

連載

心理社会的要因の測定(4) 「尺度の開発II 尺度の編集と標準化」

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4回では、第3回に引き続き、尺度開発の後半部分にあたる尺度の編集と標準化について述べる(図1)¹⁾。尺度を開発する場合も、使用する場合も、正確な測定を行うための手続きとポイントがあることを理解していただきたい。

1. 尺度全体の編集

1) 項目数と信頼性・妥当性の関係

項目数は多ければ多いほど信頼性と妥当性を上げることができる。また、妥当性の増加の割合は尺度の信頼性に依存する。尺度の妥当性を向上させようとするとき、尺度の信頼性が低ければ項目数を増やすことも意味があるが、すでに尺度の信頼性が一定

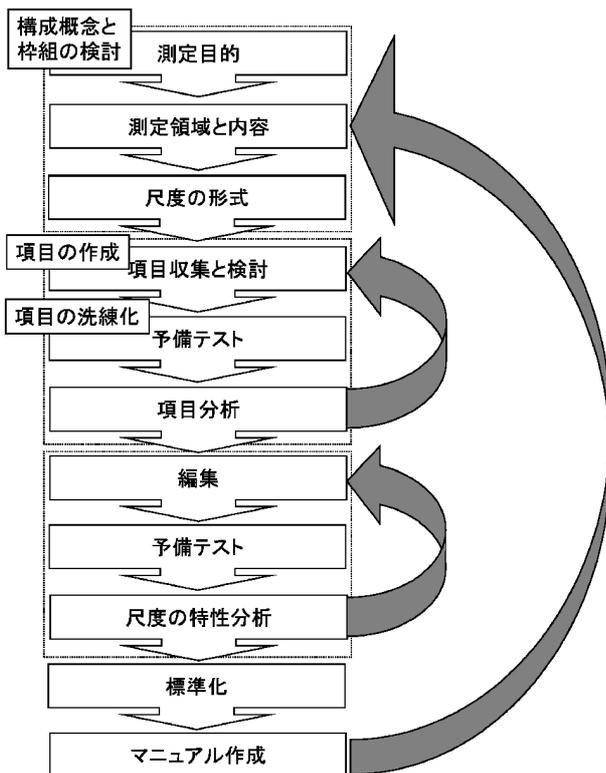
以上高いということであれば(たとえば、0.8以上)、項目を増やすことは、妥当性の向上を図るという意味では、ほとんど意味がない。項目数が多いと、被験者の負担も増えるし、回答のモチベーションにも影響する。ある程度の信頼性が維持できる範囲で、項目数が決まることになる。

一般にひとつの構成概念の内容的妥当性を補償するには、一尺度内にはある程度以上の項目数が必要である。また、ほんの2~3項目では、十分な信頼性を確保することは(数学的に)かなり難しい²⁾。一方で、自覚的健康観などは、ひとつの項目でも、予測妥当性が高いことがわかっている³⁾。

2) 尺度の目的に応じた項目の選択

少し難しい話になるが、項目の統計的な特性を活用することで、尺度の得点分布をどのような形にするか編集することも可能である(表1)。測定目的にふさわしい分布型の得点が得られるように、必要な特性を有する項目を収集して工夫される。

図1 尺度開発の手順



(文献1より改変)

表1 尺度の得点分布をどのような形にするか⁴⁾

尺度の目的	項目の選択
記述のための尺度	すべての能力水準においてまんべんなく識別力を高めるために、信頼性の高い項目で、項目の困難度をいろいろの水準に選択する(例:知能テスト)。
選抜のための尺度	特定の能力水準においてのみ、識別力を高めるように工夫された尺度。 能力の高い被験者の間においてできるだけ個人差を明らかにし、識別するような尺度を作りたいときは、困難度の高い項目を多用する。
判別のための尺度	判別の境界となる能力水準の前後において得点の変化が大きくなるように工夫する。 妥当性の高い項目で、判別の境界となる能力の水準において正答率が急激に変化する項目を集める(例:診断用テスト)。

2. 標準化とマニュアルの作成

1) 標準化 standardization

尺度の得点を正しく評定するための基準をその尺度が対象とする集団の中で設定していくことを標準化といい、このような基準によって作成された尺度を標準尺度という⁶⁾。標準化の意義は、評定の基準を定めることにある。

標準得点には、内容的に標準化された得点 (content standard score) と基準集団による標準得点 (normative standard score) の2種類がある。前者は、尺度項目を設定する基準が明白で (すなわち、項目の困難度や識別度が確定しており) 得点そのものに意味がある。そのために、尺度を適用した対象、時期、地域を問わず尺度得点の比較が可能となる。TOEFLの得点などがその例である。このような標準化には、項目反応理論が応用されている。

