

令和 2 年 6 月 6 日現在

機関番号：11301

研究種目：基盤研究(B)（一般）

研究期間：2016～2019

課題番号：16H03731

研究課題名（和文）発達段階をトレースできる到達度評価のためのIRT垂直尺度構成の試み

研究課題名（英文）Development of vertical scale for measuring student growth in multiple level achievement tests

研究代表者

柴山 直（Shibayama, Tadashi）

東北大学・教育学研究科・教授

研究者番号：70240752

交付決定額（研究期間全体）：（直接経費） 13,700,000円

研究成果の概要（和文）：的確な学習指導のためには、学習者の理解度や習熟度、そこに至る個々人の学習プロセスや発達段階の把握は必須の条件である。本研究では、対象集団に小学校4年生から中学校2年生の5学年、対象科目に算数/数学、国語を取り上げ、コホートの調査デザインのもと、学力の発達をトレースできる到達度尺度の構成を、項目反応理論(IRT)モデルにもとづく垂直尺度化により実現した。構成された尺度を使うことにより小学校4年生から中学校2年生までの学力発達を追跡することが可能となった。

研究成果の学術的意義や社会的意義

児童生徒の学力を評価する場合、数年にわたる教育課程の中で、個々の子どもたちがどの程度成長するかを数値化してモニターすることは学習指導上重要である。その際、年度ごとの相対評価ではなく、個々の児童生徒がその子の中でどれだけ成長できたかの到達度評価であることが望ましい。本研究で開発された垂直尺度は、学年を越えて子どもたちの学力を追跡できるメリットをもつ。その範囲は小学校4年生から中学校2年生におよぶ。公開予定のデータを使えば学力の成長に伴う質的変化の考察などの研究への2次的成果も期待できる。

研究成果の概要（英文）：In order to effectively teach students, it is essential to understand where they are located in the developmental stage using developmental scale scores or vertical scale. In this study we developed vertical scale for measuring student growth in multiple level achievement tests using IRT model as psychometric model. The target groups are from the 4th grade elementary school to the 2nd grade of junior high school. A cohort design was used as the survey design. The survey period was three years from 2017 to 2019. By using the developed scale, it is possible to trace the academic development from the 4th grade of elementary school to the 2nd grade of junior high school.

研究分野：教育測定学

キーワード：IRT 垂直尺度 発達 学力 項目反応理論 項目応答理論 到達度評価 形成的アセスメント

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

同じ構成概念で定義される心理学的な特性(学力や知的能力、適性、パーソナリティなど)を測定するテスト内に、複数の冊子(版、フォーム、分冊ともよばれる)が存在するとき、異なる冊子から得られたスコアであっても、互いに比較ができるようにしておく必要がある(例:TOEFL)。これを一般にテスト得点の等化(equating)とよぶ。また、定義される構成概念が同等であっても、設計仕様が互いに異なるテスト間のスコアに対してこの操作を行う場合を対応づけとよぶ。さらに、この2つを含む上位概念としてリンクング(linking)という用語が使われることがある(Holland & Dorans,2006; Dorans,Pommerich & Holland,2007; 柴山他,2012)。

1) 応募者がまず対応づけの問題に直面したのが、法科大学院入学者選抜試験においてであった。そこでは日弁連法務研究財団と大学入試センターが互いに異なる設計仕様で実施していた2種類の適性試験が存在し、そのスコアの比較を等パーセントイル等化法の援用によって実現した。この手法は両試験が統一実施される2011年まで利用された(柴山・野口,2004)。また、実施団体が法科大学院協会、公益財団法人日弁連法務研究財団、社団法人商事法務研究会が共同して設置する適性試験管理委員会に統合された2012年以降は、この方法をさらに発展させた手法が佐藤・柴山(2011)によって提案され、年度内2回実施のテストのスコアを共通尺度上で表現するための等化方法(分割等化法)として現在もハイスタークスの現場で実用に供されている。

2) その一方で、原理的には学力分布の経年比較ができない文部科学省全国学力学習状況調査の調査方法の抜本的な改善のため、応募者らの研究グループは、IRT等化ベースの重複テスト分冊法の導入(柴山他,2011)、経年比較と国語の記述問題へのIRT対応の実現(柴山他,2012)、釣合い型不完備ブロック計画(BIBD; Balanced Incomplete Block Design)や推算値(Plausible Values)の採用による、PISA等でも採用されている本格的なマトリックス・サンプリングの実現(柴山他,2013)、さらには、そこで開発された調査方法に基づく東日本大震災の学力への影響の経年比較分析(柴山他,2014)等の一連の開発・適用経験の積み上げを行い、その成果は、2018年度より文部科学省の経年比較調査の基盤手法として正式に採用されることとなった。

