

心理尺度の開発と利用における問題の所在

小杉考司¹

Problems in the Development and Use of Psychological Scales

Koji Kosugi¹

Abstract：心理学の研究方法として心理尺度を用いたものは数多くみられるが、心理尺度が何を表現しうるかについて、その方法論的根拠が明確でないものが多く含まれる。このような問題が生じる原因は、尺度作成が手続的には非常に容易になったこと、客観的に判断する数値基準が容易に達成できることがあると考えられる。そこで本稿ではデモンストレーションとして、身長と体重を測定する客観的な基準を満たす尺度を作成した。この心理尺度の作成を通じて、あたかも心理学的な測定ができたかのような議論の展開ができること、こうした試みが原理的に排除できないことを示した。また心理学的議論がミスリードされないようにするために、心理尺度の仮定や適用の限界を再確認し、今後必要な議論の展望を論じた。

Keywords：心理尺度、尺度構成法、信頼性、妥当性、ボディイメージ

問題

現代心理学の研究においては、心理尺度や質問項目において調査対象者の心理的状态を尋ね、その反応を数値化して分析するという手続きが非常に多く見られる。例えば日本心理学会が2021年度に刊行した「心理学研究」の92巻では、全6分冊を通じて33件の研究報告（資料論文、特集論文を含む）があり、そこには81件の言葉による測定、184の下位概念、因子が含まれている（小杉, 2022）。これらほとんどの研究において、数値化を前提とした言葉による測定がなされている。実験における反応時間やパターンだけを取り扱っていたり、集計されたデータだけを分析対象とした研究は小山・大久保（2022）や福田ほか（2021）など数件にとどまる。また、ほとんどの研究例において回答者自身が尺度に反応する自記式であり、他記式の研究例はわずか1件（永谷ほか, 2022）のみである。このような傾向はおそらく「心理学研究」誌だけにみられるものではなく、現代心理学研究の多くは「本人に心理尺度に回答してもらう」形でデータ化されたものを研究対象にしていると考えられる。

本稿では心理尺度が多用されるこのような現状を鑑み、あらためて心理尺度の原理および方法論的特性を検証し、問題点を指摘するとともに今後の展望を論じる。

心理尺度は多くの研究で用いられているだけでなく、その開発も盛んである。例えば論文検索システム Google Scholar において、キーワード「尺度作成の試

み」で検索したところ、2022年度以降に限定しても1220件の論文や報告がヒットする（2022年10月14日現在）。年間1000件を超える開発報告があることから、心理尺度は乱立状態にあるともいえるだろう。尺度の多さ、作られすぎ問題については批判もあり、その論点は主に構成概念妥当性の側面（吉田・石井・南風原, 2012；宇佐美, 2016）や信頼性の側面（高本・服部, 2015；岡田, 2015）からのもの、また尺度作成の手続きが形式化していることを指摘するもの（吉田ほか, 2012；仲嶺・上條, 2019）である。

一般的な尺度作成手続きは、本来非常に手間暇のかかる作業である。特に統計パッケージが今ほど発展していなかった時代は、探索的因子分析を大型計算機で実行するだけでも大変な時間を必要とした。しかし昨今、PCなど小型計算機が普及し、また統計モデルやパッケージが発展したことにより、複雑な数値計算も瞬時に結果を出すことが可能になった。かつて因子分析といえは探索的に行うしかなかった時代に比べ、昨今は構造方程式モデリングが統計パッケージとして簡単に利用できることから、検証的因子分析の利用も数多く見られるようになった。項目反応理論を使った項目特性と潜在特性の精緻な検証なども、専門的なソフトウェアを必要とせず、統計環境 R (Ihaka, and Gentleman, 1996) のパッケージ (Rizopoulos (2006) など) が利用できる。また Web を使った調査にすれば、質問紙のデザイン、印刷、配布、回収、入力といった労力もほとんど必要としない。こうした実践上のコスト低下が尺度作成を容易なものとし、心理尺度の乱立とも呼べる現状の大きな要因になっていることは間違いないだろう。

しかし計算や実践上のコストが下がったこと、つまり

受稿日2022年11月14日 受理日2022年11月28日

1 専修大学人間科学部心理学科 (Department of Psychology, Senshu University)

心理尺度を容易に作れることがすなわち、測定が容易になったことを意味するものではない。むしろ安易に測定「した」ことで、「できた」ものと勘違いし、心理学研究を誤った方向に導く危険性が増している可能性が高い。続く節ではこのことについて、具体的なデータを用いてデモンストレーションを試みる。そこでの目的は、一般的な尺度作成手続に即して、 α 係数が0.8以上の内的整合性を持ち、仮説通りの因子構造と、十分に高い基準関連妥当性を持つ心理尺度を作成できることを示すことにある。

もちろん心理尺度の評価は尺度作成手続の適切さだけでなく、尺度が必要とされる理論的背景や、適用される領域での有用性なども含めてなされるべきものである。しかし理論的な背景や運用時の有用性については、個別の研究領域におけるドメイン知識や実態に依存するものであるから、客観的な評価の対象となりにくい。逆にいえば、手続的に大きな問題がなく数値的な基準を達成した尺度開発研究が示されたとき、適切な心理尺度として公開され、既成事実として利用されてしまうことがあるのではないだろうか。

ここで示すのは、内容的な妥当性の検証を避けても、一定の手続きを踏めばある程度の数値的精度をもった尺度が開発可能なことである。逆説的ではあるが、このことは一般的な尺度構成手続きの頑健さを示すことになるかもしれない。しかし本論の目的は、心理尺度開発として形式的に成立していることが、適切な測定を意味するものではないことを示すことにある。

