

加法モデルを用いた TIMSS カリキュラム質問紙の分析：シミュレーションと実データの分析

○萩原康仁，松原憲治
(国立教育政策研究所)

1. 問題

次期学習指導要領の改訂に向けた動きがなされている。その中で、日本でも長い間実施されてきた国際的な学力調査の一つである TIMSS の結果は、量的なデータからカリキュラムの改訂のためのエビデンスを与えうるものとして重要である。TIMSS では、第 4 学年の児童と第 8 学年の生徒を対象として算数・数学と理科が実施され、その教育到達度が測定される（国立教育政策研究所，2013）。また、この調査は国レベルの意図されたカリキュラム，学校や教師レベルの実施されたカリキュラム，そして児童生徒の成果を示す達成されたカリキュラムからなる三層構造のカリキュラムモデルに基づく（Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009）。その中でも、達成されたカリキュラムに相当する児童生徒のスコアの国際比較が一般的には関心の中心となっているといえよう。

一方で、TIMSS では、意図されたカリキュラムに相当するものとして、調査参加国／地域（以下、調査参加国あるいは国とする）の各国調査責任者にカリキュラムに関する質問紙（以下、カリキュラム質問紙とする）が実施されている。そこでは、算数／数学及び理科の構成や内容について主に尋ねた質問項目群が並んでいる。TIMSS では、各内容領域（例えば中学校理科における内容領域は生物・化学・物理・地学）内で細分化されたトピック（Mullis et al., 2009）と呼ばれる単元におおむね相当する水準のものがある。カリキュラム質問紙では、就学前から第 12 学年にいたるまでの、どの学年のカリキュラムにそれぞれのトピックが扱われているかについて尋ねた問がある。

Matsubara, Hagiwara, and Saruta (2015) は、TIMSS2011 の小学校理科における生物の内容領域にある六つのトピックが、参加 59 か国の小学校第 1 学年から第 6 学年までの各学年のカリキュラムで扱われているかどうかについてのデータに対し、多特性多方法（MTMM）データに適用される確認的因子分析モデルの加法モデルのうち、CT-CM モデル（e.g., Widaman, 1985）を用いて分析した。具体的には、トピックを特性の相、学年を方法の相とみなし、それぞれの相内において六つずつの互いに相関のある因子を仮定して分析した。ただし、各観測変数を二値変数とみなしたため、カテゴリカル因子分析モデルを用いた。また、サンプルサイズが小さいため、マルコフ連鎖モンテカルロ法によるベイズ推定を行った。その結果、「人間の健康」と「環境の変化」のトピックについて、参加国の平均に比べて日本の小学校理科カリキュラムで扱われていないことが示唆された。この結果について、Matsubara et al. (2015) は、前者は理科ではなく体育等の他の教科や学校全体の教育活動で扱われていると考えられ、児童の当該項目群の通過率も高いままであったこと、一方で後者においては、当該項目群の多くにおいて通過率が必ずしも高いとは言えず、ESD（持続可能な開発のための教育）との関連を踏まえれば、このトピックに関連する内容をカリキュラムに含める余地が残されていることを指摘した。

しかしながら、サンプルサイズが小さいデータでベイズ推定をする際には事前分布の影響が強いことが知られている（e.g., Asparouhov & Muthén, 2010b）ものの、Matsubara et al. (2015) では、その影響は示されていない。また、MTMM データを分析する他のモデルとの比較をしていない。

そこで本研究では、CT-CM モデルの他に、学年（方法）の相における一つの因子を除いた CT-C(M-1) モデル（Eid, 2000）を用いる。これら二つのモデルについて、因子負荷量と閾値の事前分布について二つの異なるパターンを設定した、小規模なシミュレーション研究を行う。併せて、Matsubara et al. (2015) のデータをこれらのモデルで再分析し、結果を確認する。

2. シミュレーション

60 の国に六つのトピックの扱いの有無を 6 学年分尋ねること（36 の二値の観測変数）を想定した。CT-CM 及び CT-C(M-1) モデルについて、各モデルを真とした場合のデータを 200 ずつ発生させた。各因子に（多変量）正規分布を仮定し、平均を 0，分散を 1 と固定した。因子負荷量の真値はすべて 1，因子間相関の真値はすべて 0.6 とし、閾値の真値は -1.5, -0.9, -0.3, 0.3, 0.9, 1.5 の 6 種類を六つずつ用意した。本研究の分析には、Mplus（Muthén & Muthén, 1998–2012）version 7.3 を使い、マルコフ連鎖モンテカルロ法を用いたベイズ推定を行った。因子負荷量と閾値の事前分布については、 $N(0, 2.25)$ の条件と $N(0, 1)$ の条件を設定した。因子間の共分散（相関）行列の事前分布については、ソフトウェアのデフォルト（ $IW(I, 7)$ ），ただし CT-C(M-1) モデルの学年（方法）因子については $IW(I, 6)$ ）とした。後半の二つおきの値を計算に含めることとし、2 本の連鎖を最低 10 万個ずつ発生させた。収束の基準については、PSR（potential scale reduction）の指標値が最大で 1.1 未満であることとした。

