

# 自由記述回答の定量評価に向けた画像説明タスクの提案： 進捗バー速度と回答欄サイズの影響比較

畠中 健壱<sup>1</sup> 中村 聰史<sup>1</sup>

**概要：**自由記述設問は、選択式設問に比べて回答者の思考や理由付けをより詳細に引き出せるため、アンケートにおいて重要な役割を担っている。しかし実際には、「特になし」や無意味な文字列といった不真面目回答が一定数生じることが知られており、これらを抑制するためにさまざまな工夫が提案されてきた。一方、既存研究で用いられてきた自由記述設問は、設問内容によっては回答者が十分に回答できない場合があり、提示された手法を純粋に評価することが難しいという問題がある。つまり、すべての回答者が回答可能でありつつ、記述量に自然な個人差が生じる設問設計が求められる。そこで本研究では、画像内で複数の出来事が同時に起こっている「画像説明タスク」を提案し、誰でも回答可能でありながら記述量の差が明確に現れる状況を意図的に構築した。またこのタスクを用いて、進捗状況バーの表示速度および回答欄の大きさというUI要因が自由記述回答に与える影響を比較し、画像説明タスクの有用性と、不真面目回答の抑制や回答質の向上に寄与する要因を検証した。実験の結果、画像説明タスクはUIデザインの効果を多角的に検証するための手法として有効であることが示された。

**キーワード：**アンケート、自由記述、離脱、回答欄、画像説明

## 1. はじめに

Web アンケートは紙ベースのアンケートに比べ安価で[1,2]、手軽に多くの回答を集めることができる[3]。また Web アンケートは瞬時に集計を行うことができる[4,5]ことなどから、広く利用されているアンケート手法の一つとなっている[6,7]。その中で自由記述形式の設問は選択形式の設問と比べ、回答者の態度や評価に関わる重要な洞察を提供することができ[8]、また多様な回答を得ることができる[9-11]ためアンケートにおいて必要不可欠な要素となっている。

しかし、自由記述設問は選択形式の設問よりも回答する負担が大きい。これは選択肢による意味推測ができないこと[12]や、回答を自身の言葉で構成する必要があること[13]から、選択設問と比べ回答に時間がかかること[14-16]、などが原因である。その結果、無回答や意味をもたない語句などの不真面目回答が生じやすいことが知られている[9,17,18]。実際に Holland ら[19]は、自由記述形式の設問で、そもそも何も回答しない無回答や、「わからない」、「fdjkgf」、「xxx」、「ただ何となく」などの意味をもたない回答が集まったと報告している。こうした不真面目回答が

多く集まってしまうと、データ除外の手間や有効データの減少といった問題が生じる。そのため、不真面目回答が集まらないようにすることが求められる。

自由記述設問における回答の質を高めるための研究はこれまで多く行われている。例えば、設問文のわかりやすさが回答の質に影響することを示した研究[20]や、回答欄の大きさが記述量や回答特性に及ぼす影響を調べた研究[18,21]、さらに設問の配置や順序が不真面目回答や回答時間に影響することを示した研究[22,23]など、様々な観点から自由記述設問を改善する取り組みが行われている。

こうした研究において、アンケートテーマや設問の内容そのものが回答しやすさに大きく影響してしまう課題がある。特定のカテゴリやテーマに関するアンケートや自由記述設問は、回答者の興味・知識・経験の有無によって記述量が大きく変化する。実際に、学生生活をテーマにしたもの[18]や漫画をテーマにしたもの[24]などを用いて評価を行う研究も存在するが、こうしたテーマは回答者の関心の有無によって記述できる量が大きく異なる。そのため記述量の違いがUIの効果によるものなのか、回答者や設問の性質によるものなのかを切り分けることが難しい。そこで、全ての回答者が無理なく回答できる一方で、記述量に自然な個人差が生じるような設問を用いることが望ましい。具体的には、特定のテーマや知識に依存しない共通の

<sup>1</sup> 明治大学  
Meiji University



図 1: Cookie Theft Picture ([25] より)

刺激を用いることで、回答者の背景知識や経験の差による影響を最小限に抑え UI の違いが回答行動に及ぼす影響をより純粹に比較することが可能となる。

このような要件を満たす設問形式として、画像を説明するタスクが挙げられる。画像説明タスクは、提示された画像に含まれる出来事や状況を回答者が文章で説明するものであり、その一例として Boston Diagnostic Aphasia Examination の Cookie Theft Picture [25] (図 1) が知られている。この種のタスクは、回答者が画像から読み取った情報をもとに説明を行うため、回答が特定のテーマに関する知識や興味の有無に左右されにくい。そのため全ての回答者が必ず何らかの説明を行うことが可能であり、またどの範囲に注目するかやどれだけ詳しく説明するかといった点で自然な個人差が現れる。このような特徴から、画像説明タスクは自由記述設問の UI などの影響を比較するのに適したタスクであるといえる。

