

日本語版拡散連想課題の開発と妥当性

石黒 千晶^{1,*} 鈴木 崇太郎² 服部 雅史³ 安陪 梨沙⁴ 楊 鯤昊⁵

¹ 聖心女子大学現代教養学部心理学科 ² 金沢工業大学大学院工学研究科 ³ 立命館大学総合心理学部
⁴ 立命館大学大学院人間科学研究科 ⁵ 山口大学大学院創成科学研究科

Development and validation of Japanese version of Divergent Association Task

Chiaki Ishiguro^{1,*} Sotaro Suzuki² Masasi Hattori³ Lisa Abe⁴ Kunhao Yang⁵

¹ *Department of Psychology, Faculty of Liberal Arts, University of the Sacred Heart*

² *Graduate school of Engineering, Kanazawa Institute of Technology*

³ *College of Comprehensive Psychology, Ritsumeikan University*

⁴ *Graduate School of Human Science, Ritsumeikan University*

⁵ *Graduate School of Sciences and Technology for Innovation, Yamaguchi*

Divergent thinking (DT) has long been studied as a part of creative potentials; however, the extensive scoring process for DT poses issues. A recent solution gaining attention is automated scoring by calculating semantic distance. The Divergent Association Task (DAT), a simplified version of the existing DT task, allows participants to engage in DT tasks within a shorter time, with immediate and automatic scoring. This study developed and validated the Japanese version of the DAT (DAT-J). Twelve candidate word vectors derived from the Japanese corpus were prepared to calculate DAT scores, and three surveys were conducted to examine the validity and reliability of the DAT-J. In Survey 1, 95 art majors completed an online questionnaire, including the DAT-J and creativity-related scales. Because Survey 1 revealed that the DAT-J may include inappropriate responses for automatic scoring, we modified the DAT-J instructions and developed a web application to validate inappropriate responses. In Survey 2, 358 students participated in the modified version of the DAT-J, the existing DT tasks, and the creativity-related questionnaire. The DAT-J was retested two weeks later. In Survey 3, an experiment was conducted comparing DAT-J responses by handwritten, web app, and web app with validation. Survey results demonstrate the basic validity and reliability of the DAT-J, underscoring its potential utility for assessing DT. Furthermore, the DAT-J answering format did not affect DAT scores. Finally, limitations of the DAT-J and future improvements for measurement accuracy were discussed.

Keywords: divergent thinking (拡散的思考), creativity (創造性), semantic distance (意味的距離), computational scoring (計算論的スコアリング), validation (妥当性確認)

1. はじめに

Guilford (1967) が物事や課題に対して様々な解決法を生み出す力を拡散的思考 (divergent thinking; DT) として定義してから, DT は創造的潜在能力として長年研究されてきた (Runco & Acar, 2012). 代

表的な DT 課題である Alternative Uses Task (AUT; Guilford, 1967) は「れんが」などのオブジェクトの普段とは異なる使い方を考える課題である。回答者は複数のアイデアを回答し, そのアイデアは流暢性 (アイデアの数)・柔軟性 (アイデアのカテゴリ数)・独自性 (各アイデアの特異性)・精緻性 (アイデアの詳細さの程度) などの観点で評価される。これらの採点は頻度集計か熟練した採点者による評価によって行われるが (Reiter-Palmon et al., 2019), 2つの集計方法にはそれぞれ特有の課題がある。頻度

本研究は, 石黒・鈴木 (2022) による探索的研究を発展させ, 実際のデータにより各種妥当性の検証を行ったものである。また, 本研究の関連データは OSF で公開している (<https://osf.io/pdw4m/>)。*

* 責任著者。E-mail: ishigurochiaki37@gmail.com

集計を行う場合は、DT 課題の回答プールが必要になる。例えば、柔軟性の評価には回答のカテゴリ表が必要であるが、事前に DT 課題の回答プールから作成する必要がある。採点者による評価の場合は、採点経験者か採点ルールを十分理解している人を複数人集める必要がある。しかも、この採点は回答者の人数・課題の数・評価観点・回答者のアイデア数を乗じた数だけ行う必要があるため、採点のコストが非常に高い。DT 課題の実施には、大量のデータの蓄積や人的リソースの確保が大きなハードルになるのである。

DT 課題の採点に関する問題を解決するアプローチとして近年注目されているのが自動採点である (Acar & Runco, 2019)。例えば、潜在意味分析 (latent semantic analysis: LSA) は新聞や辞書などの膨大なテキスト情報から作成された言語コーパスによって、単語間の類似度を計算することができる (秋山・内海, 2010; 猪原・楠見, 2011 を参照)。DT 課題においても、LSA を用いた独自性や柔軟性の採点が行われ、人間による採点と比較しても同程度、あるいは、それ以上の信頼性を担保できることが示されている (Acar & Runco, 2014, 2019)。自動採点のオンラインプラットフォームも複数公開されている (e.g., Beaty & Johnson, 2021; Beaty et al., 2021)。最近提案された拡散連想課題 (divergent association task; DAT; Olson et al., 2021) は互いに意味や使い方が異なる名詞を 10 個考えるという DT 課題である。平均 1 分半で実施でき、自動採点も即座に行うことができるより簡便な DT 課題であり、数百名の実験室実験と約 9000 人の大規模データで妥当性が示されている。

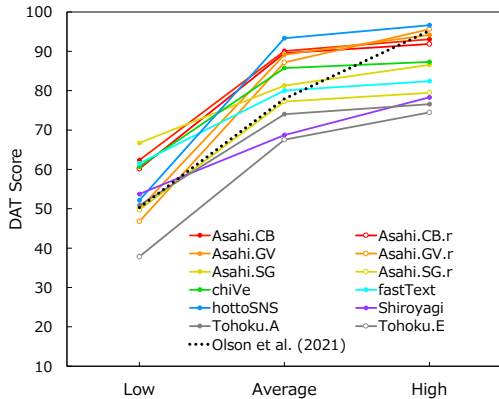
DAT の日本語版を開発する試みも始まっている。石黒・鈴木 (2022) は日本語版 DAT を作成し、Olson et al. (2021) が示した低スコア・平均スコア・高スコアの回答例を当てはめたところ、類似するスコアを算出できることを示している。しかし、日本では DT 課題の自動採点の妥当性はほとんど検証されていない。例えば、自動採点で用いられる分布意味モデル (distributional semantic model), すなわち単語ベクトル (word vectors, 以下, WV) としては、どういったものが適切だろうか。石黒・鈴木 (2022) の日本語版 DAT では、Olson et al. (2021) に倣って、単語分散表現埋め込み手法 GloVe (Pennington, 2014) による WV を用いたが、他の WV を使用する可能

性は検討していない。また、日本語の調査データを用いた自動採点は行なっておらず、日本語版 DAT の妥当性の検討も行われていない。そのため、本研究では日本語版 DAT の開発と妥当性の検証を目指して、様々な日本語 WV の使用可能性を検討するとともに (予備的検討)、調査データを用いて DAT スコアの採点が行えるかどうかを検討する (調査 1)。その上で日本語版 DAT の改良を行いながら、既存の DT 課題や創造性関連指標との関連性があるかどうかを確認し、DAT スコアが DT 指標として基準関連妥当性や再検査信頼性を持っているかどうかを検討する (調査 2)。

なお、調査 2 では既存の DT 課題として SA 創造性検査 (創造性心理研究会, 1969) を用いる。SA 創造性検査は DT 課題の一つである Torrance Test of Creative Thinking (TCTT; Torrance, 1966) をもとに開発された。SA 創造性検査はペーパーテスト形式の DT 課題で、検査会社による採点が行われる。創造性心理研究会 (1969) によると、既存のアイデアプールに基づく頻度集計を行っており、それに基づいて経験のある採点者が採点する。SA 創造性検査は開発当時に信頼性や妥当性が検証されているが、最近の研究においても創造的潜在能力として創造活動や達成、創造性に関わる生物学的指標とも関係が示されている (e.g., Ishiguro et al., 2022; Takeuchi et al., 2010)。SA 創造性検査で得た DT スコア (特に、柔軟性や独自性) と日本語版 DAT の間に、Olson et al. (2021) が示したような正の相関が見られるかどうかを検討することで、日本語版 DAT の妥当性を明らかにすることができる。さらに、創造的自己や開放性などの性格 (Haase et al., 2018; Grajzel et al., 2023; Puryear et al., 2017)、創造活動や達成、行動については DT 課題と正の弱い相関があることが示されており (Said-Metwaly et al., 2022)、開放性以外の性格は DT 課題と負の相関あるいは無相関の関係性が指摘されている (Grajzel et al., 2023; Puryear et al., 2017)。DT 課題の自動採点研究ではこうした創造性関連指標との相関も検討されていることから (e.g., Beaty & Johnson, 2021)、調査 2 では DAT スコアと創造性関連指標との関係も検討する。

