

デザイン教育における自己調整学習を支援する
対話エージェントの開発

Development of Dialogue Agent to Support
Self-regulated Learning for Design Education

2024 年 1 月
廣瀬 伸行

論文内容の要旨

本研究の目的は、デザイン教育において学生毎にパーソナライズした自動学習支援手法を確立することである。そのために、近年発展の目覚ましい大規模言語モデルを用いることで、学生個別のアドバイスを生成し、自動的に学習支援する対話エージェントを開発した。具体的には、学生個別の学習状況に基づいて、自己調整学習における学習方略の利用と調整のアドバイスを自動的に提示する方法を提案した。

自己調整学習では、学生が効率よく主体的に学習を進めていくためには、以下に示す3つのプロセスについて、学習方略を改善する循環プロセスを継続的に行うことが求められる。

予見 具体的な学習計画立案

遂行コントロール 実行中の自己モニタリングと学習方略の適宜調整

自己省察 活動後の振り返り

特にデザインの教育では、学生が個別のテーマに取り組む段階になると、主体的に学習活動を進められるスキルを修得していることが重要である。しかし、一部の学生は主体的な学習活動をうまく行えず、支援が必要であることが認識されている。そのため、学習の困難や挫折を防ぐためには、具体的な学習活動への支援が必要である。しかし、教員だけで学生個別のテーマと状況に応じた学習計画や振り返りの支援を行いつつ、個別指導と授業進行を同時に行うことは負担が大きい。そこで、デザインを学ぶ学生の自己調整学習における学習実態を捉え、学習方略の利用や調整を自動的に支援する機構を実現したい。ところが、客観的に学習傾向を捉え、学習傾向に最適な定型的助言や資料を自動提示するだけでは、学生個別の学習実態に沿った支援ができていなかった。

そのために、学生毎にパーソナライズした自動助言による自動学習支援手法を実現するために以下の研究を行った。

第3章 課題遂行段階に対する学習スキルの自動助言

第4章 学生の記述内容から学習スキルの自動評価

第5章 学生個別の記述内容に追従した助言自動生成

第3章では、自己調整学習の課題遂行段階を支援するための、学習スキルを自動助言する

手法を提案した。学習活動中の自己モニタリングと学習方略の利用や調整を促すために、自動的な学習状況の聞き取りとアドバイスによる支援方法について評価した。

評価の結果、学生が目標達成の見込みが低いと回答した場合に限定すると、課題プロセスの経過を把握する学習方略を使用したという自己申告の割合が増加することがわかった。このことから、課題遂行段階を自動支援する手法として一定条件下での可能性が示された。

しかし、学習方略の適宜調整など、大半の学習方略への効果については明らかにすることができなかった。そのため、さらに分析を進め、学習方略の相互関係に基づく改善点を明らかにした。そこで、実験結果の詳細な分析による新しい仮説として、学習内容を得意と感じさせる対話をすることで、学習経過の把握、振り返り、学習方法の工夫などの方略行動を増やすことができるという可能性が得られた。

第4章では、自己調整学習の予見と自己省察の学習スキルを自動評価するための、学生個別の記述内容をループリックを用いて自動評価する手法を提案した。第4章は、第5章のアドバイス自動生成の前提となる。具体的には、学生が記述した学習計画と振り返りの記述内容を自動評価するためにループリックを用いたデータで GPT-3 をファインチューニングした大規模言語モデルを用いた手法を評価した。

評価の結果、教員の手動評価と比較して、記述の具体性については順位相関係数 $+0.815$ 、躓き度については順位相関係数 $+0.876$ であった。このことから、十分に実用に耐える性能と考えられるが、躓き度については経緯を考慮した推定ができない場合がある等の課題も明らかになった。

第5章では、予見と自己省察の助言を自動生成するための、学生個別の記述内容に追従した助言を自動生成するための手法を提案した。予見と自己省察の支援として、学生の学習計画と振り返りの記述内容に追従し、具体性を高めるアドバイスを自動生成して提示することによって学生の学習方略の利用と調整を促す支援手法を提案した。具体的には、学生が記述した学習計画と振り返りの記述内容に加え、テーマやタスク情報、ループリックをプロンプトに含め、大規模言語モデル GPT-4 のゼロショット学習を用いた手法を提案した。そこで、自動生成されたアドバイスが学生にとって理解しやすく受け入れできるものであったかどうかを評価した。

評価の結果、7 件法評価において平均が適合性 5.27、理解性 5.16、実行性 5.0 であった。つまり、全体的には 5= ややそう思うという評価が得られたため、テーマやタスク情報を併用してアドバイスを自動生成する方法の可能性を示した。ただし、テーマを単独で用いたほうが、テーマやタスク情報を併用して用いた場合よりも学生にとって分かりやすいアドバイスができる可能性があることがわかった。その一方、テーマやタスク情報を併用して用いた場合は、学生にとっての分かりやすさが低下するという問題があった。そのため、今後の課題として、学生にわかりやすいアドバイスを生成するために、テーマに対するアドバイスと

タスク情報に対するアドバイスの2つに書き分けて生成する方法を検討することにした。

以上のことから、本研究では、デザイン教育における自己調整学習の自動支援の方法として、以下の3つを明らかにした。

第3章：課題遂行段階に対する学習スキルの自動助言

定形式と決定木を用いた自動的な学習状況の聞き取りとアドバイス手法について、一定条件の下でのみ可能性を示した。

第4章：学生の記述内容から学習スキルの自動評価

GPT-3 のファインチューニングを用いた学習計画と振り返りの自動評価手法について、実用可能な性能を示した。

第5章：学生個別の記述内容に追従した助言自動生成

GPT-4 を用いた学生毎にパーソナライズしたアドバイスの自動生成手法について、一定の可能性を示した。

ただし、第4章と第5章の評価は、学生がアドバイス内容を読んだ結果、適合性、理解性、実行性があると評価したに留まる。そのため、今後の課題として、自動生成されたアドバイスによって、実際に学生が記述内容を改善したかどうか、また、記述内容を改善することで自己調整学習のスキルが身につくかどうかについて実証的な評価を行う。

abstract

The purpose of this study is to establish an automatic learning support method personalized for each student in design education. For this purpose, we developed a dialogue agent that automatically supports learning by generating personalized advice for each student using a large-scale language model, which has been remarkably developed in recent years. Specifically, we proposed a method for automatically presenting advice on the use and adjustment of learning strategies in self-regulated learning based on the learning status of individual students.

In self-regulated learning, in order for students to learn efficiently and independently, a continuous cyclical process of improving learning strategies is required for the three processes listed below.

Foresight Planning a specific learning plan

Execution control Self-monitoring during execution and adjusting the learning strategy as necessary

Self-reflection Reflection after the activity

Especially in design education, it is important for students to acquire the skills to independently carry out learning activities when they reach the stage of working on individual themes. However, it has been recognized that some students are not able to engage well in independent learning activities and need support. Therefore, specific support for learning activities is necessary to prevent learning difficulties and setbacks. However, it is burdensome for teachers alone to provide support for learning plans and reflections based on the themes and situations of individual students, while simultaneously providing individual guidance and facilitating classes. Therefore, we aim to realize a mechanism that automatically supports the use and adjustment of learning strategies by capturing the actual learning status of design students in their self-regulated learning. However, simply by objectively capturing their learning tendencies and automatically presenting the most appropriate standardized advice and materials, we have not been

able to support students in accordance with their individual learning situations.

To this end, the following studies were conducted to realize qualitative and student-specific automated advising.

Chapter 3 : Automatic advice on learning skills for learning execution phase

Chapter 4 : Automatic evaluation of learning skills based on students' descriptions

Chapter 5 : Automatic generation of advice that follows individual student descriptions

In Chapter 3, we propose a method for automated advising of learning skills for learning execution phase in order to support the learning execution phase of self-regulated learning. We evaluated a method to support learning by automatically listening to learning status and providing advice to encourage self-monitoring and the use and adjustment of learning strategies during learning activities.

The results of the evaluation revealed that the percentage of students self-reporting that they used learning strategies to track the progress of the assignment process increased when limited to when students responded "rather unlikely" to the prospect of achieving the goal. However, the effects on most of the learning strategies could not be clarified. Therefore, further analysis of the results yielded points for improvement based on the interrelationships among the learning strategies. The new hypothesis, based on a detailed analysis of the experimental results, posits that dialogue that makes students feel proficient in what they are learning will increase their strategic behavior, such as understanding the progress of their learning, reflecting on their learning, and devising new learning methods.

In Chapter 4, we propose a method for automatically assessing the learning skills of anticipation and self-reflection of self-regulated learning from students' descriptions. Chapter 4 is the premise for the automatic generation of advice in Chapter 5. The evaluation results showed that the rank correlation coefficient was +0.815 for the specificity of the description and +0.876 for the degree of stumbling compared to the teacher's manual evaluation. These results indicate that the performance of the model is sufficient for practical use. However, there are some issues, such as the fact that the estimation of the stumbling degree cannot be done in some cases considering the history of the stumbling degree.

