

## 連載

## 心理社会的要因の測定(3) 「尺度の開発 I 手順と項目分析」

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(3)と(4)では、あるひとつの構成概念を測定しようとする心理尺度を想定して、尺度の開発手順の概略を紹介する。前半の(3)で、尺度開発の手順と項目分析について、後半の(4)で、尺度の編集と外国で開発された尺度の利用に関して述べる。実際の調査票の開発手順を通して心理尺度の特性についての理解を深めることが目的である。

### 1. 尺度開発の手順

尺度開発には複数の手続きが関わり、開発方法は必ずしも一様ではない<sup>1)</sup>。図1に、尺度開発の標準的な手順を示す<sup>2)</sup>。

#### 1) 測定の目的の確定

まずは、測定の目的を確定する。すなわち、何を測ろうとするのか、構成概念を明確に思い描けるようにする。尺度の要件のひとつである妥当性はこの段階に決定的に依存している。

#### 2) 測定領域と内容

次に、測定の内容領域 domain を設定する。測定の範囲を確定するとともに、その内部の下位領域も定める。このプロセスは、開発しようとする尺度の内容的妥当性を評価する枠組みを決定することに相当する。たとえば、健康に関連する包括的なQOLの尺度を開発しようとする場合、身体的機能にとどまらず、精神的な健康状態も含む下位領域を設定するような手続きが取られる。被験者の侮りや不信、不快の念を抱かせるものは避ける、といった表面的妥当性にも留意する。

#### 3) 尺度の形式の決定

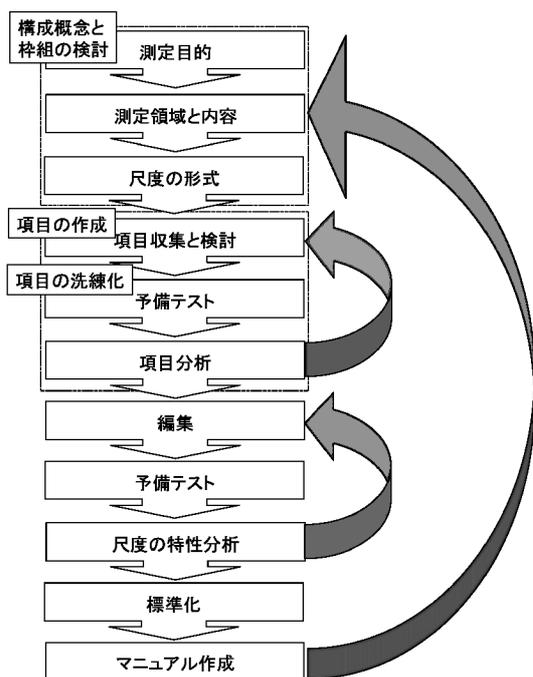
どのような形式で測定を行うのか、測定の形式 format の決定を行う。連載(1)で述べたように、調査方法には、質問紙法、面接法、投影法などがあり、最近はオンラインでの調査も行われている。採点処理の手順なども定めておく。

#### 4) 項目の作成：項目の収集と検討

このプロセスは、一般に、項目候補の収集とその検討の二段階からなる。まずは、項目の候補となる記述を、選別は後回しにして、できるだけ多く、できるだけバリエーション豊かに収集し、項目のワーキングサンプルを作成する。各側面について、取りこぼしなく項目をそろえることにより内容的妥当性の保証につながる。収集方法は、多くの関連文献に当たるとともに、自分で考えるだけでなく、同僚や調査対象者へのインタビューを利用するなどすると実りが多い。

続いて、項目の重複や、ワーディング、項目の受け入れやすさなどに留意して、収集段階では保留していた項目候補の検討を行う。尺度の内容を充足する包括性に留意しつつ、重複した項目は取捨を行い、不快な印象を与える項目は削除する。必要に応じてワーディングを修正して分かりやすく、誤解のない表現にする。項目収集の段階同様、同僚（研究者間）による少人数のグループにおける作業や、対象者によるチェックも活用する。この時点での情報収集方法は、必ずしも体系的な手法をとる必要はなく、これ以上新しい情報が得られないという段階ま

