

加齢研究における測定についての一考察

—行動学的検査項目の再検査信頼性について—

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

A study on measurements in aging studies: Retest reliability of behavioral test items
Takeshi Hatta (Department of Health Science, Kansai University of Welfare Sciences, Japan)

Abstract

The reliability by the test-retest examination of the cognitive test items that have been used in the Yakumo Study was examined using 40-55 years old community dwellers. These participants' performances can be regarded that the possible influence of cognitive decline due to aging can be ignored. As a result of the test-retest examination, reliability of the subjects who participated in the Yakumo Study health examination for successive consecutive years during the 40-55 years old were as follows; D-CAT1 and D-CAT3, Stroop (dot) and troop (letter) showed a high correlation of 0.75 or more. The logical memory test had a correlation of less than 0.6, the letter fluency test had a correlation of 0.6 or more, but the semantic fluency test had a correlation of 0.4 or less. From these results, it was clarified that the D-CAT and Stroop tests are reliable measures for examining the intra-individual fluctuation of frontal lobe executive function with aging.

Key words

longitudinal aging study, test-retest reliability, digit cancellation test, Stroop test, executive function

1. はじめに

筆者らは1982年に始まった「Yakumo Study」と称される研究グループの構成員として、2001年から前頭葉機能の測定・評価を担当し、中高年者の脳機能特性の検討、脳機能と心理社会的要因との関連、さらに「Yakumo Study」を構成する他の研究班、整形外科班、耳鼻科班、泌尿器科班の資料との連携による共同研究に携わってきた。

2020年初頭からの新型コロナウイルス感染の拡大が終息せず、例年8月に実施している「Yakumo Study」の核である住民健診事業は中止となった。この中断で筆者の属する心理班が2001年から収集してきた縦断的資料は2019年で一旦途切れることとなったが、これまでに6,893名の40歳以上の健常成人(健診の結果の判定でMCIを疑える若干数が含まれる)の蓄積があることを確認した。住民の参加は自由意志なので毎年参加する者、隔年で参加の者など形態は多様であるが、これらの縦断的資料の検討を本格的に始めることにした。幸いなことに2001年から実施した検査バッテリー(NU-CAB)を構成する中核的な注意、記憶、言語を対象とする検査項目は初年度から不変であり縦断的検討が可能である。なお、この検査バッテリーについては、項目や実施手順プロトコルを含めて公開してある(八田, 2004)ので詳細を紹介することはここでは省略する。

言うまでもなく縦断研究に費やされる費用や労力は膨大で、公的機関や自治体との連携・支援で行われ、加齢をテーマにするものには著名な研究が多い。Martin and Ortuno (2019)は最近10年間の縦断的加齢研究の一覧を

記載し、縦断研究の利点を集約している。筆者らの研究も公的資金と自治体の支援で継続できており、期間だけを取り上げれば、Baltes, B. P.が率いる加齢研究の嚆矢であるBerlin Studyよりも筆者らの19年という期間は長く、資料は貴重なものと考えている。この貴重な資料の縦断的検討を本格化するにあたって、検査バッテリーNU-CABの特性について検討・確認するのが本研究の目的である。

検査成績の加齢に伴う個人内変化を追跡する際に考慮すべきは、行動学的指標の信頼性についての確認である。大抵の加齢研究に含まれる医学的検査指標(例えば、身長、体重、血圧、血液中の物質)は物理量であるために、一般的に信頼性が高いものとみなされている。つまり、 X 年次の測定値と $X + \alpha$ 年次の測定値の差分は α 年間の加齢現象がもたらす変化と解釈され、議論が進められる。一方で、認知機能を測定する行動学的指標の場合に、測定値の差分が直ちに加齢現象がもたらす変化とみなせるかには注意が必要と考える。この間には長年多様な中高年対象者に検査を実施してきた経験からのものであり、同様の指摘をしている研究者らもいる(Loevden et al., 2007)。例えば、意味流暢性検査では、オリンピック年であればスポーツに関する反応は他の年に比べて多くなる。このようなコホート効果は横断的検討の場合には対象者に共通の要因として相殺もできようが、縦断的な検討では厄介な問題を孕むことになる。意味流暢性検査でも、ある単語の再生に拘って、再生数が大きく減少することがある。そこで、縦断的な検討を本格化するにあたり、NU-CABの主要な検査項目の信頼性を検討しようと企図した。

再検査信頼性については、過去に報告した検査項目もある。例えば、D-CATについてはHatta et al. (2012)で報告している。そこでは、3つの大学と2つの短期大学から310人の学生(18~22歳)を対象者に選び、同じ検査者