デモンストレーションのための題材として、客観的な指標との対比を行いやすい身体特徴の主観的捉え方をとりあげ、その最も単純な例として、身長と体重に関する心理尺度の作成を考える。身長と体重を心理尺度で測定した場合、物理的な数値についての主体的な評価であるから、そもそもの物理量と強い相関を持つことは容易に想定される。自分自身の身長と体重については誰しも測定された経験があり、ある程度その客観的情報を有していると考えられ、これを基準関連妥当性を検証するための材料としたい。

身体的特徴については個人間で比較することも可能であり、またその主観的イメージも個人間で比較することができると仮定するのに無理はない。そこで社会的態度のように、その要素として認知、感情、行動の三成分（藤原，2001）を想定し、それぞれに対応した項目を考える。認知成分については、身体的特徴を自身でどのように把握しているかとして、「自分は人より背が高い方

である」といった項目などで測定できると考えた。感情成分については、身体特徴に対する自分の好悪感情であるとして、「背の高い人を見ると羨ましいと思う」といった項目などで測定できると考えた。行動成分については、身体的特徴に基づいてどのような意図を持ち行為を企図するかを示すものとして、「靴底の厚いものを履くことがある」といった項目などで測定できると考えた。

分析方法としては、これらの尺度項目に対して探索的因子分析を行い、3因子構造になることを確認する。また尺度得点と言語報告による身長と体重の値の相関関係を検討することとする。

方法

参加者および手続き

調査はすべて Web を介して行われた。調査フォームの作成には、jsPsych (de Leeuw, 2015) および Pavlov-ia を用い、参加者の募集にはクラウドソーシングサービスのクラウドワークス (<https://crowdworks.jp/>) を用いた。

調査対象者はクラウドワークスを通じて、「ボディイメージとライフスタイルについての調査」と題した調査として1000名が募集された。クラウドワークス上で研究同意書に同意を求めた後、リンク先に用意された調査フォームに誘導し、回答してもらった。参加者には回答終了後、パスワードが表示され、それをクラウドワークス上から入力するよう指示し、パスワードが符合していた場合にのみ100円の報酬が支払われた。調査期間は2020年11月10日から2020年11月12日であった。なおこの調査は、専修大学人間科学部心理学科「人を対象とした研究倫理委員会」の承認を受けて行われたものである（申請番号 19-S004-1）。調査を最後まで完了せず、途中でやめた場合でもサーバ上に回答データが記録されるため、回答データはすべて合わせると1102名（うち男性398、女性679名、答えたくない25名）であった。このなかで、パスワードが合致し報酬が支払われたのが1000名であり、これを分析対象のデータとする。また回答者のなかで Satisficer 検出項目（三浦・小林，2018）によって不適切と判断されたものや、実際の身長が100cm 以下や250cm 以上の回答をしていたり、年齢が100歳を超えていたものは真面目に回答されなかったものと判断して分析からは除外した。その結果、最終的な調査対象者は954名（うち男性342名、女性596名、答えたくない16名）であった。

調査項目

調査項目は、基本属性（年齢、身長、体重、性別、最終学歴、回答に使用している端末、回答環境）および、Satisficer 検出項目（三浦・小林, 2018）と、次に示す尺度項目である。

尺度項目

項目群は身長および体重それぞれについての心理的イメージを測定するため、筆者が独自に考案した各30項目である。30項目の内訳は、身長と体重について事実の記述である認知次元10項目（自分は人より背が高い方である、など）と、好悪や良し悪しの価値判断からなる感情次元10項目（背の高い人を見ると羨ましいと思う、など）、経験や行動についての記述からなる行動次元10項目（他人を見上げて話すことが多い、など）からなる。回答者はそれぞれの項目に対して「非常によく当てはまる」から「全く当てはまらない」まで5件法で回答を求めた。

結果

分析には R (Ihaka and Gentleman, 1996) および psych パッケージ (Revelle, 2021) を用いた。また本研究に用いたデータや分析コード、詳細な分析結果は、Open Science Framework にて公開されている¹。

データの基本的特徴

調査回答者の基本属性について Table 1 に示す。なお回答は「非常によく当てはまる」を5、「やや当てはまる」を4、「どちらともいえない」を3、「あまり当てはまらない」を2、「全く当てはまらない」を1と数値化した。身長尺度について、 α 係数は0.848、 ω 係数は0.881であり、尺度項目全体としての内的整合性信頼性は十分であると判断した。同様に体重尺度について、 α 係数は0.850、 ω 係数は0.885であり、こちらも尺度項目全体として十分な信頼性が認められた。

二つの尺度の探索的因子分析

身長および体重尺度それぞれについて、ポリコリック相関係数に基づく並行分析によって因子数を求めたところ、いずれも7因子モデルが提案された。両尺度のスクリープロットを Figure 1 に示す。固有値の減衰状況は、身長尺度において7.185, 2.881, 1.144, 0.839, …であり、同様に体重尺度において7.133, 2.666, 1.640, 0.655, …となっていた。固有値1.0以上という基準、および仮説に基づく解釈可能性から、3因子解とした。

因子抽出には最小二乗法と同等の解を示す最小残差 (minres) 法、回転はより単純構造に近くなるように回転させる斜交ジオミン回転を選択した。身長尺度の記述統計量と因子分析結果を Table 2 に、体重尺度の結果を Table 3 に示す。

またこのなかで、共通性が0.3よりも小さい項目を除外し再度探索的因子分析を行った。Table 4、Table 5 にこの結果と、下位因子ごとの信頼性係数を示す。

二つの尺度のスコアと相関関係

ここまでで得られた、3つの因子得点と自己申告の値の相関係数を Table 6 に示す。ここでは下位因子に分けずに尺度全体を1因子として考えた時の推定値、また逆転項目に配慮して項目得点を単純加算して得られた簡便的因子得点との相関係数も合わせて示している。またここでの因子得点の算出には、Table 2、3 に示した当初の仮説的因子構造に基づいている。なお1因子モデルとして分析するにあたっては、段階反応モデル (Samejima, 1997) を適用し、ltm パッケージ (Rizopoulos, 2006) の grm 関数を用いて潜在得点を推定した。