そこで本研究では、画像説明タスクが自由記述設問の評価に適した形式となり得るかを検証する。そのうえで、本タスクに対する UI 要因の影響を確認するため、回答欄の大きさと進捗バーの速度を操作し、回答行動や記述量の変化を比較する。

## 2. 関連研究

### 2.1 データ品質に関する研究

クラウドソーシングや Web アンケートは手軽に多くの回答を集めることができるのである一方で、データ品質に課題が挙げられる。こうしたデータ品質の課題についてさまざまな研究で報告されている。Brühlmann ら [26] は、クラウドソーシングサンプル (N=394) において、不注意な行動を系統的に評価した。その結果参加者の 45.9% が何らかの不注意な行動をしていることを明らかにしている。同様に Burns ら [27] は、MTurk で募集したサンプルの内、およそ 30% がチェック質問に失敗していたと報告している。このようにクラウドソーシングのサンプルは不真面目回答が集まるという問題がある。

こうした現状に対し、Gottfried [28] は 2022 年に心理学分野で発表された 3298 の論文について、55% の論文がデータの品質の評価を行っていないと指摘している。データの

品質が悪いデータが含まれていると、相関を弱めたり、本来明らかになるべき項目の関係性や構造が正確に観測・評価できなくなる [29, 30]。そのため、本研究ではこうしたデータ品質の問題について解決することを目指す。

### 2.2 不真面目回答に関する研究

不真面目回答が集まらないようにする取り組みも行われている。特にアンケートの設計は回答の質に影響を与えること [31] から重要であり、さまざまな研究が行われている。

アンケートにおける設問文や事前に回答者に提示される文章に着目した研究も多く行われている。Lenzner ら [20] は、理解しにくい質問を受けた人は、理解しやすい質問を受けた人より、離脱率や回答時間、回答の一貫性などの指標に基づく回答の質が低いと報告している。また Chaudhary [32] は、設問での重要性に関する記述を追加することで、設問の回答率が増加することを明らかにしている。さらに Clifford [33] は、無作為に抽出した 4 つの異なるスタイルの警告メッセージが回答者の行動にどのように影響しているのか分析を行った。その結果、3 種類の警告が回答者の意欲を大幅に向上させることが明らかになった。一方、一部の人にとっては社会的に望ましい回答を増加させるという結果も得られたとしている。このように、設問文のわかりやすさや重要性の記述は質の良い回答を得るために有効な手法であるため、アンケートのデザインだけでなく、設問文についても注意し設計する必要がある。

本研究で着目している自由記述設問についても、回答欄の大きさや設問順序に関する研究はさまざま行われている。Smyth ら [18] は、大学生に向けた Web アンケートにおいて、自由記述設問の回答欄の大きさを変更することが回答にどのような影響を与えるのか調査を行った。その結果、回答欄の大きさが大きく表示された人は、小さく表示されていた人より、回答の質が高くなることを示している。また Smith ら [21] は、大きい回答欄を使用することで回答が長くなり、実際の口頭表現に近い回答が得られるようになったとしている。その他にも同様に大きい回答欄は小さい回答欄と比べ文字数や単語数が多くなると報告されており [32, 34, 35]、大きい回答欄は質の良い回答を得るための有効な手段である。

このようにアンケートにおいて良い回答を得るために研究は多く行われており、本研究はこうしたアンケートの UI などに与える影響を調査する際に有効なタスクを検討するものである。

## 3. 画像説明タスク

### 3.1 自由記述設問の定量評価に求められる要件

自由記述設問の UI などを評価する際、全ての回答者が無理なく回答できることに加え、記述量に自然な個人差が生じることが求められる。また設問テーマに対する興味や

知識の有無によって回答しやすさが左右されないことも重要である。これらの要件を満たす設問を用いることで、UIなどの違いによる影響をより純粋に比較しやすくなる。

### 3.2 画像説明タスクの特徴

3.1 節に示す条件を満たすタスクとして図1に示すような複数の出来事が含まれる画像を説明する画像説明タスクが有効であると考えられる。この形式のタスクは、画像が提示され説明をさせるため、全ての回答者が何らかの説明を行うことが可能である。また複数の出来事が起こっているため、どの要素に注目するか、どの程度詳しく説明するかといった点で自然な個人差が生じる。さらに、場面を説明するだけであるため設問の興味や知識の有無に左右されにくい。

### 3.3 本研究で用いる画像刺激の設計

本研究では画像説明タスクとして、Cookie Theft Picture(図1)の構造を参考にしつつ、回答者が理解しやすい日常場面をもとに新たな画像刺激を作成した。具体的には、まずVLMを用いてCookie Theft Pictureの説明文を生成した。次に、その説明文と同様の構造をもつ新たな日常場面の説明文を生成し、最後にその説明文を基に画像生成を行った。本研究で実際に用いた画像は図2であり、後述する実験において刺激として使用する。

