

測定についての一考察

—認知と筋運動機能との関連実験の資料から—

八田 武志 (関西福祉科学大学 健康福祉学部, hatta@fuksi-kagk-u.ac.jp)

What is a proper scale for cognitive assessment in psychological studies?:

A consideration on scientific measurement based upon an experimental data

Takeshi Hatta (School of Health Sciences, Kansai University of Welfare Sciences, Japan)

要約

顔面筋肉の運動を操作し、顔の模式図の表情を評価する実験でのデータを用いて、人間の認知行動の適切な測定・評価はどうあるべきかの問題を検討した。実験対象者は鉛筆を歯でくわえる群、鉛筆を唇で保持する群、及び統制群の3群にランダムに配分され、顔の模式図の表情をポジティブからネガティブまで7件法と17件法で評価することが求められた。7件法の結果と17件法での結果は異なるものとなった。このような実験結果に基づいて心理学的尺度における認知評価尺度はどうあるべきかを考察した。

キーワード

測定, 尺度, 不思議の数7, 認知評価, 妥当性と信頼性

1. はじめに

科学・技術研究の根本は測定である。測定とは、何らかの属性に数値を割り付ける手続きであると定義できよう。当然のことながら、数値だけで科学・技術研究がその内容を深化できるわけではなく、属性に割り付けたその数値を価値基準に照らし合わせる評価というプロセスを加味することで研究は実行でき、進展する。したがって、測定と評価は現代の科学・技術の根幹であると言うことができ、このことには異論が生じる余地は少ないはずである。

このような、科学的手続きを用いることが近代以降の自然科学、そしてその影響を受けてのちに人文科学や社会科学にも適用されるようになった。これらの信念が、現代に生きる人間の共通理解・時代精神となっている。

グローバルな観点から見ても科学的手続きを共通言語として様々な科学・技術分野での議論が進んでいくと言えよう。

このような考え方を基本的に共有するとしても、人間行動の様々な側面に測定、すなわち数値を割り付ける作業の実際は簡単ではない。とりわけ情動、性格、知能、学力、適性などの構成概念を共通言語で語り合うことには様々な制約が伴う。

「こころ」をあつかう心理学は1879年にライプツヒ大学に実験心理学教室を開設したウイルヘルム・ヴントを現代心理学の創始者と見なしている。彼は人間行動を如何にして自然科学的な測定をもって取り扱うかに苦心し、一定の方向性を明らかにしたことが創始者として歴史的に重要な人物として顕彰されることにつながっている。

心理学史を繙くと、ヴントの実験心理学の主流は1910年代から50年間ほどに渡る行動主義心理学の時代を経て、今日の認知心理学へとつながったと指摘できる。認知心理学の研究者として世界中で何十万人という人間が大学や研究所等で

生活の糧を得ているが、人間行動を科学的に測定・評価することまつわる問題がすべて解決済みというわけではない。

そこで、本稿では現時点でも人間行動を科学的に取り扱う上で内包する問題は残されたままであり、未解決ではないのかという疑問を投げかけようと企図した。

具体的には、Cretent and Dru (2009) および Dru and Cretenet (2011) の2つの論文に刺激されて行った行動実験を基礎に、心理学的測定における尺度構成の問題を取り上げ、測定とは何か、どうあるべきなのかを再考することが目的である。

Cretent and Dru (2009) の実験は実験心理学分野で最も権威があると見なされる、アメリカ心理学会発刊の *Journal of Experimental Psychology* 誌に掲載されており、その影響力は看過できないものがある。ただ、もう一つの Dru and Cretenet (2011) の2編を読み比べてみると、使用されているデータは同一のものである可能性が疑え、2重投稿の疑義が出そうな論文である。しかし、この種の倫理性を議論するのが本稿の主目的ではないので、これ以上は言及しない。

彼等の実験は2編とも5実験から構成されているが基本的な検討課題は一つであり、筋運動が認知機能に影響を与えることの立証を目指している。そこでは2種の手の運動動作と認知課題を同時に遂行させている。たとえば、手のひらを机に押しつける動作や手のひらを引き上げる動作と顔の模式図の表情を評価させるという具合である。このときの顔の模式図表情の評価に17件法が用いられている。この実験の背景や仮説の詳細は後述することにして結論だけ記載すると、仮説に見事に合致する結果を5つのすべての実験で得ている。

この筋運動が認知機能に与える影響の検討課題は、最近のホットトピックの一つでもあるが、Strack et al. (1988) によれば、筋運動が認知機能に影響するかの問題の端緒は Darwin (1872) に遡るもので、Darwin は人間の情動経験は表情筋の活動により強めることも弱めることもできるとした。