考察

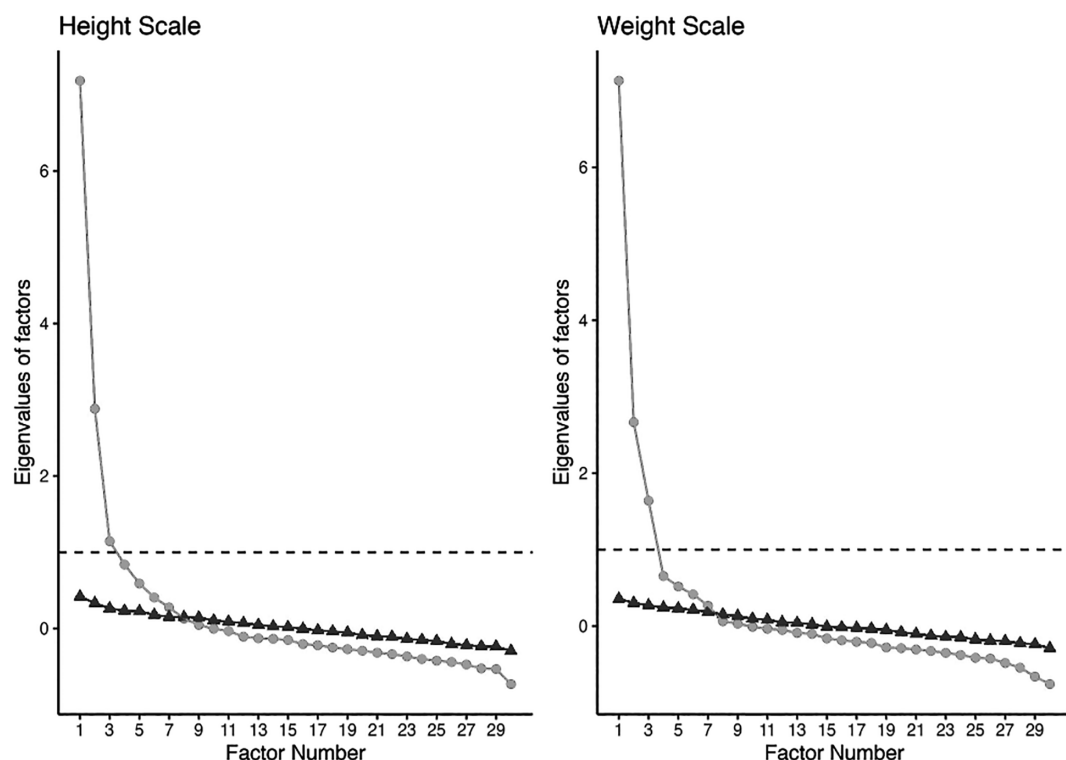
結果が示すもの

結果で示された通り、身長尺度も体重尺度も、尺度全体での α 係数が0.8以上であり、これらの尺度の内的整合性は十分高いことが確認された。一般的に、下位因子が認められる場合は下位因子ごとに α 係数を算出することがある。Table 4 および Table 5 にあるように、下位

Table 1. 性別ごとの記述統計量（平均値と括弧内は標準偏差）

変数	女性, N = 596	男性, N = 342	答えたくない, N = 16
年齢	36.51 (9.295)	39.89 (10.579)	38.62 (8.245)
身長	158.55 (5.806)	171.60 (5.626)	161.94 (7.122)
体重	52.12 (8.442)	67.47 (12.580)	55.31 (8.792)

Figure 1 身長尺度（左）と体重尺度（右）のスクリープロット



因子ごとにみると一部この基準を達成していない因子（身長尺度の第二、第三因子、体重尺度の第三因子）が見られるが、 α 係数よりも精度が高いとされる ω 係数（清水，2007）ではこれらの基準を満たしている。

また、3因子構造を想定した項目群を用意した結果、身長尺度と体重尺度のいずれにおいても3因子解が妥当であると判断された。並行分析の結果は必ずしもこの構造を支持しなかったが、因子構造の決定についてはいくつかの基準があり（小杉，2019）、固有値の減衰状況からみるという基準では仮説を支持する結果であった。

さらに因子得点と言語報告の相関関係は、身長尺度において最大の絶対値が0.678、体重尺度において最大の絶対値が0.376であることが示された。調査研究における基準関連妥当性として考えた場合、これは中程度からやや大きい相関係数であるから、この尺度で身体的特徴の言語報告を予測することがある程度は可能であるといえるだろう。

一方、この試みの問題点もいくつか挙げられる。まず尺度項目が筆者一人によって考案されたものであることから、そもそも準備された項目群の選出根拠が個人の主観であり一般性に欠くものであったかもしれない。第二に、3因子構造ではあるが因子構造は必ずしも認知、感情、行動の態度成分に分割されていないことが挙げられる。特に共通性の基準などから項目を除外した場合、因

子ごとに寄与する項目の数や種類が異なっている。洗練された尺度とするためには、項目の入れ替えなど吟味を重ねる必要があるだろう。第三に、身長と体重についての言語報告と尺度得点の相関が小さくないことを認めたとしても、妥当性の検証としてはごく一面的であるとの批判は避け得ない。心理尺度の信頼性、妥当性の検証として、より多角的な検証や指標、論拠が必要であろう。

