

連載

心理社会的要因の測定(2)

「心理特性II 妥当性」

産業医科大学産業医実務研修センター 堤 明純

連載「心理社会的要因の測定」の第2回は、心理尺度の妥当性について解説する。妥当性を表現する3つの考え方である内容的妥当性・基準関連妥当性・構成概念妥当性を抑えていただきたい。尺度の開発に携わらない限り、研究者が尺度の妥当性を実際に検討することはあまりない。しかし、妥当な測定がなされているか否かを基に、論文(研究)の評価や尺度の選択ができるようになることが望ましい。

1. 心理社会的要因の測定における妥当性の概念

測定の妥当性とは、測りたいものを測っているかどうかを示す概念で、測定の正確さ accuracy を表す概念のうち、系統誤差—真の値から常に一定方向に偏って生ずる誤差—に言及する。妥当性は測定の意義 meaningfulness を問う概念で、心理尺度の最も本質的な問題に関わる(図1)。

連載の第1回でも記したように、心理尺度を用いて測定される心理社会的な概念には参照すべき至適基準 gold standard がない。したがって、測定された概念が的を射ているかどうかという評価は定量的には困難であり、表1に挙げたような内容が質的に検討されることになる¹⁾。

内容妥当性 content validity とは、尺度に用いられる項目の内容が、それをを用いて結論しようとしている測定内容の良い見本となっているかを示す。

基準関連妥当性 criterion (-related) validity とは、測定しようとしている現象と関連のある外部変

数 external variable あるいは基準測度 criterion measure と測定値が相関する度合いをいい、併存妥当性 concurrent validity と予測妥当性 predictive validity がある。併存的妥当性は同時に測定した関連構成体の測定値との相関をいう。予測妥当性は、測定値が将来のアウトカムをよく予測するかを検証したものである。

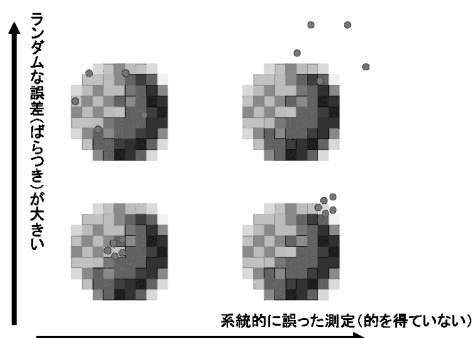
構成概念妥当性 construct validity とは、測定値が測定しようとしている現象の理論的な概念(構成概念)を反映していることで、測定用具としての尺度が持たなければならないもっとも重要な妥当性である。

2. 妥当性の検討

冒頭で述べたように、尺度の開発に携わらない限り、研究者や読者が自ら、尺度の妥当性を検討することは少ないが、測定が正確になされているか否かは、論文(研究)を評価する際に、常に重要なポイントとなる。

尺度開発の過程では、尺度を構成する各項目が、測定しようとしている内容の枠組みに属しているかどうか、測定しようとしている概念を包括的にとらえているか、偏りなく選ばれているか、という点について、専門家や関連の仕事に携わっている人たちの意見が求められる。以上が内容的妥当性の確認の手続きであり、一般に、複数の専門家による独立し

図1 測定の正確度



一貫した結果(信頼性のある測定)でも、的はずれなものはない(右下図)

表1 妥当性の分類

種類	内容
内容的妥当性	測定したい領域 domain の枠組みの中に、個々の項目が属するものか。
基準関連妥当性	問題としている特性と関連のある外部変数あるいは基準測度との間に期待される関連が観察されるか。併存的妥当性と、予測妥当性がある。
構成概念妥当性	その尺度が、あらかじめ計画された心理学的構成体をどの程度測定しているかということ。

た評価の一致をもって、内容的妥当性が保証される。

関連する妥当性に表面的妥当性 face validity がある。表面的妥当性は、直感的・印象的な感じを基に言及される妥当性で、厳密な科学的手段あるいは具体的手段にうたえて確かめられた妥当性ではない。しかし、とくに新しく開発されようとする尺度についてはないがしろにはできないもので、たとえば、「〇〇式方能適性テスト」など、中身の伴わない見かけ倒れの標題ではないかとか、測定をしようとしている概念と明らかに異なるものについての質問ではないかなどといったことは、専門家でなくても、論文の記述中からうかがい知れるものである。

