レスポンス・スタイルは文化的等価性を毀損するのか¹

田崎 勝也 申 知元 青山学院大学

Response styles as a threat to cultural equivalence

Katsuya Tasaki and Jiwon Shin (Aoyama Gakuin University)

This study aimed to identify how a response style (RS) affects differential item functioning (DIF) detections in cross-cultural comparison studies. RS is an individual's systematic response tendency to survey questions. Following the representative indicators response style means and covariance structure (RIRSMACS) model introduced by Weijters et al.(2008), Tasaki and Shin (2017) developed an RS measurement model and quantified three RS types (acquiescence, extreme, and mid-point) in three nations (Japan, Korea, U.S.). Using the same data set, this study adapts their RS model to control for RS-related nuisance variances and to see if different DIF items emerge with or without the RS model. As a result of DIF analysis through the MGMACS approach, three nonuniform and five uniform DIF items were flagged on a 10-item psychological scale. By applying the RS model to DIF analysis, all the nonuniform DIF items and three uniform DIF items turned out to be non-DIF items. These results imply that RS-related nuisance variances may be a cause of DIF, and RIRSMACS can be used to control for RS.

Key words: response style, differential item functioning (DIF), cross-cultural studies, representative indicators response style means and covariance structure (RIRSMACS).

The Japanese Journal of Psychology

2019, Vol. 90, No. 2, pp. 194–200

J-STAGE Advanced published date: February 28, 2019, <https://doi.org/10.4992/jjpsy.90.18302>

文化的背景の異なる人々の心的特性や行動形態などの同異を検討する国際・文化比較調査では、文化的等価性 (cultural equivalence) が調査の妥当性を決める争点になることが多い (Steenkamp & Baumgartner, 1998)。文化的等価性を毀損する要因はさまざまで、たとえば、対象となる概念がそもそも比較集団間に存在するかといった概念的な要因、サンプルの代表性や調査環境の不変性を問題にする方法論的な要因、質問項目の翻訳ミスなどによって生じる設問・项目的な要因などがある (van de Vijver & Poortinga, 1997)。

文化的等価性を統計的に検証する手段として DIF 分析が広く用いられている。DIF は “differential item

functioning” の略で「特異項目機能」と訳され、項目や設問に対する反応のグループ差を示す (藤本, 1999)。特異項目機能の検証においては、識別力 (discrimination) や困難度 (difficulty) といった2種類の項目機能について測定の等価性を検討する²。本研究で用いる解析法を含め、現今の DIF 分析は “matching principle” (Angoff, 1993) と呼ばれる基本原理に立ち、所属グループは異なるが同程度の特性をもつ個人が項目にどういった反応を示すかを検討する。つまり DIF 分析の眼目は、測定対象となる心理特性を集団間で統制した時に現れる項目機能の差となる。項目機能は集団により異なることがあり、DIF 項目がすべてバイアス項目になるわけではない (藤本, 1999)。しかし、測定対象が統制され検出される DIF は、次元性の一致をみた同一の特性値尺度において、集団間で推定値が異なることになるため、基本的にバイアス項目とな

Correspondence concerning this article should be sent to: Katsuya Tasaki, Department of International Communication, Aoyama Gakuin University, Shibuya, Shibuya-ku, Tokyo 150-8366, Japan. (E-mail: tasaki@siweb.aoyama.ac.jp)

¹ 本論文の研究結果の一部は、多文化関係学会第 11 回年次大会および日本社会心理学会第 55 回大会で発表された。

² 識別力や困難度は、因子分析の枠組みでは、パス係数や切片に相当する (Ferrando, 1996)。

る。

国際・文化比較調査において、尺度によって測定した代表値（平均値）が比較可能性を持つためには、項目機能が集団間で一致していなければならない。項目が同一の機能を有していなければ、仮に得点差が観測されたとしても、それが当該概念の差異を示すのか、それとも測定におけるバイアスを反映しているのか特定できない（Steenkamp & Baumgartner, 1998）ためである。

文化的等価性の棄損要因については前述の通りだが、特定の調査環境下でどの要因が顕著化して DIF を導くのか、不明なこともいまだに多く、さらなる検証が必要である。本研究では、DIF の原因の 1 つとして議論されるレスポンス・スタイル（response style: 以下、RS とする）に焦点を当て、国際調査において参加者の文化的背景によって異なる RS が DIF に関与するのかを検討する。