3) この間、柴山・千葉(2015)、佐藤・柴山(2014,2013)らは、「パフォーマンス評価におけるIRT尺度を利用した信頼性向上のための基礎研究(科研基盤(C))」を遂行する中で、学年をまたいだ学力の伸びをトレースする必要性に迫られ、理科と算数とにおいて、新潟県全県学力調査で得られたスコアが学年間で比較できるように、IRTにもとづく尺度の構成を探索的に試みた。この観点からの尺度の構成は、たとえば、同じ数学的尺度で学力をトレースするとはいっても、算数/数学の学習内容が小学生と大学生とは質・量・難度とも大きく異なっていることなどから、上の等化とは区別して、特に垂直尺度化(Vertical Scaling)と呼ばれる。

2. 研究の目的

類似のアイデアとしては、国外においては、たとえばPISAにおける習熟度尺度(Proficiency Scale)などに見られるが、これらは、そもそも、集団統計量による国レベルの比較が目的であって個人スコアに着目しているわけではない。また、個人の学力の発達をトレースできるように設計されているわけでもない。一方、国内においては新しいテスト理論としてIRTモデルが注目され始めた頃から、IRTのもつ柔軟性を生かした、相対評価にかわる到達度評価のための尺度構成の方法として、当初は垂直等化(Vertical Equating)という用語によって理解され期待されてきた。

しかしながら、その実現には、国内における教育界へのIRTモデルの普及と、それと密接に関連した等化や対応づけ、BIBID、推算値といった周辺技術の開発、垂直等化から垂直尺度化への概念の変遷に見られるように、この方法に関する本質的理解の深化を待たねばならなかった。本研究の目的は、このような経過の中で、「発達段階をトレースできる到達度評価のためのIRT垂直尺度の構成」に着手し、これが実現することにある。

3. 研究の方法

平成28年度は1)研究体制の確立、2)先行研究に関する文献調査、3)研究遂行に必要なソフト等の準備、4)テスト設計・分冊設計、5)研究補助者等の雇用、6)垂直尺度化に関する理論的検討、7)最適な垂直尺度構成のための項目配置の決定、8)次期指導要領改定にともなう項目の入れ替えのメカニズム構築は問題項目の作成・開発も含めていずれも計画通り実施できた。具体的には以下の通り。2)については平成29年度日本教育心理学会第59回総会にて報告の予定である。また、6)については佐藤・柴山(2016)および坂本・柴山(印刷中)により公開した。佐藤・柴山(2016)では「IRTモデルにもとづく学力評価ルーブリックの作成手法の試み 小学校算数・面積学習領域を対象として」をテーマに、学力テストとは別にパフォーマンス課題を用意し、「どの学力レベルにいる者が具体的にどのような解答のふるまいをするか、どこ

でつまづきやすいのか」という学力特性値()のレベルと実際の解答パフォーマンスとの関連について探索的検討を行い、IRT モデルにもとづく学力評価ルーブリック作成のための手法を開発し、学力レベルと実際の解答パフォーマンスとの関連について検討し、学力評価のためのルーブリック作成の手法を提案できた。坂本・柴山(印刷中)では「学力テストの下位領域に関する多次元 IRT 分析」をテーマに、「テスト(項目)はそもそも何を測っているのか」という構成概念に関する精緻な検証方法が求められていることを念頭に、多次元 IRT を使って下位領域による項目への影響について定量的に検証した例により、多次元 IRT 分析にもとづくより精緻な妥当性検証が可能であることを示すことができた。