基準関連妥当性の程度は、問題としている特性と関連のある外部変数あるいは基準測度と尺度得点を比較することにより判定される。厳密には、基準がゴールド・スタンダードであることをもって(のみ)基準関連妥当性と呼ばれることがある。たとえば、大うつ病のスクリーニング尺度を評価するときに精神科医の確定診断を基準として、その尺度の特性を評価するときなどである。しかし、多くは、類似の概念を測定する確立された尺度を同時に使用するなどして検討することが多い。ADL(日常生活動作)の尺度を構成するのであれば、同時に測定した併存疾患や障害の程度などとの間に有意な(負の)相関を有していたい(併存妥当性)。予測妥当性の確認には、一般に時間的な猶予(前向き研究の成果が待たれること)と複数の知見の集積を必要とする。多くの研究が「ある職業性ストレス尺度で測定されたストレスが高いグループは将来心筋梗塞を発症するリスクが高い」という所見を一貫して提示していれば、「これらの研究で使用されている職業性ストレス尺度は心筋梗塞の発症をよく予測する(予測妥当性)」という評価になる。

構成概念妥当性の検証は、測定値が理論的に正しい測定であるかの検証である。たとえば、年齢が上がるにつれ尺度得点が上昇するような概念であれば、測定値と年齢の間にそのような関係が見られない場合は正しい測定がされていない可能性がある。構成概念妥当性の検証には、構成概念の異なる特性と、異なる測定方法により得られる測定値間の相関行列である多特性・多方法の行列 multitrait-multimethod matrix における収束妥当性 convergent validity と弁別的妥当性 discriminant or divergent validity を用いた検証方法がある²⁾。

収束的妥当性は、同一の特性を異なる方法で測定した場合の相関で、同一特性内の相互相関は相対的に高くならなければならない。そうでなければ、異なる方法はそれぞれ別の構成概念を測定していると

考えられる。逆に、ひとつの構成概念は、他の構成概念とはっきり区別されて意味がある。弁別妥当性は、異なる特性を同一方法によって測定したとき、測定値間の相関が低いことを持って確認される。

今、自記式質問紙と面接法のふたつの方法を用いて、新規に開発した抑うつ尺度の妥当性を検討する(図2)。対比する特性として「活気」を取り上げる。自記式質問紙と面接法で評価される抑うつの間には高い相関が期待される(収束妥当性)。一方、抑うつと活気の間には、関連がないか、関連があっても負の相関があること予測され(弁別妥当性)、その関連の強さ(相関係数の絶対値)は同一の特性を測定する場合よりも小さい。このようなパターンが観察されない場合、開発した尺度の妥当性に疑問が生じることとなる。

因子妥当性 factorial validity は、尺度の持つ因子構造から推測される妥当性で、構成概念妥当性の一部として紹介されることが多い。因子分析の手法を用いて、尺度が当初から想定された概念通りの因子構造を有しているかを確認する。因子分析の詳細と解析の手続きについては成書にゆずる^{3,4)}が、その「手法」には、大きく分けて探索的因子分析と確証的因子分析がある。

探索的因子分析では、因子構造を規定せずに得られたデータに因子分析を適用し、データの特徴と項目間の関係(因子構造)が探究される。尺度を構成

図2 多特性・多方法の行列による構成概念妥当性の検討(想定例)

方法	特性	自記式質問紙		面接法	
		抑うつ	活気	抑うつ	活気
自記式質問紙	抑うつ	(.91)			
	活気	-.42	(.90)		
面接法	抑うつ	[.71]	-.19	(.88)	
	活気	-.20	[.69]	-.39	(.89)

ある特性を複数の方法で測定した場合、それぞれの方法で測定される特性間の関連の強さ(相関の絶対値)には、以下のような関係が観察されることが期待される:

同一特性を同一方法で測定した時の関連の強さ(再テスト法による信頼性係数: 図中()内)

> 同一特性を異なる方法で測定した時の関連の強さ(収束妥当性: 図中[]内)