## 4. 実験

### 4.1 実験概要

本研究では、画像説明タスクが自由記述設問の評価に適した形式となり得るかを検証するとともに、UI要因が自由記述回答に与える影響を明らかにすることを目的とし実験を行う。実験では、3.3節の手順で生成した画像(図2)を提示し、その内容を説明する自由記述設問を設けた。また回答欄の大きさと進捗バーの表示速度を操作し、得られた自由記述の量や回答行動に違いが生じるかを比較した。

### 4.2 実験設計

本研究では、回答欄の大きさと進捗バーの表示速度を独立変数とする $3 \times 3$ の要因計画を用いた。回答欄の大きさはlarge, medium, smallの3水準とし、進捗バーの速度はfast→slow, none, slow→fastの3水準とした。これにより9条件が構成され、参加者はそのうちのいずれかにランダムに割り当てられた。

進捗バーの挙動は、先行研究[36]を参考に図3に示す通り設計した。進捗バーはすべての条件で0%から開始し、ページ番号を $x$ 、全ページ数を $N$ としたときの関数 $y(x)$ に従って進むように設定した。fast→slow条件では以下の対数関数を用い、序盤で急速に進み後半で緩やかに進む挙動とした。

$$y(x) = \frac{\log(x)}{\log(N)}$$

slow→fast条件では以下の関数を用い、序盤でほとんど進まず後半で加速する挙動とした。

$$y(x) = \left(\frac{x}{N}\right)^e$$

またnone条件ではバーを表示しないものとし、全ての条件において全体の設問数、現在の進捗(%)は回答者に提示しないこととした。

回答欄の大きさは高さのみを条件間で操作し、largeを240px、mediumを160px、smallを80pxとし、その他のUI要素(フォントサイズ、行間、余白など)は全条件で統一した。

実験で提示した設問は、3枚の画像(図2)それぞれに対して自由記述1問と選択設問3問を提示し、合計で12問とした。設問内容の一覧を表1に示す。自由記述設問では、画像から読み取れる出来事をできるだけ多く記述するよう指示し、回答者の記述量が十分に観測できるようにした。一方、選択設問については、本研究の操作要因(回答欄の大きさや進捗バー)に参加者が注意を向かないよう、一般的なアンケートにおける構成に近い形で設問を配置した。

アンケートシステムはNext.jsを用いて実装した。実際のシステムイメージを図4に示す。参加者が設問順序の影響を受けずに回答できるように、回答中に前のページへ戻ることはできない仕様とした。また全ての設問において、回答が入力されるまで「次へ」ボタンは押せないようにし、未回答のまま進行することを防いだ。

実験参加者の募集にはYahoo!クラウドソーシングを利用した。実験参加者の男女比が同程度となるように、男性・女性と分け、800人ずつ計1600人募集した。実験参加者の対象は男性・女性ともに成人以上である。

### 4.3 実験手順

実験参加者は、スマートフォンを用いSafariもしくはGoogle Chromeを利用し、Yahoo!クラウドソーシングのページからアンケートシステムにアクセスする。参加者は注意事項を読み回答を開始するボタンをクリックとアンケート開始となり、1つ目の設問が表示される。参加者が設問への回答を入力もしくは選択し、「次へ」のボタンを押すことで2つ目の質問が表示される。このようにして、参加者が計12問の設問に対して回答を終えると、システムはアンケートの終了を知らせるページに遷移する。

終了ページでは実験コードと参加者ごとにユニークなIDが表示され、参加者はそのコードとIDをクラウドソーシングの設問の画面に入力し、その後自身の年齢区分を選択することで終了する。参加者は正しいコードを入力することで報酬を受け取ることができる。



(a) リビングの場面



(b) 公園の場面



(c) 学校の場面

図 2: 本研究で生成し実験に用いた画像刺激

表 1: 画像説明タスクにおける設問内容一覧

設問番号	設問内容	設問形式
Q1, Q5, Q9	画像を見て起こっている出来事について、できるだけ多く記述してください。	自由記述
Q2, Q6, Q10	この場面には、全体としてまとまりや一体感があるように感じる。	5 択 (そう思う～そう思わない)
Q3, Q7, Q11	この場面には、緊張感や慌ただしさがあるように感じる。	5 択 (そう思う～そう思わない)
Q4, Q8, Q12	この場面には、日常の一コマらしさを感じる。	5 択 (そう思う～そう思わない)

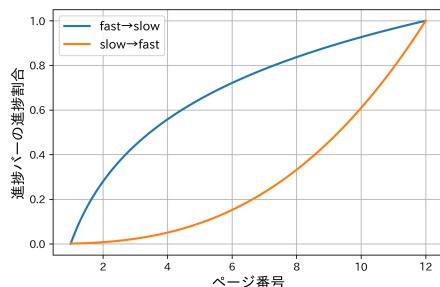


図 3: 進歩バーの挙動



図 4: 実際のシステム (slow→fast, large 条件の設問 5)

