はしがき

この報告は，平成 27 年度「学力調査を活用した専門的な課題分析に関する調査研究」の「A．全国学力・学習状況調査における経年変化分析調査の年度間等化に関する調査研究」に応募し，技術審査会等を経て採用された調査研究の成果をまとめたものである。

この調査研究では，平成 25 年度に実施され，また 28 年度にも実施予定の「経年変化分析調査」に関して，以下の点を目的とした。

（1）平成 25 年度と平成 28 年度の経年変化分析調査データの得点を比較可能にする等化分析について，どのような手法を用いるのが良いのかをシミュレーション・データに基づき検討する。
（2）全国学力・学習状況調査の全数データと経年変化分析調査データとを結びづける，「対応づけ」分析を行う。
（3）共通項目や共通受検者が存在するテストデザインを実施・運営していく状況において，想定される様々な課題について，他機関への訪問調査により検討する。
（4）平成 25 年度全国学力・学習状況調査データに対して，名義反応モデルや多次元モデルといった応用的 IRT 分析の実施可能性を検討する。

我が国の教育測定場面において，IRT や等化分析に関しての注目が高くなり，そのための基礎研究の重要性が増している。本報告書が，それらの研究の一助となれば，研究代表者として幸いである。

研究代表者　　熊谷　龍一

事業概要

事 業 名	学力調査を活用した専門的な課題分析に関する調査研究
事業内容	Ａ．全国学力・学習状況調査における経年変化分析調査の年度間等化に関する調査研究
委託期間	平成 27 年 7 月 13 日から平成 28 年 3 月 31 日
事業者名	国立大学法人東北大学大学院教育学研究科長　　高橋　満
事 業 費	5,463 千円

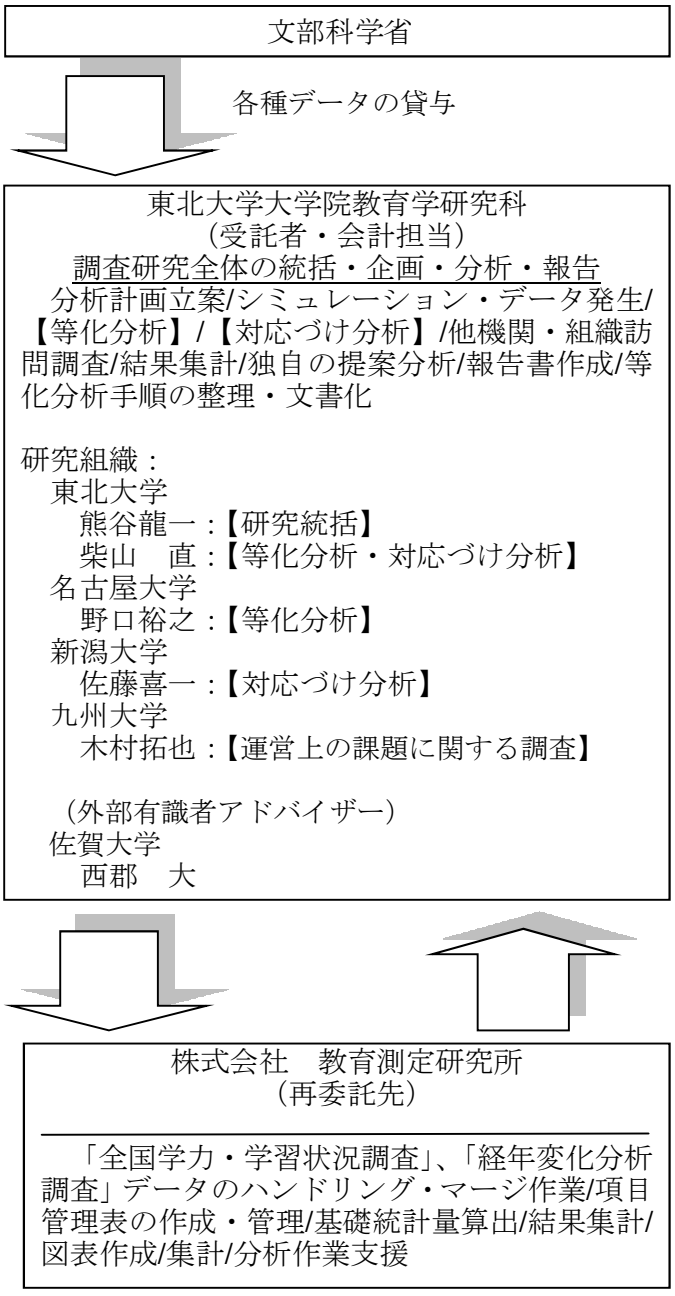
研究組織

研究代表	熊谷　龍一	東北大学大学院教育学研究科（全体統括，第 2, 5 章）
研究協力	野口　裕之	名古屋大学大学院教育発達科学研究科（第 2 章）
	柴山　直	東北大学大学院教育学研究科（第 3 章，付録）
	佐藤　喜一	新潟大学・教育・学生支援機構（第 3 章）
	木村　拓也	九州大学基幹教育院（第 4 章）
共同研究	西郡　大	佐賀大学　アドミッションセンター（第 4 章）
研究助手	新川　壮光	東北大学大学院教育学研究科
資料整理	阿部　竜士	東北大学教育学部
	加藤　周平	東北大学教育学部
	米林　巽	東北大学教育学部
	澁谷　拓巳	東北大学教育学部
事務担当	紙屋　雅子	

実施集計　　株式会社　教育測定研究所

（所属は平成 28 年 3 月 31 日現在）

実施体制


文部科学省
各種データの貸与
東北大学大学院教育学研究科
（受託者・会計担当）
調査研究全体の統括・企画・分析・報告
分析計画立案/シミュレーション・データ発生/【等化分析】/【対応づけ分析】/他機関・組織訪問調査/結果集計/独自の提案分析/報告書作成/等化分析手順の整理・文書化
研究組織：
東北大学
熊谷龍一：【研究統括】
柴山　直：【等化分析・対応づけ分析】
名古屋大学
野口裕之：【等化分析】
新潟大学
佐藤喜一：【対応づけ分析】
九州大学
木村拓也：【運営上の課題に関する調査】
（外部有識者アドバイザー）
佐賀大学
西郡　大
株式会社　教育測定研究所
（再委託先）
「全国学力・学習状況調査」、「経年変化分析調査」データのハンドリング・マージ作業/項目管理表の作成・管理/基礎統計量算出/結果集計/図表作成/集計/分析作業支援

実施日程

平成 27 年	7 月 13 日	事業開始
	8 月 12 日	第 1 回全体研究会
	9 月 14 日～15 日	第 1 回経年比較部会
	10 月 4 日	第 2 回全体研究会
	11 月 13 日～15 日	第 2 回経年比較部会
	11 月 29 日 ～12 月 1 日	第 3 回経年比較部会
	12 月 23 日	測定部門合同会議
平成 28 年	1 月 10 日～13 日	Educational Testing Service（米国）訪問
	1 月 31 日	第 3 回全体研究会
	2 月 13 日	第 4 回経年比較部会
	2 月 20 日	第 4 回全体研究会
	3 月 17 日	第 5 回全体研究会
	3 月 31 日	事業終了

※　各部会内での分析，報告書執筆などは随時作業が行われた。

目　次

1.　調査研究の概要

1.1.　問題と目的

全国学力・学習状況調査における経年変化分析調査の第一回調査が，平成 25 年度に実施された。平成 28 年度に実施が予定されている第二回の経年変化分析調査で得られるデータとあわせて，平成 25 年度から 28 年度にかけての学力の経年変化を調べることができる。ただし，この経年変化分析調査においては，平成 25 年度のテスト項目（テスト問題のことで，以後「項目」と呼ぶ）と平成 28 年度の項目は，一部が重複しているものの，それぞれに独自の項目によってテストが構成されている。含まれている項目が異なる複数のテストについて，それぞれ独立した受検者集団に実施した場合，そのままではテスト得点を相互に比較することはできない。仮にテストＡの平均得点がテストＢのそれよりも高かったとしても，それはテストＡの受検者集団の能力が高かったのか，若しくはテストＡが易しいものだったのかの判別ができないからである。

このように中身が異なる複数のテストの得点を比較する手続は「リンキング（linking）」と呼ばれ，さらにその目的やテストの性質により，「予測（predicting）」，「尺度調整（scale aligning）」，「等化（test equating）」に分類される（Holland & Dorans, 2006）。平成 25 年度と 28 年度の経年変化分析調査においては，この中の「等化」に当たる手続が採用される。

この等化に関しても，テストのデザイン（テストにおいて項目がどのように配置されているかや，複数のテストをどの受検者にどのように実施するかなどの構造）により，様々な分析手法が提案されており，唯一絶対の方法が存在するわけではない。等化の手続を決定するためには，数値基準のみならず，テストデザインを良く吟味し，継続性なども考慮しながら決定する必要がある（熊谷・荘島，2015）。

本研究では，平成 25 年度と 28 年度の経年変化分析調査において，最適な方法を検討することを主目的とする。さらには，このような経年変化調査を運営・実施する場合における様々な課題について検討を行う。

1.2.　テストデザイン

平成 25 年度，平成 28 年度における全国学力・学習状況調査及び経年変化分析調査のテストデザインを図 1.1 に示す。

テストデザインの概略は次の通りである。

- 平成 25 年度，28 年度ともに，全国学力・学習状況調査及び経年比較調査によるテストデータが存在する。
- 全国学力・学習状況調査は悉皆（しっかい）調査（以下，全数データと呼ぶ），経年変化分析調査データは，その中からの抽出調査である。両調査を受検した者を「共通受検者」と呼ぶ。
- 経年変化分析調査テストにおいては，平成 25 年度と 28 年度の間に同一の項目（以下，共通項目と呼ぶ）がある。

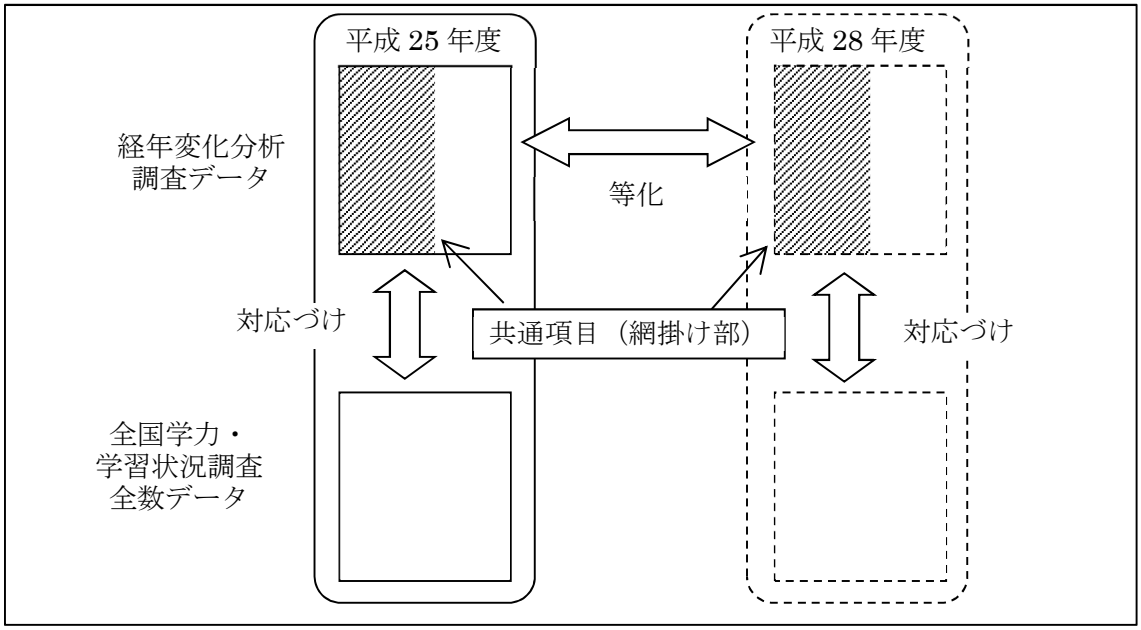


図 1.1　テストデザイン

1.3.　分析概要と本報告書の構成

本章に続く第 2 章では，平成 25 年度と平成 28 年度の経年変化分析調査データの得点を比較可能にする等化について，どのような手法を用いるのが良いのかを検討することが目的となる。等化においては，項目反応理論（item response theory，以下 IRT と呼ぶ）と呼ばれるテスト理論による方法を採用することとする。分析手法としては，調査研究時点では平成 28 年度の経年変化分析調査データについては未実施で入手できないため，様々な状況を想定したシミュレーション・データによる検討を行うこととする。

第 3 章では，全国学力・学習状況調査の全数データと経年変化分析調査データとを結びづける，「対応づけ」分析を行う。これには既にデータが入手されている平成 25 年度の実データを用いて分析を行う。

第 4 章では，共通項目や共通受検者が存在するテストデザインを実施・運営していく状況において，想定される様々な課題について，他機関を訪問調査し，その結果をまとめる。

第 5 章では，平成 25 年度全国学力・学習状況調査データに対して，名義反応モデルや多次元モデルといった応用的 IRT 分析の実施可能性を検討する。

また付録として，等化を行うテストデザインにおいて非常に多数の項目を扱うことが必須となるが，その際にそれらの管理に必要となる「項目管理ツール」について，本研究で構築した試作版を元に，その概要について述べる。

文献

Holland, P. W., & Dorans, N. J. (2006) Linking and Equating. In R. L. Brennan (Ed.), *Educational Measurement*. 4th ed. Westport, CT: American Council on Education

and Praeger Publishers. pp. 187-220.
熊谷龍一・荘島宏二郎　(2015)　教育心理学のための統計学 ーテストでココロをはかるー　誠信書房.