レスポンス・スタイルとは

RS は設問や質問項目に回答する際にみられる反応バイアスの一種で、測定内容に関係なく、特定の選択肢を好意的に選択する回答行動を意味する。国際・文化比較調査で問題になる RS には、盲目的に項目に同意する黙従反応（acquiescence response style: 以下、ARS とする）、極端な選択肢を選ぶ極端反応（extreme response style: 以下、ERS とする）、中間の段階評価を選択する中間反応（mid-point response style: 以下、MRS とする）の 3 つの形態がある（Yang, Harkness, Chin, & Villar, 2010）。

RS を生じさせる要因としては（a）調査協力者の属性やパーソナリティ特性、（b）測定内容と認知的負担の程度、（c）リカート法や複数回答形式など回答方法の特徴、（d）留置き法や Web 調査に代表される調査モードによる特性、（e）測定時間などデータ収集の環境的要因、（f）質問項目のワーディングや標識の有無などが挙げられている（Vaerenberg & Thomas, 2013）。

尺度によって構成概念を測定する心理学研究において、RS は尺度得点の解釈を妨げるノイズとなるばかりか、文化比較に際しては DIF に関与する要因にもなり得る。測定に伴う誤差には、偶発的に発生する非系統的誤差と事象や時間を超えて常に現れる系統的誤差とがあるが、RS は後者の誤差を生む要因となる（Podsakoff, MacKenzie, Lee, & Podsakoff, 2003）。また、系統的誤差は構成概念外の分散（construct irrelevant variance）とも呼ばれ（平井, 2006）、DIF の生じにはこうした局外分散/因子の関与が疑われている（Reise, Smith, & Furr, 2001）。つまり、RS により生じた分散が測定対象とは無関係な因子を形成し、こうした局外因子により DIF が生起すると考えられる。

回答行動の個人差を説明／予測する変数として、国

籍や民族的背景が挙げられている。たとえば、日本人をはじめ東アジア人は、5 件法のリカート尺度で 3 の選択肢を選ぶといった高 MRS 的 RS を示す（Harzing, 2006）。「文化」によって規定される回答行動は、各国から算出した得点を基に知見を得る国際調査において、比較可能性を阻み、調査の妥当性を毀損する要因になる（Kankaraš & Moors, 2011）。

レスポンス・スタイルの同定

ところで RS 研究では、まず RS を測定しなければならないが、設問や質問項目への反応のうち、どこまでが真の回答で、どこまでが反応バイアスによる回答なのかを見極めづらいといった問題が指摘されてきた（Yang et al., 2010）。こうした問題に対して、近年共分散構造分析（structural equation modeling: SEM）が採用されるようになった。いくつかのタイプはあるものの、SEM を用いた方法では、RS を一種の性格特性と捉え、項目反応に潜む反応傾向を因子として抽出する（Billiet & McClendon, 2000）。

現時点で最も精緻とされる Weijters, Schillewaert, & Geuens (2008) の方法（representative indicators response style means and covariance structure: 以下、RIRSMACS とする）では、ランダムに配置した測定内容の異なる項目群を使用して RS 因子を抽出する。たとえば田崎・申 (2017) は、RIRSMACS を用いて、日本・米国・韓国の ARS, ERS, MRS を比較検討している。この調査では測定内容の異なる 3 尺度から得た 51 項目を無作為に 3 グループ・17 項目に分け、反応バイアスの定義に従って 5 件法リカート尺度の評価基準の選択数を集計した。Figure 1 の点線下部はこうして得た 3 つの RS 変数を基に 3 種の RS 因子を測定するモデルを描いている。RIRSMACS はランダム化した測定内容の異なる項目反応を基に RS を因子として抽出することで、測定内容との交絡を避けた RS の測定が可能になる。

制御モデルとしての RIRSMACS と本研究の目的

その他 SEM をベースとした測定法には、反応バイアスの制御モデルとしても優れた特徴がある。たとえば、4 項目で任意の因子（content factor）を測定する場合には（Figure 1 点線上部）、RS 因子から content factor を測る各観測変数（item 1 から item 4）にもパスを引く。こうすることで、項目反応に潜む反応バイアスの影響を考慮した content factor の測定が可能になる。

