

# タブレット版 Stroop 検査作成の試み

## —信頼性及び妥当性の検討—

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

八田 武俊 (岐阜医療科学大学 保健科学部, [hatta@u-gifu-ms.ac.jp](mailto:hatta@u-gifu-ms.ac.jp))

木村 貴彦 (関西福祉科学大学 健康福祉学部, [takimura@tamatetama.ac.jp](mailto:takimura@tamatetama.ac.jp))

重森 健太 (関西福祉科学大学 保健医療学部, [shigemori@tamatetama.ac.jp](mailto:shigemori@tamatetama.ac.jp))

岩原 昭彦 (京都女子大学 発達教育学部, [iwahara@kyoto-wu.ac.jp](mailto:iwahara@kyoto-wu.ac.jp))

An attempt to develop a tablet version of the Stroop Test:

Examination of reliability and validity

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

Taketoshi Hatta (Department of Health Science, Gifu University of Medical Sciences, Japan)

Takahiko Kimura (Department of Health Science, Kansai University of Welfare Sciences, Japan)

Kenta Shigemori (Department of Rehabilitation Science, Kansai University of Welfare Sciences, Japan)

Akihiko Iwahara (Department of Education, Kyoto Women's University, Japan)

### Abstract

The reliability and validity of Digital Stroop Test (Dig. Str.) that was developed for a screening test of the executive function were evaluated by two studies. The Dig. Str. was developed to evaluate both Stroop effect and Reverse Stroop effect using a computer tablet. In Study 1, the reliability of the Dig. Str. was evaluated using test-retest paradigm. High correlations between scores (mean response time per correct response) on the two tests were obtained. In Study 2, the construct validity was examined using a sample of healthy old-aged community dwellers. Old-aged people showed significantly slower mean response times than college students. Based on these findings, the Dig. Str. can be regarded as a reliable screening equipment for executive function.

### Key words

Digital Stroop Test, Stroop effect, reversed Stroop effect, test-reliability, test-validity

### 1. はじめに

本研究はタブレット版遂行機能検査 (Digital Stroop Test 以下: Dig. Str.) 開発のための基礎的検討が目的である。具体的には試作版の信頼性及び妥当性の検討を行う。Stroop 検査 (ストロープ検査) は Stroop (1935) が発見した現象を検出するための検査である。その現象とは、色名を示す文字情報と文字を描くインクの色という 2 種の属性が含まれる刺激が提示され、インクの色を口頭で回答させた際、その二つの属性が一致する場合よりも、不一致の場合の方が反応時間を要するというもので、発見者の名前を冠してストロープ効果と呼ばれている。2 種の属性であれば、文字とインクの色だけでなく、たとえば、イヌの線画の中にネコの文字がある図形で文字を読む課題や、男性と女性の話者が「おとこ」「おんな」と読み上げ話者の性別を判断するなど、視覚でも聴覚材料でもストロープ効果が生じることが報告されている (MacLeod, 1991)。

ストロープ検査は書かれた文字のインクの色を口頭で答えるプロトタイプと、印刷された検査用紙で回答を選択するというマッチング法を用いたタイプとがある。前者は基本的には検者と対象者が 1 対 1 であるが、後者は

集団での検査が可能というメリットがある (奈良・石橋・染谷他, 2010; 箱田・佐々木, 1990)。本研究の対象である Dig. Str. は、高齢者を対象とする遂行機能の加齢による発達の变化の検出利用などを企図しているので、マッチング法による反応を採用している。

なお、色の文字情報と一致しない色のインクで印刷されたストロープ干渉 (Stroop interference: 以下、SI) 課題では、インクの色について回答することを求めるが、逆ストロープ干渉 (Reverse-Stroop interference: 以下、RI) 課題では、文字とインクの色が不一致の場合に、文字の読みを求めるというように反応方法に違いがある。一般に、RI は、SI ほどストロープ効果は強く見られないことが知られているが、箱田・佐々木 (1990) はマッチング法では、SI よりも RI の方が 2 つの属性での干渉効果は大きいと報告している。マッチング法を採用した SI の得点は、5～24 歳の児童および成人を対象とした研究において、不注意・多動性—衝動性の指標となる ADHD 尺度得点との関係が顕著な一方で、RI 課題の得点と ADHD 得点との関連は乏しかったことが報告されており、2 つの課題は異なる認知過程を反映している可能性があることから (Ikeda, Okazumi, & Kokubun, 2013)、Dig. Str. の開発を企画した際に、SI と RI 両方の課題を採用し、併せて両者の感度の比較を目指した。

