

ウェブ調査におけるパラデータをを用いた 心理尺度回答への『つまずき』検出法の開発

代表研究者	分 寺 杏 介	神戸大学大学院経営学研究科	准教授
共同研究者	馬 場 絢 子	奈良女子大学生生活環境学部	助教
	福 島 綾	一宮メンタルクリニック	心理職

1 本研究の背景

心理学をはじめとした、主に社会科学の諸領域では、対象者の心理的な特性（性格・気分・思考など）を捉えるために心理尺度が幅広く用いられてきた。また臨床心理学の業務においては、心理尺度への回答内容がクライアントの状態や特性の診断等に用いられることもあり、研究以外の場面でも心理尺度は用いられている。心理尺度のなかでも代表的な形式であるリッカート尺度(Likert, 1932)では、表 1 に示すように、各選択肢に程度表現を用いて、回答者自身がどの程度当てはまるかを回答する。

表 1 リッカート尺度の例

	全く 当てはまらない	当てはまらない	あまり 当てはまらない	やや 当てはまる	当てはまる	非常に 当てはまる
初対面の人とすぐに打ち解ける	1	2	3	4	5	6
今の仕事が好きだ	1	2	3	4	5	6

一般的に、研究者は心理尺度を利用する際に、回答された内容自体を疑うことは多くない。もちろん適当に回答している場合 (Satisficing, 努力の最小限化, Krosnick, 1991) や、自分をより良く見せようと意図的に回答を歪める (faking, e.g., Benton & Kornhauser, 1948 年; Ones & Viswesvaran, 1998) といった回答行動を検出するような試みは多数存在する (e.g., Blazek & Siegel, 2023 年; Goldammer et al., 2023; Mazza et al., 2020)。ただし、少なくとも回答者が誠実に質問項目に反応していると信じた場合には「回答者は質問項目を正しく理解し、聞かれた内容に対して自分の状態を正しく把握し、正しく反映した選択肢を選んで」ことを暗に仮定して、回答データを利用していることが多いだろう。

しかし実際には、心理尺度を含む質問項目に正しく回答するためには、高度な認知能力が要求される。具体的にどのような認知プロセスが含まれるかに関しては、Cognitive aspects of survey methodology (CASM: Jabine et al., 1984; Schwarz, 2007) と呼ばれる研究領域として確立されている。なかでも有名なプロセスモデルの 1 つが Tourangeau らによる 4 段階モデル (Tourangeau, 1984; Tourangeau et al., 2000) である。Cannell et al. (1981) のモデルに着想を得た 4 段階モデルでは、質問項目への回答までに「理解 (Comprehension)」「想起 (Retrieval)」「判断 (Judgment)」「回答 (Response)」という 4 つのプロセスを経ると考えている。まず理解の段階では、質問項目の意味を文字通り「理解」し、その意図を捉え、正しく回答するためにはどのような情報が必要となるかを特定する、などの処理が行われる。続く想起の段階では、実際に回答を決定するために必要な情報を自身の記憶などから引き出すなどの処理を行う。ここで集められた情報は、判断の過程で重み付けや集計がなされ、もしも十分な情報を想起できないような場合には、部分的な情報に基づく推測などが行われる。最後の回答の段階では、自身の状態の判断の結果を、提示された選択肢に当てはめる作業が行われ、場合によっては意図的に選択肢を変更することも行われ、これが faking などにつながっていく。4 段階モデルに示された認知過程は、必ずしも一方通行で連続的に行われるものではなく、場合によっては前の段階に戻って考え直したり、特定の過程をスキップするといった可能性もあるが、基本的に「正しく」回答している場合には、4 つの過程を経ていると考えられる。そして、例えば質問項目を理解する能力を回答者が十分に持っていなかったり、すべての認知過程を丁寧にたどるだけのモチベーションが無いような場合に、satisficing が発生すると考えられている (Krosnick, 1991)。

以上のように、質問項目に回答するという行為は高度に知的な作業であると言える。この点に対して馬場他 (2025) では、臨床場面で心理尺度を実施した経験を持つ臨床心理士・公認心理師を対象とした調査によって、一部の回答者は質問項目に正しく回答できていない可能性 (「つまずき」) を明らかにしている。自記式の心理尺度が、上述のように「誠実に回答しているならば正しく理解しているはずだ」という前提のもとで実施され分析される場合、誠実に回答しようとしたが正しく理解できなかった (あるいは他の「つまずき」

が生じた) 回答結果は、適切に解釈されない場合がある。例えば成人の ADHD を診断するための尺度である CAARS (Conners' Adult ADHD Rating Scales, Conners et al., 1999) には「矛盾指標」というものが含まれている。これは、例えば「1 つの場所に長時間いることが難しい」と「じっと座っているときでも、内心は落ち着かない」のように、内容的に類似しており、正しく理解して回答したならば似たような選択肢を選ぶと考えられる項目のペアについて、回答内容の不一致度を算出するものである。この矛盾指標の得点が高い場合には、不適切な回答とみなされてしまうこともあるが、もしも項目の内容を正しく理解できていないなど、別の理由で回答が一致していないとしたら、特に臨床場面ではその事実自体が重要な情報となる。

心理尺度項目に対して感じた「つまずき」を調べるための最もナイブな方法は、回答直後に直接尋ねることであろう。例えば American Educational Research Association et al. (2011) による心理・教育測定のスランダードでは、心理尺度の妥当性の検証方法の一つとして、回答中の思考内容を尋ねる認知インタビュー(e.g., Meadows, 2021; Willis, 2005)が挙げられている。ここではあくまでも測定項目の妥当性評価の目的で挙げられているが、個人レベルの「つまずき」を特定するための方法として利用することも可能である。しかし、臨床場面や学術調査において、一人ひとりに認知インタビューを行うことは現実的に不可能である。このため、回答者にできる限り追加の負荷を与えることなく「つまずき」を検出・評価する方法の開発が望まれる。

回答者に追加の負荷を与えずに、より豊富な情報を引き出すためには、回答内容以外の情報を利用することが有望である。特に社会調査の領域では、回答時間やアクセスログをはじめとした、回答内容以外の情報をパラデータ (paradata: Huvila et al., 2024; Kreuter, 2013) と呼び、主に調査の改善に役立てている。また、認知心理学などでは課題遂行中のマウスカーソルの動きを分析に利用する研究も数多く見られる(e.g., Di Palma et al., 2022; Yamauchi & Xiao, 2018)。この背景には、回答内容や回答時間といった「結果」の情報では、回答に至る認知・思考の「過程」に迫ることは難しいという考えがあり、また手の動きと脳活動の間に見られる密接な関係性より、マウスカーソル軌跡の利用可能性が期待されている (Freeman, 2018)。