図1 尺度開発の手順



(文献2より改変)

で情報を集める (rule of redundancy; stop at no new information!))。

ここで、尺度開発のひとつの要素であるワーディングについて、少し解説する。

項目を記述するに当たっては、いくつかの約束事がある。言わずもがなであるが、わかりやすい用語を使用する。地域住民を対象とした調査であれば、「悪性新生物」というより「がん」というほうが、住民にとっては分かりやすい。また、「あなたの彼は強くてやさしいですか?」といった、複数の問いかけをするような質問にならないようにする。強いけど優しくなければ、何と答えてよいか分からない (yes-no では答えられない)。頻度に関する問いはあいまいにならないようにすることが望ましいとされる。たとえば、「しばしば」という副詞に対して有するイメージは被験者により異なる。「週に1回」などといった明確な指示が望ましい。このほか、無理強いをするような質問や誘導的な質問にならないように留意する。前者の例として、収入や性的なことがらを訪ねる場合などは、それなりの配慮が必要である。質問票であれば、最後に配置する、面接調査では、ラポールがとれてから尋ねる、といった尋ねる順番に配慮することもある。

ワーディングは、研究者自身で編集する 경우가多いが、同僚や予備テストの対象者に「答えにくい質問はないか」などと尋ねてもよい。ワーディングは、適宜改良を重ねていく。ただし、尺度の心理特性を確認する最終のテスト段階前に完了しておくことが求められる。古典的テスト理論に基づく尺度開発では、ワーディングが変わると、厳密には、テストをやり直さなければ心理特性が確定できないからである。

#### 5) 項目の洗練化：予備テストと項目分析

被験者の応答によりあらわれる項目得点の統計的性質を記述したり、諸性質の間の関係を分析したりして、項目の作成や改良、および尺度の編集に必要な情報を得る手続きを項目分析という<sup>3)</sup>。

本稿では、ひとつの構成概念を測定しようとする心理尺度を想定している。一般に、このような尺度を構成する場合、収集した項目ワーキングサンプルの中から、条件に合わない項目を削り、一貫性 (等質性) の高い項目群に集約していく作業をいう。言うまでもないことだが、留意したいのは、このプロセスの目的は被験者を検査することではなく、項目を検査することである。

項目分析を行うための予備テストの要件を表1に挙げる。調査対象は、できるだけ開発しようとする尺度が適用される集団を選択する。開発初期の段階

表1 項目分析のための (理想的な) 予備テストの要件<sup>3)</sup>

標本	その尺度が完成して用いられるときの対象者の母集団を考慮する。
標本数	400くらいあると十分。
実施条件	テスト時間は十分に長くし、ほとんどの被験者がすべての項目をやり終えられるように配慮する。 項目の配列順序の異なったバージョンをいくつか用意するなどして、やりのこしが同じ特定の項目だけに偏らないよう工夫する。

では、この条件の充足は困難なこともあるが、小児用の尺度を作成するのに成人でテストをするのはおかしい。標本数は、400ほどあると統計学的取り扱いに好都合だと言われるが<sup>3)</sup>、必ずしも、充足できるとは限らない。たとえば、因子分析では、一般にサンプル数が多くなるほど、安定した解が得られるが<sup>4)</sup>、因子分析に必要なサンプル数の算出方法を示した研究もあり、興味のある研究者は参照されたい<sup>5)</sup>。

項目分析に用いる予備テストのテスト時間は十分に長くってよく、ほとんどの被験者がすべての項目をやり終えるように配慮する。項目分析は、不適当な項目を削る作業でもある。除外する項目の要件には、答えにくさなども含まれる。回答しにくい項目は、無回答が多いことであたりをつけることができる。一方で、時間的な制約があると、後半に回答されない項目が集中する可能性がある。やりのこしが特定の順番の項目だけに偏らないように、項目の配列順序が異なるバージョンをいくつか用意するなどの工夫をすることがある。