## 5. 結果

### 5.1 データ処理

収集された男性 800 名、女性 800 名の計 1600 名のうち、クラウドソーシング上のタスク完了時に入力された ID に不備があり、その結果システムログとの照合が不可能であった 19 名を除外した。最終的な有効回答数は  $N = 1581$  であった。参加者の属性内訳は、男性 790 名、女性 791 名であり、年代別では 20 代が 109 名、30 代が 211 名、40 代

表 2: 各条件への割り当て人数

回答欄	進歩バーの速度		
	fast → slow	none	slow → fast
large	174	166	166
medium	179	191	183
small	178	177	167

が 451 名、50 代が 528 名、60 代が 212 名、70 代以上が 70 名であった。また、各実験条件（回答欄の大きさ × 進歩バーの速度）への参加者の割り当て人数を表 2 に示す。各条件にはランダムに割り当てられており、概ね均等な人数配置となっていた。

次に、自由記述回答の情報量を定量的に評価するため、本研究では記述内に含まれる「出来事の言及数」を指標とした。なお、評価の基準となる「出来事」のリストは、著者らが事前に各画像刺激を確認し、客観的に視認可能な要素を抽出して作成したものである。

この言及数を効率的かつ正確に計測するために、専用のアノテーションシステム（図 5）を構築した。本システムでは、各画像刺激に対して作成した「出来事（イベント）」のリストが表示され、回答テキスト中に該当する記述がある場合、評価者がチェックボックスを選択することで言及数を集計した。なお、アノテーション作業は 1 名の評価者によって行われた。評価の公平性を保つため、評価者には回答者の ID や割り当てられた実験条件（回答欄の大きさや進歩バーの速度）が一切表示されないブラインド状態で実施した。

### 5.2 離脱率の分析

本節では、実験条件（UI デザイン）による途中離脱率の違



図 5: 記述数を集計するアノテーションシステムの画面

いを検証する。前節で述べた有効回答データ ( $N = 1581$ ) は最後まで回答を完了した参加者に限られるが、離脱率の算出にあたっては、指定期間内に実験システムへアクセスし、最初の設問 (Q1) 以降のログが記録された全参加者を分析対象とした。対象となる総参加者数は  $N = 2568$  名であり、そのうち最後まで回答を完了しなかった参加者は 650 名 (全体の 25.31%) であった。各条件間における離脱率の差を検証するため、カイ二乗検定および調整済み標準化残差による事後分析を行った。

進捗バーの種類による離脱率の違いを表 3 に示す。検定の結果、条件間に有意な差が認められた ( $p < .05$ )。調整済み標準化残差による分析の結果、fast → slow 条件では離脱率が 22.7% と有意に低く (残差 -2.14)、slow → fast 条件では 28.6% と有意に高い値を示した (残差 2.76)。

回答欄の大きさによる離脱率の違いを表 4 に示す。検定の結果、統計的に有意な差には至らなかったものの ( $p < .10$ )、回答欄のサイズが大きくなるにつれて離脱率が増加する傾向が確認された。特に large 条件では離脱率が 27.8% と最も高く、調整済み標準化残差は 2.03 を示した。

回答欄の大きさと進捗バーの組み合わせ (計 9 条件) による離脱率の分析結果を表 5 に示す。検定の結果、条件の組み合わせ間においても有意な差が認められた ( $p < .05$ )。特に large × slow→fast 条件では離脱率が 32.9% と最も高く、調整済み標準化残差は 3.08 を示し、基準値 (1.96) を大きく上回っていた。

また、設問ごとの累積離脱率の推移を図 6 に示す。全条件に共通して、最初の自由記述設問 (Q1) において離脱率の増加が確認された。その後、選択式設問 (Q2–Q4) の区間では増加が緩やかになる一方で、2 回目の自由記述設問である Q5 において再び離脱率の上昇が観測された。

特に slow→fast 条件を含む系列では、Q5 における離脱率の増加幅が他条件と比べて大きく、条件間の差がこの時点から拡大する傾向が見られた。

表 3: 進捗バーの種類と離脱率

条件	離脱者数	母数	離脱率 (%)
fast → slow	191	842	22.7
none	210	856	24.5
slow → fast	249	870	28.6