以上より本研究では、パラデータとしてマウスカーソル軌跡を利用して、回答者が「つまずき」を感じていたか、具体的にどのような「つまずき」を感じていたかを検出するための方法論開発を目的として、いくつかの基礎的な研究を行った。具体的には、研究 1 として、まず回答者自身が感じた「答えにくさ」を主観的に評価できるのかを検証した。これは、最終段階で利用する教師データの作成可能性の検討も兼ねている。研究 2 では、発話思考法を用いて、実際に「つまずき」を感じた際にどのような思考プロセスが発生しているかを検証する実験を行った。研究 3 ではマウスカーソルの軌跡を用いて、回答者が「つまずき」を感じていたかを予測する機械学習モデルを試行し、性能の評価を行った。

2 研究 1：主観的な答えにくさの評価

2-1 研究の背景と目的

研究 1 では、回答者自身が各質問項目に対して具体的にどのような答えにくさを感じたかを評価してもらうための評価項目の設計を目指す。これにより、最終目的である多様な回答困難感の検出モデル構築のための教師データをより精度良く収集することが期待された。

主に心理検査において、作成した項目の妥当性や「答えやすさ」に言及したチェックリストは、これまでもいくつか提案されてきた (e.g., COSMIN; Mokkink et al., 2024; QAS-99; Willis & Lessler, 1999)。しかし QAS-99 は項目作成者の自己評価に基づくものであり、回答者目線での評価が十分にできない。COSMIN では妥当性や feasibility 等の評価のために回答者本人に対する認知インタビューを推奨しているが、前述の通りこれをすべての回答者に実施することは難しい。

本研究では、心理尺度に対する「つまずき」の主な原因と考えられる要因を定量的に評価する項目 (以下「評価項目」) の作成を目的とする。作成した評価項目および特定の「つまずき」を誘発すると想定される項目 (以下「回答困難項目」) を用いたウェブ調査を行い、作成した評価項目による分析例を示す。

2-2 予備調査

予備調査は 2 度に渡って実施した。2-3 では、予備調査の結果を踏まえて改良した調査形式を採用している。

2-2-1 第 1 回予備調査

調査協力者 クラウドソーシングサービス (クラウドワークス) を通じて、パソコンを用いて参加可能な

18-69 歳の日本語母語話者を募集した。500 名の募集に対して、有効回答数は 491 件（男性 251 件、女性 236 件、その他 4 件）であった。

評価項目 馬場他（2025）の結果を参考に、心理尺度項目への回答の際に想定される認知的な「つまずき」を直接尋ねる個別評価項目（B1-7）を 7 項目作成した。また、システムユーザビリティスケール日本語版（山内，2015）を参考に「答えやすさ」の総合評価項目（A1-3）を 3 項目作成した（表 2）。

表 2 予備調査第 1 回で使用した評価項目の一覧

B1	この項目にはわかりにくい言葉が入っていると感じた
B2	この項目への回答は、その時々で変わりやすいと感じた
B3	この項目の意味がわかりにくかった
B4	この項目は選択肢の違いがわかりにくかった
B5	この項目が指す内容は想像しにくかった
B6	この項目に関連することを思い返すのが大変だった
B7	この項目に正直に回答するのをためらった
A1	この項目は回答しやすいと感じた
A2	この項目はたいいていの人が簡単に回答できるだろうと感じた
A3	この項目に正しく回答できたと確信している

回答困難項目 評価項目（B1-7）に対応すると考えられる 8 項目を独自に作成した。これに加えて、心理尺度や質問紙調査の教科書等で言及されることが多いダブルバーレルを含む項目と、曖昧な副詞を用いた項目を 1 項目ずつ作成した（表 3）。

表 3 予備調査第 1 回で使用した回答困難項目の一覧

1	生活が光風霽月でないと感じることはない
2	出かけるときは身体が重い
3	いつもより、それはどうしたんだろうと思う
4	野鳥観察をすると心が安らぐ
5	飛行機を操縦するとき、緊張より楽しみのほうが大きい
6	先週の水曜日は良い一日を過ごせたと思う
7	オペラを見るなど、趣味や好きなことをしたい気持ちがある
8	友人知人に嫌味を言いたくなることがある
9	読書をしたり、体を動かしたりするのが好きだ
10	いつも朝食をしっかり食べている

調査の手続き 調査は 2 つのセクションによって構成された。セクション 1 では、回答困難項目に不適切回答検出用の項目（私はパソコンを持っている）を加えた計 11 項目をランダムに提示し、「全く当てはまらない」「当てはまらない」「あまり当てはまらない」「やや当てはまる」「当てはまる」「非常に当てはまる」の 6 件法で回答を求めた（図 1）。続くセクション 2 では、画面上部に回答困難項目と、セクション 1 で選んだ選択肢を 1 問ずつ表示し、ランダムな順序で提示された各評価項目について 1（まったくそう思わない）から 5（非常にそう思う）の数字から選択を求めた。

以下の文は、あなたにどの程度当てはまりますか

いつもより、それはどうしたんだろうと思う

全く 当てはまらない	当てはまらない	あまり 当てはまらない	やや 当てはまる	当てはまる	非常に 当てはまる
---------------	---------	----------------	-------------	-------	--------------

図 1 予備調査第 1 回のセクション 1 の回答画面

いつもより、それはどうしたんだろうと思う

全く
当てはまらない

当てはまらない

あまり
当てはまらない

**やや
当てはまる**

当てはまる

非常に
当てはまる

上記の項目について、以下の質問に感じたことを
1(まったくそう思わない)から5(非常にそう思う)までの
5段階評価の数字で教えてください。

すべての質問に、考え込まないで直感的に評価してください。
また、答えられない質問には3をつけてください。

	まったくそう思わない			非常にそう思う	
	1	2	3	4	5
この項目への回答は、その時々で変わりやすいと感じた	☐	☐	☐	☐	☐
この項目の意味がわかりにくかった	☐	☐	☐	☐	☐
この項目に関連することを思い返すのが大変だった	☐	☐	☐	☐	☐
この項目は選択肢の違いがわかりにくかった	☐	☐	☐	☐	☐
この項目にはわかりにくい言葉が入っていると感じた	☐	☐	☐	☐	☐
この項目に正直に回答するのをためらった	☐	☐	☐	☐	☐
この項目が指す内容は想像しにくかった	☐	☐	☐	☐	☐
この項目に正しく回答できたと確信している	☐	☐	☐	☐	☐
この項目はたいていの人が簡単に回答できだろうと感じた	☐	☐	☐	☐	☐
この項目は回答しやすいと感じた	☐	☐	☐	☐	☐
	1	2	3	4	5