これらの批判に対して真面目に答えるならば、尺度項目を複数の専門家によって考案し、因子構造が仮説通りになるように項目構成を洗練させ、また言語報告だけではなくそのほかの指標や基準で精緻化しなければならない、となるだろう。しかしこの試みはあくまでもデモンストレーションであり、筆者の目的は身長や体重を測定するための精緻な心理尺度を作ることではない。むしろ結果で示したような数値目標が達成できてしまったことに注目したい。すなわち、この心理尺度で何を測ろうとしていたか、ひいては何が測れるのかという点を明確しないまま、測定だけでできてしまうことを警戒すべきなのである。筆者が本稿において心理的身長や心理的体重を発見した、あるいはそれを想定することで実効的な意義があると論じることは、はたして適切なことだろうか。

以下ではこの問題を踏まえ、心理尺度の開発と利用の背後にある本質的な問題の所在について、今後の検討課

Table 2. 身長尺度の因子負荷量ならびに記述統計量

項目	第一因子	第二因子	第三因子	共通性	M	SD
因子負荷量						
小さいサイズの専門店があれば良いのと思う	0.836	-0.134	0.251	0.528	2.215	1.124
他人を見上げて話すことが多い	0.714	0.134	-0.041	0.649	2.521	1.111
服はだいたい S サイズである	0.703	-0.185	-0.049	0.420	1.941	1.235
自分は人より背が低いわけではない	-0.655	-0.113	0.197	0.652	3.354	1.178
高いところに手が届かないことがある	0.608	0.065	-0.288	0.625	3.367	1.137
他人と肩を組むと肩の位置が合わないことがある	0.602	0.007	0.415	0.362	3.093	1.043
人混みの中では周囲が見渡せない	0.597	0.093	-0.257	0.602	2.891	1.168
自分の靴のサイズは平均より小さい	0.592	-0.125	-0.092	0.334	2.553	1.189
靴底の厚いものを履くことがある	0.341	0.162	0.232	0.187	2.040	1.149
電車ではつり革より手すりを持つことが多い	0.298	0.050	-0.101	0.139	3.306	1.117
自分は中肉中背である	-0.211	0.158	-0.131	0.039	2.955	1.178
背の高い人を見ると羨ましいと思う	-0.024	0.884	-0.099	0.791	3.138	1.216
私の背がもう少し高ければ良いのと思う	0.114	0.724	-0.145	0.680	3.295	1.299
背が高い人は得をしていると思う	-0.034	0.678	-0.032	0.444	3.253	0.941
背が低いことで損をしていることがあると思う	0.326	0.530	-0.006	0.564	2.646	1.215
身長で社会的な評価が変わってくると思う	0.064	0.486	0.216	0.282	2.494	1.073
よく寝て身長を伸ばそうとしている	0.147	0.471	0.367	0.365	1.704	0.906
背の高さを気にしたことがない	-0.230	-0.411	-0.186	0.303	2.925	1.227
自分の体型はずんぐりむっくりである	0.049	0.355	-0.004	0.147	2.512	1.249
背が高い人のことが好きである	-0.019	0.328	0.101	0.104	3.422	0.940
身長が低くてもそれを活かした活躍はできると思う	-0.120	-0.273	-0.019	0.120	3.988	0.775
猫背になりがちである	-0.108	0.146	0.061	0.023	3.591	1.202
人混みの中では目立つほうである	-0.077	0.045	0.720	0.554	2.097	1.025
古い建物などで頭をぶつけることがある	-0.022	-0.005	0.659	0.445	1.955	1.087
自分は人より背が高い方である	-0.457	-0.049	0.610	0.810	2.458	1.224
電球の取り替えを頼まれることが多い	-0.274	-0.018	0.524	0.459	2.041	1.129
大きいサイズの洋服専門店に行くことがある	-0.075	0.199	0.505	0.286	1.603	1.039
自分は他人より腕が長いほうだ	-0.207	-0.043	0.500	0.383	2.603	1.158
サイズの合う服がみつからないことがある	0.460	0.039	0.470	0.295	2.780	1.276
身長のことを褒められると嬉しい	-0.115	0.315	0.354	0.201	2.615	1.028
因子負荷量平方和						
	4.971	3.489	3.334			
因子間相関						
	1.000	0.507	-0.352			
		1.000	-0.129			
			1.000			

題を展望として論じたい。

展望

心理尺度が測定しているものは何か

ここであらためて、心理尺度とは何を測定しうるものなのかを確認しておこう。

心理尺度の多くは、リッカート法と呼ばれる数段階のカテゴリを選択させる形式である。Likert (1932) は Thurstone (1928) の社会的態度を測定する方法の簡便的手法としてこの方法を開発しており、そのシグマ法による数値化は社会的態度が正規分布で近似できるという

仮定のうえに成り立っていたことを再度確認しておかなければならない。

社会的態度は行動の準備傾向 (preparation or readiness for response) であって (Allport, 1967), 表出される意見 (opinion) に反映される心理的実体である態度は個人によってその強度が異なる, という仮定が含まれている。個人間の相対比較可能な対象で、かつその分布に基づく仮定からカテゴリに数値が割り振られるのである。この仮定からは、個人の主観的な経験などの心理的状态、あるいは個人の内にある心理学的実体を数値化し、比較可能であるとすることは導けない。もしその尺