さらに、本研究では日本語版 DAT の回答様式についても検討する。原版 DAT はウェブアプリで提供されているが、日本語ではウェブアプリで単語の回答をするにはいくつかの問題が想定される。日本

図1
サンプル回答 (Olson et al., 2021) の WV 別 DAT スコア



註) Low, Average, High の各単語リストは脚注 1 を参照。

語には平仮名・カタカナ・漢字という 3 つの表記があり、同じ単語でも表記の違いによって別の単語と認識される。また、入力時にかな漢字変換が必要で、変換時に候補に接触することが DAT の回答に影響する恐れがある。さらに、日本語には同音異義語が多いことから、適切な表記をしなければ単語の意味が特定できないことがある。こうした日本語特有の影響を検討するため、手書きやウェブアプリによる回答様式によって DAT スコアが影響を受けるかどうかを検討した。

2. 予備的検討

日本語版 DAT スコアを算出するときに、独創的で柔軟なアイデアの転換を捉えることができるのはどの WV だろうか。この問いを検討するために、日本語コーパスを用いた既存の WV で研究上利用可能なもの 12 個 (表 1 および付録 A を参照) を比較した。石黒・鈴木 (2022) と同様に、Olson et al. (2021) が示した低スコア (50 点)・平均スコア (78 点)・高スコア (95 点) の単語リストの例¹⁾を用いて各 WV による DAT スコアを算出した (以降、WV

1) 使用した訳語は以下の通りである。低スコア:「腕 (arm), 目 (eyes), 足 (feet), 手 (hand), 頭 (head), 脚 (leg), 体 (body)」, 平均スコア:「鞆 (bag), 蜂 (bee), ハンバーガー (burger), 祭 (feast), オフィス (office), 靴 (shoes), 木 (tree)」, 高スコア:「カバ (hippo), ジャンパー (jumper), 機械 (machinery), とげ (prickle), チケット (tickets), トマト (tomato), バイオリン (violin)」。ただし、chiVe による算出では、Sudachi の正規化表記に合わせて「カバ」を「かば」、「とげ」を「棘」とした。

の名称には図 1 の略称を用いる)。なお、chiVe は日本語形態素解析器 Sudachi による表記統制²⁾がなされているため、以降のスコア算出は単語表記を Sudachi ver. 0.6.8 によって正規化した上で実施した。図 1 に示されるように、ほとんどの WV は低スコアと平均スコアは十分に区別していたと言える。一方で、平均スコアと高スコアの差が見られにくいものもあった。平均と高スコアの差が比較的大きいものは、Asahi.GV, Shiroyagi, Tohoku.E であった。

3. 調査 1

日本語版 DAT へどのような回答が得られるか、それらを自動採点した場合に DAT スコアが創造性に関わる指標と関係するかどうかを検討した。

3.1 方法

3.1.1 参加者

芸術を専門とする学生 95 名にウェブ上で日本語版 DAT と創造性関連尺度の調査に協力してもらった。芸術大学や芸術専攻に所属する大学教員を通じて学生に調査依頼を呼びかけた。参加者の詳細は、S 1³⁾を参照。

3.1.2 日本語版 DAT

Olson et al. (2021) の教示をもとにして、以下の日本語版の教示を行った (S 2)。具体的には、「言葉の意味や使い方ができるだけ異なる単語を 10 個考えて入力してください。その際、以下のルールを守ってください。」と教示して、以下の 5 つのルールを示した。なお、回答に制限時間は設けなかった。

- ルール 1** 日本語の単語のみを入力してください。
- ルール 2** 名詞のみを入力してください (例: モノ, 物体, 概念など)。
- ルール 3** 固有名詞は入力しないでください (例: 特定の人, 場所など)。
- ルール 4** 特殊な用語は使わないようにしてください (例: 専門用語など)。
- ルール 5** 身の回りのものを見たりせず、自分で単語を考えてください。

DAT スコアは、Olson et al. (2021) と同様、生成された 10 個の単語から WV に認識されない単語を

2) たとえば、「空き缶」「空缶」「空き罐」「空罐」「空きカン」「空きかん」はすべて正規化表記「空き缶」に統制されている (<https://github.com/WorksApplications/chiVe>)。

3) OSF の Supplementary の資料を S と通し番号で示す。

表 1

本研究で使用した学習済み日本語単語ベクトル (WV)

略称	学習方法	単語数	次元	コーパス	収集期間	形態素	版
Asahi.CB	CBOW	747,657	300	朝日新聞記事	1984.8–2017.8	MeCab	
Asahi.CB.r	CBOW+RF	747,657	300	朝日新聞記事	1984.8–2017.8	MeCab	
Asahi.GV	GloVe	747,657	300	朝日新聞記事	1984.8–2017.8	MeCab	
Asahi.GV.r	GloVe+RF	747,657	300	朝日新聞記事	1984.8–2017.8	MeCab	
Asahi.SG	Skip-gram	747,657	300	朝日新聞記事	1984.8–2017.8	MeCab	
Asahi.SG.r	Skip-gram+RF	747,657	300	朝日新聞記事	1984.8–2017.8	MeCab	
chiVe	Skip-gram	2,530,791	300	CC	2022.9–10	Sudachi	v1.3 mc5
fastText	Skip-gram	2,000,000	300	CC	2017.5	MeCab	
hottoSNS	CBOW	2,067,629	200	Blog	2015.1–2016.6	MeCab	
				Twitter	2016	Juman	
				Wiki	2015.11.23	MeCab	
				Web crawling	2009.9–2016.6	MeCab	
Shiroyagi	CBOW?	335,476	50	Wiki	NA	MeCab	
Tohoku.A	Skip-gram	1,511,782	300	Wiki	2019.5.20	MeCab	20190520
Tohoku.E	Skip-gram	760,421	300	Wiki	2019.5.20	MeCab	20190520

註) RF: Retrofitting, CC: Common Crawl (<https://commoncrawl.org/>), Wiki: 日本語 Wikipedia. 学習方法の記載がなくコードパラメータから推定したものは「?」をつけた。作成者や URL、関連文献などの情報は付録 A を参照。

除外し、先頭から 7 個の単語を使用して算出した。

3.1.3 創造性関連尺度

3.1.3.1 パーソナリティ

短縮版ビッグファイブパーソナリティ尺度 (TIPI-J; 小塩他, 2012) の 10 項目を用いて、「1: 全くそう思わない」–「7: 強くそう思う」で回答を求めた。

3.1.3.2 創造的自己

日本語版創造的自己尺度 (Ishiguro et al., 2022) の 11 項目を用いて、7 件法 («1: 全くそう思わない」–「7: 強くそう思う」) で回答を求めた。

3.1.3.3 創造的行動

創造的行動尺度 (Batey, 2007; Batey & Furnham, 2008) を日本語訳して用いた。過去 12 ヶ月間で積極的に関わった活動として、「短編小説を書いた」「実用品を発案して、作った」など 43 個の創造的行動の中から当てはまるものを選択するように求めた。回答の際には「仕事・学業 (大学や専門学校) や趣味など場面は問いません」という説明を追加した。

