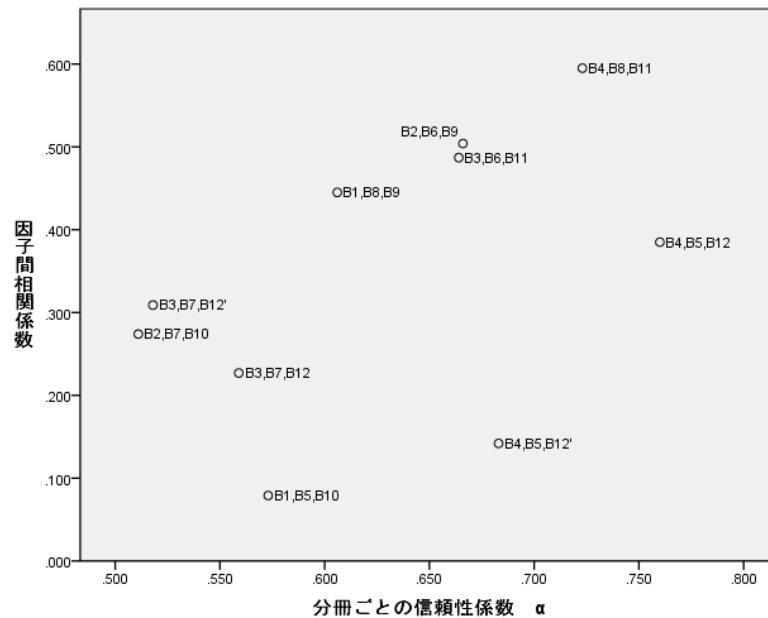


図 5.2.6 分冊ごとの信頼性係数と分冊ごとの因子間相関係数の平均との散布図

次に点双列相関係数であるが、これはその項目とその項目が含まれているテストの得点との相関を計算したものである。その意味で項目-テスト得点 (Item-Test) 間相関、あるいは IT 相関と呼ばれる

こともある。この指標はいわばその項目がそのテストで測定している特性をどの程度とらえているかを見るものであり、識別力とよばれる統計量の1つでもある。なお、点双列相関を計算する際のテスト得点の中に、当該項目の正誤情報を含む場合と、それを除外する場合があり、前者の方がその値は高くなる。本研究では後者の得点を利用している。

図 5.2.6 には、この点双列相関係数と正答率との散布図が描かれている。一般的に点双列相関係数の値は 0.2 以上が望ましいとされることが多いが、その基準で見ると、良好な結果が得られた項目が多い。しかし、点双列相関係数が 0.2 を下回った項目が 9 項目見られ、係数の値が小さい順に、B7_2, B7_4, B1_4, B3_3, B10_2, B2_2, B1_1, B11_4, B1_2 であった。すなわち、これらの 9 項目はテストで測定している特性をうまく捉えることができなかつた項目と言えよう。これらの項目が含まれていたブロック 1, 2, 3, 7, 10, 11 はブロックごとの信頼性係数が 0.4 以下と比較的低かった。点双列相関係数が特に低かつた B7_2, B7_4 が含まれていたブロック 7 は、先に述べたように信頼性係数が 0.162 と低かつた。さらに、ブロック 7 が含まれた分冊 2, 分冊 7, 7B は、他の分冊に比べて信頼性係数が低かつた。このように項目の点双列相関係数は、その項目が含まれるブロックや分冊の信頼性とも関係するものである。

また、図 5.2.6 より、例えば B1_4 や B2_2, B7_2 など正答率も低いいため、テストで測定しようとした能力を測定するには難しすぎる項目であったことが考えられる。それに対し、B1_1 や B1_2 などの項目は点双列相関係数が 0.2 を下回っているものの、正答率は高い。すなわち、そもそも本研究で測定しようとした能力とは異なる能力に関係した項目であったといえよう。

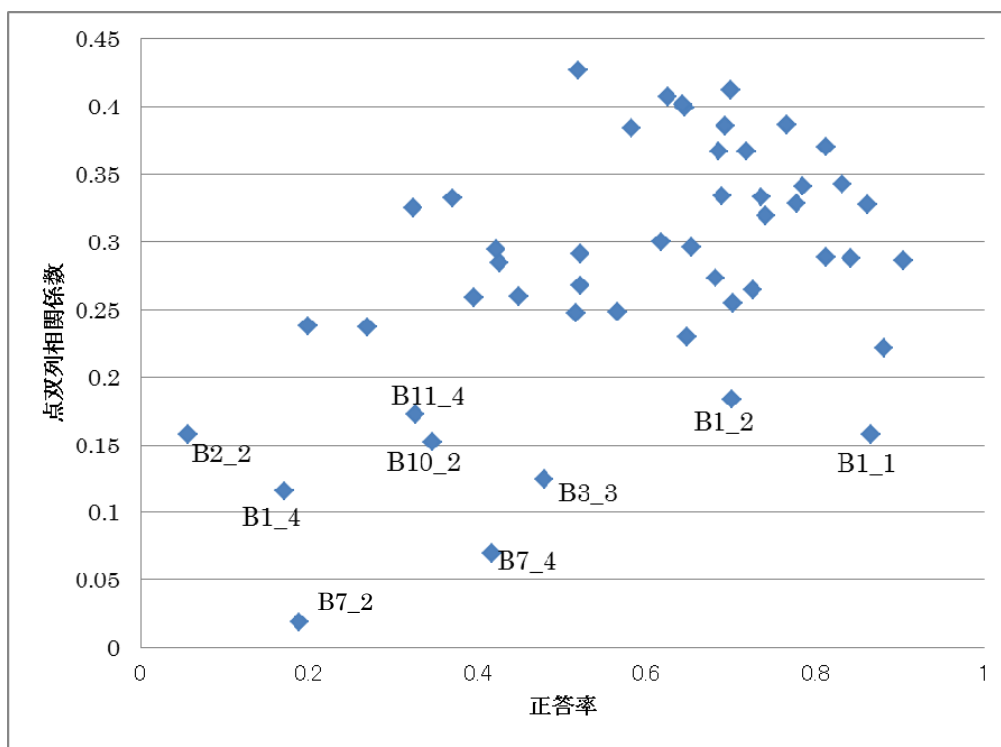


図 5.2.6 国語の正答率と点双列相関係数の散布図

点双列相関係数と因子負荷との関係から、項目の性質を検討する。データ行列の一次元性が強い（矩形行列における最大特異値が他の特異値に比して著しく大きい）場合には点双列相関係数と因子負荷はほぼ同じ情報をもつ。このデータの場合も図 5.2.7 に示すように、全ての分冊に関して求めた両者の値をプロットすると直線状にならぶ。しかし、点双列相関係数の点からも因子負荷の点からも値が著しく低く、外れ値となった項目が6項目見られた。すなわち、分冊 K7 及び分冊 K7B の B3_3、分冊 K2 及び分冊 K7 の B7_2、分冊 K7B の B7_4 及び分冊 K7B の B12_2 であった。逆に、点双列相関係数と因子負荷のいずれの点からみても値が最大であったのは分冊 K4 の B5_2 であった。

