

回答者はリッカート尺度に本当に 正しく回答できているのか？

—心理尺度回答における「つまずき」から回答者視点のテストを考える—

分寺 杏介



bunji@bear.kobe-u.ac.jp

神戸大学 経営学研究科

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「受験者視点から問い直すテストのあり方」



※本スライドは、クリエイティブ・コモンズ 表示-非営利 4.0 国際 ライセンス (CC BY-NC 4.0)に従って利用が可能です。

Outline

1 問題意識

2 研究紹介

3 将来の展望

■ 心理学研究の主流である割には回答の質は高くない気がする

Morren & Paas (2020): 短いIMC (DQSなど) でも2-3割引|っかかることもある

三浦・小林 (2016): 場合によってはIMC違反率が8割を超えることも

Chmielewski & Kucker (2020): MTurkの質が2018-19年くらいから急激に悪化

【実体験】某調査会社でおこなったとき

Q この調査について、ご意見や気付いたことなどがあればご自由に記入してください（自由回答）

えgrkpねがいgjりあvもにえあwgfういえsrfんf

実体験をもとに再現したものです。
実際の回答とは異なります。

夢ならばどれほどよかったでしょう 未だにあなたのことを夢にみる
忘れた物を取りに帰るように 古びた思い出の埃を払う

なんだかとても暑いなあと思いました。

もちろん複合的であることは承知の上で……

Fowler et al. (2022)

MTurkを対象に、フラストレーションの理由を自由記述で回答

▶ 調査デザインや質問項目デザインの不備が半数以上

山田・五十嵐(2025; 2026, 日本心理学会大会)

同様の調査を日本(Yahooクラウドソーシング・ランサーズ)にて実施

Howe could academic surveys on MTurk be designed so that frustration for Turkers is reduced/minimized?



調査参加経験の浅い人では
調査・項目の設計に関する意見が多い

ベテランになるとお金やシステムの話が増える

Q. 以下の食べ物について抱くイメージをすべて選択してください

	上品な	信頼できる	安心できる	本格的な	おいしい	あつさりした	コクのある	食べやすい	健康的な	安全な	親しみやすい	ワクワクする	手頃な	洗練された	
カレー	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	...
ラーメン	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
チャーハン	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
ハンバーグ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
寿司	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
ステーキ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
からあげ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
焼き肉	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	
⋮															

知らない人と話すのを躊躇することがある



まったく
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



とても
当てはまる

「あまり」ってどれくらいだ……？



「躊躇」ってなんて読むんだ……？

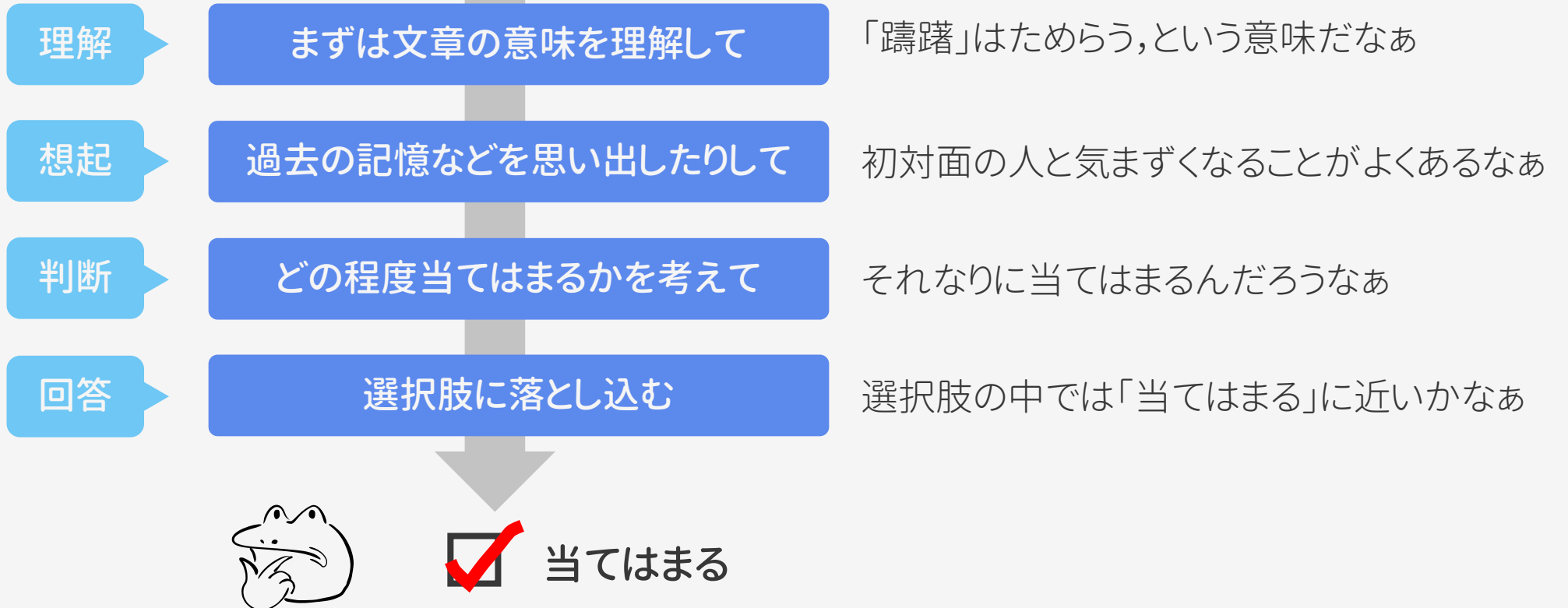
どんな場面を想定してるんだ……？

▲ Powerpointに入っていたフリー素材のひと

■ 様々な認知過程を経て回答を決定している (e.g., Tourangeau, 2018)

知らない人と話すのを躊躇することがある

Tourangeau et al.(2000)
のフレームワーク



■ 認知処理の失敗 (Cannell, 1981)

知らない人と話すのを躊躇することがある

謝礼安すぎ!

「躊躇」ってどういう意味だ?

理解

まずは文章の意味を理解して

そんなこといきなり聞かれても分からないよ

想起

過去の記憶などを思い出したりして

時と場合によるからなあ

判断

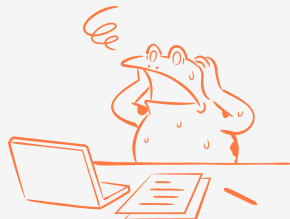
どの程度当てはまるかを考えて

「やや」と「すこし」はどう違うんだ?