図 2 予備調査第 1 回のセクション 2 の回答画面

結果 表 4 に、回答困難項目に対する各評価項目の平均値を示した。表中の太字および下線は、それぞれ全 11 項目の中で最も「答えにくい」「答えやすい」と判断された項目を指す。なお、評価項目 B については、「5：非常にそう思う」に近いほど答えにくいと考えられるため、評価項目 A とは値の意味が逆転している点に注意されたい。結果から、まず客観的事実を尋ねており答えにくさがないと考えられる不適切回答検出用の項目（S）が、ほとんどの評価項目で最も「答えやすい」と評価されたことが分かる。特に評価項目 A における平均値がいずれもほとんど 5 に近い値を示しており、回答者は少なくとも提示された質問項目に対して答えやすいかどうかの判断は十分にできると考えられる。一方で、具体的な「つまずき」の内容を問う項目である B1-B7 については、すべての項目について回答困難項目 1 または 3 が最大値となった。この結果は、回答者が「つまずき」の詳細な内容までは弁別できていない可能性を示唆している。この結果は、各評価項目間のマルチレベル相関係数（表 5）からも明らかである。評価項目 B1-B7 間の相関係数は非常に高く、「答えにくさ」に対して、全体的な印象と個別的な要因を十分に区別することは容易ではないと考えられる。

表 4 予備調査第 1 回の回答困難項目に対する各評価項目の平均値

	1	2	3	4	5	6	7	8	9	10	S
B1	4.72	1.49	3.29	1.38	1.71	1.44	1.67	1.39	1.67	1.19	<u>1.10</u>
B2	3.02	3.49	3.77	2.19	2.21	3.34	2.42	3.10	2.61	2.13	<u>1.41</u>
B3	4.57	1.64	4.30	1.39	2.26	1.51	1.76	1.44	2.15	1.15	<u>1.12</u>
B4	3.24	1.54	3.44	1.51	1.94	1.69	1.67	1.54	2.14	<u>1.30</u>	1.59
B5	4.37	1.73	4.26	1.74	3.16	2.64	1.86	1.81	2.10	1.24	<u>1.11</u>
B6	3.69	1.74	3.79	1.82	2.80	3.59	1.67	2.07	1.80	1.22	<u>1.07</u>
B7	3.07	1.52	2.94	1.38	2.19	1.77	1.49	2.07	1.80	1.18	<u>1.09</u>
A1	1.49	3.94	1.79	4.19	2.89	3.10	4.06	3.71	3.54	4.71	<u>4.85</u>
A2	1.68	3.87	1.94	3.99	2.83	3.00	4.01	3.74	3.56	4.54	<u>4.81</u>
A3	2.08	4.21	2.44	4.33	3.50	3.55	4.26	4.07	3.87	4.71	<u>4.88</u>

表 5 各評価項目間のマルチレベル相関係数（上三角は Within, 下三角は Between, 対角成分は ICG）

	B1	B2	B3	B4	B5	B6	B7	A1	A2	A3
B1	.570	.167	.507	.399	.397	.355	.368	-.879	-.864	-.917
B2	.376	.185	.169	.231	.219	.269	.202	-.532	-.514	-.520
B3	.959	.438	.594	.494	.578	.414	.392	-.928	-.913	-.949
B4	.914	.481	.987	.324	.419	.381	.376	-.926	-.911	-.943
B5	.878	.449	.936	.932	.471	.588	.427	-.988	-.992	-.986
B6	.701	.578	.745	.763	.911	.416	.414	-.920	-.938	-.906
B7	.882	.510	.931	.924	.949	.834	.238	-.970	-.955	-.965
A1	-.332	-.201	-.504	-.371	-.577	-.488	-.394	.453	.751	.696
A2	-.322	-.177	-.452	-.325	-.508	-.432	-.331	.998	.425	.614
A3	-.299	-.174	-.432	-.356	-.508	-.455	-.381	.995	.991	.395

まとめ 第1回の予備調査からは、回答者が総合的な印象としての「答えやすさ」は十分に評価できる可能性が高い一方で、個別の要素を弁別することは難しい可能性が示された。これは、ハロー効果(Phelps et al, 1986)などの知見とも整合的と言えるだろう。しかし本研究の目的としては、個人ごとに異なる「つまずき」を回答できるようにすることが望ましいため、次回では項目の提示方法を変更することとした。

2-2-2 第2回予備調査

調査協力者 クラウドソーシングサービス（クラウドワークス）を通じて、パソコンを用いて参加可能な18-69歳の日本語母語話者を募集した。1000名の募集に対して、有効回答数は991件（男性433件、女性547件、その他11件）であった。

評価項目 予備調査第1回の結果からは、一部の評価項目の表現が曖昧であった可能性を踏まえ、評価項目の修正を行った。具体的な修正箇所は、表6に示されている。

表 6 予備調査第2回で使用した評価項目の一覧（修正項目は太字で表示）

B1	この項目にはわかりにくい言葉が入っていると感じた
B2	この項目に関する自分の状態がその時々で変わりやすいために、回答を迷った
B3	この項目の意味がわかりにくかった
B4	この項目は解釈の仕方によって、回答が大きく変わりうると感じた
B5	この項目が指す状況に置かれている自分を想像することが難しかった
B6	この項目に関連することを思い返すのが難しい
B7	この項目に正直に回答するのをためらった
B8	この項目では、各選択肢がどの程度を指すのかわかりにくかった
A1	この項目は回答しやすいと感じた
A2	この項目はたいていの人が簡単に回答できるだろうと感じた
A3	この項目に正しく回答できたと確信している

回答困難項目 評価項目の修正に合わせて、一部の項目を差し替えたものを使用した（表7）。なおE1-E3は、対比のために回答困難感が生じないと考えられる項目を追加したものである。

表 7 予備調査第2回で使用した回答困難項目の一覧（新規項目は太字で表示）

1	自分には頑固な部分があると思う
2	出かけるときは身体が重い
3	いつもより、それはどうしたんだろうと思う
4	人と関わるのは嫌いではないと思わないわけではない
5	飛行機を操縦するとき、緊張より楽しみのほうが大きい
6	先週の水曜日は良い一日を過ごせたと思う
7	最近、性的欲求を感じることもある
8	読書をしたり、体を動かしたりするのが好きだ
9	貯金額は人並み以上にある
E1	1人で過ごす時間は気楽だ
E2	いま、悲しいと感じる
E3	自分には助けてくれる人がいる

