

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 6月 11日現在

機関番号：62601

研究種目：若手研究（B）

研究期間：2010～2012

課題番号：22730652

研究課題名（和文） カリキュラムに基づいた学習の習得を測定する大規模テストと分析モデルとの対応の研究

研究課題名（英文） A research on the relationships between large-scale assessment of the achievement based on the curriculum and the analysis models

研究代表者

萩原 康仁（HAGIWARA YASUHITO）

国立教育政策研究所・教育課程研究センター基礎研究部・総括研究官

研究者番号：30373187

研究成果の概要（和文）：カリキュラムに基づいた学力調査において、異なるカリキュラムの状況を背景を持った集団間の比較を行う際に、この違いを考慮した分析を行った。国内の学力調査の分析の結果、調査実施の時期が項目の特性にわずかではあるが影響を与えていることが示唆された。国際的な学力調査の分析の結果、特に知識領域において、項目の内容が国のカリキュラムに含まれていないことに伴って解きにくい状況が見られ易くなることが示唆された。

研究成果の概要（英文）：In this study, the achievement data were analyzed, taking the curricular differences between groups into account. The results of the data analysis within the country showed that the time of the testing would affect the characteristics of the items to a minor extent. The results of the international achievement data analysis revealed that the effect of the coverage of the national curriculum on the item difficulty was particularly apparent in the domain of knowing.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	500,000	0	500,000
2011年度	500,000	0	500,000
2012年度	500,000	0	500,000
年度			
年度			
総計	1,500,000	0	1,500,000

研究分野：教育評価

科研費の分科・細目：教育学・教育学

キーワード：教育評価、テスト、カリキュラム

## 1. 研究開始当初の背景

学力をカリキュラムに基づいた学習の習得の程度と定義した上で、異なるカリキュラムの状況を背景を持った集団間の学力の比較を行う際に、こうした違いを考慮して行うことの妥当性を示すことを、研究開始当初の動機とした。

## 2. 研究の目的

本研究では、学校（学級）－児童生徒のよりに階層的な構造を持った学習に関するテストの既存のデータの再分析を行う。その際には、項目反応モデル等の分析モデルについて

て、カリキュラムの実施状況等について何らかの形で考慮できるように拡張されたモデルを用いる。これによって、カリキュラムの実施状況等による背景の違いを考慮した分析モデルの妥当性について検討することを主として目指すものである。

### 3. 研究の方法

具体的には、主として以下の二つについて検討した。

(1) テストで出題される内容について学習される時期の違いに着目した研究を行った。具体的には、受検する当該年度までに学習済みの内容の場合と、当該年度で学習する場合とで、出題された各項目の困難度に関する母数において、学校レベルでの散らばりの大きさに違いが見られるかを検討した。

平成18年度に行われた学力調査の算数・数学のテストのデータを用いた。対象とした学年は、小学校第2学年・小学校第5学年・中学校第2学年であった。この調査は同一年度内に2時点にわたり実施されたものであった。1時点目は6～7月に、2時点目は12月に行われた。前者は、前学年における学習内容が範囲であり、後者は、当該学年の2学期まで（実施時期が12月のため、特に11月まで）に履修すると考えられる学習内容が範囲であった。2回分の調査で実施された項目について、領域や内容、あるいはテストの形式や項目の数に関して各学年の中で条件をそろえるようにした。このため、選択式の項目や文章題、及びその項目においてのみ他の領域の知識を明らかに要するようなものを分析の対象外とした。その上で、両時点ともに項目の数が同程度多く確保される領域や内容について焦点化した。この結果、小学校第2学年においては求答式の計算の項目、小学校第5学年においては求答式の図形の項目、中学校第2学年においては求答式の数と式の項目に焦点化した上で分析を行った。

分析には、児童生徒の能力特性の分散について学校間レベルと学校内の児童生徒間レベルに分解し、さらに学校間レベルにおいて能力特性によらない独自分散を項目ごとに仮定したマルチレベルの潜在変数モデル (Asparouhov & Muthén, 2007) を用いて分析した。

(2) TIMSS 調査において、各項目におけるカリキュラムの被覆状況の国間での違いに着目した研究を行った。具体的には、TIMSS2007 理科調査（物理領域）のデータを用いて、同じ出題項目でもその統計的な特性が国間で異なることをモデル化した。なお、分析の際には分析対象国の標本抽出デザインを考慮

した。この分析結果に対して、知識・応用・推論からなる認知的領域別に、各項目のカリキュラムの被覆状況の国間での異同によって整理した。

同じ出題項目でもその統計的な特性が国間で異なることについては、項目反応理論 (e.g., Lord, 1980) の枠組みで特異項目機能 (Differential Item Functioning, DIF; Holland & Wainer, 1993) の分析を行った。なお、ある項目について DIF が検出されたならば、能力特性が同水準であるにもかかわらず、集団（ここでは国）によって当該項目への正答確率が異なることになる。

また、層化の変数、クラスタリングの変数、及び標本加重の変数（ここではその和が標本の大きさとなるものを用いた）を分析に組み込むことによって、分析対象国の標本抽出デザインに対応した形で推定値の標準誤差を推定できるようにした。