基準集団による得点の標準化は、基準集団の得点分布の中で相対的に評定される得点を定めることで、古典的テスト理論に基づいて、一般的に行われてきた手続きである。尺度の要約統計量 (表2) は、被験者・状況・尺度の種類、の三つの要素の基に検

表2 標準化の過程で明示されるべき要約統計量の例

被験者数
平均, 分散, 標準偏差
歪度, 尖度
最高点, 最低点, 中央値
第1四分位, 第3四分位, 範囲, 四分位範囲
最頻値

討される。尺度の目的に応じて最も有効な情報を与えてくれる集団を設定すること (基準集団の設定) が、尺度の妥当性を確定し、評価の基準を明確にする上で非常に重要となる。年齢・性別・地域差・社会経済的状况などを考慮した標本抽出が求められる。パーセンタイルは、分布表現としてよく用いられる。また、平均と標準偏差に基づく標準得点は、心理領域でよく用いられる。平均0, 標準偏差1とした標準得点がよく求められ (平均50, SD10と換算したものが、いわゆる偏差値である)、同一集団内の各個人の相対的位置関係を見ることができる (表3)。必要に応じて、標準得点に基づいて、素点から解釈できる数値 (もしくはカテゴリ) に割り当てる換算表が作成される。

2) 採点法: 項目得点の和を尺度得点とすることについて

ここで、尺度の得点化について、コメントを加える。多くの尺度が、4件法や5件法で回答された項目に与えられた点数の和で得点化されているが、この手続きには項目群が等質 (一次元) であるという前提がある。すなわち、項目が等質であれば各項目の和を尺度得点 (代表値) としてよいが、等質でなければ和を尺度得点としてはいけない (意味の異なるものを加算しても意味がない)。等質性 (一次元性) をもっている限り、各部分得点 (項目得点) を合計しても意味があり、ここに等質な尺度を開発する意義がある (図2)⁷⁾。

次に得点の取り扱いであるが、上記のようにして算出された得点の水準は、高々順序尺度の水準であり、このようなデータで平均や標準偏差を求めた

表3 標準化の指標としてのパーセンタイルと平均と標準偏差に基づく標準得点の意義

	パーセンタイル	平均と標準偏差に基づく標準得点
算出	対象とする数値群を小さい順にソートし、指定された個数番目にある値を代表値とするもの	$z_i = (y_i - \bar{y}) / s_y$ $y_i: \text{被験者 } i \text{ の尺度得点, } \bar{y}: \text{平均点, } s_y: \text{標準偏差}$ 平均0, 標準偏差1 $Z_i = z_i * 10 + 50 \text{ (M = 50, SD = 10)}$
長所	分布が非常に違っている得点を比較できる 同一の集団では、形式の異なるふたつの尺度得点と同じパーセンタイル順位で対応づけられる 得点分布が正規分布に近似している場合には標準得点と一定の関係を持つ	得点が規準集団の平均から何σ離れているかを示すと、得点分布上の位置がある程度明らかになる 同一集団の中で、どの程度の能力 (特性) があるのかが分かる
短所	得点の順位関係のみによって決まる序数であるから加算や平均はできない	疫学領域ではあまり使用しない 分布型と無関係にいつも適切な比較ができるわけではない (上・下限ができる) 異なる集団でそれぞれ標準化したものを比較しても無意味

図2 項目が等質であるとき、各項目の得点の和を取る意味がある

以下のような質問に、「はい」か「いいえ」で回答する尺度について、質問項目と測定対象が一次的に対応しているときの回答パターン（はいに1点、いいえに0点を与える）：

- q(1) あなたの身長は155 cm 以上ですか
- q(2) あなたの身長は160 cm 以上ですか
- q(3) あなたの身長は165 cm 以上ですか
- q(4) あなたの身長は170 cm 以上ですか

回答者の身長	回 答								尺度得点
	いいえ				はい				
	q(1)	q(2)	q(3)	q(4)	q(1)	q(2)	q(3)	q(4)	
152 cm	○	○	○	○					0
156 cm		○	○	○	○				1
161 cm			○	○	○	○			2
168 cm				○	○	○	○		3
173 cm					○	○	○	○	4