調査の手続き 調査は2つのセクションによって構成された。セクション1では、表7の全項目に不適切回答検出用の項目（私はパソコンを持っている）を加えた計13項目をランダムに提示し、「全く当てはまらない」「当てはまらない」「あまり当てはまらない」「やや当てはまる」「当てはまる」「非常に当てはまる」の6件法で回答を求めた（図1）。この際、選択肢に「わからない」を追加し、質問の意味が分からない場合や回答したくないと感じた場合にこれを選択するよう教示した。セクション2に入る前には、質問項目に対する答えにくさの多様性について簡単な説明を行い、セクション2においてこれらを弁別して回答してもらうようにした。セクション2では、ハロー効果のように全体的な印象によって評価項目間が似てしまうことを抑制するため、評価項目ごとに画面を構成するように変更した。すなわち、画面上部に評価項目を表示し、ランダムな順序で提示された各回答困難項目について1（まったくそう思わない）から5（非常にそう思う）の数字から選択を求めた。なお、回答者の疲労にともなう Satisficing を抑制するため、セクション2では、全11の調査項目のうち3項目のみをランダムに提示した。

図 3 予備調査第2回のセクション1の回答画面

先ほどのようなアンケートに回答する際、迷ったり困ったりしたことはありませんか？

アンケートに関する心理学研究では、以下のような答えにくさがあることが示されています。

- ・ わかりにくい言葉(単語・慣用句など)が入っている
- ・ 自分の状態がその時々(思い浮かべるタイミング・場面)で変わりやすい
- ・ (難しい言葉が入っているわけではないが)項目の意味がわかりにくい
- ・ 解釈の仕方によって、回答が大きく変わりうる
- ・ 項目が指す状況に置かれている自分を想像することが難しい(そのような経験や馴染みがないために)
- ・ 項目に関連することを正確に思い出すのが難しい
- ・ (項目の内容的に)正直に回答するのをためらう
- ・ 各選択肢(「やや」「とても」など)がどの程度を指すのかわかりにくい

図 4 セクション2の前に表示した画面

セクション1のアンケートについて、**枠内**のように感じる程度を
1(まったくそう思わない)から5(非常にそう思う)までの
5段階評価の数字で答えてください。

すべての項目に、考え込まないで直感的に評価してください。
また、答えられない項目には3をつけてください。

この項目に正しく回答できたと確信している

上記の質問について	まったくそう思わない			非常にそう思う	
	1	2	3	4	5
いま、悲しいと感じる	☐	☐	☐	☐	☐
人と関わるのは嫌いではないと思わないわけではない	☐	☐	☐	☐	☐
読書をしたり、体を動かしたりするのが好きだ	☐	☐	☐	☐	☐
自分には助けてくれる人がいる	☐	☐	☐	☐	☐
飛行機を操縦するとき、緊張より楽しみのほうが大きい	☐	☐	☐	☐	☐
出かけるときは身体が重い	☐	☐	☐	☐	☐
1人で過ごす時間は気楽だ	☐	☐	☐	☐	☐
自分には頑善はない部分があると思う	☐	☐	☐	☐	☐
いつもより、それはどうしたんだろうと思う	☐	☐	☐	☐	☐
貯金額は人並み以上にある	☐	☐	☐	☐	☐
先週の水曜日は良い一日を過ごせたと思う	☐	☐	☐	☐	☐
最近、性的欲求を感じることもある	☐	☐	☐	☐	☐
	1	2	3	4	5

図 5 予備調査第2回のセクション2の回答画面

結果 表8に、回答困難項目に対する各評価項目の平均値を示した。「DK」行は、セクション1において「わからない」が選択された割合を示している。また枠で囲まれた箇所は、事前に想定していた「つまずき」の対応を表している。例えば回答困難項目1「自分には頑善はない部分があると思う」は、「頑善はない」という一般的には認知度の低い単語を含んでいることから、評価項目B1が高い値になることが期待されていた。

結果から、予備調査第1回と同様に、答えやすいと考えられる項目(E1-E3)が、ほとんどの評価項目で最も「答えやすい」と評価されており、また「わからない」選択率も低くなっていることが分かる。一方で、各評価項目(B1-B8)において最も「答えにくい」と判断されたのは回答困難項目3「いつもより、それはどうしたんだろうと思う」であった。したがって、一般的な回答者は、ある「つまずき」の要素が提示されて各項目がこれに該当するかを判断する際に、その特定の要素の有無を弁別的に評価するというよりも、全体的な印象に基づいて評価している可能性を示唆している。

表8 予備調査第2回の回答困難項目に対する各評価項目の平均値および「わからない」選択率

	1	2	3	4	5	6	7	8	9	E1	E2	E3
B1	4.24	1.40	3.50	3.71	1.77	1.63	1.33	1.52	1.37	<u>1.29</u>	1.29	1.30
B2	2.63	2.71	3.20	2.95	2.15	2.59	2.47	2.15	<u>1.72</u>	1.88	2.30	1.30
B3	3.89	1.59	4.35	4.21	2.50	1.94	1.49	1.98	1.41	<u>1.33</u>	1.45	1.39
B4	3.37	2.39	4.09	3.88	2.79	2.54	2.06	2.53	2.31	<u>1.94</u>	2.05	2.10
B5	3.58	1.75	4.03	3.50	3.48	2.56	1.87	1.94	1.64	<u>1.51</u>	1.55	1.77
B6	2.93	1.81	3.39	2.98	2.94	3.28	1.85	1.77	1.67	<u>1.64</u>	1.55	1.78
B7	3.01	1.63	3.06	3.01	2.30	2.01	2.68	1.73	2.11	<u>1.45</u>	1.54	1.73
B8	3.41	2.18	4.03	3.59	2.79	2.48	2.24	2.21	2.31	<u>1.78</u>	1.93	2.02
A1	2.17	3.86	1.57	1.79	2.55	3.05	3.55	3.73	3.76	<u>4.63</u>	4.06	4.19
A2	2.02	3.74	1.73	1.93	2.55	3.02	3.34	3.82	3.53	<u>4.48</u>	3.93	3.96
A3	2.78	4.05	2.38	2.71	3.22	3.58	4.05	3.97	4.06	<u>4.67</u>	4.20	4.32
DK	34.6	0.6	45.2	11.2	27.3	12.1	2.2	1.6	4.1	<u>0.1</u>	0.3	0.5

ウェブ調査における Satisficing の影響を検証するため、感度分析として合計回答時間が短い k% の回答者を除外して評価項目 B1-B8 間の相関係数の算出を行い、中央値および 25-75% 区間を求めた。図 6 に、この手続きを k=1 から k=100 まで繰り返した結果をまとめている。図から、回答時間が短い、すなわち相対的にモチベーションが低いと考えられる回答者を優先的に除外したとしても評価項目間の相関係数の分布がほとんど変わらないことが分かる。したがって、評価項目間の相関の高さは、Satisficing によるものというよりも、回答者が評価項目間を十分に弁別できていない可能性が示唆されたと言える。

