

国際比較／文化比較調査における測定の等価性／不変性の研究*

——多集団確証的因子分析（MGCFA）を中心として——

真 鍋 一 史**
前 田 忠 彦***
清 水 香 基****

I. はじめに

本稿は、国際比較／文化比較調査における「測定の等価性／不変性」のこれまでの研究の内容とその系譜を、伝統的な「文献研究」の手法にもとづきながら、実証的に描き出すことを試みる、いわゆる「レビュー論文」である。そこで、本稿で取りあげるテーマは、以下の4点となる。

1. この研究領域における諸文献を、「統計的方法に関する研究」「統計的方法の国際比較／文化比較研究への応用に関する研究」「国際比較調査の実践と方法に関する研究」に分けるとともに、この領域における研究の系譜を、これら3種類の研究の相互の影響という視座から概観する。
2. 測定の等価性／不変性の実証的な確認を可能にした「多集団確証的因子分析（MGCFA）」と呼ばれる統計的技法のアイデア・技法・手続きを、上述の諸文献にもとづきながら整理する。
3. 以上のようなMGCFAを用いた測定の等価性／不変性の実証的な検定は、“exact approach”という形で方法論的に完成されるが、その後、さらに“approximate approach”へと新しい展開を示すことになる。そのような新しいアイデア・技法・手続きのまとめが第3のテーマである。
4. このようなMGCFAという統計的技法は、

1990年代ごろから、日本でも広く紹介されることになった——前田の共著書（1992）も「共分散構造分析」や「確証的因子分析」を紹介する役割を担った文献の1つであったといえる——。それにもかかわらず、そのような技法の「国際比較／文化比較研究」への「応用研究」は、日本においてはほとんどなされてきていない。そこで、最後に、MGCFAの日本における受容とその応用の諸相の概観を試みる。

では、なぜ、このような「レビュー論文」の試みを行なうかという、実証科学の知的営為にとって、このような試みには大きな「意義」があると考えられるからにほかならない。

このような「意義」について説明しようとする場合、宇宙物理学者の池内了（2014）が、同じく物理学者の寺田寅彦の研究への取り組みの姿勢について述べた、つぎのような一文が、きわめて示唆的である。

「（寺田）寅彦は仕事に取り掛かるまえには洋書を何冊も読んで問題の全体像を把握し、飛躍の種を見つけようという努力を積み重ねていたのである。」（p.326）

繰り返しになるが、本稿は「レビュー論文」である。ここで、このような「文献レビュー」を試

*キーワード：測定の等価性／不変性、国際比較／文化比較調査、多集団確証的因子分析

**関西学院大学名誉教授、青山学院大学名誉教授、統計数理研究所データ科学研究系客員教授

***統計数理研究所データ科学研究系准教授

****北海道大学大学院文学研究科助教

みる所以は、「国際比較／文化比較調査における測定の等価性／不変性」に関するこれまでの研究の、まさに「全体像」の把握というところにある。そして、このような把握は、それを踏まえて、この研究領域における新しい「飛躍の種」を見つけたするためのものにほかならない。さらにいうならば、以上のようなこの領域の「研究の全体像の把握」と、それにもとづく「新しい飛躍の種の発見」は、単にわれわれの共同研究者間においてのみ共有されるものではない。いうまでもなく、それは、広く関連研究領域のアカデミック・コミュニティにおいて共有されるべきものといわなければならない。

こうして、本稿は、このようなテーマで共同研究を続けてきた3人の共同執筆の形で構成されることになった。I、II、III、Vは真鍋、IVは前田、VIは清水によって原稿が準備され、共同討議の繰り返しをとおして、現在の形へとまとめ上げられてきたのである。

II. 国際比較／文化比較調査における測定の等価性／不変性の研究の系譜

社会科学における「測定の数学的基礎の研究」の確立をめざした Paul F. Lazarsfeld は、その著書、*Qualitative Analysis: Historical and Critical Essays* (1972=1984) ——日本語の標題は『質的分析法——社会学論集——』——において、「概念形成史小論」と題する「歴史的考察」を展開している。じつは、本稿のアイデアは、この Lazarsfeld の「ひそみにならった」ものにほかならない。Lazarsfeld は「比較的新しい分野で研究をするときは、必ずといっていいほどその分野の歴史に関心がつながっている」(訳書、p.4) という。確かに、本稿で取りあげる「測定の等価性／不変性」というテーマは、社会科学の領域においては、比較的新しいものといえる。そこで、そのようなテーマの研究を概観するにあたり、本稿でも「研究の歴史」という視座を取り入れる。では、なぜそうするかというと、この点については生物学者の福岡伸一が、『新・ドリトル先生物語』のなかで、つぎのように書いている一文が示唆的である。

「ドリトル先生は言いました。『算数を勉強するには算数の歴史を勉強することから始めなさい。そうすると、なぜそのとき、こんな考え方が出てきたのか、といういきさつがわかる。いきさつがわかると理解も深まるというものだよ』」(福岡伸一の新・ドリトル先生物語)『朝日新聞』2021年4月5日(月)なお、……は筆者ら)

このような問題関心から、本稿では、このテーマをめぐる関連領域の学術書・学術論文として発表されてきた文献を時系列的な一覧表の形で整理するとともに、これら文献におけるいくつかの「レビュー論文」にもとづきながら、それら文献のつながりを跡付けていく。そして、その上で、そのような研究史におけるいくつかの問題を取りあげ、若干のコメントを加えていきたい。

ただ、以上のような一覧表の作成については、つぎの点を記しておかなければならない。

(1) ここでの関連文献の収集は、体系的なものでも、網羅的なものでもないということである。これら文献は、真鍋(1985, p.261)の用語でいうならば、「雪だるま方式」ともいべき方法で収集されたものである。ここで「雪だるま方式」とは、文献研究の伝統的なテクニックを継承するものであり、既存の文献目録や、よく知られた文献をまず手がかりとして、これらの文献にあげられているほかの文献に当たるといように、つぎつぎに雪だるま式に文献をたぐっていくという方式である。

では、ここでの文献収集において、そのような「雪だるま方式」の出発点となったのがどの文献だったかということ、それはつぎのものであった。

Eldad Davidov, Peter Schmidt, Jaak Billiet, and Bart Meuleman eds., *Cross-Cultural Analysis: Methods and Applications* [2nd Edition] Routledge, 2018.

では、なぜ、この文献が出発点となったかについては、Merton (1977=1983) の用語でいうところの Episodic Memoir——日本語訳では「エピソード

ソードで綴る回想録」——が必要となる。しかし、この点については、すでに別のところで詳細な記録を残している（真鍋、2018, 2020 a, 2020 b) ので、ここでは重複を避ける。いずれにしても、この文献との出会いをとおして、われわれは「国際比較／文化比較調査における測定の等価性／不変性の研究」へと導かれることになったのである（「国際比較」と「文化比較」という用語については、真鍋（1985, p.5）を参照）。

(2) 今回の文献収集は、このような方式で進められてきたが、こうして収集された文献の数は膨大なものとなった。この成果は、それ自体できわめて興味深いものといわなければならない。なぜならば、このような文献の数の多さは、この研究領域への研究者の関心の高さ／広さを示すものであると考えられるからである。しかし、他方で、このような大量の文献の精査のためには、膨大な時間とエネルギーが必要となる。そこで、今回の文献の時系列的な一覧表の作成に当っては、そのような文献選択の基準ともいべきものを準備した。具体的にいうならば、それはつぎの3つである。

①以上で述べた「雪だるま方式」の文献収集をとおして、いわゆる「参照回数」の多い文献が明らかとなってくる。このような「参照回数」の多さを文献選択の基準の1つとする。

②「雪だるま方式」で収集された文献のなかには、この領域の「研究動向」を概観する、いわゆる「レビュー論文」が含まれている。それら「レビュー論文」では、それぞれの選択基準にもとづいて、そこで取りあげる文献が選ばれているので、それらの基準を手がかりとする。

③文献選択の第3の基準として、この研究テーマをめぐる関連国際学会——「国際社会学会」「国際心理学会議」「国際比較文化心理学会」「ヨーロッパ社会調査学会」——における、いわば「参与観察的」ともいべきさまざまな体験——シンポジウムやセッションへの参加、見聞、そして研究者交流——があげられる。それらをとおして得られた、この領域における研究動向の把握にもとづいて、文献を選択する。

(3) 以上において、この研究テーマに関する文献の時系列的な一覧表の作成のための、文献選択の基準について、かなり詳細に説明してきた。しかし、このような一覧表の作成については、さらにもう一点、説明しておくべきことが残されている。それは、そのような文献の「種類」といったことである。しかし、学術書・学術論文の「種類」についてはさまざまなものが考えられる。ここでは、上述の Davidov et al. eds. (2018) で提案された「分類基準」を、本稿の目的に合わせて、若干改訂したものを採用する。

それは、これら文献を、

①統計的方法に関する文献

②国際比較／文化比較調査の実践と方法に関する文献

③統計的方法の国際比較／文化比較調査のデータ分析への応用に関する文献

の3種類に分類するというものである。

以上のような「手続き」にしたがって作成したのが表1の一覧表である。因みに、これら文献のうち□で囲んだ文献は、いわゆる「レビュー論文」という性格のものである。

以下においては、この一覧表と、これら「レビュー論文」の記述にもとづきながら、この領域における「研究動向」をコンパクトな形でまとめることを試みる。したがって、そのような「まとめ」は、上述のように分類された諸領域の研究の相互の影響の過程に焦点を合わせるものとなる。この過程は、図1の「仮説的モデル」に示される。

1. 本稿の問題関心が、「国際比較／文化比較研究」は、いかにして測定の等価性／不変性という問題に取り組んできたか」といったところにあるところから、ここでの「まとめ」は「国際比較／文化比較調査の実践と方法」の領域における「研究動向」から始める。

Johnson (2015) は、Public Opinion Quarterly (POQ) の79・S1の巻頭言ともいべきものとして、「調査方法論における比較文化の諸問題 (Cross-Cultural Issues in Survey Methodology)」と題する

表 1 関連諸文献の時系列一覧表

統計的方法に関する文献	統計的方法の応用に関する文献	国際比較／文化比較調査の 実践と方法に関する文献
		1958 : Lerner
1969 : Jöreskog		1963 : Almond & Verba 1968 : Rokkan ; Lipset
1971 : Jöreskog 1973 : Jöreskog		1970 : Holt & Turner 1970s : Eurobarometer
1984 : LISREL 6 (Jöreskog & Sörbom) 1989 : LISREL 7 (Jöreskog & Sörbom) 1989 : Byrne et al.		1980s : EVS WVS ISSP 1988 : Smelser
1993 : Meredith 1996 : LISREL 8 (Jöreskog & Sörbom)	1998 : Steenkamp & Baumgartner	1994 : Alwin et al. 1998 : Johnson
2006 : AMOS 2007 : Millsap & Meredith	2000 : Vandenberg & Lance	2002 : ESS 2003 : Harkness et al. eds. 2005 : ESRA 2009 : ISSP 1984-2009 2010 : Harkness et al. eds.
2012 : Muthén & Asparouhov 2013 : van de Schoot et al. 2013 : Mplus 2014 : Asparouhov & Muthén 2015 : STATA 2017 : Muthén & Asparouhov	2010 : Braun & Johnson 2011 : Davidov et al. eds. (First) 2012 : Davidov et al. 2014 : Davidov et al. 2014 : Cieciuch et al. 2015 : van de Schoot et al. 2015 : Zercher et al. 2017 : Meitinger 2018 : Davidov et al. 2018 : Davidov et al. eds. (Second) 2018 : Braun & Johnson 2018 : Cieciuch et al. 2018 : Seddig & Leitgöb 2019 : Cieciuch et al.	2015 : Johnson 2019 : Johnson et al.
	2020 : Meitinger	

注：□は「レビュー論文」を示している

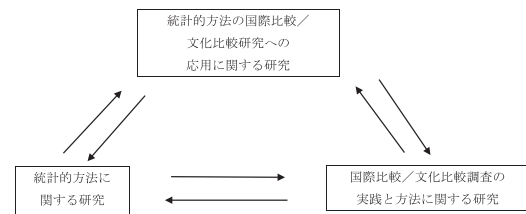


図 1 測定の等価性／不変性に関する 3つの研究領域の相互の影響の過程

小論を書いている。この小論の要点は、つぎのようなものである。

①文化比較に関する実証的な研究は、すでに半世紀を越える歴史を持っている。

②その嚆矢として、Daniel Lerner の *The Passing of Traditional Society* (1958) や Gabriel Almond and Sidney Verva の *The Civic Culture* (1963) があげられる。

③これらは、この研究領域における「実証的な調査研究」の古典として位置づけられるものであるが、その後、このような「調査研究の方法論」に焦点を合わせた著書が現れることになる。Rok-

kan (1968)、Holt and Turner (1970) などがその例である。

注：Johnson (2015) の小論においては取りあげられていないが、国際比較／文化比較という視座に関しては、「トクヴィル・デュルケム・ウェーバー」をめぐる方法論的研究も看過できない。そのような問題関心から、一覧表には、Lipset (1968=1972) と Smelser (1988=1996) を加えた。因みに、「測定と等価性」の問題は、前者(訳書) pp.1-30、後者(訳書) pp.211-218、において取り扱われている。