先に述べたように分冊 7、7B は信頼性係数が低く、一因子性が低かった。図 5.2.7 で外れ値とされている項目のほとんどがこの二つの分冊に含まれていた。このことから、分冊 7 及び分冊 7B では、このテストで測定しようとしていた国語能力とは必ずしも一致しないものを測定する項目が含まれていたと結論づけることができる。

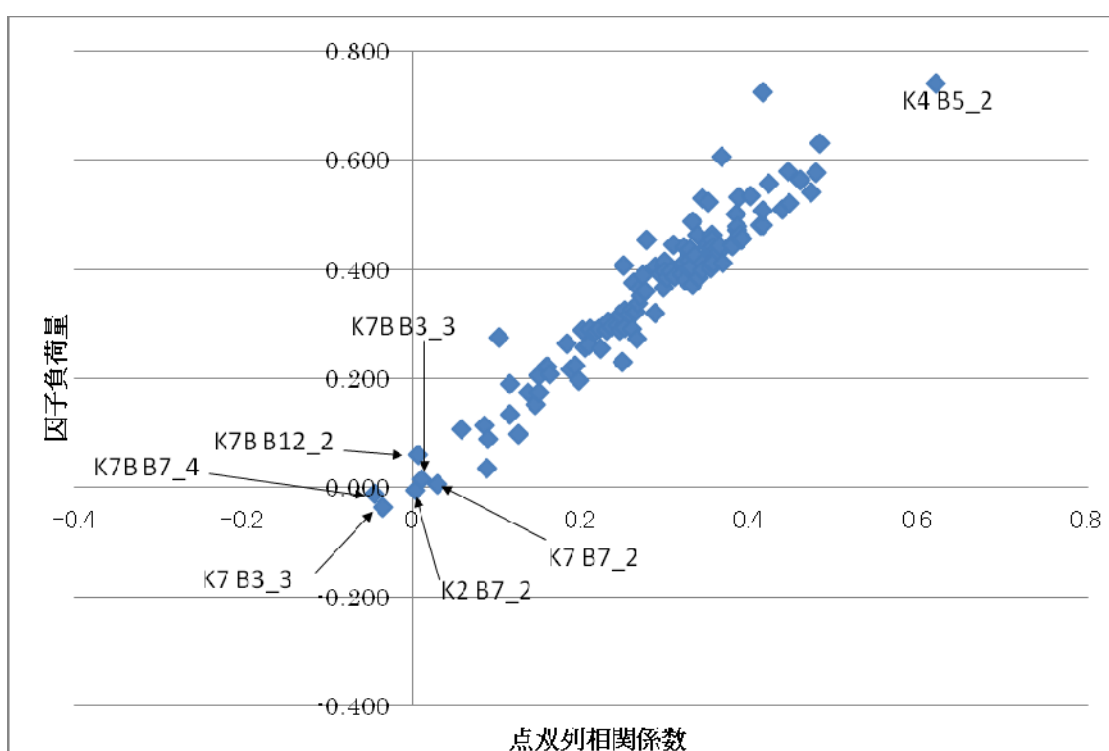


図 5.2.7 点双列相関係数と因子負荷量の散布図

5.2.2 項目分析

テストの信頼性の検討によって、国語の能力は多様であり、その多様な能力を測定するか、それとも一部の能力に焦点を当てるかによってテストの精度の測定が異なることが指摘された。次に考察すべきはそのテストを構成している個々の項目がそれぞれ期待通りの機能を果たしているかどうかの確認である。検討すべき統計量が膨大になるので、この考察に必要な数値は全て資料編を参照されたい。まず、資料 2.1.1 及び 2.1.2 には国語データで使ったテストの各項目のブロック番号、変数名、学年、領域、単元、内容の対照表が示されている。次に資料 2.2～資料 2.4 では各項目の受検者数、正答率、標準偏差、点双列相関係数が記載されている。このうち正答率はその項目を受けた全受検者中何割が正答したかの比率であるので、この値が高いほどその項目は易しかったことになる。

上のような考察をさらに詳細に行うものが資料 2.6 に示す項目分析である。項目分析に使われる図の見方について項目 B1_1 を例にとって説明する。各項目の上段には図 5.2.8a に示すような回答パターンごとの回答率を表したヒストグラムが描かれている。この例の場合、選択肢が 5 種類、無答及びそれ以外の回答をあわせたものの計 7 種類のパターンに分類してヒストグラムが描かれている。項目によっては誤答のパターンがこれよりも増減する。さらに、この B1_1 は選択問題であるが、記述問題の場合、部分得点を与えることがあるため正答パターンが複数存在することがある。そのため、項目によって反応パターンの合計が異なることには注意が必要である。

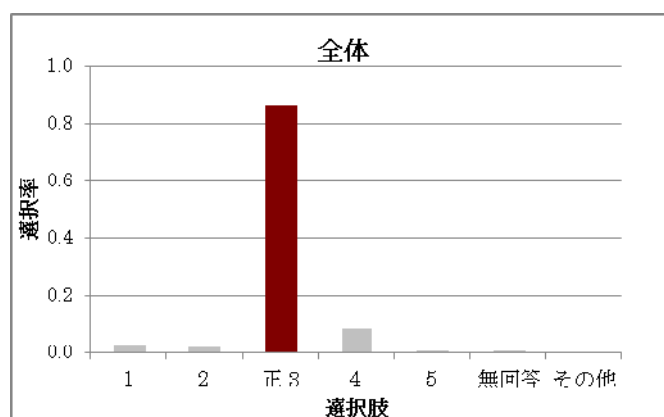


図 5.2.8a B1_1 の回答パターンのヒストグラム

次に中段には図 5.2.8b に示すような受検者の学力レベルを 5 段階に分類して段階ごとにそれぞれの回答パターンがどのくらいの比率でいるかを示した GP 分析図 (Good-Poor Analysis) が描かれている。学力レベルが上がるにつれて正答率も上昇していることがわかる。逆に誤答した比率は学力レベルの上昇とともに下がっていることもわかる。なお、これらの GP 分析図における点双列相関係数は各項目と後述する学力特性値 (θ) との相関をもとめているので、上で求めた点双列相関係数の値とは異なっている。

