

心理テストは信用できるのか

早稲田大学文学学術院 教授

小塩真司 (おしお あつし)

Profile—小塩真司

2000年、名古屋大学教育学研究科博士後期課程修了。中部大学人文学部准教授、早稲田大学文学学術院准教授を経て、2014年より現職。博士（教育心理学）。専門はパーソナリティ心理学、発達心理学。著書は『人間関係の生涯発達心理学』（共著、丸善）、『Progress & Application パーソナリティ心理学』（サイエンス社）、『はじめての共分散構造分析 第2版』『大学生ミライの統計的日常』（いずれも東京図書）、『性格を科学する心理学のはなし』（新曜社）、『自己愛の心理学』（共編、金子書房）、『パーソナリティ心理学ハンドブック』（共編、福村出版）など。



サイコパス診断

インターネットに出回っている、次のような問題をご存知だろうか。もし知らないようであれば、読んで答えてみてほしい。どうして女性は妹の命を奪ってしまうのだろうか。あなたならどのような理由を考えるだろうか。

ある女性が妹と一緒に母親の葬儀に出た。そこでこの女性は、ある男性に出会った。彼女は、この男性が自分の運命の人だと確信し、一目惚れをした。しかし彼女は、この男性の連絡先を尋ねることをせず、葬儀が終わったあとも連絡を取ることができなかった。そして数日後、この女性は妹を殺した。どうしてだろうか？

さて、この簡単な問題に答えることで、あなたがサイコパスであるかがわかるのだという。サイコパスとはサイコパシー（精神病質）傾向の強い人物のことであり、サイコパシーとは、良心や恐怖心の欠如、他者からの評価を気にしない態度、非情さや衝動性などの諸特徴をもつパーソナリティ（性格）傾向のことである。

インターネット上の解説には、「もしもあなたがサイコパスならば、『妹を殺せばまた葬儀の際にその男性と会えると思ったから』と答えるはず」と書いてある。なぜならサイコパスは、自分の目的を達成するためなら非情になり手段を選ばないからだという。さてあなたの考え方は、このようなものだったでしょうか。それとも違う考え方だったでしょうか。

この問題と同じ内容のものは、海外でも知られている。イギリスの心理学者ケヴィン・ダットンはこの問題に興味を抱き、本物の複数のサイコパス（様々な犯罪を犯し、かつ専門家の検査によってサイコパスと診断された人々）にこの問題を出してみた。その結果、「妹が死ねばもう一度葬儀をしてその男性に会えるから」と答えた人はひとりもおらず、ほとんどが「恋愛関係のもつれ」が動機だと回答したのだという（ダットン, 2013）。ダットンの著書には、この問題を出题されたサイコパスの一人が次のようにコメントしたことが書かれている。「おれは正気じゃないかもしれない……だけど、ばかじゃないぜ」（ダットン, 2013 p.64）。

1項目で測定する

では、次の質問に答えてみてほしい。あなたは回答欄に、何点を記入するだろうか。

・あなたは次の文章にどの程度同意するでしょうか：「自分はサイコパスである」

（注：「サイコパス」とは、誰かに迷惑をかけても良心がとがめることがほとんどなく、何事にも恐怖心を示さず、他者の心の痛みを感じないことを意味する）

・「1.全く当てはまらない」から「7.非常によく当てはまる」まで、1から7のいずれかの数字で答えてください。

・あなたの回答：（ ）

これは、得点が高いほどサイコパシー傾向が強い人物だということになる……はずである。ただしこの質問は、実際に心理学の研究で使用されているものではない。したがって、1点から7点のうち何点と回答したからといって、回答者のサイコパシー傾向が確実に把握できるわけではないのであしからずご了承ください。

「たった1つの質問で、サイコパシーのような複雑なパーソナリティ傾向など測定できるはずがない」という感想をもつ方もいるだろう。ところが、これと全く同じ形式の質問で、「ナルシズム」（自分自身に過剰に自信をもち、自己中心的でうぬぼれやすい傾向）の程度を測定する尺度が開発されている（Konrath, Meier & Bushman, 2014）。その尺度はSingle Item Narcissism Scale (SINS) と名づけられ、尺度の開発過程が学術専門雑誌に掲載されている。SINSは、先の枠内の「サイコパス」の部分で「ナルシズム」に変え、括弧内の注をナルシズムの説明に変えただけの尺度である。ナルシズムもサイコパシーに負けず劣らず、複雑な機能と構造からなるパーソナリティ特性である。ところがその複雑なパーソナリティ傾向を、たった1項目で測定する尺度が作成され、それが堂々と学術専門雑誌に掲載されているのである。

超短縮版尺度が開発される理由

表1にいくつかの例を示しているが、これまでに、非常に数少ない質問項目で心理学的概念の測定を試みる尺度の開発が盛んに行われている。日本でも近年、自尊感情を2項目で測定するTISE (Two-Item Self-Esteem scale; 箕浦・成田, 2013) や、Big Fiveパーソナリティを10項目で測定するTIPI-J (小塩・阿部・カトロニ, 2012) など、超短縮版と呼ぶにふさわしい尺度が開発されてきている。このような超短縮版尺度が開発される理由とは何だろうか。