④1980年代になって、国際比較／文化比較の研究の「形態 (form)」に大きな変化が現れる。それは、このような実証的な研究の「拡大化・組織化・制度化」ともいべき方向であった。具体的にいうならば、「ユーロバロメーター (Eurobarometer)」をはじめとする「ヨーロッパ価値観調査 (European Values Study: ESS)」「世界価値観調査 (World Values Survey: WVS)」「国際社会調査プログラム (International Social Survey Programme: ISSP)」「ヨーロッパ社会調査 (European Social Survey: EVS)」など、「質問紙法にもとづく多くの国ぐにを対象とする大規模な国際比較調査」の出現である。このような国際比較／文化比較の視座からする大規模な社会調査の実践は、現代の社会科学の領域における最も大きな出来事の1つとされるものである(このような研究動向については、Johnson et al., 2019 を参照)。

注：ここで、再び、Johnson (2015) において触れられていないもう1つの研究動向についても述べておかなければならない。それは、以上のような国際比較／文化比較調査の実践活動にともなって、いわゆる「調査データの二次分析 (secondary analysis of survey data)」の試みが大きく発展することになったということである。このような発展については、「Herbert Hyman の *Secondary Analysis of Sample Survey* (Wiley: 1972; Wesleyan University Press: 1987) から Max Haller, Roger Jowell and Tom W. Smith eds. の *The International Social Survey Programme 1984-2009* :

Charting the Globe (Routledge: 2009) へ」として捉えることができる(真鍋, 2012)。

因みに、International Social Survey Programme: ISSP (国際社会調査プログラム) は、もともとドイツ・マンハイムの「調査・方法・分析センター (ZUMA)」の総合社会調査 ALLBUS と、米国シカゴ大学の「全国世論調査センター (NORC)」の総合社会調査 GSS が中心となり、そこに世界各国の大学・研究所・調査機関が加わり、1984年に設立された質問紙法による大規模な国際比較調査プログラムである。Haller et al. eds. (2009) は、その設立25周年(1984~2009年)を記念して、企画・執筆・編集されたISSP調査データの二次分析の論文集であり、アメリカ社会学会の国際比較社会学部門において「最優秀出版賞 (Best Publication Award)」を受賞した。この受賞は社会調査データの二次分析をとおして、国ぐに間、時点間の比較を行なうという実証的な研究の形態と、その成果が高く評価されたことを示すものにほかならない。

しかし、このような国際比較調査と、そのデータの二次分析のスタイル／形態が社会科学の領域において、確固たる位置を獲得するに到る過程には、もう1つ忘れることのできない大きな出来事があった。それは、これら社会調査データの「収集・整理・保管・活用」の機能を担う「データアーカイヴ」の設立である(データアーカイヴに関しては、真鍋, 2014 を参照されたい)。こうして、「国際比較調査」と「データアーカイヴ」と「二次分析」には、もはや切っても切れない緊密な関係が築かれることになってきたのである。

以上のような時代的な背景のもとに、「国際比較／文化比較調査における測定の等価性／不変性」という方法論的な「問題関心」が登場することになるのである。

2. Johnson (2015) に示された慧眼の1つは、このような研究領域の発展過程における、「実証的な研究」と、それをめぐる「方法論的な研究」との相互の影響というところに焦点を合わせたということである。こうして、Lerner や Almond and Verva による「実証的な研究」に続いて、Rokkan (1968) や Holt and Turner (1970) による「方法

論的な研究」が登場してきたのと同様、上述の「国際比較査の実践活動」とほぼ時を同じくして、Alwin et al. (1994)、Johnson (1998) などによる「国際比較／文化比較をめぐる方法論的な研究」が出現してきたのである。ここで、Alwin et al. と Johnson のこれら2つの文献が、上述のドイツの研究機関 ZUMA を基盤としてでてきたものであることは、上述の社会調査のデータアーカイブの最大のものが、ドイツ・ケルン大学の ZA であったことと併せて、とくに注目されるべきことといわなければならない。日本においては、ドイツ社会学という、フランクフルト学派の「批判的 sociology (critical sociology)」が想起される傾向があるが、Lazarsfeld (1972=1984, p.4) も指摘しているように、アメリカ社会学の影響を受けた「実証的 sociology (empirical sociology)」の伝統もすでに築かれていたのである (真鍋, 2014)。

そして、本稿にとって重要であるのは、そのような研究の系譜のなかから「国際比較／文化比較研究」への「方法論的関心」が醸成されてくることになり、それが Harkness et al. eds. (2003)、Harkness et al. eds. (2010) を経て、後述する Eldad Davidov らの「国際比較／文化比較調査における測定等の等価性／不変性の研究」へとつながっていったということである。

しかし、それは、いきなり Davidov et al. eds. (2011) の出版という形で現れてきたわけではない。それに先立って、社会科学の歴史に残る、もう1つの大きな出来事があった。それは、近年のヨーロッパにおける多彩な社会調査の実践活動の交流とその知的成果の共有をめざして、2005年、「ヨーロッパ社会調査学会 (European Survey Research Association: ESRA)」が設立されたということである。真鍋はこの学会の設立当初からのメンバーであるが、そこでの毎回の研究発表の中心は、国際比較調査データの二次分析によって占められてきており、その一翼を担ってきたのが、Davidov らの研究グループによる「測定等の等価性／不変性に関する研究」であったのである。そして、その1つの集大成こそが、上述の本の出版であったのである。

さて、以上において、一覧表の右端の欄に整理された「国際比較／文化比較調査の実践と方法の

研究」の系譜を辿るとともに、それら実践活動と方法論研究の流れを、Davidov らの「国際比較／文化比較調査における測定等の等価性／不変性に関する研究」の背景として位置づけてきた。このような研究動向についての整理の仕方をするのは、本稿で、Davidov らの研究を「問題関心」の中心に置いているからであって、別の「問題関心」を設定するならば、おのずから本稿とは別の整理の仕方が可能になることはいうまでもない。

3. Davidov らの研究は、本稿で中心的な位置づけがなされるものであり、章を改めて取りあげることになるが、ここで、そのような研究の特徴を簡単にまとめておかなければ、それはつぎのようになる。

①社会科学の領域では、その研究の対象としての社会的現実 (social reality) を「客観的現実 (objective reality)」と「主観的現実 (subjective reality)」に概念的に区別する (Berger and Luckmann, 1966=1977)。具体的にいうならば、社会の「制度・組織・構造」といった側面は前者に、そして、その社会の「意識・精神・性格」といった側面は後者に対応するものとして位置づけられる。

この後者の側面は、より一般的な用語でいうならば、「人びとのものの見方・考え方・感じ方・行動の仕方、あるいは、人びとの社会意識・心理特性・価値観・エトス・国民性・行動傾向」などと呼ばれるものであり、Davidov らの研究が焦点を合わせたのは、まさにこのような社会的現実の側面——つまり「主観的現実」——であった。

②そして、このような「主観的現実」を捉える視座が、「比較」——Lipset (1968=1972) の「諸国民の価値やエトスや国民性などを解明しようとする作業は、その分析が比較という形でなされるほかはない」(訳書, p.33) という指摘は重要である——であり、より具体的にいうならば、その比較は「時間的な比較」——つまり「時点間比較」——と「空間的な比較」——つまり「国際比較／文化比較」——という形 (form) をとることになる。

③以上のような「比較」という視座は「比較可能性 (comparability)」を前提としている。では、

そのような「比較可能性」は、どのような実証的な手続きによって確認できるであろうか。このような「問い」に答えるのが、Jöreskog (1969, 1971, 1973) の考案した「多集団確証的因子分析 (multiple group—あるいは、multi-group—confirmatory factor analysis: MGCFA) と呼ばれる統計的な技法である。Davidov ら (2011) は、自らの研究を「応用研究 (applied studies)」と性格づけているが、それは、このような統計的な技法を、人びとの「ものの見方・考え方・感じ方・行動の仕方」に関する「国際比較／文化比較調査」の「データ分析」に、まさに応用したからにはかならない。そして、そうであるからして、Davidov らは、自らの研究のスタイルを“non-technical”とも性格づけているのである。

④ Davidov らの研究の特徴の最後は、それらの研究がよって立つところの「個別科学 (discipline)」ということである。この点からするならば、Davidov とその研究グループのメンバーのそれは、「社会学・社会心理学・心理学」ということになるのであろう。ここでの問題関心からするならば、「社会学・社会心理学・心理学」という領域は、「計量心理学 (psychometrics)・数理心理学 (mathematical psychology)」とは、個別科学の観点からして、ひとまず区別されるものであろう。いうまでもなく、ここでの目標は、これら個別科学の形式的な分類にあるわけではない。そもそも、このような区別は、さまざまな具体的な研究事例からするならば、およそ不可能とならざるをえないものであるかもしれない。では、ここでの「ひとまずの区別」は、その判断の基準として何を考えているのであろうか。そして、そのような区別には、どのような意味があるのであろうか。これらの「問い」に対しては、以下のように「答える」ことができるであろう。

まず、「ひとまずの区別」については、2つの判断の基準を考えている。

i) その研究が technical であるか、それとも、non-technical であるか、という基準。この基準からするならば、「社会学・社会心理学・心理学」が、どちらかといえば non-technical であるのに対して、「計量心理学・数理心理学」は、どちらかといえば technical であるといえる。“technical”

と“non-technical”という用語については、真鍋 (2021, p.11) を参照。

ii) technical であるか、それとも、non-technical であるかの区別は、その研究者が統計的技法の creator であるか、それとも user であるか、という区別に対応している。「社会学・社会心理学・心理学」が人びとの社会的な行動の法則の定立や、それらの説明の体系としての理論の構築といった substantive な側面に関心を持つものに対して、「計量心理学・数理心理学」は、そのような法則の定立・理論の構築のための統計的な技法・モデルといった methodological な側面に関心を示す。そのため、後者は technical であり、methodological であり、creator である。

つぎに、以上のような区別が、どこまでも「ひとまず」のものであり、「便宜的」なものであるにもかかわらず、あえてこのような区別をしておくことには、どのような意味があるのであろうか。ここでは、creator という用語を「統計的技法・モデルの開発者」という意味で用いた。しかし、いうまでもなく、科学と呼ばれる人間の知的営為においては、creative という用語は、より広い意味内容を含んでいる。確かに、Davidov らは、新しい「統計技法・モデル」の開発という点からするならば、creator ではない。しかし、そのような「統計技法・モデル」を用いた新しい「応用研究」の展開という点からするならば、creator である。本稿のめざすところは、このような「応用研究」における新しい creativity——創造性——の可能性の探究であり、このような点からするかぎりにおいて、以上のような「ひとまず」の「便宜的」な区別は意味のあるものとなるのである。

4. 以上において、一覧表の「応用研究」の欄に位置づけられる Davidov らの一連の研究の特徴について述べてきた。しかし、いうまでもなく、「応用研究」というものは、「社会学・社会心理学・心理学」における「応用」に限られるものではない。そのようなことを示唆する文献として、この一覧表の同じく真ん中の欄に、Steenkamp and Baumgartner (1998) と Vandenberg and Lance (2000) をあげておいた。前者は、消費者行動学 (consumer behavior science)、そして後者は、組織

科学 (organizational science) という substantive な研究領域——そして、それと同時に、学際的な研究領域——における研究成果というものである。いずれの場合も、「多集団確証的因子分析」という統計的な技法が導入されるのは、「実証科学的な研究成果の質の向上」(Steenkamp and Baumgartner 1998, p.88) が目標とされるからにほかならない。このような意味においては、「応用研究」はさまざまな個別科学、そして研究領域において試みられるはずのものである。Davidov ら (2018) は、そのような個別科学/研究領域として、「社会学」「心理学」「社会心理学」「政治学」「経営学」「マーケティング」「経済学」「教育学」「人文地理学」「公衆衛生学」「社会疫学」などをあげている。こうして、これらそれぞれの個別科学/研究領域における「応用」のあり方については、それぞれの「系譜研究」がありうるであろう。しかし、本稿の一覧表においては、「応用研究」という場合のその「応用」の範囲は、すでに述べたように、「社会学・社会心理学・心理学」の領域に限っている。それこそが、本稿の研究テーマであるからにほかならない。では、なぜ、本稿の一覧表において、上述の2つの文献だけは含めたかという点、それはこの2つがすぐれた「レビュー論文」であって、「多集団確証的因子分析」に関する文献において、たびたび参照されてきたものであるからである。