Table 3. 体重尺度の因子負荷量ならびに記述統計量

項目	第一因子	第二因子	第三因子	共通性	M	SD
因子付加量						
自分は肥満体型である	0.864	0.018	-0.018	0.749	2.421	1.315
健康のために痩せなければいけないと思う	0.846	0.047	-0.138	0.701	3.202	1.288
自分は太りやすい体質である	0.785	0.131	-0.113	0.666	3.060	1.296
自分は今、理想体重から離れてしまっている	0.750	-0.004	0.020	0.569	3.295	1.236
自分の体型が嫌いだ	0.749	-0.091	0.117	0.590	3.280	1.126
自分は昔の体型を維持できている	-0.747	-0.007	-0.004	0.563	2.752	1.258
自分は体型のことで悩むことはない	-0.730	-0.020	0.007	0.539	2.468	1.136
いろいろなダイエットに挑戦したことがある	0.643	0.100	-0.231	0.442	2.781	1.300
健康診断の前に食事制限をすることがある	0.547	-0.012	-0.136	0.274	2.592	1.246
体型を強調しない服装を選ぶようにしている	0.506	0.042	0.150	0.333	3.206	1.133
前屈みになる体勢が苦しいことがある	0.458	-0.048	0.120	0.243	2.445	1.193
自分の体を思うように動かせないことがある	0.363	0.005	0.334	0.309	3.015	1.158
靴下を履く時はどこかに腰掛けるようにしている	0.214	-0.082	0.189	0.102	2.316	1.197
太っている人は自制心がないと思う	0.045	-0.022	0.017	0.003	3.221	1.036
自分は食べるのが好きである	-0.031	0.963	0.081	0.900	4.112	0.911
食事があまり楽しいことだと思わない	0.190	-0.836	0.036	0.648	1.897	0.978
自分は食欲が旺盛な方である	0.181	0.712	-0.021	0.620	3.519	1.062
食べることでストレスを解消している	0.194	0.628	0.165	0.530	3.310	1.222
痩せたい気持ちよりも食欲が勝る	0.198	0.625	0.285	0.581	3.290	1.089
美味しく食事することは精神衛生上よいことだ	-0.163	0.559	0.103	0.273	4.444	0.679
食べすぎることがあるのは仕方がないと思う	0.003	0.407	0.194	0.189	3.616	0.894
自分は何を食べるかにあまりこだわらない方である	0.012	-0.359	0.118	0.150	2.866	1.157
忙しくて食事を抜くことがある	0.040	-0.238	0.076	0.064	2.543	1.230
痩せていると不幸せそうに見えると思う	-0.100	-0.178	0.068	0.056	2.420	1.017
運動することが嫌いだ	0.056	0.002	0.823	0.705	2.955	1.272
自分は運動があまり得意ではない	0.087	0.025	0.674	0.492	3.343	1.294
習慣的に運動するようにしている	0.059	0.001	-0.654	0.412	2.904	1.263
わかっていても食習慣は変えられないと思う	0.103	0.268	0.378	0.243	3.351	1.020
毎日体重を測るようにしている	0.295	-0.036	-0.353	0.148	2.388	1.442
自分は栄養不足気味である	-0.063	-0.236	0.237	0.128	2.469	1.152
因子負荷量平方和						
	6.042	3.842	2.338			
因子間相関						
	1.000	0.308	0.266			
		1.000	-0.101			
			1.000			

度が社会的態度とみなせないものであるなら、数値化の根拠を別途示さなければならないはずである。

そこであらためて、尺度を項目に対する人の反応のセットとして形式的に捉えてみると、心理学で実際に使われている尺度の用法は大別して3つあると考えられる。第一にはテスト理論(池田, 1994; van der Linden, and Hambleton, 1996; 豊田, 2012; 加藤・山田・川端, 2014)に基づく、学力などの個人特性の測定である。これは項目への反応における正答が個人の内部ではなく外的に定められており、測定されるものは検査対象者が有していると仮定される性質(学力・能力)である。項目

への反応パターン、対象者の正答パターンから、項目と回答者はそれぞれ序列化され、正答パターンが当該対象者の内的な性質を反映していると考えられる。この性質、すなわち学力は正規分布に従うと仮定されるが、測定対象の性質上、この仮定は妥当なものとして受け入れられやすい。また測定したとされる対象は知識量や記憶量といった、調査対象の内部で時間的に安定した状態であることも指摘しておきたい。

第二の用例は正答が外的に定義できないもので、反応パターンから対象者を分類するために用いられる尺度であり、パーソナリティ検査などがそれにあたる。そうし

Table 4. 共通性の低い項目を除外した身長尺度の因子負荷量

項目	第一因子	第二因子	第三因子
因子負荷量			
小さいサイズの専門店があれば良いのと思う	0.840	-0.091	0.311
他人を見上げて話すことが多い	0.743	0.100	-0.017
服はだいたいSサイズである	0.656	-0.110	-0.025
自分は人より背が低いわけではない	-0.655	-0.148	0.161
自分の靴のサイズは平均より小さい	0.640	-0.137	-0.040
人混みの中では周囲が見渡せない	0.634	0.055	-0.229
高いところに手が届かないことがある	0.616	0.048	-0.283
他人と肩を組むと肩の位置が合わないことがある	0.561	0.072	0.432
背の高い人を見ると羨ましいと思う	-0.075	0.947	-0.064
私の背がもう少し高ければ良いのと思う	0.047	0.805	-0.129
背が高い人は得をしていると思う	-0.017	0.650	0.013
背が低いことで損をしていることがあると思う	0.330	0.534	0.044
背の高さを気にしたことがない	-0.154	-0.485	-0.186
よく寝て身長を伸ばそうとしている	0.130	0.443	0.344
人混みの中では目立つほうである	-0.124	0.070	0.699
古い建物などで頭をぶつけることがある	-0.015	-0.003	0.676
自分は人より背が高いほうである	-0.478	-0.048	0.591
電球の取り替えを頼まれることが多い	-0.267	0.006	0.558
自分は他人より腕が長いほうだ	-0.245	-0.019	0.476
因子負荷量平方和			
	4.515	2.926	2.631
因子間相関			
	1.000	0.525	-0.360
		1.000	-0.182
			1.000
信頼性係数			
α 係数	0.840	0.797	0.789
ω 係数	0.872	0.866	0.825