研究 1 では、Dig. Str. について再検査信頼性を検討し、研究 2 では妥当性の検討を目的とした。

## 2. 研究 1

心理測定の信頼性について、岡田（2015）は本邦の査読学術誌から 159 編の関連論文を調べ、信頼性の検討は、再テストの煩雑さや類似問題を作る手間を省くために平行検査法や折半法が信頼性の検討法として生み出され、クロンバック  $\alpha$  係数が広く用いられるが、作業が煩雑であることや学習効果がマイナス要因となりうるので、再検査法が信頼性検討の基本であるとしている。そこで、本研究では、再検査法で信頼性を検討することとした。八田ら（2022）はタブレット版注意機能スクリーニング検査 Dig. D-CAT の開発に関して、大学生を対象に再検査法で信頼性の検討を報告している。そこでは、検査用紙版での D-CAT (Pap. D-CAT) を大学生に 7 日間の間隔で同じ条件下で 2 度実施した。算出した再検査信頼性係数は、1 文字抹消課題では 0.78、3 文字抹消課題では 0.83 であった。Hatta et al. (2012) での Pap. D-CAT 再検査信頼性の検討では、大・短大生が対象者で、1 文字抹消課題での再検査信頼性係数は 0.81 で、3 文字抹消課題の場合は 0.75 であった。再検査までの期間は 2 週間と報告されており、再検査期間や対象者は異なるが、両方の報告ともに Pap. D-CAT は高い相関係数を示しており、確かな信頼性を有することを確認している。しかしながら、同様の手続きで行ったタブレット版の D-CAT (Dig. D-CAT) では、1 文字抹消課題では 0.58、3 文字抹消課題では 0.45 と、相関係数の値として中程度の相関関係しか見られず、反応媒体であるタッチペンでの反応と鉛筆での斜線での反応との違いに原因がある可能性を論じている。したがって、Dig. Str. が Dig. D-CAT での再検査信頼性と同等レベルにとどまるのかは重要な関心事であり、研究 1 では、作業仮説を「Dig. Str. は高い再検査信頼係数を示す」とした。心理検査での信頼性については、小塩（2016）が指摘するように、コンセンサスのある数値があるわけではない。望ましい信頼性レベルはどの辺りかについて評論した彼によれば、0.70 ~ 0.80 程度を十分高い係数と指摘しているので、ここでの作業仮説はより具体的なものとして、①「再検査信頼性係数は 0.7 以上である」、および「RI 課題」と「SI 課題」を比較した場合にはマッチング法を採用したので、箱田・佐々木（1990）が指摘するように②「RI 課題は SI 課題よりも干渉効果は大きい」を採用した。

## 3. 方法

### 3.1 対象者

49 名の大学生（男性 33 名、女性 16 名）で、平均年齢は 19.47 ( $SD = .50$ ) 歳である。1 週間の間隔を置いて、同じ手続きで Dig. Str. を実施した。対象者は視力や手指運動機能を始めとする認知能力や身体機能に問題を持たない大学生であった。

### 3.2 検査材料と検査方法

Dig. Str. は筆者らと協議し改良を加えてきた（株）FIS の開発した試作版 Ver.4 を使用した。Dig. Str. Ver.4 では、検査日時、氏名、性別、年齢、検査時間のタイプ（60 s/30 s）

に加えて、すべての課題における反応数、正答数、誤答数、正答の平均反応時間（s）及びその標準偏差、ストループ干渉効果の Z 値、実施した母集団内での評価（正常範囲、要注意）コメントが自動的に記録される仕様となっている。原則として対象者自身が自分のペースで検査課題を実施していく仕様であり、反応様式は音声ではなくタッチペンによるマッチング法が採用された。本研究では、Babble Com (P-001) 社製のタッチペンを用い、Dig. Str. は以下のような構成のサイズ 12.3 インチのタブレットで実施した。

### 3.3 手続き

対象者は、机上のタブレット画面に提示される指示に基づいて課題を実施した。第 1 回目の検査と 2 回目の検査は、同じ場所で 7 日間の時間間隔において実施した。

### 3.4 検査手順

- 最初にタブレット画面に氏名、年齢、性別などの対象者の属性を画面に記入すると、「練習」画面が提示される。
- 「練習」画面では、タッチペンでの反応方法を習得するために、1、2、3、4、5 の数字が提示され、タッチペンで数字を押して色を変えろという作業を行う。
- ついで、2 cm 大の四角形 5 色（緑、赤、青、黄、黒）が横一列に提示され、RI 統制課題の練習が始まる。画面は図 1 に示すように、5 色の四角形の列と、その上部に 20 ポイントのゴシック体ひらがなで色名（例えば、青色インクで、きいろ）が提示される。そして、画面上部にある色名を示す「ことば」と合致するインクの色を押すよう指示があり、「「ことば」は「きいろ」ですから、右端の四角形（黄色）を押してください」との教示が続く。なお、この練習課題は必要なだけ繰り返すことができる。

