

## ChatGPT との対話を通じた質問力育成支援の枠組みの検討

## Study of a Framework for Supporting the Development of Questioning Skills through Dialogue with ChatGPT

阿部 颯斗, 本多 昂生, 松居 辰則  
Hayato ABE, Koki HONDA, Tatsunori MATSUI  
早稲田大学  
WASEDA University  
Email: a.smile.ngt410@toki.waseda.jp

**あらまし:** 若者の質問力育成を目的に、生成 AI である ChatGPT を活用した実験を行った。まず、ChatGPT が人間の回答について代替可能かを検証するためにチューリングテストを行い、その後 ChatGPT を用いた質問力向上の実験を行った。チューリングテストでは、ChatGPT の回答が人間の回答と見分けがつかない場合が多く、一定の代替可能性が確認された。一方で、質問力向上は確認されなかったため、今後の改善が必要である。

**キーワード:** 質問力、ChatGPT、チューリングテスト、満足感

## 1. はじめに

近年、大学生の批判的思考力の低下が指摘されている<sup>(1)</sup>。そのため、質問力等のコミュニケーション力の育成が必要となっている。質問力は、通常疑問を提示するだけに留まるが、自分の意見を提示して相手の思いを引き出す方法<sup>(2)</sup>もある。そこで、本研究における質問力の定義を「不明点を明らかにした上で、質問者回答者双方が満足感を感じられるようなやり取りを生み出す能力」とした。

また昨今注目されている生成 AI の ChatGPT は、自然言語での対話が可能であり、教育分野での活用が期待されている。本研究では、ChatGPT を活用して質問力を育成する枠組みを構築することを目的とする。

## 2. ChatGPT の利用可能性を目的とした実験

## 2.1 実験概要

本実験は、ChatGPT が人間の回答を代替可能かを検証した。そのため、チューリングテストにおいて ChatGPT の回答を人間の回答と判断した確率が 50% 以上の時に、チューリングテストに合格する<sup>(3)</sup>として、ChatGPT が人間のような回答を生成できると判断した。

対話は全て Slack にて実施した。本実験では、大学一般入試未経験者が経験者に対して大学受験の様子について質問を行った。実験の流れは、初め被験者に質問相手となる実験協力者の受験勉強に関する情報を提示し、その情報を踏まえて相手が満足感を得られるような質問を作成、送信する。次に、その質問に対する返答について、人間と AI のどちらが生成したものかを瞬時に判断して送信し、返答に対する質問を作成して送信する。今回は質問項目を勉強面(科目)、勉強面(メンタル)、生活面、受験後の 4 つに設定している。そのため、質問は 1 人当たり全部で 20 回 (1 項目の質問回数 5 回×質問 4 項目) 実

施した。対象者は、質問者である被験者 6 名、人間の回答を作成する実験協力者 1 名である。

実験後には、質問紙調査と実験中の思考についてインタビュー調査を行い、被験者自身が思う人間と AI の回答の印象を SD 法にて調査し、表 1 にある 20 個の形容詞対について因子分析を行なった。

表 1 SD 法で用いた形容詞 20 対

曖昧な-明確な	丁寧な-粗雑な	心温まる-冷たい	的確な-不的確な
友好的な-敵対的な	幼稚な-成熟した	かっこいい-ださい	重要な-不要な
真実味がある-虚偽の	信頼できる-疑わしい	好ましい-好ましくない	(知識が)一般的な-(知識が)専門的な
大袈裟な-控えめな	攻撃的な-穏やかな	社交的な-内向的な	面白い-つまらない
上品な-下品な	単純な-複雑な	安全な-危険な	客観的な-主観的な

## 2.2 実験結果

本実験全体における ChatGPT の回答を人間の回答と判断した被験者全体平均は 42%と、50%を達成することはできなかった。しかし、被験者 6 名中 3 名は 50%以上であり(70%, 50%, 50%)、概ね ChatGPT が人間のような回答を生成できると言える。

また SD 法による因子分析より、人間の回答の印象について「カリスマ性」「親近感」の因子として抽出された。一方、AI の回答の印象については「信頼性」「魅力的な曖昧さ」の因子として抽出された。

## 2.3 実験考察

本実験では概ね ChatGPT が人間のような回答を出

力できると結論付けた。しかし、全体平均が 50%を超えなかった理由として、実験後のインタビュー調査において AI の回答に具体的な内容が少ないことや統一的な言い回しがあることが指摘された。人間の回答の印象で親近感の因子があることから、より詳細な情報を取り入れる必要があることが明らかとなった。

### 3. ChatGPT を用いた質問力向上の実証実験

#### 3.1 実験概要

前実験を踏まえて、人間のような回答を生成する ChatGPT との対話を繰り返し行い、ChatGPT の質問力評価を参考にすることによって被験者の質問力向上を図る。

