

健康・スポーツ科学における「はかる」ことの意義

—測定評価・統計リテラシー教育のススメ—

鈴木 宏 哉

東亜大学 総合人間・文化学部 健康科学研究室
E-mail: suzuki@toua-u.ac.jp

奥 本 正

東亜大学 総合人間・文化学部 健康科学研究室
E-mail: okumoto@toua-u.ac.jp

江 橋 博

東亜大学 総合人間・文化学部 健康科学研究室
E-mail: eba@toua-u.ac.jp

安 陪 大治郎

東亜大学 総合人間・文化学部 スポーツ学研究室
E-mail: daijiro@toua-u.ac.jp

石 井 信 輝

東亜大学 総合人間・文化学部 スポーツ学研究室
E-mail: ishii@toua-u.ac.jp

櫛 田 芳 美

東亜大学 総合人間・文化学部 スポーツ学研究室
E-mail: kushi443@toua-u.ac.jp

大 森 一 伸

東亜大学 総合人間・文化学部 スポーツ学研究室
E-mail: ohmori@toua-u.ac.jp

岡 村 豊太郎

東亜大学 総合人間・文化学部 スポーツ学研究室
E-mail: okamura@toua-u.ac.jp

1. はじめに

地球環境と人間の健康は密接に関連しており、特に地球の温暖化が生活環境に負の影響を及ぼすことは周知の事実である。「CO₂（二酸化炭素）削減どう算定：電力・ガス業界綱引き」という見出しの記事が朝日新聞（2006年1

月10日朝刊）に掲載された。記事の内容は、コージェネレーションシステム（電気と同時に有効に利用できる熱を発生し、エネルギーを多段的に活用する省エネルギーシステムのこと、エネルギーの高効率利用が特徴）が、CO₂をどれだけ削減することに役立つのかを判断するために用いられる基準について、夏などの電気需要が多い時期に増加する火力発電所

からの CO₂ 排出量のみを基準とするか、すべての種類の発電所からの CO₂ 排出量を基準とするかをめぐり、電力業界とガス業界が対立しているという内容であった。対立の背景には、都市ガスなどを利用するコージェネレーションシステムを積極的に導入したいガス業界とそれに対して慎重な電力業界のパワーゲーム的な争いがある。このように、CO₂ 削減量の算定ひとつをとってみても、その算出方法の決定には、真に地球環境を守ろうとする純粋な観点だけでなく、経済的あるいは政治的観点が関与しており、その指標の決定は今後の日本経済だけでなく、地球規模の環境や社会システムに多大な影響を及ぼすものであるということを伺い知ることができる。それだけに、ある実態・事象をどのように「はかる」か、ということは重要であることが分かる。

今回の特集のテーマである「はかる」は、上述した例でいうと、CO₂ 削減量の算出式を作成し、それに基づいて CO₂ 削減量を求めるということになる。「はかる」という意味にはこの他に様々な意味を含んでおり、統計学、測定評価学（あるいは計量学）の研究者が様々な観点から定義し、論じている。本稿では、健康・スポーツ科学分野において「はかる」とは如何なることかということを通して、「はかる」ことの意義、「はかられたもの」の見方（読み取る際の心構え）について述べ、健康・スポーツ関連情報（データ）が氾濫する現代において、データの本質を見抜く力の必要性を提案する。

2. 本稿で扱う「はかる」

本稿では、「はかる(こと)」を「測定評価」という概念に置き換え、一般に用いられる「測定」と「評価」に関する統合的概念として「はかる(こと)」という用語を議論する。また、「はかる」には「統計(学)」の根源的意味でもある「統べ計る」という意味も含まれる(図1)。したがって、これ以降は統計学的視点の一部を含む測定評価学的視点から「はかる(こと)」を議論することとする。

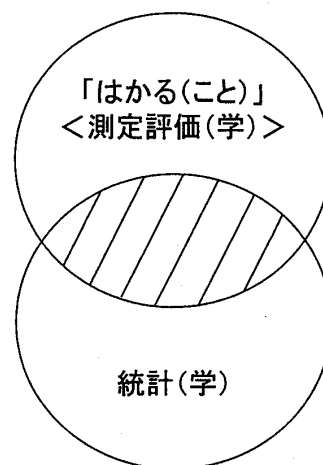


図1. 「はかる(こと)」・測定評価学・統計学概念的包含関係

(注) ここでの「はかる」は「計る・量る・測る」に該当する。

3. 健康・スポーツ科学における測定と評価

筆者が考える一般的な(健康・スポーツ科学に限定されない)測定及び評価の定義は次のようである。測定(measurement)とは「ある目的にしたがってデータ(data)を収集すること」であり、「データはある現象に対して何らかの物差し(scale)をあてがうことによって収集される」。評価(evaluation)とは「収集されたデータに価値(意味)を与えること」である。この定義は代表的な計量学者らによる定義(付録1)の共通部分を取り出した表現といえよう。

体力・運動能力を例に挙げると、体力・運動能力を測ったときに、それが一般大学生として低い値か否か、発育発達段階に照らし合わせて標準的か否かなどの判断を行おうとするとき、ある物差し(scale)を用いて体力・運動能力を測る部分は「測定」であり、低いあるいは標準的であるなどの価値判断の部分は「評価」であるといえる。

測定及び評価の概念は如何なる科学分野においても同様である。異なる点は測定対象であるといえる。健康・スポーツ科学における測定対象は身体及び精神機能・身体能力、それと身体運動現象(パフォーマンス)である。機能・能