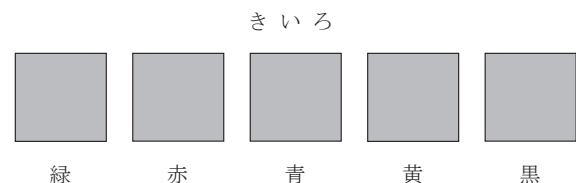


図 1：逆ストループ課題（RI）の画面例

注：図形の下側に記載された色名は実際の刺激提示画面では表示されない。

- 4.3 秒後に本試行へと移行し、色名を示す「ことば」に合致する色で描かれた四角形を選択するという課題を 60 秒間行う（対象者の反応速度に実行課題数は依存する）。
- RI 統制課題の本試行が終了すると、対象者は「次へ」のタブ押しにより、RI 課題の練習に移行する。ここでの教示は、「「ことば（色名）」とインクの色がチグハグですが、インクの色に惑わされないように、「ことば」が表すインクの色をの四角形を押してください。「ことば」

は、青色で「きいろ」となっていますので（きいろの文字は青色で表示されている）、右から3番目の黄色の四角形を押してください。”である。

- 練習試行を3回行い、それ以上の練習が必要ないとした場合には、RI課題の本試行が始まる。試行数は対象者の反応速度に依存するが、60秒で終了する。
- ついで、SI統制課題に移行する。この課題では、対象者が図2に示すような、画面上部に提示された四角形の色に合致する色名を示す「ことば」を選択する。具体的には、図2に示したように、四角形の色は「みどり」なので、右端の「みどり」の色名をタッチペンで押すことが求められる。

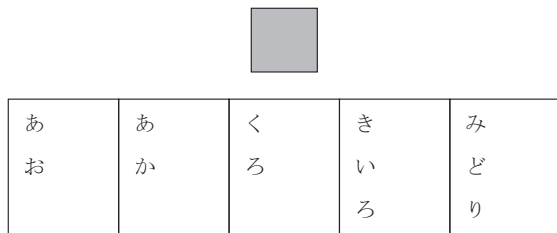


図2：ストロープ課題（SI）の画面例

注：実際の刺激提示画面では上部の四角形はみどり色で表示されている。

- 練習試行を5回行った後、3秒後にSI統制課題の本試行が60秒間実施される（図3）。画面に表示される教示は、“画面上部にあることば（色名）が表すインクの色を押してください。例題では「みどり」ですから、右端の「みどり」を押してください。”であり、このとき四角形はみどり色で表示されている。試行数は対象者の反応速度に依存する。

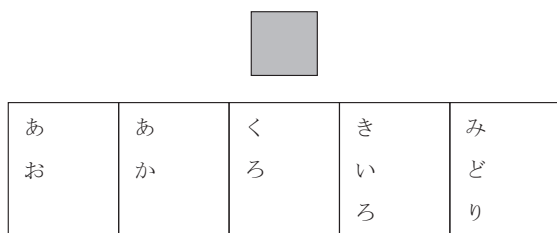


図3：SI統制課題の画面例

注：実際の刺激提示画面では上部の四角形はみどり色で表示されている。

- 最後に、SI課題に移行する（図4）。この課題では、インクの色名を示す「ことば」が表す色で描かれた四角形を選択することが求められる。具体的には、図4とともに教示が提示される。「ことば」はみどり色で描かれているが、「ことば」が示す色名は「くろ」なので、中央の黒い四角形を押さねばならない（くろの文字は黒色で表示されている）。教示画面3秒後から5回の練習試行が終わったのち、60秒間のSI課題の本試行に移行する。なお、全ての教示画面には、できるだけ早く反応することを求める文章が付記されており、5つの色の配置は試行毎にランダムに変化するように設計

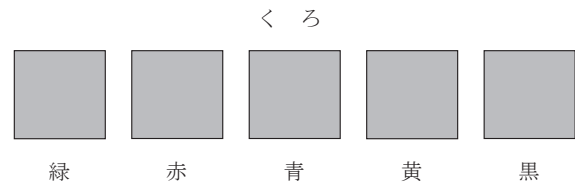


図4：SI統制課題の画面例

注：実際の刺激提示画面ではみどり色で表示されている。

されている。

対象者は「RI統制課題」、「RI課題」、「SI統制課題」、「SI課題」の順に自分のペースで実施した。

#### 4. 結果

Dig. Str. 検査の指標には、平均反応時間 (Sec.) を用いた。各課題は60秒で終了するので、対象者による反応数は30～50回程度に分布した。そこで、1反応あたりの時間を計算し、その平均値を指標とした。平均反応時間が短いことは、作業成績が優れることになる。Dig. Str. の平均反応時間と標準偏差は表1の通りである。表1に示すように、再検査相関係数はRIでは0.76と0.79、SIでは0.88と0.86と高い値となった。したがって、作業仮説①「再検査信頼性係数は0.7以上である」は支持されたことになる。

