経年変化分析調査との対応づけによる
本体調査の年度間比較の試み

平成 29 年度文部科学省委託研究
「学力調査を活用した専門的課題分析に関する調査研究」
研究成果報告書

平成 30 年３月 30 日

国立大学法人東北大学

はしがき

この報告は，平成 29 年度文部科学省企画公募研究「学力調査を活用した専門的な課題分析に関する調査研究」の『B.経年変化分析調査を活用した本体調査の教科の設問の解答状況に関する調査研究』に応募し，技術審査会等を経て採用された調査研究，「経年変化分析調査との対応づけによる本体調査の年度間比較の試み」の成果をまとめたものである。

全国学力・学習状況調査のうち，本体調査においては指導目的等のため毎年問題が公開され，かつ実施対象となる児童生徒集団も年度ごとに入れかわっていくため，原理的には経年比較ができない構造になっている。一方，平成 25 年度に導入され，平成 28 年度は２回目の実施となった経年変化分析調査は，大規模学力調査における国際標準の調査方式である重複テスト分冊法を採用し，かつ問題非公開のため，共通項目デザインのもと平成 25 年度項目，平成 28 年度項目とも基本的にはすべて項目が IRT モデルにより共通尺度上に等化可能であり，経年変化を追うことができる。

本事業においては両調査のもつそれぞれの利点と，項目反応理論（Item Response Theory：IRT）モデルや等化（equating），対応づけ（linking）などの学力調査の測定技術を組み合わせることで，

１）平成 25 年度児童生徒集団が仮に平成 28 年度本体調査を受検したとした場合に生成できる復元得点分布（IRT observed score distribution）による両年度間の学力分布の変化を捉える方法,

２）IRT モデルなどの複雑な方法を利用しなくても都道府県レベルで，上と同等のことが実現できる，平成 25 年度本体調査の得点を平成 28 年度本体調査規準の得点へ変換するための換算表（対応表：concordance table）を作成する方法,

３）項目数が少ないため安定的な経年比較ができない科目内領域ごとの年度間比較のために，推算値（Plausible Values:PVs）を利用する方法,

の三つを開発・提案することを目的とした。ここで開発・提案した手法を使えば，全国学力・学習状況調査の本体調査であっても，経年変化分析調査データを経由することで，都道府県単位ごとの年度間比較が可能となる。年度内限りでの都道府県間の相対比較にともすれば陥りがちであったこれまでの本体調査の結果利用の競争原理的な在り方を超えて，全国学力・学習状況調査の本来の目的の一つである政策立案（Evidence-Based Policy Making：EBPM）のための総合的かつ客観的根拠を供給する技術的基盤の一つとなるものと考えている。

研究代表者　柴山　直

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	経年変化分析調査を活用した本体調査の教科の設問の解答状況に関する調査研究
委託期間	平成 29 年 7 月 10 日　～　平成 30 年 3 月 30 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・工藤　与志文
事 業 費	2,302 千円

研究組織

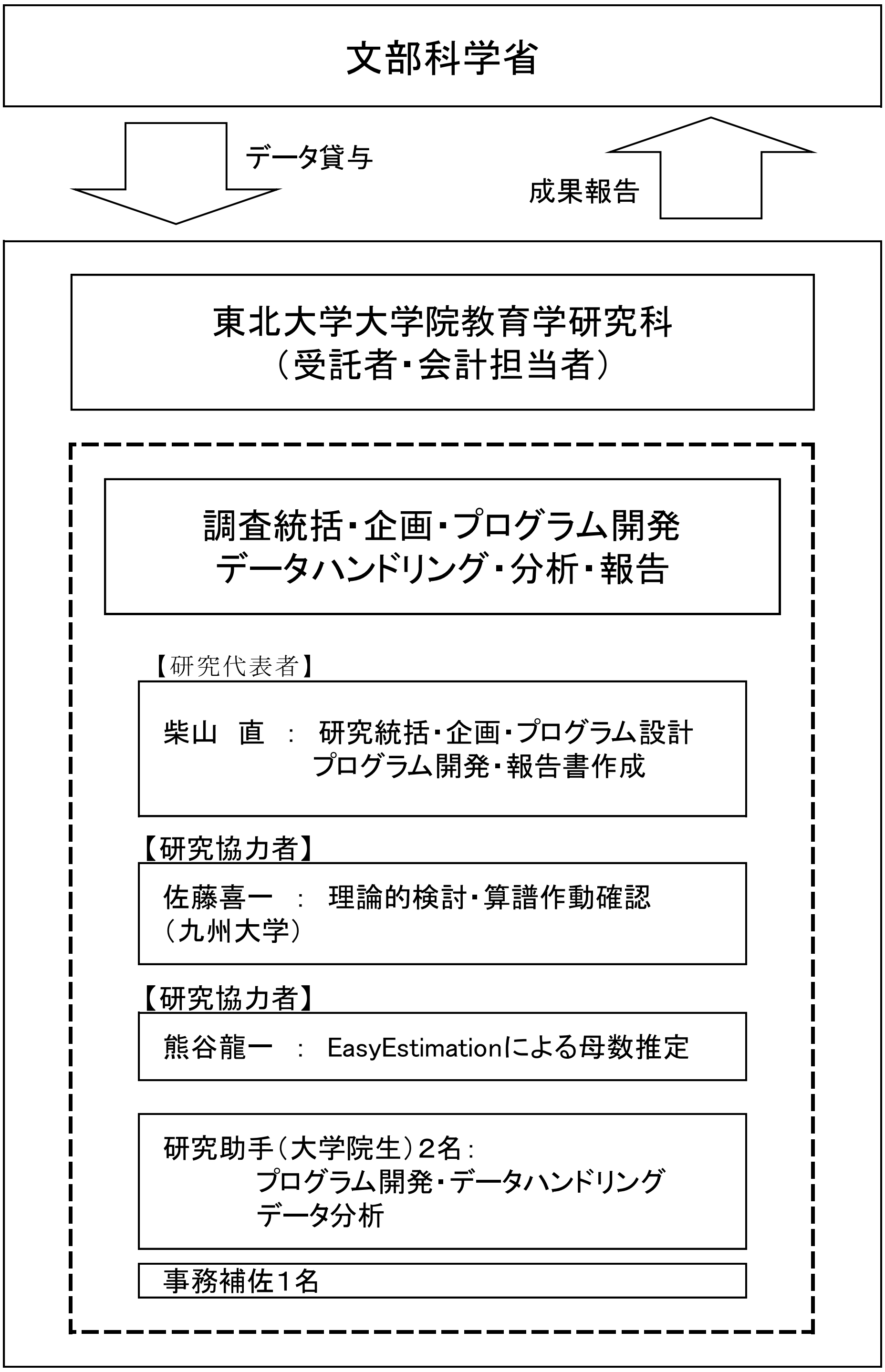
研究代表	柴山　直	東北大学大学院教育学研究科
研究協力	佐藤　喜一	九州大学基幹教育院人文社会科学部門
	熊谷　龍一	東北大学大学院教育学研究科
研究助手	澁谷　拓巳	東北大学大学院教育学研究科
	板宮　千尋	東北大学大学院教育学研究科
研究補助	江尻　大亮	東北大学教育学部