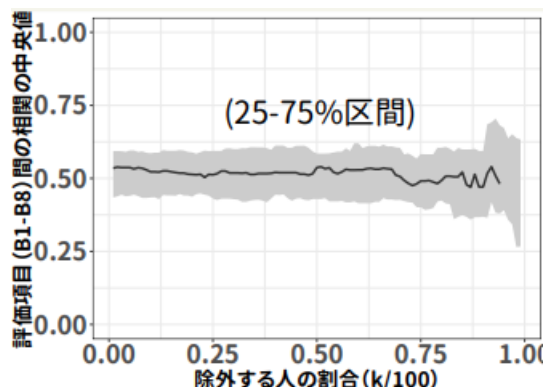


図 6 回答時間が短い回答者を除外して求めた評価項目 B1-B8 間の相関係数の分布

まとめ 表 8 の結果を縦方向に見ると、例えば回答困難項目 3 に対しては評価項目 B3 が最も高い値となっている。このように見ると、回答困難項目の多くにおいて、設計時に想定していた要素が最も高い値になっていることが分かった。あくまでも集計された結果の上での話だが、この結果は、複数の「つまずき」要因に対する相対的な評価はある程度できている可能性を示唆するものと言える。これを踏まえて次回調査では、再び評価項目の提示方法を変更することとした。

2-3 本調査

本調査では、回答の安定性を評価する目的で、2 回の調査を行った。

2-3-1 第 1 回本調査

調査協力者 クラウドソーシングサービス（クラウドワークス）を通じて、パソコンを用いて参加可能な 18-69 歳の日本語母語話者を募集した。1000 名の募集に対して、有効回答数は 997 件であった。

評価項目および回答困難項目 予備調査第 2 回と同一のものを使用した。ただし評価項目については、後述するように提示の方法を変更した。

調査の手続き セクション 1 およびセクション 2 の直前に示す答えにくさの教示画面は、予備調査第 2 回と同一であった。セクション 2 では、画面上部に回答困難項目を提示し、評価項目 B1-B8（表 6）を選択肢として表示し、この中から最大 3 つを選択させた。また、選択肢には「いずれも当てはまらない」を追加し、この選択肢は他の選択肢と同時に選択できないように設定した。図 7 に、セクション 2 の画面を例示する。

いつもより、それはどうしたんだろうと思う

上記の質問項目について、以下の中から当てはまると思ったものを、**最大3つまで** 選んでください。

☐ この項目にはわかりにくい言葉が入っていると感じた

☐ この項目の意味がわかりにくかった

☐ この項目に関する自分の状態がその時々で変わりやすいために、回答を迷った

☐ この項目は解釈の仕方によって、回答が大きく変わらうと感じた

☐ この項目が指す状況に置かれている自分を想像することが難しかった

☐ この項目に関連することを正確に思い出すのが難しい

☐ この項目では、各選択肢がどの程度を指すのかわかりにくかった

☐ この項目に正直に回答するのをためらった

☐ いずれも当てはまらない

図 7 本調査のセクション 2 の回答画面

結果 表9に、各回答困難項目に対して、それぞれの選択肢（評価項目）が選ばれた割合を示した。なお「NA」行は、「いずれも当てはまらない」を選択した割合を示しており、太字は各回答困難項目に対して「いずれも当てはまらない（NA）」以外で最も多く選ばれた選択肢を示している。表からは、回答が容易と考えられる項目 E1-E3 において、回答困難項目と比べて「いずれも当てはまらない」の選択率が高いことがうかがえる。この結果は予備調査第2回と整合的であったといえる。また、多くの回答困難項目において、事前に想定していた（枠で囲われている）「つまずき」要素の選択率が最も高くなっていた。あくまでも平均的な傾向ではあるが、この結果は回答者が「つまずき」要素を相対的には評価できる可能性が高いことを示している。なお、回答困難項目7「最近、性的欲求を感じることもある」については、評価項目 B7「この項目に正直に回答するのをためらった」と対応していると予想していたが、本調査が個人を特定されるおそれのないウェブ調査であったことから、ためらいを感じる人が少なかったと考えられる。同時に、この項目は頻度を問わずわざわざ性的欲求を感じる必要がある場合には「当てはまる」と判断できる、すなわち二値的な項目であったため、「あまり」や「非常に」といった程度の副詞の適用範囲が分かりづらくなっていた可能性が、評価項目 B8 の選択率の高さに表れている。

全体を通して見ると、項目 E1-E3 についても「いずれも当てはまらない」の選択率は最大 89.0%であり、どのような項目でも一定の割合で何らかの回答困難感を抱く回答者が存在することがうかがえる。ただしこの結果は、一定の割合で適当に回答しているものが混ざっているだけでなく、本調査で使用したプラットフォーム特有のバイアスが存在する可能性がある。クラウドワークスでは、謝礼の支払いの前に、提出された回答内容をレビューして承認・非承認を選択するプロセスが含まれているため、「いずれも当てはまらない」を選択することに対する心理的なブレーキがかかっていた可能性が考えられる。そして E1-E3 における選択率を見ると、「いずれも当てはまらない」以外を選択する場合には、評価項目 B2「この項目に関する自分の状態がその時々で変わりやすいために、回答を迷った」の選択率が高くなりやすい可能性が示唆された。実際に、B2「いま、悲しいと感じる」や B3「自分には助けしてくれる人がいる」といった項目は、「いま」の捉え方やライフステージの変化等に応じて回答が変わりうるとも考えられる項目である。これらの項目について、筆者らは「つまずき」は発生しないと予想していたものの、今回の結果は、心理尺度を実施する際に、時間のフレームを明確に指示することの重要性を示すものと言えるだろう。

表 9 各回答困難項目に対する選択率

	1	2	3	4	5	6	7	8	9	E1	E2	E3
B1	0.818	0.005	0.181	0.329	0.010	0.002	0.004	0.018	0.009	0.002	0.001	0.003
B2	0.038	0.323	0.080	0.078	0.033	0.051	0.105	0.209	0.038	0.086	0.106	0.105
B3	0.432	0.040	0.892	0.856	0.080	0.010	0.012	0.077	0.004	0.002	0.010	0.003
B4	0.037	0.091	0.236	0.121	0.065	0.047	0.046	0.164	0.160	0.031	0.056	0.088
B5	0.045	0.032	0.172	0.038	0.714	0.132	0.031	0.040	0.022	0.013	0.027	0.031
B6	0.019	0.026	0.054	0.023	0.053	0.594	0.015	0.012	0.025	0.003	0.009	0.022
B7	0.039	0.081	0.109	0.074	0.022	0.040	0.073	0.120	0.207	0.022	0.061	0.075
B8	0.028	0.010	0.045	0.035	0.037	0.008	0.401	0.012	0.245	0.003	0.031	0.023
NA	0.061	0.561	0.019	0.030	0.197	0.318	0.482	0.555	0.488	0.890	0.790	0.769