この古典的問題は、1980年頃の Buck をきっかけに、質

問誌法、記憶実験、自律神経系の反応を用いる生理心理学的な方法で様々な研究により検討された。有名な Zajonc et al. (1989) の研究は、ドイツ語のüという口をすぼめて発音する母音を多用したテキストと、そうでないテキストを実験参加者に読ませたものである。üは口をすぼめて発音するため、しかめ面に似た表情の動きをすることになり、üを多用したテキストを読む間のほうが参加者は否定的な感情を示した。これは皮膚温を指標にしているため、生理心理学研究に分類できる。

同一のテーマに簡単な実験を用いて質問紙手続きでこの問題に取り組んだ研究に Strack et al. (1988) がある。彼らは、実験対象者に漫画のおもしろさについての評価を2つの条件下で求めている。一つの条件は鉛筆を歯でくわえる条件であり、もう一つの条件は鉛筆を縦にしてその先を加える条件である。前者は笑うときの口角が上がった表情筋を必然的に生じさせる条件であり、後者は不平を言うときの口先と咄める表情筋を生じさせる条件として設定している。結果は、同じ顔の図を見ても鉛筆を横に加えた条件では、漫画をよりpositiveなものに感じたとして報告した。鉛筆を縦にくわえた条件ではnegativeな評価へと認知評定は移行したというものである。これらから、筋肉の運動慣性が情動生起時の認知にバイアスを生じさせると解釈した。この認知評価には10件尺度による測定が行われている。

以上の評定尺度を指標とした研究での測定には、Grue and Cretent (2011) では17件法、Strack et al. (1988) では10件法が用いられ、評定幅が広いのが特徴である。一方で、一般的な社会心理学での通常の調査研究では、質問紙での意見や認知の測定は「はい・いいえ」の2件法、「はい・どちらでもない・いいえ」の3件法のような評定幅の狭いものが一般的である。

果たしてどのような尺度構成の測定が適当なのであろうか。本稿は Strack et al. (1988) の実験を追試する形式を採り、尺度構成により異なる結論が導き出されることを検証し、心理学的測定の在り方について考える。

2. 実験

2.1 対象者

実験の対象者となったのは、男子大学生24名と女子大学生21名の合計45名である。年齢は20歳から31歳までの範囲に入り、視覚をはじめとする感覚能力に特別な困難を持つものは含まれていない。対象者はランダムに鉛筆を横にして歯でくわえる条件群に15名(男子8名、女子7名)、鉛筆を唇で縦にくわえる条件群に15名(男子8名、女子7名)、鉛筆を加えるなどの負荷条件のない統制群に15名(男子8名、女子7名)を配分した。

歯でくわえる群は Strack et al. (1988) に倣ったもので顔面の口角が必然的に上がり、笑う表情にともなう筋運動を誘導するためであり、口でくわえる群は不平で口をすぼめる表情にともなう筋運動を誘導するためである。表情は文化があることは当然であり、「Yes」と言う口角が上がる肯定的言語表現や「No」と言う口角が上がらず唇をとがめる否定的言語表現と同じように、日本人が笑う表情で口角を上げたり、否定的な感情の際に口をすぼめたりするのが一般的であるかは疑問

がないわけではないが、本実験の主たる狙いは別にあるので問題とはしないことにする。

2.2 刺激材料と手続き

実験ではパソコン画面上に提示される漫画の顔43種の刺激図に対して表情の評価を7件法と17件法の2種のやり方で求めた。

顔刺激図は Grue and Cretent (2011) の実験1で用いたものを使用した。図1はその刺激図の例である。

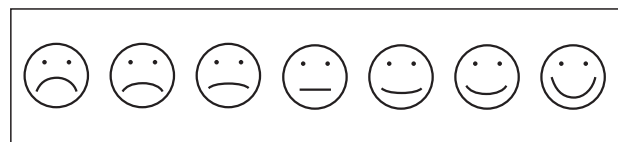


図1：顔刺激

手続きの説明を受けた対象者はパソコンに面して座り、パソコン画面上に1つ提示される刺激図にたいして評価を行うのが課題であった。刺激図は7秒間提示され、つづいて回答を促す指示が画面に呈された。