3.1.3.4 創造活動と達成

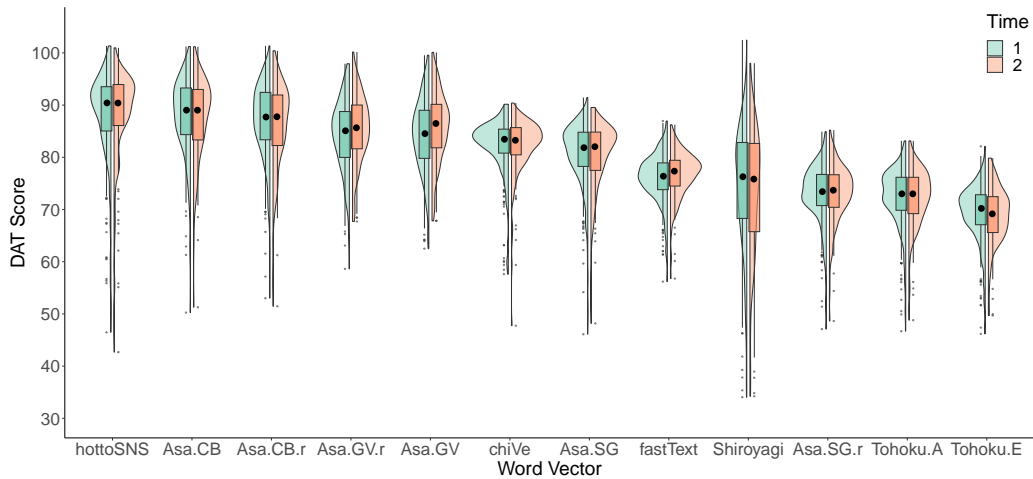
日本語版創造活動・達成尺度 (Ishiguro et al., 2024) を用いた。この尺度は文学、音楽、アート・クラフト、創作料理、スポーツ、視覚芸術、上演芸術、科

学工学の 8 領域それぞれについて、過去 10 年間で実施した創造活動の回数、各領域で達成したレベルを回答してもらった。例えば、文学領域での創造活動は「短い文学作品 (詩、短編小説など) を書いた」「ブログ記事を書いた」などの項目について「0: 一度もない」「1: 1–2 回」「2: 3–5 回」「3: 6–10 回」「4: 10 回より多い」の 5 個の選択肢から回答を求めた。また、創造的達成については「0: 私はこの分野に関わったことがない」「1: この分野の創作活動を一度やってみたことがある」「2: この分野で少なくとも 1 つはオリジナル作品を作成したことがある」などから「10: この分野で自分の作品を販売したことがある」というレベルから、当てはまるものをすべて回答してもらった。

3.1.4 手続き

すべてオンラインで実施した。参加者は、まず Qualtrics[®]XM のウェブフォーム上で研究目的や手続きの説明を読み、同意した場合のみ、創造性関連尺度に順に回答した。続いて、同じウェブフォーム上で日本語 DAT に回答した。本調査は、金沢工業大学 (第 1 著者の調査当時の所属機関) の倫理審査委員会の承認 (承認番号 22-05-023) を受けた上で実施した。

図 2
DAT スコアの分布 (調査 2)



註) Time 1 は, DAT を少なくとも 1 回実施した参加者 ($n = 315$) の初回スコア, Time 2 は, 2 回実施した参加者 ($n = 172$) の 2 回目のスコアを表す. バイオリン図はカーネル密度を, 箱ひげ図は, 中央値, 25/75 パーセンタイル値, 最大値から最小値の範囲 (ただし, パーセンタイル範囲の 1.5 倍を超える外れ値は除外して点描) を表す.

表 2
調査ごとの DAT スコアの記述統計

WV	調査 1 ($n = 95$)			調査 2 ($n = 315$)		
	M (SD)	n_v		M (SD)	n_v	
Asahi.CB	88.2 (6.80)	88		88.1 (7.16)	315	
Asahi.CB.r	86.9 (7.15)	88		87.0 (7.23)	315	
Asahi.GV	85.5 (8.43)	88		84.2 (7.01)	315	
Asahi.GV.r	84.9 (8.70)	88		84.3 (6.85)	315	
Asahi.SG	81.2 (5.08)	88		80.8 (5.82)	315	
Asahi.SG.r	72.9 (6.19)	88		73.1 (5.38)	315	
chiVe	82.5 (4.69)	95		82.3 (5.23)	315	
fastText	77.1 (4.91)	88		76.1 (4.52)	315	
hottoSNS	88.4 (8.14)	93		88.6 (8.08)	315	
Shiroyagi	76.0 (11.13)	92		74.6 (11.43)	315	
Tohoku.A	72.7 (6.29)	92		69.3 (5.38)	315	
Tohoku.E	69.6 (6.27)	87		72.3 (5.56)	315	

註) 調査 2 は初回実施分のみ. WV: WV 略称. n_v : DAT スコア算出可能サンプル数. 調査 1 と 2 のスコアの差はすべて非有意, $ps > .09$.

3.2 結果

3.2.1 DAT 回答の適切性の確認

回答のうち, DAT スコアの算出過程に影響のある回答が数種類見られた. 一つ目は, 「あめ」などの同音異義語を含むひらがなの回答である (5 件). 例えば, 「あめ」には「雨」「飴」など複数の異なる意味の名詞を指す可能性があり, どちらを回答者が意図しているかによって DAT スコアが変わる. 二つ目は, 名詞以外の回答である (4 件). 例えば, 「明

るい (形容詞)」「したたか (副詞)」などの回答が見られた. こうした回答は品詞が異なることで他の回答と DAT スコア算出過程が変わるおそれがある. 三つ目は, 一つの回答欄に複数の名詞を記入したり (1 件), 単一の名詞ではなく「椅子の裏側」のような連体修飾語を書き込むといった回答である (2 件). WV で単語と認識されない不適切な回答は, WV ごとに個別回答をスコア算出対象から除外してスコアを計算した.

3.2.2 WV ごとの DAT スコア

図 2, 表 2 に各 WV による DAT スコアの平均点を示した. 例えば, hottoSNS や Asahi.CB など予備的検討で平均スコアと高スコアの弁別が明確でなかった WV は平均点が高スコアになりやすい傾向が示唆されている. 一方, Tohoku.E や Tohoku.A など低スコアが低得点になりやすい WV ではデータ全体が低いスコアになりやすかった.

3.2.3 DAT スコアと創造性関連尺度の関係

各 WV によって算出された DAT スコアと創造性関連尺度との相関行列 (Pearson の相関係数による, 以下, 相関係数は同様) を算出した (表 3). その結果, DAT スコアと創造的自己, 性格との間に有意な相関関係は見られなかった. 一方で, WV の中には

表3
DAT スコアと創造性関連尺度の相関係数（調査1）

WV	CS	Big-Five					CB	CAct	CAch
		E	A	C	N	O			
Asahi.CB	.00	-.11	-.04	-.25	.13	-.01	.12	.08	.12
Asahi.CB.r	-.01	-.13	-.02	-.27	.14	-.01	.11	.05	.12
Asahi.GV	.11	-.08	-.13	-.11	.11	.07	.27**	.06	.24*
Asahi.GV.r	.12	-.11	-.11	-.11	.09	.09	.26**	.06	.23*
Asahi.SG	-.04	-.13	.01	-.22	.06	-.06	.14	.10	.16
Asahi.SG.r	.01	-.17	.02	-.21	.07	.01	.13	.01	.10
chiVe	.02	-.01	.09	-.06	.08	.00	.19*	.04	.17
fastText	.06	-.16	-.05	-.14	.07	.07	.18*	.07	.25*
hottoSNS	.08	-.06	-.04	-.18	.13	.06	.09	.03	.13
Shiroyagi	.06	.02	.03	-.14	.05	.05	.08	.02	.12
Tohoku.A	.05	-.04	.03	-.03	.06	.08	.10	.04	.18*
Tohoku.E	.08	.08	-.05	-.04	.06	.10	.13	.03	.04

註) CS: Creative self, E: Extraversion, A: Agreeableness, C: Conscientiousness, N: Neuroticism, O: Openness, CB: Creative behavior, CAct: Creative activities, CAch: Creative achievements. 相関の検定は Olson et al. (2021) に従い片側検定. * $p < .05$, ** $p < .01$.