本研究では、RIRSMACS によって、RS 由来の局外分散/因子を制御すると、通常の因子モデルの測定にどのような変化がみられるかを検討する。RS は国際・文化比較調査では DIF の原因となり得ることから、特に RS と DIF の関係に着目し、田崎・申 (2017) に

よって見出された RS の測定モデルを適用した場合と
しなかった場合で、DIF の検証結果が変わるのかを検
見する。

方 法

分析データ

本研究では「レスポンス・スタイルの 3 文化調査」³
のデータの一部を利用した。この調査データは日本・
米国・韓国の合計 813 名の大学生から収集したもの
である。各国の参加人数、年齢の平均および標準偏差は、
日本は 328 名 ($M = 19.49$, $SD = 1.29$), 米国は 164 名 ($M = 21.54$, $SD = 4.95$), 韓国は 321 名 ($M = 19.73$, $SD = 1.79$) で、
そのうち男性の参加者数および割合は、日本は 112 名 (34.1%), 米国は 56 名 (34.1%), 韓国は 97 名 (30.2%) だった。

分析には Mplus (Muthén & Muthén, 2012) を用いた。

レスポンス・スタイルの測定モデル

田崎・申 (2017) は同データに含まれる質問項目を
利用して RIRSMACS モデルを作成しているが、本研
究ではこのモデルを流用し RS の同定に用いた。ラン
ダムに配置された 17 項目に関して、5 件法のリカー
ト尺度の各評価基準 (1 = 「同意しない」から 5 = 「同
意する」) の選択回数を基に RS 変数を得た。
Baumgartner & Steenkamp (2001) の操作的定義に準じ

て、ARS は 4 もしくは 5 の選択数⁴、ERS は 1 もしく
は 5 の選択数、MRS は 3 の選択数をグループ毎に集
計し、項目数 17 で除して RS 変数を算出した。

RIRSMACS には RS モデルの妥当性と信頼性を事前
に確認できるという利点がある (Vaerenberg & Thomas,
2013)。田崎・申 (2017) は、多母集団分析を通して、3 カ
国間の文化的等価性を確認し、当モデルは十分な妥当
性と信頼性を有するとしている。

かなしい映画尺度

DIF の検証には、15 項目から構成されるかなしい
映画尺度 (The Sad-Film Scale (SFS) : Oliver, 1993) を
用いた。5 件法のリカート尺度 (1 = 「同意しない」か
ら 5 = 「同意する」) によって回答を得た当該尺度には
「かなしい映画は私に大きな感動を与えてくれるから
好きだ」, 「かなしい映画は大抵とても退屈で楽しめな
い」, 「私はかなしい映画のキャラクターに感情移入す
ることを楽しむ」などの項目が含まれ、娯楽としての
かなしい映画の好恶を尋ねており、いずれの文化の参
加者に対しても、特定の信条や理念、価値観が反映さ
れにくい内容である。クロンバックの信頼性係数は .94
— .95 と報告されている。

実際の DIF 分析は 15 項目のうち 10 項目を対象に

⁴ Baumgartner & Steenkamp (2001) は ARS 変数のみ以下のよ
うに重み付けをして算出している。

ARS 変数 = $[f(4) \times 1 + f(5) \times 2] / k$ ($f(o)$ = 選択肢の選択数, k =
項目数)。

Figure 1. RIRSMACS モデル (Weijters et al., 2008)。

³ CHS#19958-"Assessing Response Styles in Three Cultures"

行った。今回はRSに由来する局外分散／因子がDIFの特定に与える影響を検討することが目的であるため、RS以外のDIFに関与する要因は可能な限り排除して分析に臨んだ。まずバックトラスレーションにより翻訳ミスが示された「私は映画に娯楽とか刺激を求めているからかなしい映画は楽しめない」(“I don't like sad movies, because they are too light and oftentimes fun”)を削除した。韓国語訳では質問の主旨が「かなしい映画は刺激的なので好きでない」となり、日本語・英語の項目とはかなしい映画を嫌う理由が逆転していたことが削除理由である。

次に、探索的因子分析を実行し、因子構造を確認した。文化的等価性を棄損する要因は多岐にわたるため、DIF分析では対象となる因子の1次元性が分析の前提になる(Ainsworth, 2007)。因子分析の結果、かなしい映画尺度に想定された1因子構造が認められたが、因子負荷量.40以下の項目が2項目示された。より強固な1因子構造を確保するため、これら2項目を分析から外した(「私はかなしい映画から悲しさより愚かさを感じる」, 「かなしい映画は大抵とても退屈で楽しめない」)。