各項目のカリキュラムの被覆状況については、TIMSS 調査における各国の調査責任者によって回答された、Test-Curriculum Matching Analysis (TCMA; Beaton, 1998) のデータを用いた。分析対象国については、その教育制度が“日本の学校制度とほぼ同じである”（藤村, 2009, p.51）韓国を取り上げ、日本と比較した。

### 4. 研究成果

3. (1)、(2)の結果をそれぞれ(1)、(2)に記す。

(1) 小学校第2学年では計算、小学校第5学年では図形、中学校第2学年では数と式の求答式の項目からなる上記の学力調査のデータを分析した。この結果、いずれの学年においても、調査実施の前学年における学習内容が範囲である場合よりも、当該学年の調査実施時期までに履修されると考えられる学習内容が範囲である場合の方が、わずかな違いではあったものの、学校間レベルにおける能力特性によらない項目独自の散らばりが認められ易いことが示唆された。

この結果は以下のように解釈できるだろう。すなわち、学習指導要領の水準から見れば、調査実施時点において既に取り扱われていたと仮定することができる範囲から出題された内容よりも、当該学年の特に11月までに履修すると考えられるものの、学習指導要領の水準から見れば（学年内でいつの時期に実施するとは規定されていないために）、全ての学校において既に取り扱われていたとは必ずしも仮定できるとは限らない範囲から出題された内容の方が、調査実施時点での履修状況のばらつきが学校間でわずかに大きいと考えられる、ということである。

ただし、複数のデータの間で結果は一貫し

てはいたものの、両者の違いは必ずしも大きくはなかった点、及び領域や内容及び問題の形式について焦点化したデータ分析を行っていた点から、本研究で得られた結果の一般化可能性については課題が残されたと言えるであろう。

(2) 予備的な分析の結果、ここでは4項目(4点)については分析の対象外とし、残りの50項目(53点)について分析の対象とした。被覆状況のパターン別にDIFの有意性検定の結果の分布を認知的領域ごとに整理したものを表1に示す。

これを見ると、まず知識領域において、両国においてカリキュラムに含まれているもののうち、能力特性が同水準でも日本の生徒にとってより難しいと判断されたものは50%であったのに対し、日本においてのみカリキュラムに含まれていないもののうち、日本の生徒にとってより難しいと判断されたものは80%であった。このことから、知識領域においては、日本と韓国の間では、日本においてのみカリキュラムに含まれていないことに伴って、能力特性が同水準であっても日本の生徒にとってより解きにくい項目が見られるようになることが示唆された。

次に、応用領域においては、両国においてカリキュラムに含まれているもののうち、能力特性が同水準でも日本の生徒にとってより易しいと判断されたものは50%であったのに対し、日本においてのみカリキュラムに含まれていないもののうち、日本の生徒にとってより易しいと判断されたものは25%であった。つまり、両国においてカリキュラムに含まれているものについては、能力特性が同水準であっても日本の生徒にとってより解き易い項目が半数程度あったが、日本においてのみカリキュラムに含まれていないものについては、日本にとってより解き易い項目の割合は小さくなった。このことから、応用領域においては、日本と韓国の間では、日本においてのみカリキュラムに含まれていないことに伴って、能力特性が同水準であっても日本の生徒にとってより解き易い項目が見られにくくなることが示唆された。

表1. 物理の被覆状況別にみた認知的領域ごとの均一DIFの検定結果

認知的領域	被覆状況	項目数	日本に難	有意差なし	日本に易
知識	日本○韓国○	6	3	1	2
	日本○韓国×	1	1	0	0
	日本×韓国○	5	4	0	1
	日本×韓国×	0	0	0	0
応用	日本○韓国○	22	3	8	11
	日本○韓国×	0	0	0	0
	日本×韓国○	8	1	5	2
	日本×韓国×	1	1	0	0
推論	日本○韓国○	7	1	4	2
	日本○韓国×	0	0	0	0
	日本×韓国○	3	1	0	2
	日本×韓国×	0	0	0	0

最後に、推論領域においては、他の認知的領域で挙げたような特徴的な日本の状況は見られなかった。

この結果は以下のように解釈できるだろう。すなわち、ある項目で問われた内容についてその知識を再生すること、あるいはそうした知識に基づいて応用することは、その項目で出題上示された状況から科学的に推論することに比べて、項目の内容が国のカリキュラムに含まれているかどうか及びその影響は大きいと考えられる。このことに鑑みれば、TIMSS2007調査の物理領域についての日本と韓国との比較において、日本においてのみカリキュラムに含まれていないことに伴う不利な状況が特に知識領域において見られ易くなり、応用領域においてはこのことに伴って有利な状況が見られにくくなるということが示された一方で、推論領域ではこのような傾向は見られなかったという本研究の結果は妥当なものであると言えよう。