平成 29 年度は前年度に引きつづき各研究ユニットの課題を遂行した。問題開発、コホートデザインの調整、テスト仕様の決定などの昨年度までの準備に基づき、本年度より調査が本格的に実施された。6 月から 7 月にかけては、調査対象校の設置者たる自治体の教育委員会に調査協力依頼と実施の具体的な打ち合わせを行った。項目のテストへの組み上げ、問題冊子の印刷、マトリックス・サンプリング手法に則した特殊な実施ノウハウが必要なため、実施にあたっては、外注により専門企業のサポートをうけた。実施対象の自治体数は 2、小学校数 7 校、中学校数 5、参考の小学校 4 年生 415 名、小学校 5 年生 397 名、小学校 6 年生 399 名、中学校 1 年生 433 名、中学校 2 年生 432 名であった。実施時期は 10 月から 11 月にかけての学校行事とぶつからない日に設定した。なお、コホート・デザインによりこの年度の実施対象学年が一番多い。実施後採点作業を開始したが、記述式問題の採点基準の見直し作業なども入り、見込みよりもデータ化するのに時間がかかり、最終的な結果が納品されたのは平成 30 年 2 月末となった。また、答案用紙そのものは今後の検討に資するためすべて画像情報に変換した上で保存した。その後、年度内に得られた結果の初回の項目分析および IRT 分析によるひとまずの垂直尺度の構成を行った。成果の公表としては 10 月に澁谷・柴山(2017)、また本研究より派生的に得られた成果として、多次元 IRT モデルにもとづく学力構造の下位領域分析を坂本・柴山(2017)として公刊した。

平成 29 年度調査で得られたデータは 1 年分のみであるため、平成 30 年度においては小学校 4 年生から中学校 2 年生の 5 学年横断型の分析を行った。その結果は教育心理学会第 60 回総会にて研究発表をおこなった。また、IRT モデルからの能力値の分布だけでは、協力校・教育委員会にとってはわかりにくいいため、IRT observation scores による等化をおこない、参加児童・生徒が仮に 5 学年全部のテストを受けたとしたときの正答数得点分布を生成するアルゴリズムを開発し、その成果も教育心理学会第 60 回総会で報告した。この成果は「素得点実分布の IRT observed score による復元分布と IRT true score による復元分布の精度比較」として投稿・印刷中である。さらに、次年度以降に必須となる IRT 垂直尺度化に適したサンプル数ならびに尺度調整に関する系統的かつ詳細・包括的なシミュレーション研究をおこない、日本テスト学会第 16 回大会にてその結果を発表した。あらたなデータ収集については、昨年度実施とは異なる内容をもつ分冊のもとで、小学校 5 年生から中学 2 年生に対して計画通り調査を実施した。調査対象校は昨年度と同一のため参加児童・生徒を個別に追跡できるコホートデザインとなっている。ただし、個人情報保護のため、ID と個人情報との対照表自体は各学校のみ保持して、研究組織側では ID のみの情報しか分からないようにした。なお、日本テスト学会第 16 回大会における柴山(2018)「対応づけの開発と小論文採点事業化の試み、その後の展開」、熊谷(2018)第 13 回 ICPSR 国内利用協議会・統計セミナーにおける「項目反応理論入門」および、第 12 回『日本テスト学会賞』記念講演・ワークショップにおける熊谷(2018)「EasyEstimation を用いたテストデータの IRT 分析」は本研究の派生的な成果である。

4. 研究成果

研究最終年度においては、4 月～9 月期を調査実施準備とし、1) 小学校 6 年生から中学校 2 年生向けのテスト項目の内容精査、2) 実施協力校への昨年度結果のフィードバック、3) 実施協力校の確定、4) 実施対象校の実情に合わせた調査デザインの調整、5) 同じく重複テストデザインの調整、を行った。また、10 月～3 月期を調査実施・データ分析・協力校へのフィードバックにあて、6) 調査実施、7) データ集計と採点結果の数値化作業、8) データ分析、9) 分析結果の検討、10) 協力校への速報によるフィードバックを予定していたが、新型コロナウイルスによる感染拡大防止対応のため、7) までしか完了できなかった。本年度で研究期間は終了するが、8) 以降の計画は引き続き進めていく予定である。

しかしながら、これまでの研究成果としては、本研究で開発された垂直尺度が、学年を越えて子どもたちの学力を追跡できるメリットをもつこと、その範囲は小学校 4 年生から中学校 2 年生におよぶことが指摘できる。また、公開予定のデータを使えば学力の成長に伴う質的変化の考察などの研究への 2 次的成果も期待できる。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 柴山直・澁谷拓巳・板宮千尋	4. 巻 67
2. 論文標題 素得点実分布のIRT observed scoreによる復元分布とIRT true scoreによる復元分布の精度比較	5. 発行年 2019年
3. 雑誌名 東北大学大学院教育学研究科研究年報	6. 最初と最後の頁 印刷中
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 坂本佑太郎・柴山直	4. 巻 6
2. 論文標題 学力テストの下位領域に関する多次元IRT分析	5. 発行年 2017年
3. 雑誌名 データ分析の理論と応用	6. 最初と最後の頁 31-45
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 坂本佑太郎・柴山直	4. 巻 6
2. 論文標題 学力テストの下位領域に関する多次元IRT分析	5. 発行年 2017年
3. 雑誌名 データ分析の理論と応用	6. 最初と最後の頁 印刷中
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 佐藤誠子・柴山直	4. 巻 -
2. 論文標題 IRTモデルにもとづく学力評価ルーブリックの作成手法の試み	5. 発行年 2016年
3. 雑誌名 日本テスト学会第14回発表論文抄録集	6. 最初と最後の頁 140-143
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 坂本佑太郎・柴山直	4. 巻 -
2. 論文標題 学力テストの下位領域に関するIRTを使った測定論的特徴の把握	5. 発行年 2016年
3. 雑誌名 日本教育心理学会第58回総会発表論文集	6. 最初と最後の頁 PB87
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 1件 / うち国際学会 0件）