> 異なる特性を同一方法で測定した時の関連の強さ(弁別妥当性: 図中アンダーライン)

> 異なる特性を異なる方法で測定した時の関連の強さ期待されるような関係が観察されない場合、正しい測定がなされていない可能性がある。実際にはこのような関係の一部を観察して妥当性を示していることが多い。

する項目は、相関の強い項目群が共通の因子としてまとまるように分解される（節約の原理）ので、分解された構造が了解可能なものであるかを検証する。たとえば、ソーシャルサポートを測定する尺度の項目が、手段の供与として分類される手段的サポートを測定する項目群と情緒的サポートを測定する項目群に分かれて抽出されるのであれば、（それなりに）妥当な測定がなされているといえる（情緒的サポートを測定することを期待している項目が、手段的サポートの因子として抽出されるようであれば、その項目が妥当なものか否か検討の余地がある）。もし、すべての項目が単一の共通部分だけからなると解釈されれば、その尺度は因子的に純粋な（等質性の高い）尺度と考えられる。現在では、多くの統計パッケージが探索的因子分析のコマンドを用意している。

確証的因子分析は、あらかじめいくつかの因子構造を設定して因子分析を実行し、モデルの当てはまり具合を確認するものである。研究者が想定している項目間の関係（因子構造）が、ほかに想定される因子間の関係（モデル）よりも当てはまりがよいという仮説を、実際のデータによって検証しようとするものである。

このほかの妥当性検証の例として、尺度の反応性や交差妥当性などが検討されることがある。

社会的な変動の影響や治療による臨床症状の改善など、とらえたい変化を尺度がとらえているかというのは、尺度の妥当性の一面である。たとえば、異動や部門の統廃合など経済不況に伴うリストラクチャリングに関連して過酷な就業状況の変化を経験した労働者ほど尺度得点に悪化が見られるとすると、その尺度は、組織のストレスフルな組織の変化をよくとらえているといえる（尺度の反応性）。

また、ある場面（対象）で確認された妥当性が、他の場面でも同様に成り立つかといった交差妥当性 cross validity が検討されることがある。具体的には、異なる対象でも十分な信頼性が得られ、基準となった対象と同様の外的基準との相関（基準関連妥当性）や因子構造（因子妥当性）が確認されることが多い。

表2 心理特性Ⅱ 妥当性のまとめ

測定の妥当性の検証とは、測定したい概念を測定できているか否かを確認することを示す。読者は、「その論文（研究）で捉えられておかなければならない概念が正しく測定されているか」という視点をもって論文を読むとよい。研究者（尺度のユーザー）は、研究課題に沿って妥当性が高い尺度を選択し、自らの測定の妥当性を確認するようにする。

3. まとめ

ある心理尺度が何を測定しているかを洗練させていくには、いろいろな概念を測定した値の間の相互関係が明らかになり、重複した概念が除かれ、あいまいな概念が区別されるといった多くの研究の蓄積が必要で、そうしてできた概念が「構成概念」とよばれる。完璧な構成概念を有する心理尺度は存在しないとも言える¹⁾。

尺度のユーザーは、その研究課題に合わせて、なるべく妥当性の確立した尺度を選択するとよい。そして、実際に調査を行ってからは、妥当な測定がなされているかどうか常に確認することが望ましい。具体的には、解析の過程で、尺度の測定値が他の指標との間に予測していたとおりの関連を有しているか否かを意識しておく。妥当性のみならず、測定の信頼性も含めて、測定の誤差とそれが所見に与える影響の方向と大きさは得られた所見を評価する際、貴重な考察の材料を与えてくれる。

文 献

- 1) 池田 央. 妥当性の諸概念. 心理学研究法 8 東京: 東大出版会, 1973; 177-206.
- 2) 池田 央. 測定と数量化. 行動科学の方法 東京: 東大出版会, 1971; 123-157.
- 3) 柳井晴夫, 繁榊算男, 前川眞一, 他. 因子分析. 東京: 朝倉書店, 1990.
- 4) Ferguson E, Cox T. Exploratory factor analysis: a user's guide. International Journal of Selection and Assessment 1993; 1 (2): 84-94.