第一に、調査場面の広がりや多様化である。心理学の調査フィールドはより広く、大規模になりつつある。広い年代を対象としたアンケート調査やインターネットを介した調査、疫学的

な調査や長期的なコホート調査などでは、多くの質問項目が同時に実施される。そのような中、回答時間も質問が占める空間も少ない超短縮版尺度の重要度が増してくる。

第二に、省スペースで測定できるということは、探索的に他の変数との関連を検討したい場合に使用しやすいことを意味する。これはいわば「ついでに測定しておく」ということであり、本来の研究のあるべき姿ではないかもしれない。しかしたとえば、探索的に検討した変数間の関連から、新たな研究が進展するかもしれない。また、ある変数を統制するため「念のために測定しておく」という使い方も、超短縮版の尺度の場合には利用しやすいだろう。

表1 超短縮版尺度の例

測定概念	項目数	研究・尺度名
自尊感情	1項目	<i>Single-Item Self-Esteem Scale</i> (SISE; Robins, Hendin & Trzeniewski, 2001)
抑うつ	1項目	Watkins, et al. (2001) ; Skoogh et al. (2010)
生活満足度	1項目	Schimmack & Oishi (2005)
幸福感	1項目	Abdel-Khalek (2006)
不安	1項目	数学不安: <i>Single-Item Math Anxiety scale</i> (SIMA; Nunez-Pena, Guilera, Suarez-Pellicioni, 2014) 歯科不安: <i>Dental Anxiety Question (DAQ; Neverlien, 1990)</i>
自己愛	1項目	<i>Single Item Narcissism Scale</i> (SINS; Konrath, Meier & Bushman, 2014)
Big Five パーソナリティ	5項目	<i>Five-item measure of the Big Five</i> (Aronson, Reilly & Lynn, 2006) ; <i>Your Personality Form</i> (Bernard, Walsh, & Mills, 2005) ; <i>Single-Item Measures of Personality</i> (SIMP; Woods & Hampson, 2005)
	10項目	<i>Ten Item Personality Inventory</i> (TIPI; Gosling, Rentfrow, & Swann, 2003) ; <i>BFI-10</i> (Rammstedt & John, 2007)

第三に、簡便な尺度であることによって、特殊な状況下での測定に有効になる場合がある。たとえば、ある得点の1日の変動を細かく測定するような、同一人物が繰り返し回答を行う調査である。またたとえば、複数の対象を次々と評価するような調査・実験である。コンピュータスクリーン上に測定対象（人や物）が次々と現れ、それぞれについてある基準で評価を行っていくと、必然的に回答する項目数が増加してしまう。このような場合にも、超短縮版尺度が有用となる。

妥当性

心理学では、直接的に観察が困難な構成概念を扱う。サイコパシーもナルシズムも構成概念であり、そのものを直接見ることはできない。同じように知能も学力も外向性も運動能力も構成概念であり、直接把握することはできないものである。それは、明るい性格を、物理的な明るさ（照度）で測定することができないのと同じである。そのような文脈において重要な考え方のひとつが、妥当性（validity）である。妥当性とは、測定された内容が構成概念を適切に反映しているかという問題のことである。

尺度の妥当性検討でよく用いられてきたのは、内容的妥当性（測定された内容が概念を過不足なく反映しているか）、基準関連妥当性（測定された内容が外部基準に関連したり、予測したりするか）、構成概念妥当性（測定された内容と他の変数との関連が概念間の構造を反映しているか）という三つである。しかし近年では、構成概念妥当性こそが統合的な妥当性の考え方であり、構成概念妥当性を確かめるために様々な証拠が検討されるという考え方がとられるようになってきている（Messick, 1995; 村山, 2012）。概念の構造が尺度の因子構造に反映されているか、他の概念との間で予想される関連が実際に観察されるか、将来生じる結果を実際に予測できるか、他の集団でも妥当な結果が生じるかなど、一口に妥当性の検証と言ってもその内容は多岐にわたる。決して、ある決まった手順通りのことをすれば、妥当性の検証

が終わるというものではない。

違いはどこに

どうして冒頭のサイコパス診断は心理学の研究で使われることがなく、それよりも簡単に見える超短縮版尺度は論文に掲載され、心理学の研究で「使用可能」となってしまうのだろうか。サイコパス診断については妥当性について書かれた論文が見当たらず、ダットン報告によればこの診断は本物のサイコパスを見分けることが難しそうなので妥当性にも疑問が残る。それに対し、SINSは多岐にわたる研究から妥当性が検討されて、その内容は論文に掲載されている。つまり両者は、「妥当性を検討しているかどうか」そして「妥当性の情報が公開されているかどうか」という2点が大きく異なるのである。