回答

選択肢に落とし込む

回答スタイルに基づく回答
周囲の情報からの推測
サティスファイシング



当てはまる

- 「回答者は、心理尺度を**正しく理解して**(まじめに)回答している」

あるいは、正しく理解して回答したものとして扱っている

- 仮定の逸脱は様々な問題に繋がらうる

【研究】データの質が低下するかもしれない

【臨床】回答をもとに判断してしまうと誤った診断を下してしまうかもしれない



- 様々なレベルでの回答困難感が報告されている (馬場他, 2025)

(臨床場面での心理尺度の利用に関して)

「やや」のような表現の程度に迷って全然答えを決められない

質問の向き(「当てはまる」)を反対に解釈して間違えてしまう

etc.

- 「答えにくい項目」は避けるべきであることは前提として
COSMINでは理解可能性の評価に認知インタビューを要求している
▶ 一般論として「わかりにくい項目」は避けるべき
- ただ同じ項目でも人によってつまずきポイントは異なるはずで

知らない人と話すのを躊躇することがある

「躊躇」ってどう読むんだ？



仕事とプライベートで変わる
場合はどう答えたら良いの？

イージーな質問だね！

回答者×項目ごとに異なる「答えにくさ」を評価できたほうが良い

Outline

1 問題意識

2 研究紹介

3 将来の展望

※本研究は、電気通信普及財団研究調査助成を受けて実施しました。

2つの研究をご紹介します

1 定量的な評価方法の開発

(分寺他, 2026, 日本心理学会大会)

各項目について、どのような「答えにくさ」を感じたかを直接回答してもらう

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、回答が大きく変わりうると感じた
- ☐ 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

2 発話思考法による回答プロセスの探索

(福島他, 2026, 論文投稿中; 日本心理学会大会)

答えにくい項目に発話しながら回答してもらう



事後インタビューでさらに深堀り



- ウェブ調査などでも利用できる方法として
- 【大問題】本当に「答えにくさ」の細かい違いを自覚的に弁別できるのか？

パイロット的な検討：どうやら**失敗** (分寺他, 2024; 2025, 日本心理学会大会)

読書をしたり, 体を動かしたりするのが好きだ

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段階評価の数字で答えてください。

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
この項目に関連することを思い返すのが大変だった	☐	☐	☐	☐	☐
この項目の意味がわかりにくかった	☐	☐	☐	☐	☐
この項目に正直に回答するのをためらった	☐	☐	☐	☐	☐

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

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
自分には助けしてくれる人がある	☐	☐	☐	☐	☐
いつもより, それはどうしたんだろうと思う	☐	☐	☐	☐	☐
自分には頑実はない部分があると思う	☐	☐	☐	☐	☐
いま, 悲しいと感じる	☐	☐	☐	☐	☐
先週の水曜日は良い一日を過ごせたと思う	☐	☐	☐	☐	☐

どちらの形式でも項目間の相関がかなり高くなっていた

個別具体の要素の評価が
全体的な印象に引っ張られてしまった？

そもそもリッカート型項目の難しさを
リッカート型で尋ねても良いのか？
(オープンエンドな問い)

質問に回答したあとだと
「いちおう回答できた」ことが影響する？

「答えにくさ」にも様々な要素があることを
回答者は認識していないのでは？

すべての要素の評価が
一律で高くないような
回答形式に変更しよう

「回答できない」ということも
きちんと許容しておく必要がある

「答えにくさ」を回答してもらう前に
簡単に説明をしたほうが良さそう

質問に回答してもらう

以下の文は、あなたにどの程度当てはまりますか

人と関わるのは嫌いではないと思わないわけではない

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

わからない

すこし説明する

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
- (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
- 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
- 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
- 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
- (項目の内容的に)正直に回答するのをためらう

■ 上限付きのCATA (Check-All-That-Apply)

人と関わるのは嫌いではないと思わないわけではない

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、回答が大きく変わりうると感じた
- ☐ 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

1つの項目が4つも5つも同時に含むことはさすがにないだろうということで

最大**3**つまで

◀ 排他選択肢として用意

■ 対応をもたせるように作成してみました

回答困難項目(+ 回答しやすい項目 : E)

- 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

評価項目

- 1 この項目にはわかりにくい言葉が入っていると感じた
- 2 この項目の意味がわかりにくかった
- 3 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
- 4 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
- 5 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
- 6 この項目に関連することを正確に思い出すのが難しい
- 7 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
- 8 この項目に正直に回答するのをためらった

まとめ

集団レベルでは概ね弁別できていそう

答えやすい項目では確かに低くなっている

要素3(変わりやすい)は一定どこでも出現する

基本的には「分からない」選択率と相関するが、一部では違う傾向も見られる

回答困難項目(+ 回答しやすい項目 : E)	
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

「答えにくさ」評価要素	1	2	3	4	5	6	7	8	9	E1	E2	E3
1 この項目にはわかりにくい言葉が入っていると感じた	0.82	0.01	0.18	0.33	0.01	0.00	0.00	0.02	0.01	0.00	0.00	0.00
2 この項目の意味がわかりにくかった	0.43	0.04	0.89	0.86	0.08	0.01	0.01	0.08	0.00	0.00	0.01	0.00
3 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	0.04	0.32	0.08	0.08	0.03	0.05	0.11	0.21	0.04	0.09	0.11	0.11
4 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	0.04	0.09	0.24	0.12	0.07	0.05	0.05	0.16	0.16	0.03	0.06	0.09
5 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	0.05	0.03	0.17	0.04	0.71	0.13	0.03	0.04	0.02	0.01	0.03	0.03
6 この項目に関連することを正確に思い出すのが難しい	0.02	0.03	0.05	0.02	0.05	0.59	0.02	0.01	0.03	0.00	0.01	0.02
7 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	0.04	0.08	0.11	0.07	0.02	0.04	0.07	0.12	0.21	0.02	0.06	0.08
8 この項目に正直に回答するのをためらった	0.03	0.01	0.05	0.04	0.04	0.01	0.40	0.01	0.24	0.00	0.03	0.02
いずれも当てはまらない	0.06	0.56	0.02	0.03	0.20	0.32	0.48	0.55	0.49	0.89	0.79	0.77
「わからない」選択率(%)	32.2	0.4	42.8	9.2	29.3	10.9	1.1	1.2	4.0	0.1	0.4	0.6