対象者は1ページごとに回答欄が印刷されているブックレットの回答欄に、楽しい表情であるか悲しい表情であるかを両極端とする尺度上の、最も適切と考える個所に○印を記入する課題を行うように求められた。7件法の場合には中央の値を0に左右それぞれに3つの目盛りを記入し、両端に、楽しいまたは悲しいの表記が印刷されていた。したがって、評定値は最大で±3となる。一方の17件法では中央の値を0とし、左右にそれぞれ8つの目盛りを記入し両端に、楽しいまたは悲しいの表記が印刷されていた。したがって、評定値は最大で±8となる。

実験者は対象者の右横の位置に座り、回答が記入されると、次に進んでよいかを確認後、パソコン画面を次の画面へと切り替えることを行った。したがって、対象者ペースで刺激提示が行われたと見なすことができよう。

本実験では対象者は同一の刺激図に対し7件法での評価と17件法での評価の両方を行ったが、両者はランダムな順序で提示するように構成されていた。ただし、刺激図の提示順序は被験者間でランダム化はされておらず、対象者は同一の提示順序で合計86回刺激図を見たことになる。

3. 結果

本実験の目的は7件法での測定と17件法での測定結果を比較することである。

性別に7件法での評定結果を示したのが表1である。表1の結果を男女と条件を要因とする2要因分散分析で効果測定を検証したところ、性差については $F(1, 39) = 0.72$ であった。有意確率は0.40であり、差異は認められなかった。つまり評定に性差はないということになる。条件の要因については、 $F(2, 39) = 4.19$ であり有意確率は0.023と有意となった。つまり、鉛筆のくわえ方の要因により評定値は異なることが明らかとなった。ただし、両要因の交互作用は $F(2, 39) = 0.06$ で

表1：性別にみた3つの実験条件での平均評定値と標準偏差

		男子	女子
歯で横にくわえる条件	M	0.12	0.03
	SD	0.32	0.11
唇で縦にくわえる条件	M	0.22	0.16
	SD	0.22	0.17
統制条件	M	0.32	0.31
	SD	0.11	0.29

あり有意ではなかった。

次に17件法での評定結果を表2に示す。表2の結果を男女と条件を要因とする2要因分散分析で効果測定を検証した。性差については $F(1, 39) = 3.99$ であり、有意確率は0.053であり、差異の存在が示唆されること、つまり評定に性差がある可能性が高いことになる。条件の要因については、 $F(2, 39) = 21.569$ であり有意確率は0.001未満の数値なり有意となった。つまり、鉛筆のくわえ方の要因により評定値は顕著に異なることが明らかとなった。両要因の交互作用は $F(2, 39) = 5.01$ であり、有意確率は0.0012で有意となった。

表2：性別にみた3つの実験条件での平均評定値と標準偏差

		男子	女子
歯で横にくわえる条件	M	-0.30	-0.96
	SD	0.52	0.22
唇で縦にくわえる条件	M	-0.07	-0.17
	SD	0.24	0.41
統制条件	M	0.16	0.29
	SD	0.31	0.27

この2つの分散分析の結果は、統計的に異なる様相を示すことが明らかである。つまり、尺度の幅が異なると違う結論が得られることを示している。異なることをより分かり易く表すための7件法と17件法の結果を図に示すと図2のようになる。

4. 考察

図2から分かるように、同じ刺激材料についての感情評価を2つの異なる尺度評価法で同一の対象者に実施した結果は同じものではなかった。

7件法での結果からは性差は認められないこと、鉛筆を2つの形態でくわえる条件にも違いはないことが明らかとなり、統制条件と実験2条件の間には評定値に違いがあるという結果となった。これに対して、17件法では2つの実験要因の両方が有意な効果を示し、くわえて両者の交互作用も有意となる統計処置の結果であった。通常実験計画法では2要因の計画が最も基本的で2つの要因の間に交互作用が有意に立

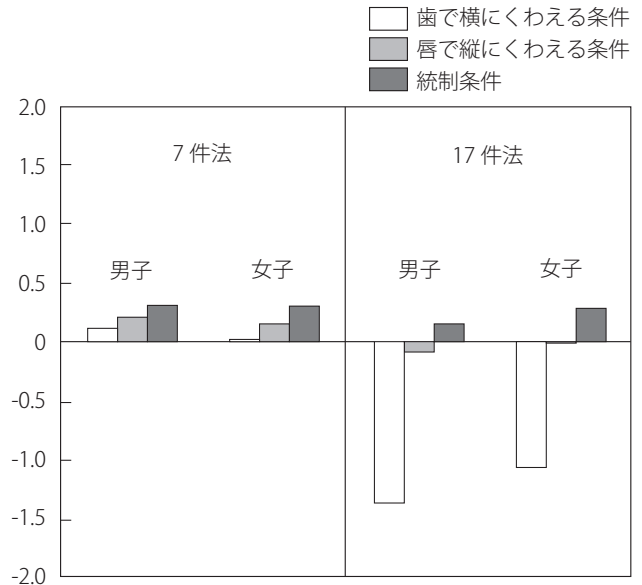


図2：2種の評定尺度での結果のまとめ

証できるか否かが重要な問題となる。つまり、ある条件ではそうだが、別の条件ではそうではない、という結論が得られるかどうかを検証するのが実験計画法の中心的な狙いである。