項目分析に用いる統計学的手続きと項目統計量の例を表2に挙げる<sup>3)</sup>。

## 2. 項目の選択と改良

### 1) 尺度の信頼性を高めるための項目の選択

実際に尺度を開発する時には、尺度の信頼性を左右する項目の等質性、いわゆる内的整合性を高めることが大きな目標となる。そのためには、まず、互いに相関関係の高い項目を選び出すことが必要になる。

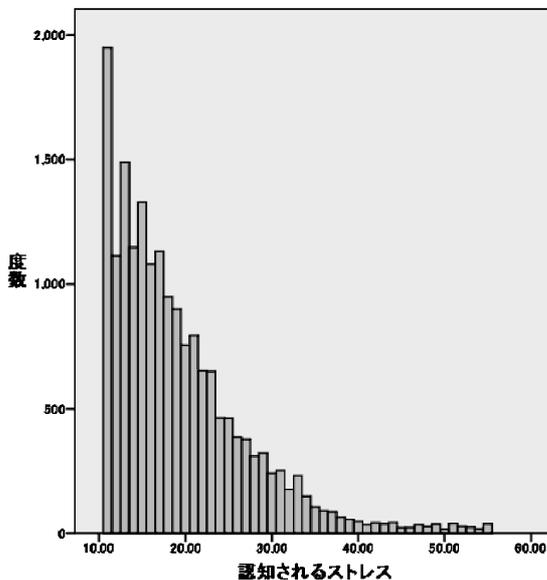
項目と尺度の間の相関は、その項目が測定しようとする対象と、尺度全体の測定しようとする対象との類似性を保証するための手続きで、項目の妥当性を統計的に確認できる数少ない手法のひとつである。相関が小さいことは、削除の基準のひとつとなる。

1組の項目群は互いに様々な相関係数によって関係しあっている。これを総合的に分析し、項目間の

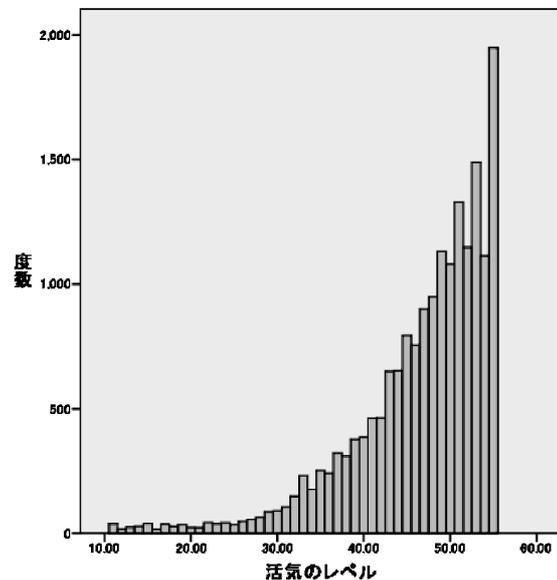
表2 項目統計量 (抜粋)<sup>3)</sup>

種類	内容	意義
項目困難度	<p>被験者が全体としてその項目にどう応答したかを示すもの：</p> $p_j = \frac{1}{N} \sum_{i=1}^N u_{ij}$ <p><math>u_{ij}</math> : <math>i</math> 番目の被験者が <math>j</math> 番目の項目において得た項目の得点  <math>N</math> : 被験者の人数</p>	<p>尺度得点の平均をどの程度にするかは、適当な困難度の項目を選ぶことによって調節することができる。正答率・項目の通過率・項目得点の平均、などとも呼ばれる。</p>
項目得点の分散	<p><math>\sigma_j^2 = p_j(1-p_j)</math>  項目困難度だけによって定まる (1/2 のとき最大)。</p>	<p>尺度得点の分散は、項目得点の分散と項目間の相関係数によって決まる。</p>
項目間相関係数	<p><math>\phi</math> (ファイ) 係数 : 1-0 データの場合。  四分相関係数 : 仮定された 2 変量正規分布をする 2 変量の相関係数を、ふたつの項目のそれぞれの分割点によって分けられた 1 と 0 との応答の比率から推定。</p>	<p>項目間相関の増加により、尺度得点の分散が大きくなり、尺度の信頼性が高められる。  四分相関係数を用いると、項目困難度と関係なく、項目間の相関係数を表すことができる。</p>
項目と尺度の間の相関	<p>項目の得点と全体の得点との相関係数を検討する。Pearson の積率相関係数が使用可能。  項目が 1-0 データの時は、点双列相関係数を計算できる。</p>	<p>項目得点とテスト得点との類似性を確認する手続きは、尺度全体で測定しようとする対象と、その項目が測定しようとする対象との類似性を保証するための手続き。</p>