2-3-2 第2回本調査

調査協力者 第1回本調査に参加し有効な回答を提出した回答者に直接参加を依頼した。997名のうち852名から有効回答を得た。

調査の手続き セクション2のみを、第1回本調査と同じ手順で実施した。ただし回答困難項目の提示順は第1回とは無関係に再度ランダムに決定した。

分析方法 この調査の目的は、第1回との回答の一貫性（再検査信頼性）を評価することにある。ただし評価項目の回答方法が、最大数に制限のある複数選択形式であることを踏まえて、Dice 係数および RV 係数 (Robert & Escoufier, 1976) を用いた評価を行うこととした。Dice 係数は、2つの集合の要素の一致度の係

数である。今回の場合では第1回と第2回のそれぞれにおいて、どの回答困難項目に対してどの評価項目を選んだか、を集合とみなす。このとき、Dice 係数は、2つの集合の共通要素の総数の2倍を、各集合の要素数の合計で割った値である。本研究ではこれを、個人ごとおよび項目ごとに計算し、平均値・中央値および分布を調べる。RV 係数は、ピアソンの積率相関係数の二乗を多変量に一般化した指標であり、行列間の類似度を示す指標である。これによって、項目ごとに選択しやすい評価項目の傾向の一致度（選択率の安定性）を確認できる。

結果 （報告書提出時点で論文執筆中のため省略）

2-4 実際の心理検査を用いた検証

調査項目 馬場他（2025）のアンケートで尋ねた「臨床で使用している心理尺度」において回答数の多かったものの中から、AQ 日本語版（若林他，2007）、LSAS-J（朝倉他，2002）、YGPI（日本心理テスト研究所，2025）の3つを使用した。

調査の手続き 基本的な流れは本調査第1回（2-3-1 節）と同一であった。なおセクション1においては、もとの心理検査と同じ教示および選択肢に加えて「わからない」に相当する選択肢を用意した。また強制速度法が採用されている YGPI に関しては、一定の時間が経過すると「できるだけ素早く回答してください」というメッセージを表示するようにした。一人が回答する項目数は、回答の負荷を軽減するため、各検査の中からランダムに選ばれた15項目とした。

調査協力者 クラウドソーシングサービス（クラウドワークス）を通じて、パソコンを用いて参加可能な18-69歳の日本語母語話者を募集した。後述するように3つの心理検査（AQ 日本語版、LSAS-J^{注1}、YG 性格検査^{注2}）それぞれについて募集した。サンプルサイズについては、第1回本調査の結果を参考に、一般的な項目でも生じうる選択率10%を基準として、これよりも有意に高い選択率を示す項目を検出することを目指し、次の手順で決定した。はじめに、選択率の真値が20%である項目において検出力が0.9に達するために必要な各項目の回答数を、 $\text{pwr.2p.test}(h = \text{ES.h}(p1 = 0.1, p2 = 0.2), \text{power} = 0.9)$ によって求めたところ、261であった。各項目が回答者ごとに選択されるかは、確率15/I（Iは各検査の項目数：AQは50、LSAS-Jは24、YGPIは120）のベルヌーイ試行であることから、二項分布 $B(n, 15/I)$ において、上位95パーセンタイル点の値が261に近くなるnの値（AQは946、LSAS-Jは444、YGPIは2296）を、各検査のサンプルサイズの目安とした。最終的に、述べて有効回答数は3674件（AQ936名、LSAS-J442名、YGPI2296名、複数の検査に回答した参加者を含む）であった。

分析方法 はじめに、各評価項目に対して高い選択率が見られた検査項目をいくつか取り上げ、評価項目が確かに検査項目に含まれる「つまずき」の要素を反映していることを定性的に確認した。続いて、調査項目×回答者を分析単位に、評価項目の選択有無を特徴量とした潜在クラス分析によって、主観的に報告される「つまずき」のパターンを調べ、どの「つまずき」が共起しやすいかを確認した。

結果・まとめ （報告書提出時点で論文執筆中のため省略）

3 研究2：発話思考を用いた回答プロセスの探索的検討

3-1 研究の目的

本研究では、回答者が答えにくい項目に回答する際にどのような思考プロセスが現れるかを、リアルタイムかつ対話的な手法を用いて明らかにすることを目的とする。これにより、回答者が感じているつまずきをより詳細に理解し、今後の心理尺度の開発や回答者支援の一助とすることを目指す。同時に、またマウスカーソル軌跡から抽出できる特徴量の参考になる知見を抽出することが期待された。

3-2 方法

実験協力者 日本語話者を参加条件とし、大学生および大学院生23名を対象として対面で本実験を行った。このうち、実施者側の不備によって正しく実験を完了できなかった1名を除いた、22名分のデータを分析に使用した。

実験の手続き 実験は2つのセクションで行われた。セクション1では、モニターに表示される調査項目に対してマウスを用いて回答しながら、発話思考法（Ericsson & Simon, 1980; 海保・原田, 1993）を行うよう求めた。表7に示された回答困難項目および回答容易な項目から、対面では侵襲性が高いと考えられる項目7を除いた全11項目を提示し、発話思考を求めた。セクション2では、セクション1で表示された調査項

目を再度提示し、1項目ずつ回答プロセスについて振り返りながら詳細を確認する認知インタビューを行った。特定の情報や認知プロセスについて掘り下げる質問（verbal probings）を事前に用意し、主にステップ1のなかで見られた特徴的な発話をはじめ、回答時の様子の中で筆者らが気になった点について質問を投げかけた。以上の実験過程はすべて小型マイクにて録音し、同時に参加者の後ろから画面や手元を録画した。

分析方法 セクション1の逐語録を分析対象とし、適宜録音・録画データやステップ2の認知インタビューの内容を参照した。データは、プロトコル分析の手法を援用したミックスメソッドにより分析された。具体的には、まず、参加者と調査項目の組み合わせごとの逐語録を単位として、思考プロセスの遷移に注目して切片化した。その後、各切片にラベルおよび4段階モデルに基づくコードを付与した。得られたラベルについては、逐語録ごとにラベルの内容と並びを検討し、その系列を短く表現したカテゴリにまとめた。またコードについては、系列間の距離を編集距離に基づいて定義し、階層的クラスター分析によってグルーピングを行い、大まかにどのような思考プロセスが見られるかを調べた。

3-3 結果

（報告書提出時点で論文執筆中のため省略）

4 研究3：マウスカーソル軌跡を用いた「答えにくさ」の検出

4-1 研究の目的

本研究では、回答者が調査項目に対して抱く困難感を検出するため、回答内容以外のパラデータとして回答時間およびマウスカーソルの軌跡のデータに着目し、種々の基礎的な特徴量と、主観的に知覚された回答困難感の関係について考える。

