

Title	QOLD 評価測定尺度に関する基礎的研究(V) : QOL 評価測定尺度の信頼性と妥当性
Author(s)	丸山, 久美子
Citation	聖学院大学論叢, 15(1): 75-87
URL	http://serve.seigakuin-univ.ac.jp/rep/modules/xoonips/detail.php?item_id=196
Rights	

聖学院学術情報発信システム : SERVE

SEigakuin Repository for academic archiVE

QOLD評価測定尺度に関する基礎的研究 (V)

——QOL評価測定尺度の信頼性と妥当性——

丸 山 久美子

A Study of the Scaling of Life and Death (V)

——On the Problems of Reliability and Validity——

Kumiko MARUYAMA

Quality of life (QOL) is an ill-defined term. The World Health Organization (WHO, 1948) had declared good health to be a state of complete physical, social, mental, and spiritual well-being, not merely the absence of disease. Recently, WHO has placed special emphasis on spirituality. However, since an attempt is being made to measure something that is ill-defined, unobservable, and latent of QOL scaling, it can only be inferred that the instrument is valid insofar as it correlates spirituality with other observable behavior. The validation process in this instance consists of a number of stages, in which it is hoped that convincing evidence is collected where the instrument taps into the intended constructs and produces useful measurements accurately reflecting patients' QOL scaling.

Naturally, reliability, validity, and sensitivity (responsiveness) are interrelated, yet each is independently important in its own right. The assessment of reliability consists of determining that a particular scale or measurement yields reproducible and internally consistent results. Reliability is based upon the analysis of correlations between repeated measurements, where the measurements are either repeated over time, through test-retest reliability, Bayesian probability, observers, or by different variants of the instruments.

A number of different measures have been proposed herein and the associated problems discussed from psychological and statistical points of view.

Key words; Test-Retest Reliability, Validity, Spirituality, Internal Consistency, QOL Scaling

は じ め に

国家が、当面、国民の生活水準を測定するために用いる手法は「各個人の経済的生活の満足度」とも言うべき指標「QOL（Quantity of Life）」を作成する事である。つまり、人々の生活に関わる所得、消費財の所有数などを測定して彼等がどれ程生活に満足しているかを知る事である。しかし、今日、民族紛争などの小規模の戦争やテロの多発する地域以外では経済が豊かになり、食物が豊富で平和が約束されてくるようになると、人々は、必ずしも物資が容易に手に入るだけでは生活の満足を得ることが出来なくなった。人間は衣食住が足りるだけで満足する存在ではない。幸福の度合いは物質的量の大小だけでは保証されないのである。

従って、20世紀後半からQOLの概念は「Quality of Life」に変わり、生活の質の問題が議論されるようになった。単に物資だけではなく「心の安らぎ」や「精神的充足」、「生きがい」など、人の心模様が取り沙汰される時代になったのである。

近年、高齢化社会を迎えるに際して、困難な疾病対策も然る事ながら、高齢者の生活の満足度が医療関連分野に導入されるようになった。この傾向はこれまでの治療第1主義から患者第1主義への移行を促進させ、QOL評価測定尺度がこの分野に浸透し取り込まれ、WHOは世界万国に共通の普遍的なQOL評価測定尺度の作成を模索するようになった。

しかし、身体的痛みや社会経済的状況、人間関係の有様は民族特有のパターンがあり、人類共通の普遍的な尺度でそれらを測定することは困難である。民族はそれぞれの価値観を持ち、生活習慣や言語体系、宗教の種類が混在している。

「身体的痛み」に関しても「痛み」の感じ方さえ千差万別であり、「痛み」を測定する尺度を構成することは無意味であると主張する研究者さえ存在する。特に医療においては、病人の痛みを一つの物差しで測ることを好まず、「症例」毎に痛みの塩梅を目測し、治療することが本来の医療であると言われている。

従って病人の数だけ「痛み」や「苦悩」が存在する。症例研究が重視され、統計的測定尺度が敬遠される所以である。だが、異なる症例の数だけを追っていると、全体の病状が見えにくく、病名の判断が独善的になる可能性がある。WHOが率先してQOL評価測定尺度の活用を喧伝する事情は、出来るだけ客観的に疾病の全体像を把握する事や個々の症例を全体的に見つめ直す上で重要な主観的測定尺度の客観性を保証するためである。かくして、患者の主観的満足度が総合的相対的にどの程度のものであるかを想定するために普遍的な測定尺度の開発が必要になった。QOL評価測定尺度が医療・福祉関連分野に浸透したのは以上のような理由による。