図2 項目統計量の応用：項目困難度



困難度の高い項目を集めて、高得点を取りにくい尺度を作成すると、高得点のグループをよく識別する尺度を開発できる。選抜試験などに应用される



困難度の低い項目を集めると、一定水準に到達しないグループをよく弁別できる尺度を開発できる。資格試験などに应用される

統計的構造を明らかにする方法に因子分析がある。項目間因子分析の結果得られた主因子解の第1因子負荷の高い項目群ほど等質である。因子負荷の低い項目を削除することにより、等質性の向上とともに

尺度の因子妥当性が保証されることにもなる。

項目選択後は、同様の分析を繰り返し、選ばれた項目の等質性を再確認する。

表3 尺度の開発と項目分析のまとめ

尺度開発の手順は、測定目的、測定領域と内容、尺度と項目の形式の決定を経て、適宜テストを繰り返しながら、項目分析、編集、尺度の特性分析を行い、標準化していく過程を含む。

開発の出発点は、構成概念の設定である。

主に尺度の等質性（内的整合性）を高める目的で、不要な項目を落として研ぎ澄ませていく。

項目は、ワーディングにも気を配って作成する。

## 2) 尺度の妥当性を高めるための項目の選択

尺度の基準関連妥当性を高めるために項目の選択が検討されることもある<sup>6)</sup>。外部基準との相関を高めようとする作業は、数学的には信頼性係数を高めようとする作業と背反する<sup>3)</sup>もので、多数の被験者について確定できる外部基準が少ないこともあり、診断用性格検査等の尺度の開発を行う機会などに限られている。

社会的に望ましいとされる方向へと偏る回答者の反応性バイアスを測定する尺度得点<sup>7)</sup>と高い相関を有する項目を削除して、このような反応バイアスを発生させにくい尺度を開発することもある。

## 3) 尺度の目的に応じた項目の選択：項目困難度の応用

項目困難度は、被験者が全体としてその項目にどう応答したかを示すものである。正答率は、被験者の能力分布に依存するが、困難度は、被験者の能力分布とは独立に与えられる概念で、特定の標本に対してではなく、項目の特性を一般的に表す概念である。項目困難度は、尺度全体の難易度を設定する際に応用できる。全体に難しい尺度を作成したけれ

ば、困難度の高い項目を選ぶことで高得点をとる被験者が減少する。有害反応のスクリーニングなどに適用される。一方で、資格試験などでは、ある一定のレベルをクリアできていればよく、困難度の低い項目を集めることで、点数の低いグループを弁別する尺度を作成することができる（図2）。

## 3. まとめ

尺度は、いくつかのプロセスを経て開発され、その出発点は構成概念の決定である。実際のテストを通して、被験者の反応により得られる項目の統計学的な性質を利用して、尺度を洗練させていく。互いに相関の高い項目を選ぶことにより、尺度得点の分散が大きくなり、尺度の信頼性が高められる。

## 文 献

- 1) Guyatte GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. *CMAJ* 1986; 134: 889-895.
- 2) 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.
- 3) 芝 祐順. 項目分析. 肥田野 直編. テスト1, 心理学研究法7. 東京: 東大出版会, 1972; 53-91.
- 4) Ferguson E, Cox T. Exploratory factor analysis: a user's guide. *International Journal of Selection and Assessment* 1993; 1(2): 84-94.
- 5) Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. *Psychological Bulletin* 1988; 103: 265-275.
- 6) 池田 央. テストの作成. 心理学研究法8. 東京: 東大出版会, 1973; 237-284.
- 7) 北村俊則, 鈴木忠治. 日本語版 Social Desirability Scale について. *社会精神医学* 1986; 9(2): 173-180.