た研究目的の一つは説明次元の数や構造の解明であり、個人差の記述であると考えられる。測定の対象であるパーソナリティとは、個人のなかであまり変動しない行動傾向である（小塩，2020）。ここでの目的は、測定対象が社会全体において示す傾向の記述、それに基づく標準化や個々人の相対的比較であり、個々人の主観的な経験や個人的な感情を測定することではない。測定されるスコアの母集団分布に正規分布を仮定することもまた、こうした社会的な分布に準拠することや、パーソナリティが無数の要因の結果として醸成され则认为られるのであれば、正当化されるものである。

第三の用法は、主観的経験の数値化、平たく言えば「何をどう感じているか」を定量的に評価するものである。例えば川崎・小塩（2021）では、病理的自己愛目録日本語版として「私はよく、自分の偉業が認められるという空想をする」といった項目を用いている。小野・及川・及川（2021）は感情反応を測定する項目として、

「やる気がわく」「怒りを感じる」といった項目を用いている²。天井（2021）は中学生を対象とした情緒的援助期待尺度として、「自分はいい人間だと思わせてほしい」「がんばれと言ってほしい」といった項目を用いている。このように、自分の意図、信念、自己評価、願望、感情などを測定するものとして尺度上への反応が用いられている事例は枚挙にいとまがない。自分が感じていることに最も近い項目カテゴリの選択を求めるこの用法は、当然のことながら客観的に正答・誤答を判定できるものではない。またこうした主観的経験についての反応を集め、他者との相対比較で個々人を位置づけることが可能であるとする理論的根拠も明らかではない。多用されている方法ではあるが、なぜそれがリッカート法で測定しうるのか、また測定の結果因子として抽出されたものが何なのかについて、改めて検討する必要がある。

Table 5. 共通性の低い項目を除外した体重尺度の因子負荷量

項目	第一因子	第二因子	第三因子
因子負荷量			
自分は肥満体型である	0.843	0.035	0.006
健康のために痩せなければいけないと思う	0.802	0.072	-0.080
自分の体型が嫌いだ	0.788	-0.123	0.132
自分は昔の体型を維持できている	-0.781	0.026	-0.008
自分は今、理想体重から離れてしまっている	0.781	-0.031	0.010
自分は体型のことで悩むことはない	-0.762	0.028	-0.030
自分は太りやすい体質である	0.754	0.129	-0.049
いろいろなダイエットに挑戦したことがある	0.586	0.110	-0.134
体型を強調しない服装を選ぶようにしている	0.491	0.051	0.155
自分の体を思うように動かせないことがある	0.352	-0.001	0.316
自分は食べるのが好きである	-0.081	0.978	0.028
食事があまり楽しいことだと思わない	0.197	-0.783	0.060
自分は食欲が旺盛なほうである	0.126	0.766	-0.069
痩せたい気持ちよりも食欲が勝る	0.160	0.649	0.181
食べることでストレスを解消している	0.153	0.639	0.101
運動することが嫌いだ	-0.018	0.022	0.965
自分は運動があまり得意ではない	0.058	0.005	0.765
習慣的に運動するようにしている	0.047	0.017	-0.620
因子負荷量平方和			
	5.285	3.109	2.170
因子間相関			
	1.000	0.379	0.249
		1.000	-0.011
			1.000
信頼性係数			
α 係数	0.890	0.845	0.783
ω 係数	0.911	0.887	0.837

Table 6. 身長体重尺度の推定されたスコアと合計得点及び言語報告との相関

	第一因子	第二因子	第三因子	GRM 推定値	項目素点合計	言語報告
身長尺度						
第一因子	1.000	0.532	-0.397	0.935	0.912	-0.663
第二因子		1.000	-0.158	0.669	0.757	-0.274
第三因子			1.000	-0.579	-0.438	0.532
GRM 推定値				1.000	0.946	-0.678
項目素点合計					1.000	-0.643
体重尺度						
第一因子	1.000	0.298	0.280	0.971	0.892	0.376
第二因子		1.000	-0.098	0.454	0.572	0.110
第三因子			1.000	0.282	0.366	-0.087
GRM 推定値				1.000	0.941	0.346
項目素点合計					1.000	0.268

心理尺度で測定できるものは何か

測定とは一定のルールに従って対象に数値を与えること、とされる。リッカート法による尺度の数値化は、社会的態度が正規分布に従うと仮定されることから正当化されるのであった。またテスト理論やパーソナリティ検

査でも同様に、潜在変数に正規分布を想定することは妥当なものであろう。必ずしも正規分布でなくても良いが、カテゴリに数値を与えること、またそれが間隔尺度水準であることにこうした理論的背景があったことは、改めて周知されるべきである。よくできた尺度項目はお

おもねカテゴリが等間隔になるという研社会的態度研究の歴史的経緯を経て、シグマ法ではなく5つの段階を順に1, 2, 3, 4, 5と数値化しても近似としては十分である、と許容されてきたに過ぎない。分布の形状が不明確なものに、時には報告すらされないカテゴリを付与して、端から順に整数を割り振ることには、いかなる数理的根拠もないといっているだろう。

また測定されると想定されたものが、学力やパーソナリティ、社会的態度のように、他者と比較可能なものであることが理論的に正当化される必要がある。自分の願望や評価、感情が他者と同じ程度かどうかをいかにして判断できるのかについては、何らかの哲学的決断を含んでいると考えられる。例えばそれは、項目や反応カテゴリが言語という共通基盤に立脚し、その意味内容が誰に対しても均一であるとみなす、という大胆な仮定であるかもしれない。しかし仮にそこまで受け入れたとしても、そうした共通化されたラベルに対する反応パターンから得られる、潜在変数とはどこにある何を表すのであろうか。