力と身体運動現象の間には図2に示した関係がある(注1)。機能が関与した結果として運動現象が表出される。あるいは、能力が発揮された結果としてパフォーマンスが成就される。

健康・スポーツ科学における測定対象を図2のように仮定した場合、健康・スポーツ科学における測定の立場は2つに分類することができる。1つは、身体運動現象は身体及び精神機能・身体能力の複合的関与及び複合的発揮の結果であり、身体及び精神機能・身体能力を測定することによって、身体運動現象は説明されると考える立場である(図3)。長距離走の成績を全身持久力から説明(予測)しようとし、全身持久力の指標として最大酸素摂取量を測定することなどがこの立場である。もう1つの立場は、身体運動現象(パフォーマンス)を測定し、そのパフォーマンス成就に関連の強い機能・能力を推定しようとする立場である(図4)。サッカーのドリブルテストやシューティングテストなどからサッカー技能(注2)を推定しようとするなどがこの立場である。2つの立場は前者をラボラトリテスト(実験室テスト)による測定、後者をフィールドテストによる測定のように簡約することもできる。いずれの立場においても身体及び精神機能・身体能力と身体運動現象との関連性を生理学的メカニズムなどの理論的背景から十分考察しておく必要がある。なぜならば、関連性が不確かであるとき、その身体運動現象からどんな能力を推定できるのかを判断できなくなってしまうからである。筆者はこのことに関連する不思議な論文を目にしたことがある。その論文は、調整力の下位概念である敏捷性を測定するテストとして反復横跳びを用いた研究であったが、反復横跳びは脚伸展筋力の高いものほど優れた成績を示したというものであった。断定はできな

いものの、この結果は反復横跳びの成績が筋力に左右されるということを示しているのであるならば、反復横跳びは敏捷性を測定するテストとしては適切ではないことを意味していると考えるか、あるいは敏捷性は筋力を向上させることで向上するということの意味していると考えしかるべき。前者であるならば、明らかに研究者の前提を自分自身で否定していることになるし、後者であるならば、一般的な体力・運動能力の構造を考えるならば、敏捷性と筋力は並列的な概念であり(池上, 1990)、因果的関係にある概念ではないため、一般的な了解は得られない。

このような事例は決して稀なものではなく、だれもがややもすると陥りかねない現象である。能力と身体運動現象の関連性を十分考察し

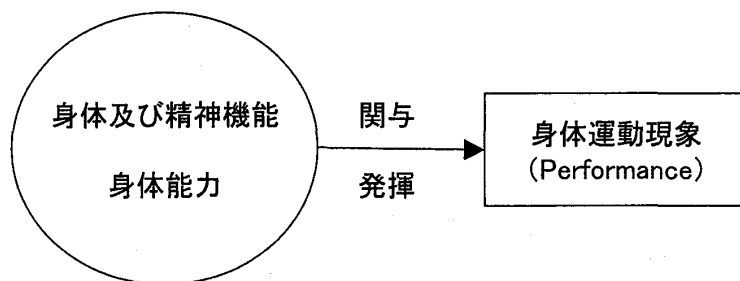


図2. 身体及び精神機能・身体能力と身体運動現象の関係

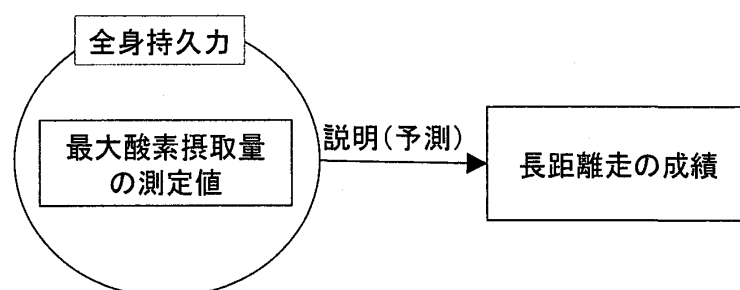


図3. 最大酸素摂取量の測定値と長距離走の成績の関係

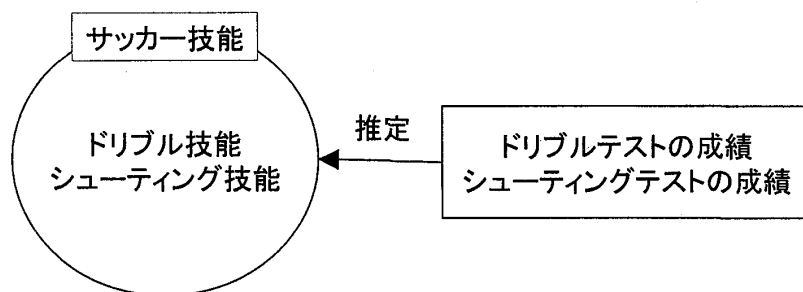


図4. フィールドテストの成績と技能の関係

ておくということは、このような事態を招かないためにも必要である。少なくとも、この例では敏捷性を測定するテストとしての反復横跳びの妥当性を事前に検討するということが必要である。不確かな概念と不確かな概念の関連性からは不確かな結論のみが生まれるものである。不確かな概念（機能・能力）を客観的で確かなデータに変換する手続きが測定といえる。テストの確からしさを評価する概念には妥当性 (validity) がある。

4. 「はかりたい」ものと「はかられた」ものとのギャップ

体力・運動能力、生活の質 (Quality of Life: QoL)、健康観など、様々な概念を測定するためにフィールドテストや質問紙テストが用いられる。しかしながら、これらの概念をテストはどの程度適切に測定できているのだろうか。言い換えると、「はかりたい」ものと「はかられた」ものとの間にはどの程度違いがあるのだろうか。