ここで、具体的な項目例として、図1に公開問題(正答コードは④)を示しておく。この項目は、金属の粒子の配列に関する応用領域の問題であり、TCMAのデータによると、日本では調査実施学年までのカリキュラムに含まれておらず、韓国では含まれている項目であった。これについての両国の項目特性曲線を図2に示す。これを見ると、横軸の能力特性が同水準であっても、日本の方が縦軸の正答確率は低いことが分かる。

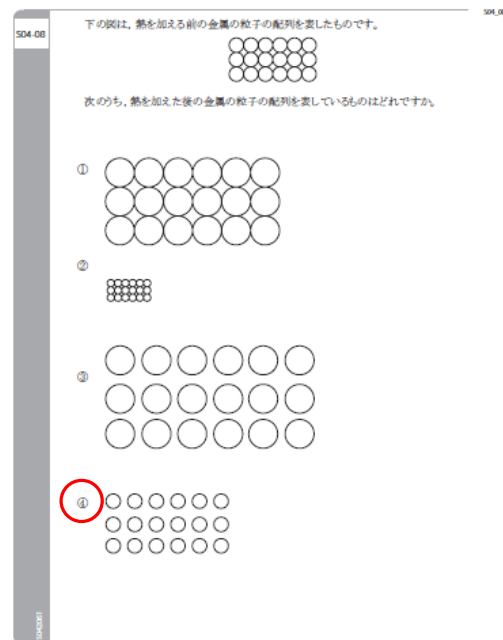


図1. 項目 S042061 (国立教育政策研究所, 2009 より転載し作成)

本研究の課題として、生物・化学・地学といった他の内容的領域については分析されていない点、比較の対象とした国が一つしかない点、調査言語の違いの影響 (cf. Ercikan

& McCreith, 2002) については考慮していない点が挙げられる。最初の2点については、内容的領域や分析対象国を今後の分析で充実していくことが考えられる。一方で、最後の点については、別途工夫が必要だと考えられる。具体的には、まず日本以外で調査言語が同一で地理的条件に近い国間での比較を行い、同様の結果が得られるか確認することが挙げられる。また、日本と他国の比較の際に調査言語の違いの影響についても検証する場合は、Ercikan & McCreith でもなされている主観的な評価に加えて、テキストマイニング等の方法も用いることについても検討することが望まれるであろう。

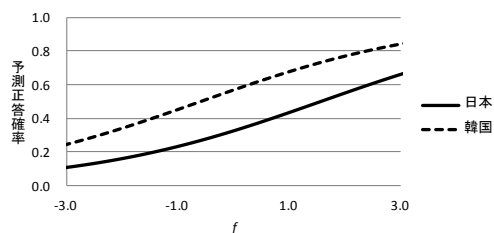


図2. S042061の項目特性曲線

## 文献

- Asparouhov, T., & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. Proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology. Retrieved November 4, 2009, from <http://www.statmodel.com/download/JSM2007000746.pdf>
- Beaton, A. E. (1998). Comparing cross-national student performance on TIMSS using different test items. *International Journal of Educational Research*, 29, 529-542.
- Ercikan, K., & McCreith, T. (2002). Effects of adaptations on comparability of test items and test scores. In D. F. Robitaille & A. E. Beaton (Eds.), *Secondary analysis of the TIMSS data* (pp. 391-405). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- 藤村和男 (2009). II. 教科書制度と教育事情 9. 韓国 国立教育政策研究所 第3期科学技術基本計画のフォローアップ「理数教育部分」に係る調査研究—理数教科書に関する国際比較調査結果報告—2009年3月 pp. 51-57.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.
- 国立教育政策研究所 (2009). TIMSS2007 理科教育の国際比較 (平成 21 年 8 月改訂版).
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Acknowledgment:  
The following data source and the released item were used: TIMSS 2007 Assessment. Copyright © 2009 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 4 件)

- ① 萩原康仁 中学校第 2 学年理科の物理領域を対象とした TIMSS データの DIF 検出の一分析—カリキュラムの被覆状況の違いに着目して— 日本教育心理学会第 54 回総会 琉球大学 2012 年 11 月 24 日
- ② 萩原康仁 カリキュラムの被覆状況を考慮した TIMSS データの DIF 検出の一分析—中学校第 2 学年理科の物理領域を例として— 日本テスト学会第 10 回大会 東京医科歯科大学 2012 年 8 月 21 日
- ③ 萩原康仁 学力調査の実施時期と学習内容の履修時期との関連についての検討—中学校第 2 学年に実施した数学のテストを用いて— 日本テスト学会第 9 回大会 岡山大学 2011 年 9 月 11 日
- ④ 萩原康仁 学力調査の実施時期と学習内容の履修時期との関連についての検討—小学校第 2 学年の算数の計算問題に着目して— 日本教育心理学会第 53 回総会 かねて 2・7 2011 年 7 月 24 日

## 6. 研究組織

- (1) 研究代表者  
萩原 康仁 (HAGIWARA YASUHITO)  
国立教育政策研究所・教育課程研究センター基礎研究部・総括研究官  
研究者番号: 30373187
- (2) 研究分担者  
該当者なし
- (3) 連携研究者  
該当者なし