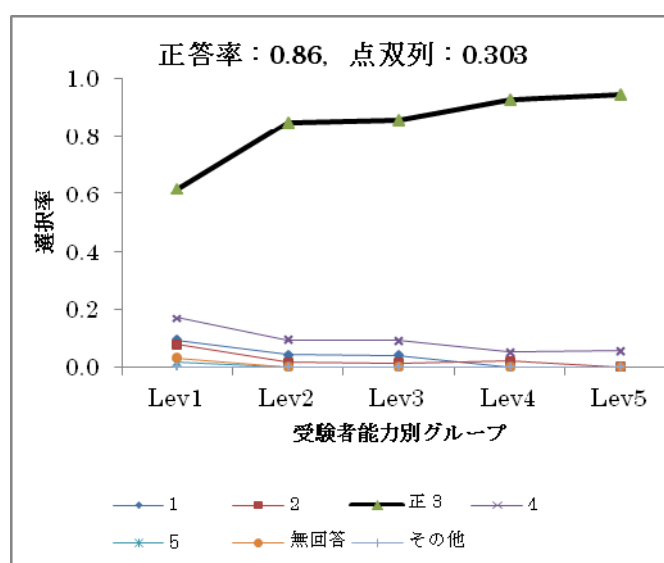


図 5.2.8 b B1_1 の GP 分析図

さらに下段には図 5.2.8c に示すような上記のグラフを描くための正答率の数値データが埋め込まれている。したがって各項目の良否を検討するうえで本質的な情報をわかりやすくとらえているのは GP 分析図ということになる。

レベル別解答選択率						
解答類型	全体	Lev1	Lev2	Lev3	Lev4	Lev5
1	0.027	0.092	0.042	0.039	0.000	0.000
2	0.019	0.077	0.017	0.013	0.020	0.000
正 3	0.865	0.615	0.849	0.857	0.929	0.944
4	0.083	0.169	0.092	0.091	0.051	0.056
5	0.002	0.015	0.000	0.000	0.000	0.000
無回答	0.004	0.031	0.000	0.000	0.000	0.000
その他	0.000	0.000	0.000	0.000	0.000	0.000
人数	520	65	119	77	98	161

図 5.2.8c B1_1 のレベルごとの各解答類型の反応率

例えば B1_1 は正答率が 0.86 と極めて易しい、したがって多くの生徒が正答している項目であるが、その GP 分析図は図 5.2.8b に示すようになる。当然のことながら、レベル 1 のように学力が低い層の識別に役立つ項目ということになる。ただし図に示すレベル 2 以上の学力層ではほとんど全員が正答しているため、これらの層の学力の識別にはあまり機能していない項目であることがわかる。

ここで、これまで信頼性係数の推定値や因子分析の結果、点双列相関係数から、外れ値である可能性が指摘されていた B7_2, B7_4, の項目分析結果を見てみる。

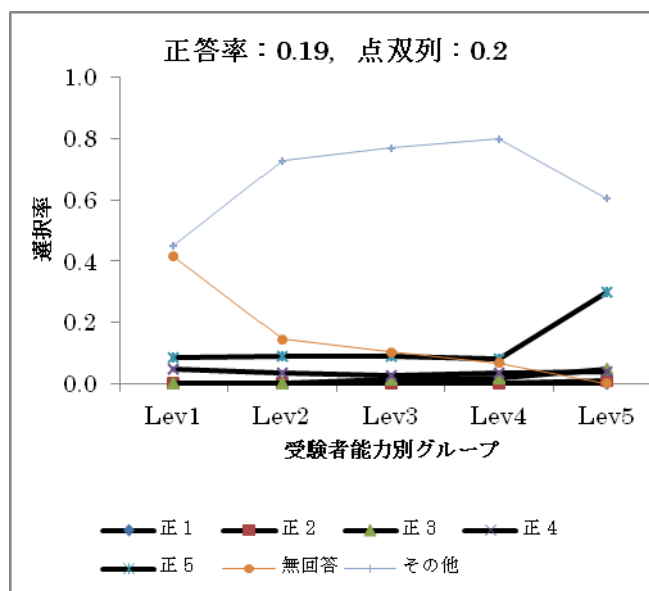


図 5.2.9 B7_2 の GP 分析図

B7_2 は記述問題であったため、正答パターンが 5 種類ある。しかしながら、図 5.2.9 より、いずれのパターンも選択率が低く、このことからわかるように正答率が 0.19 と低く、難しい問題であった

ことが伺える。レベル4と5で正答の選択率が急激に変化していることから、学力の高いグループでの識別力の高い項目であった可能性を指摘できる。しかしレベル5においても誤答率が正答率を上回っていることから、難しすぎる問題であったことは否めない。

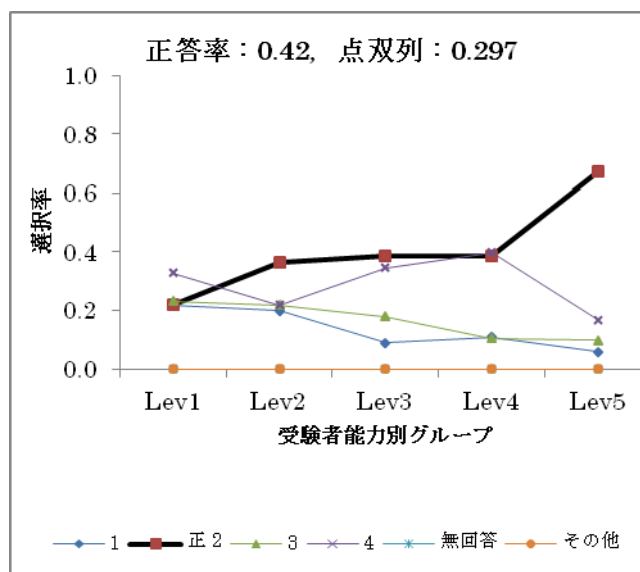


図 5.2.10 B7_4 の GP 分析図

次に B7_4 の GP 分析図を図 5.2.10 に示す。図 5.2.10 より、レベル2からレベル4にかけて正答率が横ばいになっている。このことから、この項目は学力レベルの高い者と低い者を識別するのに適しているものの、中程度の者の学力は識別しにくい項目であったといえよう。この正答率のみをみるとよい項目であったように見えるが、選択肢“4”の選択率、すなわち誤答率が、レベル2から4にかけてあがっている。このことから、この項目の選択肢4は学力レベルの中程度の生徒を“ひっかける”ようなものであったことが考えられる。