例えば、サッカーの授業を受講する学生に優劣をつけるためにサッカー技能を測定したいと考える。教師は試合中のプレーを観察して、学生の優劣をつけようと考えたが、複雑すぎて客観的な指標に落とし込むことができない。しかしその教師はサッカーの試合中にドリブルをする場面が多いことに気づき、ドリブルパフォーマンスを客観的に測定するために、等間隔に並べたコーンをジグザグにドリブルしたときの所要時間 (ドリブルテスト) によってサッカー技能を推定した。

この例では適切にサッカー技能を測定できているだろうか。サッカー技能を構成する下位技能にはドリブル技能以外にも、シューティング技能、守備に関する技能、味方とのコンビネーション技能などがあると考えられる。したがって、ドリブルテストによって推定されるドリブル技能はサッカー技能の一部でしかない (図5)。図に示すサッカー技能の円全体を網羅するテストのほうがより適切なテスト (妥当性が

高いテスト) であるといえる。しかしながら、健康・スポーツに関する諸現象を測定しようとする場合、テストの妥当性について、概念図に示したように、ことは単純にはいかない。なぜならば、能力は目に見えない概念であるからである。身長などの形態は目に見える。身長計は身長を測定するテスト (測定器具) として妥当性が高いことはいうまでもない。ところが、サッカー技能、健康観、QoL などは概念であるために、妥当性の高さは図5に示す円をどのように仮定するか (定義するか) に依存することになる。QoL の概念は研究者や研究分野、使用される文脈により様々に解釈されている (田中他, 2004)。QoL を身体的健康、心理的状态、自立レベル、社会的関係、環境の特徴に対するこれらの関連の5領域から構成されると仮定する場合 (WHO QoL Group, 1993) と、身体の質的側面、精神の質的側面、環境の質的側面、経済の質的側面の4領域であると仮定する場合 (Spiriduso, 1995) とでは、妥当性が高いと判断されるテストは異なる。最近では、健康度チェックというような見出しで、生活習慣の質問に答えるもの、血液流動性 (俗に「血液サラサラ度」といわれている指標) の検査を行うもの、骨密度を測定するものなど枚挙にいとまがない。健康という言葉から受けるイメージは人それぞれ異なる。World Health Organization (WHO) 憲章の定義によれば、健康は身体的、精神的、社会的に良好な状態を意味し、上述したような測定はごく一部を説明しているに過ぎない。しかし多くの人びとは、テストそのものではなく、各個人がテストによって推定される (と信じている) 概念 (この例では健康) に目を奪われ、一部分を説明しているに過ぎない検査結果を鵜呑みにしているきらいがある。ここには、テスト作成者による「はかりたい」ものと実際に「はかられている」ものとのギャップだけでなく、テスト受験者による「はかっている」と錯覚しているものと、実際に「はかられている」ものとのギャップが混在する。

この傾向はデジタルな出力であるほど強い傾

向にあると思われる。最近では多くの家庭に普及している体脂肪計や体組成計などの類が典型であろう。ヒトの身体組成を直接測定するには人体解剖をする以外にないことから、簡易に間接測定する方法が多く開発されてきた (Heymsfield et al., 2005)。その中に、生体電気抵抗法 (bioelectrical impedance method: BI 法) を用いた身体組成推定方法がある。BI 法の問題点は多くの論文で指摘されているものの (奥野他, 2000; 大河原他, 2003)、その簡便性から一般家庭に広く普及している。一方で古くから活用されてきた方法に皮下脂肪をキャリ

パーを用いてつまみ、その皮下脂肪厚から推定する方法 (皮脂厚法) がある。いずれの方法においても実用面における長所・短所はあるが、推定精度には大きな違いがないことが確認されている (奥野他, 2000)^(注3)。しかし、推定原理を理解している一部のを除いては、人の手で測定を行う方法よりも計測器が自動測定する方法のほうが推定精度に優れると考える傾向がある^(注4)。

我々は「はかられたもの」と「はかられている」であろう概念との間にはギャップがあるということを認識しなければならない。しかし、上述したように測定 (テスト) が妥当であるかの判断は、一般には判断しにくいことが多い。アメリカではこのような実態に即し、米国心理学会 (American Psychological Association: APA)、米国教育学会 (American Educational Research Association: AERA)、全米教育測定協議会 (National Council on Measurement in Education: NCME) が合同で教育・心理検査法のスタンダードを作成し、その中でテスト開発者と出版者がテストやテスト得点が誤解・誤用される見込みを減らすのに役立つ情報をテスト利用者に提供するべきであるということを明記している (AERA et al., 1999)。テスト