2. 等化分析

2.1. はじめに

2.1.1. テストデザイン

本章では，平成 25 年度及び 28 年度実施の経年変化分析調査データについての等化分析について，どのような手法を採用するのかについて検討することが目的となる。経年変化分析調査データの概要は第 1 章の図 1.1 に示したとおりであるが，経年変化分析調査データ部分について，より詳細に示したものが図 2.1 である。

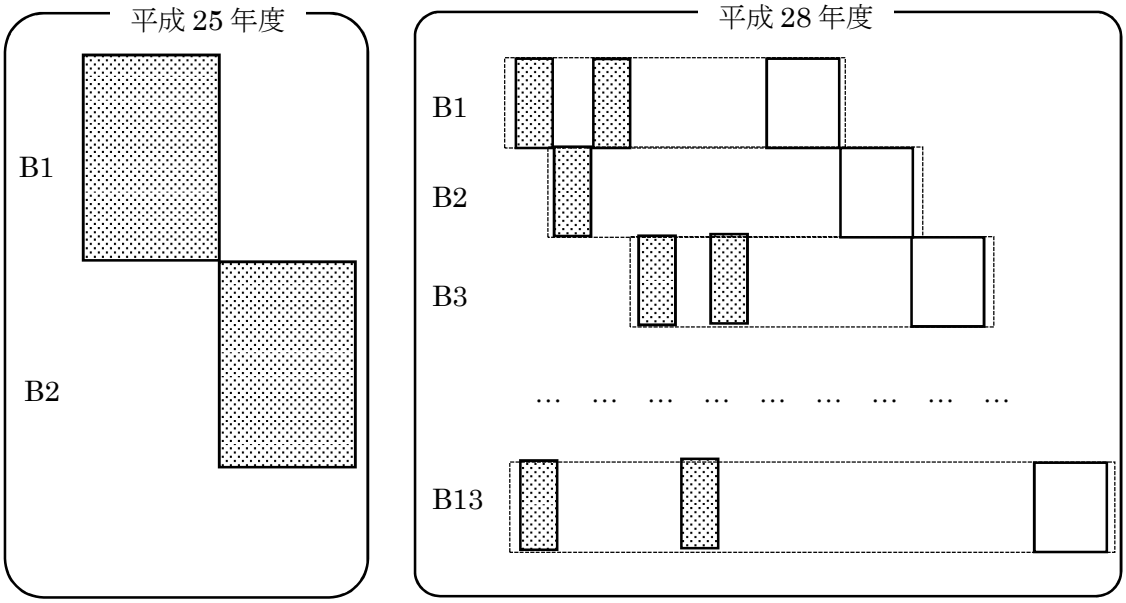


図 2.1　経年変化分析調査データ詳細デザイン

図 2.1 で，平成 25 年度は 2 つの版（冊子）があり，それぞれ B1，B2 と区別する。B1 と B2 の間には共通項目が存在しないため，矩形に重なりがない形で表現されている。平成 28 年度は，B1 から B13 までの 13 版が存在している。平成 28 年度には，平成 25 年度で出題された項目が再出題されており，図では網掛けで表現されている。これを「共通項目」と呼ぶ。また，共通項目は B1 から B13 の間でも共通に出題されているものがある。さらに平成 28 年度で新規に出題された項目があり，図では白抜きの矩形で表現されている。これを新規項目と呼ぶ。

2.1.2. 等化分析

図 2.1 で示したテストにおいて，平成 25 年度のテスト得点と平成 28 年度のテスト得点については，テストに含まれている項目が（一部）異なるため，相互に得点を比較することはできない。このように，異なる版のテスト得点を相互に比較可能にする手続が，「等化

分析」と呼ばれるものである。等化分析にも様々な手続があるが，本調査研究では，項目反応理論（IRT）を利用した分析を行うこととする。また IRT に基づいた等化分析といっても，その手法が様々に存在する（各手法については後述する）。本章での研究目的は，幾つかある等化分析の中で，どの方法を用いるのが良いのかを検討することである。

2.2. 分析方針

本調査分析時には，平成 25 年度のデータは存在するものの，平成 28 年度データは未実施のため存在しない。そこで本研究では，平成 28 年度データ部分について，様々な設定のもとでコンピュータ・シミュレーションによりデータ生成を行うことによる，シミュレーション研究を行うこととする。

2.3. 分析教科の決定

平成 25 年度経年変化分析調査は，小学校 6 年生の国語・算数（以下，小学校国語，小学校算数）及び中学校 3 年生の国語・数学（以下，中学校国語，中学校数学）の 4 教科が実施された。本研究で，この 4 教科全てを分析対象とするかどうかを決定するために，平成 25 年度データを IRT 分析し，各教科の特徴を検討することとした。

2.3.1. 1 次元性の確認

IRT 分析を行うためには，そのテストが測定しようとしている能力が 1 次元であるという仮定が必要となる（この仮定を必要としない多次元 IRT モデルも存在する）。そこでこの仮定が成り立っているかを調べるために，各教科及び冊子ごとに，項目間テトラコリック相関係数行列を用いたスクリープロットを描いた（図 2.2）。同一教科においても冊子間で共通項目が存在しないため，相関係数を算出することができないことから冊子ごとに分析を行っている。また，各教科・冊子ごとに項目数が異なることから，項目数に対する固有値の比率を用いている。

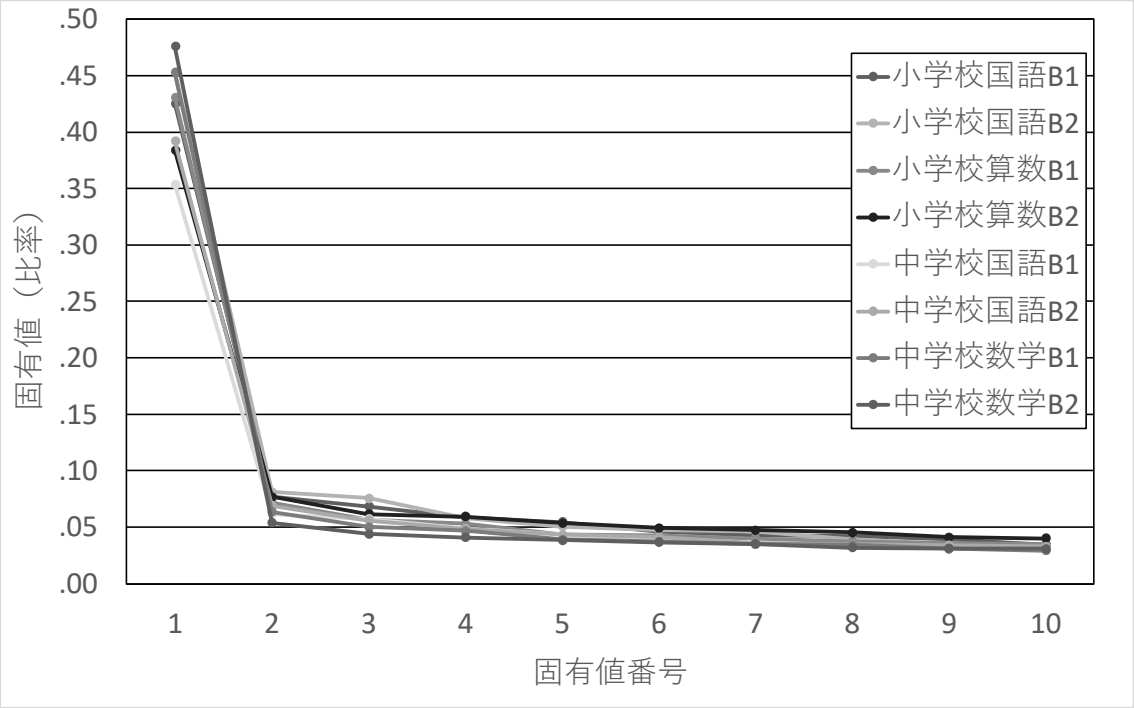


図 2.2　教科・冊子ごとのスクリープロット

図 2.2 より，どの教科・冊子においても，第 1 固有値の値が大きく，第 2 固有値以降の減少率が非常に小さくなっているため，1 次元性の仮定が成り立っていることが示唆された。

2.3.2.　単独 IRT 分析

先の分析で，どの教科においても 1 次元性の仮定が成り立っていることが示唆されたため，各教科のデータセットごとに単独での IRT 分析を行い，項目母数を比較することとした。利用した IRT モデルは，2 パラメタ・ロジスティック・モデルとした。

図 2.3 は，教科ごとの項目母数の散布図である。横軸は困難度母数，縦軸は識別力母数である。散布図から，小学校国語，小学校算数，中学校国語については教科ごとの差異がそれほど見られなかった。中学校数学においては，困難度が－1～＋1 の範囲において，他の教科よりも識別力が高いことが見てとれた。

以上の分析から，本研究においては，「小学校国語」，「中学校数学」の 2 教科について分析の対象とすることにした。

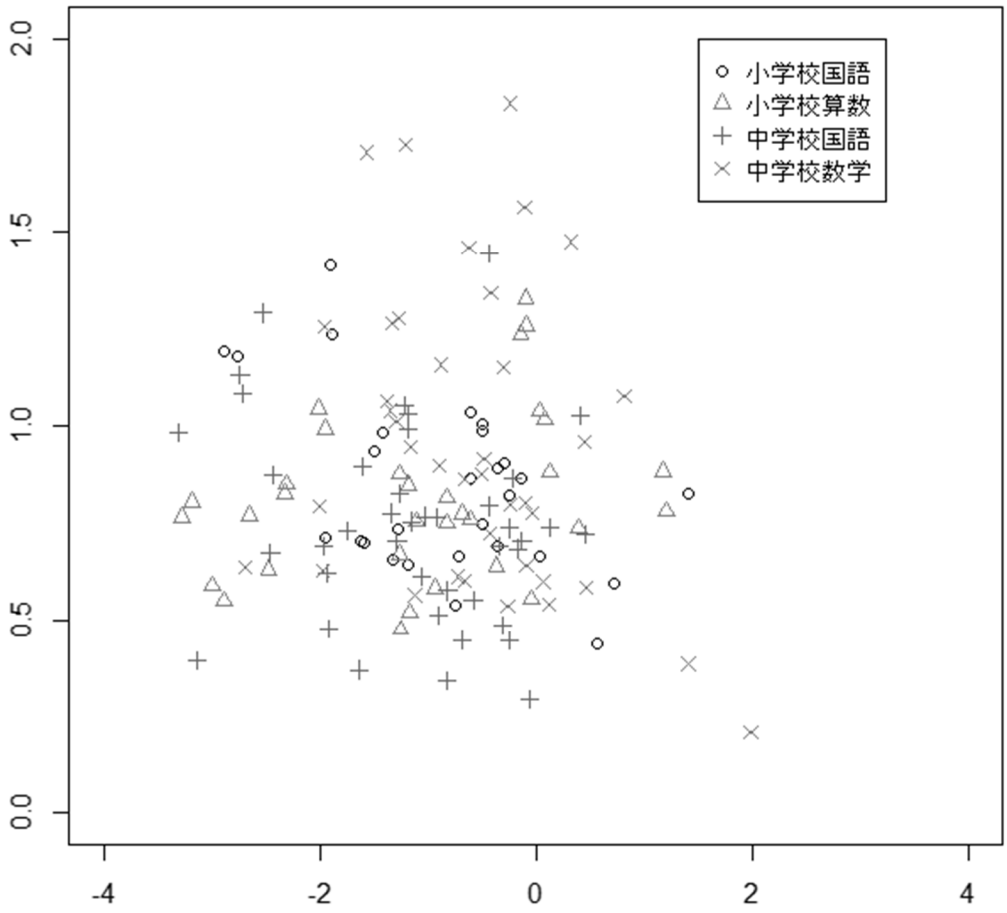


図 2.3　平成 25 年度データの項目母数比較

2.4.　等化手法について

本研究で利用する等化手法について述べる。ただし，各手法の具体的な計算方法などの詳細は，各文献等を参照されたい。

2.4.1.　共通項目デザインによる方法

図 2.1 にあるとおり，本調査のテストデザインは，平成 25 年度と 28 年度に共通な項目が含まれる共通項目デザインとなっている。共通項目デザインによる方法は，共通項目部分の項目母数を用いて，等化係数を推定する方法となる。本研究では，以下の 3 つの手法を扱うこととした。

・Mean & Sigma 法（Marco, 1977）
・Mean & Mean 法（Loyd & Hoover, 1980）
　なお項目識別力の平均には算術平均ではなく，幾何平均を採用する方法（Mislevy, & Bock, 1990）を採用した。
・Haebara 法（Haebara, 1980）

なお，Haebara 法と同系列の手法として Stocking & Lord（1983）による方法も存在するが，本研究では扱わなかった。

2.4.2. 共通受検者デザインによる方法

前項でも述べたとおり本調査のテストデザインは共通項目デザインであるが，分析の比較対象として，共通受検者デザインを応用したものについても分析対象とした。図 2.1 のテストデザインにおいて共通受検者デザインを応用した手続を以下に示す。

① 平成 28 年度データのみを用いて IRT 分析を行い，潜在特性尺度値 θ_{H28} を推定する。
② 平成 28 年度データの共通項目部分（図 2.1 の網掛け部）について，平成 25 年度データから推定された項目母数を用いて，潜在特性尺度値 θ_{H25} を推定する。
③ θ_{H28} 及び θ_{H25} を用いて，Mean & Sigma 法（Marco, 1977）を用いて等化係数を推定する。

なお，上述の潜在特性尺度値 θ の代わりに母集団分布推定値 $\widehat{g(\theta)}$ を用いる熊谷・野口（2012）の方法も分析対象とした。

2.4.3. 項目固定法

平成 28 年度のデータを用いて項目母数を推定する際に，平成 25 年度との共通項目部においては，平成 25 年度データから推定された項目母数に値を固定して等化を行う方法が項目固定法である。このとき，等化係数の推定は行わないが，他の方法との比較のために，推定母集団分布の平均値及び標準偏差を用いて，等化係数の代わりとした。

本研究で採用した等化手法を表 2.1 にまとめた。本研究で扱う等化手法は以下の 6 種である。

表 2.1　本研究で採用した等化手法一覧

デザイン	等化手法
共通項目デザイン	Mean & Sigma 法
	Mean & Mean 法
	Haebara 法
共通受検者デザイン	Mean & Sigma 法
	熊谷・野口の方法
項目固定	項目固定法

2.5.　シミュレーション分析 1 ：データ生成時に誤差を混入させない場合

本研究でのシミュレーション分析の方針は，平成 28 年度経年変化分析調査データ部分について，様々な設定値のもとでシミュレーションデータを生成し，平成 25 年度データ（こちらは実データである）とともに等化分析を行い，等化係数を推定することとする。この手続を各条件（設定値）のもとで 100 回繰り返し，等化係数の平均や標準偏差，RMSE（root mean square error；平均二乗誤差の平方根）などを計算し，各手法の比較を行う。

2.5.1.　シミュレーションデータの生成方法

平成 28 年度経年変化分析調査に関するシミュレーションデータは，以下の手続で生成した。

① 等化係数の真値の設定：等化係数 K 及び L について，次の 3 通りを設定値とした。（K, L）=（1.0, 0.0），（1.3, 0.6），（0.5, -0.6）。なおここでの等化係数 K 及び L は，潜在特性尺度値 θ について $\theta^* = K\theta + L$ と線形変換する各係数のことである。