評価測定尺度の信頼性・妥当性に関する研究の系譜

評価測定尺度法は計量心理学 (Psychometrics) におけるテスト理論の構成から始まった。より正確に言えば、人間の心 (行動) を測定する場合の物差し (Scale, Measure) の作成であり、その基本的主題は主観的な人間の心を一本の物差しで客観的に測定するための方法論にある。つまり、測定評価法とはある対象を客観的に測定するに際して用いられる尺度 (Scale) を操作的定義によって考える事である。多くは官能検査、人格テスト、知能テスト、適正検査、社会調査における測定評価尺度、あるいは分類カテゴリーの作り方に論点が絞られ、研究されてきた。測定されたデータの尺度の単位の原点 (基準点) が様々に変わるのであれば種々のデータを比較する上で、不便であり、容易にデータを比較考察することができない。そこで、一般性のある同一の測定尺度や基準点を持つ尺度を考える必要がある。この場合、古典的には「比較判断の法則」(Thurston, L. L. 1927) の考えを踏襲することが多い。それによれば、測定対象は特定の属性について、ガウス分布 (正規分布) を仮定し、測定評価尺度の基準点は $\pm \sigma$, $\pm 2 \sigma$, $\pm 3 \sigma$ となる。つまり、ガウス分布に従うとき、平均値 0 を中心とする $\pm 1 \sigma \sim 3 \sigma$ の範囲にそれぞれの度数が含まれていると仮定するのである。一般の分布では、平均より数えて $\pm 1 \sigma$ の 3 倍以上に及ぶものは極めてまれであり、もっとも度数が少ない。ある度数分布が間隔尺度⁽¹⁾として与えられたものと考え、分布得点を線形変換 (正規変換) するのであるが、こうした考えに底在するのは、感覚事象は無数の類似の微小事象からなるという前提がある。すると、極限中心定理が当てはまり、変数が無限に増大するに従って、ガウス分布に収束すると考える。又、ガウス分布を仮定すれば、変数の操作が簡単であると言う実際的理由もある。ところが、必ずしも、分類尺度基準点が $\pm 1 \sigma \sim 3 \sigma$ に存在するかどうかわからない場合が多々ある。つまり、評定者が n 個の対象をガウス分布による基準点、ある種の分類の基準点、カテゴリーの境界点に分類する時、個人によっては変動する場合がある。これは、個人の反応を反復繰り返し評定させると、分類の基準点が評定点毎にずれるという結果によるものである。このような評定分類における誤差変動の問題は専ら統計学上の問題として処理され、心理学的測定法では、余り関心の及ぶところではなく、結果の検定は χ^2 検定、尺度項目関連係数 a などを算出して処理する事が多い。分類の基準点の動揺が、その測定法への不信に繋がると考えれば、単に基準点が個人の評定毎に動揺する理由を偶然誤差が原因であるとして処理するだけではすまされない。そこで、何らかの一連の恒常誤差が作用していると考える事も出来る。このような問題意識は、既に分類カテゴリーの信頼性の研究として取り上げられている (林, 1967, 丸山, 林, 1968, 田中, 1969)。

特に分類評定尺度の誤差処理に関する研究は反応 (回答) の選択肢との関係で論じられることが多い。これらの研究の論点は同じ尺度で測定しても反応の状態が安定しているのは、どのような評

価尺度基準点を持つ尺度であるのか、又は、評定項目尺度の反応選択肢の数がどれ位であれば、繰り返し測定において、反応の変動が起こらないのかという事である(林, 1967, 丸山, 林, 1968, Maruyama, K. 1968, 1971, 丸山, 1971, 鈴木, 1968)。

Osgood等(1956)のSD法(Semantic Differential Method)における尺度分類について、中点評定に変動の原因があるとする研究がある。(Maruyama, 1968, 1971, 丸山, 1971, 鈴木, 1968)。男女40名の評定者に11色に関する3色配色の組合せ165枚の図形の中からランダムに選んだ25枚の3色配色の図形にSD評定尺度で3回にわたり繰り返し測定する。尺度の段階数(基準点)は $7 \cdot 6 \cdot 5 \cdot 3 \cdot 2$ である。参考までに表1に結果を示す。それによれば、繰り返し測定の単純一致率では7段階評定(中点を含む)は5分類、3分類に詰め、6段階評定(中点を含まない)は2分類に詰めた場合の平均値である。それによると、6段階2分類の単純一致率が最も信頼性の高い値を示していることが分かる。つまり、中点を含まない1、又は0の反応の変動は低いという事になる。中点を含む3分類と中点を含まない2分類の平均値の差は10%以上である。反応の選択肢の数と尺度分類の問題が反応傾向に一つの方向性を与え、それが反応変動の決定要因になる可能性があることの事例である。