そもそも実務で使われている尺度にそんなに「答えにくさ」はあるのか？

■ 対象尺度：AQ日本語版*，LSAS-J*，YGPI性格検査**

全50項目

全24項目

全120項目

* 尺度の使用に際して、株式会社三京房の許諾を受けました。

** 尺度の使用に際して、日本心理テストセンター株式会社の許諾を受けました。

■ 手続き：先程の調査と同じ（回答→「答えにくさ」説明→「答えにくさ」評価）

負担軽減のため、一人あたり特定の尺度からランダムに15問回答してもらった

サンプルサイズ：述べ3674名

10%程度まではノイズで変動する

▶ 10%と20%の差の検出力が0.85/0.9になるサンプルサイズ=223/261

▶ これに95%の確率で到達するよう尺度ごとに設定（予算も考慮して）

■ 分析：回答者×項目を単位として

潜在クラス分析，ネットワーク分析をやってみました

【ちゅうい】具体的な尺度・項目の問題を名指しで指摘する意図はないので、各項目における「答えにくさ」要素選択率は割愛しています。気になる方は個人的に聞いてください。

「理解」の難しさ ▶

ケースバイケース(「判断」に近い?) ▶

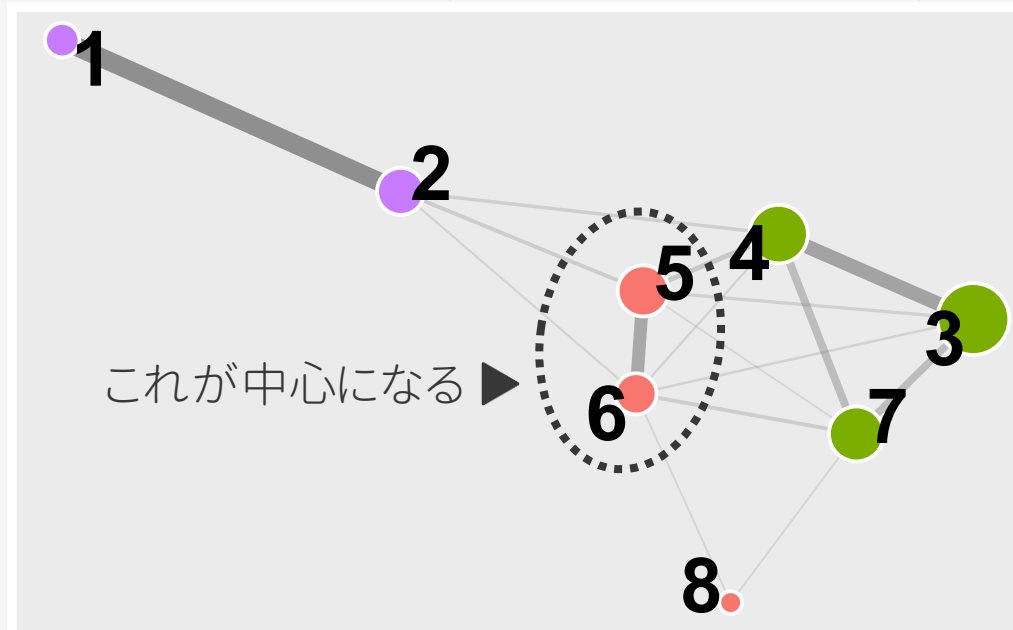
「想起」の難しさ ▶

「答えにくさ」評価要素

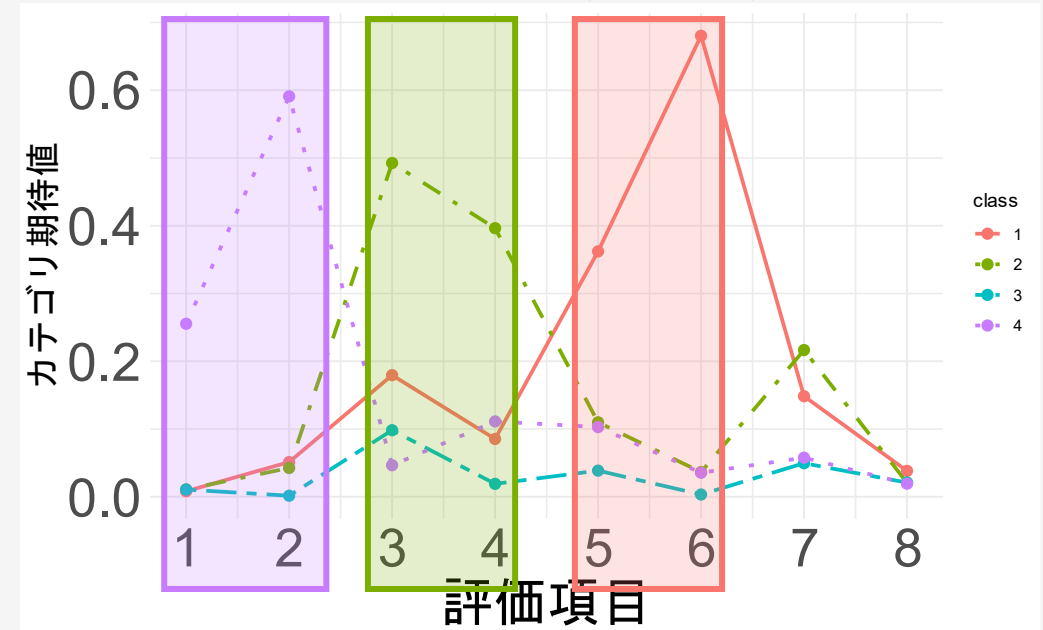
平均選択率

1 この項目にはわかりにくい言葉が入っていると感じた	0.030
2 この項目の意味がわかりにくかった	0.059
3 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	0.174
4 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	0.102
5 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	0.072
6 この項目に関連することを正確に思い出すのが難しい	0.043
7 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	0.087
8 この項目に正直に回答するのをためらった	0.021

ネットワーク分析(Louvain法にて色分け)



潜在クラス分析(4クラス)



- 回答者に「答えにくさ」を主観的に尋ねることはある程度できそう
- 実際の心理検査でも、答えにくいと感じられることはそこそこある
特に「ケースバイケース」というのは一般的に起こりやすそう
- ウェブ調査なのでどの程度まともに回答していたかは不明
1ヶ月後に再度同じ回答をしてもらったが、一致率(Dice係数)は0.6程度、ほぼ0の人も
- 回答に困難感を抱えるような特性の人が正しく答えられるかは要検証
なかなか高度なメタ思考が要求されてしまうような気がする
答えられるのであれば、この方法はいろいろな応用が効きそう

■ 「答えにくい」項目に直面したとき,どのように考えて回答に到達するのか?