5. さて、以上において、このような「社会学・社会心理学・心理学」という個別科学/研究領域における「国際比較/文化比較研究」の系譜を辿ることをとおして、Davidov らの Jöreskog に始まる統計的技法——「多集団確証的因子分析」——の「応用研究」へと記述を進めてきた。ここで、ようやく Jöreskog らの統計的技法の系譜——一覧表の左端の欄に時系列的に記された文献の流れ——に移ることになる。しかし、このような記述に関しては、それに先立って、つぎの点について述べておかなければならない。一覧表の左端の欄の標題は、「統計的方法に関する研究」としたが、この欄に整理した文献も、厳密に言えば、さらにいくつかの種類に分けられるであろう。それは、

i) 新しい統計的技法の開発研究

ii) そのソフトウェア・プログラムの開発研究
iii) その技法の開発者と利用者のインターフェイスの役割を担う研究

というような研究のタイプが区別されるのではなからうか、ということである。例えば、多次元尺度法の開発の先駆者である Louis Guttman は、「最小空間分析 (Smallest Space Analysis)」と呼ばれる統計的技法を考案したが、そのコンピュータ・ソフトウェア・プログラムは J. C. Lingoes によって作成され——J. C. Lingoes (1973). *The Guttman-Lingoes Nonmetric Program Series*, Mathesis Press——、さらに、P. Porrat によって、そのプログラムの「利用の手引き」が編集された——R. Porrat, Compiled (1974). *Laboratory Manual for the Guttman-Lingoes Nonmetric Computer Program*, The Israel Institute of Applied Social Research. —— ことによって、はじめてその統計的技法の一般的な利用が可能となった。そして、その上で、I. Borg, S. Shye, D. Canter, S. Levy らによる社会調査のデータ分析への「応用研究」が進められることになったのである (真鍋、1993、2002、2016、2021 などを参照されたい)。

ただ、本稿においては、具体的にいうならば、一覧表の左端の欄に整理した文献に関しては、以上のような研究内容にもとづく分類は行っていない。それは、今後の課題としたい。

6. 社会科学——とくに、「社会学・社会心理学・心理学」——の領域における人びとの主観的意識の実証的な研究では、従来から、測定の実証的道具 (measurement instrument) ——つまり質問紙調査の場合でいうならば、質問諸項目——の「信頼性 (reliability)」と「妥当性 (validity)」の検討が、データ分析のための第1の課題とされてきた。しかし、国際比較/文化比較のデータ分析の場合は、それに加えて、そのような測定の「比較可能性」の検討が重要な課題となってきた。そのような課題こそが、「測定の等価性/不変性の検定 (testing)」と呼ばれる「手続き (procedure)」である。このような研究動向については、以下の点を確認しておくことが必要であろう。

①人びとの主観的意識の測定においては、従来

から、質問諸項目の「信頼性」と「妥当性」の検討がデータ分析の第1の課題とされてきたと述べた。では、それは、どのようになされてきたであろうか。ここでは、人びとの主観的意識を捉える質問諸項目／諸スケールを総合的・体系的・概括的に検討した3つの文献をあげおきたい。

Ian McDowell and Claire Newell (1987). *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press.

Johon P. Robinson, Phillip R. Shaver, Lawrence S. Wrightsman eds. (1991). *Measures of Personality and Social Psychological Attitudes*. Academic Press.

Peter C. Hill and Ralph W. Hood Jr. eds. (1999). *Measures of Religiosity*. Religious Education Press.

これらの文献においては、人びとの主観的意識をめぐるさまざまな substantive なテーマに関する実証的研究において作成されてきた、measurement instruments についての「信頼性」と「妥当性」がきわめて詳細に検討されている。

しかし、それにもかかわらず、これらの文献においては、「測定の不変性／不変性」という問題はまったく取り扱われていない。そのような問題がこの研究領域における重要な課題となるのは、時間的にもう少し後——ほぼ2000年代——になってからのことである。「測定の不変性／不変性」という問題を「新しい研究動向」の1つとして位置づけた所以が、まさにここにあるのである。

②以上においては、「測定の不変性」という用語と、「測定の不変性」という用語を、それぞれ定義することなく、両者を並べて用いてきた。それは、本稿の筆者らの判断というわけではなく、Davidov et al. (2011, 2014, 2018)、Cieciuch et al. (2019)、van de Schoot et al. (2015)、Meitiger (2020) などの文献において、そのような用語法がとられているからにはほかならない。

しかし、ほぼ20年前に真鍋が実施した国際比較／文化比較に関する文献研究の結果（真鍋、

1999, 2004 a, 2004 b）からするならば、当時の文献では、もっぱら「測定の不変性」という用語が用いられており、その用語は、さまざまな国／文化において実施される質問紙調査が「同じもの」を捉えているかどうかの確認というコンテキストで使用されていた。より具体的にいうならば、それは、国際比較／文化比較調査の実践の諸フェーズ——「質問紙の作成」「質問紙の翻訳」「調査対象のサンプリング」「実査」「データの収集・整理・コーディング」「データ分析」などの諸フェーズ——において、いかにして「同じ調査」「同じデータ処理」「同じデータ分析」を行なうかといったところに焦点が合わされていたということである。

それとは対照的に、近年の文献において、新しく用いられるようになった「測定の不変性」という用語には、上述のような「測定の不変性」という意味内容に加えて、あるいは、それに替わって、新しく、人びとの主観的意識についての構成概念とその測定のための質問諸項目／諸スケールの国際比較／文化比較の可能性（cross-national/cross-cultural comparability）についての実証的／統計的な検定（empirical/statistical test）という意味内容が含まれることとなった。

こうして、後者の「測定の不変性」という問題は、この領域の「研究動向」における、まさに「新しい問題」として登場したといえるのである。

7. 以上のような経緯で、国際比較／文化比較の視座からするデータ分析において、「測定の不変性」の確認という手続きがきわめて重要な課題となってきた。測定の不変性は、実証的に「検定（test）」を行なうことができる。測定の不変性の検定のために、最もよく利用されている技法は、「多集団確証的因子分析」である（Davidiv et al., 2014; Cieciuch et al., 2019; Meitinger et al., 2020）。そして、その技法の考案者が、Jöreskogであった。Jöreskogの参照文献としては、Jöreskog (1969, 1971, 1973) があげられる。この研究領域における日本の第一人者の一人と目される狩野裕 (2002) は、「構造方程式モデリング（structural equation modeling: SEM）」の端緒は、Jöreskogが「確証的因子分析」を提案（1969）

し、さらに因子分析に「因果分析」の考え方を取り入れる (1973) ことによって、開かれたとして、これら2つの文献に注目する。しかし、すでに述べた「応用研究」の領域の研究者たちは、「多集団確証的因子分析」の出発点となった Jöreskog (1971) を「レファレンス・スタディ (reference study)」とする。

8. その後、このような Jöreskog の SEM・確証的因子分析・多集団同時分析の理論とアイデアは、Sörbom との共同開発 (1984, 1989, 1996) をとおして、LISREL と呼ばれるコンピュータ・ソフトウェア・プログラムの形へと発展していく。そして、それは、いまでは、主要な SEM のソフトウェア・プログラムのすべてに広く行き渡るようになった。

9. こうして、MGCEFA と呼ばれる統計的技法についての伝統的な形 (traditional form) が完成する。その後、MGCEA を用いた応用研究は大きく変化していくが、「統計学の近年の発展は、このような研究領域に対して、新しい分析のアイデアと道具 (tools) を提供し続けている」(van de Schoot et al., 2015, p.2)。思うに、MGCEFA をめぐる研究の発展は、このような「応用研究」と「統計的方法の研究」との相互影響の連続過程として捉えられるものに違いない。そのような過程の詳細な実証的な解明は、今後に残されたきわめて興味深い課題といえよう。ここでは、このような問題の所在を指摘するにとどめる。

Ⅲ. MGCEFA を用いた測定の等価性／不変性の検定

——そのアイデア・技法・手続き——

以上においては、国際比較／文化比較調査における測定の等価性／不変性の研究の系譜を概観してきた。そこで、以上の概観を踏まえて、つぎに、このような測定の等価性／不変性の実証的な検定を可能にする MGCEA のアイデア・技法・手続きを、一覧表に掲げた諸文献、とくに Davidov et al. (2014)、Ciecuch et al. (2019)、van de Schoot et al. (2015)、Meitinger et al. (2020) な

どのレビュー論文にもとづきながら、再びコンパクトな形でまとめることを試みる。

1. 測定の等価性／不変性とは何か

「測定の等価性／不変性」という用語は、すでに本稿において、たびたび使用してきたし、その意味内容についても、簡単な解説は加えてきた。ここでは、諸文献の記述にもとづきながら、そのような内容を、もう一度、整理しておきたい。

「測定の等価性／不変性」とは、異なる集団 (different groups) ——さまざまな国／文化、あるいは、時点——において、同じ質問紙 (questionnaire) を用いて、同じ構成概念 (constructs) を、同じように測定する (measure in the same way) ということを意味する (Ciecuch et al., 2019; Davidov et al. 2014)。「測定の等価性／不変性」の確認がなされなければ、諸集団の比較には意味がない (not be meaningful)。それは、集団間に観測される (observed) 差異が「真の差異 (true difference)」であるか、それとも、「方法論的な人工物 (methodological artifact)」にすぎないのか、を判断することができないからである (Ciecuch et al., 2019; Davidov et al., 2014)。

このような考え方にもとづいて、「測定の等価性／不変性」を確認するために、以下のようなさまざまな統計的技法が開発されてきた。

- ・探索的因子分析 (exploratory factor analysis: EFA)
- ・多集団確証的因子分析 (multigroup——あるいは multiple group——confirmatory factor analysis: MGCEFA)
- ・多次元尺度法 (multidimensional scaling: MDS)
- ・項目反応理論 (item response theory: IRT)
- ・潜在クラス分析 (latent class analysis: LCA)

これら統計的技法の歴史的な概観とそれぞれの technical な特徴 (feature) については、Millsap and Meredith (2007)、Braun and Jhonson (2010) などを参照されたい。ここでは、国際比較／文化比較というコンテキストにおいて、これまで最もよく用いられてきた「多集団確証的因子分析 (MGCEFA)」に焦点を合わせる。

では、MGCFA と呼ばれる統計的技法は、どのように解説されるであろうか。Davidov et al. (2014) は、この技法が「一般化潜在変数アプローチ (a generalized latent variable approach)」を取るものとして位置づける。このアプローチは、社会科学の理論的な概念 (theoretical concepts) というものは、一般に、「構成概念」であって、直接に観察できるものではない (つまり、「潜在変数」である) という考え方にもとづいている。したがって、それは、そのような潜在変数を測定する「多重観測 (顕在) 指標 (multiple observed (manifest) indicators)」から推測 (inferred) されるものであるとする。このような考え方は、以下の図2を用いて説明される。

この図では、ある理論的な概念 (構成概念) である η を、A と B の2つのグループで比較するものとしている。ここで、その理論的な概念 (η) は、直接に測定されるのではなく、3つの質問項目 (χ_1 - χ_3) からなる「バッテリー (battery) 型質問」の形に操作化 (operationalized) される。 λ は「因子負荷量 (factor loading)」、 τ は「切片 (intercept)」、 δ は「測定誤差 (measurement error)」をそれぞれ示している。

こうして、「潜在変数アプローチ」の鍵となる考え方は、「測定の等価性/不変性」は2つのグループにおいて、「潜在変数」と「諸指標 (indicators)」との実証的な関係 (empirical relations) を比較することによって、統計的に検定することができるというものである。そして、グループ間におけるその関係の類似性 (similarity) は、測定のパラメータ (parameter) ——「因子負荷量」と「切片」——によって判断され、それが「測定の等価性/不変性」の実証的な証拠 (empirical evidence) とされるのである。

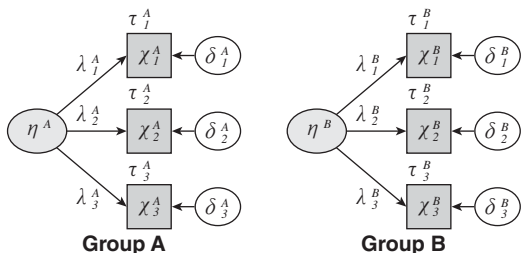


図2 多集団確証的因子分析モデルの図形的表現

Davidiv et al. (2014) は、このように、「潜在変数アプローチ」の基本的な考え方を確認した上で、つぎに、MGCFA の technical な解説へと進む。そこで取りあげられるのが、Jöreskog (1971) の「線形回帰方程式 (linear regression equation)」である。

$$x_i^g = \tau_i^g + \lambda_i^g \eta^g + \delta_i^g.$$

この数式において、「結果となる変数 (結果変数: 被説明変数)」は左辺に、そして、その「原因となる変数 (原因変数: 説明変数)」は右辺に配置されている。それぞれの記号については、 x_i^g は集団 g の i 番目の人の観測スコア (observed score)、 τ_i^g は同じく回帰モデルの切片 (intercept of the regression model)、 λ_i^g は因子負荷量 (factor loading)、 η は潜在因子スコア (latent factor score)、 δ_i^g は測定誤差を、それぞれ表している。

以上のような、MGCFA についての technical な定義にもとづいて、「測定の等価性/不変性」は実証的に確認することが可能となるのである。

2. 測定の等価性/不変性の検定はどのように行なうか

測定の等価性/不変性は、「多集団確証的因子分析 (MGCFA)」という統計的な技法の確立によって、実証的に検定を行なうことが可能となった。

しかし、このような「測定の等価性/不変性」の検定は、単にそのような等価性/不変性が「確認できるか」それとも「確認できないか」という二分法的な判断ではない。それは、そのような等価性/不変性の「レベル」を問題にする。「より高いレベルの等価性/不変性」が確認されるならば、「より広い形 (form) での国際比較/文化比較および時点間比較」が可能となる。そして、「より低いレベルの等価性/不変性」の場合は、「より狭い形での国際比較/文化比較および時点間比較」にとどまる。このような、等価性/不変性のレベルは上述の「線形回帰方程式」の形で定式化されるモデルに課せられる「制約 (constraint) の多少」——制約が多いか、それとも、