表1：Dig. Stroop Test の平均反応時間

	1回目	2回目	信頼性係数
RI 統制課題	0.87 (-0.13)	0.92 (-0.01)	.79
RI 課題	1.07 (-0.14)	1.02 (-0.16)	.76
SI 統制課題	1.32 (-0.13)	1.20 (-0.15)	.86
SI 課題	1.55 (-0.22)	1.40 (-0.19)	.88

注：単位はsecで、括弧内は標準偏差)および再検査信頼性係数。

#### 5. 研究2

研究2の目的は、Dig. Str. 検査の妥当性の検討である。妥当性の検討は、測ろうとするものを測れているかの検証であり、その具体的方法はさまざまである。例えば、D-CAT 検査（用紙版）の妥当性の検証は、健常者と頭部外傷者（Traumatic Brain Injury：以下TBI）を対象に検討している（Hatta, Yoshizaki, Ito et al., 2012）。そこでの作業仮説は、TBI患者は前頭葉機能に障害を残存させることが多くの研究で明らかにされており（Prigatano, 1986; Whyte, Schuster, Polansky et al., 2000）、とりわけ、情報処理速度や注意機能低下が特徴であることにはコンセンサスがあるとみなされる。例えば、Lezak (2004) はTBI後の患者に最も一般的に残存する症状に注意及び集中力の障害があるとし、Mateer (2000) も同様にTBI患者は外傷後、集中力の欠如、気が散ってしまうこと、物忘れ、同時に複

数のことができないなどの訴えが一般的であるとしている。そこで、「TBI患者と健常者にD-CATを実施すれば、前者の成績は有意に後者より劣る」と作業仮説に設定した。年齢、性別、教育歴を統制し、外傷後平均7.7年経過したTBI患者と健常者を対象にD-CATを実施している。その結果は、1文字条件では平均検索文字数は220対378、文字条件では平均検索文字数127対224とTBI患者は有意に成績が劣っていた。見逃し数では1文字条件で2.0%対1.0%、3文字条件では13.2%対3.6%であり、数字の作業記憶への負担が増えるとTBI患者は見逃しエラーが有意に増えていた。これらの結果から、D-CATは両群の注意機能を判別でき、その測定には妥当性があると結論している。

そこで、同様のロジックで本研究では大学生と高齢者の2群の対象者にDig. Str.を実施し、成績を比較することで妥当性を検証する。Hatta, Yoshizaki, Ito et al. (2012)は、注意機能低下の対象者としてTBI患者を実験群に、問題がない者として健常成人を統制群として成績を比較したが、研究では遂行機能が低下している者として高齢者を実験群として採用し、問題がない者として大学生を統制群とした。

高齢者の遂行機能が青年期のそれよりも低下することには膨大な報告があり、コンセンサスを得ている事実とみなすことができる(Caballero et al., 2021; Dempster, 1992; Kadmbi, Abdalla, & Loh, 2020; Furguson, Brunson, & Bradford, 2021; Loaiza & Souza, 2019; Madden, Whiting, & Huettel, 2005; 永原・伊藤・岩原他, 2012; Parkin & Walter, 1992; 吉崎, 2009; Zhuravleva, Alperin, Haring et al., 2019)。

したがって、研究2での作業仮説は、① Dig. Str. のRIおよびSIでの作業成績である平均所要時間は、実験群は統制群よりも劣り、②誤反応数でも、実験群の成績は統制群よりも劣る、であった。

## 6. 方法

### 6.1 対象者

実験群として30名の高齢者(男性4名、女性26名で、64～86歳：平均年齢74.0歳)、統制群として50名の大学生(男性34名、女性16名で、19～20歳：平均年齢19.47歳)を対象にDig. Str.を実施した。すべての対象者は視力、手指運動機能を始めとする認知能力や身体機能に特に問題を持たなかった。高齢者の認知機能については、別に実施した認知機能検査(文章記憶検査、D-CAT)の結果から、MCIを疑うような結果は得られていない。統制群についても大学生であり、認知機能障害は考えにくい。

### 6.2 検査材料と検査方法

Dig. Str.Ver.4が実装されているタブレットを机上において、対象者自身が自分のペースで検査者の指示のもとで、課題を実施した。実施の方法等は研究1と同じである。実験群の対象者への検査実施は、2020年10月と2022年10月に行い、統制群の対象者への検査実施は2022年10