▶ 特に臨床場面での,個別具体的なサポートにつなげるための知見を得たい

■ 参加者: とある国立大学の 大学生・大学院生22名

■ 手続き: 回答困難項目に発話しながら回答してもらう→事後でインタビュー

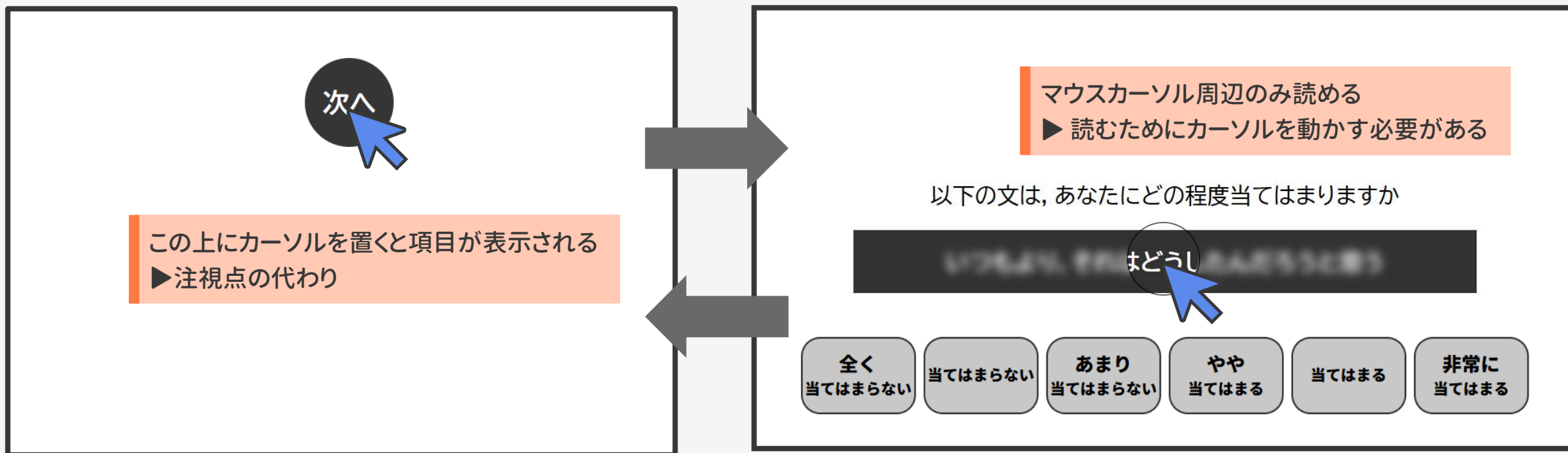
回答困難項目(+ 回答しやすい項目: E)	
1	自分には頑固な部分があると思う
2	出かけるときは身体が重い
3	いつもより,それはどうしたんだろうと思う
4	人と関わるのは嫌いではないと思わないわけではない
5	飛行機を操縦するとき,緊張より楽しみのほうが大きい
6	先週の水曜日は良い一日を過ごせたと思う
7	最近,性的欲求を感じることもある
8	読書をしたり,体を動かしたりするのが好きだ
9	貯金額は人並み以上にある
E1	1人で過ごす時間は気楽だ
E2	いま,悲しいと感じる
E3	自分には助けってくれる人がいる

必ず最後に提示(読めないと発話が抑制されてしまうため)

センシティブ過ぎるので不使用

■ マウスカーソルの研究に使えないかと思って入れた仕掛け

ただ、まだうまく使えていないのでこれ以降でお話しすることはありません。



ミックスメソッドを採用

…と言いつつ,今回は定量的な分析の結果を報告します

1

発話を切片化して,ラベル・コード・カテゴリを付与する

Tourangeauのモデルに基づく

切片	ラベル	コード					カテゴリ
		理解	想起	判断	回答	該当なし	
先週の水曜日は,良い一日を過ごしたと思う。	質問項目を読み上げる	1					該当日のことをよく思い出せないまま判断する
水曜日,全然何してたか思い出せんけど,	過去の出来事を振り返る		1				
悪いってイメージはないから,	過去の出来事に対する感覚を振り返る			1			
やや当てはまるぐらいかな	選択肢に当てはめる				1		

■ ミックスメソッドを採用

2 コードの系列をもとにクラスター分析を行う

すべての回答者×項目 ($n = 22 \times 11 = 242$) について、コードの系列を求める

▲ 前ページの例では{理解,想起,判断,回答}

系列間の距離を「編集距離」に基づいて決定

{理解,想起,判断,回答}

↓ (想起を1つ追加する)

↓ (判断を理解に置換する)

{理解,想起,想起,理解,回答}

遷移の起こりやすさをもとにコストを算出

From / To	理解	想起	判断	回答
理解	-	0.2	0.4	0.5
想起	0.6	-	0.1	0.3
判断	0.8	0.4	-	0.2
回答	1.1	0.7	0.8	-

■ ミックスメソッドを採用

3 回答者が各クラスターに割り当てられた項目数をもとにグルーピング

回答者の傾向のようなものを捉えられないか？

項目レベルでは同じ分析は行っていません
▲ 項目自体が恣意的に作られたものであるため

回答者ID	クラスター					計
	1	2	3	4	5	
A	4	3	3	1	0	11
B	8	0	0	3	0	11
C	2	1	0	1	7	11
⋮						11

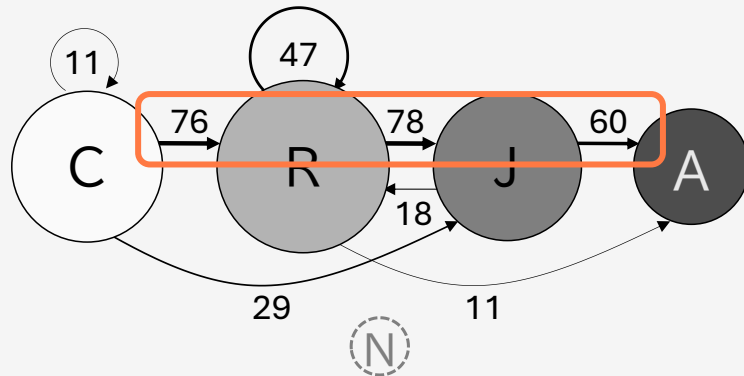
Hellinger Distance (Legendre & Gallagher, 2001)

$$H(A, B) = \sqrt{\sum_{k=1}^5 \left(\sqrt{\frac{n_{Ak}}{11}} - \sqrt{\frac{n_{Bk}}{11}} \right)^2}$$