DAT スコアと創造的行動、創造的達成の間に弱い正の相関が見られるものがあった。特に、Asahi.GV, Asahi.GV.r, fastText, chiVe, Tohoku.A で算出された DAT スコアとの間で弱い正の相関が見られた。

3.3 考察

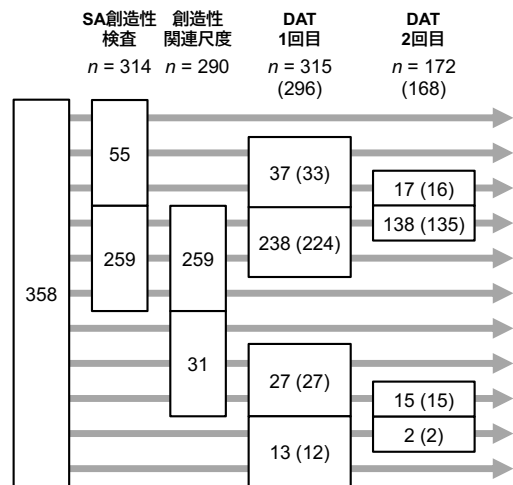
調査1では、自由記述形式で日本語版 DAT を行うことで、実際の回答傾向から DAT スコアの計算可能性を検討した。その結果、自由記述形式ではひらがなの回答によって単語が特定できない場合があること、名詞以外の品詞で回答されるケースがあることが示された。対策としては、教示を付け加えたり、入力時に回答の適切性をリアルタイムにチェックしてフィードバックすることでルールを逸脱しない回答を強制することが考えられる。

また、調査1では DAT スコアと自己報告式の創造性関連指標との関係について予備的な知見を得ることもできた。その結果、一部の WV で算出した DAT スコアと創造的行動、創造的達成との間に正の弱い相関が見られた。これらの結果から、日本語版 DAT は創造性に関係する指標として機能する可能性が示唆された。

4. 調査2

予備的検討と調査1の結果を踏まえて、改良版日本語版 DAT として、教示を調整し、バリデーション機能（後述）を付け加えたウェブアプリを開発し

図3
各課題を実施した参加者の数（調査2）



註) 括弧内の数値はスクリーニング後のサンプルサイズを示す。「DAT 1回目」($n = 315$)には期間2に1回目を実施した参加者 ($n = 55$) を含む。

た4)。調査2では、改良版 DAT が DT 課題として十分な妥当性・信頼性を持つかどうかを検討した。Olson et al. (2021) に従い、(1) 既存の DT 課題との関係、(2) 創造性関連指標との関係、(3) 再検査信頼性を検討した。さらに、既存の DT 課題と DAT の

4) OSF に示した URL にて公開している。公開版は DAT スコアの計算結果を画面表示するようになっており、その点のみ本実験で使用したものと異なる。

回答の印象（楽しさ、興味深さなど）を比較した。

4.1 方法

4.1.1 参加者

調査2では芸術を専門としない大学生を対象に、SA創造性検査とDATを実施する集団実験と質問紙調査を実施した。事前のサンプルサイズ分析から、Olson et al. (2021) で示されたDATとAUTの相関の下限値 $r = .20$ の検定に十分なサンプルサイズとして $N = 193$ ($\alpha = .05$, 検定力 $1 - \beta = .80$) が必要であった。調査2では再検査も含めて欠損値が出ることも想定して200名から300名の回答者を得ることを目標とした。図3にSA創造検査・DAT・質問紙調査それぞれに参加した人数を示した。

4.1.2 改良版日本語版 DAT

改良版日本語版DATによる検査を実施した。参加者は、本調査のために独自に作成したウェブアプリにアクセスして、10個の単語を回答し、性別や年齢などを回答し、最後に回答中にどのようなことを考えていたかを自由記述で回答した。

改良版日本語版DATではOlson et al. (2021) の手続きをもとにしてより詳細な教示(S3)を行うとともに、集団での調査実施を可能にするために(1)計算時間の改善、(2)計算可能な単語のチェックの2点の改良を行った。まず、教示の改善点として、「この検査では名詞の意味が一つに定まるように、できるだけ漢字やカタカナを用いて入力してください」という教示を追加した。この教示によって、ひらがな表記によって同音異義語を回答する確率が低くなり、参加者が想起した名詞どうしの意味的距離をより正確に算出することが可能になる。

次に、参加者の回答した単語が名詞であるかどうかを入力時にチェックして、不適切な回答ではないかをチェックし、不適切であった場合に当該単語を再入力させる（以下、バリデーション）機能を追加した。この機能は原版DATにも用いられており、参加者の不適切な回答を事前に防ぐことができる。バリデーションでは、あらかじめWVに含まれる名詞リスト（辞書）を作成して回答がそのリストに含まれているかをチェックする。辞書は、調査時点で原版DATのWVと類似していると判断したAsahi.GV.r等の朝日新聞WVをベースに作成した。この辞書に含まれる単語で、日本語形態素解析

器 MeCab 0.996 (IPA 辞書 2.7.0)⁵⁾ を用いて形態素解析を行ったときに形態素のいずれかが、(1)「名詞、一般」(2)「名詞、固有名詞、一般」を含むものを抽出した。他にも単語が平仮名表記で入力された際、同音異義語が存在するかを確認するバリデーションを行った（詳細なプログラムコードはOSFに掲載した）。ただし、すべての単語についてバリデーションのクリアを要求すると回答修正が必要以上に繰り返されるおそれがあるため、10個の名詞のうちDATの計算に最低限必要な7個の単語でバリデーションがクリアされない場合のみ再入力を求めるようにした。

4.1.3 SA 創造性検査

SA創造検査のA版（創造性心理研究会、1969）を実施した。SA創造性検査は新聞紙などものの普通とは違う使い方を考えるテスト1、テレビなどがどのようなものであったらよいか考えるテスト2、この世からネズミがいなくなったらどのようなことが起きるのかを考えるテスト3に別れ、それぞれ既存のDT課題のAUT、Product Improvement, Just Supposeに当たる(Guilford, 1967)。検査では3つのテストそれぞれで、練習問題を2分間で回答した後、本題を2問合計5分間で回答を求めた。これらの回答は検査会社のエキスパートが流暢性（適切な回答の数）、柔軟性（回答のカテゴリの数）、独自性（集団内で出現頻度が低い回答の数）、精緻性（回答の具体性）の観点で採点された。なお、分析ではOlson et al. (2021) の妥当性検討の手続きに従って、AUTのみを分析対象とした。

4.1.4 創造性関連尺度

調査1と同じ心理尺度に回答する質問紙調査にオンラインで回答するよう求めた。

4.1.5 創造性に関する自己報告と2つの創造性検査の楽しさ評価

調査1と同様の内容に創造的自己、パーソナリティ、創造的行動、創造活動・達成についての回答を求めた。加えて、Olson et al.(2021)と同様に、DT課題として日本語版DATとSA創造性検査を行っているときの楽しさを測定した。具体的には「1つ目のペーパーテスト(SA-創造性検査)に回答してい

5) <http://taku910.github.io/mecab/>

ときのあなたの気持ちについて教えてください」と「2つ目のオンライン簡易連想検査に回答しているときのあなたの気持ちについて教えてください」という質問をして、「楽しい」「興味深い」「面倒くさい」という項目について「0: 全くそう思わない」～「5: 非常にそう思う」の6件法で回答を求めた。DATとSA創造性検査は回答時間が異なるため、「興味深い」の項目では課題に知的な楽しさがあるか、また、「面倒くさい」の項目では課題時間の長さによるネガティブな影響がないか確認することを目的とした。

4.1.6 手続き

調査2は2回に分かれて実施され、参加者は創造性に関連する授業、あるいは、授業終了後に研究参加を依頼された。第1回目の調査では、実験者の指示のもとSA創造性検査に回答し、続いて口頭あるいは紙面で指示を受けた上でDATに回答した(期間1)。その後、1週間以内に創造性関連尺度の調査に回答するよう依頼した。さらに、2週間後の第2回目の調査で再度DATに回答した(期間2)。本研究は、金沢工業大学(第1著者の調査当時の所属機関)の倫理審査委員会の承認(承認番号22-05-023)および、現所属の倫理審査委員会の承認(承認番号2023-5)を受けた上で実施した。参加者は、研究目的や手続きについて口頭、あるいは紙面上で説明された上で研究参加に同意した場合のみ検査や調査に回答した。