さらに、Mplusで修正指数(modification index)として出力されるLM検定(Lagrange Multiplier)を参考に、2項目(「私は映画を見ることで悲しむよりは幸せになりたいからかなしい映画を好まない」, 「かなしい映画はあまりにもドラマチックで楽しめない」)を削除した。RSモデルを含めた分析で、上記の2項目は理論的には結びつきが考えられない複数の変数との関係性が示され、パスの付加により縮減される χ^2 値も50を超えていたためである。

結 果

非修正モデルのDIF分析

RSモデルの適用を試みる前に、まずかなしい映画尺度の10項目に対して、平均構造・多母集団同時分析(multigroup mean and covariance structure: 以下、MGMACSとする)を用いてDIF分析を行った(非修正モデル)。全データによる探索的因子分析で最も因子負荷量の高かった項目2のパス係数を1に固定し、切片には等値制約を課した。また、韓国の因子平均をゼロに固定し、日米の因子平均は自由推定した。

本研究では、田崎・申(2017)に従い、2種の指標を参考にDIF項目を認定した。MGMACSによるDIF分析は、ベースラインモデルと検定モデルの間に生まれる χ^2 値の乖離度を項目ごとに検討することから、検定の繰り返しによる第1種の過誤が問題視されている(Millsap, 2011)。尤度比検定の代案として、適合度指標を用いたDIFの検出法を模索するMeade, Johnson, & Bradley(2008)は、シミュレーションの結果に基づ

きCFI値の乖離度.002以上を提案している。一方、Cheung & Rensvold(1999)は.02以上を示すなど、研究者によって基準にばらつきがある。本研究では、尤度比検定の結果とCFI値の離脱幅の双方を参照し、 χ^2 値が1%水準で有意であり、同時にCFI値に.002以上の差がある場合に、モデル間の乖離を認めDIFと判断した。

DIFの検証に際して、まずベースラインとなる配置不変モデルから確認した。モデル確定の目的以外の制約は持たない配置不変モデルは、データと十分な当てはまりを示した(Table 1)。上記の χ^2 値の検定結果とCFI値の乖離度を参照し検討したところ、すべてのパス係数に等値制約を課した測定不変モデルおよびすべての切片に等値制約を課したスカラー不変モデルは、いずれも配置不変モデルから有意に離れており、識別力機能が異なる不均一DIF項目や困難度機能が異なる均一DIF項目が存在することが示された。そこで項目ごとに順次パス係数や切片に等値制約を課し、DIF項目の特定を試みた。分析の結果、項目5, 7, 8のパス係数と項目3, 4, 5, 7, 8の切片に日米韓で違いがみられ、一部重複しているが、3項目で不均一DIF、5項目で均一DIFが検出された。

修正モデルのDIF分析

次に、RSモデルを適用した「修正モデル」において、同じ手順でDIF分析を実行し、非修正モデルで認められたDIF項目が修正モデルでもDIFとなるかを確認した。RSの影響はそれぞれの文化内で一定と考えられることから、各RS因子から伸びるパスには文化ごとに等値制約を課した。パス係数の標準化推定値は、ARSは日本が.434、米国が.289、韓国が.704、ERSは日本が-.578、米国が-.104、韓国が-.173、MRSは日本が.268、米国が.232、韓国が.626となった。ERS因子からのパスで韓国のみ有意傾向($p = .056$)となったものの、その他のパスはすべて1%水準で有意だった。

まず、修正モデルのデータとの適合度を確認した。配置不変モデルの適合度指標は、かなしい映画尺度の10項目のみをモデル化した非修正モデルと比べ、いずれの指標も改善し(χ^2/df 値: 3.74 → 1.79; CFI値: .900 → .959; RMSEA値: .101 → .054)、RSに関連する局外分散がモデル適合を妨げていた可能性が示された。続いてパス係数および切片に等値制約を課し、DIF項目の有無を検討したところ、項目5と項目7の切片を除いて、DIFは検出されなかった。非修正モデルでDIFと認められた項目5, 7, 8のパス係数および項目3, 4, 8の切片は、修正モデルでは非DIFとなった。DIFから非DIFとなった項目機能は、RSの文化差が関与していた可能性が高い。