4-2 方法

データ 研究1の予備調査1（2-2-1節）で収集したものの一部（122名分）を使用した。この調査では、ランダムに選ばれた一部の回答者に対しては、回答困難項目にぼかしをかけて、マウスカーソルを近づけた場合にのみ文字が表示されるように設定しており、この際のマウスカーソル軌跡のログをパラデータとして利用した。

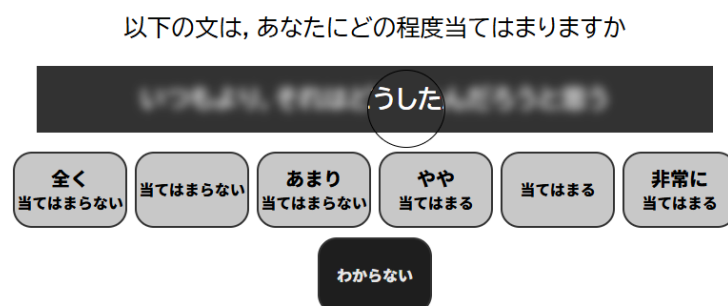


図8 セクション1の回答画面（回答困難項目にぼかしがかかっている様子）

変数 研究1の予備調査2の段階では、答えにくさの多様な要因を区別した回答が得られていないことから、本研究では、評価項目のうち総合的な答えやすさを尋ねた表2のA1-A3の3項目に対して、1因子モデルによって算出した因子得点を「答えやすさ得点」として使用した。

使用する特徴量は、複数の文献(Kieslich et al., 2019; Pino et al., 2025)を参考に、表10に示された基礎的な特徴量を算出した。

なお、マウスカーソル軌跡に関する特徴量は、まずデータを「項目が表示されている領域（図8の濃いグレー）内にいる期間」と「項目の領域から一度でも外に出た後」の2つに分割したあとで、それぞれのデータに対して個別に算出した。加えて、回数以外の特徴量については、個人内で標準化した値を別途用意した。以上の手続きにより、最終的に計48個の特徴量を使用した。

表 10 使用した特徴量の一覧

回答内容	実際の回答内容（「全く当てはまらない」から「非常に当てはまる」を1-6に数値化） 項目 ID（ダミー変数として投入）
回答時間	項目が表示されてからの合計時間 カーソルが項目文の領域上にあった時間 カーソルが項目文の領域を離れてから回答を決定するまでの時間 カーソルが実際に選んだ選択肢の領域上にあった時間 カーソルが項目文の領域を離れたあとに、再び項目文の領域に戻った時間
カーソル軌跡	カーソルの総移動距離 最大速度 最大加速度 停止していた時間 AUC（始点と終点を結んだ直線とカーソル軌跡が囲む面積） MAD（始点と終点を結んだ直線との最大距離） MD_above（始点と終点を結んだ直線より上での MAD） MD_below（始点と終点を結んだ直線より下での MAD） X 座標についてカーソルが反転した回数 カーソルが実際に選んだ選択肢の領域上に入り直した回数

分析方法 記述統計的な探索として、各特徴量と「答えやすさ得点」の相関係数を求めた。続いて、線形モデル（LASSO）および非線形モデル（XGBoost）による予測を試みた。まず、データを参加者のレベルで訓練データとテストデータに 8 : 2（97 人分 : 25 人分）の割合で分割した。その後、(1) 回答内容のみを使用したモデル、(2) 項目が表示されてからの合計時間（RT）を追加したモデル、(3) すべての特徴量を投入したモデル、の 3 種類について、それぞれ学習および予測を行った。なお、回答時間に関する特徴量のうち、モデル(2)に含めた RT 以外は、いずれもカーソル軌跡の情報があって初めて区別されるものであることから、モデル(2)はカーソル軌跡を使用せず回答時間のみを追加したモデルと位置づけられる。

4-3 結果

相関係数の絶対値が高かった特徴量は順に、「カーソルが項目文の領域上にある時間（標準化後）：-0.569」「RT：-0.507」「カーソルが項目文の領域上にある時間（標準化前）：-0.483」「カーソルの項目文の領域上の総移動距離：-0.451」「実際の回答内容：0.442」と続いた。答えやすい項目ほど回答時間が短くなるというのは当然の結果であるが、これがカーソルの移動距離などとも関連していることがうかがえる。

続いて LASSO および XGBoost による予測の精度の指標として、RMSE および真値との相関係数を算出した結果を表 11 に示した。いずれにおいても、マウスカーソル軌跡を追加したモデル(3)の精度が最も低くなった。

表 11 LASSO および XGBoost による「答えやすさ得点」の予測精度

モデル	LASSO		XGBoost	
	RMSE	相関係数	RMSE	相関係数
1	0.671	0.738	0.684	0.723
2	0.658	0.751	0.687	0.727
3	0.679	0.731	0.738	0.669

4-4 まとめ

表 11 の結果からは、本研究の範囲においては、マウスカーソル軌跡の利用は回答困難感の予測には利用できないと言わざるを得ない。ただし、本研究で使用したデータにはいくつかの問題点がある。はじめに、使用した特徴量の数に対してサンプルサイズが小さい（122 名分）ため、過学習が起きている可能性が高く、より大きなサンプルサイズでの再検討が必要である。加えて、認知心理学におけるマウスカーソルを利用した実験(e.g., Di Palma et al., 2022; Yamauchi & Xiao, 2018)と比べると、本調査では画面に表示される項目文と選択肢の距離が近かったために、軌跡の直線性（AUC の小ささなど）が正確に計測できていない可能性がある。また、マウスカーソル軌跡からの特徴量の抽出に関しても、より高度な手法を適用することで、より効果の高い特徴量が得られる可能性も考えられる。

5 まとめ

本研究課題では、心理尺度に回答する際に生じる「つまずき」あるいは「回答困難感」を検出する手法開発に向けて、いくつかの基礎的な研究を行った。はじめに教師データ生成のために、回答者が主観的に回答困難感を報告できる形式を開発する必要があったが、この点に関しては一定の成果を上げたと言える。一方で、実際に回答困難感を検出するための手法の開発についてはまだ多くの検討課題が残されている。今後は、精緻な予測を可能にするためのパラデータの利用方法の研究が望まれる。こうした基礎研究は、臨床場面や心理尺度の質向上に有用な知見となるだろう。