② 平成 28 年度経年変化分析調査の想定受検者数に従って，潜在特性尺度値 θ_{H28} を乱数により生成した。このとき乱数には先の等化係数 K，L を用いた正規乱数 $\sim N(L, K^2)$ を用いた。

③ 平成 28 年度経年変化分析調査データの共通項目部について，IRT の 2 パラメタ・ロジスティック・モデルにおける項目母数（識別力母数，困難度母数）として，平成 25 年度データを IRT 分析して得られた項目母数をそのまま用いた。新規出題部分については，乱数により生成した。このとき，識別力母数については，$logX \sim N(-0.2, 0.2^2)$となるような対数正規乱数を用いた。困難度母数については，文部科学省が事前に想定している「想定正答率」を利用して，「想定正答率」を標準正規分布の逆関数に代入し，その値に正規乱数$\sim N(0, 0.15^2)$ を足し合わせることで

生成した。

④ 前項の②，③で生成した潜在特性尺度値及び項目母数を用いて，2 パラメタ・ロジスティック・モデルに代入することで項目正答確率が計算できる。この確率と一様乱数を比較することで，「正答確率≧一様乱数」となった場合には正答，「正答確率＜一様乱数」となった場合は誤答として，シミュレーションデータを生成した。

2.5.2. 分析

前項の手続に従い，小学校国語・中学校数学それぞれに対して，等化係数の設定値が（K, L）=（1.0, 0.0），（1.3, 0.6），（0.5, -0.6）という合計 6 条件について，各 100 セットのシミュレーションデータを生成し，実際に等化分析を行った。等化分析によって得られた 100 セットの等化係数（項目固定法については，母集団分布推定値の平均値と標準偏差）について，平均，標準偏差，最大値，最小値，RMSE を計算した。

なお，IRT 分析については，熊谷（2009）による EasyEstimation を本調査用に改修したものを用いた。それ以外の計算に関しては，統計分析ソフトのRを利用した。

2.5.3. 結果

各条件における等化係数推定値に関する統計量を表 2.2，2.3，2.4，2.5 に示す。各表において，「M＆S（項）」は共通項目デザインにおける Mean & Sigma 法，「M＆M」は Mean & Mean 法，「Haebara」は Haebara 法を表す。また「M＆S（受）」は共通受検者デザインにおける Mean & Sigma 法，「KN」は熊谷・野口の方法，「Fixed Item」は項目固定法を表す。

表 2.2　小学校国語，等化係数 K（誤差混入なし）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	0.994	0.023	1.049	0.933	0.024	1.295	0.027	1.341	1.241	0.027	0.490	0.025	0.557	0.439	0.027
M & M	1.000	0.014	1.034	0.955	0.014	1.302	0.017	1.336	1.259	0.017	0.499	0.011	0.528	0.472	0.011
Haebara	0.999	0.013	1.025	0.956	0.013	1.300	0.015	1.337	1.257	0.015	0.503	0.010	0.525	0.480	0.010
M & S（受）	0.989	0.011	1.011	0.954	0.016	1.240	0.014	1.281	1.200	0.062	0.556	0.009	0.578	0.536	0.057
KN	1.025	0.017	1.055	0.977	0.031	1.228	0.017	1.274	1.183	0.074	0.505	0.009	0.528	0.482	0.010
Fixed Item	1.002	0.013	1.028	0.958	0.013	1.260	0.016	1.300	1.220	0.043	0.503	0.009	0.522	0.480	0.010

表 2.3　小学校国語，等化係数 L（誤差混入なし）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	-0.006	0.017	0.041	-0.060	0.018	0.596	0.027	0.653	0.523	0.028	-0.603	0.025	-0.554	-0.696	0.025
M & M	0.000	0.014	0.035	-0.038	0.014	0.604	0.019	0.643	0.555	0.019	-0.598	0.025	-0.539	-0.679	0.025
Haebara	-0.002	0.012	0.026	-0.039	0.012	0.600	0.018	0.637	0.559	0.018	-0.600	0.009	-0.581	-0.624	0.009
M & S（受）	-0.054	0.011	-0.029	-0.088	0.055	0.475	0.015	0.506	0.437	0.126	-0.585	0.009	-0.566	-0.612	0.017
KN	0.007	0.013	0.040	-0.032	0.015	0.582	0.018	0.627	0.534	0.026	-0.598	0.008	-0.578	-0.618	0.008
Fixed Item	-0.001	0.012	0.028	-0.039	0.012	0.583	0.017	0.616	0.541	0.024	-0.600	0.008	-0.582	-0.619	0.008

表 2.4　中学校数学，等化係数 K（誤差混入なし）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	0.998	0.016	1.039	0.961	0.016	1.297	0.015	1.331	1.251	0.016	0.498	0.017	0.534	0.468	0.017
M & M	1.000	0.008	1.019	0.982	0.008	1.301	0.011	1.325	1.271	0.011	0.499	0.005	0.508	0.484	0.005
Haebara	1.000	0.007	1.021	0.985	0.007	1.300	0.009	1.321	1.274	0.009	0.499	0.004	0.508	0.485	0.005
M & S（受）	1.071	0.006	1.086	1.053	0.071	1.280	0.009	1.303	1.257	0.022	0.596	0.004	0.604	0.587	0.096
KN	1.044	0.010	1.070	1.017	0.046	1.317	0.013	1.354	1.281	0.021	0.503	0.005	0.514	0.490	0.006
Fixed Item	1.002	0.007	1.021	0.984	0.008	1.282	0.008	1.303	1.254	0.020	0.500	0.004	0.509	0.490	0.004

表 2.5　中学校数学，等化係数 L（誤差混入なし）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	0.001	0.015	0.044	-0.033	0.014	0.597	0.015	0.634	0.545	0.016	-0.600	0.013	-0.573	-0.625	0.013
M & M	0.002	0.011	0.027	-0.020	0.011	0.601	0.011	0.629	0.565	0.011	-0.600	0.014	-0.572	-0.628	0.014
Haebara	0.001	0.007	0.018	-0.015	0.007	0.599	0.011	0.628	0.566	0.011	-0.600	0.004	-0.589	-0.610	0.004
M & S（受）	0.041	0.007	0.061	0.025	0.041	0.574	0.009	0.596	0.543	0.028	-0.561	0.005	-0.551	-0.572	0.040
KN	0.018	0.008	0.034	0.003	0.019	0.619	0.012	0.651	0.589	0.022	-0.598	0.004	-0.588	-0.607	0.004
Fixed Item	0.001	0.007	0.020	-0.015	0.007	0.593	0.010	0.620	0.563	0.013	-0.600	0.004	-0.591	-0.611	0.004

2.5.4. 考察

表 2.2，2.3，2.4，2.5 より，はじめに平均値（Mean）を見ると，共通項目デザインの 3 種の方法がほぼ設定値通りの値になっている。この数値が設定値から離れると，推定値に偏り（バイアス）が生じていることを表すが，幾つかの設定値において「M & S（受）」が相対的に設定値からずれていることが分かる。

また表の（RMSE）の値は推定値が真値からどの程度散らばっているかを示し，数値が小さい方が散らばりが小さい，すなわち真値に近い推定値が得られていることを示す。RMSE については，条件ごとに様々であるが，「Haebara」が全体的に小さな値を示している。

2.6. シミュレーション分析２：データ生成時に誤差を混入させた場合

2.6.1. 共通項目の母数における誤差の混入

前節のシミュレーションデータ生成手続③において「平成 28 年度経年変化分析調査データの共通項目部について，IRT の 2 パラメタ・ロジスティック・モデルにおける項目母数（識別力母数，困難度母数）として，平成 25 年度データを IRT 分析して得られた項目母数をそのまま用いた」とした。これは，平成 25 年度から 28 年度にかけて，問題項目の特徴が全く変わらずにいることを想定している。しかしながら，現実のテスト場面においては，実施時期，受検者集団，その他の周辺状況が異なる場合，必ずしも問題項目の特徴，すなわち項目母数が不変であることは保証できず，多少の程度はあれ，項目母数に変動が生じることもある。

そこで本分析においては，平成 25 年度データを IRT 分析して得られた項目母数をそのまま用いるのではなく，識別力母数に関しては正規乱数~$N(0, 0.2^2)$ を足し合わせ，困難度母数に関しては正規乱数~$N(0, 0.4^2)$ を足し合わせることで，項目の特徴が変化した状況をシミュレートした。またこれにあわせ，新規出題部分の識別力母数については，$logX$~$N(-0.3, 0.3^2)$となるような対数正規乱数に変更を行った。

これ以外の分析手続に関しては前節と同様である。

2.6.2. 結果

各条件における等化係数推定値に関する統計量を表 2.6，2.7，2.8，2.9 に示す。

表 2.6　小学校国語，等化係数 K（誤差混入あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	0.953	0.078	1.174	0.793	0.091	1.212	0.097	1.462	0.891	0.131	0.459	0.037	0.585	0.382	0.055
M & M	0.971	0.057	1.107	0.831	0.064	1.267	0.070	1.479	1.098	0.077	0.481	0.028	0.550	0.429	0.033
Haebara	0.960	0.059	1.094	0.819	0.071	1.239	0.076	1.464	1.062	0.098	0.467	0.036	0.557	0.397	0.048
M & S（受）	0.948	0.048	1.054	0.832	0.070	1.162	0.058	1.339	1.026	0.149	0.539	0.025	0.601	0.487	0.046
KN	0.989	0.066	1.177	0.838	0.067	1.204	0.068	1.372	1.062	0.117	0.487	0.048	0.602	0.367	0.049
Fixed Item	0.969	0.058	1.098	0.839	0.066	1.211	0.066	1.395	1.090	0.110	0.491	0.051	0.638	0.381	0.052

表 2.7　小学校国語，等化係数 L（誤差混入あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	-0.061	0.098	0.185	-0.245	0.115	0.507	0.137	0.900	0.159	0.165	-0.631	0.086	-0.450	-0.844	0.091
M & M	-0.046	0.090	0.249	-0.259	0.101	0.570	0.114	0.857	0.310	0.117	-0.619	0.090	-0.413	-0.846	0.091
Haebara	-0.070	0.094	0.152	-0.284	0.116	0.518	0.114	0.839	0.260	0.140	-0.630	0.093	-0.401	-0.853	0.097
M & S（受）	-0.126	0.082	0.096	-0.336	0.150	0.375	0.088	0.590	0.147	0.241	-0.617	0.089	-0.429	-0.835	0.091
KN	-0.062	0.089	0.153	-0.281	0.108	0.509	0.102	0.734	0.257	0.136	-0.625	0.093	-0.446	-0.853	0.096
Fixed Item	-0.074	0.089	0.169	-0.297	0.115	0.496	0.100	0.730	0.272	0.144	-0.631	0.095	-0.431	-0.854	0.099

表 2.8　中学校数学，等化係数 K（誤差混入あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	0.928	0.064	1.219	0.817	0.097	1.205	0.075	1.404	1.049	0.121	0.466	0.034	0.574	0.386	0.048
M & M	0.983	0.040	1.085	0.896	0.043	1.291	0.053	1.439	1.159	0.054	0.494	0.023	0.554	0.441	0.023
Haebara	0.925	0.068	1.174	0.752	0.101	1.208	0.074	1.410	1.021	0.118	0.462	0.033	0.545	0.380	0.050
M & S（受）	0.928	0.064	1.219	0.817	0.097	1.225	0.048	1.388	1.094	0.088	0.573	0.024	0.632	0.513	0.077
KN	0.925	0.068	1.174	0.751	0.101	1.239	0.070	1.409	1.020	0.093	0.479	0.040	0.573	0.377	0.045
Fixed Item	0.939	0.057	1.114	0.802	0.083	1.205	0.059	1.404	1.049	0.111	0.480	0.038	0.575	0.383	0.043

表 2.9　中学校数学，等化係数 L（誤差混入あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	-0.039	0.077	0.187	-0.187	0.086	0.502	0.095	0.761	0.286	0.136	-0.595	0.071	-0.415	-0.755	0.071
M & M	-0.007	0.079	0.204	-0.171	0.078	0.579	0.089	0.799	0.395	0.091	-0.596	0.075	-0.403	-0.748	0.075
Haebara	-0.060	0.081	0.152	-0.246	0.100	0.490	0.101	0.748	0.245	0.149	-0.614	0.083	-0.383	-0.826	0.084
M & S（受）	-0.014	0.074	0.176	-0.165	0.075	0.487	0.080	0.697	0.277	0.138	-0.576	0.092	-0.326	-0.786	0.095
KN	-0.044	0.077	0.162	-0.189	0.089	0.511	0.090	0.758	0.298	0.127	-0.612	0.089	-0.379	-0.819	0.090
Fixed Item	-0.057	0.074	0.122	-0.211	0.093	0.488	0.086	0.731	0.272	0.141	-0.616	0.089	-0.382	-0.823	0.090

2.6.3. 考察

前節の結果と異なり，誤差が混入することで推定値の平均（Mean）が，どの条件においても設定値からのズレが大きくなった。それにより真値からの散らばり具合を示す（RMSE）の値も前節より大きくなった。前節では，Haebara が相対的に（RMSE）の値が小さくなっていたが，本調査においては条件ごとに大小するようになり，相対的には M & M が小さな数値を示していた。

2.7. シミュレーション分析３：識別力母数に外れ値がある場合

2.7.1. 識別力母数の外れ値

前節の分析では，相対的に Mean & Mean 法の RMSE が小さくなることが分かった。しかしながら，例えば村木（2011）では，識別力母数推定の不安定さから Mean & Mean 法よりも Mean & Sigma 法がより一般的に利用されていることを述べている。

図 2.4 は，Kolen & Brennan（2004）で示された，識別力母数の外れ値の様子である。本来であれば二つのテスト間で一直線上に並ぶはずのものが，item27 のようにそこから逸脱してしまう状況であり，特に識別力母数でこれが生じやすいといわれている。