In Chapter 5, we propose a method for automatic generation of foresight and self-reflection advice that follows individual student descriptions. To support foresight and self-reflection, we proposed a method for automatically generating advice that encourages

students to use and adjust their learning strategies by analyzing their learning plans and reflections and increasing their concreteness. By applying the rubric-based automatic assessment validated in Chapter 4, the method automatically generates advice based on the scaffolding and scaffolding removal phases. Specifically, we propose the following: a method using zero-shot learning of the large-scale language model GPT-4, which includes themes, task information, and rubrics in the prompts, in addition to the student's description of the learning plan and reflections. We evaluated whether the automatically generated advice was understandable and acceptable to students.

Analysis of the student evaluations revealed a mean of 5.27 for conformance, 5.16 for comprehensibility, and 5.0 for performability on a 7-point scale.

However, it was found that the use of themes alone may produce advice that is easier for students to understand than the use of themes and task information together. On the other hand, when themes and task information were used together, the clarity of the advice for students decreased.

Therefore, as a future task, we plan to investigate a method of generating advice based on themes and task information separately, in order to provide students with easy-to-understand advice.

Based on the above, this study identified the following three methods for automatic support of self-regulated learning in design education.

Chapter 3: A method for automated advising of learning skills for learning execution process

A method of automatic listening and advising on learning status using formal sentences and decision trees was shown to be feasible only under certain conditions.

Chapter 4: A method for automatic evaluation of learning skills based on students' descriptions

A method for automatic evaluation of learning plans and reflections using GPT-3's fine tuning was demonstrated to have practical performance.

Chapter 5: A method for automatic generation of advice that follows individual student descriptions

A method for automatic generation of personalized advising for each student using GPT-4 was shown to have certain potential.

Regarding the support for goal setting and strategy planning, the possibility of automatic generation of advice based on automatic evaluation using the GPT-4 in Chapter 5 was demonstrated.

self-reflection

For the support of self-evaluation and attribution of causes, the possibility of automatic evaluation based on rubrics was shown by the automatic evaluation of learning plans and reflections using the GPT-3 in Chapter 4. For the support of self-assessment and attribution of cause, we showed the possibility of the automatic generation of advice based on the automatic assessment using the GPT-4 in Chapter 5.

However, the evaluation of the automatically generated advice is limited to the results of the student's reading of the advice content and assessing its suitability, understandability, and actionability. Therefore, it is not clear whether the automatically generated advice actually improved the students' writing, or whether the students acquired self-regulated learning skills by improving their writing. Therefore, as a future task, we plan to empirically evaluate whether the students actually improve their writing through the automatically generated advice and whether they acquire self-regulated learning skills by improving their writing.

目次

第 1 章	序論	1
1.1	研究の背景	1
1.2	研究の目的	3
1.3	本論文の構成	4
第 2 章	関連研究	7
2.1	序言	7
2.2	自己調整学習における支援の自動化研究	7
2.3	学習計画と振り返りの自動評価研究	10
2.4	学習支援におけるアドバイスの自動提示研究	12
2.5	結言	14
第 3 章	定形文による自動的な学習状況の聞き取りとアドバイス	19
3.1	序言 自己調整学習スキルの支援	19
3.2	チャットボットの機能	21
3.3	対話が破綻しない自動的な対話機能（機能 1）	24
3.4	自動的に学習状況の聞き取りを行う対話機能（機能 2）	26
3.5	自動的な学習方略のアドバイス提示機能（機能 3）	27
3.6	実験設定	31
3.7	結果	36
3.8	考察	42
3.9	結言	49
第 4 章	GPT-3 を用いた学習計画と振り返りの自動評価	69
4.1	序言学習計画と振り返り記述の評価，足場かけ，足場はずし	69
4.2	振り返りの記述の具体性評価と躓き度評価	75
4.3	具体性評価のためのループリック	78
4.4	自動評価のために GPT-3 を用いた理由	80

4.5	実験	80
4.6	評価方法	83
4.7	結果	85
4.8	考察	89
4.9	結言	95
第 5 章	GPT-4 を用いた自動評価に基づく学生個別アドバイスの自動生成	107
5.1	序言 自動評価に基づく学生個別アドバイス自動生成	107
5.2	アドバイス自動生成のための GPT-4 とプロンプトの設計	113
5.3	アドバイス自動生成の予備実験	115
5.4	学生個別のテーマとタスク情報を追加したアドバイス自動生成の評価 . . .	123
5.5	実験	125
5.6	結果	127
5.7	考察	141
5.8	結言	146
第 6 章	本研究の包括的な考察	149
第 7 章	結論	157
7.1	本研究のまとめ	157
7.2	本研究の貢献	161
7.3	今後の課題	162
参考文献		165
謝辞		168
本論文に関する研究業績		169
付録		171
付録 A	予備実験で使用したアドバイスの雛形文	171
付録 B	システムメッセージ用ループリックとアドバイス方針	175
付録 C	予備実験で使用したプロンプトの実例	179
付録 D	学習計画や振り返りを対話的に聞き取る機構	181
付録 E	プロンプトの実例	182

図目次

1.1	本論文の各章の関連図	6
3.1	デザイン学習用の Web アプリの画面構成 [1] を著者が加工して使用	22
3.2	チャットボットの概要図 [1] を著者が加工して使用	23
3.3	自己調整学習のサイクル・モデル [2] p.11, 図 1 を元に著者が手を加え作成した	24
3.4	決定木ベースのチャットボット（一部抜粋）	25
3.5	テキストデータの設定例 [1] を著者が加工して使用	26
3.6	対話の流れの一部 イメージ例 [1] を著者が加工して使用	27
3.7	決定木の編集過程 [1]	28
3.8	自動対話の流れを構成する決定木 [1] を著者が加工して使用	29
3.9	決定木の流れ一部	54
3.10	チャットボットと学生の対話例 [1] を著者が加工して使用	55
3.11	統制群ごとの実験タスク [1] を著者が加工して使用	56
3.12	自動アンケート機能	56
3.13	チャットボットとの対話の流れの自然さ [1] を著者が加工して使用	59
3.14	振り返り記録のスコア平均	60
3.15	学習方略の分析方法	62
3.16	課題プロセス把握方略（SRL06）の統制群比較 [1]	62
3.17	課題プロセス把握方略使用の初回 - 最終回比較 [1]	63
3.18	方略使用増減と対話パターン I02 出現回数の散布図 [1]	64
3.19	方略使用増減と対話パターン I02 と交絡因子	65
3.20	対話パターン I02 出現数の算出方法	65
3.21	方略増減値の算出方法	66
4.1	学習支援の概要	71
4.2	開発中の学習支援画面	73
4.3	学習支援の自動評価したい部分	74

4.4	自動評価用モデルの作成	75
4.5	授業回ごとの振り返りと計画のサイクル	82
4.6	足場はずしレベルごとの入力フォーム（上から Lv.3, 2, 1）	97
4.7	全ての具体性評価に関する手動評価と自動評価	98
4.8	項目ごとの具体性評価に関する手動評価と自動評価	99
4.9	足場はずし Lv に関する手動評価と自動評価	102
4.10	躓き度に関する手動評価と自動評価	103
5.1	評価を 1 段階向上させる足場かけと足場はずし	109
5.2	自動的なアドバイス生成による自動学習支援システム	110
5.3	学生個別のテーマ入力画面	112
5.4	学生個別のタスク進捗管理画面	112
5.5	学習計画, 振り返り確認画面	114
5.6	予備実験：ループリック評価別の頻度	118
5.7	予備実験：回答者のループリック評価別の頻度	119
5.8	アドバイスと雛形文との Embedding 間のコサイン類似度平均の比較	120
5.9	全体のループリック評価別頻度	128
5.10	回答者のループリック評価別頻度	129
5.11	全体のアドバイス種類別平均値の比較	130
5.12	記述内容の具体性が低い場合のアドバイス種類別平均値の比較	133
5.13	記述内容の具体性が高い場合のアドバイス種類別平均値の比較	136
5.14	実行性評価について記述内容具体性高低の比較	138
5.15	理解性評価について記述内容具体性高低の比較	138
6.1	自己調整学習の循環モデル [3]	150
D.1	学習計画や振り返りを対話的に聞き取る機構	186

表目次

2.1	学習支援システム研究における貢献と関連研究との関係	15
3.1	聞き取りとアドバイスの事例	51
3.2	事例データの一部 [1]	52
3.3	事例データの変数 [1]	52
3.4	方略の工夫アドバイスの変数	53
3.5	学習方略の計画入力での質問項目	57
3.6	振り返り記録入力での質問項目	58
3.7	実験タスクと実験アンケート実施人数	59
3.8	対話アンケートと対話パターン ID の相関	60
3.9	対話パターン（抜粋）	61
3.10	特に不自然さを感じたところの集計	61
3.11	条件 1: 統制群間比較と条件 2: 実験群最終初回比較	66
3.12	条件 3: 対話パターンとアンケート項目の相関	67
3.13	実験群の SRL 項目間の相関	67
3.14	振り返り記録の記述例	68
3.15	アドバイス有無と記述有無の集計	68
3.16	記述内容に変化があったと判定できた事例	68
3.17	自己効力感と学習方略の相関表 [1] を著者が修正して使用	68
4.1	教員の手動による具体性評価の例（焦点振り返り）	76
4.2	トレーニングデータの例	77
4.3	手動評価された評価項目と評価値の度数分布	84
4.4	交差検証用グループの設定	85
4.5	項目ごとの具体性評価誤差	87
4.6	手動と自動の具体性評価が一致した記述内容	100
4.7	手動と自動の具体性評価に誤差のあった記述内容	101
4.8	想定外の記述内容への自動評価事例	101