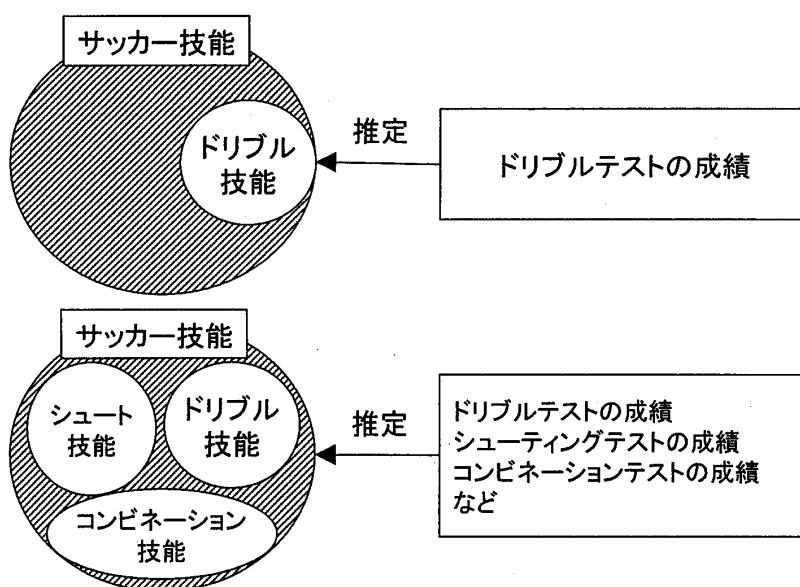


図5. テスト成績と技能の関係

(注) この例では、上図よりも下図の方がサッカー技能を妥当に測定できているといえる。

利用者はその情報に基づき、自分が適用する状況において、テストがどの程度適切に機能するかを判断することができる。日本においても健康・スポーツ科学分野のテスト開発と利用のスタンダードができることを期待したい。それ以上に、研究者やテスト開発者には、複雑な統計手法を駆使した論文で煙に巻くのではなく、テスト利用者に対する利用手引きとなる情報を提供しようとするスタンスが必要であろう。

いずれにしても、「はかる」ことによって概念が客観化され、その客観化の確からしさも客観化されたデータを用いて検証されるのである。そういう意味で、「はかる」ことはあらゆる事象を客観的に説明するための基礎である。しかし、重要なことは、どんな場合にも、概念と測定値の間にはズレがあることを認識するということであり、測定値を絶対視してはならないということである。

5. テスト選択の重要性：測定項目が持つ社会的インパクト

専門家であってもテスト利用に関する判断を誤ることがある。次に示す例は、誤用ではないものの社会的な影響を及ぼす判断をとらな

テストの選択である。

文部科学省は国民の体力・運動能力の現状を把握することなどのために、昭和39年以来、体力・運動能力調査を実施している。調査のために用いられているテストの中で、調査当初から替わることなく用いられているテストに握力テストがある。平成12年から用いられている新体力テスト（文部科学省，2000）において、小学生から高齢者までを通じて、純粋な筋力を測定する項目として共通に扱われているのが握力テストである。筋力には上肢の筋力と下肢の筋力がある。その中で、上肢の筋力を測定する握力テストによって、筋力を評価しているのにはいくつかの理由がある（文部科学省，2000）。小学生から高齢者までだれでも容易に実施できることや、背筋力や脚筋力と握力の測定値に直線的な関係がある（高い相関がある）ことが主な理由である。高齢者が単なる生活の自立にとどまらず、積極的な社会参加や社会貢献ができるためには、元気で余裕のある歩行能力を維持すること、そして、歩行能力の維持のために下肢の筋力が重要であり、下肢の筋力が老化を反映していることは多くの先行研究によって明らかにされている（福永，2000；金ほか，2000）。アスリートにおいても脚筋力はいうまでもなく重要である。握力テストを軽視するわけではないし、利点はたくさんあるが、握力テストを運動指導の現場で用いる場合に生じる問題点を指摘したい。

筆者は中高齢者の運動指導において、参加者の体力の現状を把握することと、運動指導の効果を確認するために新体力テストを用いることがある。新体力テストを利用する利点は、簡便であることと大標本から求められた全国の標準値があることである（標準値に関する解釈の注意点については後述する）。運動指導前と運動指導半年後の2回体力テストを実施し、改善の程度を確認すると、運動開始前は全国値以下だった参加者が、6分間歩行、開眼片足立ち、長座体前屈、10m障害物歩行において全国値かそれ以上の値を示し、握力だけがあまり変化しないことがある。参加者の中にはそのことを

気にして、握力を鍛えようと普段の生活で意識して重い物を持ったりするようになる者もいる。このこと事態は全く問題ではなく、むしろ握力を測定したことによって参加者の行動をプラスに変容させることができた例にも見える。しかし、指導する側の意図するところは、高齢になっても活動的に生活できるために、参加者の下肢筋力を向上させることであり、そのための簡易筋力測定として握力測定を用いたのである。また、握力測定をしなければ、総合体力評価の算出ができないから握力測定を行ったのである。

測定の目的は様々である。集団の特徴を把握することやトレーニングの効果を判定するためなどである。テストを実施する側は測定の目的以外にも考慮に入れるべきことがある。それは、前述の例に示したように、測定にはテストを受けた側の行動を変容させる効果があることである。この例では一個人の行動変容に影響を与えた例であるが、大学入試センター試験におけるリスニングテストの導入などのように数万人規模の受験生と教育現場の変容に多大な影響を与えるテストもある。したがって、テスト作成者は、測定項目の選択が個人の意志決定に大きな影響を及ぼすことを肝に銘じて測定項目を選択する必要がある。

このようにテストの妥当性に関する概念は、テストの確からしさからだけでなく、そのテストを使用したことによってもたらされる受験者の行動変容についても含まれなければならない。このことは1999年に刊行された教育・心理検査法のスタンダード（AERA et al., 1999）にも妥当性の概念の中に「Evidence based on consequences of testing」として登場し、改訂前の1985年版（AERA et al., 1985）にはない要素が時代の要請に合わせて追加されたことから、その重要性が伺える。

テストはある限られた目的のために作成されていることが多く、その目的以外に利用しようとする場合には特に慎重にならなければならない。文部科学省新体力テストは国民の体力・運動能力を把握することが目的で作成されたもの