しかし、質問項目の性質によっては、必ずしも中点評定を設けても、変動の要因にならず、反って反応の安定性を示すという研究もある(鈴木, 1968)。その意味において、中点評定は反応の変動性と安定性の両面を持っており、問題の性質によっては、どちらに重点をおくかを十分に見極める必要がある。

さらに、あるテストや調査、実験に関する評価尺度分類において、その質問尺度項目分類における回答選択肢のおき方が、果たしてそれらの実験やテストの目的から見て妥当なものであるかどうかを考慮する必要がある。当該現象において、測定尺度基準点が個人にとって、種々の意味を持っていることへの配慮が払われるべきであろう。例えば、前述のSD測定尺度では尺度の基準点が7段階であるとしたとき、それぞれの段階点は任意にガウス分布を仮定し、1点、2点、3点……7点とし、中点が4点、あるいは0を中点として左右に ± 3 点を与えて結果を処理するのが常套である。所謂、点と点の間は等間隔である。ところが、必ずしも中間点とその測定尺度の中間に位置していない場合がある。両端に示される形容詞対語(明るい—暗い)が対語として相関係数 -1.0 であるかどうかは分からない。多くは常識的に明るいと言う形容詞の対語は暗いと言う形容詞であると推定して測定尺度を構成する。しかし、完全に逆相関を示す形容詞対語はSD法においては存在しない。せいぜい $\rho = -0.8$ (64%の説明力)近傍の逆相関を示す。残りの36%は誤差として蓄積される。従って、測定尺度の基準点が等間隔に並び、形容詞対語が完全に逆相関になるために、「明るい」という形容詞の対語は反対語の「明るくない」とする方が妥当であるとする研究がある。いずれにしても、SD尺度点の中間点は必ずしも中間点を示すわけではなく、Messick(1957)によれば、SD測定尺度の0点は0点から0.2程離れているという。評定者から見れば、そのような曖昧な

測定尺度で、当該現象を測定するのであるから、その反応状態も極めて多義的であり、反応誤差の大きいものであって現象の真実を伝えるものではない。しかも心理学的測定値には評定者の不確実性要因も加味されており測定尺度が本来持っている恒常誤差（丸山，1981）との関係から論ずるのが妥当である。

評定尺度法に於ける個人の反応の不確実性を含む尺度項目の反応状況から、その確率的変動を考慮し、反応誤差モデルを考えてみよう。

以下に、A：ベイズ確率モデル、B：反応の一致率モデルを取り上げる。

A：個人の反応誤差におけるベイズ確率モデル

ベイズ確率とは事前確率と事後確率の比率である。つまり、今からある調査をしよう。調査をするとき事前にある回答が出る確率を知っているとす。テスト—再テストによる結果を事前確率とすれば、3回目のテスト結果において事前確率を重みとして計算すれば、より一層調査の真実に迫ることができると考える。因にベイズ確率の一般式は

$$P(B_i|A) = P(A|B_i) P(B_i) / \sum (P(A|B_i) P(B_i))$$

$P(B_i)$ は事前確率、 $P(B_i|A)$ は事後確率を示す。

詳細は参考文献（丸山・林，1968）

B：反応の一致率モデル

今、測定対象についての測定尺度が評価尺度分類として与えられているものとする。そこで得られた測定値の確率表は以下のように書ける。

反応一致率マトリックス								
後 前								
	1	2	3	i	n	
1	P_{11}	P_{12}	P_{13}					$P_{1\cdot}$
2	P_{21}	P_{22}						$P_{2\cdot}$
3								$P_{3\cdot}$
.....								
					$P_{\cdot i}$			$P_{\cdot i}$
.....								
n							$P_{\cdot n}$	$P_{\cdot n}$
	$P_{\cdot 1}$	$P_{\cdot 2}$	$P_{\cdot 3}$	$P_{\cdot i}$	$P_{\cdot n}$	N