具体的には7_4は以下のような項目であった。

- 問い この文章の特徴を述べたものとして不適切なものを次の中から一つ選びなさい。
- 1 具体例が述べられている
 - 2 比喩を使って述べられている
 - 3 重要な点が繰り返し述べられている
 - 4 二つのことについて対比しながら述べられている

このように「不適切なもの」を選択させる問題は、生徒の混乱を招く危険があるため、一般的には避けるべき表現とされる。実際に本調査では、B7_4の横に記載されていたB7_3は「正しいもの」を選ぶ項目であったため、項目間での混同があった可能性は否定できない。さらに、文章中に「隣の誰かに騙されない」という比喩的な表現が使われているため、“2”を選択しなかった者もいたと考えられ

る。このように、B7_4 は生徒の混乱を招いたため、図 5.2.7 などでは外れ値として表れてきたのであろう。

このように、GP 分析図の結果を参照し、実際の問題に立ち戻ることによって、問題のよしあし及びその理由を考察することが可能になる。

6 数学における経年比較の実際

6.1 概要

ここでは、2010年に実施された数学テストと2011年に実施された数学テストを等化することにより、それらのテスト結果を（経年）比較することが目的となる。

2章で述べたとおり、テストを等化する場合には、はじめに等化のデザインをどのようにするのが問題となる。本研究では、2010年度テストと2011年度テストに共通の項目群を配置する「共通項目デザイン」を採用している。実際のテストデザインは図6.1.1のようになる。2010年度のテストと2011年度テストに共通項目部（網掛け部分）がある。より詳細には、各年度のテストは分冊テスト法により、小冊子部分に分割されるが、この図では省略している。

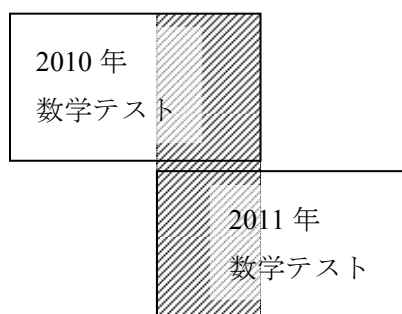


図 6.1.1 数学テストの等化デザイン図

6.2 等化の計算方法の比較

等化デザインが決定され、それに従いテストデータが収集された後に、等化の計算が行なわれる。すでに述べたとおり、等化の計算手続きも複数の方法が存在し、必ずしも等化デザインと1対1でそれが決まるものでもない。ここでは、現在得られているテストデータに対して複数の等化計算を行ない、その数値結果及び今後の研究継続可能性なども考慮しながら、各方法の比較検討を行なっていく。

今回採用した計算手続き

現在得られているテストデータについて、全項目数と共通項目数の割合、各テスト受検者数などを考慮し、以下の計算手続きを採用することとした。

- (1) 同時尺度調整法 (concurrent calibration)
- (2) 項目固定法 (fixed items method)
- (3) 共通項目情報を用いた mean / sigma 法
- (4-a) 共通受検者情報を用いた mean / sigma 法
- (4-b) 共通受検者情報を用いた mean / sigma 法に野口・熊谷 (2011) の補正を加えた方法
- (5) 熊谷・野口 (印刷中) による推定母集団分布情報を用いた mean / sigma 法

今回のテストデータに対する各計算手続きの概要は以下の通りである。

- (1) 同時尺度調整法 (concurrent calibration)

本テストデータにおける同時尺度調整法は、図 6.1.1 の状態のテストデータ全体に対してそのまま IRT 分析を行うことで実施することができる。この際、他の方法との比較を可能にするため、多母集団モデルを適用し、2010 年度受検者の母集団分布について、その平均が 0.0、標準偏差が 1.0 となるような尺度の基準化を行なった。

(2) 項目固定法 (fixed items method)

本テストデータにおける項目固定法は、はじめに図 6.1.1 の 2010 年度テストデータだけを用いて、項目母数の推定を行う。次に、2011 年度テストデータだけを用いて項目母数の推定を行うが、この際、2010 年度テストとの共通項目部に関しては、先に 2010 年度テストデータの分析で得られた項目母数の値を既知として、推定計算を行う (2010 年度との共通項目部については推定計算が行なわれない)。このようにすることで、2011 年度テストデータを用いて得られた項目母数は、全て 2010 年度テストの尺度上で表現される。

(3) 共通項目情報を用いた mean / sigma 法

本テストデータにおける共通項目情報を用いた mean / sigma 法は、はじめに図 6.1.1 の 2010 年度テストデータ部及び 2011 年度テストデータ部に対して、それぞれ独立に項目母数の推定を行う。これにより、2010 年度と 2011 年度の共通項目については、2010 年度テストデータを用いて推定された項目母数と、2011 年度テストデータを用いて推定された項目母数の 2 種類が算出される。この 2 種類の項目母数を用いて、2 章で見たように項目困難度の平均・標準偏差から等化係数の計算を行う。

(4-a) 共通受検者情報を用いた mean / sigma 法

本テストデータにおける共通受検者情報を用いた mean / sigma 法は、はじめに (3) と同様に、2010 年度テストデータ部及び 2011 年度テストデータ部に対して、それぞれ独立に項目母数の推定を行う。次に、2011 年度テストデータ部を用いて、受検者の潜在特性尺度値 θ の推定を行う。この際、2010 年度との共通項目部分のテストデータ及び 2010 年度テストデータから得られた項目母数を用いて推定した $\hat{\theta}_{2010}$ と、2011 年度新規項目部分のテストデータ及び 2011 年度テストデータから得られた項目母数を用いて推定した $\hat{\theta}_{2011}$ の 2 種類の潜在特性尺度値を算出する。この 2 種類の潜在特性尺度値の平均・標準偏差から、等化係数の計算を行う。