4.9	手動評価と自動評価に関する足場はずし Lv. 件数	102
4.10	手動評価と自動評価に関する足場はずし Lv. 誤差	103
4.11	躓き度評価の誤差	104
4.12	手動と自動による躓き度評価が一致した記述例	104
4.13	手動と自動による躓き度評価に誤差（手動 4, 自動 2）のあった記述例 . .	105
5.1	予備実験：アドバイスと雛形文のコサイン類似度	120
5.2	予備実験：目標計画ループリック評価ごとの自動生成アドバイス	121
5.3	予備実験：教員手書きアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較	122
5.4	予備実験：GPT-3 のアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較	122
5.5	予備実験：（具体度が高いデータ）教員手書きアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較	123
5.6	予備実験：（具体度が高いデータ）GPT-3 のアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較	123
5.7	実験結果に用いた回答件数の内訳	127
5.8	全体のアドバイス種類別平均値と有意な差	131
5.9	記述項目別のアドバイス種類別平均値と有意な差	132
5.10	記述内容の具体性が低い場合のアドバイス種類別平均値	134
5.11	記述内容の具体性が低い場合の記述項目別のアドバイス種類別平均値と有意な差	135
5.12	記述内容の具体性が高い場合のアドバイス種類別平均値	137

第 1 章

序論

本研究の目的は、デザイン教育において学生ごとにパーソナライズした自動学習支援手法を確立することである。そのために、近年発展が目覚ましい大規模言語モデルを用いることで、学生個別のアドバイスを生成し、自動的に学習支援する対話エージェントの開発を目指す。

第1章では、デザイン教育における自己調整学習の重要性を背景に、学生が主体的な学習活動を行えるよう自動的な支援方法の確立を目的としていることを述べる。この目的を達成するために、本研究では学生個別の学習実態を自動的に捉える事が可能な学習支援システムの開発を行い、その有効性を検証する。

1.1 研究の背景

学生ごとに最適な学習方法は異なる。主体的な学習活動では、学生個別の学習状況や特性に応じた個別の学習方法などの実態を学生が自ら考えて調整する必要がある。学習支援においては、このような学生個別の学習実態を捉えて行う必要がある。

従来の自動的な学習支援は、学習時間、小テスト点数、学習行動の数、記述内容の単語頻度などから学習傾向を測る定量的なアプローチに基づいている。定量的なアプローチでは、学生個別の学習状況や傾向を客観的に捉えることができ、学習活動を最適化する指針となる助言を示すことができるが、学習内容の実態を捉えきれておらず、提示される助言も汎化されるため学生個別の学習実態と離れがちになる。

学生個別の学習実態を捉えるためには、何を目標に、何に着目し、どんな方法という内容を捉える定性的かつ学生個別のアプローチを行う必要がある。学生個別の学習支援を行うためには、学生個別の学習実態を捉える定性的なアプローチも重要である。

したがって、本研究は、定性的かつ学生個別のアプローチによる自動助言手法の実現を目指す。

近年発展が目覚ましい大規模言語モデル（LLM）によって、学生の記述したテキストの

内容を自動的に解釈させて扱うことができるようになってきている。そこで、大規模言語モデルを用いることで、学生個別の学習内容の実態を自動的に捉えた学習スキルの助言を実現する。

なお、本研究における定性的なアプローチとは、学生個別の学習実態について、何を目標に、何に着目し、どんな方法という学習内容を捉えて扱うアプローチであり、定量的なアプローチとは、学生個別の学習実態について、学習時間、小テスト点数、学習行動の数、記述内容の単語頻度などの学習傾向を捉えて扱うアプローチである。

1.1.1 デザイン学習と自己調整学習の支援

デザイン学習と主体的な学習活動の関係性を示し、主体的な学習活動の学習支援について述べる。

デザイン学習では、独創性や新規性が重要視される。このため、学生には効率的なデザインプロセスを進めるための主体的な学習活動が求められる。具体的には、問題の発見、課題の定義、解決方法の検討、実現方法の提案、具体化のための制作と改善の実行のプロセスにおいて、テーマへの理解、提案内容の精査、提案内容を可視化する表現手法の研鑽が求められる。主体的な学習活動ができる学生は、デザインに必要なスキルを効率よく獲得していくことができる。一方、主体的な学習活動に慣れていない学生は、自身の学習活動を効果的に制御できないことがある。そのため、デザイン学習における主体的な活動が滞り、学習の困難や中断が生じる可能性がある。したがって、デザイン教育における主体的な学習活動の支援は重要な課題である。

主体的な学習活動については、学生が自分の学習活動を認知的、能動的、行動的にどのように制御するのかに焦点を当てた自己調整学習 [4] (Self regulated learning : SRL) の研究が行われている。そして、学生が自身の学習の見通しを立て、学習を遂行し、自身の学習状況を把握して、学習方略を調整していくのに役立つとされている [5]。自己調整学習の循環モデル [3] は、自己調整を効率よく運用するための循環プロセスとして示されている。

自己調整学習は、デザイン教育に求められる主体的な学習活動を効率よく運用するためのプロセスとして当てはめることが可能である。なぜならば、自己調整学習の焦点が個人の能力よりも学習方略の調整に焦点があるため、デザイン活動の問題解決の探索に方略の調整を重ねる活動と親和性が見込めるためである。

自己調整学習は、学生が主体的に目標、焦点、方略を計画、遂行し、調整を重ねていくプロセスである。そのために、自己調整学習がうまくできない学生に行う支援では、学生個別の目標、焦点、方略について、どのように計画、遂行し、調整しているのかという学習実態を捉えた個別の支援が求められる。

しかし、教員 1 名では学生数が多くなると学生個別の学習実態を細かく捉えて支援を行うことが困難になる。したがって、学生個別の学習実態に沿った個別支援の自動化が課題となる。

1.1.2 主体的な学習を支援する学習支援システムの要求

学生の学習活動を補助するために学習支援システムの研究が行われている。これらのシステムは、コンピューターやインターネットなどのテクノロジーを活用して、学生の学習活動を支援する。そのため、利用実態としては教材の配布やレポート、小テストの実施、そして、学生と教員間の指導、さらには、学生間の学習活動における利便性を高める目的で用いられることが多い。

一方、自己調整学習のスキル向上を目指す学習支援システムの研究では、学生の学習活動や内容を支援することで学習の質への効果を高める研究がなされている。中でも、学生の学習状況の分析に基づいて自動的な支援に関する研究がある [6, 7, 8]。教員の教育活動や学生の学習活動について、学習状況を定量化して分析し、その結果から最適化した支援を行う方法について提案されている。具体的には、学生の学習状況を学習時間、成績、学習行動の数、記述内容の単語頻度などの定量化に基づき、学生個別の学習傾向を客観的に捉えることができる。そして、学習を最適化するための資料の提示や、課題の提示、定形文アドバイス、学生が入力した学習状況の記録などを提示する手法が示されている。

しかし、既定の答えや解法が定まっていない学習内容では、定量化しただけでは、学生個別の学習実態から離れがちになる。そのため、学生は、学習支援として提示された内容を自身の状況に適用することが困難な場合がある。なぜならば、自己調整学習がうまくできない学生は自身の学習実態を自身で把握できていないことが多いためである。そこで、課題となるのは、学生が認知する学習実態に基づいて、それを広げる、もしくは深めることと、そこに、適用できる学習方法を組み合わせる助言の提示である。そのためには、学習実態としての学習目標、焦点、方略の計画、遂行、調整の中身を扱うことのできる定性的なアプローチが必要となる。

1.2 研究の目的

本研究の主要な目的は、デザイン教育などに見られる学生個別に答えや遂行過程の定まらない学生個別の学習活動を支援するとき、何を目標に、何に着目し、どんな方法で取り組むのかという学生個別の学習実態に基づいた自動支援手法を確立することである。そのために、従来の学習時間、小テスト点数、学習行動の数、記述内容の単語頻度などの定量的かつ定型的なアプローチだけでは困難であった何を目標に、何に着目し、どんな方法の学習なの

かという実態を捉えた自動支援の実現のために、定性的かつ学生個別なアプローチによる自動助言手法を明らかにする。

また、デザイン教育のような問題解決の為に様々手法を試行錯誤する既定の答えや解法が定まっていない探索型の学習活動を対象として、自己調整学習に基づいた定性的かつ学生個別なアプローチによる自動助言手法を評価する。

自己調整学習の循環モデルを効果的に活用するためには、学生が学習方略を具体的に利用し、調整できるように支援する必要がある。支援では、学生が具体的に学習方略を利用や調整できるよう、学生個別の学習状況に基づいた具体的な改善アドバイスを自動的に提示する必要がある。そのためには、学生個別の学習状況と学習方略の実態を自動的に捉える方法と、評価した学習状況と学習方略の利用状況に応じて学習方略の利用を改善するアドバイスを自動的に提示する定性的かつ学生個別なアプローチ方法が必要になる。