4.2 結果

DATに回答した参加者のうち、期間1に実施したDATに不参加で期間2のみに参加した者($n = 55$)の回答は、期間2の回答を「1回目」のものとして扱った。SA創造性検査と2回のDAT、創造性関連尺度調査の参加者の重なり具合は、図3に示す通りである。なお、Olson et al. (2021)に従って、指示に従っていない回答者を自由記述の回答(例:身の回りにあるものを書いた)をもとにスクリーニングした。DATとSA創造性検査、および、創造性関連尺度調査とのスクリーニング前のデータでの相関分析の結果は、S5,6に記載した。

4.2.1 WVごとのDATスコア

調査1と同様にWVごとにDATスコアを算出し

表4

DATスコアとSA創造性検査の相関係数(調査2)

WV	T	AUT			
		Flu	Flx	Org	Ela
Asahi.CB	1	.10	.19**	.16**	.07
	2	.02	.08**	.18**	.04
Asahi.CB.r	1	.07	.16**	.14*	.03
	2	.02	.08**	.17**	.05
Asahi.GV	1	.13*	.19**	.15**	.15**
	2	.18*	.19**	.15**	.21**
Asahi.GV.r	1	.10	.17**	.16**	.14*
	2	.19	.19**	.13**	.22**
Asahi.SG	1	.09	.19**	.18**	.09
	2	.05	.10**	.20**	.09
Asahi.SG.r	1	.06	.18**	.17**	.06
	2	.04	.12**	.14**	.10
chiVe	1	.02	.12*	.12*	.06
	2	.02	.08*	.09*	.11
fastText	1	.07	.14*	.16**	.09
	2	.12	.14*	.19**	.19
hottoSNS	1	.02	.15**	.12*	.07
	2	.03	.12*	.16*	.11
Shiroyagi	1	.04	.12*	.14*	.00
	2	.02	.07*	.19*	.05
Tohoku.A	1	.06	.14*	.13*	.05
	2	.03	.07*	.17*	.08
Tohoku.E	1	.06	.18**	.19**	.10
	2	.04	.08**	.16**	.05

註) T1/2: 1/2 回目の実施。Flu: Fluency, Flx: Flexibility, Org: Originality, Ela: Elaboration。相関係数はスクリーニング後のデータを実施回別にリストワイズ除去して算出した(T1: $n = 257$, T2: $n = 151$)。相関分析の検定はOlson et al.(2021)に従い片側検定。
* $p < .05$, ** $p < .01$ 。

た(図2)。WVごとのDATスコアは調査1と調査2で一貫しており、回答者が芸術などの創造活動を専攻しているかどうかによっても変わることがなかった(表2)。なお、本研究の調査は大学生を対象にしているが、いずれのWVにおいても性別によるスコアに違いはなく($ps > .05$)、年齢とも有意な相関関係はみられなかった($ps > .05$)。

4.2.2 DATスコアとSA創造性検査の関係

DATスコアとSA創造性検査のテスト1(AUT)と各評価観点の相関係数を算出した(表4)。その結果、AUTの独自性と柔軟性スコアとDATスコアとの間に一貫して有意な正の相関が見られた。一方、流暢性と精緻性とDATスコアの間で有意な正の相関が見られるものは少なかった(例外として、流暢性ではAsahi.GV、精緻性ではAsahi.GV.r, Asahi.GV[いずれも両回目]との間に正の弱い相関が見られ

た)。この傾向はスクリーニング前のデータでも類似していたが、スクリーニング前では WV によっては 2 回目に独自性・柔軟性スコアとの相関が見られなくなるものもあった (S 5)。Olson et al. (2021) で報告された DAT と独自性の相関係数は .50 (スクリーニング前: .13-.35) であったのに対して、本研究の結果では DAT と独自性の相関係数は .09-.20 (スクリーニング前: .10-.21) と弱い相関であった。

4.2.3 DAT スコアと創造性関連尺度の関係

日本語版 DAT スコアと創造性関連尺度との相関係数を算出した (表 5)。調査 2 で得られた SA 創造性検査における AUT の下位尺度得点は、創造的自己、開放性、創造的行動、創造活動・達成のいずれとも有意な正の弱い相関関係が確認された (S 4)。表 5 に示された DAT スコアとの相関関係においても、創造的自己、開放性、創造的行動、創造的活動・達成は DAT スコアと正の弱い相関がみられたものが多く、開放性以外の性格とは負の相関、あるいは、有意な相関関係がみられなかった。WV の中では Asahi.GV, Asahi.GV.r, Tohoku.E, fastText, chiVe では 1 回目も 2 回目も一貫して創造的自己、開放性、創造的行動・達成と有意な正の相関がみられた (S 6 に示すスクリーニング前データでも同様の傾向)。

4.2.4 DAT スコアの再検査信頼性

再検査信頼性については、 r の値が .41-.63 (スクリーニング前データでは .37-.62) であり、検査としては十分とは言いがたい (表 6)。しかし、Olson et al. (2021) が報告した再検査信頼性 $r(48) = .73$ [.57, .84], $p < .001$ を鑑みると、chiVe や hottoSNS では信頼区間の下限を超える再検査信頼性が得られていた。AUT の信頼性と比較した場合には、評定者ほどの信頼性は hottoSNS と chiVe のみでしか得られなかったが、アルゴリズムの場合の信頼性の下限には概ねおさまっていると言える (評定者の場合 $r = .61-.70$, アルゴリズムの場合 $r = .49-.80$, Stevenson et al., 2020)。

4.2.5 DAT と SA 創造性検査の楽しさ評価

SA 創造性検査と DAT について、回答中の楽しさや興味深さ、面倒くささについて参加者に評価してもらった。SA 創造性検査 (以下、SA)、創造性

関連尺度、DAT 1 回目 (以下、DAT) への回答が確認できているデータ ($n = 238$) について、SA と DAT の評価を比較した (S 7)。その結果、楽しさの平均値 (標準偏差) は、DAT で 3.13 (1.31), SA で 3.47 (1.32), 興味深さは、DAT で 3.44 (1.30), SA で 3.80 (1.36), 面倒さは、DAT で 3.53 (1.38), SA で 3.62 (1.35) であった。Olson et al. (2021) では、楽しさと興味深さは DAT の方が AUT よりも高かったが、本調査では DAT よりも SA 創造性検査の方が楽しく、 $t(237) = 4.52, d = 0.26, p < .001$, 興味深い、 $t(237) = 5.49, d = 0.26, p < .001$ と回答された。面倒さに関しては評価に違いは見られなかった、 $t(237) = 1.12, d = 0.07, p = .26$ 。

4.3 考察

調査 2 では日本語版 DAT を改良して、Olson et al. (2021) を踏まえながら、日本語版 DAT の DT 課題としての妥当性を検討した。日本語版 DAT は既存の DT 課題における AUT の柔軟性、独自性と弱い正の相関関係にあることが明らかになった。日本語版 DAT の AUT の相関は、Olson et al. (2021) で示されている相関関係 (独創性: $r = .32-.50$, 柔軟性: $r = .34-.51$) ほど強くないものの、どの WV で計算しても、1 回目の DAT スコアと AUT の柔軟性・独創性の間に正の相関がみられた。この結果は、日本語版 DAT が DT をある程度捉えることが可能であることを示唆している。また、既存の DT 課題として SA 創造性検査における AUT は先行研究と同様に創造的自己、創造的行動、創造的活動・達成と正の弱い相関関係にあったが、DAT スコアにおいても同様の関係がみられた。これらの結果は日本語版 DAT が DT 課題として妥当性を持つことを示唆している。

ただし、日本語版 DAT は、現時点では再検査信頼性という点で限界があるかもしれない。日本語版 DAT の再検査信頼性は原版よりも低かった。chiVe で $r = .63$, hottoSNS で $r = .62$ など、WV によっては評定者信頼性の下限 ($r = .61$) を超えるものもみられたが、日本語版 DAT スコアと創造性関連尺度との正の相関は 1 回目の DAT スコアで認められても、2 回目では有意な相関が見られなくなるものもあった。この結果は、日本語版 DAT は一定期間あけた測定では DT を十分に反映できなくなる恐れを示唆している。また、日本語版 DAT は短時間で

表 5

DAT スコアと創造性関連尺度の相関係数 (調査 2)