少ないかということ——に対応している。いうまでもなく、「より制約の多いモデル (more constrained model)」は、「より高いレベルの等価性／不変性」に、そして「より制約の少ないモデル (less constrained model)」は、「より低いレベルの等価性／不変性」に、それぞれ対応している。

こうして、「制約の少ないモデル」から「制約の多いモデル」へと「階層モデル (hierarchical model)」が構成される。それは、configural → metric → scalar の3つの段階的なレベルのモデルである。因みに、これらの用語については、この研究領域の日本のパイオニアの一人とされる豊田秀樹による日本語訳の試みがある（例えば、豊田、2007など）が、いまだ定訳とはなっていない。そこで、本稿では原語をそのまま用いることにする。

そのような制約の多少に対応するレベルは、具体的にいうならば、つぎのとおりである。

configural level : 因子負荷行列のゼロでない自由パラメータの位置が同じであるということ以外、何らの制約も置かれない。

metric level : 「因子負荷量」が国／文化間で等しい (equal) という制約が置かれる。

scalar level : 「因子負荷量」と「切片」の両方が国／文化間で等しいという制約が置かれる。

このような制約とモデルとの関係を、Ciecuch et al., (2019) は、図3を用いて説明する。この図において、X軸は潜在変数 (latent variable) の平均値 (mean value) を示し、Y軸は潜在変数を

測定する質問紙調査の項目 (survey question item) に対する回答の値を示す。対角線 (diagonal) は、ここで分析に取りあげる2つの国における、潜在変数と調査項目に対する回答との関係を示している。

(a) は、configural invariance のレベルを示している。ここでは、同じ潜在変数を捉えることをめざしながらも、2か国についての2つの斜線の勾配 (つまり、因子負荷量) も、これら2つの斜線とY軸との交点 (つまり、切片) も、いずれも異なる値を示している。

(b) は、configural invariance とともに、metric invariance のレベルを示している。ここでは、2か国の斜線の勾配 (つまり、因子負荷量) が同じであるが、これら2つの斜線とY軸との交点 (つまり、切片) は異なる。

(c) は、configural invariance、metric invariance とともに、scalar invariance のレベルを示している。ここでは、因子負荷量も、切片も、いずれも2か国で等しい。したがって、斜線は一本となっている。つまり、2か国において、潜在変数の平均値も、調査項目に対する回答の値も、いずれも、同じであることを示しているのである。

確かに、このような「図式化」による解説は、上述の「数式化」の場合にくらべて、より直観的な理解を可能にするものといえよう。

いずれにしても、「測定の等価性／不変性」については、以上の3つのレベルが区別される。それでは、ある特定の研究で取りあげる国／文化、あるいは時点のデータにおいて、どのレベルの「測定の等価性／不変性」——configural invari-

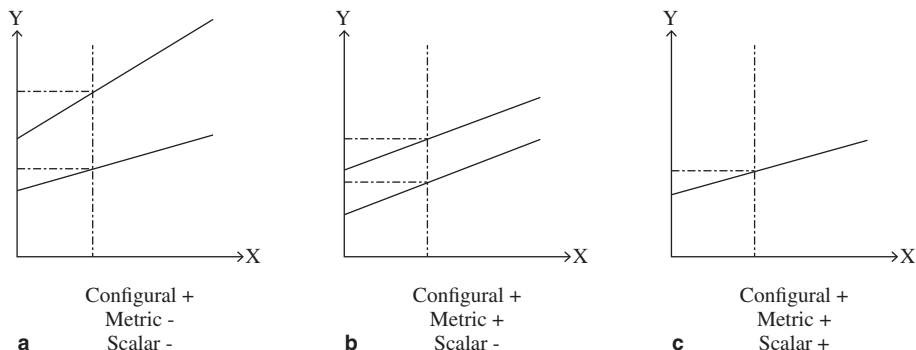


図3 Configural Invariance, Metric Invariance, Scalar Invariance の説明図 (Ciecuch et al. 2019)

ance, metric invariance, scalar invariance——が確認されるかは、どのようにして判断されるであろうか。それは、統計的な検定の「手続き (procedure)」をとおしてなされる。

その「手続き」は、具体的な例をあげていうならば、例えば、あるデータにおいて、まず configural invariance が確認されたとするならば、つぎに、metric invariance に対応する制約を導入し、そのモデル適合度指数 (model fit index) が低下するということがなければ、さらに scalar invariance の制約を導入し、それでモデル適合度指数が低下することになったとするならば、そのデータについては metric invariance モデルのレベルまでが支持されたとして、その測定の不変性のレベルが確定することになる (Cieciuch et al., 2019)。この進め方は、bottom-up approach と呼ばれる。

しかし、これとは逆に、scalar レベルが達成されるかどうかから始めて、それが達成されない場合は、つぎに metric レベルは達成されるかどうか、というように確認を進めていくやり方もある (Jagodzinski and Dülmer, 2019)。この進め方は top-down approach と呼ばれる。

では、以上のような判断の基準となる「モデル適合度指数」には、どのようなものがあるであろうか。かつては、カイ 2 乗値が用いられていた。しかし、モデル適合度の判断のための「カイ 2 乗検定 (chi-squared test)」には、それがデータの数の影響を受けるなどの問題点が指摘され、それに替わって、「近似平方根平均平方誤差 (root mean square error of approximation: RMSEA) や「比較適合度指数 (comparative fit index: CFI)」などの利用が一般的になってきている (Davidov et al., 2014)。

以上のような「手続き」を取ることによって、MGCFI という統計的技法を用いて、測定の不変性のレベルを確定することが可能となる。そのような手続きの確立は、この領域における重要な「方法論的研究」——Davidov らの用語であり、筆者らの「統計的方法の研究」とほぼ同じ内容のものといえる——の成果というべきものである。ところが、このような手続きを、さまざまな substantive なテーマをめぐる具体的な「応用研究」に当てはめてみるならば、高いレベル——例

えば、scalar レベル——の測定の不変性が確認されるのは、きわめて稀であるということがわかってきた。Cieciuch et al. (2019) は、以下のような研究事例をあげている。

- ・ Lomazzi (2018) は、性役割についての人びとの態度 (gender-role attitudes) は、「世界価値観調査 (WVS)」のすべての調査対象国について比較できるわけではないことを例証した。
- ・ Sokolov (2018)、そして Aleman and Woods (2016) は、Inglehart の価値観スケールでは、分析に取りあげたすべての国の比較ができるわけではないことを示した。
- ・ Coromina and Davidov (2013) は、社会的・政治的な信頼 (social and political trust) は、さまざまな国と、さまざまな時点の両方において、あるいは、そのどちらかにおいて、常に比較が可能であるわけではないという結論に到った。
- ・ Ariely and Davidov (2010) は、民主主義に対する人びとの支持 (public support for democracy) は、「世界価値観調査 (WVS)」の国ごとに比較することはできないことを見出した。
- ・ Davidov (2009) は、ナショナリズムと愛国心 (nationalism and patriotism) のスケールの平均値は、「国際社会調査プログラム (ISSP)」の参加国間で比較できないことを発見した。
- ・ Davidov (2008) は、「ヨーロッパ社会調査 (ESS)」による人びとの価値観 (human values) の平均値は、すべての国について比較できるわけではないことを示した。

3. 測定の非不変性という問題にどのように対処するか

以上に述べてきたように、測定の等価性／不変性に関する「方法論的研究」の成果を国際比較／文化比較の領域に導入するさまざまな「応用研究」の試みをとおして、実際の社会調査——「実践研究」——のデータにおいて、高いレベルの測定の等価性／不変性を確認することは、きわめて困難であることがわかってきた。では、このような問題——高いレベルの測定の等価性／不変性が確認できないという問題：これは「測定の非不変性 (measurement non-invariance)」の問題といわ

れる——は、どう考えればいいのであろうか。このような問題については、つぎの2つの考え方がある (Cieciuch et al., 2019)。

①測定の等価性／不変性についての伝統的な考え方：測定の等価性／不変性の検定における厳格な制約 (strict constraints) を保持 (preserve) しようとする考え方 (Steinmetz, 2018 ; De Beuckelaer and Swinnen, 2018 など)。

②測定の等価性／不変性についての現実的な考え方：測定の等価性／不変性の検定における厳格な制約を緩和 (release/relax/tolerate) しようとする考え方。

そして、②の考え方から、測定の不変性を full invariance と partial invariance に区別し、国際／文化比較という課題にとっては、partial invariance で十分であるとする考え方がでてくる。こうして、以下の4つのモデルが提案されることになる (Byrne et al., 1989 ; Steenkamp and Baumgartner, 1998)。

- ・ partial metric invariance : 構成概念の少なくとも2つの質問項目の「因子負荷量」が、複数の国／文化間で等しいという制約を置いたモデル。
- ・ partial scalar invariance : 構成概念の少なくとも2つの質問項目の「因子負荷量」と「切片」が、複数の国／文化間で等しいという制約を置いたモデル。
- ・ full metric invariance : 構成概念のすべての質問項目の「因子負荷量」が、複数の国／文化間で等しいという制約を置いたモデル。
- ・ full scalar invariance : 構成概念のすべての質問項目の「因子負荷量」と「切片」が、複数の国／文化間で等しいという制約を置いたモデル。

因みに、「S. Schwartz の概念・理論枠組みにもとづく価値観の国際比較——ドイツと日本における『大学生調査』のデータ分析——」と題するドイツ・ケルン大学との共同研究 (真鍋, 2018 ; 真鍋ほか, 2020 a, 2020 b) において、Jagodzinski and Dülmer (2019) は、Schwartz の19の価値観スケールのうちの16については、ドイツと日本

の大学生で、少なくとも partial scalar invariance のレベルが確認されたことを報告している。

このような知見 (findings) は、文字どおり、上述のような「方法論的研究」の成果が「応用研究」に取り入れられた結果を示すものといえよう。では、「応用研究」の成果が「方法論的研究」に取り入れられた事例には、どのようなものがあるであろうか。このような探索は、今後に残された重要な課題の1つといえよう。

いずれにしても、測定の等価性／不変性の研究をめぐる「レビュー論文」によるならば、MGCFA という技法を用いた検定の以上のような「手続き」は、“traditional approach” と呼ばれるものであり、それは、その後の “approximate approach” とは区別されるものであるという。“approximate approach” と呼ばれる新しいアイデア・技法・手続きがでてくる背景には、“traditional approach” に対する方法論的な批判があった。それは、これまでの MGCFA を用いた測定 of 等価性／不変性の検定が、①あまりにも「厳密 (exact)」を求めすぎる、②あまりにも「厳格 (strict)」になされすぎる、という批判であった。そのため、“traditional approach” は、“exact approach” あるいは “strict approach” とも呼ばれたのである。

このような批判に応えるために、再び、さまざまな「統計的な解決策 (statistical solution)」が提案されてきた。それらのなかで、現在、最も注目されているのが、①Bayesian approximate method と、②alignment optimization method の2つである (Meitinger et al., 2020)。前者に関する文献としては、Muthén and Asparouhov (2012), van de Schoot et al. (2013) が、後者に関する文献としては、Asparouhov and Muthén (2014)、Muthén and Asparouhov (2017) があげられる。

ここでも、本稿の「問題関心」からして重要であるのは、このような「方法論的研究」の成果に続いて、それとほぼ時を同じくして、それらのアプローチを国際比較／文化比較調査のデータ分析に取り入れる「応用研究」がでてきたということである。前者①については、Cieciuch et al. (2014)、Zercher et al. (2015)、Seddig and Leitgöb (2018) などが、そして、後者②については、Cieciuch et

al. (2018)、Cieciuch et al. (2019) などがある。

以上においては、MGCFA という技法を用いた測定の等価性／不変性の検定について、それがいったん “exact approach” という形で方法論的に完成されるが、その後、さらに “approximate approach” へと展開していく過程を概観してきた。このような “approximate approach” のアイデア・技法・手続きについては、章を改めて、解説していくことにする。

IV. 測定の等価性／不変性の検討法の新しい展開

—Exact MI から Approximate MI へ—

本章では、厳密な測定等価性／不変性から、近似的なそれに向けてのいくつか統計的な対処のうち、①ベジアンBayesianの近似的等価性検討法と、②Alignment 最適化による方法、の概要を、前述の文献から簡単に紹介しよう。本稿では、前者については、Seddig and Leitgöb (2018) を、後者についてはオリジナルの手法提案論文である Asparouhov and Muthén (2014) とそれを解説した Cieciuch, Davidov and Schmidt (2018) を、それぞれ主に参照する。

これらの新しいアプローチと対比される形で traditional と呼ばれるアプローチは、前章に述べたようにその厳密 (exact) さ、厳格 (strict) さがいわば「足かせ」となっているのであるが、その内実は、metric invariance であれ、scalar invariance であれ、それがグループ間で厳密にパラメータが等しい値であるという仮説の形で提示され、検討される点にある。これが全てのパラメータについて成立する状態 (full invariance) は、規模の大きな国際比較調査のデータなどでは考えにくく、現実的な妥協案として何らかの緩和が必要となり、そこに部分的な不変性 (partial invariance) の考え方が導入される点までを前章に述べた。ここで、その部分的な不変性を前提とし、最終的な目的 (の一つ) である、因子平均のグループ間比較が意味のあるものであることを担保するために、具体的に導入される方法として上に述べた二つのアプローチが注目を浴びているのであ