学生個別の学習実態を自動的に捉える定性的かつ学生個別なアプローチによる学習支援では、学生個別の学習活動の実態として、何を目標に、何に着眼し、どんな方法で学習しているのかを自動的に捉えて評価し、学生個別の学習実態に追従しつつ、デザイン教育に適した学習方略の利用や調整を促すための具体的なアドバイスを自動提供する手法が必要である。

そのために、以下の 3 つの定性的かつ学生個別な自動助言手法を提案した。

1. 課題遂行段階における自動助言を行う、定形文による自動的な学習状況の聞き取りとアドバイス
2. 学生の記述内容から学習スキルの自動評価を行う、GPT-3 を用いた学習計画と振り返りの自動評価
3. 学生個別の記述内容に追従した助言を自動生成を行う、GPT-4 を用いた自動評価に基づく学生個別アドバイスの自動生成

1.3 本論文の構成

本論文の構成を以下に示す。まず、第2章では、本研究の背景となる関連研究について述べる。また、自己調整学習における定性的かつ学生個別なアプローチによる自動支援の必要性を検討する。そして、学習支援システムの研究における本研究の貢献について検討する。

第3章、第4章、第5章の各章は自己調整学習の循環モデルの支援に対して、定性的かつ学生個別なアプローチによる自動支援を段階的に深化させる一連のプロセスとして構成されている。

第3章では、課題遂行段階における学習スキルの自動助言を実現するために、定形文と決定木を用いた自動的な学習状況の聞き取りとアドバイス手法により、自己調整学習の循環モ

デルの遂行コントロールにおいて学生の自己調整学習が促進されたかどうかを評価した。

第4章では、学生の記述内容から学習スキルの自動評価を実現するために、GPT-3 を用いた学習計画と振り返りの自動評価手法により、予見と自己省察における学生の記述内容から学習スキルを自動評価したときの性能を評価した。

第5章では、学生個別の記述内容に追従した助言の自動生成を実現するために、GPT-4 を用いた自動評価に基づく学生個別アドバイスの自動生成手法により、学生個別の記述内容に追従したアドバイスを自動生成し、自動生成したアドバイスに対して学生評価の結果を分析した。

第6章では、本研究の目的に対する結果と限界、今後の課題について包括的な考察を行う。最後に第7章では、今後の課題、そして今後の展望についてまとめる。

図1.1に本論文の構成を各章の関連図として示す。

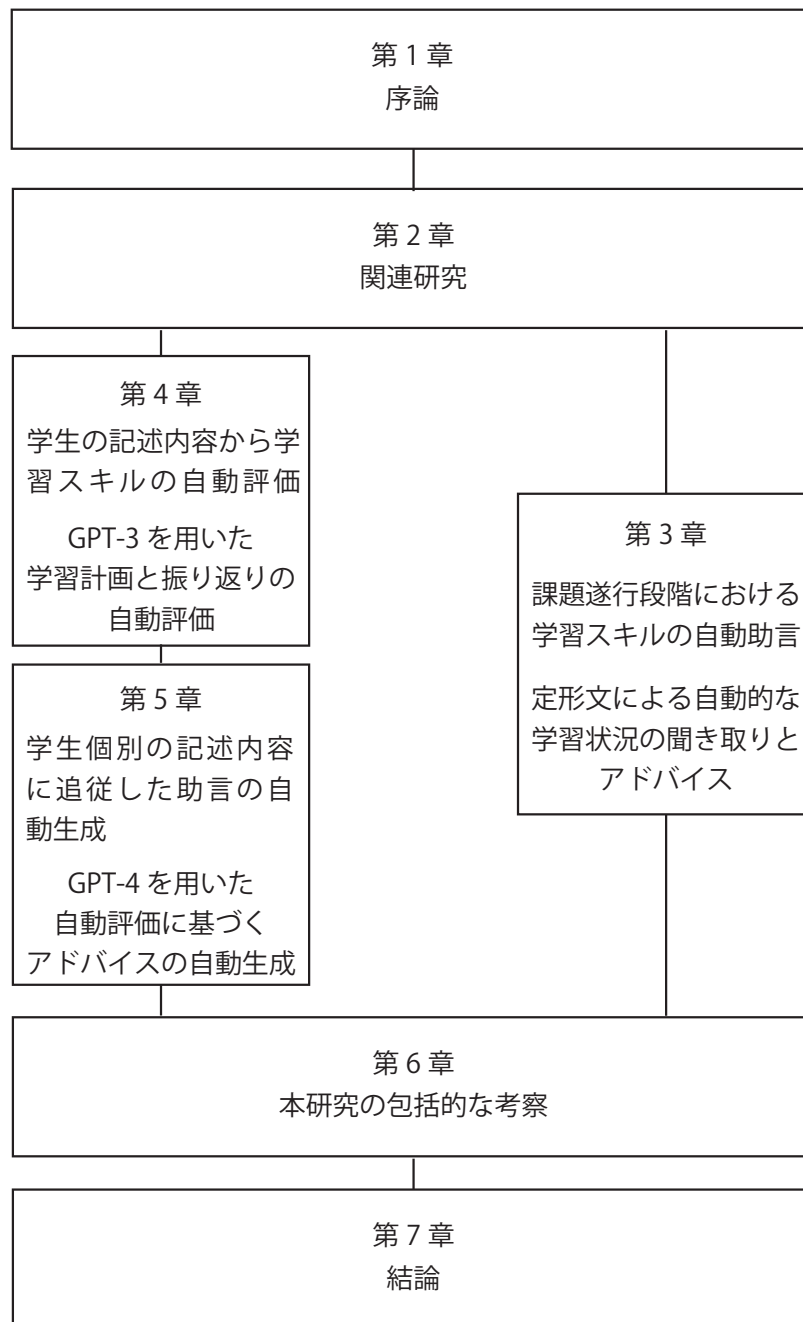


図1.1: 本論文の各章の関連図

第 2 章

関連研究

2.1 序言

本章では、本研究の背景となる関連研究について述べる。

まず、2.2節では、本研究の研究対象であるデザイン教育における自己調整学習と自動的な支援手法の関連研究について述べる。次に、2.3節で、自動的な支援において、学生個別かつ品質の揃ったアドバイスを実現するために必要となる学生個別の学習振り返りや学習計画の自動評価手法の関連研究について述べた後に、2.4節で、自動評価に基づいて、学生個別かつ品質の揃ったアドバイスを自動生成する手法の関連研究について述べる。最後に、節で、本章の各節で述べた関連研究との差異と学習支援システム研究への貢献について表に示しながら言及し、本章をまとめる。

2.2 自己調整学習における支援の自動化研究

2.2節では、自己調整学習における自動支援について論じる。

デザイン教育では、主体的な学習活動が求められる。なぜならば、与えられたテーマに基づいて、学生は個別の提案が求められ、提案内容を具体化していく過程に、テーマへの理解、提案内容の精査、提案内容を可視化する表現手法の研鑽が求められるためである。主体的な学習活動ができる学生は、効率的にこれらのデザインスキルを獲得していくことができる。しかし、主体的な学習活動に苦手意識を持つ学生は、学習の躓きや停滞を経験すると、学習を諦めやすいという傾向がある。そのため主体的な学習方略を支援することが重要である。

主体的な学習活動の重要性は、デザイン教育に限らない。主体的な学習活動を実施できているかの確認は、個別に教員が把握する必要があるが、現実の教育現場では時間が限られている。そのため、人の教員による支援を代行する自動支援に関する研究が進められている。なかでも、自己調整学習の理論を用いた自動支援手法は、主体的な学習活動を支援する研究課題として注目されている。自己調整学習の循環モデル [3] は、デザイン学習の学習過程にも

適用できる。

自己調整学習 [5] は、学生が自身の学習の見通しを立て、学習を遂行し、自身の学習状況を自己評価をして、学習方略を調整していくのに役立つ。そして、自己調整学習の循環モデルは、学習方略を効率よく運用するための指針を提供する。具体的には、次のような言及がなされている。予見の目標設定では、学生が自身の学習目標を明確にし、その達成のために具体的な学習計画を支援する。遂行コントロールの自己モニタリングでは、学生が自分の進捗を定期的にチェックし、必要に応じて学習計画を調整することを促す。また、自己省察では、学生が自分の学習結果を評価し、成功や失敗の原因を理解することで、次の学習活動に活かすことができる。これらのプロセスを通じて、学生は内発的な興味や価値を見出し、自己効力感を高めることができるため、学習への動機づけが強化される [4, 2, 9]。

また、自己調整学習では、主体的な学習に不慣れな学生は、自律的に自身の学習状況を把握する行動や学習を計画する方略の調整について行動に移すことが困難であるとされている [2]。

つまり、自己調整学習を支援するためには、学生個別の学習状況や学習方法の実態を捉え、個別の学習目的、焦点、方略の状況に沿って、個別の学習スキルに応じた支援を行う必要がある。そのため、学習時間や成績、学習行動数や記述単語の頻度などによって学習傾向を捉える定量的なアプローチのみでは、学生個別に何を目指し、着目し、どういう方法で学習しているのかという学習実態を捉えきれない。そこで、学習実態を捉える定性的かつ学生個別なアプローチによる自動支援の実現が課題となっている。