であり、大集団の体力・運動能力の実態を簡易にいち早く、できるだけ正確に把握するという観点において、優れたテストである。しかし、トレーニングの成果を良く反映し、対象者の行動変容を促進するテストとして優れているということではない。そのような目的で体力を測定するのであれば、その目的に特化して妥当性の高い体力テストを実施すればよいのである。もしもトレーニングの成果を確認したいのであれば、そのトレーニングによって変化するであろう要素を鋭敏に測定するテストであるか否かを事前に検討しておく必要がある。

いずれの場合においても、テスト利用者は「万能なテストは存在しない」ということを認識し、テスト作成者は「用いる目的に応じた妥当性の検証が必要である」ということ、そして「測定項目の選択が個人の意志決定に大きな影響を及ぼす」ということを認識しておく必要がある。

6. 測定された標準値や基準値の解釈の 注意点

身長や体重、体力測定値には、ある集団における平均的な値（標準値）がある。ある測定値をもとに個人を評価するためには標準値が必要となる。身長や体重などは充実した官庁統計が存在する。体力・運動能力についても文部科学省が毎年、数万人規模の測定を実施している。しかし、体力・運動能力のテストについては標準値を作成するための基礎となる集団に偏りがあることに注意しなければならない。

文部科学省体力・運動能力調査報告書における各年齢の標準値について、小・中学生の測定値から得られる標準値には大きな問題はないと考えられるが、高齢者においては注意が必要である。小・中学生は、義務教育のもとに日本に住む全児童・生徒が対象（母集団^(注5)）となり、層化多段的に無作為抽出が行われ、学内行事や授業の一貫として半強制的に実施される。しかし、高齢者においては、測定場所に足を運び、そこで、新体力テストを実施することがで

きる体力を持つ高齢者のみが対象となり、集団の測定値が得られる。したがって、測定が困難なほどの低体力者の測定値は得られない。加えて、高齢者の中で体力テストを受けてみようと思う人は、普段から健康に気をつけている人や自分の体力に自信のある人であることが多い。そのような集団から得られた測定値から標準値を作成しているために、報告書における値は真の高齢者全体よりも高い標準値になっている。この標準値に従って運動教室参加者の体力を評価すると、十分元気に見える高齢者でも標準値以下と判定されることがある。しかしこの評価が真に示しているのは、体力テストを実施できるくらいの元気な高齢者集団と比較するとやや劣るということだけであり、日本人の20%を占める65歳以上全員と比較して劣るわけではない。

『統計はこうしてウソをつく』の著者ベスト(2002)は、著書の中で、良い社会統計が満たす条件として、調査された標本^(注6)が大きいことと、抽出された標本が母集団を代表していることを挙げている。体力テストの例では、標本の大きさに関して、65歳以上の標本が3千人弱であり、標本から得られた平均値と母集団から得られる平均値との誤差には大きな問題はない^(注7)。しかし標本の代表性については、無作為標本抽出とはほど遠い標本に基づいているために十分とはいえない。体力テストは特に、測定の制約上、無作為抽出ができないことはやむを得ない。しかし、ベストが述べるように、「その数字がどのようにして出てきたかが語られているということ (p. 85)」が良い統計の目安であり、それらの情報を利用者が把握した上で、標準値を解釈することが正しい意志決定をするために必要となることを理解しなければならない。

次に、基準値を解釈する上での注意点について述べる。高齢者や低体力者が健康づくりのために行う有酸素運動の安全基準として、最大酸素摂取量の最低30%から50%強度程度での運動が推奨されている(ACSM, 2005)。この強度は何らかの運動負荷試験により最大値を求め

ること得られる。しかし高齢者の運動負荷試験においては最大酸素摂取量を求めるために必要となる酸素摂取量のレベリングオフ（負荷を上げて酸素摂取量が上昇しなくなる状態）が現れることは稀である。したがって、最大酸素摂取量よりも低い値を最大値と見なして、運動強度を設定してしまっているため、安全であるものの効果が得られるほどの強度になっていないことが多い。そのため、田中（2003）は運動負荷試験による強度設定が必ずしも妥当であるということではなく、ひとつの目安として利用すべきであると述べている。また、実際の運動では数値にこだわり過ぎるとよくない。テスト（測定）は思っているほど有効でないことがあるとさえ述べている。

標準値や基準値にはさまざまな、やむを得ない限界を含んでいる。したがって、以上述べてきたことは新体力テストの標準値や運動負荷試験から得られる基準値を否定するものではなく、利用する側が測定値の限界を理解した上で活用すべきであることを示したのである。

7. 「はかる(こと)」がもたらす弊害

証拠に基づく医療（Evidence Based Medicine: EBM）に代表されるように、エビデンススペースという言葉が、近年強調されるようになった。それは、ある概念を客観化するために測定を行い、そこから得られるデータ（証拠）に基づいて事象を説明するということである。アメリカなどの保証社会ではエビデンススペースが思考の根底にある。「はかる(こと)」は曖昧さをなくすための最良の方法である。しかし、「はかる(こと)」は同時に「はかっていない」事象を生む。別の意味では、「はかる(こと)」ができない事象はエビデンススペースの社会では無能であるといえる。哲学者ヴィトゲンシュタインは「われわれはゲームのルールやテクニクを確立する。しかし、このルールにしたがおうとしても、予想通りには事が運ばない。したがって、いうなれば、われわれは自分たち自身のルールに囚われてしまうのである。（ストラ