月であった。

## 7. 結果

Dig. Str. 検査の指標は、研究1と同じで、所要時間(Sec.)を用いた。各課題は60秒で終了し、対象者による反応数は30～50回程度に分布した。そこで、1反応あたりの時間を計算し、その平均値を作業成績の指標とした。平均反応時間が短いことは、長い場合よりも、優れる成績ということになる。Dig. Str. の平均反応時間と標準偏差は表2の通りである。別の指標には、60秒間に生じた誤反応数を用いた。その結果も表2に記載の通りである。誤反応数が少ないことが作業成績は優れることになる。

表2：高齢者と若年者のDig. Str. 検査における統制課題及びストループ課題での平均反応時間(Sec.)と平均誤反応数、および標準偏差

		高齢者群	若年者群
統制課題	反応時間	2.02 (0.06)	1.49 (0.05)
	誤反応数	0.73 (0.11)	0.96 (10.3)
ストループ課題	反応時間	2.60 (0.12)	1.29 (0.09)
	誤反応数	0.88 (0.12)	0.60 (0.93)

作業仮説①および②を検証するために、2要因混合型分散分析をおこなった。対象者間要因は高齢者群・若年者群であり、対象者内要因は統制課題とストループ課題である。

平均反応時間を指標とした分散分析結果は、すべての要因で有意差があり、高齢者群は若年者群よりも反応時間が長かった( $F_{1,78} = 9.54, p < .01$ )、また、ストループ課題では統制課題よりも反応時間は長く、ストループ干渉効果は顕著に存在したことになる( $F_{1,78} = 78.27, p < .01$ )。

誤反応数を指標とした分散分析では、主要因に有意差はなく、交互作用は有意傾向( $F_{1,78} = 3.10, p = .08$ )にとどまった。つまり、群間に有意な差異は認められなかった。したがって、作業仮説①は支持されたが、②は支持されなかったことになる。誤反応数の指標については若年者も老年者も対象者は健常であり、変量とみなすには問題があるレベルの出現数であったことが原因と考えられる。

## 8. 考察

まず、遂行機能を測定するとされる行動学的検査には、SAS (supervisory attentional system) 理論に基づく、BADS (Behavioral Assessment for Dysexecutive Syndrome)、Tripartite 理論に関連するWCST (Wisconsin Card Sorting test) や Trail Making Test、作業記憶理論によるN-BACK検査など、幾つもの検査がある。なぜ、その中でストループ検査に着目するのか、その理由を上げておく(遂行機能理論とその測定評価の詳細については、八田(2022))

を参照されたい)。Tripartite 理論は前部網様体賦活系、拡散視床投射系、そして前頭-視床ゲート系の3者を想定し、前2者は注意の覚醒維持を担い、残りは行為下での注意の調整、計画、刺激-反応の選択、日常生活でモニタリングを担うとする理論である。ストループ検査は、この理論と関連し、八田(2007)で指摘した注意検査の要件、①特別な機械装置・器具を必要としないこと、②対象者への心的負担が過度にならない、③検査者が実施と結果の判定に特別な訓練を要しないこと、④検査結果の採点・評価が簡単である、⑤機能評価が科学的に裏付けられた方法で可能、を満たすことがある。とりわけ、対象者が検査を受けた後にも、ストループ検査は「面白かった」などの報告があるように、不快感で検査が終了して、繰り返し検査への忌避感を生じない実用面での妥当性を持つ点にある。さらに、他の遂行機能検査に比べて、所要時間が短いことや時間制限法で検査所要時間が計算できる点でも、継続して遂行機能の変化を検討する目的の場合には優れている。

筆者らは、従来のストループ検査版への口頭による反応で測定するよりも、正確な時間測定が可能となること、結果処理の簡便さなどを追求してデジタル版の検査の開発を目指しており、試作機の信頼性、妥当性の検証を行うのが本研究の狙いであった。