表 4: 回答欄の大きさと離脱率

条件	離脱者数	母数	離脱率 (%)
large	237	853	27.8
medium	216	868	24.9
small	197	847	23.3

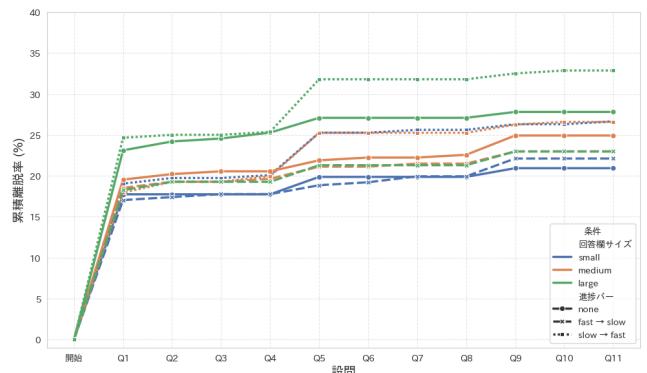


図 6: 設問ごとの累積離脱率の推移

表 5: 各条件の組み合わせにおける最終的な離脱率 (%)

回答欄	進捗バーの種類		
	fast → slow	none	slow → fast
large	23.0	27.8	<b>32.9</b>
medium	23.0	24.9	26.6
small	22.1	20.9	26.6

### 5.3 記述量の分析

UI 条件が自由記述回答の記述量に与える影響を検証するため、総文字数を指標として分析を行った (分析対象データ数  $N = 4438$ )。各条件における総文字数の平均値および標準偏差を表 6 に示す。

二元配置分散分析の結果、回答欄の大きさの主効果が有意であった ( $F(2, 4429) = 5.86, p = .0029$ )。一方、進捗バーの速度の主効果 ( $p = .1167$ ) および回答欄の大きさと進捗バーの速度の交互作用 ( $p = .1830$ ) は認められなかった。

多重比較 (Tukey HSD) の結果、large 条件は small 条件と比較して、総文字数の平均値が有意に高いことが確認された ( $p = .0021$ )。medium 条件と他条件との間には、有意な差は認められなかった。

### 5.4 言及数の分析

画像内の出来事に対する言及数について分析を行った。各条件における言及数の平均値および標準偏差を表 6 に示す。

表 6: 要因ごとの記述指標の平均値と標準偏差 (SD)

要因		文字数		言及数		詳細度	
		平均	SD	平均	SD	平均	SD
回答欄	large	70.69	53.20	3.65	1.72	18.34	8.06
	medium	68.24	45.80	3.66	1.70	17.86	7.01
	small	64.69	43.17	3.61	1.69	17.23	6.78
進捗バー	fast → slow	67.52	48.11	3.66	1.70	17.65	7.45
	none	69.75	47.30	3.68	1.66	18.28	7.49
	slow → fast	66.21	47.12	3.59	1.75	17.47	6.91

二元配置分散分析の結果、回答欄の大きさおよび進捗バーの速度の主効果はいずれも有意ではなかった（それぞれ  $p = .6499, p = .3713$ ）。一方、回答欄の大きさと進捗バーの速度の交互作用が有意であった ( $F(4, 4429) = 2.94, p = .0194$ ).

ただし、多重比較の結果、回答欄の大きさおよび進捗バーの速度の各水準間において、一貫した有意差は確認されなかった。

## 5.5 記述の詳細度の分析

記述内容の内訳に着目するため、1 言及あたりの文字数を指標として分析を行った。各条件における 1 言及あたりの文字数の平均値および標準偏差を表 6 に示す。

二元配置分散分析の結果、回答欄の大きさの主効果 ( $F(2, 4429) = 8.52, p = .0002$ )、および進捗バーの速度の主効果 ( $F(2, 4429) = 5.12, p = .0060$ ) がいずれも有意であった。一方、交互作用は認められなかった ( $p = .9156$ )。

多重比較の結果、回答欄の大きさについては、large 条件および medium 条件が small 条件と比較して 1 言及あたりの文字数の平均値が有意に高いことが確認された。進捗バーの速度については、none 条件が fast → slow 条件よりも高く、none 条件は slow → fast 条件よりも有意に高い値を示した。

## 6. 考察

### 6.1 離脱行動に与える影響

回答欄の大きさについて、大きな回答欄ほど離脱率が高くなる傾向が確認された。統計的には有意傾向にとどまったものの、large 条件では他条件と比較して一貫して高い離脱率が観測されており、視覚的に大きな入力領域が提示されること自体が、回答者にとって心理的負担として作用した可能性が示唆される。この結果は、回答欄の拡大が記述量を増加させる一方で、回答開始前のハードルを高め、離脱を招きやすくなるとする先行研究 [24] の知見と合致している。

次に、進捗バーの速度については、fast → slow 条件で離脱率が最も低く、slow → fast 条件で最も高いという明確な差が確認された。特に slow → fast 条件では、初期段

階で進捗がほとんど進まない表示がなされるため、回答者が「タスクが長い」「負担が大きい」と認識しやすく、その結果として離脱が促進されたと考えられる。一方、fast → slow 条件では、初期段階で進捗が大きく進むことで、タスク全体が短く感じられ、回答継続への動機づけが維持された可能性がある。この点についても、進捗バーの速度が回答行動に強く影響することを示した先行研究 [36] と一致している。