【参考文献】

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2011). *Standards for educational and psychological testing*. American Educational Research Association.
- 朝倉 聡・井上 誠士郎・佐々木 史・佐々木 幸哉・北川 信樹・井上 猛・傳田 健三・伊藤 ますみ・松原 良次・小山 司 (2002). Liebowitz Social Anxiety Scale (LSAS) 日本語版の信頼性および妥当性の検討 精神医学, 44(10), 1077-1084. <https://doi.org/10.11477/mf.1405902721>
- 馬場 絢子・分寺 杏介・福島 綾 (2025). 心理尺度回答におけるクライアントの「つまずき」と心理職のサポート 心理学研究, 96(1), 48-54. <https://doi.org/10.4992/jjpsy.96.23321>
- Benton, A. L., & Kornhauser, S. I. (1948). A study of "score faking" on a medical interest test. *Academic Medicine*, 23(1), 57-60. <https://doi.org/10.1097/00001888-194801000-00007>
- Blazek, D. R., & Siegel, J. T. (2023). Preventing satisficing: A narrative review. *International Journal of Social Research Methodology*, 27, 635-648. <https://doi.org/10.1080/13645579.2023.2239086>
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, 12, 389-437. <https://doi.org/10.2307/270748>
- Conners, C. K., Erhardt, D., & Sparrow, E. P. (1999). *Conners' adult ADHD rating scales (CAARS): Technical manual*. Multi-Health Systems.
- Di Palma, M., Carioti, D., Arcangeli, E., Rosazza, C., Ambrogini, P., Cuppini, R., Minelli, A., & Berlingeri, M. (2022). The biased hand: Mouse-tracking metrics to examine the conflict processing in a race-implicit association test. *PLOS ONE*, 17(7), e0271748. <https://doi.org/10.1371/journal.pone.0271748>
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251. <https://doi.org/10.1037/0033-295X.87.3.215>
- Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, 27(5), 315-323. <https://doi.org/10.1177/0963721417746793>
- Goldammer, P., Stöckli, P. L., Escher, Y. A., Annen, H., & Jonas, K. (2023). On the utility of indirect methods for detecting faking. *Educational and Psychological Measurement*, 84(5), 841-868. <https://doi.org/10.1177/00131644231209520>
- Huvila, I., Andersson, L., & Sköld, O. (Eds.). (2024). *Perspectives on paradata: Research and practice of documenting process knowledge (Vol. 13)*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-53946-6>
- Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. National Academies Press.
- 海保 博之・原田 悦子 (1993). プロトコル分析入門：発話データから何を讀むか 新曜社
- Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). *Mouse-tracking: A practical guide to implementation and analysis*. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A Handbook of Process Tracing Methods* (pp. 111-130). Routledge. <https://osf.io/zuvqa>
- Kreuter, F. (Ed.). (2013). *Improving surveys with paradata: Analytic use of process information*. John Wiley & Sons.

- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213–236. <https://doi.org/10.1002/acp.2350050305>
- Likert, R. (1932). *A technique for the measurement of attitudes* (R. S. Woodworth, Ed.). New York.
- Mazza, C., Monaro, M., Burla, F., Colasanti, M., Orrù, G., Ferracuti, S., & Roma, P. (2020). Use of mouse-tracking software to detect faking-good behavior on personality questionnaires: An explorative study. *Scientific Reports*, 10(1), Article 1. <https://doi.org/10.1038/s41598-020-61636-5>
- Meadows, K. (2021). Cognitive interviewing methodologies. *Clinical Nursing Research*, 30(4), 375–379. <https://doi.org/10.1177/10547738211014099>
- Mokkink, L. B., Elsmans, E. B. M., & Terwee, C. B. (2024). COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Quality of Life Research*, 33(11), 2929–2939. <https://doi.org/10.1007/s11136-024-03761-6>
- 日本心理テスト研究所 (2025) . YGPI オペレーションガイド
- Ones, D. S., & Viswesvaran, C. (1998). The effect of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11, 245–269. <https://doi.org/10.1080/08959285.1998.9668033>
- Phelps, L., Schmitz, C. D., & Boatright, B. (1986). The effects of halo and leniency on cooperating teacher reports using likert-type rating scales. *The Journal of Educational Research*, 79(3), 151–154. <https://doi.org/10.1080/00220671.1986.10885668>
- Pino, A., Papatheodorou, N., Kouroupetroglou, G., Giannopoulos, P.-A., Makris, G., & Papageorgiou, C. (2025). Hand dexterity evaluation grounded on cursor trajectory investigation in children with ADHD using a mouse and a joystick. *Technologies*, 13(3), Article 3. <https://doi.org/10.3390/technologies13030099>
- Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Applied Statistics*, 25(3), 257. <https://doi.org/10.2307/2347233>
- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21(2), 277–287. <https://doi.org/10.1002/acp.1340>
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). National Academies Press.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge University Press.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. SAGE.
- Willis, G. B., & Lessler, J. T. (1999). *Question appraisal system—QAS-99*. http://www.websm.org/uploadi/editor/1364216022Willis_Lessler_1999_QAS_99.pdf
- 山内 繁 (2015). エンジニアのための 人を対象とする研究計画入門 丸善出版
- Yamauchi, T., & Xiao, K. (2018). Reading emotion from mouse cursor motions: Affective computing approach. *Cognitive Science*, 42(3), 771–819. <https://doi.org/10.1111/cogs.12557>
- 若林 明雄・内山 登起夫・東條 吉邦・吉田 友子・黒田 美保・バロン-コーエン サイモン・ウィールライト サリー (2007). 自閉症スペクトラム指数 (AQ) 児童用・日本語版の標準化——高機能自閉症・アスペルガー障害児と定型発達児による検討 心理学研究, 77(6), 534–540. <https://doi.org/10.4992/jjpsy.77.534>

(注 1) AQ 日本語版および LSAS-J の利用にあたっては、株式会社三京房の許諾を受けました。

(注 2) YGPI の利用にあたっては、日本心理テスト研究所株式会社の許諾を受けました。

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
心理尺度に対する「答えにくい」感覚を捉える：回答者個人の主観を問う定量的評価の試み	日本心理学会第 88 回大会	2024 年 9 月
質問項目に対する回答困難感を反映するパラデータ指標の探索的研究	日本行動計量学会第 53 回大会	2025 年 9 月
心理尺度に対する「答えにくさ」は弁別できるか	日本心理学会第 89 回大会	2025 年 9 月
質問紙法の心理検査における「答えにくさ」の主観的弁別：回答者の自己報告に基づく困難感の多次元構造的検討	日本心理学会第 90 回大会	2026 年 9 月（予定）
リッカート型項目への回答はどこで止まり、逸れるのか：回答時の体験およびつまずきのプロセスに着目して	日本心理学会第 90 回大会	2026 年 9 月（予定）
受験者体験から問い直すテストのあり方(仮題)	日本テスト学会第 24 回大会 (大会実行委員会企画シンポジウム)	2026 年 9 月（予定）