ザーン, 1997)」と述べている。エビデンスベースの社会では、ヴィトゲンシュタインが述べるように、はかられたものだけで議論せざるを得ない状態を生み出してしまったのである。

エビデンスベースの社会では、ある運動を行った際の健康に対する効果を示すエビデンスがなければ、健康づくりのための運動として、その運動を推奨することはできない。逆説的には、エビデンスが用意できる運動のみが推奨される。2002年にアメリカで、「Walking Games and Activities」という著書が出版された（Decker and Mize, 2002）。子どもを対象に、健康づくりの一貫としてウォーキング遊びを紹介する内容である。健康づくりに有効なウォーキングは単調な運動であるから、ゲーム性を取り入れて楽しく歩こうという趣旨である。これは、エビデンスベース思考がもたらした望ましくない結果を表現しているように思える。歩行動作は種々の条件をコントロールすることが容易であり、研究対象として扱いやすいため、様々な視点から歩行運動の効果や安全性を検証した研究が多数報告されている。したがって、それらのエビデンスに基づいてウォーキングを推奨しようとしているのである。子どもはウォーキングよりも野球やサッカーなどのボールゲームのほうが楽しんで実施できるはずである。しかし、ボールゲームはチームスポーツであり、対戦チームとの攻防であるため、種々の条件をコントロールすることが難しい。したがって、画一的な運動の効果を示すエビデンスを用意しにくい。根源的な運動欲求を刺激するようなボールゲームでは運動の効果のエビデンスが用意できないので、エビデンスの豊富な単純運動を推奨するということは、本末転倒であり、「はかる(こと)」がもたらした弊害といえるのではないだろうか。このことは一概に間違っているとはいいい切れない。ノーベル物理学賞受賞者のP.W.Bridgmanは、測定したものによってその概念を定義するという操作主義の提唱者として知られている。このことは実証的研究の基本である。しかし、はかる(こと)ができない部分に本質が潜んでいることも忘れるべき

ではない。我々はその概念の一部を測定しているに過ぎないことを常に肝に銘じなければならない。客観的（厳密）に測定しようと思えば思うほど、真実（真にとらえたい事象）から遠ざかっている可能性があることを理解しなければならない。

8. 「はかる（こと）」の本質

2005年10月にアメリカで「Walking for Health: Measurement and Research Issues and Challenges」というシンポジウムが開催され、身体活動量測定に関する様々なテーマが議論された。筆者は世界各国の測定評価学の専門家が集うこの場所に参加し、International Society for Aging and Physical Activity会長の Wojtek Chodzko-Zajko 博士と身体活動量計（例えば、歩数計や加速度計）について議論した際に、印象的なアドバイスを受けた。彼はどのような身体活動量計が最も望ましいかの判断基準は、測定の精度ではなく、どんな身体活動量計であっても、装着することによって、運動に対する動機づけがなされ、今まで運動していなかった者が運動をするようになればそれで良いのではないかという意見であった。

「はかる（こと）」あるいは「精度良くはかる（こと）」そのものは本質ではない。はかることによって健康の意識が高まったり、運動の継続が促進されたりするなどの行動変容をもたらす可能性があるのであれば、はかる必要があるだろうし、用いる状況に応じた意味のある測定精度が存在するのであって、測定精度が高ければよいというものではない。はかることが目的ではない。例えば、健康になるという目的のために「はかる（こと）」が必要な手段であるならば、はかるべきであり、はかることは目的を達成するための数ある手段のひとつである。これまで述べてきたように、完璧な測定は存在しないのであるから、客観化したい概念とのズレが生じたとしても概念を測定値として客観化する価値があれば、測定し、ある概念を主観的なまま用いることで目的が達成されるのであれ

ば、あえて測定する必要はない。運動を実践する際に、運動強度を把握するために心拍数や血中乳酸値を測定することがあるが、健康づくりのための運動実践においては、心拍数による運動強度設定よりも実践者の主観的な判断に頼るほうが妥当なときがある（田中，2003）、ということからもこのことは理解できる。

9. 「はかる（こと）」の必要性

体育・スポーツは、はかることによって科学として認められるようになったといっても良い。様々な「はかる（こと）」の問題や限界はあるにせよ、体育学、スポーツ科学が独立科学としての性格を確固たるものにするためには、「はかる（こと）」をしなければならない。さらに、他の研究領域と同様に体育・スポーツに独自の研究手法、測定技術、手法、道具を提供する分科領域を持たなければならない（松浦，1970）。心理学の分野では計量心理学（psychometrics）、経済学の分野では計量経済学（econometrics）、社会学の分野では計量社会学（sociometrics）という分科領域ができたことによって当該分野が発展してきた歴史があることから、体育学及びスポーツ科学の分野においても計量運動学（Kinesimetrics）といえるような分科領域が必要であると松浦（1970）が30年以上も前に提案している。近年、アメリカではイリノイ大学アバナ・シャンペーン校に Kinesimetrics^(注8) という名前の研究室が登場し、体育・スポーツ科学の計量学的研究が体系的に行われ始めたところである。日本には未だ計量運動学という学問領域は存在しないが、諸外国の現状を鑑みても、その学問領域の確立が急務であろう。そして、体育学及びスポーツ科学が独立科学として発展するために「はかる（こと）」が必要不可欠なのである。

10. 測定評価・統計リテラシー教育のススメ

これまでに述べてきた「はかる（こと）」に関わる諸問題は、測定そのものに帰属する問題

というよりも、測定値を解釈する側に帰属する問題である。「はかる(こと)」が必要不可欠である以上、解決しなければならない問題は測定値を解釈する側、すなわち、利用者に帰属する問題である。