そこで、単純一致率 (Coefficient of Reliability) CRは

$$CR = \sum_{i=1}^n P_{ii} / N \quad (i=1, 2, 3, \dots, n)$$

で与えられる。Nはサンプル数を表す。ところが、CRは反応分布の集中度 ($CP = \max. P_{ij} / N$ として与えられる) や反応の選択肢の数と関連して相当変わってくるから、多くの尺度項目を比較検討する場合、CRのみでは様子が分からず、これを標準化して用いるのが望ましい。つまり、CRから前後の結果は全く独立であるとし、偶然で一致する部分だけ減じ、この値が前後の実験実施のとき、各人が2回とも一致した反応をするときに1となるように修正した場合の反応一致率を考えるのである。これをCoefficient of Stability, CSと記し、修正した単純一致率、または単に反応一致率と呼び、以下のように与えられているものとする。

$$CS = \sum_{i=1}^n P_{ii} - \sum_{i=1}^n P_i / 1 - \sum_{i=1}^n \bar{P}_i$$

ここに P_{ii} は前後で一致した比率、 P_i はその周辺分布の比率を示し、 $\bar{P}_i = P_i + P_i / 2$ で \bar{P}_i は P_i と P_i の平均を取るものとする。

表1 3色配色の一致率尺度分類

刺激図形	7	5	3	6	2	平均
1	36	44	64	49	80	55.0
2	38	39	52	44	70	48.6
3	36	42	64	39	75	51.2
4	43	45	57	44	70	51.8
5	46	48	61	43	74	54.4
	39.8	43.6	59.6	43.8	74.2	

注：尺度分類についての項目の数

尺度分類	選択肢の数
7	(1), (2), (3), (4), (5), (6), (7)
5	(1), (2, 3), (4), (5, 6), (7)
3	(1, 2, 3), (4), (5, 6, 7)
6	(1), (2), (3), (5), (6), (7)
2	(1, 2, 3), (5, 6, 7)

心理学的測定尺度の構成は予備調査が必要であり、それがテスト (質問項目、尺度項目) の信頼性に繋がり、事前確率から事後確率を予測するようなベイズ確率法や反応一致率を調べるための反復繰り返し測定データは時間を要し、複雑な分析を実施するために、現場サイドでは敬遠される傾向がある。しかし、心理学的測定法の古典的な信頼性尺度や妥当性の方法が今日、医療分野に浸透するとき、分析の本質を見失った方法を用いて簡単にデータ解析を行っているケースが多々見られる。そのような場合は心理学的測定法やデータ解析、統計的概念の理解を深めるために、データ・マイニングの学習が中学・高校における徹底した統計教育が現場で行われる事が必要である。⁽²⁾

QOL評価測定尺度の妥当性の問題

心理学的測定法における知能や人格などのテスト項目の妥当性問題を以下に述べる。テストによる測定尺度項目ではじめに考えなければならない基本課題は測定尺度の妥当性問題である。これは、一言で言えば、当該測定尺度項目は様々な側面において妥当なものであるかどうか、所謂正当なものであるかどうか、を決定する事である。多くの場合、抽象的な構成概念の間接測定である。例えば、人間の不安傾向に関する測定項目を構成する時、測定内容を具体的に規定している外延的定義が明確でなければならない。これは、表面的妥当性といわれ、見かけの正当性を包括的に表現したものである。しかし、意図した概念の外延を十分に満たしているかどうかは分からない。一つのQOL評価測定尺度に対して、他の類似の多くの尺度項目がいくつも考えられるし、疾病の内容によって測定尺度は若干変化してくる。WHOが要請する第4の側面、Spiritualityを測定するためのQOL評価測定尺度を構成するとしよう⁽³⁾。内容的妥当性が高い測定尺度であるためには、まず、Spiritualityとは何かという外延的定義が必要である。

Spiritualityは他聞に宗教・哲学用語である。最も一般的な日本語訳はキリスト教的霊性、実存的霊性、靈魂、彼岸の世界等である。言葉の認知から始めて、実存的霊性の内容、を列挙する。人間とは何か、自己の存在の証明、靈魂の存在、死後の生命、死者の世界、死のイメージ、死への関心・恐怖、人生の指針等霊性に関する測定項目をリスト・アップして、測定内容を明確にする。これは測定尺度を構成するための手続きである。ことに手続きが明確に整っていれば、QOL評価測定尺度の「内容的妥当性」が保証される。この種の沢山ある評価測定尺度の内容的妥当性を高めるためには、はじめにリスト・アップした当面の尺度を羅列するだけでは項目数が多数存在して到底実際の測定値を得ることは不可能に近い。そこで、変数選択、あるいは尺度の分類の手段として因子分析、数量化によるパタン分析、クラスター分析などの解析技法を用いる。