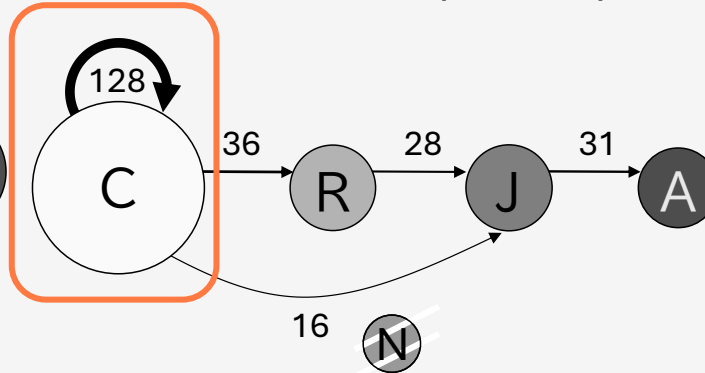
階層的クラスター分析

■ 解釈可能性から5つのクラスターに分類することに

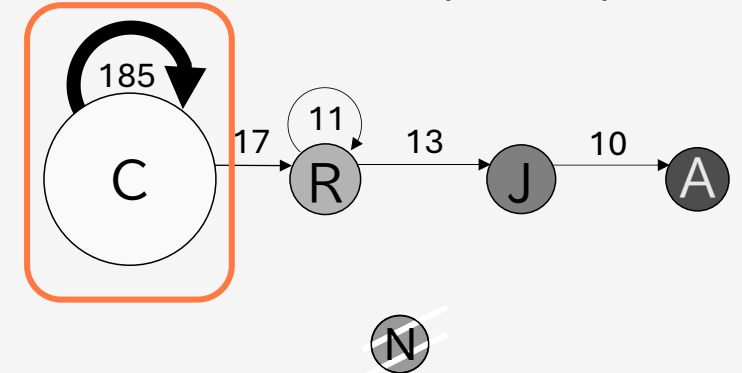
順番に進む

クラスター1 ($n = 109$)

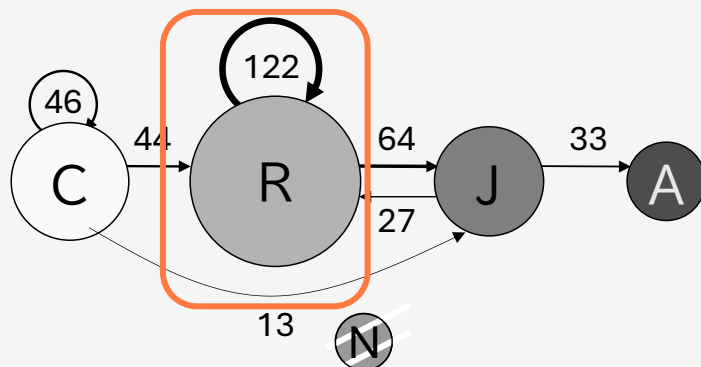
理解で止まる

クラスター2 ($n = 42$)

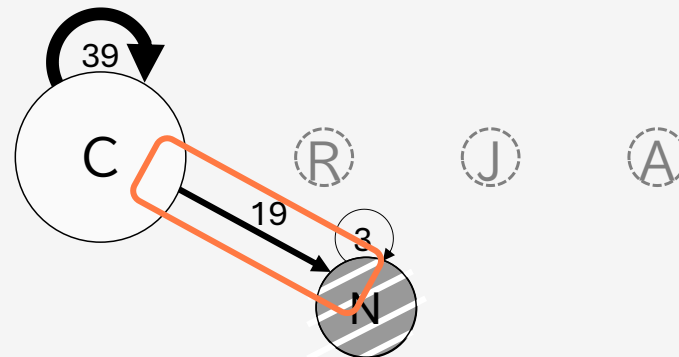
もっと理解で止まる

クラスター3 ($n = 20$)