専門的指導助言

前川　眞一	東京工業大大学院社会理工学研究科
安永　和央	九州大学基幹教育院人文社会科学部門

事務担当　　紙屋　雅子

事業の実施体制図


文部科学省
データ貸与
成果報告
東北大学大学院教育学研究科
（受託者・会計担当者）
調査統括・企画・プログラム開発
データハンドリング・分析・報告
【研究代表者】
柴山　直　：　研究統括・企画・プログラム設計
プログラム開発・報告書作成
【研究協力者】
佐藤喜一　：　理論的検討・算譜作動確認
（九州大学）
【研究協力者】
熊谷龍一　：　EasyEstimationによる母数推定
研究助手（大学院生）２名：
プログラム開発・データハンドリング
データ分析
事務補佐１名

実施経過

実施日程	計画事項
6 月	H25 年度と H28 年度に実施された「全国学力・学習状況調査」,「経年変化分析調査」データならびに必要情報の貸与・提供を受ける。 データの前処理の開始 必要なプログラムの仕様・設計 リンキング，等化，R 言語等に関する文献収集
7 月	対応づけプログラムの開発開始 得点分布生成プログラムの関数定義作業開始 推算値生成プログラムの関数定義作業開始
8 月	対応づけプログラム等の開発継続
9 月	対応づけプログラム等のデバッギング
10 月	等化および対応づけ実験開始 専門的知識の第 3 者からの提供（於東京工業大学/10 月 12 日）
11 月	等化および対応づけ実験継続
12 月	等化および対応づけ実験結果の検討 進捗状況確認のための研究会（於九州大学/12 月 18 日）
1 月	追加提案手続の実施
2 月	追加提案手続の結果検討
3 月	報告書完成

目次

はしがき i

事業概要 ii

研究組織 ii

事業の実施体制図 iii

実施経過 iv

はじめに 1

1. 等パーセンタイル等化法のアルゴリズム 5

1.1 等化と対応づけ 5
1.2 等パーセンタイル等化法の概念 5
1.3 等パーセンタイル関数の理論的準備 6

2. Lord & Wingersky（1984）の Recursion Formula による復元得点分布の生成 9

2.1 IRT（項目反応理論）モデル 9
2.2 IRT モデルに基づく得点分布の生成 9
2.2.1 IRT true score とその分布の生成 10
2.2.2 IRT observed score とその分布の生成 11
2.3 復元得点分布の利用 14
2.4 R における計算アルゴリズム 15

3. フォンノイマン棄却法を用いた推算値（Plausivle Values）算出のアルゴリズム 17

3.1 推算値 17
3.2 VON NEUMANN の棄却法（REJECTION METHOD） 17
3.3 推定事後分布 18
3.3.1 EAP 推定値と周辺分布 18
3.3.2 MAP 推定値 19
3.4 推算値の利用 20
3.5 アルゴリズムの概略 20

4. 対応づけによる本体調査の年度間比較の実際 21

4.1 分析に利用したデータ 21
4.2 データ収集デザイン 22
4.3 対応表の作成手順 24
4.3.1 利用したデータとプログラム 24
4.3.2 等化実行前の準備 25
4.3.3 項目分析および項目母数の推定 25
4.3.4 共通項目法による異なる年度の経年変化調査の等化 25

4.3.5　共通受検者法による本体調査と経年変化調査の等化（対応づけ）.......... 26
4.3.5　IRT 母数による復元得点分布の生成 28
4.3.6　対応表の計算 29
4.4　年度間比較の実施 30
4.4.1　テストの信頼性と対応づけ可能性の検討 30
4.4.2　対応づけの実際 32

5. 推算値をもちいた下位領域ごとの得点比較の試み 38

5.1 下位領域ごとの年度間比較の難しさ 38
5.2 推算値利用の可能性 40
5.3 推算値を利用した下位領域ごとの年度間比較の適用例 42

文　献 47

R スクリプト 49

付録１　等パーセンタイル等化法の R プログラム 50
付録２　RECURSION FORMULA を用いた復元得点分布の生成のための R プログラム 53
付録３　フォンノイマン棄却法を用いた推算値算出のための R プログラム 63
付録４　機能テスト 68
等パーセンタイル等化関数“epe” 68
推算値発生プログラム“plausible” 73
復元得点分布生成プログラム“score.dist”, “dist.data” 83