が間隔を2週間にして再検査を行なっている。D-CAT1 (1つの数字を抹消する試行)、およびD-CAT3 (3つの数字を抹消する試行)の2回の検査作業量の相関は、それぞれ $r=0.81$ 、および $r=0.75$ であった。相関は非常に大きく、統計的に有意であった ($p<.001$)。この検討でD-CAT1およびD-CAT3の再検査信頼性は高いと結論している。このような再検査信頼性の検討方法は適正ではあるが、参加した対象者は知的レベルのばらつきを考慮してはあるものの特定の特性を持つ若者が母集団であると指摘することは可能である。つまり、実際に適用する母集団(中高年)とは異質の対象者での信頼性検討に問題はないかと指摘することは可能であろう。実際に検査対象者となる母集団で再び検査を行い比較検討する方が望ましいが、再検査信頼性の検討だけを目的に100~200名の中高年者に2度にわたって同じ検査を実施すべきというのは、机上の空論的指摘である。そこで、類似性の高い検査環境で、加齢による影響を考えずに済む対象者の2度の検査結果を比較する方法を採用して、NU-CABの信頼性検討を意図して本研究を計画した。

2. 方法

2.1 対象者

再検査信頼性を検討した対象者は、自治体が主催する健診事業に自主的に2001年から2019年までに参加し、心理班のNU-CABを受診した者である。したがって、40歳以上の、自立して生活を送っている健常者である(心理班のデータに基づき、MCIが疑われる住民が参加した場合があるが、1%以下である)。NU-CABを受診した対象者の検査結果は電子化されデータベースとして保管されている。

このデータベースからの選択に設けた基準は1年後に同一検査項目を受診していることで、この条件を満たす対象者の検査成績を比較する方法を採用した。例えば、40歳と41歳での検査成績の比較、あるいは42歳と43歳での検査成績の比較であり、再検査の時間間隔は1年ということになり、学習経験の影響を考慮する必要がない間隔とみなせる。1年の間の変化には加齢に伴う変化要素が加味される可能性は考えられうるが、筆者らは加齢に伴う認知機能成績は55歳以降に顕在化することを本研究と同じ母集団の資料から明らかにしており、55歳までであれば1年後の検査結果に加齢による認知の機能的変化要素は考慮せねばならないほど大きくないと考えた(八田, 2011)。なお、NU-CABはMMSE (Mini Mental-State Examination: Folstein et al., 1975)と、下記の紹介する認知機能検査項目から構成される検査バッテリーであるが、MMSEは満点が30点となるように構成された一般知能測定を目的とし、得点が正規分布する性質をもたない。再検査信頼性の統計処理に適さないため、以下の検討からは除外した。

八雲研究では60歳代~70歳代の対象者の参加が多数を占め、40歳頃から毎年受診する対象者は多くはない。当初から長期の縦断研究を意図していたわけではないの

で、検査項目によっては対象者のサイズに多寡があるのは止むを得ないが、相応の母集団サイズと考えている。つまり、2001~2019年の間に2年連続して受診している40歳以上55歳までの対象者は男子139人、女子215人での合計354人であった。男子の対象者は39.3%であり、女子の方が連続年で多く健診に参加していることになる。分析対象者の平均年齢は男子50.6歳($SD=4.2$)、女子50.4歳($SD=3.8$)で性差は見られない。連続年で受診をした対象者の場合は、年齢の若い連続年を統計解析の対象とした。つまり、2001、2002、2004、2005年というように連続して受診した場合には、2001と2002年の間の結果を用いた。そのために、354人全体が1年後の再検査結果の分析対象というわけではない。検査によって欠損や未実施年があるために、対象標本数はD-CAT1 ($N=233$)、D-CAT3 ($N=233$)、論理記憶 ($N=203$)、文字流暢性 ($N=102$)、意味流暢性 ($N=101$)、Money道路図検査 ($N=80$)、Stroop (dot) $N=145$ 、Stroop (Letter) $N=144$ である。

なお、NU-CABの各検査項目の概要は下記に示す通りで、実施手続きについては前述した八田(2004)に記載がある。

- D-CAT1: ランダムな順序で1桁の数字が印刷された用紙を与えられ、「6」をできるだけ早く見落としがないように抹消することを求める課題。
- D-CAT3: ランダムな順序で1桁の数字が印刷された用紙を与えられ、「8、3、7」をできるだけ早く見落としがないように抹消することを求める課題。