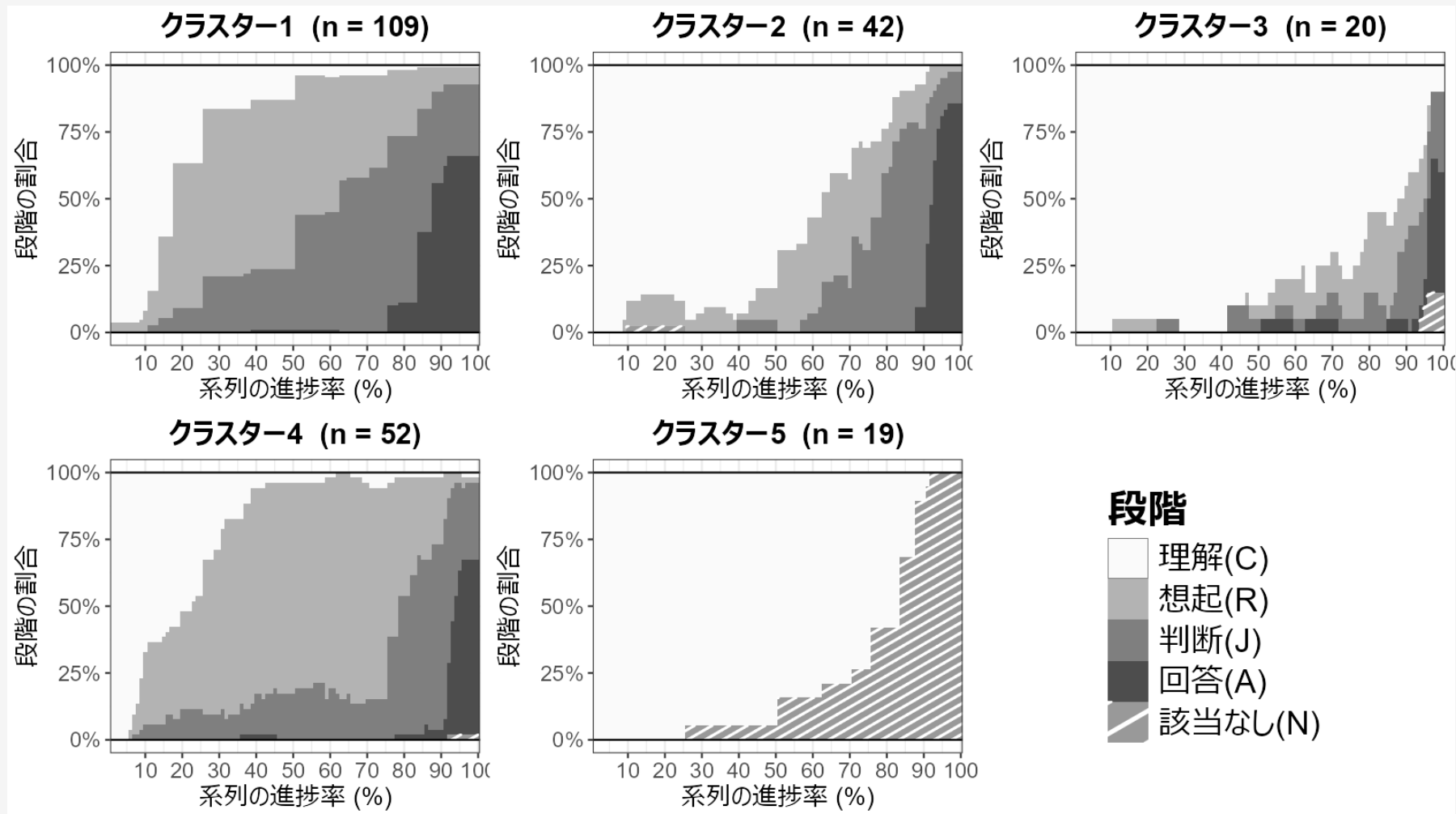
想起で止まる

クラスター4 ($n = 52$)

回答をあきらめる

クラスター5 ($n = 19$)

■ クラスター別・相対時間に対する各コードの割合

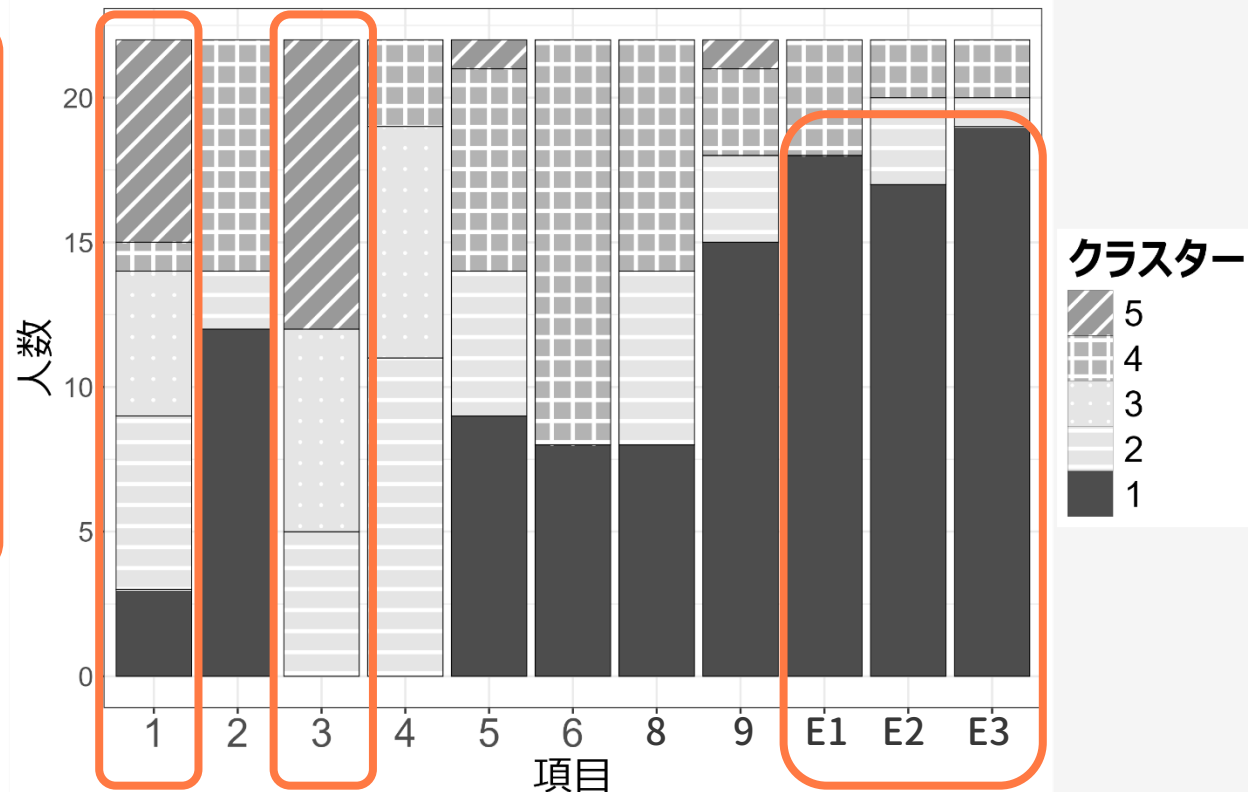
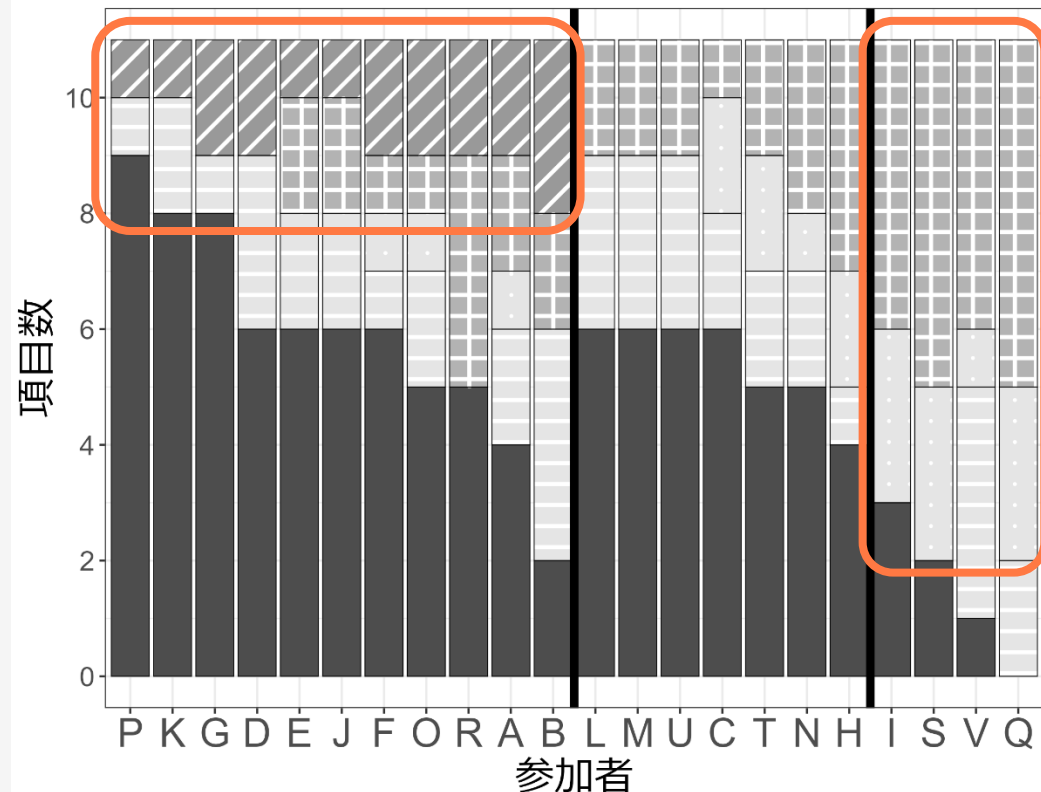