資　料 90

資料１　平成 29 年 10 月 12 日東工大前川先生指導助言メモ 91
資料２　文部科学省委託研究九州大ミーティング発表資料 92
資料 2-1　研究会の概略 93
資料 2-2　等パーセンタイル等化法のためのプログラム 94
資料 2-3　Recursion Formula による復元得点分布の生成 99
資料 2-4　推算値（Plausible Values;PVs））.......... 105
資料３　H25 年度集団を H28 年度に復元した得点と H28 年度集団の素得点の度数分布表 109
資料 3-1　小学校国語 110
資料 3-2　小学校算数 111
資料 3-3　中学校国語 112
資料 3-4　中学校数学 113
資料４　対応づけ可能性分析 114
資料 4-1　平成 25 年度本体調査・経年変化分析調査の項目数・信頼性・相関 115
資料 4-2　平成 28 年度本体調査・経年変化分析調査の項目数・信頼性・相関（小学校）.......... 116
資料 4-3　平成 28 年度本体調査・経年変化分析調査の項目数・信頼性・相関（中学校）.......... 117
資料 4-4　平成 25 年度対応づけ得点の信頼性の下界と上界 118
資料 4-5　平成 28 年度対応づけ得点の信頼性の下界と上界（小学校）.......... 119

資料 4-6　平成 28 年度対応づけ得点の信頼性の下界と上界（中学校） 120
資料５　H25 年度得点を H28 年度得点に換算するための対応表 121
資料 5-1　対応表：小学校国語 122
資料 5-2　対応表：小学校算数 123
資料 5-3　対応表：中学校国語 124
資料 5-4a　対応表：中学校数学（前半） 125
資料 5-4b　対応表：中学校数学（後半） 126
資料６　復元得点分布を利用した全国における学力分布の年度間比較の例 127
表記法 128
資料 6-1　全国の年度間比較　小学校・国語 129
資料 6-2　全国の年度間比較　小学校・算数 129
資料 6-3　全国の年度間比較　中学校・国語 130
資料 6-4　全国の年度間比較　中学校・数学 130
資料７　復元得点分布を利用した都道府県（匿名化済み）における学力分布の年度間比較の例 131
資料 7-1　都道府県別の年度間比較　小学校・国語 132
資料 7-2　都道府県別の年度間比較　小学校・算数 140
資料 7-3　都道府県別の年度間比較　中学校・国語 148
資料 7-4　都道府県別の年度間比較　中学校・数学 156

執筆編集等分担 164

プログラミング等分担 164

経年変化分析調査との対応づけによる本体調査の年度間比較の試み

はじめに

全国学力・学習状況調査のうち，学力センサスとでもいうべき本体調査においては，指導目的等のため毎年問題が公開され，かつ実施対象となる児童生徒集団も年度ごとに入れかわっていくため，原理的には経年比較ができない構造になっている。この性質は，本体調査で得られたデータを利用しての総合的かつ客観的根拠に基づく政策立案（Evidence-Based Policy Making：EBPM）を推進するにあたっての越えるべき壁ともなっている（図１）。