そこで，本調査では，このような識別力母数に外れ値がある状況をシミュレートして，等化係数への影響を調べる。

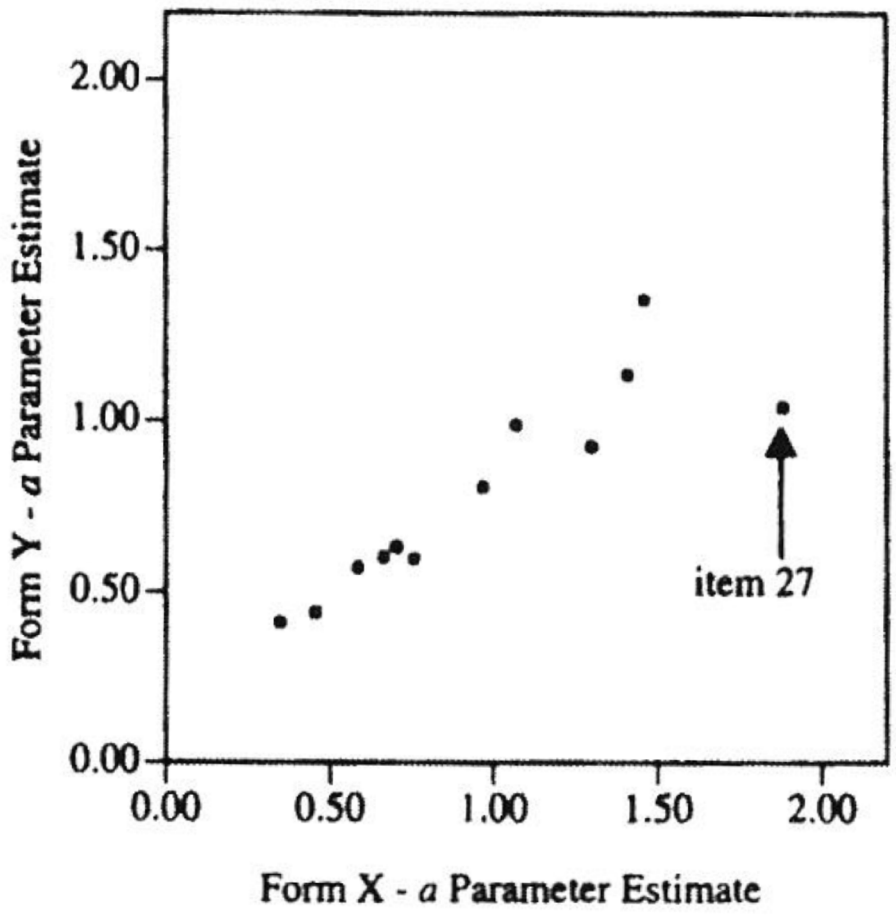


図 2.4　識別力母数の外れ値（Kolen & Brennan（2004）より）

2.7.2. シミュレーション方法

識別力母数の外れ値をシミュレートするために，前節の分析で用いたシミュレーション方法について，共通項目の識別力母数について，小学校国語では上位二つ，中学校数学では上位三つの数値を 0.5 に固定した。それ以外の条件は，前節と同様である。

2.7.3. 結果

各条件における等化係数推定値に関する統計量を表 2.10，2.11，2.12，2.13 に示す。

表 2.10　小学校国語，等化係数 K（識別力母数外れ値あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	0.939	0.073	1.204	0.774	0.094	1.212	0.095	1.545	1.021	0.129	0.459	0.040	0.555	0.381	0.057
M & M	0.901	0.049	1.065	0.786	0.111	1.173	0.066	1.392	1.011	0.143	0.445	0.024	0.494	0.387	0.060
Haebara	0.958	0.063	1.138	0.810	0.076	1.240	0.080	1.583	1.072	0.100	0.471	0.034	0.561	0.392	0.045
M & S（受）	0.961	0.045	1.083	0.832	0.060	1.181	0.058	1.338	1.078	0.132	0.550	0.023	0.600	0.478	0.055
KN	1.018	0.068	1.222	0.863	0.070	1.218	0.063	1.396	1.048	0.103	0.530	0.042	0.638	0.422	0.051
Fixed Item	0.995	0.058	1.167	0.858	0.058	1.235	0.064	1.468	1.124	0.091	0.545	0.045	0.646	0.444	0.063

表 2.11　小学校国語，等化係数 L（識別力母数外れ値あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	-0.046	0.099	0.233	-0.228	0.109	0.494	0.123	0.922	0.257	0.162	-0.634	0.074	-0.452	-0.777	0.082
M & M	-0.080	0.088	0.154	-0.251	0.119	0.451	0.107	0.713	0.231	0.183	-0.643	0.067	-0.468	-0.775	0.079
Haebara	-0.066	0.093	0.139	-0.258	0.114	0.496	0.111	0.860	0.274	0.152	-0.654	0.085	-0.439	-0.843	0.101
M & S（受）	-0.173	0.080	-0.008	-0.350	0.191	0.309	0.089	0.528	0.143	0.304	-0.692	0.075	-0.498	-0.868	0.119
KN	-0.100	0.092	0.112	-0.298	0.136	0.440	0.096	0.702	0.230	0.186	-0.698	0.083	-0.475	-0.892	0.128
Fixed Item	-0.135	0.087	0.037	-0.355	0.161	0.413	0.100	0.710	0.222	0.212	-0.721	0.082	-0.517	-0.916	0.146

表 2.12　中学校数学，等化係数 K（識別力母数外れ値あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	0.922	0.055	1.037	0.798	0.095	1.204	0.077	1.422	1.024	0.122	0.465	0.033	0.559	0.392	0.048
M & M	0.891	0.039	0.972	0.797	0.115	1.158	0.052	1.263	1.037	0.151	0.446	0.020	0.491	0.400	0.058
Haebara	0.900	0.060	1.060	0.791	0.117	1.173	0.076	1.341	1.003	0.148	0.447	0.029	0.512	0.368	0.060
M & S（受）	0.970	0.034	1.065	0.902	0.045	1.174	0.045	1.283	1.067	0.134	0.548	0.020	0.604	0.489	0.051
KN	0.936	0.055	1.055	0.801	0.084	1.237	0.067	1.371	1.073	0.092	0.459	0.029	0.523	0.370	0.050
Fixed Item	0.884	0.044	0.993	0.791	0.124	1.147	0.061	1.296	1.015	0.164	0.463	0.029	0.532	0.374	0.047

表 2.13　中学校数学，等化係数 L（識別力母数外れ値あり）

	K=1.0, L=0.0					K=1.3, L=0.6					K=0.5, L=-0.6				
	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE	Mean	SD	Max	Min	RMSE
M & S（項）	-0.054	0.066	0.114	-0.221	0.084	0.513	0.091	0.741	0.297	0.125	-0.591	0.063	-0.421	-0.741	0.063
M & M	-0.069	0.066	0.089	-0.211	0.096	0.471	0.073	0.681	0.308	0.148	-0.589	0.062	-0.417	-0.758	0.062
Haebara	-0.090	0.065	0.099	-0.263	0.111	0.460	0.094	0.748	0.263	0.168	-0.614	0.070	-0.427	-0.760	0.071
M & S（受）	-0.097	0.065	0.074	-0.250	0.116	0.391	0.074	0.595	0.215	0.221	-0.602	0.072	-0.401	-0.737	0.071
KN	-0.122	0.063	0.052	-0.294	0.138	0.431	0.079	0.658	0.257	0.187	-0.641	0.066	-0.444	-0.766	0.078
Fixed Item	-0.145	0.063	0.025	-0.298	0.158	0.386	0.078	0.620	0.207	0.228	-0.644	0.068	-0.456	-0.769	0.081

2.7.4. 考察

前節の条件では，相対的に（RMSE）の値が小さかったＭ & Ｍについて，幾つかの条件では，その数値が大きくなった。これについては，識別力母数の外れ値の影響であると考えられる。（RMSE）については，どれかの方法が相対的に小さくなるようなことはなく，条件ごとに様々であることが分かった。

2.8. まとめと提案

これまでに見た三つのシミュレーション研究から，条件により各手法の特徴は様々であり，どの条件においても最適であるような方法は存在しないことが分かった。Kolen & Brennan（2004）においても示唆されているように，複数の方法を常に実施，比較することが重要である。本調査のまとめとして，経年変化分析調査データの等化分析においては以下のような方針を提案する。

・元々のテストデザインに従い，共通項目を利用した等化を採用する。
・等化係数の算出については，Mean & Sigma 法，Mean & Mean 法，Haebara 法（Stocking & Lord 法を追加しても良い）を全て行い，結果を比較検討する必要がある。
・上記で検討が困難な場合，項目固定法，また共通受検者デザインによる熊谷・野口（2012）の方法なども実施して，検討の参考とする。

参考文献

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, *22*, 144-149.

Kolen, M. G., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd ed.). New York: Springer – Verlag.

熊谷龍一 (2009). 初学者向けの項目反応理論分析プログラム EasyEstimation シリーズの開発　日本テスト学会誌, *5*, 107-118 .

熊谷龍一・野口裕之 (2012). 推定母集団分布を利用した共通受験者法による等化係数の推定　日本テスト学会誌, *8*, 9-18.

Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, *17*, 179-193.

Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, *14*, 139-160.

Mislevy, R. J., & Bock, R. D. (1990). *BILOG3. Item Analysis and Test Scoring with Binary Logistic Models* (2nd ed.). Mooresville, IN：Scientific Software International.

村木英治 (2011). 項目反応理論. 朝倉書店.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*, 201-210.

3. 対応づけ分析

3.1. はじめに

本章では，異なる二つのテスト間のスコアを結びつけることを対応づけ（linking）[1)]と呼ぶことにする。特に，二つのテストが同一の構成概念を測定しているなどの条件を満たす場合，その対応づけは等化（equating）と呼ばれる。等化の例として，TOEFL の異なる二つの版（同一テストの異なる版）の対応づけがあげられる。

一方，実際のテストの運用場面では，上述した等化の枠組みに収まらないスコア間の対応づけが必要とされるようになってきた。その例として，法科大学院の入学者選抜に利用される実施主体の異なる二つの適性試験の対応づけがあげられる（柴山・野口，2004）。さらに教育現場でも，全国的な学力調査と地方自治体における学力調査との対応づけなどに関心が高まりつつある（石井，2008 など）。

今後の全国学力・学習状況調査においても，経年調査間の等化（本報告書の第 2 章）だけでなく，全国学力調査と経年調査との対応づけを検討しておくことは一考の価値がある。そこで本章では，共通受検者のスコアを用いて平成 25 年度全国学力調査と平成 25 年度経年調査（経年調査Ⅰ＋経年調査Ⅱ）との対応づけを試みる。小学校国語を例として，図 3-1 のように全国学力調査を経年調査に対応づけするための一つの方法を提示する。

なお，全国的な学力調査のうち，悉皆（しっかい）で実施する調査を本章では全国学力調査と呼ぶことにする。また，いわゆる経年変化分析調査を簡単のために経年調査と呼ぶことにする。さらに，平成 25 年度経年調査のうち，重複テスト分冊法を採用しなかった 2 分冊（共通項目なし）をそれぞれ経年調査Ⅰ，Ⅱと呼称する。

本章の構成は，以下のとおりである。3.2 節では，対応づけに用いたデータについて詳細を述べる。3.3 節では，経年調査の実施時期の問題について考察する。3.4 節では，全国学力調査における共通受検者の代表性について考察する。3.5 節では，経年調査のスコア分布を生成する方法について述べる。3.6 節では，全国学力調査と経年調査の対応づけ可能性について検討する。3.7 節では，実際の対応づけ結果を示す。3.8 節では，本章の内容を総括する。

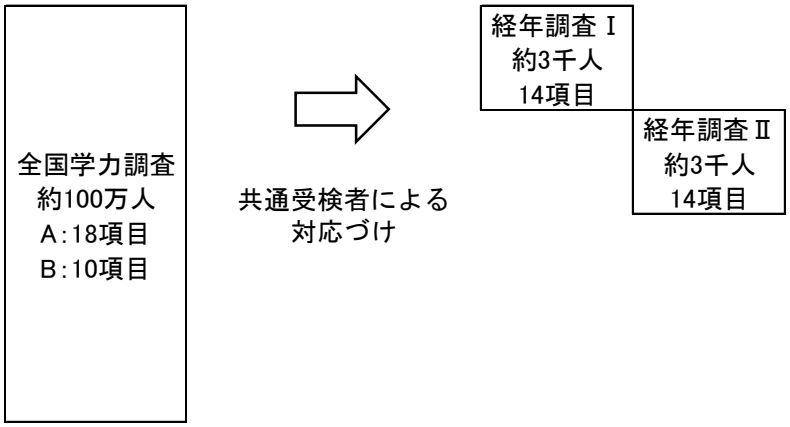


図 3-1　本章の対応づけ

3.2.　対応づけデータ

平成 25 年度全国学力調査（小学校国語）は，平成 25 年 4 月 24 日（水）を中心に実施された。調査対象は，全国の小学校第 6 学年及び特別支援学校小学部第 6 学年の児童生徒であった。悉皆(しっかい)調査として全国 20,746 校から 1,157,235 人が受検した。教科に関する調査では，国語 A（「主に知識に関する問題」，解答時間 20 分，18 項目）と国語 B（「主に活用に関する問題」，解答時間 40 分，10 項目）が出題された。さらなる詳細は，文部科学省・国立教育政策研究所（2013）を参照されたい。

都合により，一部の学校では 4 月 25 日以降に試験が実施された。また，一部の受検者には，特別対応として点字問題冊子，拡大文字問題冊子，日本語指導，時間延長が必要であった。本章の分析では，簡便のため，それらの該当者を分析データから削除することとした。その結果，全国学力調査の受検者数は 1,118,392 人，項目数は A 問題＋B 問題で 28 項目となった。

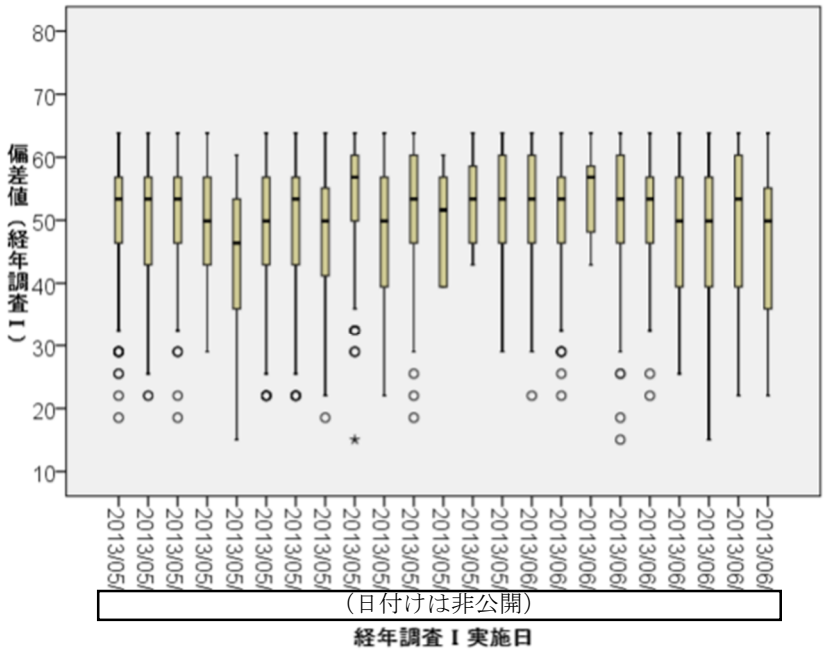
平成 25 年度経年調査（小学校国語）は，平成 25 年 5 月中旬から 6 月下旬にかけて協力校において実施された。全国学力調査と同様に，経年調査にも点字問題冊子，拡大文字問題冊子，日本語指導，時間延長の特別対応がとられた。やはり簡便のため，該当者を分析データから削除することとした。その結果，経年調査Ⅰの受検者数は 2,915 人，項目数は 14 項目となった。経年調査Ⅱの受検者は 2,943 人，項目数は 14 項目となった。