設問ごとの離脱推移に着目すると、すべての条件に共通して、最初の自由記述設問である Q1 において大きな初期離脱が発生していた。また、Q5 といった後続の自由記述設問においても、離脱率の再上昇が確認された。この結果は、自由記述設問が選択式設問と比べて高い認知的負荷を伴うことを示しており、離脱行動の主要な発生要因が自由記述タスクにあることを改めて示唆している。特に注目すべき点は、Q5 における離脱率の増加が、slow → fast 条件で顕著であった点である。Q5 は 2 回目の自由記述設問であり、回答者がタスク全体の負担を具体的に認識し始める段階に相当する。このタイミングで進捗の遅れが強調される slow → fast 条件では、「まだ多くの作業が残っている」という印象が強まり、離脱が集中的に発生したと考えられる。

一方、UI 要因の組み合わせ（交互作用）に着目すると、回答欄が large であっても、fast → slow 条件と組み合わせた場合には、離脱率が small 条件と同程度まで抑制されていた。これは、大きな回答欄によって生じる記述負担の印象が、初期進捗を速く見せるフィードバックによって相殺された可能性を示している。つまり、単一の UI 要因だけでなく、複数の UI 要因を適切に組み合わせることによって、回答者の離脱行動を制御できる余地があることが示唆された。

以上の結果から、回答欄の大きさおよび進捗バーの速度は、それぞれ独立に離脱行動へ影響を及ぼすだけでなく、その組み合わせによって影響の強さが変化することが明らかとなった。特に、large × slow → fast 条件のように、記述負担と進捗遅延が同時に強調される設計は離脱を顕著に増加させる一方で、large × fast → slow 条件のように、進捗の初期印象を工夫することで、記述負担の大きさを補完できる可能性が示された。

### 6.2 記述量および記述内容に与える影響

総文字数に着目すると、回答欄の大きさに有意な主効果が認められ、大きな回答欄ほど記述量が増加することが明らかとなった。この結果は、回答欄の視覚的な大きさが、回答者に対して「多く書くことが期待されている」という印象を与え、結果として記述量を増加させるという先行研究 [24] の知見と一致している。特に本研究では、回答中に画像刺激を提示する形式であったため、スマートフォン

等の画面サイズによっては、入力中に回答欄全体が視認できないケースも多かったと考えられる。それにもかかわらず、回答欄の大きさによる記述量の差が確認された点は、回答者が実際の入力操作中だけでなく、設問提示時に視認した回答欄の大きさから、求められている記述量を事前に認識し、行動を調整していた可能性を示唆している。一方、進捗バーの速度については、総文字数に対する明確な主効果は認められず、記述量そのものを直接的に増減させる要因とはならなかった。

言及数に着目した結果、回答欄の大きさおよび進捗バーの速度のいずれについても、有意な主効果は認められなかった。このことから、回答欄を大きくしたとしても、回答者が新たな出来事に注意を向けるようになったわけではなく、画像内で着目される出来事の範囲自体は条件間で大きく変化していなかったと考えられる。また、1言及あたりの文字数（詳細度）に着目すると、回答欄の大きさおよび進捗バーの速度の双方において有意な主効果が認められた。具体的には、回答欄が大きい条件ほど、1つの出来事に対する記述が有意に長くなる傾向が確認された。この結果は、回答欄の拡大が、回答者に対して新たな出来事を探索するのではなく、既に着目した出来事について、より丁寧に説明すること（精緻化）を促していたことを示唆している。なお、進捗バーの影響については、進捗表示のない（none）条件と比較して、進捗バーがある条件では詳細度がやや低下する傾向が見られた。これは進捗の可視化が効率化への圧力を生んだためと考えられるが、fast → slow 条件は slow → fast 条件よりも高い詳細度を維持しており、離脱率の低さと記述品質のバランスにおいて優れていることが確認された。

### 6.3 画像説明タスクの有効性

画像説明タスクが自由記述設問の UI 評価において有効な代替手法となり得るかについて、回答欄の大きさに関する結果に着目すると、画像説明タスクにおいても、回答欄が大きい条件ほど総文字数が増加するという明確な傾向が確認された。この結果は、従来の自由記述設問（感想記述など）を用いた研究において報告してきた知見と一致しており、画像説明タスクにおいても、回答欄の視覚的な大きさが記述量に影響することが示された。つまり、画像説明タスクは、回答欄サイズという UI 要因に対して、従来の自由記述設問と同様の感度を有しているといえる。