WV	T	CS	Big-Five				CB	CAct	CAch	
			E	A	C	N				O
Asahi.CB	1	.08	.02	.02	-.13	-.11	.10	.07	.01	.16**
	2	.12	.05	.01	-.05	-.07	.09	.05	-.04	.07
Asahi.CB.r	1	.08	.01	.02	-.11	-.11	.09	.06	.00	.15**
	2	.10	.06	.02	-.04	-.08	.08	.05	-.06	.06
Asahi.GV	1	.15**	.03	.05	-.10	-.15	.17**	.04	.09	.17**
	2	.17**	.04	.09	-.02	-.08	.14*	.13	.24**	.22**
Asahi.GV.r	1	.15**	.03	.06	-.10	-.13	.17**	.06	.08	.17**
	2	.15**	.03	.10	-.01	-.09	.14*	.13	.21**	.23**
Asahi.SG	1	.14*	.02	.03	-.14	-.10	.16**	.10*	.04	.19**
	2	.12	.06	.03	-.05	-.10	.10	.08	-.03	.10
Asahi.SG.r	1	.15**	.00	.03	-.12	-.12	.16**	.03	.04	.17**
	2	.12	.04	.04	-.01	-.07	.09	.07	.02	.11
chiVe	1	.14*	.02	.04	-.03	-.13	.19**	.08	.03	.15**
	2	.18*	.09	.04	-.01	-.12	.23**	.10	.02	.16*
fastText	1	.20**	.04	.04	-.07	-.16	.23**	.17**	.08	.28**
	2	.19**	.13	.12	.07	-.19	.19*	.16*	.09	.23**
hottoSNS	1	.09	.04	.03	-.02	-.14	.13*	.10*	.06	.16**
	2	.17*	.10	-.02	-.04	-.09	.19**	.08	-.05	.11
Shiroyagi	1	.10	.04	.03	-.05	-.08	.14*	.11*	.02	.17**
	2	.11	.09	.03	-.06	-.03	.08	.06	.03	.09
Tohoku.A	1	.17**	.07	.05	-.02	-.14	.19**	.10	.03	.18**
	2	.17*	.14*	.02	-.02	-.08	.15*	.07	.08	.12
Tohoku.E	1	.16**	.03	.04	-.09	-.05	.16**	.16**	.08	.19**
	2	.24**	.13	.03	-.06	.00	.18*	.12	.17	.18*

註) CS: Creative self, E: Extraversion, A: Agreeableness, C: Conscientiousness, N: Neuroticism, O: Openness, CB: Creative Behavior, CAct: Creative activities, CAch: Creative Achievements. 相関係数はスクリーニング後のデータを実施回別にリストワイズ除去して算出した (T1: $n = 251$, T2: $n = 150$). 相関分析の検定は Olson et al.(2021) に従い片側検定. * $p < .05$, ** $p < .01$, *** $p < .001$.

回答できるというメリットがあるものの、その課題を行うこと自体の楽しさは既存の課題のほうが大きいことが示唆された。

5. 調査 3

調査 1 (手書き) と調査 2 (ウェブ) では DAT の回答様式が異なっていた。調査 2 ではバリデーション機能も導入したが、回答様式によってスコアに大きな影響がないことを確認するため調査 3 を実施した。同時に、日本語特有の問題として、ウェブ等を使って日本語入力する際に必要となる仮名漢字変換の影響も検討した。調査 1 の「手書き」(バリデーションも仮名漢字変換もなし) と調査 2 の改良日本語 DAT (ウェブ+バリデーションあり) に、仮名漢字変換の効果を検証するための「ウェブ+バリデーションなし」の条件を加えて、3 条件間の DAT スコアの違いを検討した。

5.1 方法

芸術を専門としない大学生 165 名 (男性: 46 名, 女性: 110 名, 不明: 9 名; 年齢: $M = 20.0$, $SD = 3.8$) を対象に集団実験を実施した。過去に類似の実験がないため効果量を中程度 ($\eta^2 = .06$) と仮定し, $\alpha = .05$ とすると, $N = 159$ で検定力 ($1 - \beta$) が .8 を超えるため, 160 名程度のサンプル数の確保を目標とした。参加者はランダムに (1) 手書き ($n = 62$), (2) ウェブ+バリデーションなし ($n = 50$), (3) ウェブ+バリデーションあり ($n = 53$) の 3 条件に割り振られた。条件 (1) では, 調査 1 と同様に教示と回答欄が印刷された用紙に回答を記入した。条件 (2), (3) では, 用紙に印刷された QR コードからウェブにアクセスし, 条件 (3) では調査 2 と同様のバリデーションが設けられた。なお, 調査 3 での DAT スコア算出には, 調査 1, 2 で AUT と創造性関連尺度との妥当性が最もバランスよく確認され, 再検査信頼性が高かった chiVe を用いた。なお, 調査 3 は, 第

表 6
DAT スコアの再検査信頼性 (調査 2)

WV	<i>n</i>	<i>r</i>	
chiVe	172 (159)	.63 [.52, .71]	(.59 [.52, .71])
hottoSNS	172 (159)	.62 [.52, .71]	(.63 [.52, .71])
Asahi.SG	172 (159)	.59 [.48, .68]	(.58 [.47, .68])
Asahi.SG.r	172 (159)	.56 [.45, .66]	(.57 [.45, .67])
Asahi.CB.r	172 (159)	.55 [.44, .65]	(.54 [.42, .64])
Asahi.CB	172 (159)	.54 [.42, .64]	(.53 [.41, .63])
fastText	172 (159)	.53 [.41, .63]	(.53 [.41, .63])
Tohoku.A	172 (159)	.48 [.36, .59]	(.49 [.36, .60])
Shiroyagi	171 (158)	.47 [.34, .58]	(.46 [.32, .57])
Asahi.GV.r	172 (159)	.42 [.29, .54]	(.41 [.28, .53])
Asahi.GV	172 (159)	.41 [.28, .53]	(.41 [.27, .53])
Tohoku.E	168 (156)	.37 [.24, .50]	(.42 [.28, .54])

註) カッコ内はスクリーニング後のデータによる値を示す。Pearson 相関係数 *r* は、WV によって DAT スコアが出せないケースをペアワイズ除去して計算した。範囲は 95%信頼区間を示す。すべての相関は有意 ($p < .001$)。

1 著者の現所属の倫理審査委員会の承認 (承認番号 2023-5) を受けた上で実施した。インフォームドコンセントの手続きは調査 2 と同様である。

5.2 結果と考察

回答した有効単語数が 7 個に満たなかった条件 (1) の 1 名を分析から除外した。各条件の DAT スコアの平均値 (標準偏差) は、条件 (1), (2), (3) の順に、81.94 (5.88), 82.77 (6.05), 81.08 (7.13) で、有意差は認められなかった、 $F(2, 161) = 0.91, p = .41, \eta^2 = .01$ 。

続いて、調査 1 と 2 の間に回答様式による違いがないことを確認するため、条件 (1) と (3) の間で 2 回片側検定 (two one-sided test; TOST) による同等性検定 (equivalence test; Lakens, 2017) を実施した。最小効果量 (smallest effect size of interest; SESOI) を Cohen's $d = \pm 0.50$ とすると⁶⁾、両条件の同等性が示された、 $t(101) = -1.94, p = .027$ 。

さらに、仮名漢字変換による影響がないことを確認するため、条件 (1) と (2) の間で同様の検定を行ったところ同等性が示された、 $t(103.54) = 1.895, p = .031$ 。ただし、条件 (2) と (3) の同等性については、 $t(99.9) = 1.30, p = .197$ となり、仮名漢字変換

を前提とするとバリデーション機能が DAT スコアを下げる効果を否定はできない結果となった。回答が受け付けられない場合に、容易に思いつく単語に変更してスコアが低下する可能性は否定できない。仮名漢字変換などの日本語版 DAT 特有の問題については、引き続き検討が必要であろう。