2.2 論理記憶

ウェクスラー記憶検査日本版に含まれる散文を検者が2度読み上げ、直後に再生を求める課題。25の文節、句に分けて採点する。

2.3 文字流暢性

指示する音節(文字)「あ」、「か」、「し」から始まる普通名詞を1分間にできるだけ多く産出するように求める課題。

2.4 意味流暢性

指示するカテゴリ一名「スポーツ」、「職業」、「動物」に属する普通名詞を1分間にできるだけ多く産出するように求める課題。なお、文字流暢性および意味流暢性課題で用いる文字や単語は、標準化手続きを経たものである(伊藤・八田, 2004)。

2.5 Money 道路図

練習用図版と本使用図版から構成されている。図版は2cm幅の道路とみなすように指示される線分が12箇所左右に曲がるように作成されている。「ここにある道を歩くつもりで、左右どちらに曲がるかを回答するように求める。身体や首は動かさないように指示される。練習試行

後、対象者のペースで本試行を行う。Butters, Soeldner and Fedio (1972) により開発された地誌的空間見当識を調べるための検査の簡略版である。

2.6 Stroop (dot)

A4 サイズの図版に 5×8 の直径 2.5 cm 大の円が赤、青、黄、緑の 4 色でランダムな順に印刷された用紙を用いて、色名をできるだけ早く、正しく読み上げる課題で、読み上げに要した時間とエラーを記録する課題である。

2.7 Stroop (letter)

A4 サイズの図版に 5×8 の「あか」、「あお」、「き」、「みどり」が色名と一致しないように印刷された用紙を用いて、色名を出来るだけ早く、間違いなく読み上げる課題で、読み上げに要した時間とエラーを記録する課題である。

3. 結果

検査項目別に受診初年度と翌年度の作業量とのピアソン相関係数を示したのが表 1 である。相関は高いものとそれほど高くないもの等があるが、統計的にはすべて有意である ($p < .001$)。

4. 考察

Martin and Ortuno (2019) は、縦断的加齢研究は病気のメカニズムや内在的能力の複数の要素の低下がどのように相互作用するか、ある要素の喪失が別の要素の経路にどのように影響するか、そしてこれらの変化がどのように機能障害につながるかを明らかにすることができると指摘し、その利点を、①遺伝学、バイオマーカー、生理学的障害、機能的な能力、ヘルスケアの使用など多くの領域でデータを提供できる、②連続したコホートの変化の軌跡についてのデータを提供し、加齢の影響、コホートの違い、および期間の影響をよりよく理解できる、③後の健康低下の早期予測因子情報を提供することができ、公衆衛生および臨床診療に役立つ、④健康への悪影響の発生における時間、場所、社会の貢献を解明するのに役立つ、⑤縦断的データは、データによって示唆される因果関係の証拠を強化または反駁する可能性がある、と集約している。改めて縦断研究の役割の大きさを確認せず

には居られない。それと同時に八雲研究を 19 年間継続できた幸運を自覚せねばなるまい。

備忘録として若干の寄り道をしたい。前頭葉機能の測定評価に関わるきっかけは、頭部外傷からの回復期にある患者の認知リハビリテーションに取り組みされていた名古屋市総合リハビリテーションセンターからの誘いによるもので、それまでにセンターで利用してきた認知機能検査のレビューを行い、注意機能の検査項目が不足している結論に至ってからである。当時の注意理論を調べ、情報系教官の協力を得て 15 分ほどで実施可能な、注意機能の階層に対応する項目から成る Windows 版 NNAA (Nagoya University Neuropsychological Assessment for Attention) を試作した。この試作検査をセンターの勉強会で披露した所、5 分以内で実施できなければ使えないと医師らからダメ出しを食らった。心理学検査に実用的妥当性 (practical validity: 利用者の使い勝手や非検査者の負担) という視点を加味すべきことを知らされた。心理学では、検査開発に信頼性と妥当性 (内容的妥当性、基準関連妥当性、構成概念妥当性) の必要性は強調するが、それだけでは不十分で、実用的妥当性が考慮されねば、実際に活用されることがない自己満足の検査に終わるのである。井の中のカエルでは学際研究に参加できないことを学んだことになる。そこで、注意機能の階層の基礎部分に焦点を当てた D-CAT を作成した。