Table 1
かなしい映画尺度の DIF の検証

		χ^2	df	CFI	RMSEA	ΔCFI	$\Delta\chi^2$	Δdf	Δp
非修正モデル	配置不変	392.380	105	.900	.101				
	測定不変	455.427	123	.884	.100	.016	63.047	18	.000
	項目 1 パス	393.111	107	.900	.099	.000	0.731	2	.694
	項目 3 パス	394.974	107	.900	.100	.000	2.594	2	.273
	項目 4 パス	395.218	107	.900	.100	.000	2.838	2	.242
	項目 5 パス	407.362	107	.895	.102	.005	14.982	2	.001
	項目 6 パス	395.111	107	.900	.100	.000	2.731	2	.255
	項目 7 パス	403.564	107	.897	.101	.003	11.184	2	.004
	項目 8 パス	403.613	107	.897	.101	.003	11.233	2	.004
	項目 9 パス	394.072	107	.900	.100	.000	1.692	2	.429
	項目 10 パス	393.235	107	.900	.099	.000	0.855	2	.652
	スカラー不変	512.575	123	.864	.108	.036	120.195	18	.000
	項目 1 切片	393.124	107	.900	.099	.000	0.744	2	.689
	項目 3 切片	403.292	107	.897	.101	.003	10.912	2	.004
	項目 4 切片	403.110	107	.897	.101	.003	10.730	2	.005
	項目 5 切片	416.805	107	.892	.103	.008	24.425	2	.000
	項目 6 切片	392.992	107	.900	.099	.000	0.612	2	.736
	項目 7 切片	438.185	107	.885	.107	.015	45.805	2	.000
	項目 8 切片	434.174	107	.886	.106	.014	41.794	2	.000
	項目 9 切片	393.710	107	.900	.100	.000	1.330	2	.514
	項目 10 切片	392.498	107	.901	.099	.001	0.118	2	.943
修正モデル	配置不変	778.567	435	.959	.054				
	測定不変	838.841	453	.954	.056	.005	60.274	18	.000
	項目 1 パス	781.450	437	.959	.054	.000	2.883	2	.237
	項目 3 パス	782.886	437	.959	.054	.000	4.319	2	.115
	項目 4 パス	779.767	437	.959	.054	.000	1.200	2	.549
	項目 5 パス	793.044	437	.958	.055	.001	14.477	2	.001
	項目 6 パス	780.055	437	.959	.054	.000	1.488	2	.475
	項目 7 パス	789.649	437	.958	.055	.001	11.082	2	.004
	項目 8 パス	789.238	437	.958	.055	.001	10.671	2	.005
	項目 9 パス	778.954	437	.959	.054	.000	0.387	2	.824
	項目 10 パス	781.958	437	.959	.054	.000	3.391	2	.184
	スカラー不変	905.316	453	.946	.061	.013	126.749	18	.000
	項目 1 切片	778.877	437	.959	.054	.000	0.310	2	.856
	項目 3 切片	791.620	437	.958	.055	.001	13.053	2	.001
	項目 4 切片	788.020	437	.958	.055	.001	9.453	2	.009
	項目 5 切片	804.334	437	.956	.056	.003	25.767	2	.000
	項目 6 切片	779.919	437	.959	.054	.000	1.352	2	.509
	項目 7 切片	820.163	437	.954	.057	.005	41.596	2	.000
	項目 8 切片	787.561	437	.958	.054	.001	8.994	2	.011
	項目 9 切片	779.123	437	.959	.054	.000	0.556	2	.757
	項目 10 切片	780.332	437	.959	.054	.000	1.765	2	.414

注) ΔCFI , $\Delta\chi^2$, Δdf , Δp はすべて配置不変モデルとの差を示し、ボールドは DIF 項目を示す。項目 2 のパス係数は 1 に固定、切片には等値制約を課した。韓国の因子平均を 0 に固定した。すべてのモデルの χ^2 値は 0.1% 水準で有意である。