自己調整学習の循環モデル [3] に基づいた学習支援システムの先行研究 [6, 10] は、学生の研究活動を対象として自己調整学習能力の獲得を目的としている。学生の研究活動とデザイン学習に見られる学生が試行錯誤を重ね独自の提案を導く研究過程と、デザイン学習において学生が自身のアイデアを表現する制作活動と類似しているため、自己調整学習をデザイン学習にも応用できる。

従来手法では、初歩の自己調整者に行う学習方略目標の階層化、自己モニタリングと自己評価のための自己記録の提示手法 [6] は、定性的なアプローチとして参考にできる。しかし、学生による自由な文字入力のため、記録すべき内容の適切な支援や適切な学習方略使用について支援を行わなければ、学生が自身で把握しておくべき学習状況や調整すべき学習方略を見落としてしまう問題がある。

そこで、システムがあらかじめ定まった学習方略の視点で聞取りを行うことで学生自身の目標に対する学習状態を学生自身に比較させる手法に着目した自動聞取り手法 [10] が参考になる。しかし、聞取り結果に基づいた具体的なアドバイスを行うなどの方法については今後の課題とされていた。

学習支援にチャットボットを用いた事例を研究した先行研究 [11] は、チャットボットによる学生個別の学習支援について、学生の学習状況を学生の解答パターンや対話結果から判別して個別に、最適な学習資料を提示するなどの支援手法が示されている。しかし、遂行段階において行われる学習スキルに関する聞き取りではなく、個々の学生が異なる内容に取り組み、かつ、一律の答えや、解法が定まらない学習において学習方略を促す手法については十分な検証がされていない。

そこで、遂行段階において必要となる効率的な自動聞き取りと助言の提示手法として、決定木学習に基づいたチャットボットによる効率的な聞き取りから結論を導く手法 [12] が応用できる。先行研究 [12] は、チャットボットが効率よく結論を導く決定木に沿った自動的な聞き取りを行い、決定木によって判断された結果を自動的に選択して提示する手法を示している。しかし、この研究は学習支援における自動的な助言を対象としていない。以上の関連研究では、以下 2 つの要件すべてを満たす方法が示されていなかった。

1. 学生個別に自己モニタリングを支援する方法
2. 学生個別の学習方略の状況に基づいて、学習方略の利用や調整のアドバイスを自動的に選択して提示する方法

自己調整学習では、学生は、自己モニタリングや振り返りによって、自身の学習状況を把握しながら、学習方略を調整することが望ましいとされている。つまり、学生が自身の学習状況を把握する支援と、学習方略の使用や調整の支援は関連して扱われることが望ましいと言える。しかし、自己調整学習の循環モデルにおける遂行コントロールのタイミングでの自己モニタリングを支援する自動的な方法については明らかにされていない。

筆者らは、これまでに分析途中の研究成果 [1] として、上記 (1) と (2) の要件を実現する機能を備えたチャットボットを試作して実験を行った。その結果、対話の流れの自然さと一部の方略利用の増加の示唆、今後の課題として学生の自信と学習方略との関係について可能性を示した [1]。

第3章は、分析途中の研究成果 [1] を含み、実験結果を精査して分析を進めたものである。方略利用の結果を精査し、自己モニタリングと方略の関係性に分析を加えることで、学習方略間の関係から着目できる結果について考察を加えた。

なお、第3章では、学生が学習方略を模倣する初期段階に焦点を当てているため、一般的に自己調整学習の学習支援システムで用いられる足場はずしを想定していない。足場かけと足場はずしの機構については、今後の課題とした。

第3章は以下の点で関連研究と異なる。学生のデザイン学習における自己調整学習の利用を促進するために、自己調整学習の循環モデル [3] の遂行コントロールのタイミングに対す

る自動的な支援方法として、学習実態を捉える定性的かつ学生個別なアプローチによる自動支援として、チャットボットを用いた対話型の自己モニタリングの自動支援と、学習方略の利用や調整のアドバイスを自動提示する方法を同時に満たす点である。

第3章の貢献は以下のとおりである。学習支援システムの研究における学習実態を捉える定性的かつ学生個別なアプローチによる自動支援として、課題遂行段階における自動助言の実現に貢献する。

第3章は、課題遂行段階における自動助言を実現するために、教員が想定した定形文の編集による学生への自動的な学習状況の聞取りとアドバイスの提示手法を用いた。提案手法は、一定の条件では、課題プロセスの経過を把握する学習方略の割合を増やすことができたことを示した。

しかし、教員が想定したデータに基づく決定木と定形文だけでは、実際の学生の多様さに対応することが困難であったことも示された。このことから、課題遂行段階における自動助言の実現への貢献は限定的であった。

2.3 学習計画と振り返りの自動評価研究

2.3節では、学習計画と振り返りの自動評価手法について論じる。

従来手法における学習スキルの分析手法は、学習行動の定量化を行うことで自動評価を実現している。例えば、学習の振り返りの記述内容を評価して定量化する手法として、ファイルサイズに着目した事例がある [13]。ファイルサイズが定量化の指標となりうる可能性について、学生が手書した振り返りの記述内容を機械的に読み取ったときのファイルサイズと学習成果評価との関係と傾向から明らかにしている。しかし、文字数やファイルサイズなどの指標を用いる従来手法は、事例の中でも説明されているように、学習成果との関係性が示されたのみで質的な因果関係が明らかでない。そのため、本研究が必要とする記述内容の具体性や質によって学習方略の利用状況を評価するのに適した手法ではない。

また、学生のノート記録内の単語数と教員が提示した単語や単語数との比較結果を特徴量として、教員によるノート評定との関係からテストの得点を予測する可能性を示した事例がある [14]。この事例では、学生のノート記録に出現する単語や単語数に着目し、教員の付けたノート評定との関係から推定したテストの得点として定量化されている。しかし、学生の記述内容と教員が提示する内容を解析して比較を特徴量として評価した結果を成績の予測に用いているが、学生個別の多様な学習実態の評価については扱われていない。

また、学生の記述内容と教員が提示する内容を解析して比較を特徴量として評価するための事前事後のデータ準備を授業回ごとに行う手間の負担が課題となる。

このように、自動的に学生個別の学習実態を捉えるための評価について、定量的なアプロ

一チだけでは課題がある．一方，学生個別の学習状況の質を捉える手法について，自然言語処理の発達と大規模言語モデル（LLM）によって可能性が高まっている．

例えば，機械学習モデルを作成して学生の振り返り内容を分類する手法を示した事例がある [15]．自然言語処理の手法を用いたテキストデータの分類モデルを再学習し，学生の振り返りの記述内容から学生の振り返り状況を分類することで，学生の自己評価と教員の評価を支援する手法を示している．学生の振り返り記述内容と分類を再学習したモデルを用いる評価手法が参考になれる．しかし，自己調整学習の学習スキルを学生個別の学習実態から自動評価するためには，複数のルーブリック毎に学習スキルの段階に対応した自動分類が必要となる．

そこで，第4章では，振り返りと計画の記述内容に見られる学生個別の学習スキルの段階を自動評価するために，振り返りと計画の記述内容と教員が評価した具体性と躓き度の5段階の評価値をトレーニングデータに用いて再学習したモデルによる自動分類手法を提案した．

第4章は，学生の記述内容から学習スキルの自動評価を実現するために，機械学習モデルを作成して学生の振り返り内容を分類する手法を示した事例 [15] を応用した．第4章は以下の点で関連研究と異なる．記述項目ごとの学習方略について具体的な言及が見られるかどうかのルーブリックに基づいた自動的な評価について，教員の人手による評価結果を詳細に比較することによって明らかにした．具体的には，以下の点が異なる．

1. 記述項目ごとに学習スキルの段階に対応したルーブリックについて自動評価を行ったこと
2. 教員の評価を模倣する自動評価に大規模言語モデル GPT-3 をファインチューニングした手法を用いたこと
3. 自動評価の結果と教員の人手による評価結果について詳細な比較を行ったこと

第4章の貢献は以下のとおりである．学習支援システムの研究における学習スキルの自動評価を行うために，大規模言語モデル GPT-3 をファインチューニングを用いた手法の有効性を示す．これにより，学習支援システムの研究における，学生個別の記述内容から学習スキルの段階を自動評価する手法の実現に貢献する．

第4章の提案手法は，学生の学習計画や振り返りの記述内容を自動評価するとき，学生の記述内容を自動評価するために，教員が人手で行ったルーブリック評価データを用いて大規模言語モデル GPT-3 をファインチューニングした．この手法は，教員が人手で行ったルーブリック評価と比較して十分に実用に耐える性能であることを示した．一方，学生の状況や経緯を含めた評価やトレーニングデータに含まれない記述内容の評価については，改善の余