1. 発表者名 板宮千尋, 澁谷拓巳, 柴山直
2. 発表標題 IRT observed scoreによる復元得点分布産出プログラムの開発とその応用について
3. 学会等名 日本教育心理学会第60回総会, 慶応大学
4. 発表年 2018年

1. 発表者名 澁谷拓巳, 柴山直
2. 発表標題 異なる学年で実施される学力テスト項目のIRT垂直尺度化
3. 学会等名 日本教育心理学会第60回総会, 慶応大学
4. 発表年 2018年

1. 発表者名 澁谷拓巳, 柴山直
2. 発表標題 学力テストのIRT垂直尺度化に適したサンプル数と尺度調整法の検討
3. 学会等名 日本テスト学会第16回大会, 東京家政大学
4. 発表年 2018年

1. 発表者名 柴山直
2. 発表標題 対応づけの開発と小論文採点事業化の試み、その後の展開 - 公平性・公正性・透明性とテストの品質保証 -
3. 学会等名 日本テスト学会第16回大会、東京家政大学（招待講演）
4. 発表年 2018年

1. 発表者名 熊谷龍一
2. 発表標題 項目反応理論入門
3. 学会等名 第13回ICPSR国内利用協議会・統計セミナー，大阪大学
4. 発表年 2018年

1. 発表者名 熊谷龍一
2. 発表標題 EasyEstimationを用いたテストデータのIRT分析
3. 学会等名 第12回『日本テスト学会賞』記念講演・ワークショップ，明治学院大学
4. 発表年 2018年

1. 発表者名 澁谷拓巳・柴山直
2. 発表標題 学力の発達を追跡するための垂直尺度化について
3. 学会等名 日本教育心理学会第59回総会
4. 発表年 2017年

1. 発表者名 佐藤誠子・柴山直
2. 発表標題 IRTモデルにもとづく学力評価ルーブリックの作成手法の試み
3. 学会等名 日本テスト学会第14回大会
4. 発表年 2016年

1. 発表者名 坂本佑太郎・柴山直
2. 発表標題 学力テストの下位領域に関するIRTを使った測定論的特徴の把握
3. 学会等名 日本教育心理学会第58回総会
4. 発表年 2016年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

濹谷拓巳.(2019).異なる難易度のテスト項目のIRT垂直尺度化 尺度化テストデザインによる垂直尺度構成 .東北大学.修士(教育学).学位論文 http://hdl.handle.net/10097/00125343

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	佐藤 喜一 (Sato Yoshikazu) (00300517)	九州大学・アドミッションセンター・教授 (17102)	

6. 研究組織（つづき）

	氏名 (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	熊谷 龍一 (Kumagai Ryuichi) (60422622)	東北大学・教育学研究科・准教授 (11301)	
研究分担者	川端 一光 (Kawahashi Ikkou) (20506159)	明治学院大学・心理学部・准教授 (32683)	
研究分担者	斉田 智里 (Saida Chisato) (50400594)	横浜国立大学・教育学部・教授 (12701)	
研究分担者	野口 裕之 (Noguchi Hiroyuki) (60114815)	名古屋大学・教育発達科学研究科・名誉教授 (13901)	
研究分担者	足立 幸子 (Adachi Sachiko) (30302285)	新潟大学・人文社会科学系・准教授 (13101)	
研究分担者	佐藤 誠子 (Sato Seiko) (20633655)	石巻専修大学・人間学部・助教 (31308)	
研究分担者	藤本 亮 (Fujimoto Akira) (80300474)	名古屋大学・法学研究科・教授 (13901)	
研究協力者	安永 和央 (Yasunaga Kazuhiro)		