そこで、因子間の独立性が保証され、相互作用を除去した整合性のある評価測定尺度が構成される。しかし、適性テストや職業テストなどに比較して、Spiritualityを測定するための内容的妥当性はかなり困難である。なぜなら、仮に職業適性テストであれば、自分はどんな職業に適しているのかを測定するという明確な目標設定を遂行できる上に、適性テストの結果、当該職業人としての適性が合致した人と適性テストをしなかった一般の人との回答結果から、統計的に有意な差があるかどうかを検討し、職業適性テストの予測妥当性が高いか低いかを検討することが出来る。さらに、入学選抜テストで、入学試験の結果と入学後の成績の関連性を比較し、入学テストの妥当性を確かめる事が出来る。ここにある種の「基準」との比較において、そのテストが目的に採った基準と深い関係があれば、そのテストの基準的関連妥当性が高いと言える。特に妥当性の高さをテストとの基準値との相関係数によって表現する時に、その相関係数をテストの妥当性係数と言う。

だが、或る基準と相関の高いテストでも別の基準とは相関が高いとは限らない。テストの妥当性係数は、このように何を基準に採ったときの相関であるかを明確にしておく必要がある。おそらくは妥当性は経験的なもので、一般に遍く広がったテストであれば、普遍性が高く、それは妥当性の高いテストであると言える。その観点からみれば、因子分析を含む評価測定尺度項目の項目分析が重要である。既に、普遍的な尺度項目は項目分析によって不適切な項目は除去されている。

テストの信頼性に関する若干の考察

テストの信頼性はある特定のテスト得点と、その後の同じテスト得点とが同一のものであるかどうかを典型的な形であらわすことである。既に述べたテスト—再テスト法は、一回目のテストと二回目の得点のペイズ確率を重み付けとして本テストに加味する場合やテスト—再テストの反応一致率を本テストに加重する方法である。但し、テスト—再テストは時間的にもコスト的にもかなり繁雑で手間がかかりすぎるという理由で今日では、現実に調査やテストを実施するにも困難な時代ではこれらの方法は敬遠される。そこで、クーダー・リチャードソン (Kuder, G. F. Richardson, M. W., 1937) の公式を用いて作られたテスト尺度項目の信頼性を考えてみよう。今、いくつかの合理的な仮定の推移値が与えられたと仮定してただ一回のテストからテスト項目間の統計量の信頼性係数を推定して行くことが出来ると考える。以下に示されるのは彼等が導いた公式である。

質問項目が「はい、いいえ」のようにデータが1, 0の2通りしか与えられていない時の質問項目の通過率（1の値をとる人の割合）を P_j とおけば、

$$a = m \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m P_j (1 - P_j)}{\sigma_x^2} \right)$$

m はテスト項目の数である。

今日、医療・関連分野で頻繁にもちいられているのは専らクロンバックの a 係数と呼ばれるもので以下の公式である。

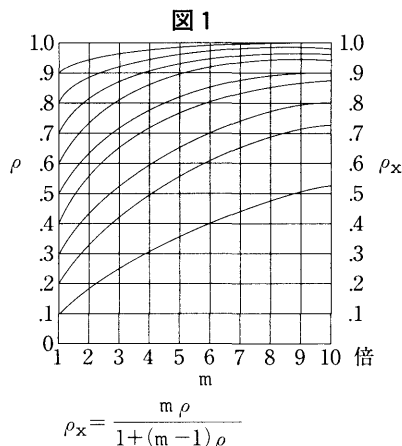
$$a = m \frac{m}{m-1} \left(1 - \frac{\overline{m \sigma_j^2}}{\sigma_x^2} \right)$$

$\overline{\sigma_j^2}$ は m の分散の平均

これに遡ってスピアマン・ブラウン (Spearman, C. 1910, Brown, W. 1910) の公式を示す。

$$a = \rho_x = \frac{m \rho}{1 + (m-1) \rho}$$

ρ が様々な値を採る時、 m の数を増やして行けば、 ρ_x がどのように変わるかを示す図が図1である。同質の問題を増やして、テスト量を増やせば、信頼性係数が上がってくることを示している。