『統計でウソをつく方法』(ハフ, 1968)、『統計のウソとホント』(キャンベル, 1975)、『統計はこうしてウソをつく』(ベスト, 2002)、『「社会調査」のウソ』(谷岡, 2000)、『「心理テスト」はウソでした。』(村上, 2005)、これらは、提示されたデータを鵜呑みにしてはならないということを警告する著書である。このような著書が現在もお売れ続けている背景には、コンピュータやインターネットの大衆化に起因する情報過多によって、情報の真贋を見抜くことが難しい社会になったことが考えられる。近年では、メディアリテラシー、情報リテラシー、リサーチリテラシー、健康リテラシーなど、さまざまな分野において、知識・教養を身につけようとする気運が高まっている。情報過多の現代社会においてまさに、必要不可欠な知識・教養となっている。測定評価学や統計学の分野においても同様な能力が要求される。それはこれまで述べてきたように、種々の測定から得られたデータの本質を理解し、適切に判断し、利用する能力や、データを使いこなすための能力であり、測定評価・統計リテラシーといえるような能力である。

測定そのものに帰属する諸問題については、研究者やテスト開発者が複雑な統計手法を駆使した論文で煙に巻くのではなく、テスト利用者に対する利用手引きとなる情報を提供しようとするスタンスを持つことによって改善させることができると考えられる。測定値を解釈する側に帰属する諸問題については、データの真贋を判断することが困難な現代において、情報(データ)の荒波に飲み込まれてしまわないためにも、テスト利用者が測定評価・統計リテラシーなる能力を身につけることによって改善させることができるのではないだろうか。

最後に、胚性幹細胞(ES細胞)研究におけるデータ捏造問題、中国製ダイエット用健康食

品「天天素」による健康被害、抗インフルエンザ薬投与が原因と噂される死亡事故などのデータや情報に関わる問題が話題になる現代社会を生き抜くためには、自己責任において、測定評価・統計リテラシーなる能力を身につけるといふことと同時に、測定評価学、統計学の研究者側が測定評価・統計リテラシー教育を提供するべきであり、その責任があることを認識する必要があるのではないだろうか。

謝辞

本稿で取り上げた内容の一部は本学総合人間・文化学部奥本正助教授との日頃の議論から生まれたものである。ここに感謝いたします。また、筑波大学名誉教授松浦義行博士には本稿を執筆するにあたり、有益な助言を頂きましたことを感謝いたします。

注

- (1) 機能と能力の関係については、Larson and Yacom (1951)に示されるような階層性が存在する。しかしながら本稿では、説明を簡単にするために、階層性を仮定しなかった。機能と能力の関係については、松浦(1983)に詳しい解説が示されているので参照されたい。
- (2) Larson and Yacom (1951)が定義する「sports motor skills (スポーツ技能)」と同義であり、サッカーに特異なスポーツ技能がサッカー技能である。
- (3) 奥野他(2000)では、男性において、BI法と皮脂厚法における顕著な差異はなく、女性において、肥満者に適用する場合には注意が必要であると報告している。したがって今回は、多くの対象者にはBI法と皮脂厚法の推定精度に違いがないと簡約し、引用したのであり、厳密にはBI法と皮脂厚法の推定精度には違いがある。
- (4) この例での問題は妥当性に関する議論というよりも、テストの精度、すなわち「信頼性」に関する誤解である。テストの適切性に関する概念は前述したテストの確からしさを意味する「妥当性」とテスト結果の安定性を意味する「信頼性」に分類される。誌面の都合上、ここでは触れないが、信頼性に関する詳細は、松浦(1983)を参照されたい。
- (5) 統計学における専門用語で、調査対象や研究対象となる集団全体のことを意味する。

- (6) 標本とは母集団を代表するものとして抽出されたデータである。
- (7) 統計学的には、誤差の大きさは標本誤差や平均値の信頼区間として求めることができる。南風原 (2002) などに分かりやすい説明があるので参照されたい。
- (8) 松浦 (1993) は Kinmetrics (カインメトリクス) という言葉を友人であるイリノイ大学の某教授に提案されたと述べているが、その単語が Kinesmetrics という名前となって具現化されたようだ。しかし、松浦 (1970) やイリノイ大学の某教授の提案するものとはスペルが違うことが分かる。少なくとも現在のアメリカでは Kinesmetrics が受け入れられている。松浦と筆者の私的な議論によれば、Kinesiology の Kin (metrics) に e を加えた kinematics (カインメトリクス) でも良いが、sio まで加えてしまうと -ology (学問、科学) の意味が残ってしまうのではないかと述べていた。また、ヨーロッパにおいては計量運動学といえる表現に最も近いといえるのは、anthropometry (人体測定学) から派生した kineanthropometry (運動測定学) であろうと述べていた。いずれにせよ、学問が体系化されつつある分野であることには違いない。