このような背景で生まれた D-CAT について、40 歳以上 55 歳までの対象者に 1 年後の再検査信頼性を検討した本研究の結果は、大学生を対象として 2 週間後に再検査を行って信頼性を検討した先行研究 (Hatta et al., 2012) と酷似した結果であった。さらに、Stroop (dot)、Stroop (letter) の検査項目も相関係数は D-CAT 検査に極めて類似したものであった。一方で、論理記憶項目は、相関は相応に高いが D-CAT や Stroop 検査項目とは差異が見られること、文字流暢性と意味流暢性の項目間には違いがあり、文字流暢性検査は高い相関を示したが、意味流暢性検査は「適度の相関」のカテゴリーに属するレベルであった。Money 道路図検査については、相関係数は極めて高いが、測定値に分散が乏しく、D-CAT や Stroop 検査と同列には考えるべきではないことが明らかとなった。Money 道路図検査は空間見当識の測定項目であるが、簡略版で 12 点満点

表 1: 検査項目別相関係数値

検査項目	N	X 年次 Mean (SD)	X + 1 年次 Mean (SD)	r 値
D-CAT1	228	31.26 (5.60)	31.52 (5.40)	0.7818
D-CAT3	228	54.13 (11.60)	54.20 (11.82)	0.7852
論理記憶	203	17.91 (4.45)	17.57 (4.31)	0.5804
文字流暢性	102	11.14 (4.14)	10.95 (4.29)	0.6200
意味流暢性	101	14.53 (5.18)	14.77 (4.78)	0.3831
Money 道路図	79	11.79 (0.58)	11.84 (1.33)	0.9439
Stroop (dot)	145	24.01 (5.46)	23.48 (4.80)	0.7599
Stroop (letter)	144	32.78 (7.98)	31.89 (7.70)	0.8164

である。実際に検査を実施してきた経験からは相応なエラーが出ている印象であったが、55歳までの対象者はほとんどエラーを生じないことが確認できた。

Dikmen et al. (1999) は、Halstead-Reitan Neuropsychological Test Battery の再検査信頼性データ（記憶、注意、運動機能）を報告し、ほとんどの検査項目は、0.7～0.9の適度に堅牢なピアソンの相関係数を示したが、記憶関連検査の相関係数は0.6未満で、記憶検査での再検査信頼性は比較的低いとしている。このことは他でも報告されており（Geffen, Butterworth, & Geffen, 1994; Rapport et al., 1997; Rasmussen, Bylsma, & Brandt, 1995; Uchiyama, D'Elia, Dellinger, & Becker, 1995）、本研究の結果はそれらに一致したものといえよう。

本研究での解析結果で、Stroop 検査の再検査信頼性の高さが確認された。Stroop (dot) については情報処理速度という共通性から D-CAT と類似レベルの相関係数を予想していたが、Stroop (letter) で 0.82 の高い値を得たことは、実行系機能を 3 要素 (updating information, shifting, inhibition) の構成であり (Miyake et al., 2000)、D-CAT と Stroop (dot) が前二者と、Stroop (letter) が inhibition 要素との対応を想定し、性質が異なるのではないかとする予想とは異なった。しかし、このことは、実行系機能が前頭葉機能の中核という見解は研究者間でコンセンサスとなっているので、加齢と前頭葉機能（とりわけ実行系機能）との関連を詳細に検討する上では強力な指標を特定できたことになる。

本研究から筆者らが加齢研究で焦点をあてようとしている実行系機能を測定する際に用いようとしている D-CAT と Stroop 検査の信頼性が高いことが確認できたことは、加齢に伴う検査成績の個人ない変動の検討を今後志向する上で裏付けを得られたことになる。検査測定値の個人内変動（バラツキ）は月・年単位の要因と日内・数日内の要因から構成されることについては (Li, Lindenberger et al., 2004) の研究などに集約がある。この種の検査測定値の変動を指標とする縦断研究での研究の関心は、特に高齢期において加齢に伴って変動が大きくなるのか、変わらないのかについての関心に焦点が当てられている。Loevden et al. (2007) は、試行間の変動性の変化が認知能力レベルの変化と関連しており、試行間の変動性が高いほど、老年期の認知能力の低下に先行し、予測するとした。MacDonald et al. (2003) では、75～89歳の対象者の6年間の個人内変動を反応時間で検討したもので、加齢に伴い変動が大きくなることは前頭葉の機能低下を反映し加齢と共に進むとしている。加齢を焦点に据えた縦断研究では身体的特徴や循環器、泌尿器などの内科的検査指標と行動学的指標が併用される。特に、筆者らが志向する行動学的指標による認知検査、とりわけ実行系機能検査結果の変動性の検証は、差し迫った認知機能低下の早期警告として役立つ可能性があり、重要な臨床的利益をもたらす可能性が高いと言えよう (Hultsch & MacDonald, 2004; MacDonald et al., 2006)。