3.3.　経年調査の実施時期

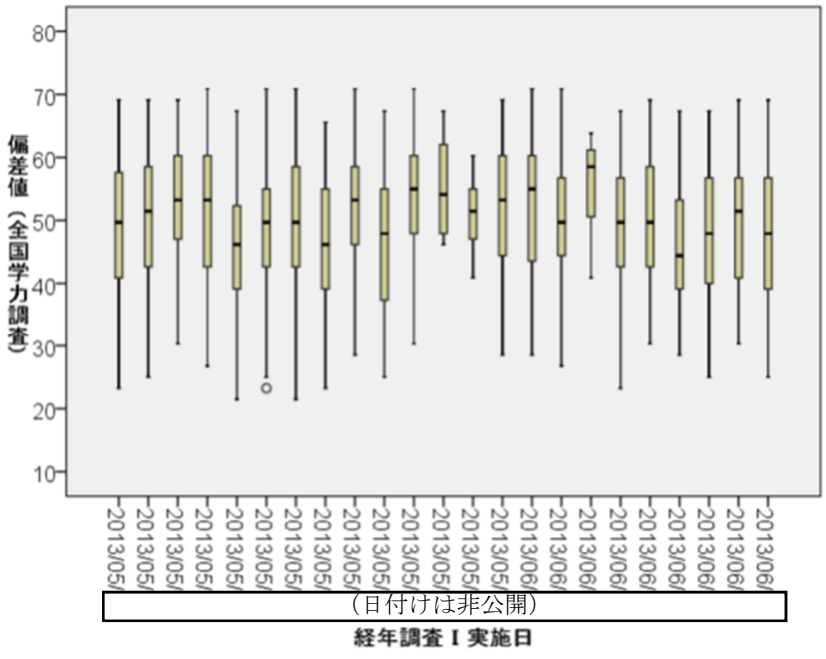
実施校の都合により，経年調査Ⅰ，Ⅱの実施時期は約 5 週間におよぶ。そのため，その

間に受検者の学力が向上し，実施時期の遅い学校ほど経年調査のスコアが高くなるという傾向が現れるかもしれない。その場合，程度によっては対応づけ結果の補正を考慮する必要が出てくる。本節では，実施日の違いが経年調査のスコアにどんな影響を与えたのかについて考察する。

全国学力調査のスコアと経年調査Ⅰ，Ⅱのスコアをそれぞれ偏差値に換算し，経年調査の実施日ごとに受検者の偏差値分布を箱ひげ図で比較した（図 3-2，図 3-3)。図 3-2 (a)と図 3-3 (a)によれば，経年調査の実施日が遅い学校ほど経年調査のスコアが一様に上昇するといった傾向は見られない。また，各図の(a)と(b)を比較すると，経年調査のスコア分布は 4 月 24 日の一日に実施された全国学力調査のスコア分布と相対的な位置関係が似ていることがわかる。これらの理由から，受検者集団の学力は約 5 週間の期間に対応づけ結果の補正が必要なほど変化しなかったと判断した。本章の分析では，経年調査Ⅰ，Ⅱの受検者をあたかも同一日に受検した一つの受検者集団として扱うこととした。

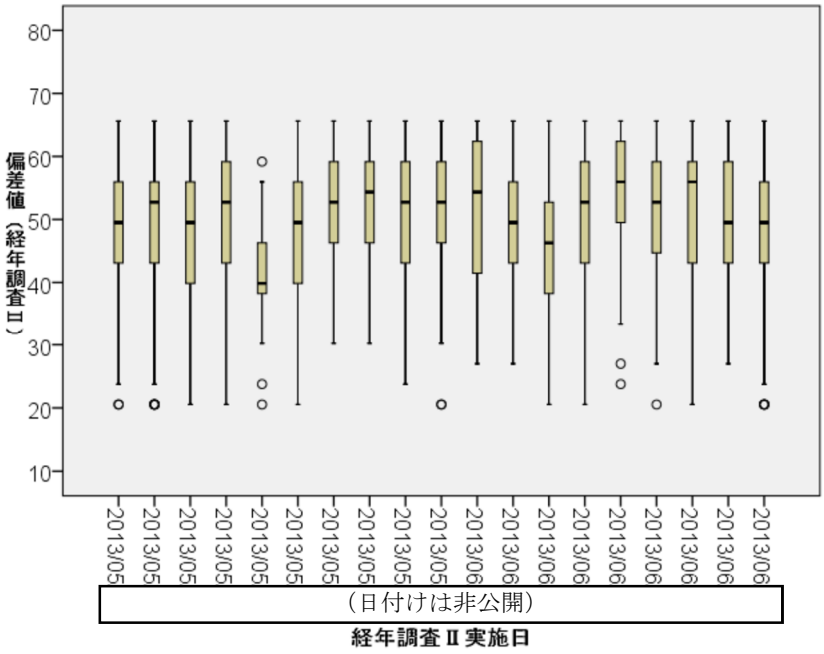


(a)

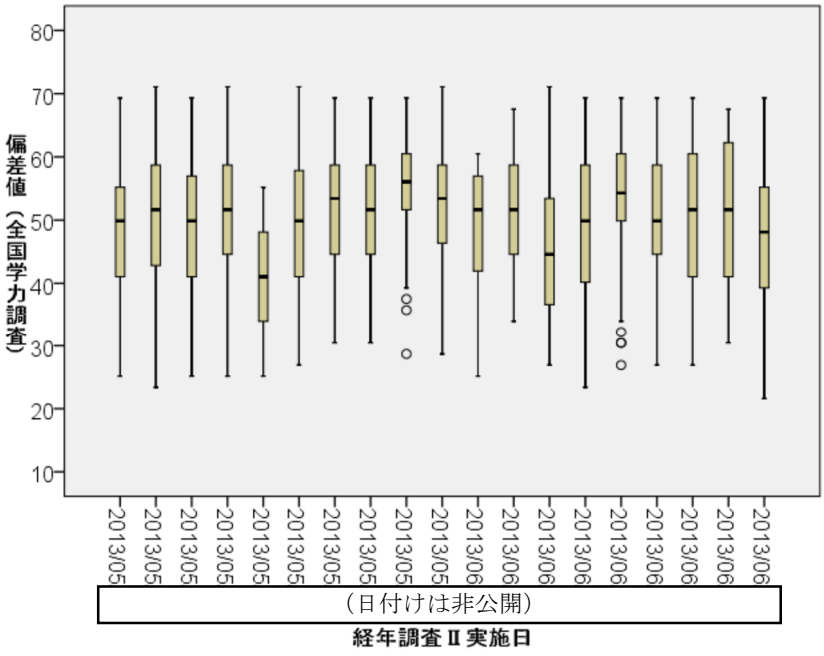


(b)

図 3-2　経年調査Ⅰ実施日による偏差値分布の比較



(a)



(b)

図 3-3　経年調査Ⅱ実施日による偏差値分布の比較

3.4.　共通受検者の代表性

全国学力調査において，共通受検者が全受検者に照らして偏りのある集団になっているような場合には，対応づけ結果の妥当性に問題が生じることがある。本節では，適性試験委員会（編）（2011，第 7 章）の要領に従い，共通受検者のスコア（正答数得点）分布が全受検者のそれを代表しているかどうかを比較検討する。

表 3-1 に，全国学力調査における共通受検者と全受検者の試験結果を示す。表中，「下から 10%」に当たる数値は，各受検者集団の下から 10%の位置（10 パーセンタイル順位）にいる受検者のスコアを示している。図 3-4 と図 3-5 に，共通受検者と全受検者のスコア分布をそれぞれ示す。図中の曲線は，データの分布と等しい平均と標準偏差をもつ正規曲線を表している。図 3-6 に，共通受検者と全受検者のスコア分布を箱ひげ図で示す。

表 3-1 をみると，共通受検者と全受検者の平均・標準偏差は僅かな差であり，各パーセンタイル順位のスコアもほとんど一致している。図 3-4，図 3-5，図 3-6 からも，両者のスコア分布はほぼ同一であるといえよう。したがって，共通受検者は全受検者の傾向を十分に反映しており，対応づけの基本データとして十分に利用可能といえる。

表 3-1　「全国学力調査」各受検者集団のスコアの比較

	人数	平均	標準偏差	下から10%	25%	50%	75%	90%
共通受検者	5,849	16.1	5.66	8	12	17	21	23
全受検者	1,118,392	16.3	5.71	8	12	17	21	24

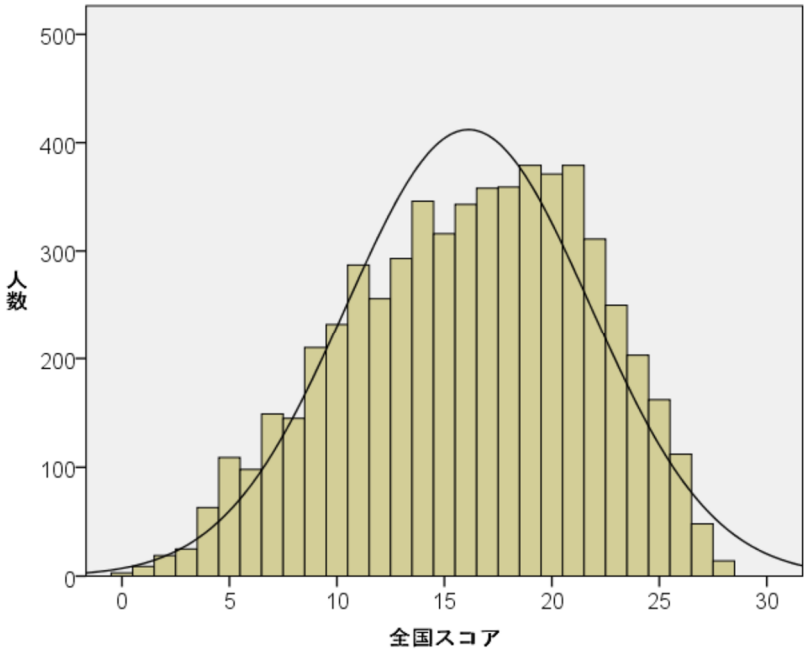


図 3-4　「全国学力調査」共通受検者のスコア分布

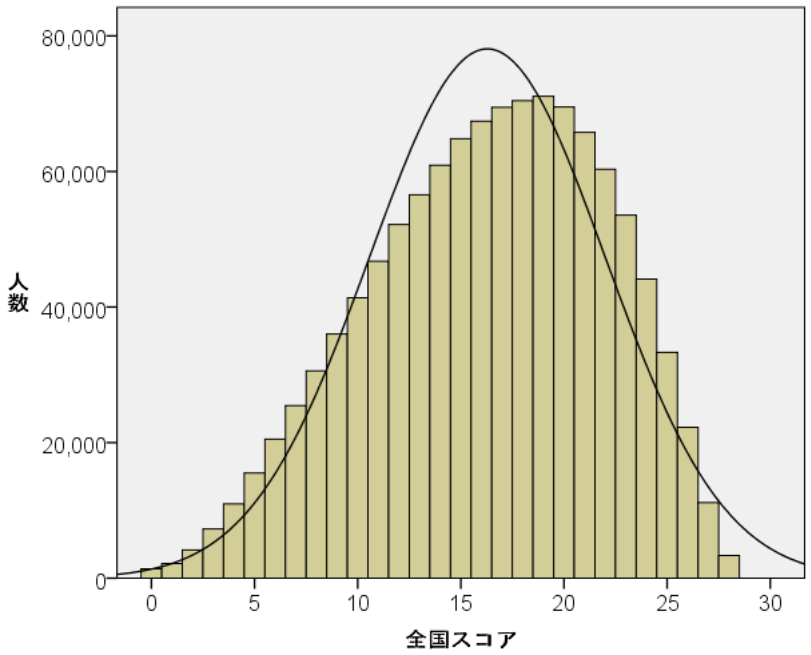


図 3-5　「全国学力調査」全受検者のスコア分布

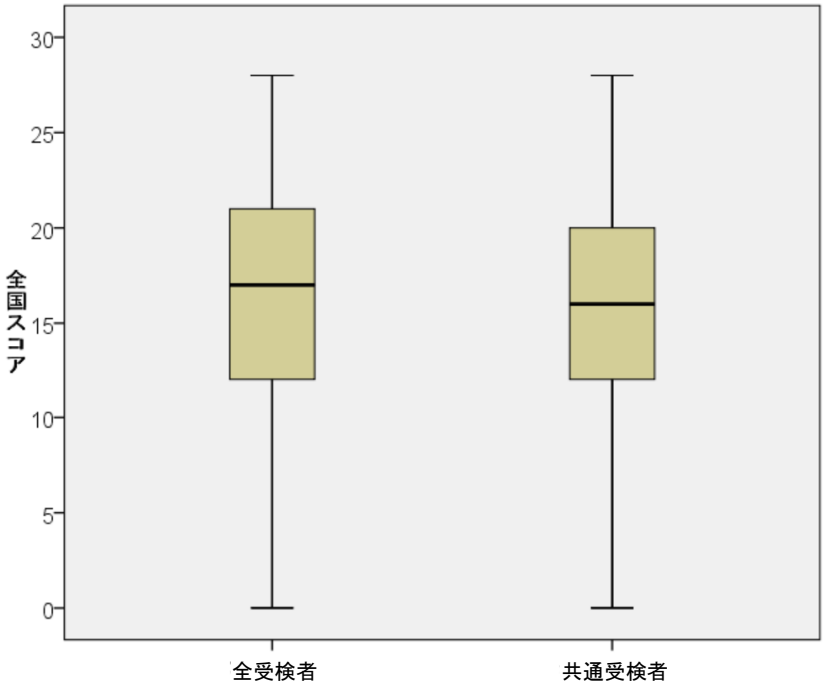


図 3-6　「全国学力調査」各受検者集団のスコア分布の比較

3.5.　経年調査のスコア分布

本章の対応づけでは，全国学力調査と経年調査における共通受検者のスコア分布を利用する。特に経年調査では，経年調査Ⅰ（14 項目）と経年調査Ⅱ（14 項目）を別々に扱うのではなく，一つの経年調査（28 項目）としてスコア分布を推定する。そのようなスコア分布は，項目反応理論（item response theory，IRT）を用いると推定可能である。本節では，経年調査Ⅰ，Ⅱの受検者が仮に全 28 項目を受検した場合のスコア分布を推定する方法について述べる。