文献リスト

- ACSM (2005), *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription 5th Ed.*, Lippincott Williams & Wilkins, Baltimore
- AERA, APA, and NCME (1985), *Standards for educational and psychological tests*, APA, Washington, DC
- AERA, APA, and NCME (1999), *Standards for Educational and Psychological Testing*, AERA, Washington, DC
- ベスト (2002) 『統計はこうしてウソをつく』白揚社
- キャンベル (1975) 『統計のウソとホント』ダイヤモンド社
- Campbell, N.R.(1938), Symposium: Measurement and its importance for philosophy, *Proceedings of the Aristotelian Society 17*, Harrison & Sons, London
- Decker, J.I., and Mize, M.G. (2002), *Walking Games and Activities*, Human Kinetics, Champaign
- 福永哲夫 (2000) 『高齢者の生活機能増進法』ナッ
プ : 39-45
- 南風原朝和 (2002) 『心理統計学の基礎 : 統合的理解のために』有斐閣アルマ
- ハフ (1968) 『統計でウソをつく法』講談社
- Heymsfield, S.B., Lohman, T.G., Wang, Z., and Going, S.B. (2005), *Human Body Composition 2nd Ed.*, Human Kinetics, Champaign
- 池田央 (1971) 『行動科学の方法』東京大学出版会
- 池上晴夫 (1990) 『新版運動処方 理論と実際』朝倉書店
- 岩原信九郎 (1965) 『新訂版教育と心理のための推計学』日本文化科学社
- 岩脇三良 (1970) 『心理検査入門』日本文化科学社
- 金俊東, 久野譜也, 相馬りか, 増田和実, 足立和隆, 西嶋尚彦, 石津政雄, 岡田守彦 (2000) 「加齢による下肢筋量の低下が歩行能力に及ぼす影響」『体力科学』49 : 589-596
- Larson, L.A., and Yacom, R.G. (1951), *Measurement and evaluation in physical health and recreation*, The C.V. Mosby, St. Louis
- 松浦義行 (1970) 「計量運動学 (Kinesiometrics) の提案」『体育の科学』20 : 643-645
- 松浦義行 (1983) 『体力測定法』朝倉書店
- 松浦義行 (1993) 『数理体力学』朝倉書店
- 文部科学省 (2000) 『新体力テスト : 有意義な活用のために』ぎょうせい
- 村上宣寛 (2005) 『「心理テスト」はウソでした。』日経 BP 社
- 大河原一憲, 田中喜代次, 中田由夫, 李東俊, 魏丞完, 中塘二三生 (2003) 「単周波数および多周波数 BI 法における身体組成評価の比較検討」『体力科学』52 : 443-454
- 奥野淳, 中塘二三生, 大蔵倫博, 田中喜代次 (2000) 「簡便法による体脂肪率の推定精度に関する検討 : 4 種類の BI 法及び皮脂厚法の比較」『体育学研究』45 : 252-261
- 繁樹算男 (1998) 『心理測定法』放送大学教育振興会
- Spiriduso, W.W.(1995), *Physical Dimensions of Aging*, Human Kinetics, Champaign
- Stevens, S.S.(1951), *Handbook of Experimental Psychology*, John Wiley, New York
- ストラザーン (1997) 『90 分でわかるヴィトゲンシュタイン』青山出版社
- 田中喜代次 (2003) 『平成 14 年度日本体育協会スポーツ医・科学研究報告 No. II 中高年者の運動プログラムに関する総合的研究 第 1 報』日本体育協会

谷岡一郎 (2000) 『「社会調査」のウソ』文藝春秋
 ラッセル (1960) 『バートランド・ラッセル著作集
 10 人間の知識Ⅱ』みすず書房

WHO QoL Group (1993), Study protocol for
 the world health organization project to
 develop a quality of life assessment in-
 strument (WHO QoL), *Quality Life Re-
 search*, 2: 153-159

Zeller, R.A., and Carmines, E.G. (1980),
*Measurement in the Social Sciences: The
 Link between Theory and Data*, Cam-
 bridge University Press, New York

付録 1

代表的な計量学者らが定義する測定と評価につ
 いて以下に示した。

Stevens (1951) :

「測定とは、ルールに従って対象物 (objects)
 や事象 (events) に数値を付与することである。」

ラッセル (1960) :

「もっとも一般的な意味で、量の測定とは、ある
 種の量と数の間に 1 対 1 の対応をつける方法のこ
 とである。」

岩原 (1965) :

「Campbell (1938) によれば測定とは一定の法
 則に従って数字を対象または事件に当てがうこと
 である。しかし数字をわれわれが実際に問題にし
 ている対象または事件にあてがうためにはこの二
 つの体系の間に同型写像的 (isomorphic) 対応
 が必要である。もしかかる同型写像が妥当しない
 ならば測定は意味がない。同型写像とは測定され
 る対象が数学における論理的体系 (の一部) と対
 応するような特性を有することで、(以下省略)」

岩脇 (1970) :

「広義に解釈すれば、ある尺度上に、ある位置を
 与えることが測定である。(中略) 測定にとって本
 質的なことは、ある標準があって、その標準を基
 礎にしてものごとを比較する規則があり、ものご
 とに数値を割り当てることである。」「一般に、あ
 るできごとやあるものごとの価値を決定するた
 めに用いられる手続きを称して評価という。」「測定
 は客観的事実を数量的に表すもので、『どれだけ
 の……』という問に答えるものである。評価は、
 到達される目標を念頭において、『どんな価値があ
 るか』という問に答えるものである。」

池田 (1971) :

「測定とは、対象の属性に対して比較可能な数値
 を与えることで、与えられた数値を測定値あるい
 は実測値という。」

Zeller and Carmines (1980) :

「抽象的な概念を経験的指標 (empirical
 indicants) に関連づける過程である。」

繁榊 (1998) :

「測定とは測定対象に対してある正の実数を対応
 させる操作である。(一般的でない言い方をすれ
 ば、測定値は、実数に限らず、ベクトル、複素数
 その他の数学的概念を含むがここでは簡単のため
 にそれぞれの性質が、正の実数によって表現され
 るとする。)」