(4-b) 共通受検者情報を用いた mean / sigma 法に野口・熊谷 (2011) の補正を加えた方法

(4-a) での 2011 年度テストデータから潜在特性尺度値の推定を行う場面において、2010 年度との共通項目数は冊子によって異なるが 19~25 項目であり、2011 年度新規項目数は 7 から 13 項目となっており、その比は 2:1~3:1 程度であった。野口・熊谷 (2011) では、このような場合、潜在特性尺度値の推定値に含まれる誤差 (標準誤差) の影響により、mean / sigma 法では適切な等化係数の推定が行えない場合があることを指摘している。またそのような場合に、誤差の影響を取り除いたうえで mean / sigma 法を適用する、補正方法を提案している。詳細な計算手順については、野口・熊谷 (2011) を参照されたい。

(5) 熊谷・野口（印刷中）による推定母集団分布情報を用いた mean / sigma 法

(4-b) における野口・熊谷 (2011) の方法が、誤差の影響を直接取り除く方法であったのに対し、熊谷・野口（印刷中）では潜在特性尺度値の推定値を用いずに、受検者の推定母集団分布の情報を用いることで、誤差の影響を低減する方法を提案している。野口・熊谷 (2011) では条件により適切な誤差成分（誤差分散）が推定できない場合があるが、熊谷・野口（印刷中）では、そのような状況でも安定した等化係数の推定を行うことができる方法となっている。詳細な計算手順については、熊谷・野口（印刷中）を参照されたい。

6.3 等化結果の比較

前節でとりあげた各方法を用いて 2010 年度テストデータと 2011 年度テストデータを等化することで得られた「等化後」の識別力母数を表 6.3.1 に、困難度母数を表 6.3.2 に示す。各表には、全方法の平均値も最終列に付してある。また (1) の方法以外は、2010 年度テストに出題された問題項目 (2011 年度テストとの共通項目も含む) の項目母数は全て同値になるため、各表には 2011 年度に新たに出題されたもののみを示してある。なお、計算には全て EasyEstimation (熊谷, 2009) を使用した。

表 6.3.1, 6.3.2 から、方法 (4-a) を除き、ほとんどの方法・項目で全体との平均から ± 0.1 の範囲に収まっており、方法間の数値が非常に近いものであることが分かる (今回のデータについて方法 (4-a) がずれる理由については、前節を参照のこと)。シミュレーション実験とは異なり、これらの方法のなかでどれが「真の状態」に最も近いのかを断定することはできない。しかしながら、たとえば (2) と (5) の方法などは、その計算手順は全く異なるものであるものの、等化後の項目母数の値は非常に近いものとなっていることなどから、適切に等化が行なわれたことが示唆される。

今後さらなるテストデータの追加などの発展可能性を考慮すると、(1) の方法はデータの追加によりすべての項目母数が再計算されてしまうため、用いない方がよいであろう。計算作業の容易さ、及び等化後項目母数が平均値に最も近いことなどを考慮し、本研究では等化方法として (2) 項目固定法によるものを採用することとする。

表 6.3.1 等化後項目母数の比較（識別力母数）

項目番号	(1)同時尺度 調整法	(2)項目 固定法	(3)共通項目 (M&S)	(4-a) 共通 受検者 (M&S)	(4-b) 共通 受検者 (NK 補正)	(5) 推定 母集団	平均値
11_2	0.98	0.93	0.88	1.01	0.86	0.92	0.93
11_8	0.67	0.64	0.61	0.70	0.59	0.63	0.64
11_9	0.83	0.80	0.77	0.88	0.75	0.80	0.80
11_12	1.30	1.22	1.16	1.34	1.13	1.22	1.23
11_19	0.56	0.53	0.51	0.59	0.50	0.54	0.54
11_20	0.91	0.86	0.83	0.96	0.81	0.87	0.87
11_21	1.06	1.02	0.95	1.10	0.93	1.00	1.01
11_22	1.05	1.01	0.96	1.10	0.93	1.00	1.01
11_31	0.13	0.13	0.12	0.14	0.12	0.13	0.13
11_32	0.84	0.81	0.78	0.90	0.76	0.82	0.82
11_39	1.28	1.23	1.15	1.33	1.12	1.20	1.22
11_42	1.13	1.08	1.03	1.19	1.00	1.08	1.08
11_44	1.96	1.89	1.81	2.09	1.77	1.89	1.90
11_52	0.25	0.25	0.23	0.27	0.23	0.24	0.25
11_54	0.90	0.86	0.82	0.95	0.80	0.86	0.87
11_56	1.38	1.31	1.21	1.40	1.19	1.27	1.29
11_58	1.68	1.55	1.47	1.70	1.44	1.54	1.56
11_61	0.82	0.78	0.75	0.86	0.73	0.78	0.79
11_63	0.65	0.63	0.59	0.69	0.58	0.62	0.63
11_64	1.11	1.06	1.02	1.18	0.99	1.07	1.07

表 6.3.2 等化後項目母数の比較（困難度母数）

項目番号	(1)同時尺度 調整法	(2)項目 固定法	(3)共通項目 (M&S)	(4-a)共通 受検者 (M&S)	(4-b)共通 受検者 (NK 補正)	(5)推定 母集団	平均値
11_2	-1.03	-1.06	-1.09	-0.99	-1.15	-1.05	-1.06
11_8	-0.83	-0.85	-0.86	-0.79	-0.91	-0.83	-0.85
11_9	-0.58	-0.58	-0.57	-0.54	-0.62	-0.56	-0.57
11_12	-0.95	-0.98	-1.00	-0.91	-1.05	-0.96	-0.98
11_19	0.20	0.22	0.26	0.19	0.24	0.24	0.23
11_20	0.22	0.25	0.30	0.22	0.27	0.27	0.26
11_21	-0.49	-0.49	-0.48	-0.46	-0.52	-0.47	-0.48
11_22	-0.47	-0.47	-0.46	-0.44	-0.50	-0.44	-0.46
11_31	1.82	1.92	2.06	1.75	2.08	1.96	1.93
11_32	0.99	1.03	1.12	0.93	1.12	1.06	1.04
11_39	-0.52	-0.52	-0.51	-0.48	-0.55	-0.50	-0.51
11_42	-0.10	-0.08	-0.04	-0.07	-0.07	-0.04	-0.07
11_44	0.15	0.17	0.23	0.16	0.21	0.21	0.19
11_52	-0.90	-0.91	-0.94	-0.86	-1.00	-0.91	-0.92
11_54	0.31	0.33	0.39	0.30	0.37	0.37	0.35
11_56	-0.85	-0.87	-0.88	-0.81	-0.94	-0.85	-0.87
11_58	-1.26	-1.32	-1.35	-1.21	-1.41	-1.29	-1.31
11_61	-0.68	-0.69	-0.70	-0.64	-0.75	-0.67	-0.69
11_63	-0.80	-0.82	-0.83	-0.76	-0.88	-0.80	-0.81
11_64	-0.04	-0.02	0.01	-0.03	-0.02	0.00	-0.02