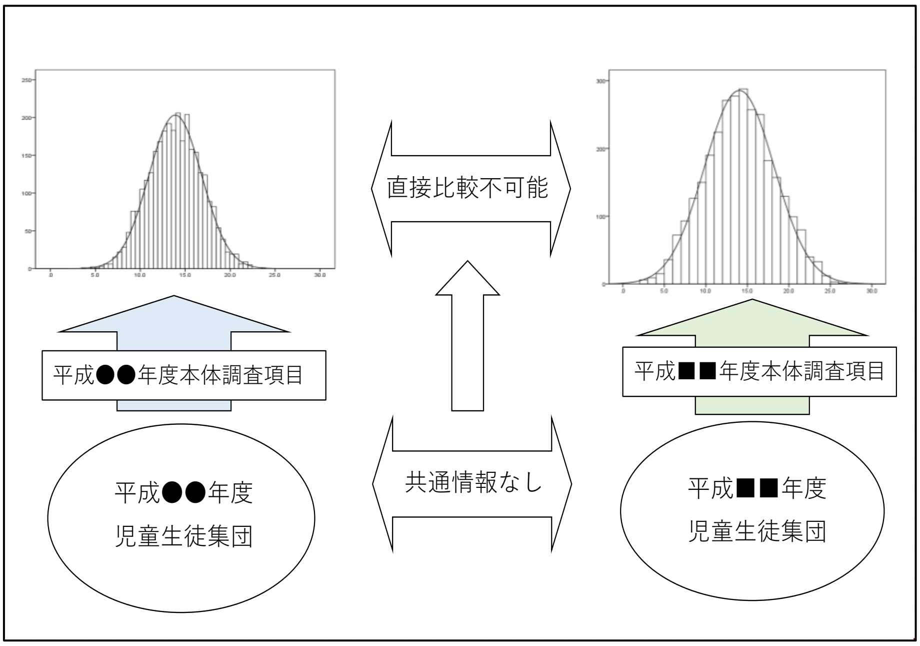


図１　全国学力学習状況調査（本体調査）の特徴

一方，平成 25 年度にプロトタイプとして導入され，２回目の平成 28 年度は本格実施となり，今後も「悉皆，かつ，毎年度実施する調査を補完する調査」として，保護者に対する調査と並んで標本抽出方式により「継続的，かつ，定期的に実施」されることとされている経年変化分析調査（図２）は，原理的に両年度の結果を比較することが可能な構造に設計され実施されている。すなわち，国全体の学力分布の変動を，両年度に共通して設定された尺度の上で測定することが実現されているのである。これは，PISA や TIMSS，NEAP 等でも採用されている，国際水準の大規模学力調査技術である重複テスト分冊法（Item Matrix Sampling Method）を採用し，かつ問題が非公開のため，共通項目デザインのもと平成 25 年度項目，平成 28 年度項目とも，基本的にはすべての項目が IRT（Item Response Theory：項目反応理論）モデルにより共通尺度上に等化可能であることによる。また，将来的に学習指導要領が改訂されても，注意深く出題問題を差し替えていくことで，この尺度自体は安定的に使えるため，今後も継続して経年比較が実現できる設計にもなっている。

本事業においては，両調査のもつそれぞれの利点と，項目反応理論（Item Response Theory：IRT）モデルや等化（equating），対応づけ（linking）などの学力調査の測定技術を組み合わせることで，

１）平成 25 年度児童生徒集団が仮に平成 28 年度本体調査を受検したとした場合に生成できる復元得点分布（IRT observed score distribution）による両年度間の学力分布の変化を捉える方法を開発し，

それに基づく，

２）平成 25 年度本体調査の得点を平成 28 年度本体調査規準の得点へ変換するための換算表（対応表：concordance table）を作成する分析手法を開発・提案する，

ことを目指す（表１）。このような換算表が準備できれば，全国学力・学習状況調査の本体調査であっても，例えば都道府県単位ごとの年度間比較を可能とすることが期待できる。

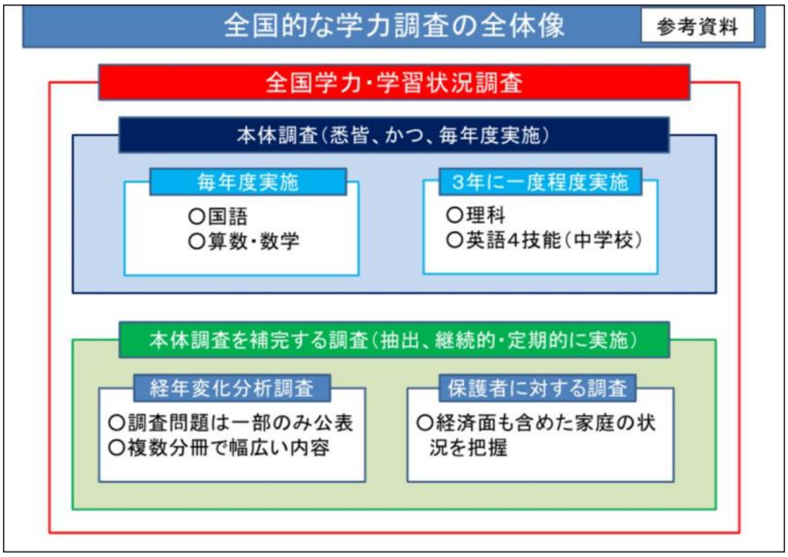


図２　全国的な学力調査の全体像[1]

H25年度得点	H28年度得点	H25年度得点	H28年度得点	H25年度得点	H28年度得点
0	0.0	10	13.2	20	23.1
1	0.5	11	14.5	21	23.9
2	1.6	12	15.7	22	24.7
3	2.9	13	16.7	23	25.4
4	4.7	14	17.8	24	26.1
5	6.8	15	18.8	25	26.8
6	8.3	16	19.8	26	27.4
7	9.6	17	20.7	27	28.0
8	10.8	18	21.5	28	28.0
9	12.0	19	22.3		

表１　全国学力・学習状況調査（本体調査）における年度間の対応表のイメージ

この目的を達成するための手順は図３のようになる。基本的には，

[1]「全国的な学力調査の今後の改善方策について「論点の整理」」参考資料，全国的な学力調査に関する専門家会議，文部科学省（平成 28 年６月 15 日）.

1）経年変化分析調査の項目を，IRT モデルを利用して共通尺度上に等化しておく。基準年度は平成 28 年度とする。
2）本体調査と経年変化分析調査をともに受検した集団を共通受検者集団とする。
3）共通受検者集団の情報を使って，本体調査項目を１）でもとめた共通尺度上に等化する。
4）共通尺度上で平成 25 年度本体調査の児童生徒の尺度値を推定する。
5）尺度値を使って IRT モデルを通して平成 25 年度受検者集団が H28 年度全国学力・学習状況調査を受検したと仮定して，平成 28 年度全体調査の得点分布を生成する（復元分布）。
6）平成 25 年度の得点分布（実データ分布）からその復元分布への対応づけおこない，平成 25 年度得点を平成 28 年度得点に変換するための対応表を求める。

というステップを踏む。この対応表を使って，各都道府県が手持ちの本体調査データを使って平成 25 年度得点を平成 28 年度得点に換算することにより，各都道府県は本体調査での年度間変化の様子が確認できる。