地があることを明らかにした。このことから、学生の記述内容から学習スキルの自動評価の実現に一定の貢献をすることができたと考える。

2.4 学習支援におけるアドバイスの自動提示研究

2.4節では、学習計画と振り返りのアドバイス自動生成手法について論じる。

従来手法における自動化した定量的かつ定型的な助言提示では、定量化した学習データの分析に基づいて、学習行動を最適化する量データの可視化や、分析状況にあてはまる学習資料やある程度汎化した助言情報を自動提示することができる。例えば、学習計画や振り返りの支援に関連する先行研究では、学生が行ったテストの評価や時間などの分析に基づいて、次の学習計画に必要な支援情報を学生個別に提示する手法 [16] や、学生個別の学習活動の結果や過程をグラフ化することで、学習の振り返りを支援する手法 [8] が報告されている。これらは、学習行動を自動的に定量化することによって自動的に学習支援を実現している。例えば学習時間の最適化、学習行動の量の最適化などの自動助言を達成する。

しかし、成績や答えが定まらない学習内容を対象とする場合、定量化に基づく自動助言だけでは、学生が個別に定める目標点や焦点や方法について学習実態に追従した調整を具体的に助言することは困難である。なぜなら、定量的なアプローチだけでは、学習計画や振り返りの量から傾向を捉えることはできるが、学生個別の具体的内容については捉えきれないためである。

学生は、特に学習計画や振り返りが不十分な場合、自身の内容と離れた定形文や一般化されたアドバイスだけでは、自身の学習実態に対して提示された学習方略を適用することが困難な場合がある。

そのため、学習計画や振り返りの不十分さを学生が自身の問題として認知し、改善する動機につながりやすく提示するためには、学生個別の記述内容に基づいた具体的な改善アドバイスの提示する手法が重要となる。

一方、学生個別の学習スキルに応じたアドバイスとしての品質も同時に満たさなければならない。なぜならば、学習スキルの段階とかけ離れたアドバイス、または方向性の揃わないアドバイスなどでは学生が混乱するためである。これは、学生個別の学習スキルに応じた段階的な足場かけ、足場はずしの重要性である。つまりは、学習スキルの段階を捉えるルーブリックに沿いつつ、個別の記述内容に基づいた具体的なアドバイスの提示が重要となる。これを実現するためには、学生個別に取り組むテーマごとの学習実態に追従するための定性的かつ学生個別なアプローチによる自動支援が必要となる。

そこで、第5章では、学生個別の記述内容に追従した助言を自動生成を実現するために、GPT-4 を用いた自動評価に基づく学生個別アドバイスの自動生成を用いて、自動評価結果

に基づいて学生個別の記述内容に追従したアドバイスを自動生成する手法を提案した。具体的には、学習方略に関する記述項目について、学生が学習計画や振り返りを記述した後に、改善アドバイスを学生の記述内容に追従して反映させることで、学習方略の利用や調整について具体的な記述を支援する手法を提案した。

関連する研究として、ループリックに沿ったアドバイスの自動生成のために、GPT-3 のファインチューニングに教員のアドバイステキストデータと記述内容の具体度評価用ループリックを用いた手法について評価を行った [17]。この手法 [17] では、教員の手動作文と同程度に類似して、ループリックに沿ったアドバイスが生成できていた可能性を示すことができた。しかし、学生の記述内容が不十分な場合に、具体的なアドバイスを生成することができていないことがわかった。具体的なアドバイスを生成できなかった具体的な理由は次の 2 点である。

1. 具体的なアドバイスを生成するための材料がなかった
2. 記述内容の具体度が低い場合に対応するトレーニングデータが不十分だった

なお、従来手法 [17] では、学生の記述内容が十分かつ、ループリック評価の上限評価の場合についても、具体的なアドバイスができていなかったことを報告している。しかし、具体度が低い場合とは異なる理由が存在するため、別のアプローチが必要であると考えられる。そのため、第5章では、学生の記述内容の具体度が十分に高い場合のアドバイス自動生成については、今後の課題とした。

第5章は、アドバイス自動生成に GPT-3 のファインチューニングに教員のアドバイステキストデータと記述内容の具体度評価用ループリックを用いた手法 [17] の改善である。

第5章は以下の点で関連研究と異なる。従来手法の量の分析に基づく最適値や関連情報の提示といった助言に対して、本研究は、学生個別の学習テーマのその時々の内容に追従した自動助言である。具体的には、学生の記述内容と学生個別のテーマ、タスク情報を併用し、さらに GPT-4 のゼロショット学習を用いて、学習計画と振り返り記述内容のループリック評価に基づいたアドバイスの自動生成を行った。

第5章の貢献は、学習支援システムの研究における学生個別の記述内容に追従した助言を自動生成する手法の実現である。

第5章の提案手法の評価結果は、学生の記述内容の具体性が不十分な場合でも、やや学生の内容にあっており、やや理解しやすく、やや改善を実行できそうだと感じさせるアドバイスを自動生成する手法を示した。自動生成したアドバイスは、学習計画や振り返りの記述内容が不十分な学生に対して、学生個別のテーマと学生のタスク進捗情報にやや沿っており、改善点がややわかりやすく、改善案をやや実行しやすいと学生に思わせることができていた。

しかし、理解性の評価については、テーマの情報を単独で併用した場合、テーマとタスクを併用した場合よりも高いという結果が示された。したがって、学生個別の記述内容に追従した助言自動生成の実現の可能性について一定の可能性を示した貢献があったと考えられる。

2.5 結言

本章では、本研究の背景となる関連研究について言及した。

2.2節では、デザイン学習の自動支援に自己調整学習の適用した理由と定性的かつ学生個別なアプローチによる自動支援の必要について述べた。そして、自己調整学習の学習支援として、学生の自己記録を自動提示主手法と学習支援にチャットボットを用いた手法 [11] について言及した。そこで、自己調整学習における自己モニタリングの自動支援と学習方略の使用や調整を促す自動支援を同時に満たす手法の必要性を述べた。

2.3節では、学生個別の学習スキルの段階を自動的に捉えるための定性的かつ学生個別なアプローチによる自動支援の必要について述べた。そして、機械学習モデルを作成して学生の振り返り内容を分類する手法を示した手法 [15] について言及した。そこで、学習の振り返りと学習計画の記述内容を対象に、ループリックについて自動評価を行う手法の必要性を述べた。

2.4節では、学生個別の記述内容を自動的に捉え、学習実態に沿ったアドバイスを自動生成するための定性的かつ学生個別なアプローチによる自動支援の必要について述べた。そして、アドバイス自動生成に GPT-3 のファインチューニングに教員のアドバイステキストデータと記述内容の具体度評価用ループリックを用いた従来手法 [17] の課題点を示した。そこで、学生の記述内容に学生個別のテーマとタスク情報を併用し、GPT-4 を用いることで、学生の記述内容が不十分な場合でも具体的なアドバイスを自動生成できる手法の必要性を述べた。

表 2.1 は、本研究の学習支援システム研究における貢献と関連研究との関係を示した。表 2.1 の「1. 課題遂行段階に対する学習スキルの自動助言」は、自己調整学習の遂行段階について自動的に助言することができる。2.2節において言及した自動聞き取り手法 [10] は、自動的な聞き取りによって、学生の学習スキルについての自動評価までの検証はされているが、自動助言については今後の課題としていた。本研究では、自動助言を目的として第3章の手法を提案した。しかし、実験結果は一定条件下のみの限定的な効果に留まったため、自己効力感と目標設定、方略計画、自己モニタリングの関係性に基づく対話設計という新しい仮説を今後の課題とした。

表 2.1 の「2. 学生個別学習状況の定量化と自動分析」は、従来手法では、学生の学習行動の時間や小テストの成績などの量に基づくデータと、学生の行動の種類や記述に含まれた

表2.1: 学習支援システム研究における貢献と関連研究との関係

項目	自動聞取り 手法 [10]	学 習 分 析 可視化手法 [16]	記述内容分 類手法 [15]	本研究
1. 課題遂行段階に対する学習スキルの自動 助言	△	×	×	△ : 第3章
2. 学生個別学習状況の定量化と自動分析	×	○	×	×
3. 学生個別の助言自動提示	×	○	×	○ : 第5章
4. 学生の記述内容の自動分類	×	×	○	○ : 第4章
5. 学生の記述内容から学習スキルの自動評 価	×	×	×	○ : 第4章
6. 学生個別の記述内容に追従した助言自動 生成	×	×	×	○ : 第5章

単語の数などの質を量に置き換えるなどの定量化データに基づく学習状況を自動分析することができる。学習行動を定量化する自動分析や評価手法は、過去データ蓄積と分析によって、定量化が可能な場合、学生の学習状況を評価する手法として特に有効である。そのため、ある程度答えが定まった学習内容を対象とすることが多い。

しかし、デザインの学習などでは、社会状況と共に変化の大きいテーマについて、学生が個別のテーマを定め、探求的に取り組む場合は、既定の答えや解法、学習過程が定まらない場合が多い。このような学習内容を自動評価する場合については定量的なアプローチだけでは学習実態を捉えきれないという課題がある。また、学生の自己調整スキルを評価するためには、学生が学習目標、焦点、方略について、学習計画や振り返りの記述や表現にどのように反映されているかという実態を捉えることが重要である。