6.4 テスト・項目の経年比較

等化された全ての項目について、2010年度だけで出題された項目、2011年度だけで出題された項目、2010年度と2011年度テストに共通に出題された項目に分けて、項目困難度と項目識別力との散布図を示したものが図 6.4.1 になる。

図 6.4.1 から、2010年度にだけ出題された項目は、困難度が低いものであることが分かる。これは、2011年度のテスト作成において、2010年度で易しかった項目を共通項目には入れなかったことを適切に反映している。

また、2010年度出題項目と2011年度出題項目をそれぞれ1つのテストとした場合のテスト情報曲線を図 6.4.2 に示す。2011年度の項目出題において、2010年度よりも難しい項目を多く出題したため、曲線が右よりになっている。これは学力のより高いレベルについて精度の良い測定ができていることをしめすものである。ただし、この曲線は各年度で出題された項目全てを1つのテストとして見なししており、分冊ごとの情報を示しているわけではないことに注意されたい。

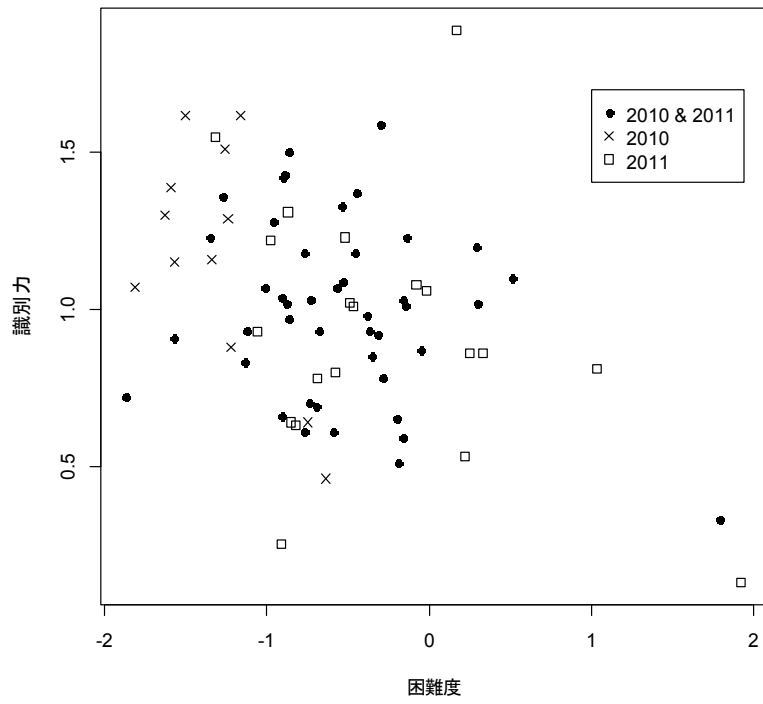


図 6.4.1 等化後項目の散布図

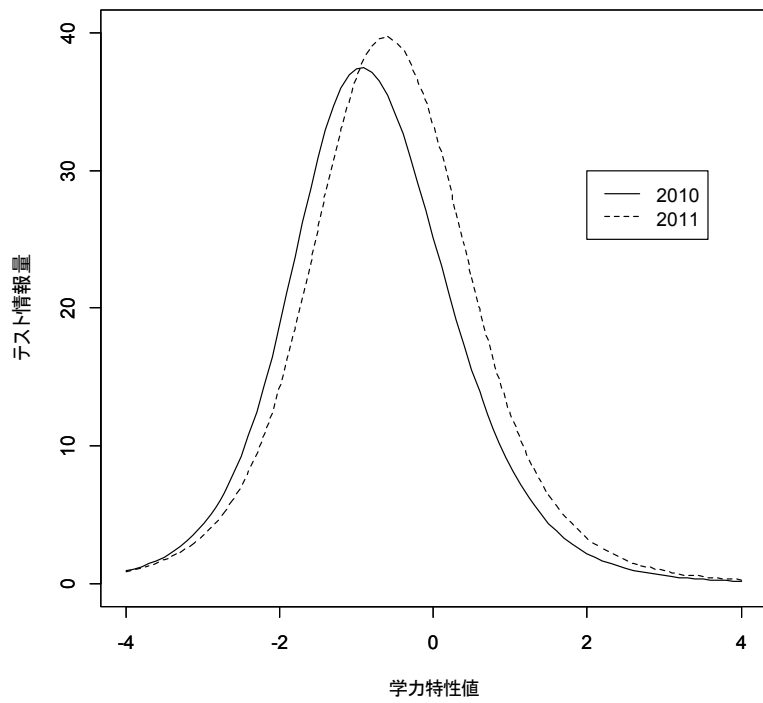


図 6.4.2 テスト情報曲線

6.5 学力特性値の年度間比較

ここでは、等化後の受検者の学力特性値の比較を行う。従来、IRTによる等化が行なわれない状況では、テストの得点は、各問題項目に正答した数で表現されてきた（配点による重み付けを行うこともある）。はじめに、2010年度及び2011年度のテスト得点を（配点をすべて1として）算出し比較してみる。分冊ごとに項目数が異なるため、テスト得点を受検した項目数で割る、いわゆる「正答率」で表現する。年度ごとの各受検者の正答率について、ヒストグラムで表示したものが図6.5.1である。通常ヒストグラムは棒グラフで表現されるが、年度間の比較を容易にするために、それを推定密度で表現したものを表示している（棒グラフを線グラフで置き換えたものとして理解されたい）。また箱ひげ図を図6.5.2に、年度ごとの正答率の平均値及び標準偏差を表6.5.1に示す。