同質の問題を同一条件でふやすことによって
増加する合成値の信頼性係数

これは、テスト項目の信頼性問題で最も一般的な条件を満たす。即ち、信頼性の高い測定値を得るには、疲労や飽きが来ない範囲内で、1) 尺度項目（問題数）が多いこと、2) テスト時間が長いこと、3) はいと答える割合や通過率（正答率）が半々に近いこと（質問の答えが一方に偏って集中する項目があれば信頼性を低める）、4) 質問の指示や意図が明確である事（曖昧な尺度項目ではないこと、質問をする前に意味不明、多義的で回答が困難な尺度項目をチェックさせる）、5) 質問形式や内容がバランスが取れていること、6) 回答に偶然的要素の入る余地が少ないこと（アチーブメント・テスト等にはこの要素が多い）、7) 採点が客観的に決められること、8) 回答者集団の分布範囲が広いこと、9) 回答者の精神的身体的条件が良好で変動が少ないこと、等々である。

このようにテスト項目の信頼性問題は教育の分野におけるテスト項目の信頼性を意図しており、一般の社会意識調査や態度測定項目とは若干ニュアンスが異なっている。上記のスピアマン・ブラウンの信頼係数、クロンバックの α 係数、クーダー・リチャードソンは上記の1, 2の条件を満たす事が狙いである。なお、この種の尺度項目の信頼性を高める問題は更に尺度の一次元性の問題をも含んでいる。多次元尺度項目を一次元尺度項目に縮小して行くプロセスは因子分析や主成分分析等である。尺度項目の多次元性を仮定しながら、多数ある尺度項目相互の相乗作用を排除する役割を持っている。集められた質問項目が一次元性尺度上になっているかどうかを確認するための分析は通常尺度分析、スケイログラム・アナリシス、ガットマン・スケール等がある。

QOLD評価測定尺度の信頼性の検討

先に論じたQOLD評価測定尺度の信頼性についてクロンバック α 係数を算出し、一般大学生と好

癖行動障害者（アルコール依存症、拒食症等）のスケールを比較してみよう。

表2 SPスケールの信頼性係数（大学生）

	現状認識 性格傾向	死の イメージ	宗教・習慣
	I	II	III
1：今の自分の生き方が気に入らない	0.615	0.143	-0.008
2：いつも孤独であると感じる	0.826	0.096	0.002
3：自分自身と向き合う事に不安を感じる	0.512	0.221	0.037
4：自分には存在価値があると思わない	0.561	0.161	0.025
5：死の恐怖に対して他者よりも強い	0.160	0.492	0.192
6：死を意識したのは他者より早い	0.145	0.651	0.034
7：自分の死について他者よりも多く考える	0.158	0.701	0.071
8：御守り御札を大切に作る	-0.023	0.032	0.507
9：先祖の霊が見守っている	0.027	0.103	0.738
λ 寄与率	0.355	0.286	0.264
α 信頼性係数	0.793	0.783	0.789

表3 SPスケールの信頼性係数（好癖行動障害者）

	現状認識	将来・不安感	死生観
	I	II	III
1：私はとにかく年齢を気にするほうだ	0.637	0.248	0.031
2：私は地位や財産、名誉を生きて行く支えにする	0.675	0.060	0.031
3：何か不幸なことが起こりはしないかいつも心配だ	0.612	0.310	0.074
4：私は生きて行く上で、名誉や財産にこだわる	0.679	0.027	0.024
5：自分には存在価値があるとは思わない	0.168	0.717	0.078
6：私の未来は絶望的である	0.150	0.898	0.152
7：死は全ての活動の終わりである	0.298	0.098	0.682
8：死んでしまったら、魂も含めて全てが消滅する	0.076	0.216	0.855
9：人は死んでも繰り返し生まれ変わるという事を信じない	0.194	0.007	0.623
λ 寄与率	0.466	0.414	0.385
α 信頼性係数	0.807	0.899	0.848

表2は23項目のQOLD評価測定尺度を基に大学生の回答から抽出された9項目の3次元空間における因子負荷量とそれらの各次元におけるクロンバック α 係数、表3は好癖行動障害者の結果である（丸山，2001）。これらを比較すると大学生は第1次元は現状認識、性格傾向で $\alpha = 0.793$ 、第2次元は死のイメージで $\alpha = 0.783$ 、第3次元は宗教・習慣行動で、 $\alpha = 0.789$ で全体の α 係数は0.788、それに対して、好癖行動障害者の第1次元は現状認識、 $\alpha = 0.807$ 、第2次元は将来に対する不安感、 $\alpha = 0.899$ 、第3次元は死生観、 $\alpha = 0.848$ で全体の α 係数は0.851である。