近年3要因や4要因、5要因の分散分析のソフトウェアが統計処理パッケージに加えられている場合があるが、実験計画法の基本的な考え方からは問題が多い。数次の交互作用についてそれが有意となった場合の解釈は不可能と考えられるからであるが、この問題も本稿の目的から外れるのでこれ以上は言及しない。

さて、本稿のような2つの異なる尺度により得られた測定結果が異なる統計処理結果を生み出した場合に、17件法が7件法よりも測定値の感受性が高い、したがって17件法が優れていると結論を出してよいのであろうか。

そこには、古典的な極めて重要な問題を内包していると考えられるのではなからうか。

たとえば、Millerの有名なMagical number seven plus minus twoを想起すれば、感情評価を17段階にも分けて正確に評価可能なのか、測定値に信頼は置けるのか、という疑問が湧くことであろう。

本稿の端緒となったCretenet and Drue (2009)の実験の具体を紹介してこの種の疑問を考えてみよう。

彼等は上肢の運動が認知判断に影響するかを検討する目的で、腕を伸ばす(extension)動作下と腕を曲げる(flexion)動作下の2つの条件に認知評価判断を2重課題として実行させている。Extension動作は拒否に付随する動作であり、Flexion動作は受容に付随する動作というのが前提である。つまり、嫌ですと腕を突き出す動作、おいで、おいでと腕を畳む動作であるというわけである。認知課題にはBreskin Rigidity Test (Breskin, 1968)を用いている。この検査は視覚的刺激物の見え方の心理的な硬さを測定しようとするパーソナリティ検査項目の一部である。

Drue and Cretenet (2011)では、認知課題にnegative、

neutral, positiveな情動を喚起するとされる写真(International affective picture system, Lang et al., 2005)を追加しているが、基本的なパラダイムはCretenet and Drué (2009)と変わらない。両方の論文で、Extension動作下とFlexion動作下で2重課題として刺激図を見せ、その情動価を対象者に評定させてそれを測定値としている。彼等は運動動作が右手か左手かを変数に加えてラテラルリティ研究として神経心理学的研究としている。右脳がnegative情動機能に関係が強く、左脳はpositive情動機能に関係が深いという1970年代後半に一時話題となったことはあるが確実な知見とまでは定まらないままの仮定に基づく実験計画なのである。更にいえば、Extension動作が拒絶でnegative、Flexion動作がpositiveという仮定も合理性があるかは疑問がないわけではない。

彼等の実験はすべての認知課題で、(右手・左手) × (Flexion・Extension)の見事に交差する交互作用を示す図を得ている。しかし、ここで指摘せねばならないのはすべての認知課題での評価が17件法であるという点である。

筆者はこれまでにたくさんの学術論文(実験研究論文が大多数であるが)を眼にしてきたが、鼻聞にして17件法による評定が行われてきた例を知らない。筆者が本稿で例示したように17件法では交互作用がでる結果も7件法であれば交互作用効果は認められないことがあるとなれば、DruéとCretenetによる実験結果が必ずしも堅牢で顕著な人間の心理学的行動かは疑わしいということになる。感受性の高い17件法を使い、7件法を使わなかったのはその研究者が優れていると賞賛されるべきなのだろうか。

統計的に有意な結果が得られたことが実験を計画する上で重要な要素であることは言うまでもないが、一般的に認知心理学実験で用いられている5件法や7件法では有意差が出ないので、感度を良くするために17件法を用いて評定を求めるとするのは、恣意的ではないかという指摘を受けても仕方がないのではないかと考えるのである。