本タスクの特筆すべき利点は、記述量の増加が「情報の網羅性（言及数）」によるものか、「情報の具体性（詳細度）」によるものかを分離して評価できる点にある。従来の自由記述設問では、文字数の増加が単なる冗長な引き伸ばしなのか、内容の深化なのかを客観的に区別することは困難であった。しかし、正解データを持つ画像説明タスクを用いることで、本研究では「回答欄の拡大は、話題の数ではなく

く、記述の深さに寄与する」という詳細な回答行動のを明らかにすることができた。このように、画像説明タスクは、従来の記述量指標だけでなく、記述内容の質的変化や、前節で述べた離脱行動といった多角的な指標を用いて、UI デザインの効果を精緻に検証するための標準タスクとして極めて有効であると考えられる。

## 7. おわりに

本研究では Web アンケートにおける自由記述設問の評価手法を確立するために、画像説明タスクを提案し、UI デザインが回答行動に与える影響を検証手法として有効か調査を行った。その結果、画像説明タスクは従来の自由記述設問と同様の機能を有していることが確認された。加えて、記述量の増加が情報の網羅性によるものか詳細度によるものかを分離して評価できたことから、画像説明タスクは UI デザインの効果を多角的に検証するための手法として有効であると考えられる。

今後は、本手法の再現性と汎用性を高めるために、対象とする UI 要素を拡張して検証していく。本研究では回答欄の大きさと進捗バーの速度という特定の要素に着目したが、他の UI デザイン要素を変更した場合においても、画像説明タスクが同様に回答行動の変化を検出できるかを検証していく。また、提示する画像刺激の作成手法についても検討を行う。画像説明タスクにおいて記述内容の質を正しく評価するためには、回答者がその画像に対して予備知識を持たず、常に初見であることが重要となる。そのため、既存の画像セットを使用するだけでなく、回答者ごとに多様なパターンの画像を自動生成できるような仕組みについても検討していく。

## 参考文献

- [1] Schleyer, T. and Forrest, J.: Research Paper: Methods for the Design and Administration of Web-based Surveys, *Journal of the American Medical Informatics Association : JAMIA*, Vol. 7 4, pp. 416–25 (2000).
- [2] Kaplowitz, M. D., Hadlock, T. D. and Levine, R.: A Comparison of Web and Mail Survey Response Rates, *Public Opinion Quarterly*, Vol. 68, No. 1, pp. 94–101 (2004).
- [3] Vergnaud, A.-C., Touvier, M., Méjean, C., Kesse-Guyot, E., Pollet, C., Malon, A., Castetbon, K. and Hercberg, S.: Agreement between web-based and paper versions of a socio-demographic questionnaire in the NutriNet-Santé study, *International Journal of Public Health*, Vol. 56, No. 4, pp. 407–417 (2011).
- [4] Hayslett, M. and Wildemuth, B.: Pixels or pencils? The relative effectiveness of Web-based versus paper surveys, *Library Information Science Research*, Vol. 26, pp. 73–93 (2004).
- [5] Heiervang, E. and Goodman, R.: Advantages and limitations of web-based surveys: evidence from a child mental health survey, *Social Psychiatry and Psychiatric Epidemiology*, Vol. 46, pp. 69–76 (2009).