より具体的には，前述した共通項目デザインのもと，平成 25 年度ならびに平成 28 年度に実施された経年変化分析調査については IRT 等化済みの共通尺度を構成することを考える。一方，本体調査はセンサスであるためデータの中に経年変化分析調査に参加した児童・生徒の集団がそれぞれの年度において含まれている。これを二種類の調査に共通する集団という意味で共通受検者集団とよぶ。平成 25 年度における共通受検者集団と平成 28 年度における共通受検者集団には共通部分は存在しない。しかしながら，経年変化分析調査で使われたすべての項目の項目母数は IRT 等化済みの共通尺度上で表現されている。この性質を利用すれば IRT モデルを媒介として，両年度の本体調査の得点分布を互いに比較可能な形で復元生成することができる。さらに，これらの得点分布を対応づけ，そこで得られた換算表（対応表）を本体調査データに適用すれば，本体調査においても年度間比較が原理的には実現可能となる。等化デザインの観点からまとめれば，経年変化分析調査側で共通項目デザイン（common item design）のもと項目等化を行い，その項目情報を使って使って復元（予測）したテスト得点の分布に基づき，本体調査側で共通受検者デザイン（single group design）のもとテスト得点の等化を行っていることになる。

また，追加提案として，領域ごとの年度間比較の実現可能性も探った。これは指導の際に，全体としての学力の変動だけではなく，領域ごとの変動もきめ細かに知りたいという学校現場からの強いニーズに応えるための試みである。ただし，テスト全体にくらべて領域ごとに細分化すると項目数が著しく減少し，古典的テスト理論的な正答数得点はいうまでもなく，IRT モデルを単純に適用するだけでは能力母数（学力）の推定精度が低下ないし推定不能となるケースが頻発することは明白である。そのため推定精度の低下を補うものとして推算値を導入することを考え，得点分布ではなく IRT による尺度スコアに基づく集団統計量による要約を追加提案については目標とした。
なお，上記を実現するためには，１）対応づけ，２）IRT モデルに基づく得点分布の生成，３）推算値の生成，のそれぞれを実行することができるアルゴリズムの開発が必須であった。これらの開発は本調査研究遂行のための最重要部分であったといっても過言ではない。また，このアルゴリズムを

統計プログラミング言語である R で記述し，これらをすべて公開し第３者による質保証を受けることを可能とし，その上で広く利用されることを意図した。

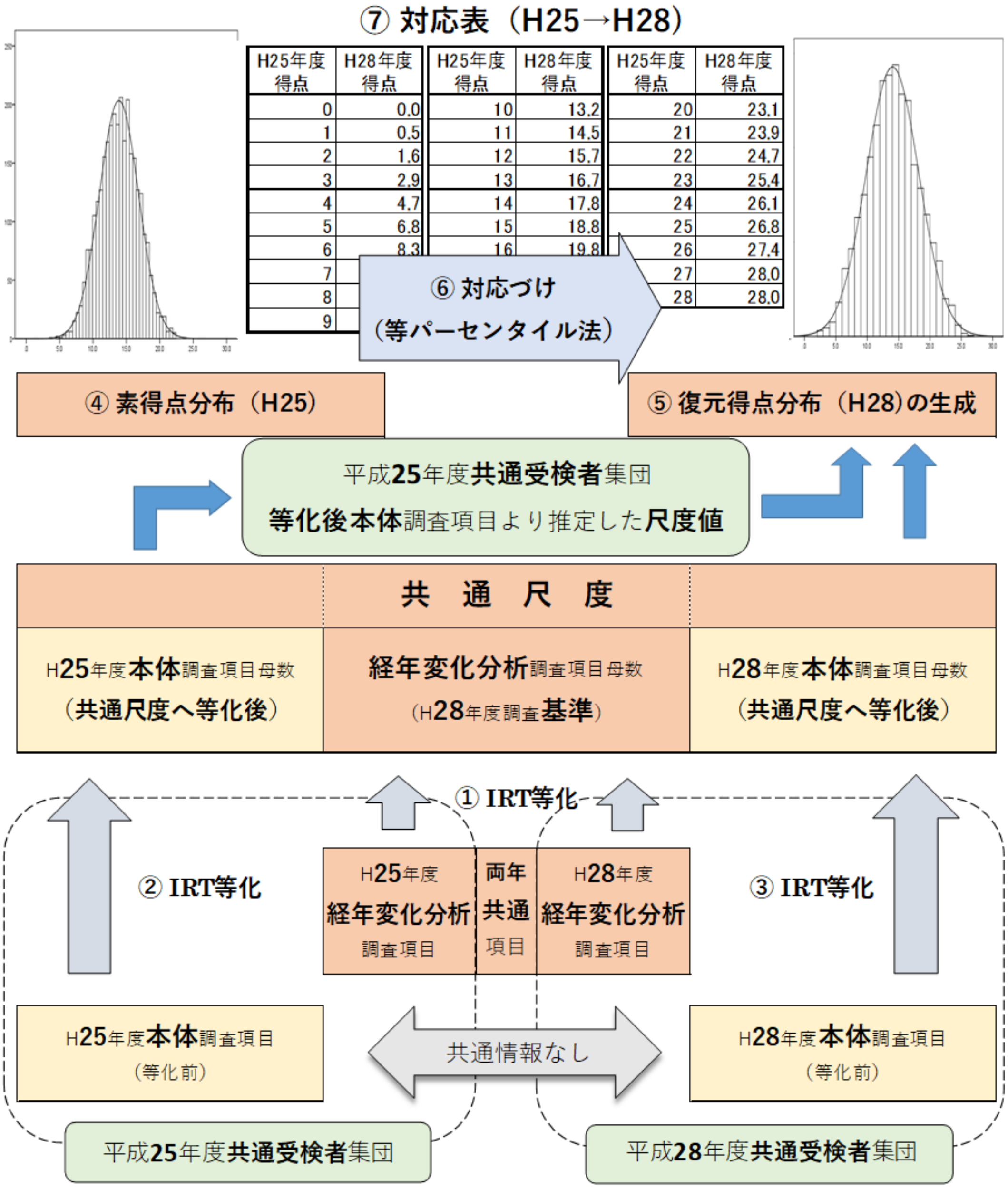


図３　開発・提案する手法の基本的な考え方（丸囲み数字は作業順を表す）

1. 等パーセンタイル等化法のアルゴリズム

1.1 等化と対応づけ

TOEFL や TOEIC のように年数回実施されるようなテストでは，受検した回ごとの得点が相互に交換可能でなければならない。これをテスト等化の問題という。等化を成立させる条件として，Dorans 他（2000）によれば，