信頼性係数としては好癖行動障害者の方が高いが、少なくとも α 係数は0.8以上である事が望ましい。QOLD評価測定尺度として、前述の23項目が妥当性があるかどうか今後の検討に期待したい。

お わ り に

高齢者医療・障害者福祉関連分野、ホスピス関連分野におけるQOL評価測定尺度の研究は医療の中でもやや一般化している疾病特異評価測定尺度（糖尿病、慢性腎疾患、癌等）に比較して初期の研究段階にあると思われる。諸般の議論は頻繁に行われているとはいうもののWHOが要求する普遍的な尺度の構築はこの分野で円滑に促進されるか否かに疑問が残る。測定値は主観的な回答から成り立っているが、認知的障害者、精神障害者、身体的障害者、急性の病気になった病人等はQOL尺度に対して回答不能に陥る事は容易に理解できる。そこで客観的に回答するのは医者や身体的な疾患から来る医学的推測や世話をしている人の憶測で、それが情報源となる。特に、WHOが要請するSpiritualityの分野には世話人や他者からの情報さえ不可能である。代理人の情報を信頼することが如何に困難な事態であるかを我々は良く知っている。そこで、倫理的な社会政策論に基づくQOLの概念や構成要素を取り込む必要がある。今日、肢体障害者や高齢者などの自立を促進するためには介護援助の仕方に多くの工夫が凝らされている。この種の議論は広く社会政策論に発展し、今後十分にQOL評価測定尺度の結果を取り込みながら、社会施策を深めて行く事を期待する。

参考文献

- Brown, W. Some experimental results in the correlation of mental abilities. Brit. J. Psychol. 3, 296-322, 1910
- Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334
- 池田 央 信頼性と妥当性の問題 田中良久（編）計量心理学 東京大学出版会 1969
- 池田 央 行動科学の方法 東京大学出版会 1971
- 池田 央 調査と測定 新躍社 1980
- 池上直巳・福原俊一・下妻晃二郎・池田俊也（編）臨床のためのQOL評価ハンドブック 医学書院 2001
- 印東太郎（編）数理心理学 東京大学出版会 1969
- Fayers, P. M., Machun, D. Quality of life-Assessment, analysis and interpretation. Wiley, 2000
- 林 知己夫 データの科学 朝倉書店 2001
- 林 文・山岡和枝 行動科学とデータの科学 朝倉書店 2002
- Kuder, G. F., Richardson, M. W. The theory of the estimation of test reliability. Psychometrika, 2, 151-160, 1937
- 丸山久美子・林知己夫 反応誤差に基づく数量解析の歪み, 「心理学研究」, 38, 297-310, 1968
- 丸山久美子・林 文・林知己夫 3色調和感にもとづく色彩空間構成の一つの試み, 「色彩研究」, 18, 8-20, 1971
- 丸山久美子 心理学的誤差の概念について, 「科学基礎論研究」, 15, 77-86, 1981
- 丸山久美子 QOLD評価測定尺度に関する基礎的研究 (I), 「聖学院大学論叢」, 第9巻, 第2号, 139-

156, 1997

丸山久美子 QOL評価測定尺度に関する基礎的研究 (IV), 「聖学院大学論叢」, 第14巻, 第1号, 101-118, 2001

万代隆 (編) QOLD評価法マニュアル 医学書院 2001

鈴木達三 面接調査における回答の安定性について, 「統計数理研究所集報」16, 47-71, 1968

Spearman, C. Correlation calculated with faulty data. Brit. J. Psychol. 3, 271-295, 1910

田中良久 心理学的測定法 東京大学出版会 1961

漆崎一朗 (監修) QOL調査と評価の手引き—調査と解析の実際とベッドサイドの生かし方—, 癌と化学療法社 1995

注

(1) 測定尺度値には4つの水準がある。

(Stevens, 1951; 吉田, 1968, 池田, 1971より転載)