因子分析は個人と変数という2相データの、個人の相を集計することでつづいた相関行列を作り、項目の作る空間の基本となる次元（基底）に注目する手法である（芝, 1979; 小杉, 2019）。つまり項目のもつ静的な構造を考えているのであり、あくまでも項目の世界を探索し、そこに個人の反応パターンを投影しているにすぎない。静的な構造、つまり測定対象が時間的また概念的に変化しにくい性質を持っていることが必要であり、不安定で個別性の高いものは因子分析で捉えきれものではない。社会的態度やパーソナリティでの因子分析の利用は、過去の出来事や社会的に共通した認識（コモンセンス）を取り出すことを目的としていることから、その安定性や共通性が正当化される。逆に言えば、時事刻々と変化する事象や個別性が高いと思われるもの、例えば感情や生理的反応の言語報告、恋愛関係や親子関係などプライベートな関係についての反応、時事的なニュースなど不安定な対象に対する評価に、因子分析を適用することは適切ではなく、適切な使用方法について改めて見直されるべきである。

また尺度項目間の基底は尺度項目の特徴であって、人の心理に実在するものでないことにも注意が必要である。個人内の心理プロセスと個人間の相対比較は、明確に区別されなければならない（村山, 2012）。社会的態度は個人間比較する次元の存在を想定しているのであり、その次元は評定者の内部ではなく外部の社会空間に

実在する共通基盤である。社会心理学における社会関係の記述や、質問紙上での変数間関係に基づく論考は、個々人の態度や意見、経験に還元して議論できるものではなく、言語的表現にみられる共通パターンにすぎない。その意味では、パーソナリティ研究における測定も、観察者の内部に性格なる心理的実体が存在すると考えるのではなく、より慎重な解釈が必要である。渡邊・佐藤（1991）や渡邊（2010）で論じられているように、われわれが他者の振る舞いに言語的な意味づけをするときに共通する評価次元を性格と呼ぶ、というほうが因子分析モデルに即した解釈である。パーソナリティや社会的態度が個人に内在する特性と考えるのは、因子分析の測定論と実在論を十分区別せずに議論していることでもある（Borsboom, 2005）。

この問題を解決しうる方略としては、次のような割り切った考え方に立脚することが考えられる。すなわち心理尺度を用いる研究はすべて、その項目が言語的な意味において社会的に共通しており、またその目的も言語あるいは意味空間における共変動を研究することである、とするのである。例えば場面想定法や未来の状態の言語評価などは、言葉で制限された仮想世界に限定でき、スクリプト処理のパターンを観測しているのだと考えれば正当化される。言葉の世界に限定されるとはいえ、言葉の世界のなかではその実在性を認め、それが社会的に実在すること、社会とは、心とはそのようなものだと前提するのである。こうした観点に立てば、本稿で示した身長・体重尺度も、「自身の身長および体重に関する態度」を測定するためのものであるとし、何らかの実在に対する心理学的背景なるものを想定することも可能になる。身長や体重に対する態度と類似する概念としては、ボディ・イメージ（柴田, 1992; Cash, and Fleming, 2002）の研究があり、これは社会的あるいは臨床的には有用な概念であろう。そこでこれを足がかりにしながら、心理的あるいは社会的身長・体重態度の実在性、有効性を議論することもできるかもしれない。あるいは因子得点と言語報告との相関も高々0.7弱でしかないことから、残りの説明できない分散に意味を見出して、物理的実体とはことなる心理的実体を想定して議論することもできるかもしれない。しかし本稿のデモンストレーションはそうした議論に寄与するものではないし、むしろ指摘したい問題はこうした詭弁を原理的に排除できないことなのである。

測定できたから実在し、実在するから意義があるというのは研究領域全体の HARKing（池田・平石, 2016）

といえるかもしれない。本稿で示したように、測定そのものや数値的基準の達成そのものは容易かもしれないが、理論的裏づけや包括的な妥当性を踏まえて適切に心理尺度を使うことは非常に困難なことなのである。心理学的実体を想定しうるかどうかは、その哲学的立場の違い、すなわち専門家集団の共通見解に依存する依存してしまうため、心理的身長、社会的体重の実在性も完全に否定することはできない。もちろん客観的、数値的な基準での判断もできないからといって、数値のような客観的な指標を避け、よく利用されているから良い尺度であるといった権威づけに依存するのも、同様に危険な方略であるだろう。

尺度作成は進んだ統計技術の応用であり、これに限らず心理学者はユーザとして統計手法を駆使して研究をかさねている。統計ツールの使い方については最近も吉田・村井（2021）で指摘されたようにその誤用が問題になることがあり、こうした指摘は心理学者にとっては耳の痛い問題であるだろう。また豊田（2020）が指摘するような統計学全体の問題も、一介のユーザである心理学者にとっては抱えきれないほど大きな問題であるかもしれない。しかしそれらよりもはるかに根本的で重要な問題である、「心とはなにであり、それを測定するとはどういうことか」という点がなござりにされているのではないか。この点こそ心理学者が直面すべき、心理学の本質的な問題関心であるはずで、測定論的な議論（Borsboom, 2005; 吉野, 2007）を踏まえて、今後取り組まなければならない大きな課題であるだろう。