a. 測定対象となる構成概念が同一であること（Equal Construct Requirement）
b. 信頼性が等しいこと（Equal Reliability Requirement））
c. 対称性が保たれていること（Symmetry Requirement）
d. どちらのテストを受けても公平であること（Equity Requirement）
e. 母集団普遍であること（Population Invariance Requirement）

の五つの条件をすべて満たす場合のみを等化とよぶ。

一方，これらの条件のいずれかを満たさない異なるテスト間の得点はもはや交換可能であるとはいえない。しかし，それでも全国学力学習状況調査におけるいわゆる全数調査として毎年実施される本体調査と，定期的に何年間に一度サンプリング調査として行われる経年変化分析調査では，その果たすべき役割や機能が異なってはいるが，測定したい学力の構成概念などは同一であることから，テストの得点は交換可能ではなくても，比較可能であるとされる。いわば等化の条件を緩めた上でのテスト間の得点を比較することが考えられる。これを対応づけ（linking）と呼んでいる。

具体的な等化法と対応づけの手法は，手続きレベルではほとんど同一であるが，いわば，その前提条件の違いによって同一手続きでも等化と対応づけが区別される。後述する手法も等パーセンタイル等化法（equipercentile equating）と呼んでいるが，前提条件の充足状況を厳密に勘案すれば，等パーセンタイル対応づけ（equipercentile linking）と呼ぶべきものである。すなわち経年変化分析調査の中においては厳密な意味での等化（この場合は IRT 等化）が行われているが，経年変化分析調査と本体調査の間は復元得点分布によって結びつけられ，その上で二つの年度のテスト得点が等化される。しかし本体調査であっても毎年問題数などに変化があるなど，その設計思想において厳密な意味での等化が意識されているとはいえず，本体調査の年度間比較を試みようとすれば，それは対応づけと呼んだ方が適切である。

ただし，手続きそれ自体を考えるときには，慣習上，わかりやすさを優先して，この報告書の中でも等化と呼ぶことにする。

1.2 等パーセンタイル等化法の概念

両年度の復元得点分布を対応づけするために Kolen & Brennan （2014）の等パーセンタイル等化法（等百分位等化法）を利用する。この手法はパーセンタイル順位（percentile rank）をもとに同じパーセンタイル順位の得点を対応づけする方法である。パーセンタイル順位とは受検者の集団内で得

点を昇順に下から並べていったときに，ある得点が下から数えて何パーセントの位置に相当するかを表す指標である。

異なるテストの得点比較においては，各テスト内での困難度の変化が一様に線形的な変化をするとは限らないため，線形等化（linear equating）よりも等パーセンタイル等化の方が一般的によく利用される。図 1 が等パーセンタイル等化の原理を示したイメージ図である。なお、両者の分布形が同じであるときは線形等化法と等パーセンタイル等化法は一致する。

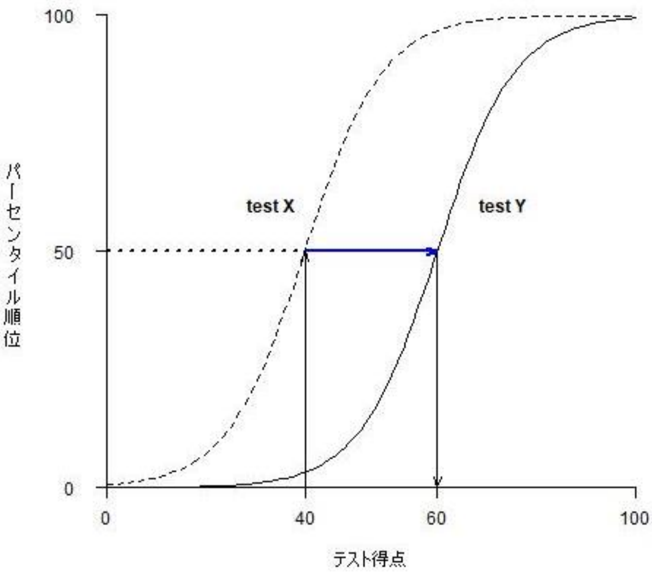


図 1.1　等パーセンタイル等化法の概念図

1.3　等パーセンタイル関数の理論的準備

あるテスト X のパーセンタイルに対応するテスト Y の得点を求めるための準備として，まずパーセンタイル順位を定義するための式を導入する。

テスト Xの項目数をm_Xとし，テスト Xによって得られる受検者の正答数得点をxとすると，$x = 0,1, \ldots\ , m_X$と表現できる。

次に受検生の数を$\mathrm{N_X}$，そのうち得点がxであった受検生の数をn_xとすると，離散密度関数（discrete density function）$f(x)$は,

$$f(x) = \frac{n_x}{N_X}\ , \qquad 1$$

で定義される。ただし,

$$f(x)\begin{cases} > 0, & x = 0,1, \ldots, m_X \\ = 0, & \text{それ以外} \end{cases}$$