(注) 水準1の尺度が成り立つ性質は水準2以上の尺度でも成り立つ。

	尺度の水準	目 的	特 徴	許される演算	尺度どうしの変換	例
定性的変数	1. 名義尺度	分類・命名符号化	$A=B$ または $A \neq B$ の決定	計数の勘定	1対1置換	氏名番号 サッカー=1 男=1 バレー=2 女=2 テニス=3 等
	2. 順位尺度	順序づけ	$A>B$, $A=B$ $A<B$ の決定	順序統計量の統計	単調増加または減少変換	健康 良 好=3 ふ つ う=2 よくない=1 等
定量的変数	3. 間隔尺度	等間隔な目盛づけ(原点・単位は任意)	$(A-B) + (B-C) = (A-C)$ の成立	和・差をもとにした統計量の算出	1次変換 $Y=aX+b$	温度 (セ氏・カ氏), 国語・英語標準得点等
	4. 比率尺度	絶対的の原点から等間隔の目盛づけ(単位のみ任意)	$A=kB$, $B=lC$ なら $A=klC$ の成立 ($k \neq 0, l \neq 0$)	加減乗除をもとにした統計量の算出	比例変換 $Y=aX$	身長・体重, 絶対温度 等

水準2の尺度で成り立つ性質は水準3以上の尺度でも成り立つ。

水準3の尺度で成り立つ性質は水準4の尺度でも成り立つ。

スピアマンの順位相関係数

$$\tau = 1 - 6 \sum (x-y) / n(n-1)$$

ピアソンの積率相関係数

$$\rho = \frac{\text{Cov. } xy}{\sigma_x^2 \cdot \sigma_y^2}$$

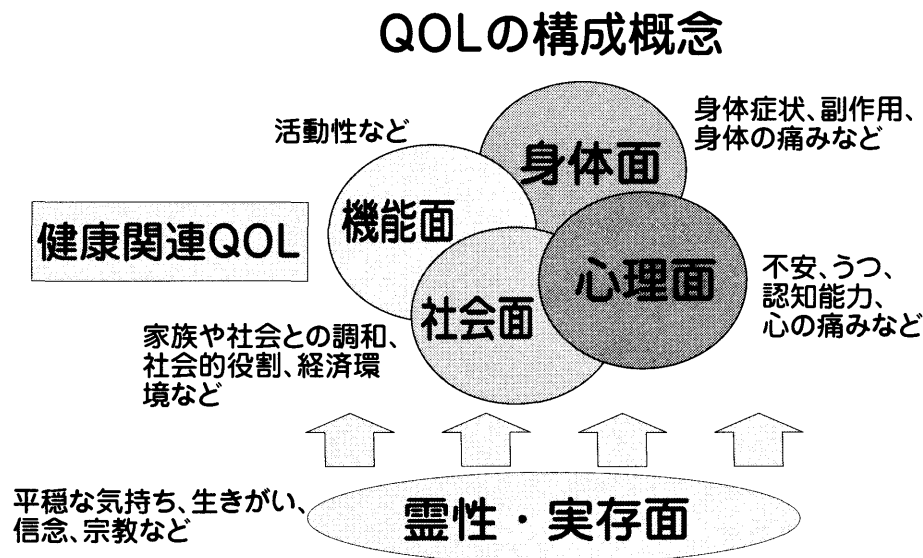
尺度水準が異なるが、スピアマンの順位相関を計算した平均値に順位をつけて順位を間隔尺度値とみなし、通常のピアソンの積率相関係数を適用しても構わない。多くは選択肢に基準点を持つSD尺度値に好きから嫌いまである段階点を順位と見做しスピアマンの順位相関係数を算出しても一向に構わない。但し、尺度水準と統計的データ解析法との関係に関しては多くの論争がある。筆者は尺度水準の定義からしてスピアマンとピアソンの相関係数を同等に扱う事に賛成できない。QOL測定尺度では順位尺度と間隔尺度を同じ水準で扱う研究が多い。この傾向は極わめて危険である。測定値の種類と対応する測定法は緊密に

結びついている。安易に測定値を的外れな測定法を用いて分析する事は慎むべきであろう。以下に測定値とそれに対応する測定法を参考までに示す。

測定値の種類		測定法
metric	計量的数量	重回帰分析, 主成分分析, etc
semi-metric	半数量, カテゴリカル・データ	数量化理論, 多次元尺度解析
ordered metric	順序のついた数量	一対比較法, スカラー順位法
non-metric	非数量, 順序データ	SSA (Guttman, 1962) MDA (Hayashi, 1974)
non-linear	非線型データ	非線型回帰分析, ロジステック分析

(2) 今日, この種の統計教育が特に日本における文部科学省の学習要綱には置き忘れられ, 日本統計学会では「統計教育賞」を設置して, 現場の優れた統計教育の実践教育者に授与し, さらなる貢献に期待している。

(3)



下妻晃二郎
医療分野におけるQOLの現状と将来について(2001)