る。②の Alignment 最適化の方法は①と相容れないということではなく、このアプローチの中でも①のようなベジアンによるアプローチも、あるいは頻度主義の王道である最尤推定法によるアプローチも採用可能である。

従来法や①では metric invariance や scalar invariance そのものを、指標変数と因子の関係のみで検討しており、その後の因子の平均比較は直接手続き内に含まれていない。これらのアプローチに比べると、②は因子平均 (と分散) の比較までを統一的な枠組で解決する点で、応用面ですぐれているように筆者には感じられる。

ともあれ、両者のアプローチについて、順に概観しておこう。

1. ベジアンBayesianの近似的等価性検討アプローチ

近似的等価性アプローチについて、本章冒頭で厳密 (厳格) さが過ぎる点での足かせについて述べたが、この点は、とくに MGCFA のような因子分析に限らない。より単純な2グループ (以上) の1変数の平均値の比較のような状況で、2グループ (以上) の平均値が等しいという帰無仮説は、サンプルサイズを非常に多くとればほとんどの場合に棄却される (2つ以上のグループにおいて母数の値が厳密に等しいという状態は事実上あり得ない) という頻度論的なアプローチによる仮説検定が常にかかえる問題と本質的に同じ状態である。例えば、たとえ1個の測定変数だけに着目したとしても、metric invariance で要請される「因子負荷量パラメータが検討している複数のグループで厳密に等しい」ということはあり得ず、サンプルサイズを非常に多くしておけば、仮説検定によるアプローチではその差が検出されてしまう。検定論の皮肉は、このようにサンプルサイズを大きくして調査の精度を上げれば上げるほど、厳密な等価性の帰無仮説は棄却されやすくなる点にある。こうした足かせへの一つの救済策をベジアンBayesianのアプローチが提供するという文脈で、以下の方法も理解することができる。

さて、前章にも述べた伝統的な MGCFA アプローチでの、「等値制約」は測定上の性質を表すパラメータである因子負荷や切片のグループ間の差が厳密に“ゼロ”であることを仮定するが、そ

れはいつも達成可能ではない。因子負荷を意味する行列 Λ や各観測変数の切片ベクトル τ の要素が、 G 個のグループ間ですべて等しい値であるという仮定は、全体的なモデル適合度の指標が十分な値にならないことで否定されてしまうことが多い。他方、より制約の緩いモデルでは実質的に全体的モデル適合度が改善するが、グループ間でパラメータが等しいとの主張を放棄せねばならない。従来よく用いられてきた分析の実際的手順としては、より多くの等値制約が入った状態のモデル、たとえば scalar invariance を仮定した状態から、修正指標 (modification index) などの統計的指標を利用して、適合度をより大きく落としてしまっているパラメータから順に等値制約を緩めていく方向で検討を進めるのが一つの方法である。

ベイジアンアプローチでは、また少し別の考え方で、厳密な等値制約よりも少し制約の弱められた状態の表現を導入する。その要点は厳密な等値制約が、近似的な等値制約に置き換えられ、パラメータのグループ間差が平均がゼロで小さな分散を持つような何らかの分布 (通常は正規分布) に従うものと仮定される点にある。ベイジアンアプローチではそのようなグループ間の差が、分析者によって事前分布の形で指定される。背後にある考えは、「両方 (厳密な等価性と高いモデル適合度) にこだわると進退窮まるので折衷的な道を見つけるべきである。受け入れ可能な適合度を犠牲にして、厳格な等価性を求めることも、比較可能性を犠牲にして、パラメータの集団間での差をあまりに認めすぎのもよくない」という点にあるように思われる。

伝統的なアプローチでは $\Lambda_1 = \Lambda_2 = \dots = \Lambda_G$ のように等号で結ばれていた各グループの因子負荷行列の値が、近似的に値が等しいことを意味する \approx 記号で結ばれた状態 (と見なすこと) の妥当性を検討することになる。近似的に等しいということは、グループ間の差がおおよそ 0 であることを意味し、「おおよそ」の程度は事前分布のバラツキ (分散) の大ききで表現しておくことになる。

本章はベイジアンによるデータ解析のアプローチ全般を説明する余裕はないが、ベイジアン推論の道具立ては、特定モデルの下での手元のデータの発生確率を表現する尤度と、モデルを特徴づ

けるパラメータの事前分布に基づき

事後分布 \propto 事前分布 \times 尤度

により計算される (パラメータの) 事後分布に基づいて構成される点に特徴がある。

複雑な統計モデルの下で、パラメータの事後分布を求めることは一般的には容易でなかったが、計算機の発達により、乱数を用いたシミュレーションにより数値的に求めることが容易になり、マルコフ連鎖モンテカルロ法 MCMC などのアルゴリズムがよく用いられるようになった。こうした具体的な技法の詳細は割愛するとして、ベイジアンアプローチでも事後分布に基づく推定値の評価は必要である。例えばグループ間のパラメータの差の事前分布にどの程度の分散を持つ分布を仮定すればよいか (小さい分散を仮定するほどグループの値が相互に近いという信念を表現したことになる)、という点についても、どの程度の仮定がよいのかという点を評価する指標が必要となる。唯一の方法があるわけではないが、ベイズ推論に適した DIC と呼ばれる情報量規準に基づくモデル選択の方法や、それ以外の指標に基づく選択が推奨されている。

どちらにせよ、グループ間の等しさの程度は、事前分布の分散が 0 (グループ間の差がないことを事前分布として表現したことになり、これは伝統的な厳密な等値制約を意味する) モデルから、グループ間の差が大きな分散を持つ (実質的に等値とは言えないような) モデルまで、分散の値を小刻みに変化させて、その中のどの程度の差 (分散の大きさ) までが受入可能であるか、という観点で最適なモデルの選択を行う。以上がベイジアンによる近似的な測定等価性アプローチの骨子となるアイデアである。

以上は一つのみパラメータ (例えば因子負荷) に関わる推論を前提に述べたが、実際にはこれに関与する全てのパラメータ (因子負荷と切片) に拡張してモデルの選択を行った結果として、近似的な測定等価性が成り立つか否かを検討するところまでを扱ったものであり、その後の因子平均の比較部分についての検討は改めて行うことになる。

2. Alignment 最適化によるアプローチ

Alignment 最適化のアプローチは、ベイジアン近似的等価性アプローチと違い、因子平均の比較までの手続きを統一的に取り扱う枠組として提案されている。Alignment Optimization の最適な日本語訳が既に提案されているとは思われないので、最初の単語は英語のままとして、Alignment 最適化と表記しておくことにしよう。

この手法を提案した Asparouhov and Muthén (2014) によれば、探索的因子分析における因子の回転に喩えることができる手続きである。探索的因子分析では、何らかの推定法で初期解を（直交解として）求めた後、回転に関する不定性というモデル上の性質を利用して、因子負荷行列に回転行列をかけることによって新しい（回転後の）因子負荷行列を求める。回転の不定性という性質は、回転の前後でモデルのデータに対する適合度が変化しないことを意味している。Alignment 最適化では、探索的因子分析では初期解に相当するのが「配置不変」configural invariance の状態、回転後の解に相当する最終的な解は、configural invariance モデルと同等の統計的な適合度を保ちつつ「部分的な metric invariance」を許容した解、ということになる。このような解は複数存在するが、適当な損失関数を定めてその最小化を図ることを以て、それらの中での最適解を求めようとするアプローチであるため、optimization の語が使われている。

もう少し具体的に述べておこう。原論文と同様に、一因子のモデルを仮定して説明する。初期解としての configural invariance の状態では、各群の因子平均は 0（分散は 1）のように制約される（より正確にはそのように再表現することが可能である）こと、そして因子負荷行列のゼロでない自由パラメータの位置が同じということ以外は、因子負荷と切片のパラメータに集団間の等値制約はおかれていない。これが因子を集団間で比較するということの出発点となる最低限の条件で、その意味での検討価値があるあるモデルの中で（最も制約が緩く）適合度が高いモデルである。ここでグループ g の因子平均 α_g と分散 φ_g というパラメータに平均 0 分散 1 というような制約が必ずしも入っていない状態を仮定しても、configural

invariance と同じ適合度を持つような各グループの“新しい（変換後の）”因子負荷ベクトルと切片ベクトルの組を考えることができ（この性質に着目した点と同論文の本質的な貢献と思われる）、この同等の適合を持つ新しい因子負荷ベクトルと切片は、configural invariance モデルのパラメータと因子平均 α_g ・因子分散 φ_g の関数として再表現できる。Alignment 最適化では、再表現後の因子負荷行列と切片に関して等値制約が満たされない状態を loss（損失）と考え、 G 個のグループから 2 個を取り出す全ての組み合わせについてこの損失を（重み付きで）加算した損失関数を導入する。この損失関数は、各グループの α_g と φ_g を変数として最適化され、損失を最小化するような解は、測定非不変性 measurement non-invariance つまり「測定不変性が成り立たない」状態が総体として最も低められた状態の解、と理解することができる。繰り返しになるが、この解が configural invariance のモデルと同等の適合度を維持している点が、提案手法のすぐれた点である。

損失関数の具体的な形については、テクニカルな話題になるから割愛することにしよう。それは「多くのグループ間で小さな差が観察される状態よりも、少数のグループの組み合わせ間で大きな差が観察され、残りの組み合わせでは等値と考えてよい状態のほうが損失が小さくなる」ような形で定義されている。このことから、この手法は「configural invariance モデルと同等の適合度を持つ中で『可能な限り少数箇所でのみ非等値性を許容し、残りでは等値性が成り立っている』ことを目標として解を求めようとするものである」と性格づけることができる。この目標は、探索的因子分析の因子の回転で「単純構造」（特定の観測変数は複数の因子に中程度の負荷量を持つより、1 個の因子にのみ大きな負荷を持ち、残りの因子には小さな負荷量しか持たない）を求めることが目標となることになぞらえることができる。小さな差が多数の箇所で見られるよりも、大きな差が少数箇所で見られる状態のほうが「単純」な構造になっているというわけである。

原論文では最適化問題の難しさを緩和するための工夫などが導入されているが、それでもなお関数の形状は複雑で、最適化手法によっては局所最

小解になっていないかを確認するために、複数の初期値から何度も最適化計算を行い、大域的最小に到達したことを確認する必要についても述べている。

Alignment というのは、英単語としては「一列に並んだ状態、整列、(兵士などの) 隊列」ほどの意味であろうが、原論文ではこの語を選んだ理由も次のように説明されている。理想状態としての full metric invariance の状態は、因子負荷と切片に関する全てのパラメータが集団間で等しい状態で、例えば2グループの場合に2次元平面上でその全パラメータの(推定) 値を、横軸にグループ1、縦軸にグループ2の値を座標としてプロットすると45度の一直線上に並ぶことになる。これが Alignment の理想が成立している状態であり、partial invariance は一部のパラメータの等値性が成り立たず一直線に並んだ状態(隊列) が乱れた状態、いわば隊列に凸凹が生じた状態に相当する。初期解の当てはまりを維持しつつその隊列の凸凹の箇所がなるべく少なくなるような(最適) 解を求めようというのが、このアプローチの目標である。

最適解を求めた後は、因子平均 α_g と分散 ϕ_g の標準誤差の推定値を求めることもでき、それらに基づく統計的な推論が可能となる。具体的には(等値性が成り立たないパラメータの箇所をなるべく少数個に抑えた上で)、因子平均がどのグループとどのグループで有意に異なるのか、因子分散はグループ間でどの程度異なるのかといったことの推論が可能になる。因子の平均や分散の検討が、国際比較調査などの多集団比較での最大の目標となるので、この時点で目標が達成されていることになるが、副産物としてどのグループ間の因子負荷や切片パラメータが等しい(等しくない)のか、という点の確認もでき、従来の measurement invariance の手続きで検討されてきた事項も事実上カバーしていることになる。

この方法でも、実際にどの程度の割合の因子負荷と切片のパラメータが等値と想定できなくてもよいのか(全体としてグループ数×観測変数の数×2のパラメータがあるうち、何箇所が他グループの値と異なるかという形で計数可能である)という点に関する明確な基準はないが、原論文では

シミュレーション研究から等値が仮定できない箇所が25%以内であれば実用上の問題は少ないだろうという主旨の提案をしている。分析を実行する上でも細部でいくらかのバリエーションがあるのだが、それらの紹介は別の機会に譲ることしよう。

以上見てきたように、この方法は実質的には、多グループ間での因子比較について非常に実用的な分析オプションを提案しており、著者らにより開発されているソフトウェア Mplus にも実装済みである。この方法には、因子負荷や切片のみならず因子そのものの平均や分散の比較に関する統一的な枠組を与えており、グループ数がかかり多い場合でも(計算量は多くなるものの)同じ考え方で対処可能であるといった利点もある。いくつか検討されるべき事柄が残っているようには思われるが、Alignment 最適化のアプローチは、国際比較における因子比較の問題に、大変魅力的なツールを提供したと評価することができよう。