ヒストグラムからも分かるとおり、2010年度に比べて2011年度の受検者の正答率は大きく左に寄っており、正答率の平均値でもおよそ7.5%も低下していることが分かる。箱ひげ図からも、両集団のずれは明確であり、通常であれば、これは非常に大きく学力が低下したととらえられるような差である。

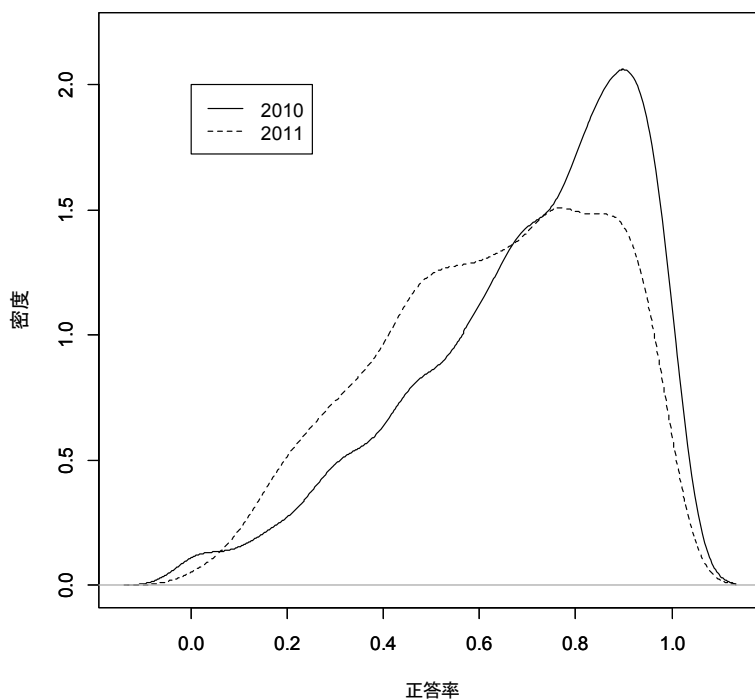


図 6.5.1 正答率のヒストグラム（推定密度関数）

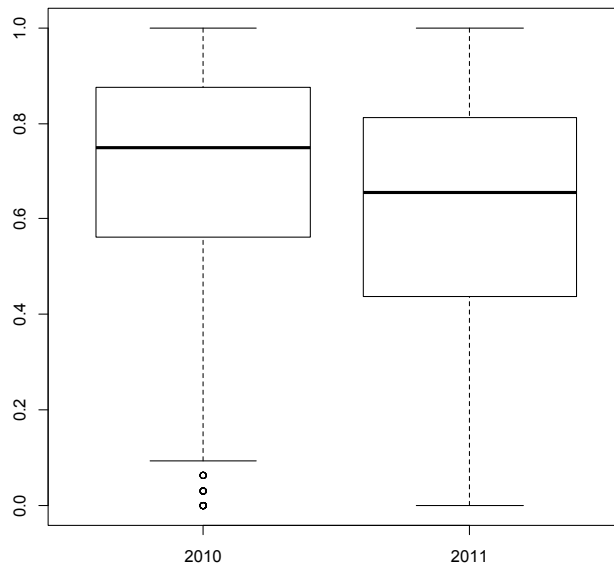


図 6.5.2 正答率の箱ひげ図

表 6.5.1 正答率の平均・標準偏差

	2010 年度	2011 年度
平均	0.696	0.624
標準偏差	0.234	0.235

次に、IRT による等化を行なったもとの学力特性値の比較を試みる。図 6.5.1, 6.5.2 と同様に、各年度の学力特性値のヒストグラムと箱ひげ図を示したのが図 6.5.3, 6.5.4 である。学力特性値の推定には最尤推定法 (MLE) を採用し、全問正答・誤答の受検者は計算から除外している。また、学力特性値の平均・標準偏差を表 6.5.2 に示す。

図 6.5.3, 6.5.4, 表 6.5.2 から分かる通り、年度間での受検者の学力特性値には大きな差は見られない。これは、先に見た「正答率」によるものと大きく異なる結果である。ここに「正答率」では見ることができなかった、等化を行うことによる学力比較の特徴が端的に表れている。

なぜ「正答率」では学力差があるように見えたのであろうか。これには、図 6.4.1 の項目困難度が大きく関係している。図 6.4.1 より、2010 年度のテストには困難度が低い、つまり易しい項目が多く含まれていることが分かった。仮に学力が等しい受検者であっても、易しい項目が多いテストでは正答率は高くなり、難しい項目が多いテストでは正答率が低くなる。図 6.5.1 で示された正答率の差は学力の差そのものではなく、項目困難度が反映してしまっているのである。

これに対して、図 6.5.3 で示された学力特性値は、IRT 分析による等化が行なわれていることから、このような項目困難度の差が学力特性値に反映することがなく、直接にその数値を比較することが可能となる。ここに、学力の経年比較を行う際の、IRT 分析及び等化の重要性が表れているといえよう。

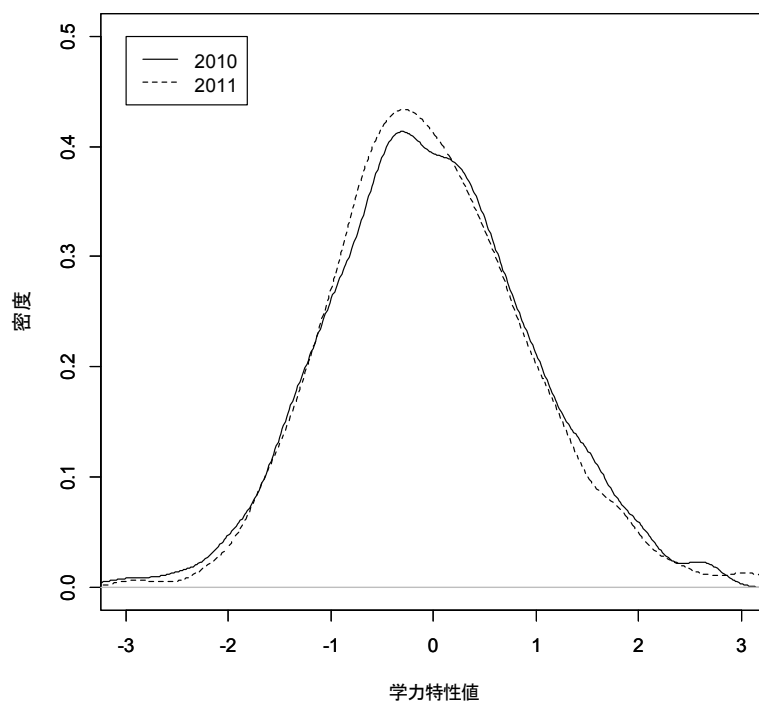


図 6.5.3 学力特性値 (MLE) のヒストグラム (推定密度関数)