以下の手順に従い，経年調査（28 項目）のスコア分布を推定した。

① 項目反応データの準備

図 3-7 のような項目反応データ（正答：1，誤答：0）を準備した。経年調査Ⅰの受検者が経年調査Ⅱを受検した場合の正誤データと経年調査Ⅱの受検者が経年調査Ⅰを受検した場合の正誤データを欠測値とみなした。

② 項目母数の推定

①の項目反応データを利用し，全 28 項目の項目母数を周辺最尤（さいゆう）推定した。項目反応モデ

ルは，2 母数ロジスティックモデル（Lord & Novick, 1969, Chapter 17）を利用した。EasyEstimation（熊谷, 2009）による項目母数の推定結果を表 3-2 に示す。ただし，尺度定数は $D=1.702$ とした。上半分が経年調査Ⅰの結果であり，下半分が経年調査Ⅱの結果である。

③ θ の母集団分布の推定

①と②の結果を利用し，経年調査Ⅰ，Ⅱの全受検者 5,858 人について，能力母数 θ の母集団分布を推定した。EasyEstimation を用いると，θ の母集団分布の推定が可能である。その結果を図 3-8 に示す。

④ スコア分布の推定

②と③の結果を利用し，経年調査Ⅰ，Ⅱの受検者が仮に全 28 項目を受検した場合のスコア分布を推定した。推定には，Lord and Wingersky (1984) Recursion Formula（Kolen & Brennan, 2014, p. 199 にも記載あり）を用いた。その結果を図 3-9 に示す。

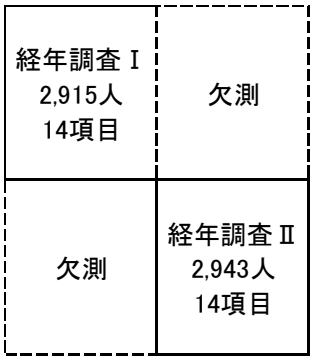


図 3-7　項目反応データのイメージ

表 3-2　経年調査の項目母数

	項目識別力		項目困難度	
項目	推定値	標準誤差	推定値	標準誤差
I-01	1.147	0.109	−2.956	0.154
I-02	1.169	0.102	−2.788	0.131
I-03	0.655	0.031	−0.751	0.046
I-04	0.648	0.034	−1.349	0.065
I-05	0.708	0.041	−1.957	0.089
I-06	0.526	0.028	−0.778	0.056
I-07	0.859	0.035	−0.140	0.030
I-08	1.005	0.040	−0.515	0.029
I-09	0.646	0.033	−1.196	0.059
I-10	0.938	0.046	−1.500	0.053
I-11	0.597	0.030	0.712	0.048
I-12	0.727	0.036	−1.292	0.057
I-13	0.736	0.033	−0.511	0.038
I-14	0.890	0.036	−0.363	0.031
II-01	1.468	0.088	−1.891	0.054
II-02	1.233	0.071	−1.904	0.060
II-03	0.680	0.037	−1.657	0.075
II-04	0.696	0.038	−1.651	0.074
II-05	1.034	0.041	−0.624	0.030
II-06	0.897	0.036	−0.295	0.030
II-07	0.994	0.047	−1.414	0.048
II-08	0.984	0.039	−0.511	0.030
II-09	0.694	0.031	−0.358	0.037
II-10	0.440	0.026	0.551	0.059
II-11	0.857	0.036	−0.625	0.035
II-12	0.815	0.034	−0.260	0.032
II-13	0.821	0.041	1.402	0.055
II-14	0.655	0.030	0.022	0.037

※D = 1.702

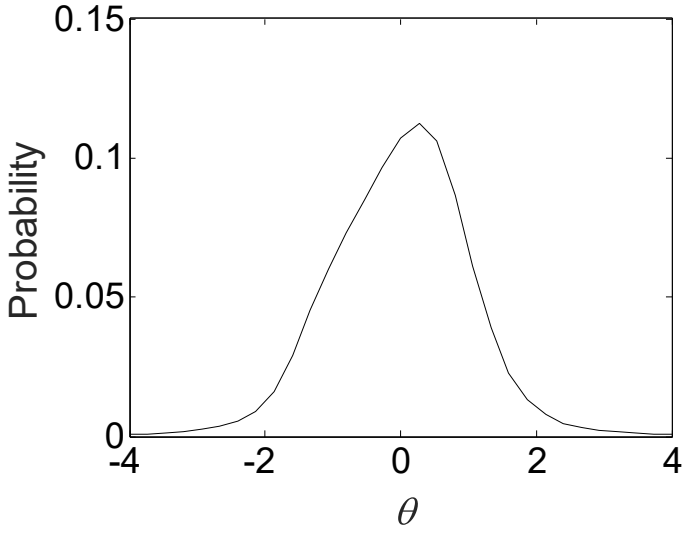


図 3-8　能力母数の母集団分布

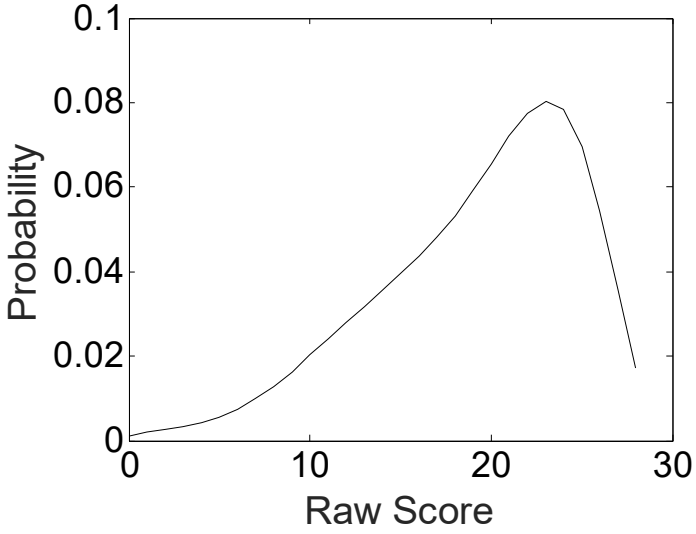


図 3-9　経年調査のスコア分布

3.6. 対応づけ可能性

Dorans（2004a）によれば，対応づけが十分な意味をもつためには，個々のテストの信頼性が高くテスト間の相関も高いことが必要である。特に，ハイステークスなテストを対応づけする際には,スコア間の相関係数が 0.866 以上であることが好ましいと述べている。この相関係数の基準について，Sawyer（2007, p. 221）は，Dorans における ACT と SAT の対応づけ結果をみると不合理な基準とはいえないものの，他の文脈においては厳しすぎる基準かもしれないと指摘している。実際には，石井（2010, p. 17）が考察するように，対応づけが可能なスコア間の相関係数としては 0.7 程度以上が経験的に一つの目安になると考えられる。

本節では，上述の基準に従い，全国学力調査と経年調査との対応づけが意味ある対応づけになるかどうかを考察した。そのための準備として，全国学力調査と経年調査の信頼性係数と両調査のスコア間の相関係数を推定する必要がある。信頼性係数の推定値として Cronbach の α 係数（Cronbach, 1951）を用いると，全国学力調査の α 係数は 0.851 と推定される。一方，経年調査はスコア分布のみ与えられているので，経年調査の α 係数とスコア間の相関係数を推定するには工夫が必要である。

以下の手順に従い，ブートストラップ法（汪・田栗，2003 など）を用いて経年調査の α 係数とスコア間の相関係数を推定した。

① $\hat{\theta}_{EAP}$の推定

経年調査Ⅰ，Ⅱの受検者 5,858 人について，能力母数の EAP（expected a posteriori）推定値を求めた。EasyEstimation により，推定可能である。ML（maximum likelihood）推定値ではなく EAP 推定値を利用した理由は，全問正答・全問誤答の場合に ML 推定値を求められない問題を回避するためである。

② R回の反復処理

- ①の結果と項目母数（表 3-2）を利用し，r回目の経年調査の項目反応データを生成した。項目反応データは，項目反応モデルと一様乱数によって生成可能である。
- r回目の経年調査の α 係数とスコア間の相関係数を推定した。

③ 推定値と標準誤差の計算

R回の平均を推定値，標準偏差をその標準誤差として， α 係数と相関係数を推定した。

表 3-3 と表 3-4 に，経年調査の α 係数とスコア間の相関係数を推定した結果を示す。反復回数R = 1,000, 3,000, 10,000 の結果が示されている。表をみると，R = 1,000 回でも安定

した推定結果が得られていることがわかる。反復回数はなるべく多い方がよいので，$R = 10{,}000$ を最終的な結果として利用することにした。

以上より，全国学力調査の α 係数は 0.851，経年調査の α 係数も 0.851，両調査のスコア間の相関係数は 0.693 と推定された。両調査の信頼性は十分に高く，スコア間の相関係数も 0.7 程度以上である。したがって，両調査は経験的に対応づけ可能と判断される。

表 3-3　経年調査の α 係数

	α 係数	
反復回数	平均	標準偏差
1,000	0.851	0.002
3,000	0.851	0.002
10,000	0.851	0.002

表 3-4　全国スコアと経年スコアの相関

	相関係数	
反復回数	平均	標準偏差
1,000	0.693	0.003
3,000	0.693	0.004
10,000	0.693	0.003

3.7.　対応づけ結果

共通受検者のスコア分布（図 3-4 と図 3-9）を用いて全国学力調査を経年調査に対応づけした結果を図 3-10 に示す。対応づけには，等パーセンタイル法（Kolen & Brennan, 2014, pp. 36－46 など）を利用した。横軸は全国学力調査のスコア（正答数得点）であり，縦軸は経年調査のスコア（正答数得点）である。

対応づけ関数を使うと，二つの試験におけるスコアのおよその対応関係がわかる。図 3-10 をみると，全国学力調査（テスト X）の 20 点は経年調査（テスト Y）の 23 点程度に相当することがわかる。それとは反対に，経年調査（テスト Y）から全国学力調査（テスト X）の対応を求めることもできる。また，Kolen and Brennan（2014, p. 252）のブートストラップ法を用いて対応づけの標準誤差 [2]を推定した。図 3-10 をみると，その最大値は DTM（Difference That Matters）[3]の 0.5 未満であり，経験的に十分な精度で対応づけされたこ

とがわかる。

対応づけ結果を利用するのに，図 3-10 のグラフから正確な対応関係を把握するのは容易でない。そこで，実際の大規模試験などでは，表 3-5 のような対応表が与えられることが多い。この対応表によると，例えば全国学力調査の 20 点は経年調査の 23.1 点に相当することがわかる。

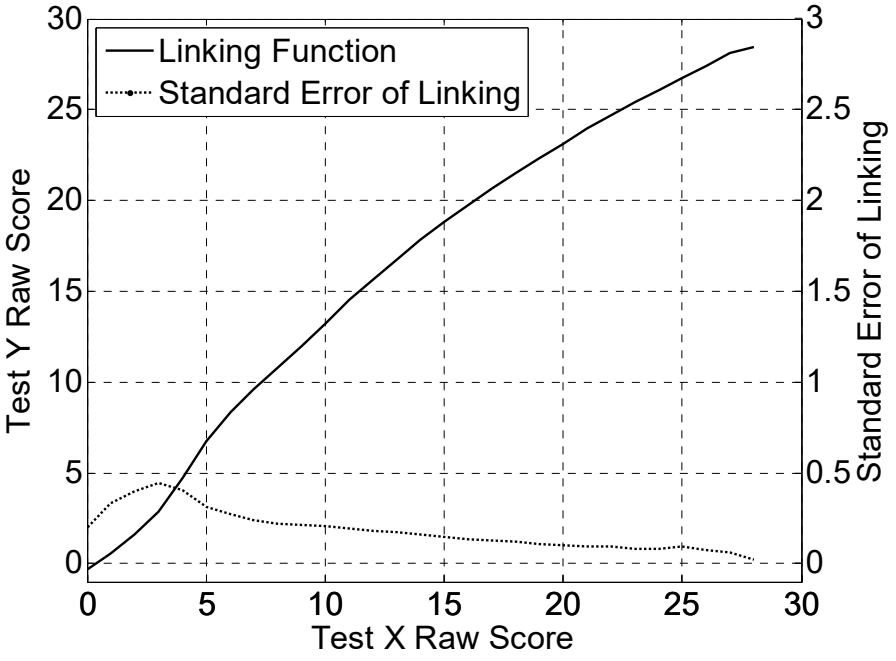


図 3-10　対応づけ関数（実線，左軸）と対応づけの標準誤差（点線，右軸）
（テスト X：全国学力調査，テスト Y：経年調査）

表 3-5　対応表

全国スコア	経年スコア
0	0.0
1	0.5
2	1.6
3	2.9
4	4.7
5	6.8
6	8.3
7	9.6
8	10.8
9	12.0
10	13.2
11	14.5
12	15.7
13	16.7
14	17.8
15	18.8
16	19.8
17	20.7
18	21.5
19	22.3
20	23.1
21	23.9
22	24.7
23	25.4
24	26.1
25	26.8
26	27.4
27	28.0
28	28.0

3.8.　おわりに

本章では，共通受検者のスコアを用いて平成 25 年度全国学力調査と平成 25 年度経年調査（経年調査Ⅰ＋経年調査Ⅱ）との対応づけを試みた。小学校国語の例を通し，全国学力調査を経年調査に対応づけするための一つの方法を提示した。本章の例では，経験的に十分な精度で対応づけ結果が得られることが示唆された。

Feuer, Holland, Green, Bertenthal, and Hemphill（1999）では，当時のクリントン大統領の要請に端を発し，対応づけ可能性に関する研究が実施された。その結果，市販あるいは州のアセスメントを相互に対応づけしたり，NAEP に対応づけしたりするのは好まし

くないという結論が得られている。このように，本章の例のような対応づけに否定的な研究結果が存在することも事実である。今後の全国学力・学習状況調査において，もし今回のような対応づけが必要な場合は，専門家による慎重な検討が必要なのはもちろんのこと，対応づけ結果の利用についても制限が設けられるべきであろう。

3.4 節において共通受検者の代表性を考察する際には，男女比，学校種の割合，都道府県の割合など，受検者の属性についても代表性を確認するのが望ましい。3.6 節の対応づけ可能性分析では，構成概念の類似性，テストの仕様，対応づけ関数の集団不変性なども含めて多面的・総合的に評価できるとなおよい。また，今回の経年調査は 2 版だけであったものの，より多くの版を利用する場合には項目反応データの欠測値の割合が増えるという問題が生じる。本章で示した方法が今後の調査分析にそのまま適用できるという保証はないものの，実データを用いて有効な対応づけ結果が得られる場合があることを示すことができたのは一つの成果といえよう。今後の全国学力・学習状況調査において，対応づけの技術がますます有効に活用されることを願ってやまない。

注釈

1) Holland and Dorans (2006)の枠組みでは, リンキング (linking) の型は等化 (equating)，尺度整列化（scale aligning），予測（prediction）という三つの基本的なカテゴリに分類されている。本章では，このような広い意味での"linking"という用語に「対応づけ」という訳語をあてた。

2) ある得点xの対応づけ得点を$l_Y(x)$とするとき，標本変動によって対応づけ誤差$l_Y(x) - \hat{l}_Y(x)$が生じる。その誤差分布の標準偏差を対応づけの標準誤差と呼ぶ。いわば等化の標準誤差（Kolen & Brennan, p. 248）の対応づけ版である。

3) DTM は，Dorans, Holland, Thayer, & Tateneni（2003）及び Dorans（2004b）において提案された指標である。報告されるスコアの 1 ユニットの半分が DTM である。本章の場合，スコアとして正答数得点を利用しているので，DTM は 0.5 点である。

参考文献

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297－334.