研究者が仮説する主張に沿う結果が捻出できる方法を探すというのは、科学研究を進めるといって現代の研究者の使命及び倫理に合致するのだろうか。

筆者は現在の認知心理学や社会心理学で採用される測定法が安易すぎる危険を孕んでいるのではないかと心配するのである。この種の安易さを内包する測定を頻繁に使い、堅牢でない結果を社会に喧伝することが続けば、心理学に対する社会からの見方が変わり、信頼できない科学性に乏しい学問領域であると見なされるようになる可能性を危惧する。そのようなことは、1879年のヴントの実験心理学教室の創設に始まる心理学の現代化(すなわち、自然科学化)のベクトルを台無しにすることにつながるわけである。

このような問題に直面するときその解決法につながる示唆は原点に戻ることであろう。この考え方は心理学だけでなく様々な社会的出来事においても共通することで、歴史に学ぶことが重要なのではないかと考える。

心理学で用いる構成概念を科学的に処理する解決法として尺度構成がすでに1950年代に欧米だけでなく我が国にも紹介されている。Guilfordの大部の著書Psychometric Methodsはその代表的なものであろう。我が国では九州大学心理学

教室創設30周年記念事業としてその訳出が行われており、Guilfordの功績を容易に参照しうる。

Guilfordの本の初版は1936年に出たと記載がある。そして、科学的心理学の始まりである、心理学徒なら誰もが知るウェーバー・フェフナーの法則で有名なFechnerの著書Elemente der Psychophysikで知覚や思考の実験的研究にもっぱら精神物理学的測定法を用いていた時期に、知覚研究分野の精神物理学的測定法と知能指数や人格研究に用いられるテスト法をつなぐことをしたのがGuilfordの著書であると紹介している。つまり、実験的方法と尺度法やテスト法のような統計学的方法の間に緊密な論理的関係を包摂したということであろう。

このGuilfordの邦訳の11章には評定尺度法がまとめられている。そこでは、評定尺度の形式の紹介や効用と限界の記載がある。たとえば、「負の評定値を用いることはすすめられない」、「数値による評定は、多くのかたよりや誤差に陥りやすいと考えられる」、「強制選択法の理論」などが記されている。つまり、「人間の行う量的判断には多くの信頼を置くように、評定尺度は、強いられるが、そこに入り込む弱みや、その判断に置ける個人的かたよりの多くの要因に対して、常に気を配っていなければならない」としている。

現代の科学・技術に関わる者は少なくとも人間行動を科学的手続きで測定することを了解し実行すべきであろう。

現時点でDruéとCretenetが発表している実験結果を否定できる根拠は筆者にはない。ただ、疑念を呈するだけである。医学の分野で用いられる痛みなどの感度の測定法であるVAS(Visual Analogue Scale)では、0から10までの線分に対象者の主観的痛みをマークしてその位置を物差しで測定してその値を測定値にするなどの方法と比べて17件法での尺度評価がそんなに問題があるのかと言われるとその反論への解答は持ち合わせてはいない。

したがって、今のところ筆者が言えるのは、何かの研究結果が示されたときに、それがどのような手続きで行われ、そこで用いられている測定には信頼性や妥当性が保証されているのかどうかを疑いつつ接することが望ましいということだけである。

謝辞

本稿に用いた行動実験での数値資料は水本和明君による、2011年度卒業研究「感情と顔面運動の関係性」を用いている。記してお礼を申し上げる。ただし、本稿で論じている議論や考察は卒業論文には含まれている内容ではない。

引用文献

- Breskin, S. (1968). Measurement of rigidity, a non verbal test. *Perceptual and Motor Skills*, Vol. 27 1203-1206.
- Darwin, C. R. (1872). *The Expression of Emotions on Man and Animals*. John Murray.
- Cretenet, J., and Drué, V. (2009). Influence of peripheral and motivational clues on rigid-flexible functioning: Perceptual, behavioral, and cognitive aspects. *Journal of Experimental Psychology: General*, Vol. 138 201-217.

- Dru, V., and Cretenet, H. (2011). Performing lateralized approach and avoidance behaviors: Effects on perceptual, affective and confidence judgement. *Journal of International Neuropsychological Society*, 17, 289-294.
- Gilford, J. P. 1956 *Psychometric Methods* (秋重義治監訳. 精神物理学の測定法. 1959, 培風館).
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2005). *International Affective Picture System (IAPS): Digitized photographs, instruction manual and affective rating. Technical Report A-6*. Gainesville, University of Florida.
- Strack, F., Martin, L., and Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, Vol. 54, 768-777.
- Zajonc, R. B., Murphy, S. T., and Inglehart, M. (1989). Feeling and facial expression: Implications of the vascular theory of emotion. *Psychological Review*, Vol. 96, 395-416.

(受稿：2012年5月1日)