対話は Slack にて実施した。本実験では、プロスポーツ選手にインタビューを行うという仮想的な場面を設定し、実験は 3 回に分けて行った。1 回目は、初めに質問相手の情報を提示し、その情報を踏まえて本研究における質問力の定義を満たすような質問を作成、送信する。そして得られた回答から再度質問を行った。2 回目以降は、質問作成前に前回の実験を踏まえた ChatGPT による質問力評価を提示して、被験者が行う質問の参考にしてもらった。3 回の実験いずれも異なるプロスポーツ選手を相手にしている。質問力評価は、各質問に対して質問力の 10 段階評価と、一言評価と改善点を提示した。各実験に対する総合評価では、質問の適切性、深掘り力、多様性、流れ、具体性、創造性の 6 項目について 10 段階評価を行い、最後に総合フィードバックを 3 点提示した。質問は全部で 24 回(1 実験の質問 8 回×3 実験)行った。

対話の内容を踏まえて、第三者である質問評価者が全部の質問について 10 段階の質問力評価と、各回実験について 10 段階の総合質問力評価を行った。各実験の総合質問力評価について分散分析(ANOVA)を行い、被験者の質問力向上を検討する。対象者は、質問者である被験者 2 名、被験者 1 の質問評価者 11 名、被験者 2 の質問評価者 9 名である。

#### 3.2 実験結果

分散分析(ANOVA)を用いて、1 回目、2 回目、3 回目における質問評価者の質問力平均値について有意差の有無を検討した。その結果、被験者 2 名共に質問力に統計的な有意差が見られなかった(被験者 1:  $F(2, 30) = 0.152, p = .86$ ) (被験者 2:  $F(2, 24) = 0.467, p = .63$ )。また、Tukey の多重比較検定を実施して、各実験における質問力平均値の差を検討した。その結果、被験者 2 名共にいずれの比較ペアにおいても統計的な有意差が見られなかった(共に  $p > .05$ )。

図 1、図 2 では、被験者の実験における質問力推移を示している。この図からも、3 回目は 1 回目よりも質問力が高いとは言えない結果となった。

#### 3.3 実験考察

本実験において ChatGPT を用いた質問力向上は見

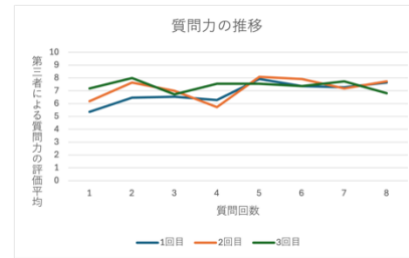


図 1 被験者 1 の実験における質問力推移

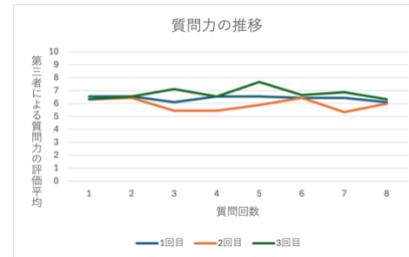


図 2 被験者 2 の実験における質問力推移

られなかった。本実験では、被験者と質問評価者を多く集められなかったこともあり、有意な結果を得づらいう状況にあったことが挙げられる。しかし、被験者は ChatGPT による質問力評価の参考度について、共に全実験で 10 段階中 8 点以上と高く評価していた。また、その後の実験において質問する際に意識したことは、被験者 2 名共に ChatGPT による評価を踏まえた内容であった。これにより、実験の試行回数を重ねることで質問力向上が見込まれる。

### 4. まとめと今後の展望

本実験において、概ね ChatGPT が人間のような回答を出力できたが、その出力された回答と ChatGPT による質問力評価を用いて質問力向上を達成することはできなかった。今後は実験体制を整え、質問力育成までの流れをシステム化することにより、質問力向上を円滑に進められることが期待される。また、相手との関係性や、質問を踏まえた感情の変化を加えることで、より実践に近い質問力を鍛えられると考えられる。その後、本研究の発展として質問力育成における長期的な効果の検証や、感情や文脈を考慮した実践的な質問力トレーニング手法の開発に関する研究を進める予定である。

#### 参考文献

- (1) 道田泰司: “批判的思考研究からメディアリテラシーへの提言”, コンピュータ&エデュケーション, 第 9 巻, pp.54-59 (2000)
- (2) いなかみライフ事務局: “【ノウハウ】インタビューがうまい人の共通点は質問力じゃなかった”, いなかみライフ, <http://inakami.net/blog/intav-3812.html> (2014) (閲覧日 2024-04-29)
- (3) Robert M. French: “The Turing Test: The first 50 years”, Trends in Cognitive Sciences, Vol.4, No.3, pp.115-121 (2000)