Dorans, N. J. (2004a). Equating, concordance, and expectation. *Applied Psychological Measurement, 28*(4), 227－246.

Dorans, N. J. (2004b). Using subpopulation invariance to assess test score equity. *Journal of Educational Measurement, 41*(1), 43－68.

Dorans, N. J., Holland, P. W., Thayer, D. T., & Tateneni, K. (2003). Invariance of score linking across gender groups for three advanced placement program exams. In N. J. Dorans (Ed.), *Population invariance of score linking: Theory and applications to advanced placement program examinations* (pp. 79-118), Research Report 03-27. Princeton, NJ: Educational Testing Serveice.

Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999). *Uncommon measures: Equivalence and linkage among educational tests* (Report of the Committee on Equivalency and Linkage of Educational Tests, National Research Council). Washington DC: National Academy Press.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 187-220). Westport, CT: American Council on Education and Praeger.

石井秀宗（2008）．全国学力・学習状況調査の項目分析的検討．平成 19 年度「全国学力・学習状況調査」分析報告書，98－111．千葉県検証改善委員会.

石井秀宗（2010）．地域におけるデータ等を補完的に用いた調査分析手法の調査研究．平成 21 年度文部科学省企画公募委託研究「学力調査を活用した専門的な課題分析に関する調査研究」報告書.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer.

熊谷龍一（2009）．初学者向けの項目反応理論分析プログラム EasyEstimation シリーズの開発．日本テスト学会誌，5(1)，107－118.

Lord, F. M., & Novik, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score “equatings”. *Applied Psychological Measurement, 8*(4), 452－461.

文部科学省・国立教育政策研究所（2013）．平成 25 年度全国学力・学習状況調査報告書―小学校国語―．http://www.nier.go.jp/13chousakekkahoukoku/data/research-report/13-p-language_2.pdf＞（2016 年 5 月 9 日）

Sawyer, R. (2007). Some further thoughts on concordance. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales.* New York: Springer.

柴山　直・熊谷龍一・後藤武俊・佐藤喜一（2014）．東日本大震災の学力への影響―IRT 推算値による経年比較分析―．平成 25 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.

柴山　直・野口裕之 (2004).「対応づけ」の理論と計算アルゴリズム. 適性試験委員会 (編), 法科大学院統一適性試験テクニカル・レポート 2004, 53－72. 商事法務.

適性試験委員会 (編) (2011). 法科大学院統一適性試験テクニカル・レポート 2009-2010. 商事法務.

汪　金芳・田栗正章 (2003). ブートストラップ法入門. 甘利俊一・竹内　啓・竹村彰通・伊庭幸人 (編), 統計科学のフロンティア 11 : 計算統計―確率計算の新しい手法―, 第Ⅰ部, 1－64. 岩波書店.

4.　等化されたテストの運用上の課題

等化されたテストを実際に運用していく場合には，例えば問題項目を非公開にするなど，従来の（等化していない）テストの運用に加えて，考慮しなければならない課題が多く存在する。そこで，実際に等化されたテストについて豊富な運用実績がある米国の Educational Testing Service を訪問し，インタビュー調査を行った。調査は平成 28 年 1 月に実施，調査対象者は Brent Bridgeman 氏であった。インタビューの概要を以下に示す。

調査対象者によれば，信頼性と妥当性をもった十分なデータがあれば，訴訟をもたらすようなトラブルが生じても対応できると考えている。大学の入学試験において，標準テストは，妥当性を高める可能性を持っているものの，スコアの検証において誤った妥当性の解釈を生じさせる脅威も認識しておかなければならないとする。インタビュー調査において，挙げられた妥当性の解釈に影響を与えるリスクを以下に整理する。

■ fatigue
- テストによって生じる疲労は，テストの妥当性の低下と受検者に不満を生じさせる可能性がある。ただし，5 時間以下のテストでは問題ないというデータがある。

■ time limit
- アメリカでは，テストの制限時間に関して様々な議論がある。
- テストにおいて問題を解くスピードが大きな影響力を持つべきではない。GRE も SAT もスピードの違いによる影響が生じないように配慮されている。
- 制限時間の違いによるテスト得点の有利・不利を検証した結果，SAT の言語テストでは，制限時間よりも多い時間を得た受検者群において得点が少しだけ高かった。ただし，学力（言語テストの総得点で算出）の低い層の受検者群では，制限時間よりも多めに時間を与えても，テスト得点が上がることはなかった。一方，数学のテストでは，やや大きな差が生じた。特に中間層の学力において大きな差が生じた。興味深いのは，学力の低い層において，時間の余裕が全く有利に作用しておらず，むしろ，点数が低くなっていたことである。

■ Guessing
- 多肢選択方式のテストでは，一部であるが「当て推量」の補正をしている。無回答の問題には点数は付与されないが，間違って解答された問題は減点される。
- 長い間，SAT では，「当て推量」に対する補正をしてきた。多くの場合，大きな相違は生じない。とはいっても，受検者集団の一部では部分的に違いが生じるし，無回答は間違いとされるために，テストの大部分が当て推量の補正になってしまう場合もある。 解答に自信がある人は積極的に推量し，自信がない人たちは，推測するのをためらう。このような自信に関する要因を私たちは測定していた。

■ Scoring Errors

- 評価者間の不一致は，評価上の問題をもたらすため，明確なルーブリックを用いたり，評価者のトレーニングやモニタリングを行ったりすることで，評価の不一致を最小化する必要がある。多くの人はルーブリックに注目しがちだが，それは余り重要ではない。なぜなら，言語のルーブリックには，「文法的な間違いをせず，流暢（りゅうちょう）に書く」ということが最初に示されている。大学生向けのルーブリックと同じものを高校２年生向けのものとして利用することは難しいだろう。

■ Lack of Test Preparation

- テスト対策の有無は，妥当性にとって重要な要因となりうる。
- 米国には，テスト対策に関する巨大な教育産業があり，かなり高い精度でテストに関する情報を収集し研究しているが十分とはいえない。SAT のような試験では，上記のような情報の影響力は大きくない。
- 受検生がテスト内容を知っていると妥当性は低下するが，受検生らが問題形式に不慣れである場合，そのテストは失敗しているといってよい。そのため，テストの基本的な形式や問題の種類といったテストの構成に関する情報は，すべての受検生に提供されるべきである。

■ Screen Size and Resolution

- CBT などコンピューターを用いたテストをする場合，画面の大小は大きな問題ではなく，解像度が重要な問題である。文章内容を考える問題である場合，高解像度の大画面であれば文全体を把握することできるが，低解像度の小さい画面では文全体の一部分しかみることができない。それゆえ，画面のスクロールが必要になり，このアクションが正確な測定に支障を来す。したがって，受検生によって使用する画面の解像度が異ならないように注意が必要である。

■ Keyboarding Skills

- 1998 年には，CBT で実施される GMAT においてキーボードに不慣れな受検者が相対的に不利になることを懸念していた。しかし，今では，幼稚園の頃から iPad のようなタブレットに触れている人たち（余り鉛筆を利用したことない人たち）に対して，ペーパーテストを受けさせることに幾分か懸念を抱いている。つまり，キーボード入力に不得手な受検者と鉛筆による手書きに不得手な受検者の両方を配慮しなければならない時代となった。

■ Memorizing

- 本当に理解しているかを評価するためには，教科書の答えの丸暗記は妨げになる。また，エッセイテストでも暗記は問題となる。実際，同じような用語や表現を用いた受検生の解答をみることが少なくない。出題する問題が異なっているときでさえ，同じような言い回しが一部の受検生にみられる。例えば，自分が記憶している内容と程遠い問題が出されたとしても，自分が記憶している内容が生かせる形に解答を作り上げている。こうした丸暗記の傾向は，中国系の人たちには特に

多い。受検生のエッセイにおいて暗記された文章を特定するには，何をすべきかを考えなければならない。

■ Underrepresentation

・ 多肢選択方式は，記述式よりもコストはかからない。とは言っても，コストを最小化するために，全てを多肢選択方式にすることが適切であるかは考えなければならない。というのも，予測的妥当性の検証のために，AP（Advanced Placement）における「歴史」のテストでエッセイ及び多肢選択方式の得点と大学入学後の歴史科目の相関関係を分析したところ，多肢選択方式とエッセイの相関係数に大きな違いはみられなかった。しかし，性別で得点差をみると，エッセイでは大きな差が見られなかったものの，多肢選択方式の得点において男子の方が高い得点がみられた。これは，大学入学後に良い成績をとりたいと考える女子にとって不公平を生じさせかねない。つまり，妥当性という一つの基準だけをみるのではなく，様々な方向性から検討することが重要なのである。

5. 発展的 IRT モデルの適用について

5.1. はじめに

本研究では，経年比較調査データの IRT 分析においては，正答－誤答の 2 値型データとして 2 パラメタ・ロジスティック・モデルを採用してきた。IRT ではこのほかにも，順序付多値型モデルや，名義反応モデル，多次元モデルなど様々な応用モデルが存在する。このうち，順序付多値型モデルは実際のテスト場面などでも利用されており，研究例も多い。対して，名義反応モデルや多次元モデルなどは，実用場面はもとより，研究事例も極めて少ない。

そこで本章では，経年比較調査データに対して，後者の名義反応モデル，多次元モデルが適用可能であるかどうかについて，検討することを目的とする。

5.2. 名義反応モデルの適用について

本研究では，受検者の反応を「正答」若しくは「誤答」の 2 値型データとして扱い，等化や対応づけについて検討を行った。本来，経年比較調査データにおける受検者の反応は，選択した選択枝（ここではテスト・スタンダード（日本テスト学会，2007）に従い，選択肢ではなく選択枝の用語を用いる）や実際に記述した解答を幾つかの類型に分けてデータ化されている。この中で特定の類型が「正答」に分類され，それ以外の類型は「誤答」として処理される。つまり，誤答に分類された受検者においては，様々な類型が含まれており，前章までで取り扱ってきた 2 値型モデルにおいては，このような誤答受検者の多様性を区分けすることはできなかった。

2 値型以上のカテゴリ数を扱うモデルは多値型モデルと呼ばれ，各カテゴリに順序が付いている場合は，順序付多値型モデルが適用される。しかしながら，今回のように誤答が幾つか類型を含んでおり，各類型に順序が仮定されない場合は，名義反応モデルを適用することとなる。本節では，経年比較調査データに対して名義反応モデルが適用できるかどうかを検討することを目的とする。なおこの目的のため，名義反応モデルそのものについての記述は本節では取り扱わない（別途 IRT に関する文献，例えば Linden & Hambleton, 1997 などを参照されたい）。また，問題項目の非公開などから，実際の問題内容の分析を目的とはしないため，実際の類型内容なども記載しない。

5.2.1. 分析結果

小学校・国語のデータについて名義反応モデルの分析を行った。なお分析については EasyNominal（熊谷，2015）を利用した。分析の目的上，実際の項目母数などは記述しないが，推定計算については問題なく繰り返し計算が終了し，標準誤差の値についても適切な範囲であった。具体的に特徴のある項目カテゴリ特性曲線について，2 項目分を図 5.1，5.2 に示す。