研究1の目的は、Dig. Str.の信頼性の検討であり、同一対象者に再検査を行った。その結果、再検査相関係数は、RI課題で.76、RI統制課題で.79、SI課題で.88、SI統制課題で.86と高い値になり、作業仮説①「再検査信頼性係数は0.7以上である」は支持された。一方で、SI課題の信頼性係数の方が高く、作業仮説②「RI課題はSI課題よりもストループ効果は大きい」は支持されなかった。両課題の信頼性係数の有意差検定結果は、RI課題で $z = -0.59$ 、SI課題で $z = -0.77$ で、いずれも有意差はない。作業仮説②が支持されなかった理由は直ちには明らかではないが、箱田らのストループ研究は紙ベースの課題であり(箱田・佐々木, 1990; 箱田・渡辺, 2005)、ペンで該当する反応枠に斜線を引く反応様式で、本研究でのタッチペンによる反応様式とは異なっている。また、宋・箱田(2011)は、パソコン画面に提示するストループテストを作成し、逆ストループ検査課題も加えて紙ベースでのストループ検査と逆ストループ検査とを比較している。彼らはパソコンキーボードでの反応と紙ベースでの反応間には高い相関があるが、紙ベースの方が干渉効果は大きいと報告している。また、反復実施による検討では紙ベースの場合には干渉効果は増す傾向にあるが、キーボード反応では減る傾向にあるとしている。八田ら(2022)は、数字抹消課題を紙ベース版とタブレット版で比較した研究において、タッチペンでは筆圧調整コントロールにおける手先の筋運動系負荷が大きく、紙に印刷された数字に斜線を引く反応様式との違いを指摘しており、箱田らの研究と同様に反応様式の違いが干渉効果に寄与することを指摘している。

再検査信頼性係数がどの程度であれば適切かについて

は、研究者間で一致があるわけではなく、 $r = .50$ レベルでも「一定の再現性がある」と記述する研究もあるが(小塩, 2016)、これらは質問紙検査を含めての論評であり、作業量を扱う心理検査での先行研究との比較をすべきであろう。D-CAT検査用紙での再検査信頼性を検討したHatta, Yoshizaki, Ito et al. (2012)では、大・短大生が対象者で、1文字抹消条件での再検査信頼性係数は $r = .81$ で、3文字抹消条件の場合は $r = .75$ あった。一定時間内での作業量を指標とする点で共通性があるクレペリン検査では、柏木(1964)によれば、作業量は $r = .96$ 、計算の誤りは $r = .88$ であったとし、生和(1972)では、作業量は $r = .87$ であったとしている。本研究で課題では、タッチペンでの反応様式であったにも関わらず、紙ベースと比肩できるレベルの高い値が得られたことになり、Dig. Str.は高い再検査信頼性を有すると結論できよう。

研究2の目的は妥当性の検討であった。研究2の前文に紹介したように、ストループ検査は遂行機能を評価するという研究は膨大であり、十分にコンセンサスを得ていると見做せるので、作業仮説「実験群である高齢者は統制群である若年者と比較して、作業検査成績は有意に劣る」の検証であった。結果は平均反応時間については、仮説は支持でき、誤反応数について群間差は認めず、支持されないとするものであった。したがって、Dig. Str.は課題の処理に要する反応時間を指標とする場合には、妥当性があると主張できる。

研究1及び2をまとめると、Dig. Str.検査は信頼性も妥当性も高いことが判明したことになり、遂行機能検査として十分な要件を満たすと結論づけることができるが、本検査ではRI検査とSI検査の両方を組み込んだ理由について触れておきたい。

なぜ、両方の課題を採用するかである。一つ目の理由は、人間の情報処理過程の個人差を調べる手段としての可能性にある。RIとSIを構成する認知過程については、「色名文字がインク色と不一致のストループ刺激」(例えば赤色インクで「青」)に対して、SIでは「アカ」と反応することを求めるが、RIでは「アオ」と反応することが求められる。つまり、文字の意味が文字の色の識別に影響を与えることがSIであり、文字の色が文字の意味の識別に影響を与える効果の場合をRIと言う。別な表現をすると、ストループ刺激は、「意味情報」と「色情報」の2つの属性を含んでおり、意味情報の処理を抑制し、色情報の処理を優先させるのが、SIでの認知処理であり、色情報の処理を抑制し、意味情報の処理を優先させるのがRIでの認知処理である。通常、私たちは文字が何色で書かれていようとも、文字の色は背景情報だからその色の意味を考えることは基本的におこなわない。一方で、文字は即座にその意味にアクセスする必要があり、SIに比べてRIは生じにくい、マッチング法という方法をとるとRI効果が検出できると報告されている(松本・箱田・渡辺, 2012; Pritchatt, 1968)。マッチング法に基づくSIでは、文字の意味を認識した後で色を識別する必要があるため、色について意識的に処理をすることになる。つまり、本

当は抑制すべき色情報の処理を意識的に行う必要が出てくる結果、文字のインクの色を無視することが難しくなり、RI が生じやすくなるという。このように、Dig. Str. に両方の課題を組み込むことにしたのは、マッチング法であれば、RI が測定でき、SI と RI とで個人差が測定できる可能性があると考えたためである。