推定値の乖離度と特異項目機能

具体的に RS によってパス係数や切片の推定値にどの程度の乖離がみられたのか、本研究では配置不変モデルにおける平均推定値を算出し検討を進めた。非修正モデルと修正モデルのパス係数の乖離度は、日本が 1.62%、米国は 0.20%、韓国が 0.63% となり、推定値は日本、韓国、米国の順で不当に高く推定されていた。一方切片は、本来あるべき推定値より米国が 2.19% 高く、日本と韓国はそれぞれ -0.12%、-0.27% と低く推定されていた。反応バイアスは、因子モデルにおける項目と因子の関係に影響を与えるため、各種推定値は不当に高く（もしくは低く）推定されることが指摘されており（Cheung & Rensvold, 2000）、RS によって生じた推定値の歪みが非修正モデルの DIF 項目に関与していた疑いがある。

考 察

バイアス項目の特定は国際・文化比較調査の成否を決める重要な要件にもかかわらず、DIF に関する知見はいまだに限定的である。文化的等価性を脅かす要因には、概念的、方法論的、設問・項目的な側面があることは冒頭で述べた通りだが、こうした諸要因が作用して発生する DIF 項目に対して、これまで理論的な裏づけをもって結果が予測されることはほとんどなかった（Chan, 2000）。仮説検証型の DIF 研究の絶対的不足に加え、DIF の理論化を妨げている背景には、検定の繰り返しによって生じる誤謬がある。項目ごとに反復される尤度比検定は、検定に伴う 2 種類の過誤によって、結果の解釈が煩雑になる（Millsap, 2011）。

本研究は DIF の関連要因として議論される RS に焦点を当て、国際・文化比較調査における DIF 項目への RS の関与を検討した。DIF 分析の結果、(a) 非修正モデルで示された DIF 項目の多くが修正モデルでは非 DIF 項目になったこと、(b) 修正モデルと非修正モデルで、パス係数や切片の推定値が異なることの 2 点が明らかになった。

特に注目すべき点としては、非修正モデルで DIF と示された 8 項目のうち 5 項目が、反応バイアスによって説明されたことである。サンプルや質問項目の独自性も考えられるが、本研究の分析では DIF 項目の約 63% は RS に関連していたことになる。RS の要因として議論される属性やパーソナリティ特性などと比較して、RS の文化差やそれによって生じる局外分散は相対的に大きいことが指摘されている（Vaerenberg & Thomas, 2013）。本研究の分析結果はこうした見解を

支持するとともに、国際調査における DIF には文化集団ごとに異なる RS や RS 差から生じる局外分散が関与していることを見出した。

一方で、DIF 項目のすべてが RS に関与しているわけでないことも明らかになった。項目 5（「かなしい映画を観ると人生について考えさせられるから好きである」）と項目 7（「かなしい映画を見ると泣きたくなる」）は非修正・修正いずれのモデルでも均一 DIF となり、RIRSMACS によって切片の推定値が補正されることはなかった。これらの DIF は RS 以外の要因が関係していた可能性が高い。本研究の分析計画ではそれを特定することはできないが、正答・誤答のない社会調査においては、項目の困難度は誘引性（attractiveness）や喚起性（evocativeness）として機能すると考えられており（Lanning, 1991）、質問内容の「なじみやすさ」の文化差がこれらの均一 DIF に関与していた疑いがある。

制御モデルとしての RIRSMACS の活用方法

質問紙の翻訳など、調査過程に得点の比較可能性を妨げる要因が多数存在する国際・文化比較調査では、文化的等価性の確保が重要な要件となる。項目機能の異なる DIF 項目や DIF 項目から得た尺度得点を文化間で比較しても意味を成さないばかりか、誤った解釈を導くこともあり得る（田崎, 2008）。既述のように現時点で DIF に関する知見は限られているが、今後はどのような要因がどういった調査環境において顕著化し DIF を生起させるのか、より精緻化した分析計画の基でさらなる実証研究と知見の蓄積が求められる。

また、こうした基礎的な DIF 研究に並行して、より実用的な観点から測定に伴う誤差を制御・統制する試みも重要になる。一般的な研究者に向けては「バイアス・フリー」な測定を行うための方法論を提示する必要がある。

本研究で用いた RIRSMACS には RS の測定法という特徴に加えて、制御モデルとしての機能がある。当モデルは柔軟性に富み、さまざまなタイプの RS のモデル化が可能で、それらを同時に制御することもできる。最後に実際の調査場面を想定し、制御モデルとしての RIRSMACS の活用法について述べておきたい。