V. おわりに

本稿は、出発点において、その目標を明らかにした。それは、ここでのテーマに関するこれまでの「研究の全体像の把握」と、この研究領域における「新しい飛躍の種の発見」ということであった。すでに、これまでの各章において、前者の側面については、いうまでもなく、後者の側面についても、そこで取りあげた問題ごとに、今後の研究の方向という形で、「新しい飛躍の種」を示唆してきた。したがって、ここで、それらを再度とりまとめるとするならば、それは、いわば「情報の redundancy」とでもいべきものになる。しかし、後者の側面については、本文中において、そのような「飛躍の種」の方向を、一般的な形で示唆するにとどまり、それを具体的な形で提示するまでには到らなかった。しかし、このような今後の「飛躍の種」に向けての研究内容の提案は、R. K. Merton (1957=1961) の用語でいうならば、「一般的な方針」としてではなく、「具体的な方略」として提示されるべきものといえよう。いうまでもなく、それが「知的生産」という視座——梅棹忠夫の『知的生産の技術』岩波新書、1969

年)を踏まえている——からして、より有効な行き方であると考えるからにはかならない。

そこで、本稿では、最後に、この研究領域における「新しい飛躍の種」の発見という視座からする、今後の研究課題として、以下の3点をあげておきたい。

1. 本稿における筆者らの中心的な問題関心は、図1に示したように、「測定の等価性／不変性」をめぐる3つの研究領域——「国際比較／文化比較調査の実践と方法に関する研究」「統計的方法に関する研究」「統計的方法の国際比較／文化比較研究への応用に関する研究」——の相互の影響の過程の解明ということである。ここで、これら「3つの研究領域の相互の影響の過程」というのは、筆者らの「仮説的命題」というものである。確かに、この「命題」は、筆者らによって独自に構成されたものである。しかし、それは筆者らの「イメージーション」だけで作られたわけではない。本稿のⅡのところでも説明したように、それは関連諸文献の渉猟と精査を踏まえた上で、提案されたものである。そこで、そのような関連諸文献における、上述の「3つの研究領域の相互の影響の過程」についての「記述のスタイル」ということが問題となる。このような点について、より具体的に説明するために、つぎの2つの例をあげてみたい。

(1) van de Schoot et al. (2015, p.2) は、MGCFAによる測定の等価性／不変性の研究領域に新しくでてきた問題——測定の非不変性という問題——の提示に続いて、「統計学における近年の発展が、測定の不変性のアセスメントのための、新しい分析の道具 (tools) を提供してきた」と述べている。

確かに、この一文は、「統計的方法の研究」と「応用研究」との相互の影響を示唆したものに違いない。しかし、この一文だけでは、事実上、①統計学の領域で、いわば「内発的に開発されてきた新しい技法」が、その後、「応用研究」の領域で、問題解決のための技法として取り入れられることになったということなのか、それとも、②「応用研究」の領域からの新しい分析技法の開発

の要請があって、その要請に応える形で、統計学の領域で、そのような開発がなされることになったということなのか、は明らかでない。このような点についての実証的な解明こそが、その研究領域に新しい飛躍をもたらす1つの契機となるのではなからうか。

(2) Millsap and Meredith (2007) は、過去数十年にわたる測定の等価性／不変性のアセスメントのためのさまざまな分析的アプローチ (analytical approaches) について、歴史的な概観を行なっているが、その結論ともいべきところで、「測定の不変性について研究する者は、同時に、すぐれた心理学者でなければならない」と述べている。この一文は、心理的尺度 (psychological scale) についての測定の不変性の研究という文脈で書かれているものである。しかし、そのようなテーマの研究においては、「統計的な道具に関する知識」とともに、「心理的尺度の理論的な内容に関する知識」も必要であるという主張は、当然といえば当然の主張というものである。ここでの議論は、このような研究の実践上の「処理的な知識」——Berger and Luckmann (1966 = 1977, 日本語訳 p.72) の用語——という問題の範疇にとどめることなく、さらに「methodological な研究」と「substantive な研究」との統合にもとづく実証科学における新しい「知の創造」への取り組みといった方向に向けて展開していくべきものであろう。この点については、詳細な議論を展開するだけの紙面の余裕はない。ここでは、Davidov ら (2012 など) の「Schwartz の価値観理論の新しい展開」をめざした「応用研究」を、このような方向を示す研究事例として位置づけておきたい (真鍋、2016 を参照)。

2. 本稿のⅡのところでも、MGCFA という統計的手法を用いた測定の等価性／不変性の研究の系譜を辿ることをとおして、その領域に、新しい研究動向が出現したことを確認した。その経緯は、MGCFA の検定の手続きを、国際比較／文化比較研究に導入した場合、そこで高いレベルでの測定の等価性／不変性が確認されることは、きわめて稀であることがわかってきたということに端を発

する。そして、そこから、それまでの MGCFA の検定があまりにも「厳密 (exact)」/「厳格 (strict)」なものでありすぎたという批判がでてくる。その結果、従来の“full or exact approach”に替わって、“partial or approximate approach”が登場することになったということである。

しかし、このような経緯については、それを単なる「便宜主義」に終わらせてしまわないためにも、もう一度、深く吟味しておくことが必要なのではなかろうか。ここでは、このような点について、以下の3つの問題を提起しておきたい。

(1) 以上に示した“full or exact approach”から“partial or approximate approach”へという方法的移行は、一言でいうならば、「統計的な検定 (statistical test) における諸要件 (requirements) の緩和 (relaxation) の方向」ということになる。しかし、このような「緩和」が、「なぜ許容されるのか」、そして、「どこまで許容されるか」といったことについては、これまでの文献研究からするかぎり、必ずしも明らかではない。これらの問題の解明は、今後に残された重要な問題の1つといわなければならない。

(2) 以上のような問題は、それをさらに突き詰めていくならば、さまざまな人間・社会・文化を対象とする「実証科学」の成立要件そのものの再考ということにまで到るであろう。その際、考慮に入れておくべきことの1つとして、Kish (1987) の方法的立場がある。

Kish によるならば、(純粋な「理論研究」の場合とくらべて——真鍋) 社会調査の研究領域においては、一般に、「望ましいこと (the desirable)」と「できること (the possible)」との間に常にギャップが存在する。社会調査の実践と方法の研究においては、この2つの間の妥協をどこまでも模索し続ける以外に方法はないというのである。

このような Kish の指摘は、とくに多数の国/文化を同時に対象とする国際比較/文化比較研究にとっては、きわめて示唆的であるといえよう。では、具体的に、どのような場合の、どのような妥協が、する価値のあるものといえるのであろうか。ここでも、具体的な事例の提示にもとづく個

別的議論の展開こそが、今後の重要な課題となるであろう。

(3) Ciecuch et al. (2019, p.77) も指摘しているように、「MGCFA の統計的検定をとおして、たとえ高いレベルでの不変性が確認されたとしても、このような『量的な技法 (quantitative technique)』では検出することのできない『意味の違い (differences in meaning)』というものが存在する可能性はある」。

そして、そうであるとするならば、この領域における研究のさらなる発展の方略としては、「量的な技法」の範囲内における技法の開発のみにとどまるのではなく、さらに、そのような「量的な技法」と「質的な技法 (qualitative technique)」とを組み合わせた“mixed-methods approach”の開発と、それにもとづく、そのような「意味の違い」の探究をめざすことこそが重要になってくるのではなかろうか。このような点からして、Meitinger (2017) や Braun and Johnson (2018) は、今後の課題のための reference studies として、きわめて大きな意味をもつものといえよう。

3. 3つ目の研究課題として、「応用研究」における creativity——創造性——の探究というテーマをあげておきたい。ここで、「応用研究」というところに焦点を合わせたのは、そのような領域においてこそ、筆者らの貢献が可能となると考えるからにはほかならない。では、そのような貢献は、いかにして可能となるであろうか。その1つの可能性は、つぎのようなところにある。

この領域における研究の概観からするならば、これまでの「測定¹⁾の等価性/不変性」に関する実証的研究の多くは、欧米の研究者によるものであり、「国際比較/文化比較」の研究事例の多くは、欧米のセッティングにおいてなされたものであった。そこで、今後、欧米の国ぐにとは歴史的・社会的・文化的な背景が大きく異なるとされる日本を射程に入れた比較研究において、「測定¹⁾の等価性/不変性」の問題がよりインテンシヴにアプローチされるとするならば、それは、この研究領域に対する独自の貢献を生み出す可能性をもっているということである。

本稿のⅡのところでも述べた「国際比較／文化比較調査の実践的／方法論的研究」においては、このような「比較調査」で用いられてきた質問諸項目が、「エティック質問 (etic questions)」と「イーミック質問 (emic questions)」に区別されてきた。これら2種類の質問項目は、具体的にいうならば、

- ・エティック質問：さまざまな国／文化を越えて普遍的な適合性 (universal relevance) をもっている質問
- ・イーミック質問：特定の国／文化に限って適合性がある質問、あるいは、あらゆる国／文化において、それぞれ固有の意味 (unique meanings) をもっている質問、ということになる (Johnson, 1998)。

以上のような、社会科学の領域におけるこれまでの諸研究の知的インヴェントリイ (inventory) を手がかりとして、それらを「測定の等価性／不変性」の問題と結びつけることで、今後の研究設計を新しい視座から準備することが可能となるであろう。

文献

- Aleman, Jose and Woods, Dwayne (2016). Value Orientations from the World Value Survey : How Comparable are They Cross-Nationally? *Comparative Political Studies* 49 : 1039-1067.
- Almond, Gabriel A. and Verba, Sidney (1963). *The Civic Culture*. Princeton University Press. (=1974, 石川一雄ほか訳『現代市民の政治文化』勁草書房).
- Alwin, Duane F., Braun, Michael, Harkness, Janet, and Scott, Jacqueline (1994). Measurement in Multi-National Surveys. In Borg, Ingwer and Mohler, Peter Ph. eds. *Trends and Perspectives in Empirical Social Research* (pp.26-39). Walter de Gruyter.
- Ariely, Gal and Davidov, Eldad (2010). Can We Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey. *Social Indicators Research* 104 : 271-286.
- Asparouhov, Tihomir and Muthén, Bengt O. (2014). Multi-Group Factor Analysis Alignment. *Structural Equation Modeling : A Multidisciplinary Journal* 21 (4) 495-508.
- Berger Peter L. and Luckmann, Thomas (1966). *The Social Construction of Reality*. Anchor Books. (=1977, 山口節郎訳『日常世界の構成』新曜社).
- Braun, Michael and Johnson, Timothy P. (2010). An Illustrative Review of Techniques for Detecting Inequivalences. In Harkness, J. et al. eds. *Survey Methods in Multinational, Multiregional, and Multicultural Context*. John Wiley and Sons, Inc.
- Braun, Michael and Johnson, Timothy P. (2018). How Should Immigrants Adapt to Their Country of Residence? : A Mixed Methods Approach to Evaluate the International Applicability of a Question from the German General Social Survey (ALLUBUS). In Davidov, E, et al. eds. *Cross-Cultural Analysis : Methods and Applications*. Routledge.
- Byrne, Barbara M., Shavelson, Richard J., and Muthén, Bengt O. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures : The Issue of Partial Measurement Invariance. *Psychological Bulletin* 105 : 456-466.
- Cieciuch, Jan, Davidov, Eldad, Schmidt, Peter, Algesheimer, René, and Schwartz, Shalom H. (2014). Comparing Results of an Exact vs. an Approximate (Bayesian) Measurement Invariance Test : A Cross-Country Illustration with a Scale to measure 19 Human Values. *Frontiers in Psychology* 5 : 1-10.
- Cieciuch, Jan, Davidov, Eldad, and Schmidt, Peter (2018). Alignment Optimization : Estimation of the Most Trustworthy Means in Cross-Culture Studies even in the Presence of Noninvariance. In Davidov et al. eds. *Cross-Cultural Analysis* [2nd Edition].
- Cieciuch, Jan, Davidov, Eldad, Schmidt, Peter, and Algesheimer, René (2019). How to Obtain Comparable Measures for Cross-National Comparisons. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 71 : 157-186.
- Coromina, Lluís and Davidov, Eldad (2013). Evaluating Measurement Invariance for Social and Political Trust in Western Europe over Four Measurement Time Points (2002- 2008). *ASK Research & Methods* 22 : 35-52.
- Davidov, Eldad (2009). Measurement Equivalence of Nationalism and Constructive Patriotism in the ISSP : 34 Countries in a Comparative Perspective. *Political Analysis* 17 : 64-82.
- Davidov, Eldad, Schmidt, Peter, and Billiet Jaak eds. (2011). *Cross-Cultural Analysis : Methods and Appli-*