そこで、本研究では、学生の自己調整学習スキルを自動評価する手法として、学生の記述内容を焦点とした第4章の自動評価手法を提案した。なお、本研究では、学習行動を定量化した自動分析については扱っていない。

表 2.1の「3. 学生個別の助言自動提示」は、学生個別の学習行動の自動評価に基づいて、学習行動を最適化する学習スキルの助言情報を自動提示することができる。2.4節において言及した学習分析可視化手法 [16] は、学生の学習行動を最適化するために、学生個別の学習状況について定量的なアプローチによる自動分析を行う。そして、学習行動を最適化するための可視化された最適値や助言情報を自動構成して提示することができる。

しかし、定量的なアプローチだけでは、学習実態の量的な傾向を捉えることはできるが、

学習実態については捉えきれない。量的な傾向にあてはまる客観的な評価によって最適な助言を提示することができるが、傾向にあてはまる学習資料や汎化された助言や指標の提示に留まる。そのため、学生は自身の学習実態と助言された内容を照らして適用することが困難な場合がある。

そこで、本研究では、学生自身が認知し記述した記述内容に沿って改善案を助言として生成する手法として、第5章の学生が記述した内容に追従する助言を生成する手法を提案した。

表 2.1の「4. 学生の記述内容の自動分類」は、あらかじめ定めた種類別に学生の記述内容を自動分類することができる。2.4節において言及した記述内容分類手法 [15] は、目的の種類別にラベリングした記述内容を用いて機械学習したモデルによって、学生の記述内容を定性的なアプローチによって自動分類を行うことができる手法である。本研究では、記述内容分類手法 [15] の概念を応用発展させることで、学生の記述内容のルーブリック自動評価として適用し、第4章の自動評価手法と第5章の学生が記述した内容に追従した助言を生成する手法を提案した。

表 2.1の「5. 学生の記述内容から学習スキルの自動評価」は、学生の記述内容を一定のルーブリックに基づいて自動評価することができる。これまでの自動的な定量的アプローチだけでは、学生個別の学習実態を捉えきことは出来ていなかった。一方、学生の記述内容の自動分類手法 [15] では、学生の記述内容を捉えて分類することができていた。しかし、学生の学習実態を捉えるには、複数のルーブリックに基づいてさらに学習スキルの段階を自動分類する手法が必要であった。

そこで、本研究では第4章において、ルーブリックに基づいて学生の記述内容から学習スキルと躓き度を自動評価する手法を提案した。第4章は、教員が人手で行ったルーブリック評価と比較して十分に実用に耐える性能であるかどうかを検証した。これにより、本研究は、学習支援システムにおける学生の記述内容について学習スキルの自動評価手法の実現に貢献した。

表 2.1の「6. 学生個別の記述内容に追従した助言自動生成」は、ルーブリックに基づいた助言を学生の記述内容に反映させて生成することができる。従来の自動分析に基づく定量的かつ定型的な助言の自動提示では、定量化した学習データの分析に基づいて、最適な量データの可視化や、分析した傾向にあてはまる学習資料やある程度汎化した助言情報を自動提示するに留まっていた。しかし、汎化された助言情報を、学生が自身の学習状況と照らして認知し、自身の状況に適用させることが困難な場合がある。

そこで、本研究では第5章において、定性的かつ学生個別の自動助言生成として、学生個別の記述内容を直接反映させつつ自己調整学習に関する助言を自動的に生成する手法を提案した。第5章は、学生の記述内容と、併用するテーマ情報やタスク情報の組み合わせパタ

ーンを変えて自動生成したアドバイスについて、学生を被験者としたアドバイスの評価を比較した。これにより、本研究は、学習支援システムにおいて学生の記述内容に追従した助言の自動生成の可能性に一定の貢献をした。

第 3 章

定形文による自動的な学習状況の聞き取りとアドバイス

3.1 序言 自己調整学習スキルの支援

自己調整学習における遂行コントロールのプロセスを自動支援するためには、学習者の学習実態を自動的な対話で聞き取りする手法と、聞き取りした内容に応じてアドバイスを自動提示する手法が必要となる。そのために、本章では、表 2.1の学習支援システム研究の項目 1 に示した、課題遂行段階に対する学習スキルの自動助言の実現を目指した。

そこで、遂行コントロールのプロセス（課題遂行段階）を自動支援するために、チャットボットを用いた学習支援システムを開発した。このシステムは、学習者が自身の学習活動を改善するための具体的なアドバイスを提供することを目的とした。具体的には、デザインを学ぶ学習者の学習方略行動を促すために、学習者に学習状況の内容を聞き取り、学習者の回答に応じて学習実態に近い助言を定形文で提示するチャットボットを試作した。そして、学習者の学習方略を促進するかどうかを評価するために、デザイン学習用に開発したスケッチブックアプリに試作チャットボットを実装して実験を行った。

近年、主体的な学習を促進する上で、自己調整学習 [5] の重要性が多方面で指摘されている。特に、学習者が自身の学習方略の状況を把握して学習方略を調整する能力が重要であるとされる。しかし、特にデザインを初めて学ぶ学生は、アイディアの検討から表現を制作していく過程の自己調整に不慣れであることが多い。そこで本研究では、学生が効率的にデザイン学習を進めていくために、学生個別の学習方略を継続的かつ自動的に支援するシステムの開発を目指した。

本研究が対象とするデザイン学習とは、一定の答えや解法の定まらない探索型の学習である。学生が個別にテーマを定め、問題発見と課題設定、解決手法について、学生がもつ自身のアイディアを文章や図、イラストによる企画提案資料、プロトタイピングなど様々な手段で可視化する表現を制作して進められる学習である。学生が自身で発見した問題について、

解決するアイディアの検討とアイディアを表現しようとする主体的な探索と制作活動の積み重ねによって進められる。

このような一定の答えや解法の定まらない探索型の学習では、主体的な学習方略の自己調整が重要である。しかし、自分の状況を把握できず、方略を自分で調整することができない学生も多い。そのような学生は、教員のフィードバックを効果的に得られず、できる学生との差が生じていく。さらに、不安や焦りからデザイン学習への自信を損ない、学習活動が停滞していく負の連鎖へと発展していく恐れが生じる。

そこで、学生が効率的にデザイン学習の活動を進めていくために、学習方略の振り返りと工夫について、教員が学生個別に継続的かつ自動的に学習を支援することが重要である。しかし、学生数が多くなると教員だけでは個別の支援が困難となる。よって本研究では、教員の役割を一部代替するチャットボットの開発を目指す。

第3章では自己モニタリングを中心とした学習方略の支援に焦点をあてる。その理由は、自己調整の大切さに学習者が気付きやすく、学習者にとって比較的やさしいとされているためである [2]。自己モニタリングは、自己調整学習の循環モデル [3] における遂行コントロールのプロセスに含まれる。遂行コントロールは、学習活動中のプロセスであり、支援方法は学生の負担を最小限にとどめることが重要である。チャットボットを用いたのは、学生の負担となる詳細な記述活動を避けるためである。

そこで、第3章は以下の 3 つを目的とする。

目的 1 破綻しない対話を実現する

目的 2 学生が学習状況の振り返りを記述できるようにする

目的 3 学生が学習方略行動を工夫できるようにする

チャットボットは、3 つの目的に対応する以下の 3 つの機能を実装する。

機能 1 対話が破綻しない自動的な対話機能（目的 1）

機能 2 自動的に学習状況の聞き取りを行う対話機能（目的 2）

機能 3 自動的な学習方略のアドバイス提示機能（目的 3）

なお、機能 2、3 は、限られた授業時間内に目的 2、3 を達成しなければならない。なぜなら、学生が個別に課題に取り組む授業時間が、自己モニタリングに適しているからである。

第3章は、デザイン学習を行う学生に対し、機能 2 で自身の学習状況を理解させ、機能 3 で学習の行動変容を目指した。

結論を先に示すと、実験の結果、機能 1 は過半数の対話で破綻がなかったことを確認できた。機能 3 は、機能 2 を併用することで、学生が目標達成の見込みに「どちらかといえば

きそうでない」と回答したときに限定すると、課題プロセスの経過を把握する方略を使用したと自己申告した割合に増加がみられた。また、アドバイスされた学習方略を工夫できていたと示唆された事例が1件あったが、6件についてはアドバイスされる前からできていた可能性があった。そのため、母数を増やした更なる検証が望まれた。機能3の大半は目的を達成しておらず、機能2の記述が増えたとする結果も得られなかった。そこで、未達成の目的について、学習方略間の関係[2]を用いた考察を加え、新しい仮説を探索するために実験データの詳細を分析した。その結果、機能2は、チャットボットが提示する選択肢を選ぶだけでは、不十分であり学生に学習状況を記述させる必要性が示唆された。機能3は、以下2つの方略の相互関係に基づいた支援機能の設計を今後の課題とした。

- 自身の課題のプロセス経過を把握する方略に対して、課題活動を行う際、学習計画に基づいて実行する方略との相互関係
- これから先、この課題内容が得意であると思う自己効力感に対して、自身の課題の現在の状況を絶えず把握する方略と、自身の課題について振り返りを行う方略、そして、課題がうまくいかないと感じたとき工夫する方略の相互関係。