■ クラスター別・相対時間に対する各コードの割合

▼ 記述統計量としての結果

クラスター5があるグループ 参加者ごと かなり悩みがち

項目ごと



表面的な意味理解の失敗が
「わからない」につながりやすい?

答えやすい項目では
理論通りの挙動

- **答えにくい質問項目に対して人がどう反応するか,の片鱗が見えた**
「ダブルバーレルは良くない」とは聞かすが,実際にどのような認知プロセスを誘発するかがわかった,という意味で重要な知見
- **答えやすいと思われた項目でも引っかかることはある**
特に真面目に答えようとするほどつまずいてしまうかも?
- **「わからない」があってもこれを使わない人が一定数いる**
もちろん対面実験の効果はあるだろうが,
ウェブ調査でも「真面目に見えるように」そのような方略を取る人はいてもおかしくない
- **(要検証)様々な属性の人でどのようなパターンが見えるか**
今回は相対的に優秀なほうの大学生・大学院生だった点には要注意

Outline

1 問題意識

2 研究紹介

3 将来の展望

「受験(受検/回答)者視点から問い直すテストのあり方」

■ 回答者の立場にたって設計しましょう

とはいえProlificの最低謝礼は
時給換算6ポンド(1200円くらい)

ウェブ調査モニターの質が低下している原因は「謝礼の安さ」だけではない
回答してくれる人を「データ源」としか考えていない調査はもうやめましょう

■ どれだけ手を尽くしてもつまずく人は一定数います

これを「不注意回答」などとして除外するかは要注意
とても真面目なのに、ただ質問項目が悪いせいで除外するとバイアスにつながるかも

■ 認知インタビューをもっと活用したほうが良い気がします

テスト開発者は量的・質的手法を使えないといけないようです。私も頑張ります。

■ マウスカーソルなどは使えないか？

元気ですか？

(予想)

「ふつう」と迷った人は
少し動きがふらふら？

(予想)

迷わなかった人は
まっすぐに進む？

全く

あまり

ふつう

やや

とても

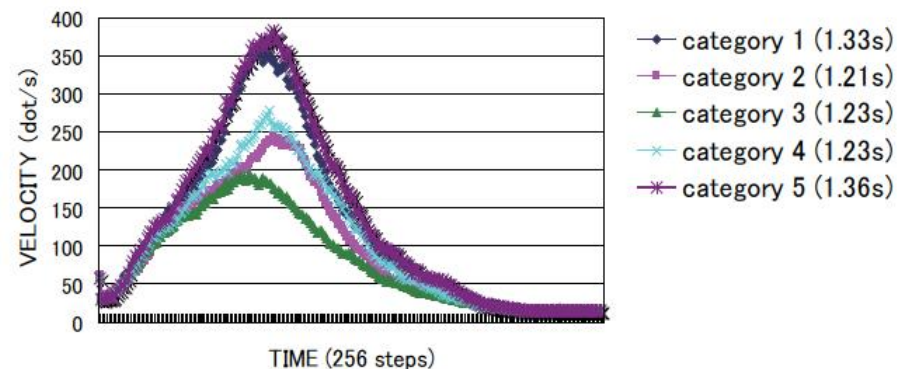


図4 接線速度の回答カテゴリーごとの平均曲線。
横軸は基準化された時間、縦軸は接線速度。
ベンチマーク課題。

▲ 椎名 (2019) より
端のカテゴリを選ぶ人ほど
瞬間最高速度が大きい

研究1によって「正解ラベル」が得られるようになった,と考えています

■ 個別的なサポートの提供や,データの偏りの補正などに使えるかも

考えられる応用例1

高齢者は選択肢の副詞に引っかかりやすい
としたら



同じ検査でも,副詞をわかりやすく改良した
バージョンを作成しよう



高齢者からも従来より正しく
回答してもらえるようになるかもしれない!

考えられる応用例2

最終学歴が低いほど
文の意味をとるのに失敗しやすい
としたら



結果として「不適切回答」と判断されて
除外される割合が高くなる



とはいえ除外しないと
正しく理解せずに回答した
データが混ざってしまって良くないかも

その分を補正することで
結果のバイアスを抑制できるかも!