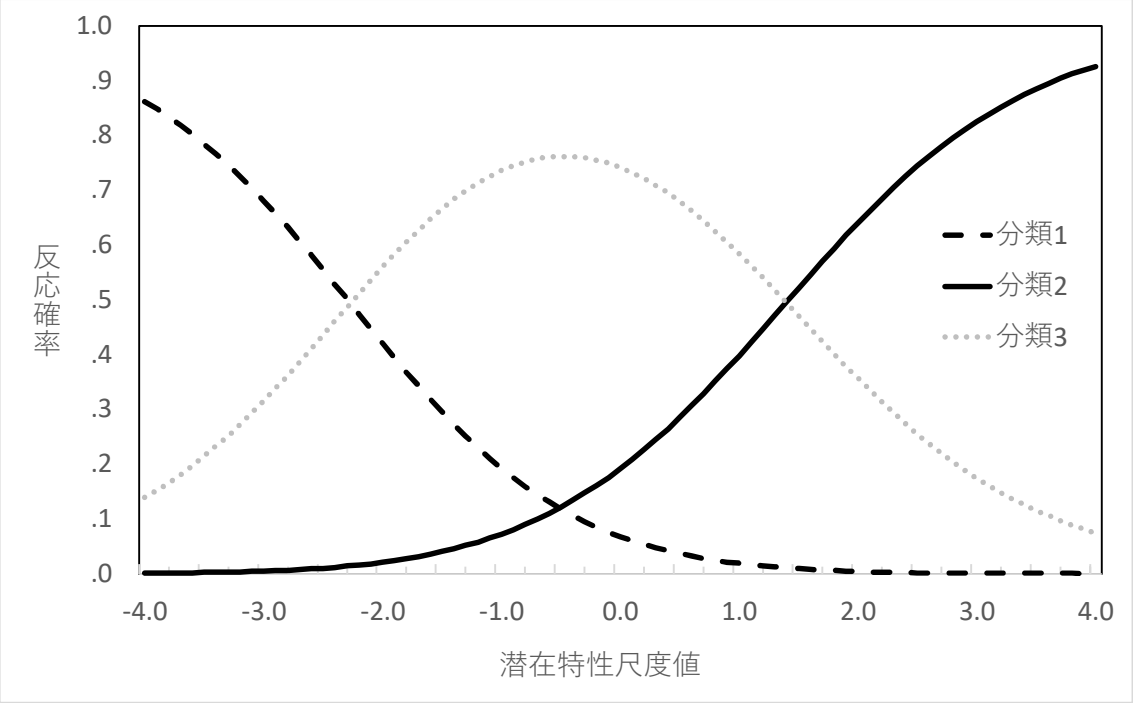


図 5.1　名義反応モデル項目カテゴリ特性曲線 1

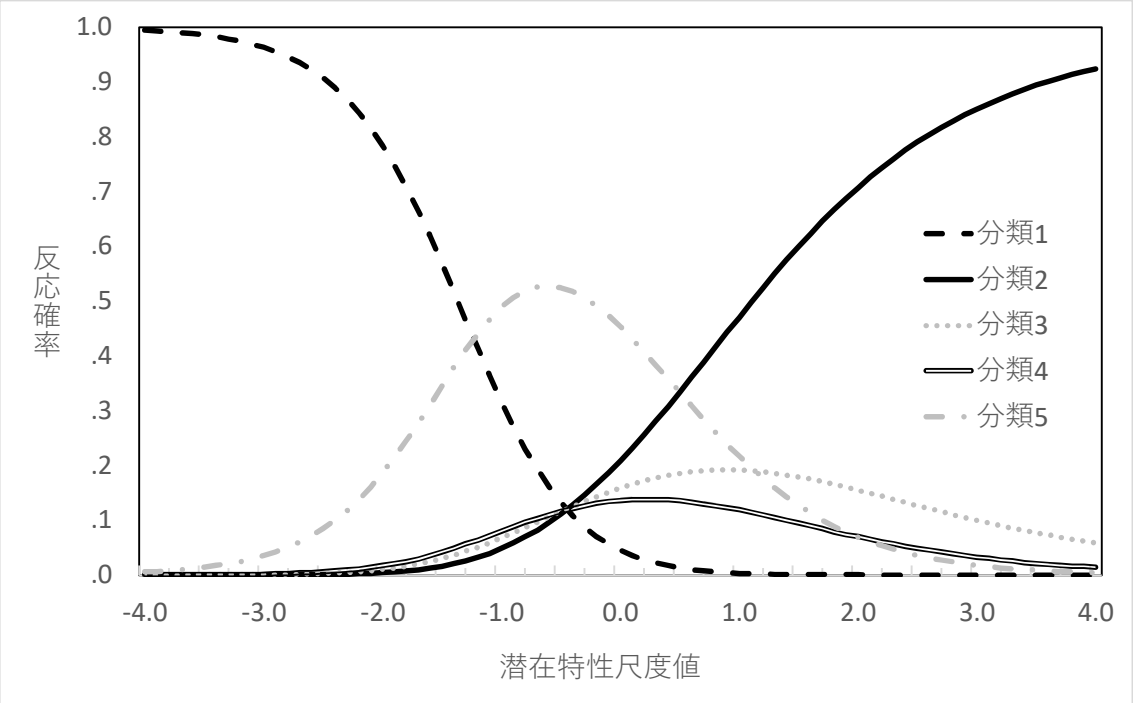


図 5.2　名義反応モデル項目カテゴリ特性曲線 2

5.2.2. 考察

小学校・国語だけではあるが，名義反応モデルによる分析は十分可能であることが示された。図 5.1，5.2 に示されたように，各類型に関する反応確率をモデル上で表現することができる。なお両図において，実線で描かれた分類 2 が正答類型である。各類型について見てみると，図 5.1 の項目では，潜在特性尺度値が低い受検者は分類 1 を，中程度の受検者は分類 3 を，高い受検者は分類 2（正答類型）を選択する様子が示されている。図 5.2 の項目では，分類 3 や 4 を選択した受検者の潜在特性尺度値もやや高い状況が見て取れる。

このように，誤答に分類された受検者の情報を加味した IRT 分析を行う可能性が示されたが，課題も多い。特に本研究の第 2 章，第 3 章で扱った等化や対応づけ分析について，名義反応モデルを採用した場合にはどのような結果になるのか，より詳細な分析が必要であろう。

5.3. 多次元モデルの適用について

第 2 章でも述べたとおり，通常の IRT モデルでは，テストが測定しようとしているものが 1 次元であるという仮定が必要となる。

経年比較調査データについて考えてみると，本体調査と同様に，A 問題，B 問題より構成されている。もしこれらの問題が測定している構成概念が別物であり，1 次元性の仮定を満たさないものであれば，通常の IRT モデルではなく，異なる構成概念に対応した多次元モデルの導入を検討しなければならない。

第 2 章ではこの点について，平成 25 年度経年比較調査データの各教科及び冊子ごとに，項目間テトラコリック相関係数行列を用いたスクリープロットにより，検討を行った（図 5.3 に再掲する）。

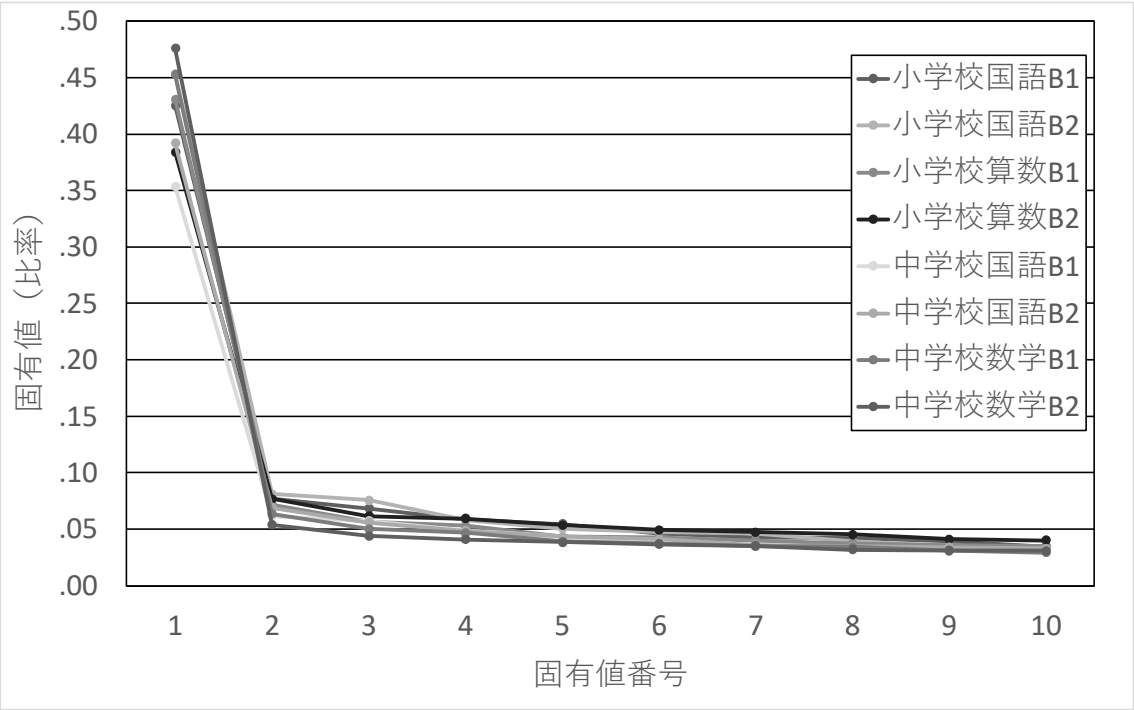


図 5.3　教科・冊子ごとのスクリープロット（再掲）

図 5.3 から明らかなように，平成 25 年度経年比較調査データに関しては，極めて高い 1 次元性が示唆される。したがって，このようなテストデータに対して，多次元モデルを適用したとしても，2 次元目で測定しているものについては明確なものは得られないと考えられる。

したがって，経年比較調査データに対しては（少なくとも平成 25 年度データと測定の構成概念が変化していないうちは），多次元モデルの適用は必要ないと考える。

文献

熊谷龍一 (2015). 第 21 章　項目反応理論分析プログラム EasyEstimation シリーズ　野口裕之・渡辺直登（編著）　組織・心理テスティングの科学－項目反応理論による組織行動の探求－　白桃書房，517-538.

日本テスト学会 (編) (2007) テスト・スタンダード－日本のテストの将来に向けて－　金子書房.

Van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.

付録　項目管理ツールについて

経年比較調査において最も重要なポイントは数十年にわたって測定尺度が固定されていることである。測定モデル（Psychometric model）として IRT を利用した場合には，全国学力・学習状況調査の枠組み内において，経年変化分析調査を今後継続する中で，項目の暴露効果による項目特性の変化や学習指導要領の改訂等に伴う項目の入替えの可能性を常に視野に入れておく必要がある。

その際，問題内容の管理とととともに項目特性の管理を同時に行うことが必須である。ここではその作業用ツールとして試作した項目管理表を示す。この管理表の作成には一般的に普及している Microsoft 社の集計ソフト Excel 2010 を利用した。

管理の具体をシミュレートするため，平成 25 年度の実資料を利用した。ただし，分冊数が 2 である H25 年度実施のものを基本に作成したため，分冊数が増えた H28 年度用に利用するためには，H28 年度の項目とあわせた上，その他の機能も含めて大幅な修正改良が必要である。

また，大問形式が使われることが多い我が国のテストの出題形式への対応も念頭に置いた仕様にしたが，プロトタイプ段階のため十分とは言えない。例えば，大問番号の ID と項目 ID との連動などの機能がそれに該当する。

そのため，本格的な運用には経年変化分析調査が安定して実施されるめどが立った段階で，それまでの運用経験・ノウハウを反映した詳細仕様にもとづく DB を構築する必要がある。

フォルダー構造

管理表本体である Microsoft Excel マクロ有効マークシートである「経年_項目管理表.xlsm」が置かれているフォルダーの下に，問題の画像ファイルが置かれた「画像一式」フォルダーがある。その下に更に学年ごと科目ごとの分冊フォルダーが置かれており，そこに実際の問題が PNG 画像ファイルとして保管されている。この画像ファイルはエクセル上の項目管理表からハイパーリンクで呼び出せるようになっている。「画像一式フォルダー」のイメージは以下の通りである。

項目管理表の実際

問題文など非公開のため一部のみを示している。

項目ID	対象	教科	出題趣旨	問題	画像パス1	画像パス2	形式	選択肢数
E6K0001	小6	国語	非公開		¥小6国語1¥1.png		短答式	
E6K0002	小6	国語			¥小6国語1¥1.png		短答式	
E6K0003	小6	国語			¥小6国語1¥1.png		短答式	
E6K0004	小6	国語			¥小6国語1¥1.png		短答式	
E6K0005	小6	国語			¥小6国語1¥2.png		短答式	
E6K0006	小6	国語			¥小6国語1¥2.png		短答式	
E6K0007	小6	国語			¥小6国語1¥3.png		選択式	5
E6K0008	小6	国語			¥小6国語1¥4.png		短答式	
E6K0009	小6	国語			¥小6国語1¥4.png		選択式	4
E6K0010	小6	国語			¥小6国語1¥5a.png	¥小6国語1¥5b.png	選択式	2
E6K0011	小6	国語			¥小6国語1¥5a.png	¥小6国語1¥5b.png	記述式	
E6K0012	小6	国語			¥小6国語1¥6a.png	¥小6国語1¥6b.png	短答式	
E6K0013	小6	国語			¥小6国語1¥6a.png	¥小6国語1¥6b.png	短答式	
E6K0014	小6	国語			¥小6国語1¥6a.png	¥小6国語1¥6b.png	記述式	
E6K0015	小6	国語			¥小6国語2¥1.png		短答式	

項目管理表の変数定義

項目管理表の変数定義は以下の通りである。

列番号	変数名		値(定義)
1	項目ID		【項目IDルール】 学年＋教科＋連番5桁 (例)E6K00001 ＜学年＞・・・小学6年(E) or 中学3年(J) ※「Elementary school」「Junior high school」の略 ＜教科＞・・・ 国語(K) or 算数/数学(S) ※「Kokugo」「Sansu/Sugaku」の略
2	対象		小6 or 中3
3	教科		国語 or 算数 or 数学
4	出題趣旨		平成25年度 経年変化分析 調査結果概要から引用 テキスト
5	問題		平成25年度 経年調査 問題冊子から引用 テキスト
6	画像パス1		実際の問題文のPNG画像ファイルへのパス1 例:¥小6国語1¥2.png
7	画像パス2		実際の問題文のPNG画像ファイルへのパス2 例:¥小6国語1¥5b.png
8	形式		選択式 or 短答式 or 記述式 (平成25年度 経年変化分析 調査結果概要から引用)
9	選択肢数		平成25年度 経年調査 問題冊子を参照
10	選択肢		平成25年度 経年調査 問題冊子から引用
11	正解		正解選択肢番号
12	採点		二値 o r多値(採点後のデータをSPSSで確認。正解/不正解のみ→二値 、部分正解が加わる場合→多値。)
13	配点		要配点情報(多値型の場合は満点の値) ※部分点がある場合は備考に書く。
14	領域		平成25年度 経年変化分析 調査結果概要から引用
15	観点		平成25年度 経年変化分析 調査結果概要から引用
16	項目パラメータa		項目識別力の値(最新)
17	項目パラメータb		項目困難度の値(最新)
18	項目パラメータc		項目母数に関する予備(当て推量母数の値など)
19	選択率		選択式の場合などに使う予定(仮)
20	点双列相関係数		選択式の場合などに使う予定(仮)
21～40	予備O1～予備2O		
41	2013年度(平成25年度)	冊子	Ⅰor Ⅱ
42		大問	その項目が所属している冊子ごとの大問番号
43		SPSS	SPSSデータにおいて対応する変数名
44	2016年度(平成28年度)	冊子	Ⅰor Ⅱ
45		大問	その項目が所属している冊子ごとの大問番号
46		SPSS	SPSSデータにおいて対応する変数名

※管理表を更新する際の注意点
・アンカー項目と同じ項目IDを用いるのは、内容が完全一致の場合のみ。リライトが入った場合には新しく項目IDを立てる。