- cations. Routledge.
- Davidov, Eldad, Meuleman, Bart, Cieciuch, Jan, Schmidt, Peter, and Billiet, Jaak (2014). Measurement Equivalence in Cross-National Research. *The Annual Review of Sociology* 40 : 3. 1-3. 21.
- Davidov, Eldad, Schmidt, Peter, and Billiet, Jaak, and Meuleman, Bart eds. (2018). *Cross-Cultural Analysis : Methods and Applications* [2nd Edition]. Routledge.
- De Beuckelaer, Alain and Swinnen, Gilbert (2018). Biased Latent Mean Comparisons Due to Measurement Noninvariance : A Simulation Study. In Davidov et al. eds. *Cross-Cultural Analysis* [2nd Edition]. Routledge.
- 福岡伸一 (2021). 「新・ドリトル先生物語」『朝日新聞』(2021年4月5日).
- Haller, Max, Jowell, Roger, and Smith, Tom W. eds. (2009). *The International Social Survey Programme 1984-2009 : Charting the Globe*. Routledge.
- Harkness, Janet A., van de Vijver F. J. R., and Mohler, Peter Ph. (2003). *Cross-Cultural Survey Methods*. Wiley-Interscience.
- Harkness, Janet A., Braun, Michael, Brad, Edwards, Johnson, Timothy P., Lyberg, Lars, Mohler, Peter Ph., Pennell, Beth-Ellen, and Smith, Tom W. (2010). Survey Methods in Multinational, Multiregional, and Multicultural Contexts. John Wiley.
- Hill, Peter C. and Hood Jr., Ralph W. eds. (1999). *Measures of Religiosity*. Religious Education Press.
- Holt, Robert T. and Turner, John E. eds. (1970). *The Methodology of Comparative Research*. Free Press. (= 1976, 内山秀夫, 柴田敏夫, 阪野亘訳『比較政治の方法』勁草書房).
- Hyman, Herbert H. (1972 ; 1987). *Secondary Analysis of Sample Surveys*. Wesleyan University Press.
- 池内了 (2014). 「解説」中谷宇吉郎『寺田寅彦——わが師の追想——』講談社学術文庫.
- Jagodzinski, Wolfgang and Dülmer, Hermann (2019). Unpublished Paper : The International Student Survey on the Basic Human Values.
- Johnson, Timothy P. (1998). Approaches to Equivalence in Cross-National Survey Research. *NACHRICHTEN SPEZIAL* (ZUMA, Mannheim, Germany) 3 : 1-40.
- . (2015). Editorial : Cross-Cultural Issues in Survey Methodology. *Public Opinion Quarterly* 79. S 1 : 211-213.
- Johnson, Timothy P., Pennell, B. E., Stoop, L., and Dorer, B. (2019). *Advances in Comparative Survey Methods*. John Wiley & Sons.
- Jöreskog, Karl G. (1969). A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika* 34 : 183-202.
- . (1971). Simultaneous Factor Analysis in Several Populations. *Psychometrika* 36 : 409-426.
- . (1973). A General Method for Estimating a Linear Structural Equation System. In Goldberger Arthur S. and Duncan, Otis Dudley eds., *Structural Equation Models in the Social Sciences* (pp.85-112). Seminar Press.
- Jöreskog, Karl G. and Sörbom, Dag (1984, 1989, 1996). *LISREL 6, 7, 8 : User's Reference Guide*. Scientific Software Inc.
- 狩野裕 (2002). 「構造方程式モデリングは、因子分析、分散分析、パス解析のすべてにとって代わるのか？」『行動計量学』29. 2 : 138-159.
- 木村通治・真鍋一史・安永幸子・横田賀英子 (2002). 『ファセット理論と解析事例』ナカニシヤ出版.
- Kish, L. (1987). *Statistical Design for Research*. John Wiley.
- Lazarsfeld, Paul F. (1972). *Qualitative Analysis : Historical and Critical Essays*. Allyn and Bacon Inc. (= 1984, 西田春彦, 高坂健次, 奥川櫻豊彦訳『質的分析法——社会学論集——』岩波書店).
- Lerner, Daniel (1958). *The Passing of Traditional Society*. Free Press.
- Lingoes, James C. (1973). *The Guttman-Lingoes Nonmetric Program Series*. Mathesis Press.
- Lipset, Seymour Martin (1968). Revolution and Counter-revolution. Basic Books, Inc. (= 1972, 鈴木広・千石好郎・篠原隆弘訳『革命と反革命』サイマル出版会).
- Lomazzi, Vera (2018). Using Alignment Optimization to Test the Measurement Invariance of Gender Role Attitudes in 59 Countries. *Methods, Data, Analyses : A Journal for Quantitative Methods and Survey Methodology* 12 : 77-103.
- 真鍋一史 (1985). 『世論の研究』慶應通信.
- . (1993). 『社会・世論調査のデータ解析』慶應通信.
- . (1999). 「質問紙法にもとづく国際比較調査の現状と課題」『国際協力論集』(神戸大学大学院国際協力研究科) 7. 2 : 67-81.
- . (2004 a). 「通文化比較調査および国際比較調査の方法論的課題——調査の等価性の問題を中心に——」『法学研究』(慶應義塾大学法学研究会) 77. 1 : 504-538.

- (2004 b). 「通文化比較調査および国際比較調査の方法論的課題——等価性確立のための方法の開発——」『関西学院大学社会学部紀要』96 : 95-110.
- (2016). 「ファセット・アプローチと価値観研究——Lois Guttman とその共同研究者の知的世界の探索——」『関西学院大学社会学部紀要』123 : 9-32.
- (2018). 「〈研究ノート〉Schwartz の『価値観研究』の方法論的な検討」『関西学院大学社会学部紀要』129 : 75-94.
- 真鍋一史, Jagodzinski, Wolfgang, Davidov, Eldad, Dülmer, Hermann, and Hommerich, Carola (2020 a). 「S. Schwartz の概念枠組みにもとづく価値観の国際比較——ドイツと日本における『大学生調査』のデータ分析——」『関西学院大学社会学部紀要』133 : 87-107.
- (2020 b). 「S. Schwartz の概念枠組みにもとづく価値観の国際比較 (II) ——ドイツと日本における『大学生調査』のデータ分析——」『関西学院大学社会学部紀要』135 : 1-20.
- 真鍋一史 (2021). 「ファセット・アプローチとウェルビーイングの研究——Louis Guttman とその共同研究者の足跡——」『関西学院大学社会学部紀要』136 : 1-28.
- McDowell, Ian and Newell, Claire (1987). *Measuring Health : A Guide to Rating Scales and Questionnaires*. Oxford University Press.
- Meitinger, Katharina, Davidov, Eldad, Schmidt, Peter, and Braun, Michael (2020). Measurement Invariance : Testing for It and Explaining Why It is Absent. *Survey Research Methods* 14. 4 : 345-349.
- Meredith, William (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika* 58 : 525-543.
- Merton, Robert K. (1957). *Social Theory and Social Structure*. Free Press. (=1961, 森東吾ほか訳『社会理論と社会構造』みすず書房).
- . (1977). *The Sociology of Science : An Episodic Memoir*. Southern Illinois University Press. (=1983, 成定薫訳『科学社会学の歩み——エピソードで綴る回想録——』サイエンス社).
- Millsap Roger, E. and Meredith, William (2007). Factorial Invariance : Historical Perspective and new problems. In Cudeck, R., MacCallum RC ed. *Factor Analysis at 100 : Historical Developments and Future Directions*. Lawrence Erlbaum.
- Millsap, Roger E. (2011). *Statistical Approach to Measurement Invariance*. Taylor & Francis.
- Muthén, Bengt O. and Asparouhov, Tihomir (2012). Bayesian Structural Equation Modeling : A More Flexible Representation of Substantive Theory. *Psychological Methods* 17 : 313-335.
- (2017). Recent Methods for the Study of Measurement Invariance with Many Groups : Alignment and Random Effects. *Sociological Methods & Research*. Advanced Online Publication.
- Porrat, Rama, comp. (1974). *A Laboratory Manual for the Guttman-Lingoes Nonmetric Computer Programs*. The Israel Institute of Applied Social Research.
- Robinson, John P., Shaver, Phillip R., and Wrightsman, Lawrence S., eds. (1991). *Measures of Personality and Social Psychological Attitudes*. Academic Press.
- Rokkan, Stein (1968). *Comparative Research Across Cultures and Nations*. Mouton.
- Seddig, Daniel and Leitgöb, Heinz (2018). Exact and Bayesian Approximate Measurement Invariance. In Davidov et al. eds. *Cross-Cultural Analysis* [2nd Edition].
- Smelser, Neil J. (1988). *Comparative Methods in the Social Sciences*. Prentice-Hall, Inc. (=1996, 山中弘訳『社会科学における比較の方法』玉川大学出版会).
- Sokolov, Boris (2018). The Index of Emancipative Values : Measurement Model Misspecifications. *American Political Science Review* 112 : 395-408.
- Sörbom, Dag (1974). A General Method for Studying Differences in Factor Means and Factor Structure Between Groups. *British Journal of Mathematical and Statistical Psychology* 27 : 229-239.
- Steenkamp, Jan-Benedict E. M. and Baumgartner, Hans (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research* 25 : 78-90.
- Steinmetz, Holger (2018). Estimation and Comparison of Latent Means Across Cultures. In Davidov et al. eds. *Cross-Cultural Analysis* [2nd Edition] Routledge.
- 豊田秀樹・前田忠彦・柳井晴夫 (1992). 『原因をさぐる統計学——共分散構造分析入門——』(ブルーバックス) 講談社.
- 梅棹忠夫 (1969). 『知的生産の技術』岩波新書.
- Vandenberg, Robert J. and Lance, Charles E. (2000). A

- Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods* 3. 1: 4-70.
- van de Schoot, Rens, Kluytmans, Anouck, Tummers, Lars, Lugtig, Peter, Hox, Joop, and Muthén, Bengt O. (2013). Facing off with Scylla and Charybdis: A Comparison of Scalar, Partial and the Novel Possibility of Approximate Measurement Invariance. *Frontiers in Psychology* 4(770): 1-15.
- van de Schoot, Rens, Schmidt, Peter, de Beucklaer, Alain, Lek, Kimberley, and Zondervan-Zwijenburg, Marielle (2015). Editorial: Measurement Invariance. *Frontiers in Psychology* 6: 1-4.
- Zercher, Florian, Schmidt, Peter, Cieciuch, Jan, and Davidov, Eldad (2015). The Comparability of the Universalism Value Over Time and Across Countries in the European Social Survey: Exact vs. Approximate Measurement Invariance. *Frontiers in Psychology* 6: 1-11.

VI. 付論：MGCFA の理論と技法の日本における受容の諸相

以下付論では、MGCFA の理論と技法の日本における受容の状況を、(1) MGCFA および関連する用語の訳語、(2) 日本における MGCFA の紹介、(3) 日本における応用研究の概況、の側面から概観する。

1. MGCFA および関連する用語の訳語

すでに本論で述べてきたとおり、本稿において筆者らが「多集団確証的因子分析」という訳語を当ててきた MGCFA という技法は、多集団において「測定の等価性／不変性」が成り立つかどうかに関心を置くものである。同様の関心に基づく技法は、すでに 1980 年代の終わりから日本でも紹介がされはじめ、LISREL、SAS、AMOS、Mplus、R といった統計解析ソフトウェアの普及とともに広まってきた。しかしじつは、この MGCFA という呼称は、日本では一般的に用いられているものではない。

1989 年の『関西大学社会学部紀要』において、清水和秋は「研究対象となった標本間において不

変な因子構造と、測定の等価性 (equality) の性質を明らかにするための出発点」となった研究として、Jöreskog (1971) と Sörbom (1974) をあげている。清水 (1989) は、前者の論文タイトルにある simultaneous factor analysis という用語に「複数標本の同時分析」という訳語を当て、「複数の標本に同一の観測変数を適用して得られたデータにおいて、これらの複数の標本から同一の因子構造が得られるかどうかを検証する理論」として紹介している (清水, 1994)。また、Jöreskog-Sörbom の multi-sample analysis という表現をとる場合には、「多群標本の同時分析」あるいは「多群同時分析」という訳語が採用されている。そして、その具体的な手続き——より「制約の少ないモデル」から「制約の多いモデル」へと段階的に分析を進め、各モデルのあてはまりを相互に比較するという仕方、どの程度の等価性／測定の不変性のレベルであるかを検討するという方法——については、清水 (1989)、豊田 (1992, 1998, 2007, 2014)、柳井 (1994) において繰り返し紹介されることになるが、その際の呼称にも若干のバリエーションがあり、「多母集団の同時分析」(豊田, 1992, 1998)、「多母集団についての同時分析」(柳井, 1994) などがあげられるが、中でも「多母集団同時分析」(豊田, 2014) が最もよく用いられている。

ただし、一般に文献で広く用いられている多母集団同時分析という用語は、MGCFA とイコールではない。多母集団同時分析という用語が用いられている文献において、対応する英語が本文中に括弧付きで明示されたり、英文要旨に載ることは稀であるが、その場合、multi-group analysis, simultaneous analysis, multi-group simultaneous analysis, multi-group SEM と書かれているのがほとんどである。多母集団同時分析とはその名のとおり、複数の集団をまたいで分析をする技法を指した広い意味を持つ用語であり、測定の普遍性／等価性を問題とする MGCFA もその中に含まれるが、両者は必ずしも一致するものではないことに留意する必要がある。

使用する分析手法が MGCFA であることを明示する場合には、「確認的因子分析の多母集団同時分析モデル」(松本, 2015)、「多母集団同時分

析による確証的因子分析」(田中・田山、2013)、「複数母集団確認的因子分析」(多田ほか、2017)などといった表現がとられることもあるが、今のところ MGCFA という言い回しに対応する定訳というべきものはない。