しかし逆に言えば、違いは「そこにしかない」とも言える。もしかしたら、インターネット上に広まっている心理ゲームの中には、検証してみたら非常に高い妥当性を示すものがあるかもしれない。ただ、誰もその妥当性を検討していないから「真偽不明」なままとなっているにすぎない。逆に、心理学の尺度の中にはそれほど高い妥当性を示さないものがあるかもしれない。ある概念を測定する心理ゲームの妥当性を調べ、それと心理学の研究で用いられる、同じ概念を測定する尺度・検査と妥当性を比べてみてはどうだろう。はたして専門家が作成した尺度は、その対決に勝利することができるのだろうか。

完璧な妥当性はあるのか

もしもある検査が完璧な妥当性をもっていて、完璧に誤差なく測定できたらどうなるだろう。たとえば1つの質問項目に答えるだけで、その人がナルシストかどうかを完璧に判定できる。たとえ本人が否定しても無駄である。何せ完璧に測定できるのだから。見た目がどうだろうと本人が否定しようと、検査の結果がナルシストならその人はナルシストなのである。

そんな質問項目の存在は、非常に危険である。そんな危険な質問項目は、専門家の管理のもとに厳重に管理すべきである。それはあたかも、イギリスのコメディ・グループ、モンティ・パイソンのコントに出てくる「殺人ジョーク」のような扱いになってしまう（小塩, 2011）。

現実には、病理的傾向を測定する検査など管理下に置かれているものが中にはあるものの、研究で用いられる尺度や検査ではここまで厳重な管理を要求されるものはほとんどない。心理学者が尺度を開発し、妥当性を検討する究極的な目標はどこにあるのだろうか。完璧な妥当性を備えた検査の開発を目指しているのだろうか。それとも、そんなことは無理だと思いつつ、完全な勝利のない戦いを続けているのだろうか。

文 献

Abdel-Khalek, A.M. (2006) Measuring happiness with a single-item scale. *Social Behavior and Personality: An International Journal*, 34, 139-150.

Aronson, Z.H., Reilly, R.R. & Lynn, G.S. (2006) The impact of leader personality on new product development teamwork and performance: The moderating role of uncertainty. *Journal of Engineering and Technology Management*, 23, 221-247.

ダットン・K. / 小林由香利 (訳) (2013) 『サイコパス：秘められた能力』NHK 出版

Bernard, L.C., Walsh, R.P. & Mills, M. (2005) Ask once, may tell: Comparative validity of single and multiple item measurement of the Big-Five personality factors. *Counseling and Clinical Psychology Journal*, 2, 40-57.

Gosling, S.D., Rentfrow, P.J. & Swann, W.B.Jr. (2003) A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504-528.

Konrath, S., Meier, B.P. & Bushman, B.J. (2014) Development and validation of the Single Item Narcissism Scale (SINS) . *PLOS ONE*, 9, e103469. doi: 10.1371/journal.pone.0103469

箕浦有希久・成田健一 (2013) 2項目自尊感情尺度の開発および信頼性・妥当性の検討. 『感情心理学研究』 21, 37-45.

Messick, S. (1995) Validity of psychological assessment. *American Psychologist*, 50, 741-749.

村山航 (2012) 妥当性：概念の歴史的変遷と心理測定

学的観点からの考察. 『教育心理学年報』 51, 118-130.

Neverlien, P.O. (1990) Assessment of a single-item dental anxiety question. *Acta Odontologica Scandinavica*, 48, 365-369.

Nunez-Pena, M.I., Guilera, G. & Suarez-Pellicioni, M. (2014) The Single-Item Math Anxiety scale (SIMA) : An alternative way of measuring mathematical anxiety. *Journal of Psychoeducational Assessment*, 32, 306-317.

小塩真司 (2011) 『性格を科学する心理学のはなし：血液型性格判断に別れを告げよう』新曜社

小塩真司・阿部晋吾・カトローニ・P. (2012) 日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み. 『パーソナリティ研究』 21, 40-52.

Rammstedt, B. & John, O.P. (2007) Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41, 203-212.

Robins, R.W., Hendin, H.M. & Trzeniewski, K.H. (2001) Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 27, 151-161.

Schimmack, U. & Oishi, S. (2005) The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology*, 89, 395-406.

Skogh, J., Ylitalo, N., Larsson Omerov, P., Hauksdottir, A., Nyberg, U., Wilderang, U., Johansson, B., Gatz, M. & Steineck, G. (2010) 'A no means no'-measuring depression using a single-item question versus Hospital Anxiety and Depression Scale (HADS-D) . *Annals of Oncology*, 21, 1905-1909.

Watkins, C., Daniels, L.L., Dickinson, H. & Van Der Broek, M. (2001) Accuracy of a single question in screening for depression in a cohort of patients after stroke: Comparative study. *British Medical Journal*, 323, 1159.

Woods, S.A. & Hampson, S.E. (2005) Measuring the Big Five with single items using bipolar response scale. *European Journal of Personality*, 19, 373-390.