6. 総合考察

本研究は、自動採点も可能な新しい DT 課題である DAT の日本語版を開発し、その妥当性・信頼性を検討するために複数の調査を行なった。予備的検討では、日本語の WV を用いた DAT の計算可能性を検討し、平均スコアと高スコアを弁別できる WV を選択する必要性を指摘した。また、調査 1 では自由記述による DAT でも DAT スコアが計算可能であり、創造的行動や達成との関係が見られることが明らかになった。しかし、同時に日本語特有の表記揺れや、名詞以外の品詞の回答によって計算可能性が下がる恐れも示唆された。調査 2 では日本語表記に関する教示と、無効な回答の再入力を要求するバリデーション機能を追加した改良版を開発した。改良した日本語版 DAT は、既存の DT 課題の得点と有意な正の相関関係にあり、DT と関連する自己報告式の創造性関連尺度とも正の相関関係にあることがわかった。ただし、日本語版 DAT の 2 週間後の再検査信頼性は原版よりも低く、WV によっては DT 課題や創造性関連尺度との関係性が維持されない恐れが示された。なお、調査 3 では日本語版 DAT の回答様式による違いも検討したが、調査 1 と 2 の回答様式の違いは DAT スコアに影響を与えないことが示された。

6.1 日本語版 DAT のメリット

日本語版 DAT を DT の簡易テストとして使用できる可能性を確認することができた。日本語版 DAT を使用することのメリットは、回答時間が短く済むため、参加者への負担が最小限に抑えられることである。そして、自動採点によって採点にかかるコストを抑えることができる。ただし、日本語版 DAT の妥当性・信頼性についてはまだ残された課題があり、現段階での使用には以下に示すような注意が必要である。

6) SESOI の決め方には議論があるが (Lakens, 2017)、たとえば、Hubert, et al. (2024) は、人間と ChatGPT を比較する実験で DAT 課題の成績の差に非常に大きい効果量 ($d = 1.57$) を得ていることから、中程度の効果量 $d = 0.50$ を SESOI とすることには一定の合理性があると考えた。

6.2 日本語版 DAT を用いる上での注意

6.2.1 現時点で反復測定は推奨されない

日本語版 DAT を使用する上で第一に注意すべきなのは、日本語版 DAT にはまだ十分な再検査信頼性が確認されていないことである。現時点では日本語版 DAT を繰り返し測定し、それを DT の向上や減衰と結びつけて解釈することは推奨されない。今後、繰り返し測定を行う実験・調査を行うためには、さらなる検討が必要である。また、DAT スコアや単語どうしの距離がわかると回答者がスコアを戦略的に高めやすくなる可能性があるため (Olson et al., 2021), DAT スコアを知らせた後に DAT や関連課題を実施しないほうがよいであろう。なお、再検査信頼性が低いことの背景には、後述する日本語版 DAT の計算精度と認知過程にトレードオフ関係が考えられるため、この課題を克服するような手続きや方法の検討も必要である。

6.2.2 標本ごとに評価基準を検討する

Olson et al.(2021) は GloVe を用いたときの低・平均・高スコアを示していたが、本研究の日本語版 DAT では WV によっても異なる傾向がみられた。WV によって DAT スコアのレベル弁別の度合いが異なり、Olson et al.(2021) が示した平均・高スコアを弁別できる WV は限られていた。現時点では、回答者集団の中で DAT スコアの高低を判断するにとどめる必要がある。また、他の指標との関係性の中で回答者の DT がどのように機能したかを議論するにとどめるべきだろう。スコアの評価基準の設定は、個人の DT 能力を検査するようなハイステークテストにつながるため、慎重に議論する必要がある。

6.2.3 使用する WV を検討する

原版 DAT は AUT の自動採点に Common Crawl をコーパスとして GloVe によって学習された WV が最も適していることから (Dumas et al., 2021), DAT の自動採点にも同様のベクトルを用いた。本研究でも同様の日本語版 WV を探したものの、現時点では同様の日本語 WV で使用可能なものを見つけることができなかった。しかし、本研究で検討した日本語 WV で計算した DAT スコアは、一貫して既存の DT 課題の独自性と柔軟性と正の相関を示していたことから、日本語版 DAT には DT 測定の妥当性

があると言えるだろう。ただし、他の創造性関連尺度との関係や再検査信頼性の結果を鑑みると、すべての妥当性・信頼性の観点で優れた WV は限られる。本研究では chiVe がバランスのとれた妥当性を示しており、再検査信頼性も最も高かった。chiVe は本研究の中で使用した WV の中では最も単語数が多く、原版 DAT と同じ Common Crawl をコーパスとしている。また、他の WV にない chiVe の大きな特徴は、日本語独特の表記の揺れを吸収するため、コーパスの単語をすべて正規化²⁾した上で意味モデルを構成している点である。こうしたことから DAT に親和性が高いのかもしれない。ただし、本研究で示された妥当性や信頼性は原版 DAT と比較すると十分に高いとは言い難い。今後も DAT 採点に有効な WV を検証する必要がある。

6.2.4 日本語版 DAT の回答様式を考慮する

翻訳版 DAT を開発するにあたって、日本語特有の表記や同音異義語の問題に直面した。特に、表記の変換で示される候補リストは DAT 回答中の認知過程に影響し、テストの実質的な妥当性を損なうおそれがあった。また、DAT はいわば逆連想ゲームのような珍しい課題であり、ルールも複数あるため課題に沿った回答が得られにくいこともある。これらの問題について検討するために、本研究では日本語版 DAT について (1) 手書き、(2) バリデーションあり/なしでのウェブアプリでの回答を比較した。回答様式による DAT スコアの違いは見られず、同質性が示されたが、回答様式によって参加者の認知過程が変わるおそれがある。回答中の認知過程の実質的妥当性という点では、手書きは日本語表記の変換候補からの刺激を防ぐことができ、バリデーションなしのアプリは教示や再試行を最低限にして課題への集中を高めることができる。一方、バリデーションありのアプリは再入力の手続きが多いと回答者のフラストレーションを感じさせてしまうが、計算可能な回答を増やすことができる。これらの回答様式の利点・欠点はそれぞれトレードオフ関係にあるため、本来測定したい柔軟で独自のアイデアを生み出す思考過程から離れない範囲でデータ欠損を減らすように研究の手続きを検討する必要がある。

6.2.5 日本語版 DAT の対象年齢を考慮する

既存の DT 検査は幼児や児童にも適用されている

が、日本語版 DAT は現時点では幼児や児童、中高生に使用可能かどうかは明らかになっていない。日本語版 DAT には表記に関する教示が追加されているため、少なくとも児童期後半以降でなければ回答が難しいだろう。児童期から青年期にかけては DT を始めとする創造的潜在能力が発達する時期でもあるため、DAT に回答できたとしてもその得点の評価については慎重になる必要がある。今後日本語版 DAT の適用範囲を広げる上では、多様な年齢層での妥当性や信頼性を検討するための調査・実験を行う必要がある。

6.3 本研究の限界と今後の展望

本研究では日本語版 DAT の改良を試み、DT 課題としての妥当性や信頼性を検討した。しかし、本研究の日本語版 DAT の妥当性検証にはいくつかの限界がある。まず、本研究の DT 課題の採点は頻度集計のみを用いて、採点者による採点との妥当性検討は行っていない。調査 2 で用いた SA 創造性検査は頻度集計を用いているため、本研究では人間の行なった採点を基準とした場合の妥当性は検討されていないことに注意されたい。今後はそうしたさまざまな採点方法を用いて妥当性を検討すべきである。

また、Olson et al. (2021) は回答そのものを楽しむことができる課題として DAT を提案し、教育現場やワークショップでの効果測定にも活用できる可能性を示唆していた。日本語版 DAT は現時点で十分な再検査信頼性が認められていないが、創造性教育やワークショップに活用するためには、DAT が既存の DT と同様に教育的介入による変化を測定できるかどうかなど、より詳しい教育測定の妥当性を検証する必要がある。創造性に関わる精度の高い指標を開発することは教育開発のための基盤となる。DAT のような簡易検査の開発はエビデンスに基づく創造性教育を支える取り組みと言えるだろう。

なお、近年は DT の自動採点を英語以外の多言語に広げる試みも行われている。DAT に関して言えば、日本語以外にはイタリア語 (Puvia et al., 2022) と中国語 (Hubert et al., 2024) での DAT が報告されているが、最新の研究では、多言語変換ニューラル・ネットワークモデルを用いて 12 言語で (残念ながら日本語は含まれていない) AUT 課題の自動採点を行うことに成功している (Patterson et al., 2023)。今後、日本での創造性研究を促進するためにも、日