である。そして，この離散密度関数の全項目における総和は,

$$\sum_{x=0}^{m_x} f(x) = 1\ , \qquad 2$$

である。さらにその離散累積分布関数は,

$$F(x) = \sum_{k=0}^{x} f(k) \ , \qquad 3$$

で定義できる。ただし，

$$0 < F(x) < 1, \qquad x = 0,1,\dots,m_X,$$

$$F(x) \begin{cases} = 0, & x < 0 \\ = 1, & x > m_X \end{cases}$$

である。

実際に受検者の項目反応データから得られる正答数得点は当然整数値に限られるが，テスト X のある正答数得点のパーセンタイルに対応するテスト Y のパーセンタイルの正答数得点は整数値で得られるとは限らない。そのため次に，xが整数値でない場合を考えて，四捨五入した値であるx^*を導入する。それは,

$$x^* - 0.5 \ \ \le x \ \ < x^* + 0.5 \ , \qquad 4$$

である。

パーセンタイル順位関数は以下のようになる。テストXの正答数得点xに対応するパーセンタイル順位関数を$P(x)$とすると，

$$P(x) = \begin{cases} 100\{F(x^* - 1) + [x - (x^* - 0.5)][F(x^*) - F(x^* - 1)]\}, \\ \quad -0.5 \le x \le m_X 0.5, \\ \quad 0, \qquad x < -0.5, \\ \quad 100, \qquad x \ge m_X + 0.5, \end{cases} \qquad 5$$

と定義される。

パーセンタイル順位関数はテストの正答数得点に対応するパーセンタイル順位を求めるための関数であった。次に，特定のパーセンタイル順位のパーセンタイルに対応するテスト得点を求めるためのパーセンタイル関数を定義する。パーセンタイル関数はパーセンタイル順位関数の逆関数P^{-1}で与えられる関数である。パーセンタイル順位p^*が与えられたとき，２種類のパーセンタイル関数が定義できるが，一部の場合を除いてどちらの定義式を用いても得られる値は同じである。

パーセンタイル順位p^*の範囲が$0 \le p^* < 100$で与えられているとき，p^*以上の累積パーセント 100 累$F(x)$に対応するxのなかで最小の整数値をx_U^*とすると，求めるパーセンタイル関数の定義式は,

$$x_U(p^*) = \frac{p^*/100 - F(x_U^* - 1)}{F(x_U^*) - F(x_U^* - 1)} + (x_U^* - 0.5) \ , \qquad 0 \le p^* < 100,$$

$$= m_X + 0.5 \ , \qquad p^* = 100, \qquad 6$$

である。

また，パーセンタイル順位p^*の範囲が$0 \le p^* < 100$で与えられているとき，p^*以下の累積パーセント 100 累$F(x)$に対応するxのなかで最大の整数値をx_L^*とすると，求めるパーセンタイル関数の定義式は,

$$x_L(p^*) = \frac{p^*/100 - F(x_L^*)}{F(x_L^* + 1) - F(x_L^*)} + (x_L^* + 0.5) \ , \qquad 0 < p^* \le 100,$$

$$= -0.5 \ , \qquad p^* = 0, \qquad 7$$

である。

このとき，もし，$f(x)=0$となるような得点xが存在している場合，つまり，全受検者において誰もその得点の受検者は存在しないような得点が存在する場合，$x_U \neq x_L$となる。この様な場合は慣例的に$x=(x_U+x_L)/2$という値がパーセンタイルとして用いられる。

次に，これまで求めてきたテスト X に関する手順を，同様にテスト Y にも適用する。このときテスト Yの項目数を$\mathrm{m_Y}$，テスト Yによって得られる正答数得点をyとおく。そして，離散密度関数を$g(y)$，離散累積分布関数を$G(y)$，パーセンタイル順位関数を$Q(y)$，パーセンタイル関数を$Q^{-1}(y)$とおく。

これらの値を用いて用いてテスト X の正答数得点xを，テスト Y の同じパーセンタイル順位に位置する得点と対応づけする。そのテスト X の正答数得点xをテスト Y の尺度上に変換するための関数$e_Y(x)$をテスト X のテスト Y への等化関数（equating function）と呼ぶ。この等化関数は，

$$e_Y(x)=y=Q^{-1}[F(x)]\ ,\quad -0.5\leq x\leq m_X+0.5\ , \tag{8}$$

で定義できる。この式 8 を，式 6 を利用して解析的に解くと，

$$\begin{aligned} e_Y(x) &= Q^{-1}[F(x)] \\ &= \frac{P(x)/100-G(y_U^*-1)}{G(y_U^*)-G(y_U^*-1)}+(y_U^*-0.5), \qquad 0\leq P(x)<100, \\ &= m_X+0.5, \qquad\qquad P(x)=100, \end{aligned} \tag{9}$$

となる。

なお，理論的にはこの等化関数はテスト X とテスト Y に関して対称な性質を持っており，

$$\begin{aligned} e_Y(x)&=e_X^{-1}(x)\quad , \\ e_Y^{-1}(y)&=e_X(y)\quad , \end{aligned} \tag{10}$$