二つ目の理由には、SI と RI ではその背景となる神経基盤が異なるとする報告があるためである。ストロープ課題の関連部位としては、これまで前帯状皮質 (ACC) との関連が指摘されてきた (たとえば、Pardo, Pardo, Janer et al., 1990; Peterson, Skudlarski, Gatenby et al., 1999)。たとえば、Ruff, Woodward, Laurens et al. (2001) は、fMRI を用いて SI と RI では脳活動が異なることを報告している。具体的には前帯状皮質 (ACC) の活性化の程度が異なる (活性化増大と減少) というものである。事象関連電位 (ERP) を用いた Atkinson, Drysdale, & Fulham (2003) は、SI 効果と RI 効果の両方が初期の側頭 N100 成分と頭頂 P100 成分に見られるが、この 2 つの成分は、RI 効果については統制条件と不一致条件で違いが見られたが、SI 効果については、そのような条件差はないと指摘し、SI と RI の干渉が異なる機能メカニズムに起因するとしている。Song & Hakoda (2015) も fMRI を用いた SI と RI での脳活動を検討した。Ruff, Woodward, Laurens et al. (2001) よりも時代が新しく、fMRI の精度や信頼性は高いと考えられる。結果は、SI が両側の中前頭回 (BA9) の活性化を誘発するのに対して、RI では両側の中前頭回 (BA9 および BA10)、内側前頭回 (BA8)、および帯状回 (BA6 および BA32) を含む他の多くの領域で有意な活性化を示したと報告し、SI と RI の干渉が異なる神経メカニズムに起因することを示唆している。また、前頭前皮質と帯状皮質がさまざまな種類の干渉に対して異なる感受性を持っていること、および前頭葉機能不全のたとえば ADHD 患者の inhibition 機能を評価するには、RI が SI よりも有用であることを指摘している。

以上 2 つの理由が、Dig. Str. 試作機に SI と RI を実装した背景である。本研究の信頼性および妥当性の両方で、SI も RI も類似したレベルの評価結果となったが、これらは健常な青年と高齢者を対象にした検討であり、多くの遂行機能評価の臨床場面での利用を想定すれば、2 つのタイプのストロープ課題を検査機器として内包させることには意味があると考えられる。というのは、ストロープ検査は遂行機能を測定・評価するのが目的の検査であり、幼児から高齢者までの発達過程での検討や遂行機能障害が疑われる対象への適用が想定されねばならないからである。もっとも、そのため妥当性の検討には、それらの機能が低下することが想定できる高齢者を対象とした検討以外に、前頭葉機能の障害と理解されている他の対象者群、たとえば、ADHD、TBI 患者を対象に妥当性を検討する必要がある。それらが未実施であることは、本研究 2 での妥当性検討の限界として記し、今後検討する必要がある。

## 謝辞

本研究は科学研究費助成事業、挑戦的萌芽研究 17K18716 (代表者：八田武志) 及び基盤研究 C20K03423 (代表者八田武俊) の助成を得て実施した。資料収集において、梅本篤樹、新宮春輝、竹村龍、田中大夢、玉置直也、寺田侑生、野々山優吾、船越颯真、山本葉大の諸氏に協力を得たことを記し、謝意を表します。

なお、本研究は関西福祉科学大学研究倫理審査委員会により承認 (承認番号 22-21) を経て実施した。

## 引用文献

- Atkinson, C. M., Drysdale, K. A., & Fulham, W. R. (2003). Event-related potentials to Stroop and reverse Stroop stimuli. *International Journal of Psychophysiology*, 47, 1-21.
- Caballero, H. S., McFall, G. P., Wiebe, S. A., & Dixon, R. A. (2021). Integrating three characteristics of executive function in non-demented aging: trajectories classification, and biomarker predictors. *Journal of International Neuropsychological Society*, 27, 158-171.
- Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, 1, 45-75.
- Ferguson, H. J., Brunson, V. E. A., & Bradford, E. E. F. (2021). The developmental trajectories of executive function from adolescent to the old age. *Scientific Report*, 11, 1382.
- 箱田裕司・佐々木めぐみ (1990). 集団用ストロープ・逆ストロープテスト—反応様式、順序、練習の効果—。教育心理学研究, 38, 389-394.
- 箱田裕司・渡辺めぐみ (2005). 新ストロープ検査 II. トーヨーフィジカル.
- 八田武志 (2007). 神経心理学研究における測定と評価について. 神経心理学, 23, 2-7.
- 八田武志 (2022). 遂行機能 (遂行機能). 岩原昭彦 (編). 神経心理学. サイエンス社. (印刷中)
- 八田武俊・重森健太・木村貴彦・八田武志・岩原昭彦 (2022). タブレット版注意機能スクリーニング検査作成の試み I—再検査信頼性の検討—. 人間環境学研究, 20, 97-102.
- Hatta, T., Yoshizaki, K., Ito, Y., Mase, M., & Kabasawa, H. (2012). Reliability and validity of the digit cancellation test: A brief screen of attention. *Psychologia*, 55, 246-256.
- Ikeda, Y., Okuzumi, H., & Kokubun, M. (2013). Stroop/reverse Stroop interference in typical development and its relation to symptoms of ADHD. *Research in Developmental Disabilities*, 34, 2391-2398.
- Kadambi, S., Adballah, M., & Loh, K. P. (2020). Multimorbidity, function, and cognition in aging. *Clinical Geriatric Medicine*, 36, 569-584.
- 柏木繁男 (1964). 内田クレペリン検査の信頼性と妥当性の客観的手法による検討. 心理学研究, 35, 93-95.
- Lezak, M. D. (2004). *Neuropsychological assessment, 4th Ed.* New York: Oxford University Press.