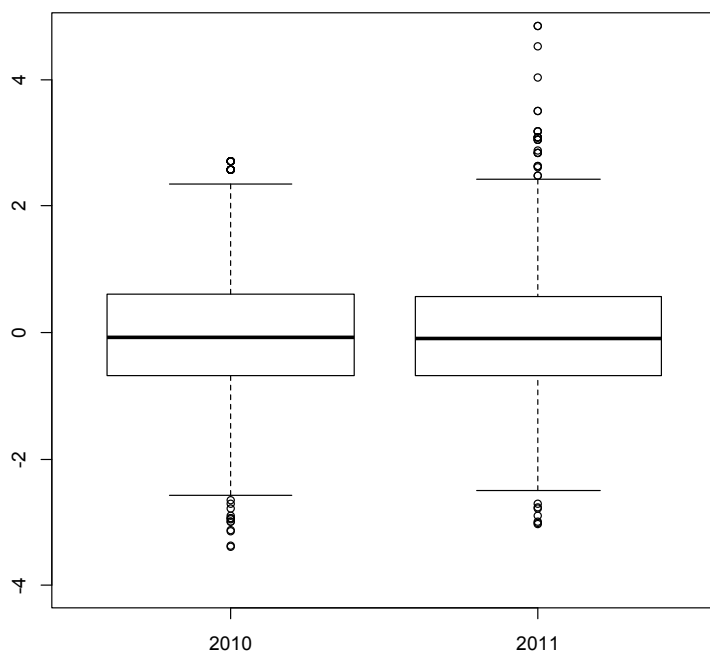


図 6.5.4 学力特性値 (MLE) の箱ひげ図

表 6.5.2 学力特性値 (MLE) の平均・標準偏差

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

最後に、学力特性値の推定方法を EAP (expected a posterior) 法で行なったものの比較を、図 6.5.5, 6.5.6, 表 6.5.5 に示す。EAP 法では、学力特性値の事前分布を仮定することにより、全問正答・誤答の受検者に対して、推定値を割り当てることができる。EAP 法による結果は、MLE 法による結果とほぼ同様のものとなっていることが分かる。

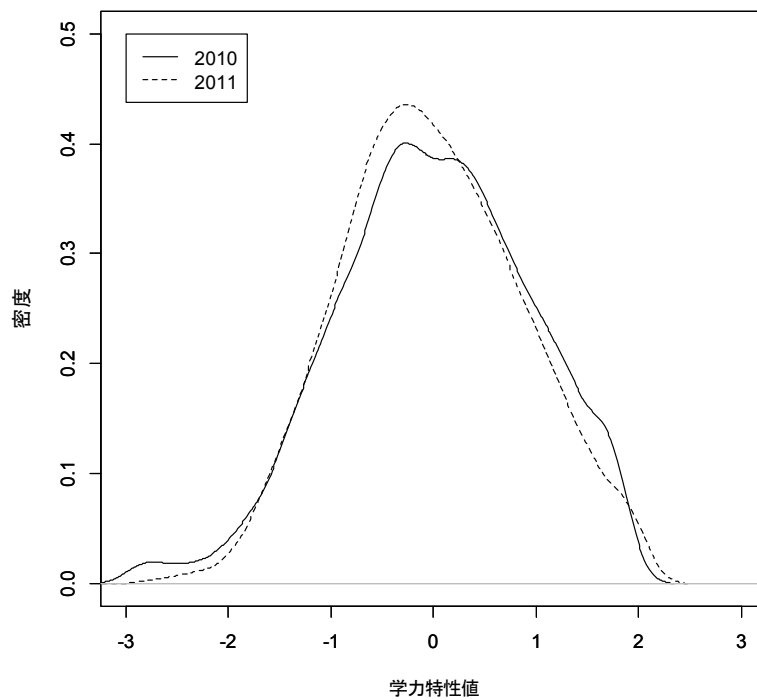


図 6.5.5 学力特性値 (EAP) のヒストグラム (推定密度関数)

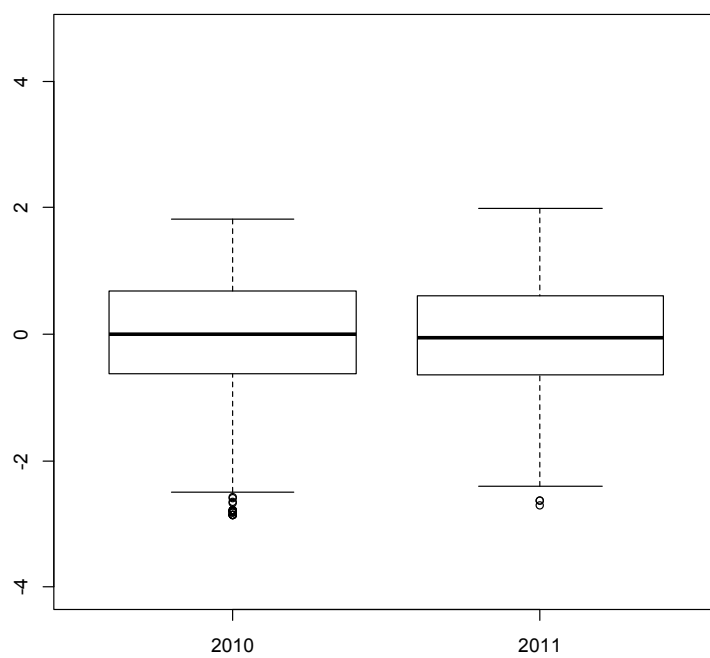


図 6.5.6 学力特性値 (EAP) の箱ひげ図

表 6.5.5 学力特性値 (EAP) の平均・標準偏差

	2010 年度	2011 年度
平均	-0.000	-0.019
標準偏差	0.945	0.945