3. シミュレーションの結果と考察

どのデータにおいても、収束の基準は満たされた。各条件における因子負荷量（特性の相）、閾値、因子間相関（特性の相）における相対バイアスの結果を表1に、これらの95%被覆率の結果を表2に示す。相対バイアスについては、事前分布が違っても同じモデル間でも相当程度異なってくることが分かる。95%被覆率については、95%に近いことが望まれるが、特にCT-CMモデルにおいては、設定した条件では事前分布が違っても母数の種類によらず大きな違いがなく、しかも95%におおむね近い被覆率であることが分かる。これらのことから、実際の分析結果を見るに当たっては、点推定ではなく区間推定を参照した方がよいこと、及び、設定した条件は少ないが、区間推定における事前分布の違いによる影響はCT-CMモデルの方がCT-C(M-1)モデルよりも幾分小さいことが示唆された。

表1. 相対バイアス(%)の最小値, 平均値, 最大値				表2. 95%被覆率(%)の最小値, 平均値, 最大値			
CT-CM	因子負荷量(特性)	閾値	因子間相関(特性)	CT-CM	因子負荷量(特性)	閾値	因子間相関(特性)
$N(0, 2.25)$	-10.6, 4.9, 15.5	-5, 7.6, 21.2	-20.1, -18.2, -15.6	$N(0, 2.25)$	93, 97.4, 100	93.5, 97.5, 100	91.5, 94.8, 97
$N(0, 1)$	-34.3, -21.2, -11.5	-19.1, -9.2, 3.2	-15.1, -13.4, -11.1	$N(0, 1)$	96, 98.4, 100	94.5, 96.5, 99	92, 93.9, 95.5
CT-C(M-1)				CT-C(M-1)			
$N(0, 2.25)$	-19.9, 7.3, 25.3	-14.3, 4.3, 30.6	-11.2, -8.4, -6.6	$N(0, 2.25)$	91.5, 95.7, 98	93.5, 96.7, 100	90.5, 93.1, 95
$N(0, 1)$	-38.5, -15.7, -0.9	-29.1, -11, 12.7	1.6, 4.1, 5.6	$N(0, 1)$	85, 96.3, 99.5	87, 95.4, 99	85, 87.3, 91.5

4. 実データの分析結果と考察

CT-CM及びCT-C(M-1)モデルの実データの分析では、因子負荷量と閾値の事前分布を $N(0, 2.25)$ とした。後半の五つおきの値を計算に含めることとし、5本の連鎖を最低10万個ずつ発生させた。残りはシミュレーションと同様の条件とした。CT-C(M-1)モデルでは、除く学年（方法）因子が6通りあるため、この6通りの分析をした。Matsubara et al. (2015)では、因子負荷量と閾値の事前分布をデフォルトの $N(0, 5)$ としており、2本の連鎖を最低5万個ずつ発生させ、後半の各連鎖の値を計算に含めていた。

分析の結果、いずれのモデルにおいても収束の基準は満たされた。事後予測 p 値はCT-CMモデルが.374、CT-C(M-1)モデルが.315から.342であり、CT-CMモデルの方が0.5により近かった。

CT-CMモデルの各国の各因子について1万個ずつ推算値を抽出した。なお、Asparouhov and Muthén (2010a)は、多数(100から500程度)の推算値を抽出した場合は個々の因子得点の推定に用いられるとしている。CT-CMモデルにおける日本の各トピック（特性）因子の推算値の95%区間を表3に示す。Matsubara et al. (2015)と同様に、「人間の健康」と「環境の変化」のトピックについて、参加国の平均に比べて日本のカリキュラムで扱われていないことが示唆された。ただし、CT-C(M-1)モデルでは、除く学年（方法）因子によって各トピック（特性）の推算値も変動し、解釈が難しくなった。

以上より、今回のデータに対して、CT-C(M-1)モデルを用いる利点はあまりないようであることが示唆された。

表3. 日本のトピック(特性)因子における推算値の95%区間(CT-CMモデル)		
因子(トピック)	2.5%	97.5%
Major body structures and their functions in humans and other organisms	-2.096	0.132
Life cycles and reproduction in plants and animals	-1.895	0.718
Physical features, behavior, and survival of organisms living in different environments	-2.328	0.010
Relationships in a given community	-2.371	0.212
Changes in environments	-2.636	-0.111
Human health	-2.474	-0.248

文献

- Asparouhov, T., & Muthén, B. (2010a, August 21). *Plausible values for latent variables using Mplus*. Technical Report.
- Asparouhov, T., & Muthén, B. (2010b, September 29). *Bayesian analysis of latent variable models using Mplus (version 4)*. Technical Report.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, **65**, 241–261.
- 国立教育政策研究所 (2013). TIMSS2011 理科教育の国際比較 明石書店
- Matsubara, K., Hagiwara, Y., & Saruta, Y. (2015, June 25). *A statistical analysis of the characteristics of the intended curriculum of Japanese primary science and its relationships with the attained curriculum*. Paper presented at the 6th IEA International Research Conference, Cape Town, South Africa.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, **9**, 1–26.

※科研費(25381060)の助成を受けた。