- [6] Fricker, R. D. and Schonlau, M.: Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature, *Field Methods*, Vol. 14, No. 4, pp. 347–367 (2002).
- [7] Evans, J. R. and Mathur, A.: The value of online surveys, *Internet Research*, Vol. 15, No. 2, pp. 195–219 (2005).
- [8] Singer, E. and Couper, M. P.: Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys, *Methods, data, analyses: a journal for quantitative methods and survey methodology (mda)*, Vol. 11, No. 2, pp. 115–134 (2017).
- [9] Reja, U., Manfreda, K. L., Hlebec, V. and Vehovar, V.: Open-ended vs. close-ended questions in web questionnaires, *Developments in applied statistics*, Vol. 19, No. 1, pp. 159–177 (2003).
- [10] Baburajan, V., de Abreu e Silva, J. and Pereira, F.: Open-Ended Versus Closed-Ended Responses: A Comparison Study Using Topic Modeling and Factor Analysis, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 22, pp. 2123–2132 (2021).
- [11] Howard Schuman, S. P.: The Open and Closed Question, *American Sociological Review*, Vol. 44, No. 5, pp. 692–712 (1979).
- [12] Dillman, D. A., Smyth, J. D. and Christian, L. M.: *Internet, phone, mail, and mixed-mode surveys: The tailored design method*, John Wiley & Sons (2014).
- [13] Keusch, F.: The influence of answer box format on response behavior on list-style open-ended questions, *Journal of Survey Statistics and Methodology*, Vol. 2, No. 3, pp. 305–322 (2014).
- [14] Couper, M. P. and Kreuter, F.: Using paradata to explore item level response times in surveys, *Journal of the Royal Statistical Society Series A: Statistics in Society*, Vol. 176, No. 1, pp. 271–286 (2013).
- [15] Yan, T. and Tourangeau, R.: Fast times and easy questions: The effects of age, experience and question complexity on web survey response times, *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, Vol. 22, No. 1, pp. 51–68 (2008).
- [16] Rotthoff, T., Baehring, T., Dicken, H.-D., Fahron, U., Richter, B., Fischer, M. and Scherbaum, W.: Comparison between Long-Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial, *BMC Medical Education*, Vol. 6, p. 50 (2006).
- [17] Revilla, M. and Ochoa, C.: Open narrative questions in PC and smartphones: is the device playing a role?, *Quality & Quantity*, Vol. 50, No. 6, pp. 2495–2513 (2016).
- [18] Smyth, J. D., Dillman, D. A., Christian, L. M. and McBride, M.: Open-ended questions in web surveys: Can increasing the size of answer boxes and providing extra verbal instructions improve response quality?, *Public Opinion Quarterly*, Vol. 73, No. 2, pp. 325–337 (2009).
- [19] Holland, J. L. and Christian, L. M.: The influence of topic interest and interactive probing on responses to open-ended questions in web surveys, *Social Science Computer Review*, Vol. 27, No. 2, pp. 196–212 (2009).
- [20] Lenzner, T.: Effects of survey question comprehensibility on response quality, *Field Methods*, Vol. 24, No. 4, pp. 409–428 (2012).
- [21] Smith, T. W.: Little things matter: A sampler of how differences in questionnaire format can affect survey responses, *American Statistical Association* (1995).
- [22] Schmidt, K., Gummer, T. and Roßmann, J.: Effects of respondent and survey characteristics on the response quality of an open-ended attitude question in web surveys, *methods, data, analyses*, Vol. 14, No. 1, p. 32 (2020).
- [23] Yamazaki, I., Hatanaka, K., Nakamura, S. and Komatsu, T.: A Basic Study to Prevent Non-earnest Responses in Web Surveys by Arranging the Order of Open-Ended Questions, *International Conference on Human-Computer Interaction*, Springer, pp. 314–326 (2023).
- [24] Yamazaki, I., Hatanaka, K., Nakamura, S. and Komatsu, T.: The Effects of Order and Text Box Size of Open-ended Questions on Withdrawal Rate and the Length of Response, *35th Australian Conference on Human-Computer Interaction*, pp. 446–453 (2023).
- [25] Goodglass, H., Kaplan, E. and Barresi, B.: *Boston Diagnostic Aphasia Examination – Third Edition*, Lippincott Williams & Wilkins, Baltimore, MD (2001).
- [26] Brühlmann, F., Petralito, S., Aeschbach, L. and Opwis, K.: The quality of data collected online: An investigation of careless responding in a crowdsourced sample, *Methods in Psychology* (2020).
- [27] Burns, G. N., Christiansen, N. D., Morris, M. B., Periard, D. A. and Coaster, J. A.: Effects of applicant personality on resume evaluations, *Journal of Business and Psychology*, Vol. 29, pp. 573–591 (2014).
- [28] Gottfried, J.: Practices in Data-Quality Evaluation: A Large-Scale Review of Online Survey Studies Published in 2022, *Advances in Methods and Practices in Psychological Science* (2024).
- [29] Maniaci, M. R. and Rogge, R. D.: Caring about carelessness: Participant inattention and its effects on research, *Journal of Research in Personality*, Vol. 48, pp. 61–83 (2014).
- [30] Credé, M.: Random responding as a threat to the validity of effect size estimates in correlational research, *Educational and Psychological Measurement*, Vol. 70, No. 4, pp. 596–612 (2010).
- [31] Ganassali, S.: The influence of the design of web survey questionnaires on the quality of responses, *Survey research methods*, Vol. 2, pp. 21–32 (2008).
- [32] Chaudhary, A. K. and Israel, G. D.: Influence of importance statements and box size on response rate and response quality of open-ended questions in web/mail mixed-mode surveys, *Journal of rural social sciences*, Vol. 31, No. 3, p. 140 (2016).
- [33] Clifford, S. and Jerit, J.: Do Attempts to Improve Respondent Attention Increase Social Desirability Bias, *Public Opinion Quarterly*, Vol. 79, pp. 790–802 (2015).
- [34] Maloshonok, N. and Terentev, E.: The impact of visual design and response formats on data quality in a web survey of MOOC students, *Comput. Hum. Behav.*, Vol. 62, pp. 506–515 (2016).
- [35] Linegang, M. and Moroney, W.: Effects of Cover Letter Subject Line and Open-ended Question Response Area on Responding to an Internet Survey, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56, pp. 1268–1272 (2012).
- [36] Conrad, F., Couper, M., Tourangeau, R. and Peytchev, A.: Impact of progress feedback on task completion: first impressions matter, *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, pp. 1921–1924 (2005).