- Loaiza, V. M. & Souza, A. S. (2019). An age-related deficit in preserving the benefits of attention in working memory. *Psychology and Aging*, 34, 282-293.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163-203.
- Madden, D. J., Whiting, W. L., & Huettel, S. A. (2005). Age-related changes in neural activity during visual perception and attention. In Cabeza, R., Nyberg, L., & Park, D. (eds.), *Cognitive Neuroscience of Aging*, Oxford: Oxford University Press, 157-185.
- Mateer, C. A. (2000). Attention. In S. A. Ranskin & C. A. Mateer (eds.), *Neuropsychological management of mild traumatic brain injury*, New York: Oxford University Press, 73-91.
- 松本亜紀・箱田裕司・渡辺めぐみ (2012). マッチング反応を用いて測定したストロープ・逆ストロープ干渉の発達変化. *心理学研究*, 83, 337-346.
- 永原直子・伊藤恵美・岩原昭彦・堀田千絵・八田武志 (2012). 認知機能スクリーニング検査としてのストロープ検査の有用性の検討. *人間環境学研究*, 10 (1), 29-33.
- 奈良雅之・石橋陽介・染谷ゆかり・丸山堅大・依田望 (2010). ストロープ・逆ストロープ課題における色の干渉効果に関する実験的研究. *目白大学心理学研究*, 6, 1-12.
- 岡田謙介 (2015). 心理学と心理測定における信頼性について. *教育心理学年報*, 54, 71-83.
- 小塩真司 (2016). 心理尺度構成における再検査信頼性係数の評価. *心理学評論*, 59, 68-83.
- Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy Sciences*, 87, 256-259.
- Parkin, A. J. & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging*, 7, 290-298.
- Peterson, B. S., Skudlarski, P., Gatenby, C., Zhang, H., Anderson, W. A., & Gore, J. C. (1999). An fMRI study of Stroop Word-Color interference: Evidence for cingulate subregions subserving multiple distributed attentional systems. *Biological Psychiatry*, 45, 1237-1258.
- Prigatano, G. P. (1986). *Neuropsychological rehabilitation after brain injury*. Baltimore: Johns Hopkins University Press. 八田武志 (訳編). 脳損傷のリハビリテーション. 医歯薬出版, 1988.
- Pritchatt, D. (1968). An investigation into some of the underlying associative verbal processes of the Stroop colour effect. *Quarterly Journal of Experimental Psychology*, 20, 351-359.
- Ruff, C. C., Woodward, T. S., Laurens, K. R., & Liddle, P. F. (2001). The role of the anterior cingulate cortex in conflict processing: Evidence from reverse stroop interference. *NeuroImage*, 14, 1150-1158.
- 生和秀敏 (1972). 内田クレペリン精神作業検査の検査反復に伴う変化について. *心理学研究*, 42, 152-164.
- 宋永寧・箱田裕司 (2011). パソコンを用いた新ストロープ・逆ストロープテスト作成および実施効果. *認知心理学研究*, 9, 19-26.
- Song, Y. & Hakoda, Y. (2015). An fMRI study of the functional mechanisms of Stroop/reversed Stroop effects. *Behavioural Brain Research*, 290, 187-196.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 28, 643-662.
- Whyte, J., Schuster, K., Polansky, M., Adam, J., & Coslett, H. B. (2000). Frequency and duration of inattentive behavior after traumatic brain injury: Effects of distraction, task, and practice. *Journal of the International Neuropsychological Society*, 6, 1-11.
- 吉崎一人 (2009). 注意する力 (2章). 唐沢かおり・八田武志 (編著). 幸せな高齢者としての生活. ナカニシヤ出版, 19-35.
- Zhuravleva, T. Y., Alperin, B. A., Haring, A. E., Renz, D. M., Holcomb, P. J., & Daffner, K. R. (2014). Age-related decline in bottom-up processing and selective attention in the very old. *Journal of Clinical Neurophysiology*, 31, 261-271.

(受稿：2022年12月4日 受理：2023年1月14日)