本語での DT や創造性の採点自動化を検討することは喫緊の課題と言える。

文 献

- Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, 26(2), 229–238. <https://doi.org/10.1080/10400419.2014.901095>
- Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 153–158. <https://doi.org/10.1037/aca0000231>
- 秋山 哲史・内海 彰 (2010). 概念間の関係に関する単語の意味空間の性質：コーパス、構築手法、文章単位による影響。認知科学, 17(1), 110–128. <https://doi.org/10.11225/jcss.17.110>
- Batey, M. (2007). A psychometric investigation of everyday creativity. Unpublished Doctoral Thesis. University College London.
- Batey, M., & Furnham, A. (2008). The relationship between measures of creativity and schizotypy. *Personality and Individual Differences*, 45, 816–821. <https://doi.org/10.1016/j.paid.2008.08.014>
- Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2), 757–780. <https://doi.org/10.3758/s13428-020-01453-w>
- Beaty, R. E., Zeitlen, D. C., Baker, B. S., & Kenett, Y. N. (2021). Forward flow and creative thought: Assessing associative cognition and its role in divergent thinking. *Thinking Skills and Creativity*, 41, 100859. <https://doi.org/10.1016/j.tsc.2021.100859>
- Dumas, D., Organisciak, P., & Doherty, M. (2021). Measuring divergent thinking originality with human raters and text-mining models: A psychometric comparison of methods. *Psychology of Aesthetics, Creativity, and the Arts*, 15(4), 645–663. <https://doi.org/10.1037/aca0000319>
- Grabe, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *Proceedings of the 11th International Conference on Language Resources and Evaluation*. <https://doi.org/10.48550/arXiv.1802.06893>
- Grajzel, K., Acar, S., & Singer, G. (2023). The Big Five and divergent thinking: A meta-analysis. *Personality and Individual Differences*, 214, Article 112338. <https://doi.org/10.1016/j.paid.2023.112338>
- Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.
- Haase, J., Hoff, E. V., Hanel, P. H., & Innes-Ker, A. (2018). A meta-analysis of the relation between creative self-efficacy and different creativity measurements. *Creativity Research Journal*, 30(1), 1–16.
- Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1), Article 3440. <https://doi.org/10.1038/s41598-024-53303-w>
- 猪原 敬介・楠見 孝 (2011). 潜在意味解析に基づく概

- 念間類似度の心理学的妥当性：言語統計解析アプローチの効用と限界 心理学評論, 54(12), 101–122. <https://doi.org/10.24602/sjpr.54.2.101>
- Ishiguro C, Sato Y, Takahashi A, Abe Y, Kato E, Takagishi H (2022) Relationships among creativity indices: Creative potential, production, achievement, and beliefs about own creative personality. *PLoS ONE*.17(9), Article e0273303.<https://doi.org/10.1371/journal.pone.0273303>
- Ishiguro, C., Matsumoto, K., Agata, T., & Okada, T. (2022). Development of the Japanese Version of the Short Scale of Creative Self. *Japanese Psychological Research*. <https://doi.org/10.1111/jpr.12418>
- Ishiguro, C., Matsumoto, K., Agata, T., & Okada, T. (2024). Development of the Japanese Version of the Inventory of Creative Activities and Achievements. [Manuscript submitted for publication]. Department of Psychology, University of the Sacred Heart.
- 石黒 千晶・鈴木 崇太郎 (2022). 拡散的思考課題の採点自動化の試み: Divergent Association Task 日本語版の開発 日本電子情報通信学会技術報告, 122(166), HCS2022-39, 18–20.
- Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>
- 真鍋 陽俊・岡 照晃・海川 祥毅・高岡 一馬・内田 佳孝・浅原 正幸 (2019). 複数粒度の分割結果に基づく日本語単語分散表現 言語処理学会第 25 回年次大会発表論文集, 1407–1410.
- 松野 省吾・水木 栄・榊 剛史 (2019). 日本語大規模 SNS+Web コーパスによる単語分散表現のモデル構築 第 33 回人工知能学会全国大会論文集, 1–3. <https://doi.org/10.11517/pjsai.JSAI2019.0.4Rin113>
- Olson, J. A., Chmoulevitch, N. D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, 118(25), Article e2022340118. <https://doi.org/10.1073/pnas.2022340118>
- 小塩 真司 & 阿部 晋吾 (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(1), 40–52. <https://doi.org/10.2132/personality.21.40>
- Patterson, J. D., Merseal, H. M., Johnson, D. R., Agnoli, S., Baas, M., Baker, B. S., Barbot, B., Benedek, M., Borhani, K., Chen, Q., Christensen, J. F., Corazza, G. E., Forthmann, B., Karwowski, M., Kazemian, N., Kreisberg-Nitzav, A., Kenett, Y. N., Link, A., Lubart, T., ... Beaty, R. E. (2023). Multilingual semantic distance: Automatic verbal creativity assessment in many languages. *Psychology of Aesthetics, Creativity, and the Arts*, 17(4), 495–507. <https://doi.org/10.1037/aca0000618>
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Puvia, E., Taibi, D., & Tressoldi, P. (2022). The role of masked solutions in the accuracy of an insight problem-solving task. *Journal of Anomalous Experience and Cognition*, 2(1), 144–157. <https://doi.org/10.31156/jaex.23304>
- Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts*, 11(1), 59–68. <https://doi.org/10.1037/aca0000079>
- Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 144–152. <https://doi.org/10.1037/aca0000227>
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity research journal*, 24(1), 66–75. <https://doi.org/10.1080/10400419.2012.652929>
- Said-Metwaly, S., Taylor, C. L., Camarda, A., & Barbot, B. (2022). Divergent thinking and creative achievement—How strong is the link? An updated meta-analysis. *Advance online publication*. <https://doi.org/10.1037/aca0000507>
- 創造性心理研究会 (編) (1969). S-A 創造性検査 <A 版> 東京心理
- Suzuki, M., Matsuda, K., Sekine, S., Okazaki, N., & Inui, K. (2018). A joint neural model for fine-grained named entity classification of wikipedia articles. *IEICE Transactions on Information and Systems*, E101.D(1), 73–81. <https://doi.org/10.1587/transinf.2017SWP0005>
- 田口 雄哉・田森 秀明・人見 雄太・西鳥羽 二郎・菊田 洸 (2021). 同義語を考慮した日本語の単語分散表現の学習 情報処理学会研究報告, 2017-NL-233(17), 1–5.
- Takeuchi H, Taki Y, Sassa Y, Hashizume H, Sekiguchi A, Fukushima A, et al. (2010). Regional gray matter volume of dopaminergic system associate with creativity: evidence from voxel-based morphometry. *Neuroimage*.51, 578–585 <https://doi.org/10.1016/j.neuroimage.2010.02.078>
- Torrance, E. P. (1966). The Torrance Tests of Creative Thinking: Norms, technical manual. Research Edition. Verbal and figural tests (Forms A and B). Prentice Hall.

付 録

A. 学習済み日本語 WV についての補足情報

- Asahi.CB/CB.r/GV/GV.r/SG/SG.r** 朝日新聞 WV (朝日新聞社/榊レトリバ; 田口他, 2017). https://cl.asahi.com/api_data/wordembedding.html
- chiVe** chiVe: Sudachi と NWJC による日本語 WV (国立国語研究所/榊ワークスアプリケーションズ; 真鍋他, 2019). <https://github.com/WorksApplications/chiVe>
- fastText** Pre-trained word vectors for 157 languages (Facebook Inc.; Grabe et al., 2018). <https://fasttext.cc/>
- hottoSNS** hottoSNS-w2v: 日本語大規模 SNS+Web コーパスによる単語分散表現モデル (榊ホットリンク; 松野他, 2019). <https://github.com/hottolink/hottoSNS-w2v>
- Shiroyagi** word2vec 学習済み日本語モデル (榊白ヤギコーポレーション) <https://ai.s.shiroyagi.co.jp/2017/02/japanese-word2vec-model-builder/>
- Tohoku.A/E** 日本語 Wikipedia エンティティベクトル (Suzuki et al., 2018). <http://www.cl.ecei.tohoku.ac.jp>

/~m-suzuki/jawiki_vector/