最後に本研究での limitation について言及する必要がある。第1は、認知検査の信頼性の検討には（疑問視する

見解もあるが、e.g. Cicchetti, 2001）400人以上の参加が望ましいという報告もあり、対象者数が十分と言えるほど大きくはないことである (Charter, 1999)。もっとも、Charter が検討対象とした再検査信頼性研究では 100 人未満が 59% ということなので、一般的には不十分なサイズということでもない。また、再検査の期間が短ければ学習効果の影響が無視できなくなるが、1年という期間についても考慮する必要はあろう。この間に加齢による要因の影響が想定できるが本研究では D-CAT での先行研究結果と変わらない相関係数値が得られたので影響は無視できる水準であると言えよう。

引用文献

- Baltes, P. B. & Mayer, K. U. (eds.) (1999). *The Berlin aging study: Aging from 70 to 100*. Cambridge University Press.
- Butters, N., Soeldner, C., & Fedio, P. (1972). Comparison of parietal and frontal lobe spatial deficits in man: Extra-personal vs personal (egocentric) space. *Perceptual and Motor Skills*, 34, 27-34.
- Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, 21, 559-566.
- Cicchetti, D. V. (2001). Methodological commentary the precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, 23, 695-700.
- Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects on Expanded Halstead-Reitan Neuropsychological Battery. *Journal of the International Neuro-psychological Society*, 5, 346-356.
- Geffen, G. M., Butterworth, P., & Geffen, L. B. (1994). Test-retest reliability of Halstead Impairment Index in hospitalized alcoholic and nonalcoholic males with mild to moderate neuropsychological impairment. *Journal of Clinical Neuropsychology*, 3, 257-269.
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Minimal state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189-198.
- 八田武志 (2004). 住民検診を対象とした認知機能検査バッテリー (NU-CAB) 作成の試み. *人間環境学研究*, 2, 15-20.
- 八田武志 (2011). 認知の個人差の脳内機構. 箱田 (編) 認知の個人差, 6 章 (130-169), 北大路書房.
- Hatta, T., Ito, Y., Yoshizaki, K., Mase, M., & Kabasawa, H. (2012). Reliability and validity of the digit cancellation test, a brief screen of attention. *Psychologia*, 55, 246-256.
- Hultsch, D. F. & MacDonald, S. W. S. (2004). Intraindividual variability in performance as a theoretical window onto cognitive aging. In R. A. Dixon, L. Baeckman, & L.-G. Nilsson (eds.), *New frontiers in cognitive aging* (pp. 65-88). Oxford,

UK: Oxford University Press.

- 伊藤恵美・八田武志 (2004). 健常成人の言語流暢性検査の結果について—生成語数と年齢、教育歴、性別の影響—. *神経心理学*, 20, 254-263.
- Li, S.-C., Huxhold, O., & Schmiedek, F. (2004). Aging and attenuated processing robustness: Evidence from cognitive and sensorimotor functioning. *Gerontology*, 50, 28-34.
- Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, 15, 155-163.
- Loevden, M., Li, S.-C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: Longitudinal data from the Berlin Aging Study. *Neuropsychologia*, 45, 2827-2838.
- MacCarrey, A. C., An, Y., Kitner-Triolo, M. H., Ferrucci, L., & Resnick, S. M. (2016). Sex differences in cognitive trajectories in clinically normal older adults. *Psychology and Aging*, 31, 166-175.
- MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria longitudinal study. *Psychology and Aging*, 18, 510-523.
- MacDonald, S. W. S., Nyberg, L., & Baeckman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*, 29, 474-480.
- Martin, F. C. & Ortuno, R. R. (2019). Longitudinal studies of ageing: from insights to impacts: Commentary to accompany themed collection on longitudinal studies. *Age and Ageing*, 448, 482-485.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.
- Rappaport, L. J., Axelrod, B. N., Theisen, M. E., Brines, D. B., Kalechstein, A. D., & Ricker, J. H. (1997). Relationship of IQ to verbal learning and memory: Test and retest. *Journal of Clinical and Experimental Neuropsychology*, 19, 655-666.
- Rasmussen, X., Bylsma, F. W., & Brandt, J. (1995). Stability of performance on Hopkins Verbal Learning Test. *Archives of Clinical Neurology*, 10, 21-26.
- Uchiyama, C. L., D’Elia, L. F., Dellinger, A. M., & Becker, J. T. (1995). Alternate form of the auditory-verbal learning test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neurology*, 10, 133-145.

(受稿：2020年10月9日 受理：2020年11月16日)