また、MGCFA によって検討される不変性のレベルが configural invariance, metric invariance, scalar invariance に大別されることは3章で述べたとおりだが、それぞれの訳語にもわずかながらバリエーションがある。たとえば configural invariance については、「布置不変」(清水、1989, 1994)、「配置不変」(豊田、2007, 2014)、「構成不変」(太郎丸、2013)といった訳語が提案されてきた。metric invariance には、「厳格な測定不変性」(清水、前掲)、「弱測定不変」(豊田、前掲)、「測定不変性」(田崎・申、2019)、「尺度不変性」(池田ほか、2017)などの訳がある。scalar invariance については、単に「スカラー不変性」(田崎・申、前掲；池田ほか、前掲)と呼ばれる場合もあるが、豊田(2014)は「強測定不変」と呼び表しており、さらにより制約の強い測定の不変性／等価性を検討するモデルとして「厳格な測定不変」「全母数が等しい」モデルをあげている。

2. 日本における MGCFA の紹介

日本において MGCFA は、主に統計解析ソフトを用いた分析方法に関するテキスト——とくに共分散構造分析(あるいは構造方程式モデリング)の方法の文脈において——の中で紹介されてきた。たとえば、上述の清水(1989)も、LISREL を用いて測定の不変性／等価性を検討するためのパラメータ制約を行う方法の解説を行ったものである。また清水(1994)も、LISREL 第7版において、多標本同時分析を前提として LISREL の基本モデルの拡張が行われたことを紹介したものであった。

その他、MGCFA の手法が紹介されている代表的なテキストとして、LISREL (豊田、1992)¹⁾、AMOS、(豊田、2007)、R (豊田、2014)、R と M-plus (小杉・清水、2014)、AMOS、EQS、LIS-

REL、SAS (狩野、1997；狩野・三浦、2002)などの解説書があげられる。これは MGCFA に限った話ではなく、その理由は豊田(1998)の巻頭言にある「構造方程式モデルの実行には SW (ソフトウェア)の使用が不可欠である。このため従来、構造方程式のモデルの教科書は特定の SW を使用することを前提に執筆されることが多かった」という記述にまとめられる。

3. 日本における応用研究の概況

最後に、日本の社会科学の領域における MGCFA の応用研究の展開状況について概観したい。MGCFA を用いた研究は、パーソナリティ研究やカウンセリング研究をはじめとした心理学的研究に集中している。とくに、尺度開発(既存尺度の検討を含む)研究の文脈においては、①まず標本を一括して探索的因子分析を行い、②その結果をもとに確証的因子分析を行った上で、③「測定の妥当性・信頼性の検討」の一環として、任意の指標を用いて標本を複数に分割して MGCFA を実施するという手続きがよく見られる。

この類の応用研究は非常に多く枚挙に暇がないが、近年発表されたもので言えば、たとえば、大学生を対象に調査を実施し、性別間での「活発な精神的活動に対する意志力の暗黙理論尺度 (Implicit Theory of Willpower for Strenuous Mental Activities Scale: ITW-M) 日本語版」の測定の不変性を検討した櫻井・渡辺(2019)、40歳から85歳の男女を対象とした調査データで、性別・年代間での「ワーク・ファミリー・バランス尺度」の測定の不変性を検討した富田ら(2019)、3か所の家庭裁判所の係属した少年たちに調査を実施し、性別、年齢層、地域、前歴の有無、保護処分の有無ごとに、「機能的攻撃性尺度 (Functional Aggression Scale)」の測定の不変性を検討した嶋田・遊間(2018)などがあげられる。これらの研究では、測定モデルの成立が特定の属性に依存していないかを確認することが、MGCFA を行う上での主要な関心とされており、本稿で取りあげてきたような異なる国／文化間での「比較」に関心

1) 豊田(1992)は、『SASによる共分散構造分析』という書名の通り、SASを用いた統計解析のテキストであるが、当時のSASはMGCFAを実施する機能が備わっておらず、MGCFAについてはLISRELでの分析の手続きが掲載されている。

を寄せたものではない。

また、一風変わった MGCFA の用い方をして
いる例もある。松川・立木 (2010) は「多母集団
同時分析とはデータの母集団が異なることを認め
たうえで、集団間に回答傾向の差がないことを証
明するための分析である」と説明し、異なる時点
で収集されたデータ間での測定の不変性を確認す
ることで、データセットをマージして分析するこ
との可否について検討を行っている。

尺度開発研究への MGCFA の応用例が豊富な
のに対して、国際比較研究への応用例もわずか
ではあるものの認められる。その代表的なもの
として、東アジア価値観国際比較調査のデータ
分析を行い、「対組織信頼」について東アジアの
8つの国・地域での MGCFA を行った松本
(2006) があげられる。松本によると「市民
的信頼」と「体制的信頼」の2因子からなる
「対組織信頼」の検証的因子分析のモデルは、
8か国で MGCFA を行った場合は configural
invariance が認められるのみであったの
に対して、中国本土の4地域とそれ以外の4
つの国・地域の2グループに分けて再度分
析を試みたところ、metric invariance が
確認される結果が得られるようになったとい
う。また、さらに日本を独立した3つ目のグ
ループとして分けることで、さらに適合度指
標の向上が見られた。metric invariance が
認められるようなグループ分けを特定する
ということは、潜在変数の影響や平均値が
比較可能とされるグループの探求と言
い換えることができる。また松本 (2016)
では、MGCFA の対象に複数のエスニック
グループが混在するシンガポールを加えた
場合、シンガポールのデータを中国系住
民に限定することで、同様に適合度指標
の向上が見られることが報告されている。

その他、国際比較への MGCFA の応用研
究の主な例として、以下に5本の論文を
あげておきたい。田崎 (2007) では、
日本人とヨーロッパ系アメリカ人を
対象に、「文化的自己観 (self-construal
scales)」の測定の不変性の検討が
行われている。また、田辺 (2011) は、
2003年のISSP ナショナル・アイ
デンティティ・モジュール調査を用
いて、「ネーションの成員条件」「文
化的プライド」「政治的プライド」「
排外性」の4因子からなるナ

ショナル・アイデンティティについて、
日本と韓国を対象とした MGCFA を
実施している。松木ら (2020) は、
日本と韓国でウェブ調査を実施し、
「連帯・積極性」「自己決定」「愛
着」「他社への依頼」という4因子
からなるコミュニティ意識について
分析を行っている。多田ら (2017)
と池田ら (2017) による2編から
なる論文では、世界価値観調査の
データ分析が行われており、アメ
リカとオーストラリアでのエコロ
ジー行動、東西ドイツ2時点での
「開放的価値観」、12か国での
「物質主義-脱物質主義」につい
て、それぞれ測定の不変性の検討
が行われている。ここにあげた
論文ではいずれも、MGCFA を
用いた測定の不変性/等価性の
確認が因子平均比較の条件とし
てあげられており、そこで用いら
れている MGCFA は、それ自体が
各国の因子構造の比較の手段で
あるという側面と、そこで言及さ
れている尺度が「因子平均の比較
に耐えるか」(田崎, 2007)とい
う、いわば比較の準備作業とい
う側面の両方を有するものとい
える。

しかしながら、すでに述べたよ
うに、心理学的尺度開発研究への
MGCFA の応用は豊富であるの
に対して、国際比較研究への応
用研究はきわめて稀である。近
年では、ISSP や世界価値観調
査をはじめとした、研究者が利
用可能な大規模な国際比較調
査のデータセットの蓄積が広く
知られるようになってきた。今
後、日本や東アジアを国際比
較研究の視野に入れていくな
らば、その方法論的前提として
の MGCFA による国際比較可
能性の検討も一層進められて
いかなければならない。

文献

池田泰成・廖舒忻・石井隆太・馮昕・庄映琮・多田
 伶・張しん妍・濱岡豊 (2017). 「グローバル・
 マーケティングと消費者の価値観：世界価値観調
 査の分析から (前篇)」『三田商学研究』60(3)：87
 -118.
 Jöreskog, K. G. (1971), Simultaneous factor analysis in
 several populations, *Psychometrika*, 36：409-426.
 狩野裕 (1997). 『グラフィカル多変量解析——目で見
 る共分散構造分析——』現代数学社 (狩野裕・三
 浦麻子 (2002). 増補版).
 小杉考司・清水裕士編 (2014), 『M-plus と R による構
 造方程式モデリング入門』北大路書房.

- 松川杏寧・立木茂雄 (2010). 「ソーシャルキャピタルと地域の防犯性因果モデルについての検討——多母集団同時分析を用いて——」『地域安全学会梗概集』27: 97-100.
- 松本祐馬・向井智哉・金信遇・木村真利子・近藤文哉 (2020). 「日韓におけるコミュニティ意識の比較」『パーソナリティ研究』29(2): 71-74.
- 松本渉 (2006). 「東アジアにおける組織に対する信頼感——国際比較のための信頼感の分析——」『行動計量学』33(1): 25-40.
- (2016). 「組織に対する信頼の国際比較可能性——東アジア価値観国際比較調査におけるシンガポール調査に注目して——」『情報研究: 関西大学総合情報学部紀要』42: 57-79.
- 櫻井良祐・渡辺匠 (2019). 「活発な精神的活動に対する意志力の暗黙理論尺度 (Implicit Theory of Willpower for Strenuous Mental Activities Scale: ITW-M) 日本語版の作成」『パーソナリティ研究』27(3): 259-262.
- 嶋田美和・遊間義一 (2018). 「家庭裁判所係属少年における機能的攻撃性尺度の妥当性の再検討」『心理学研究』89(1): 71-81.
- 清水和秋 (1989). 「検証的因子分析, LISREL そして RAM の概要」『関西大学社会学部紀要』20(1): 61-86.
- (1994). 「Jöreskog と Sörbom によるコンピュータ・プログラムと構造式モデル」『関西大学社会学部紀要』25(3): 1-41.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups, *British Journal of Mathematical and Statistical Psychology*, 27: 229-239.
- 多田伶・石井隆太・張しん妍・馮昕・池田泰成・庄映琮・廖舒忻・濱岡豊 (2017). 「グローバル・マーケティングと消費者の価値観: 世界価値観調査の分析から (前篇)」『三田商学研究』60(2): 61-97.
- 田辺俊介 (2011). 「日韓のナショナル・アイデンティティの概念構造の不変性と異質性の検討——ISSP 2003 データを用いた多母集団共分散構造分析——」『社会学評論』62(3): 284-300.
- 田中勝則・田山淳 (2013). 「確証的因子分析による日本語版 Body Image Concern Inventory の因子構造の検討」『カウンセリング研究』46(1): 11-17.
- 田崎勝也 (2007). 「文化的自己観は本当に「文化」を測っているのか: 平均構造・多母集団同時分析を用いた特異項目機能の検証」『行動計量学』34(1): 79-89.
- 太郎丸博 (2013). 「Steenkamp and Baumgartner, 1998, 国際比較研究における測定の不変性を評価する」, *Theoretical Sociology* (2021年7月7日閲覧, URL: <http://sociology.jugem.jp/?eid=740>).
- 富田真紀子・西田裕紀子・丹下智香子・大塚礼・安藤富士子・下方浩史 (2019). 「中高年者に適用可能なワーク・ファミリー・バランス尺度の構成」『心理学研究』89(6): 591-601.
- 豊田秀樹 (1992). 『SAS による共分散構造分析』東京大学出版会.
- (1998). 『共分散構造分析 [入門編]——構造方程式モデリング——』朝倉書店.
- (2007). 『共分散構造分析 [Amos 編]——構造方程式モデリング——』東京図書.
- (2014). 『共分散構造分析 [R 編]——構造方程式モデリング——』東京図書.
- 柳井晴夫 (1994). 『多変量データ解析法——理論と応用——』朝倉書店.

Measurement Equivalence/Invariance in Cross-National/Cross-Cultural Research: A Review of the Literature Focusing on MGCFA

ABSTRACT

The purpose of this paper is to examine the development of studies on measurement equivalence/invariance in cross-national/cross-cultural research, specifically the analytical approach called multiple group confirmatory factor analysis (MGCFA). This approach has been the most widely used to test for cross-national/cross-cultural equivalence/ invariance in various empirical researches.

This paper is composed of five parts:

The first part attempts to describe the history of the studies on the above-mentioned topic. According to the traditional method of literature survey, first we collected the relevant literature on the topic. Then we classified the various types of studies. The result is shown in a table.

In the second part, based on the collected literature, we attempt to summarize the basic ideas, main features, and practical procedures of the statistical approach called MGCFA.

The above is called the traditional/exact approach. After that, the new/ approximate approach has been proposed. This approach includes (1) Bayesian approximate method and (2) alignment optimization method. In the third part of the paper, we attempt to summarize the basic ideas, main features, and practical procedures of the new/ approximate approach.

The fourth part of the paper is an appendix covering the history of acceptance of the ideas, techniques and procedures on measurement equivalence/invariance in the Japanese academic community.

Finally, in the last part, we suggest a future direction for development in this field of study, and propose some empirical problems for future research.

Key Words: measurement equivalence/invariance, cross-national/cross-cultural research, multiple group confirmatory factor analysis (MGCFA)