利益相反

本論文に関して、開示すべき利益相反関連事項はない。

謝辞

本研究の着想に際して、帯広畜産大学教授、渡邊芳之氏の助力があったことをここに記し、感謝の意を表します。

引用文献

- Allport, G.W. (1967). Attitudes In M. Fishbein (Ed.), *Readings in Attitude Theory and Measurement* (pp.3-13). New York : John Wiley & Sons Inc.
- 天井響子 (2021). 青年期前期の援助評価に対する情緒的援助期待の影響 *心理学研究*, 92 (2), 140-150. 10.4992/jjpsy.92.19233
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, UK: Cambridge University Press.
- Cash, T.F. and Fleming, E.C. (2002). The impact of body image experiences: Development of the body image quality of life inventory, *International Journal of Eating Disorders*, 31 (4), 455-460. <https://doi.org/10.1002/eat.10033>
- 藤原武弘 (2001). 社会的態度の理論・測定・応用 関西学院大学出版会
- 福田恭介・水口美咲・松尾太加志・志堂寺和則・早見武人 (2021). 「喉まで出かかっている」ときの瞬目の抑制と発生 *心理学研究*, 92 (2), 122-128. 10.4992/jjpsy.92.20023
- Ihaka, R. and Gentleman, R. (1996). R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*, 5 (3), 299-314.
- 池田央 (1994). 現代テスト理論 朝倉書店
- 池田功毅・平石界 (2016). 心理学における再現可能性危機：問題の構造と解決策 *心理学評論*, 59 (1), 3-14.
- de Leeuw, J.R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser, *Behavior research methods*, 47 (1), 1-12.
- 加藤健太郎・山田剛史・川端一光 (2014). Rによる項目反応理論 オーム社
- 川崎直樹・小塩真司 (2021). 病理的自己愛目録日本語版 (PNI-J) の作成 *心理学研究*, 92 (1), 21-30. 10.4992/jjpsy.92.19217
- 小杉考司 (2019). 言葉と数式で理解する多変量解析入門 北大路書房
- 小杉考司 (2022). 広島大学集中講義資料・心理教育測定法 https://kosugitti.github.io/psychometrics_syllabus/Scaling.pdf, (2022.10.15 アクセス)
- Likert, R. (1932). A technique for the measurement of attitudes In R. Woodworth (Ed.), *Archives of Psychology* (pp.5-55). New York : Columbia University
- 三浦麻子・小林哲郎 (2018). オンライン調査における努力の最小限化が回答行動に及ぼす影響 *行動計量学*, 45 (1), 1-11.
- 村山航 (2012). 妥当性概念の歴史的変遷と心理測定学的観点からの考察 *教育心理学年報*, 51, 118-130. 10.5926/arepj.51.118
- 永谷文代・松寄順子・諏訪絵里子・上西裕之・谷池雅子・毛利育子 (2022). 教師記入式実行機能行動評定尺度の小学生に対する信頼性及び妥当性の検証 *心理学研究*, 92 (6), 554-563. 10.4992/jjpsy.92.20226
- 仲嶺真・上條菜美子 (2019). 「心理学研究」の新心理尺度作成論文に記載された尺度作成の必要性 *心理学研究*, 90 (2), 147-155.
- 岡田謙介 (2015). 心理学と心理測定における信頼性について *教育心理学年報*, 54, 71-83. 10.5926/arepj.54.71
- 小野由莉花・及川昌典・及川晴 (2021). 撤回：性的過大知覚バイアス *心理学研究*, 92 (2), 79-88. 10.4992/jjpsy.92.19042
- 小塩真司 (2020). 性格とは何か—より良く生きるための心

- 理学 中央公論新社
- 小山貴士・大久保街重 (2022). 顔と眼 心理学研究, 92 (6), 511-521. 10.4992/jjpsy.92.20011
- Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses, *Journal of Statistical Software*, 17 (5), 1-25.
- Samejima, F. (1997). Graded response model In W.J.van der Linden & R.K. Hambleton (Ed.), *Handbook of modern item response theory* (pp.85-100). NewYork : Springer.
- 芝祐順 (1979). 因子分析法 東京大学出版会
- 柴田利男 (1992). ボディ・イメージの心理学 繊維製品消費科学会誌, 33 (6), 280-284. 10.11419/senshoshi1960.33.280
- 清水和秋 (2007). a はやめて ω にしよう 日本心理学会大会 発表論文集, 71, 2PM049-2PM049. 10.4992/pacipa.71.0_2PM049
- 高本真寛・服部環 (2015). 国内の心理尺度作成論文における信頼性係数の利用動向 心理学評論, 58 (2), 220-235. 10.24602/sjpr.58.2-220
- Thurstone, L.L. (1928). Attitudes can be measured, *American journal of Sociology*, 33 (4), 529-554.
- 豊田秀樹 (2012). 項目反応理論 [入門編] 朝倉書店
- 豊田秀樹 (2020). 瀕死の統計学を救え! 朝倉書店
- 宇佐美慧 (2016). 測定・評価・研究法に関する研究の動向と展望 教育心理学年報, 55, 83-100. 10.5926/arepj.55.83
- van der Linden, W.J. and Hambleton, R.K. (1996). *Handbook of Modern Item Response Theory*. NewYork:Springer.
- 渡邊芳之 (2010). 性格とは何だったのか 新曜社
- 渡邊芳之・佐藤達哉 (1991). パーソナリティ概念を用いた行動説明にみられる方法論的問題点 人文科学論集, 25, 19-31.
- Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois:Northwestern University.
- 吉田寿夫・石井秀宗・南風原朝和 (2012). 尺度の作成・使用と妥当性の検討 教育心理学年報, 51, 213-217. 10.5926/arepj.51.213
- 吉田寿夫・村井潤一郎 (2021). 心理学的研究における重回帰分析の適用に関わる諸問題 心理学研究, 92 (3), 178-187. 10.4992/jjpsy.92.19226
- 吉野諒三 (2007). 公理的測定の基礎 吉野諒三・山岸侯彦・千野直仁 (編) 数理心理学 (pp.49・108). 培風館

注

- 1 <https://osf.io/afq5m/>
- 2 この論文はその後著者により取下げられている。