Weijters et al. (2008) は、RS 因子に必要な項目数について、感度分析の結果に基づき、以下のようなアドバイスを書いている。RS 自体に関心がある場合には、安定的に RS 因子を測定する必要があることから、各 RS 観測変数には最低 10 項目（10 項目 × 3 観測変数 = 計 30 項目）、理想的には 14 項目以上（14 項目 × 3 観測変数 = 計 42 項目）を配置する。田崎・申（2017）は、RS 変数に 17 項目を割振っているが、これはこうした指針に従ったものである。一方で RS の解明に関心がない場合には、より簡便に各 RS 変数に最低 2 項目（2

⁵ 非修正・修正モデル間のパス係数および切片の乖離度は Weijters et al. (2008) に従い、[(非修正モデルの推定値) - (修正モデルの推定値)] / (修正モデルの推定値) で計算した。

項目×3 観測変数＝計 6 項目）が用意できればよいという。6 項目なら別途心理尺度を用意しなくても、フェイスシートで尋ねるような一般的な内容の質問項目でも対応できる。特に当該の因子モデルの適合度が上がらない場合には、反応バイアスの関与を疑い RIRSMACS を積極的に導入すべきである (Weijters et al., 2008)。RS 由来の局外分散が整理され、モデルの適合度は改善し、結果的により妥当性・信頼性の高い content factor の測定が可能になる。

利益相反に関する情報開示

本論文に関して、開示すべき利益相反に関する事項はない。

引用文献

- Ainsworth, A. W. (2007). *Dimensionality and invariance: Assessing different item functioning using bifactor multiple indicator multiple cause models* (Unpublished dissertation). Los Angeles, CA: UCLA.
- Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3–23). Hillsdale, NJ: Erlbaum.
- Baumgartner, H., & Steenkamp, J. E. M. (2001). Response style in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38, 143–156.
- Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608–628.
- Chan, D. (2000). Detection of differential item functioning on the Kirton Adaption-Innovation Inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research*, 35, 169–199.
- Cheung, G. W., & Rensvold, R. B. (1999). What constitutes significant differences in evaluating measurement invariance? *Paper presented at the Conference of the Academy of Management Annual Conference* (Chicago, IL), 166.
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *Journal of Cross-Cultural Psychology*, 31, 187–212.
- Ferrando, P. (1996). Calibration of invariant item parameters in a continuous item response model using the extended Lisrel measurement submodels. *Multivariate Behavioral Research*, 31, 419–439.
- 藤本 哲史 (1999). 国際比較研究へのアプローチ——組織コミットメントの日米比較—— 渡辺 直登・野口 裕之 (編) 組織心理測定論——項目反応理論のフロンティア—— (pp. 259–276) 白桃書房
- Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, 6, 243–266.
- 平井 洋子 (2006). 測定の妥当性からみた尺度構成 吉田 寿夫 (編) 心理学研究法の新しいかたち (pp. 21–49) 誠信書房
- Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across Europe: A multigroup latent-class factor approach. *Methodology*, 7, 68–80.
- Lanning, K. (1991). *Consistency, scalability, and personality measurement*. New York: Springer.
- Meade, A. W., Johnson, E. C., & Bradley, P. W. (2008). Power and sensitivity of alternative fit indices in test of measurement invariance. *Journal of Applied Psychology*, 93, 568–592.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide version 7*. Los Angeles, CA: Muthén & Muthén.
- Oliver, M. B. (1993). Exploring the paradox of the enjoyment of sad films. *Human Communication Research*, 19, 315–342.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903.
- Reise, S. P., Smith, L., & Furr, R. M. (2001). Invariance on the NEO PI-R Neuroticism scale. *Multivariate Behavioral Research*, 36, 83–110.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–90.
- 田崎 勝也 (2008). 社会科学のための文化比較の方法——等価性と DIF 分析—— ナカニシヤ出版
- 田崎 勝也・申 知元 (2017). 日本人の回答バイアス——レスポンス・スタイルの種別間・文化間比較—— 心理学研究, 88, 32–42.
- Vaerenberg, Y. V., & Thomas, T. D. (2013). Response style in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Research*, 25, 195–217.
- van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 21–29.
- Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of Academy of Marketing Science*, 36, 409–422.
- Yang, Y., Harkness, J. A., Chin, T. Y., & Villar, A. (2010). Response styles and culture. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, ... T. W. Smith (Eds.), *Survey methods in multinational, multiregional and multicultural contexts* (pp. 203–223). Hoboken, NJ: John Wiley & Sons.