尺度得点は、回答者の身長を表現している（文献7より改変）

り、因子分析を適用したりすることは、厳密には正しい処理ではない。しかし、多くの場合、間隔尺度に近似させて取扱っていることは認識しておいてよい。

合成得点の算出法については、項目毎の重みづけも議論になることがあるが、以下のような要件を満たせば、合成得点に対する重みづけの影響は小さく、単純和による合計得点と事実上変わらないとされる；a. 項目数がある程度以上多いこと（それでも10以上といわれる）、b. 項目分析の結果不適当な項目が除かれ各要素得点の相関が高いこと、c. 重みが正（採点法が他項目と逆転する項目がない）、d. 平均に比べてばらつきの程度が小さい。とくに項目数の多い尺度で各項目得点を合成して尺度得点を定義する場合は、各項目に与える重みの違いによる影響は小さいとされる。同様のことは、回答フォーマットの違いにもあてはまり、各項目の特定の回答に1とか0の単純な数値あるいは1, 2, 3, 4等の単純な整数値を与える方法と、複雑な手続きを経て決定された数値を用いる方法で結果的に大きな差異はないことがわかっている⁸⁾。

3) マニュアル

尺度が開発されると、その目的と対象、採点・結果処理を含む実施方法、標準化の過程で算出した代表値、項目分析を経て明らかになった尺度の特性（妥当性と信頼性）を明示したマニュアル（実施手引）が作成されることになる。

表4 マニュアルに掲載される内容

項 目	内 容
名称 形式 出版（年）	著者（開発スタッフ）や版（バージョン）も示す
目的	測定する特性
対象	年齢・性など、尺度が適用される属性を明らかにする
開発過程	開発過程に関する簡単な記述
実施方法	標準化の際に用いられた条件を明記 検査者の資格 テスト時間 利用上の注意（インストラクション）
採点	採点法
結果処理	換算法 欠損値の処理方法
標準化	尺度得点の結果（数値）を解釈するための基準を明らかにする 代表値の提示 尺度の限界（例：…の対象では、信頼性が不十分、といった情報など）を挙げる
測定結果の解釈と評定	測定結果の解釈 評定のために用いられるべき尺度得点の分析や、尺度以外の被験者に関する情報（属性）を示し、これらの解釈や評定の仕方を述べる
妥当性と信頼性	代表的な妥当性に関する具体的データ（下位尺度別） 項目分析（因子分析）の結果 回答に関するバイアスに関する情報、得点誤差の発生要因とその程度、反応性 検討対象を明示したうえで信頼性係数を提示
参考文献	

尺度を開発したらマニュアルを作成し、ユーザーに正しく使用してもらうようにする。多くのユーザーに尺度を利用して得られた情報をフィードバックしてもらい尺度の改良に利用するとよい。逆に、尺度を使用する側では、マニュアルに当たって、尺度の特性をよく理解したうえで使用することが基本である。尺度の適応や手続きを誤ると正確な測定はできない。

表4にマニュアルに掲載する内容の例を挙げる⁷⁾。

3. まとめ

適切な集団において尺度得点の標準化が行われ、

表5 標準化 まとめ

尺度得点を解釈するための基準を明らかにして、その数値の利用の仕方を説明するのが、標準化の目的である。

一般的な標準化作業（基準集団による得点の標準化）は、ある基準集団における相対的位置の確定により尺度得点を評定する。

基準に取る集団が異なる場合—より厳密には、基準（母）集団から、それぞれ無作為抽出されたような集団間でなければ—尺度得点の比較は困難なものである。

テストを開発したらマニュアルを作成すること；逆に使用するときにはこれに当たり、尺度の特性をよく理解したうえで使用する。マニュアルが作成されていない尺度はあやしい。

マニュアルが作成される。尺度を使用しようとする研究者は、出版されているマニュアルを参照し、その適応や使用方法について調べておく（表5）。

文 献

- 1) 池田 央, 芝 祐順. テスト法の意義. 肥田野 直編. テスト1, 心理学研究法7 東京: 東大出版会, 1972; 1-29.
- 2) Kim JO, Muller CW. Factor analysis. Statistical methods and practical issues. Beverly Hills: Sage, 1978.
- 3) Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. J Health Soc Behav 1997; 38(1): 21-37.
- 4) 芝 祐順. テストの編集. 肥田野 直編. テスト1, 心理学研究法7 東京: 東大出版会, 1972; 93-106.
- 5) 藤田恵壘. 標準化. 肥田野 直編. テスト1, 心理学研究法7 東京: 東大出版会, 1972; 135-179.
- 6) 池田 央. 測定と数量化. 池田 央. 行動科学の方法. 東京: 東大出版会, 1971; 123-157.
- 7) 池田 央. テストの作成. 池田 央. テスト2, 心理学研究法8 東京: 東大出版会, 1973; 237-284.