と表現することができ，そして，

$$e_X(y)=F^{-1}[G(y)]\quad ,\quad -0.5\leq y\leq m_Y+0.5\quad , \tag{11}$$

であり，この等化関数はテスト Y の正答数得点yをテスト X の尺度上に変換するための関数である。

以上が，等パーセンタイル等化法の計算アルゴリズムである。これを R の計算アルゴリズムとして書き直す。その際，必要な計算は，

① 度数，離散密度，累積離散密度，を計算する関数を作成し，

② ①の値を用いてパーセンタイル順位関数を作成し，

③ ②の逆関数としてのパーセンタイル関数を作成し，

④ テスト Y のパーセンタイル関数にテスト X のパーセンタイル順位関数を代入する関数＝等化関数を求め，

⑤ テスト X の得点とテスト X のテスト Y への等化関数を表にまとめた行列の出力する，

という五つのステップとなる。

2. Lord & Wingersky（1984）の Recursion Formula による復元得点分布の生成

テスト項目・受検者集団のどちらも異なる学力調査の結果を対応づけすることを目指し，IRT（項目反応理論）モデルを介して得点の度数分布を復元することを考える。本章では，推定した項目母数と能力母数にもとづき，Lord & Wingersky（1984）の Recursion Formula（再帰式）を利用して得点分布を生成する。なお，このようにして得られた得点分布のことを英語ではIRT observed score distribution（直訳すると「IRT 観測得点分布」），predicted number-correct score distribution（「予測正答数得点分布」），expected score distribution（「期待得点分布」）等とよぶが，本研究においては直感的なわかりやすさを優先して「復元得点分布」と表現することとする。

2.1　IRT（項目反応理論）モデル

項目反応理論では，受検者が各項目に正答する確率を，測定しようとしている特性値（尺度値）の関数として表すことで，それらの項目の特徴を表現する。この関数は，項目反応関数および項目特性曲線とよばれている。項目特性曲線のモデルには，項目母数の個数の仮定による違いから，１母数・ロジスティックモデル，２母数・ロジスティックモデル，３母数・ロジスティックモデル，といったバリエーションが存在する。本研究では，項目母数として項目識別力母数a_jと項目困難度母数b_jの二つをもつ２母数・ロジスティックモデル（２PLM）を扱い，また，特性値として能力を表す尺度値θを用いる。今，項目番号を j としたとき，項目特性曲線 $P_j(\theta)$は，定数 D=1.7 として，

$$P_j(\theta) = \frac{1}{1+\exp(-Da_j(\theta - b_j))} \quad , \qquad -\infty < \theta < \infty, \qquad (1)$$

と表される。例えば，$a_j = 1.0$，$b_j = 0.0$とした場合の項目特性曲線をグラフで表すと，図１の通りとなる。

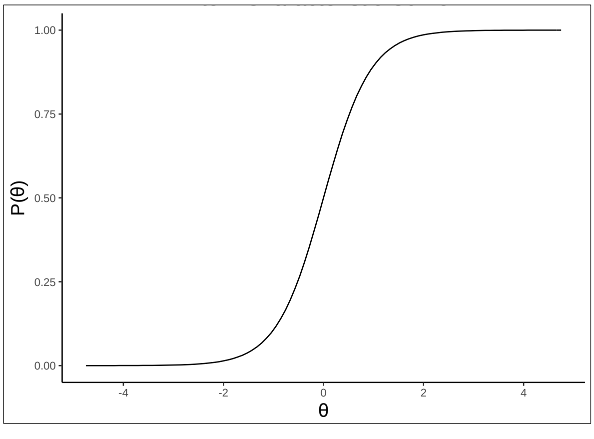


図 2.1　項目特性曲線の例

2.2　IRT モデルに基づく得点分布の生成

項目特性曲線において，$P_j(\theta_i)$は任意の能力θ_iをもつ受検者が項目 j に正答する確率を表している。この$P_j(\theta_i)$を用いることで，IRT モデルを介した得点として，任意の能力θ_iの受検者のテスト

得点を推定することができる。得点の具体的な算出方法には，期待テスト得点として真の得点（IRT true score）を求める方法と，全ての項目反応パターンを考慮することで観測得点（IRT observed score）を求める方法の２種類がある。

2.2.1　IRT true score とその分布の生成

真の得点（IRT true score）は，任意の能力 θ_i の受検者の平均的なテスト得点，すなわち能力 θ_i の受検者の得点の期待値である。今，項目数 J のテストにおいて，ある項目に正答したときは１点，誤答したときは０点とし正答数得点を数える場合，能力 θ_i の受検者がとる得点の期待値を考える。得点を x，項目番号を j としたとき，能力 θ_i の受検者の期待テスト得点 $x(\theta_i)$は以下のように表すことができる。

$$x(\theta_i) = \sum_{j=1}^{J}\{1 \times \mathrm{P_j}(\theta_i) + 0 \times (1 - \mathrm{P_j}(\theta_i))\} = \sum_{j=1}^{J} \mathrm{P_j}(\theta_i), \qquad (2)$$

（2）式からわかるように，IRT モデルにおける true score は，能力 θ_i の受検者の正答確率 $\mathrm{P_j}(\theta_i)$ を項目について足し合わせたものに等しい。

次に，IRT true score の度数分布を生成することを考える。IRT true score は，能力 θ_i の受検者に対し，ただ一つ求められる値である。つまり，IRT true score における度数分布は，能力 θ_i をもつ受検者の度数に等しくなる。よって，能力 θ_i の受検者がとりうる得点 x の度数を$f(x|\theta_i)$，受検者全体の能力分布を$\psi(\theta)$としたとき，以下の関係が成り立つ。

$$f(x|\theta_i) = \psi(\theta_i), \qquad (3)$$

$$f(x) = \psi(\theta), \qquad (4)$$

以上のようにして IRT true score の度数分布を生成する過程をイメージ図で表すと，図２のようになる。また，ある県（A 県と呼称）で実際に行われた学力調査[2]のデータを使ってヒストグラムを描画したところ，図３の通りとなった。なお，能力分布には MAP 推定法により推定した能力母数を用いた。

[2] A 県独自の調査であって全国学力学習状況調査のものではない。研究目的のためのデータ利用については A 県教育委員会より許諾済み。