「答えやすさ」は質問文だけでは決まらないはずで

同じウェブ調査でも、PCとスマホで同じ「読みやすさ」だとは思えない

▶ とある調査会社で同じ調査を
PCとスマホで見たときのレイアウト

10

以下の各文に示された行動を、あなたが普段どの程度行っているか、最も当てはまる選択肢を選んでください。*回答必須
業務中・プライベートを問わず、日常生活においてどの程度行っているかを回答してください。

	まったくしない	ほとんどしない	ときどきする	よくする	いつもする
1.長時間パソコンを使用しない場合、自動的に画面がロックされるように設定している。	☐	☐	☐	☐	☐
2.ノートパソコンやタブレットにパスワードロックをかけている。	☐	☐	☐	☐	☐
3.パソコンから離れる際は、手動でパソコンの画面をロックしている。	☐	☐	☐	☐	☐
4.携帯電話やスマートフォンのロックを解除する際はパスワードを入力している。	☐	☐	☐	☐	☐

10

以下の各文に示された行動を、あなたが普段どの程度行っているか、最も当てはまる選択肢を選んでください。*回答必須
業務中・プライベートを問わず、日常生活においてどの程度行っているかを回答してください。

1.長時間パソコンを使用しない場合、自動的に画面がロックされるように設定している。

☐ まったくしない
☐ ほとんどしない
☐ ときどきする
☐ よくする
☐ いつもする

2.ノートパソコンやタブレットにパスワードロックをかけている。

☐ まったくしない
☐ ほとんどしない
☐ ときどきする
☐ よくする
☐ いつもする

3.パソコンから離れる際は、手動でパソコンの画面をロックしている。

☐ まったくしない
☐ ほとんどしない
☐ ときどきする
☐ よくする
☐ いつもする

細かいUIの違いが細かい結果の違いにつながるのではないか

色, フォント, 要素サイズ, 間隔, ページ分割, 回答方法, 方向 (Tourangeau et al. 2004; Yan et al., 2018) ……

Couper et al. (2013): 色やフォントの違いを検証 ▶ ほとんど違いなし

適切にグリッドを分割したほうが回答の質が上がった

(余談) そもそも心理尺度のレイアウトは標準化されなくて良いのか

■ 非言語情報を軽視し過ぎでは？

以下の2つの尺度は本当に同じと言えるのか？

以下の文は、あなたにどの程度当てはまりますか。1を「全く当てはまらない」、6を「非常に当てはまる」として1つ選んでください。

朝は食欲が わかない	1		6
アウトドアな 性格だ			
	⋮		

▼ Google フォーム

以下の文は、あなたにどの程度当てはまりますか。
「非常に当てはまる」から「全く当てはまらない」
の中から1つ選んでください。

1. 朝は食欲がわかない

- ☐ 非常に当てはまる
- ☐ 当てはまる
- ☐ どちらでもない
- ☐ あまり当てはまらない
- ☐ 全く当てはまらない

2. アウトドアな性格だ

☐ 非常に当てはまる

■ 尺度・検査の開発者・ユーザーは使用したUIまで示すと良い気がしています。

最近は大いぶマシになってきているのかもしれませんが

回答者はリッカート尺度に本当に 正しく回答できているのか？

—心理尺度回答における「つまずき」から回答者視点のテストを考える—

分寺 杏介



bunji@bear.kobe-u.ac.jp

神戸大学 経営学研究科

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「受験者視点から問い直すテストのあり方」



※本スライドは、クリエイティブ・コモンズ 表示-非営利 4.0 国際 ライセンス (CC BY-NC 4.0)に従って利用が可能です。

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>

分寺杏介・馬場絢子・福島綾. (2024). 心理尺度に対する「答えにくい」感覚を捉える. 日本心理学会第88回大会

分寺杏介・馬場絢子・福島綾. (2025). 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

分寺杏介・馬場絢子・福島綾. (2026). 質問紙法の心理検査における「答えにくさ」の主観的弁別. 日本心理学会第90回大会

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, 12, 389–437.

<https://doi.org/10.2307/270748>

Couper, M. P., Tourangeau, R., Conrad, F. G., & Zhang, C. (2013). The design of grids in web surveys. *Social Science Computer Review*,

31(3), 322–345. <https://doi.org/10.1177/0894439312469865>

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4), 464–473. <https://doi.org/10.1177/1948550619875149>

Fowler, C., Jiao, J., & Pitts, M. (2022). Frustration and ennui among Amazon MTurk workers. *Behavior Research Methods*, 55(6), 3009–3025. <https://doi.org/10.3758/s13428-022-01955-9>

福島綾・分寺杏介・馬場絢子. (2026). リッカート型項目への回答はどこで止まり, 逸れるのか. 日本心理学会第90回大会

Knäuper, B., Carrière, K., Chamandy, M., Xu, Z., Schwarz, N., & Rosen, N. O. (2016). How aging affects self-reports. *European Journal of Ageing*, 13(2), 185–193. <https://doi.org/10.1007/s10433-016-0369-0>

- Kramer, J. M., & Schwartz, A. (2017). Reducing barriers to patient-reported outcome measures for people with cognitive impairments. *Archives of Physical Medicine and Rehabilitation*, 98(8), 1705–1715. <https://doi.org/10.1016/j.apmr.2017.03.011>
- Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2), 271–280. <https://doi.org/10.1007/s004420100716>
- 三浦麻子・小林哲郎. (2016). オンライン調査における努力の最小限化 (Satisfice) 傾向の比較: IMC違反率を指標として. *メディア・情報・コミュニケーション研究*, 1, 27–42.
- Morren, M., & Paas, L. J. (2020). Short and long instructional manipulation checks: What do they measure? *International Journal of Public Opinion Research*, 32(4), 790–800. <https://doi.org/10.1093/ijpor/edz046>
- Schwarz, N., Park, D., Knauper, B., & Sudman, S. (Eds.). (1998). *Cognition, aging and self-reports* (0 ed.). Psychology Press. <https://doi.org/10.4324/9780203345115>
- Tourangeau, R. (2018). The survey response process from a cognitive viewpoint. *Quality Assurance in Education*, 26(2), 169–181. <https://doi.org/10.1108/QAE-06-2017-0034>
- Tourangeau, R., Couper, M. P., & Conrad, F. G. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, 68(3), 368–393. <https://doi.org/10.1093/poq/nfh035>
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge University Press.
- 山田怜生・五十嵐祐 (2025) 参加者が考える日本のクラウドソーシング調査の改善点. 日本心理学会第89回大会
- 山田怜生・五十嵐祐 (2026) 参加者が考える学術調査の改善点は参加経験によって異なるか. 日本心理学会第90回大会
- Yan, T., Keusch, F., & He, L. (2018). The impact of question and scale characteristics on scale direction effects. *Survey Practice*, 11(2), 1–10. <https://doi.org/10.29115/SP-2018-0008>