3.2 チャットボットの機能

本研究のチャットボットは、図3.1のデザイン学習用のWebアプリ“スケッチブックアプリ”を自作して実装した。図3.1は、分析途中の研究成果[1]のFig. 2を修正した図である。学生は、スケッチ領域でアイディアの可視化を行っているときに、チャット領域でデザイン学習の支援を受けられる。

スケッチ領域は、学生がアイディアを可視化するための作図機能と図形のレイアウト機能を備える。作図は、図3.1の図形作図ウィンドウの作図ツールで行う。作成した図形や文字、画像はスケッチ領域に配置できる。さらに、作図した図形に、ラベル文字や説明などの文字情報を付加することができる機能を備えた。

チャット領域は、チャットボットチャンネルと教員チャットチャンネルの2つのチャンネルを備える。チャットボットからの学習支援は、チャットボットチャンネルで行われる。チャットボットチャンネルは、チャットボットと学生個人間のテキストチャットである。教員チャットチャンネルは、教員と学生個人間のテキストチャットである。チャンネルを分けた理由は、チャットボットとの対話時に学生が教員を意識することなく率直に回答できるように配慮したためである。そのため、教員は、学生のチャットボットチャンネルを見ることができない。

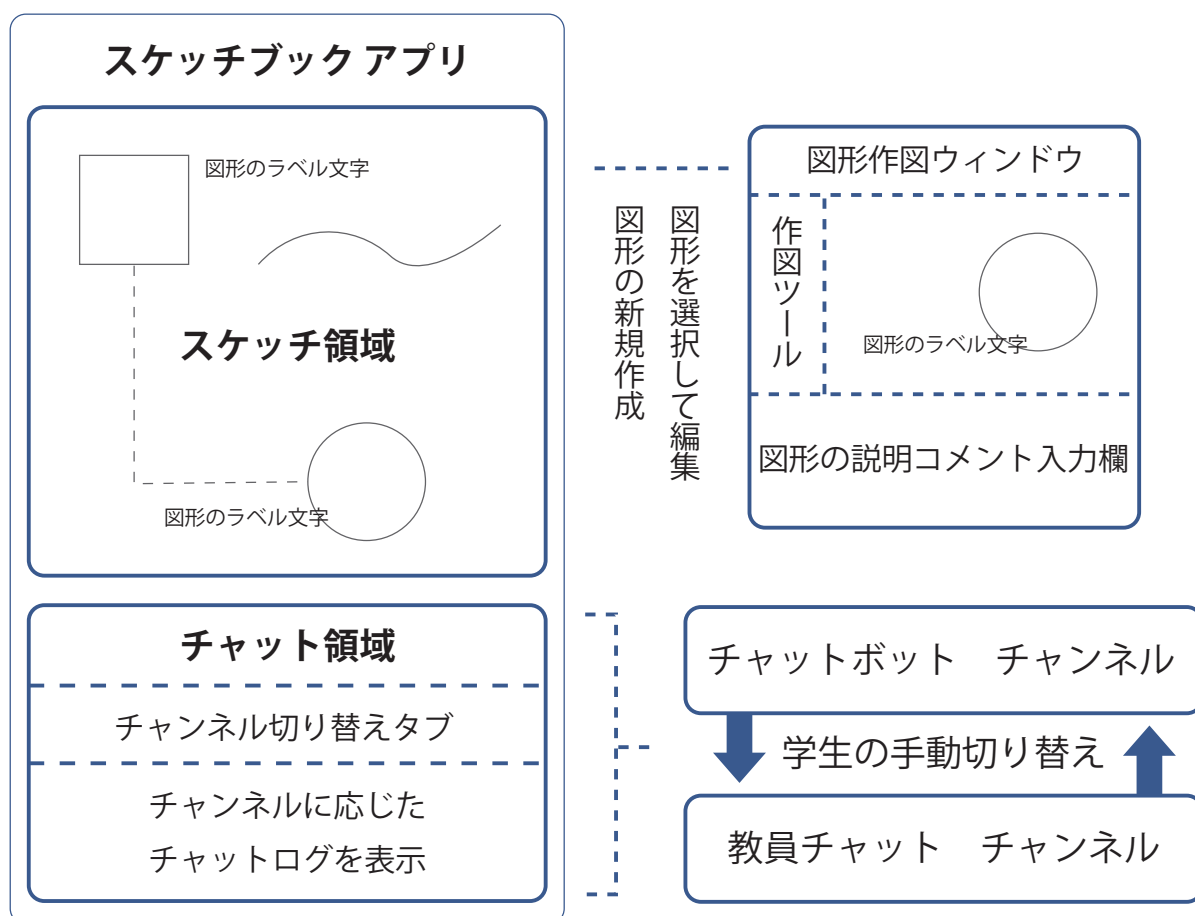


図3.1: デザイン学習用の Web アプリの画面構成 [1] を著者が加工して使用

3.2.1 チャットボットの概要

図3.2は、チャットボットの概要を示す。図3.2は、分析途中の研究成果 [1] の Fig. 3 を修正した図である。チャットボットは、自動対話と自動アンケートを備える。学習支援は、自動対話を用いて行う。

自動対話は、決定木を用いたエキスパートシステムである。チャットボットが提示した問いかけに、学生が選択肢を選んで回答することで結論を導いていく。自動対話の流れは、対話シナリオに基づいて行われる。自動対話において、チャットボットが発話するテキストデータは、あらかじめ決定木の各ノードに設定したテキストデータである。

決定木の生成と対話シナリオのテキストデータ設定は、エキスパートシステム生成ソフトウェアである、株式会社ソラ・ユニバーサルアーカイブスの Thinkeye を用いた。決定木を生成するための変数設定、事例データの入力、編集、テキストデータの設定機能、ID3 ベースに数値対応化したアルゴリズムを備えている。先行研究 [12] と類似の手法を実現でき、

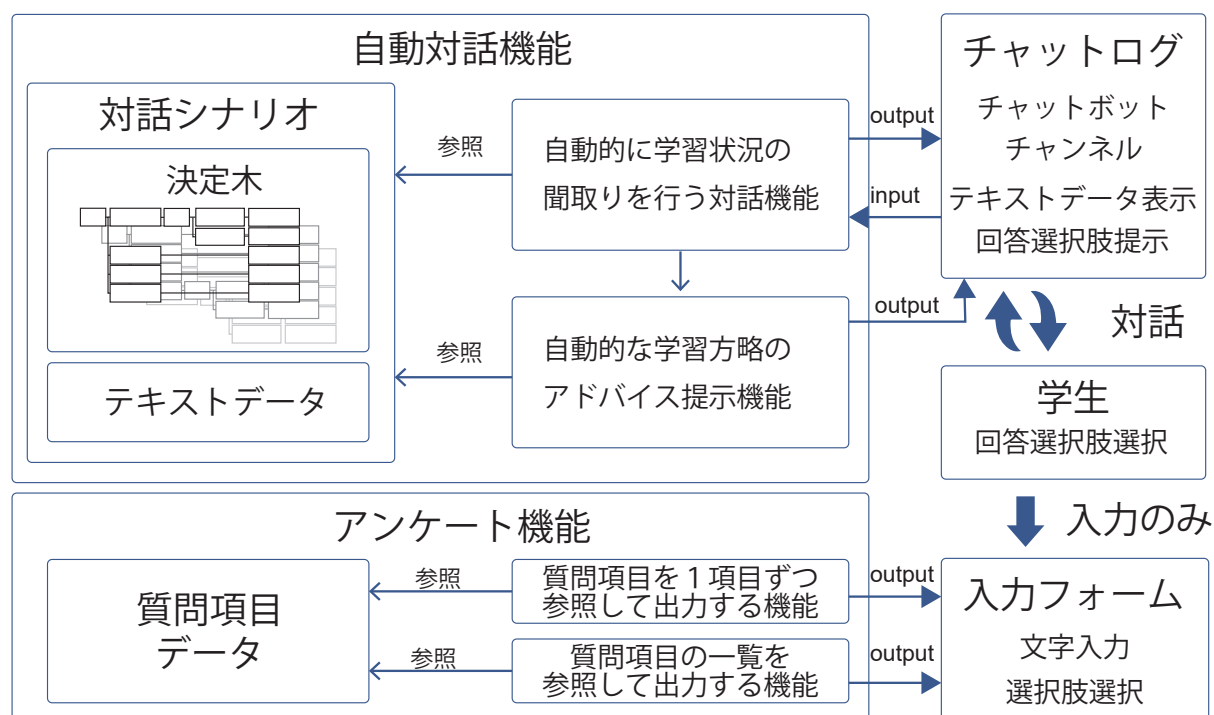


図3.2: チャットボットの概要図 [1] を著者が加工して使用

研究で使用された実績 [18] がある。

変数設定、事例データ、テキストデータ編集を繰り返す作業、そして、図3.2の対話シナリオを参照する部分の実行に適しているため採用した。

自動アンケートは、質問項目と入力欄の提示、入力されたデータの回収を行う。学生の回答についてフィードバックなどを行う対話的な対応はない。

表3.1は、機能2と3の実例として、チャットボットが自動的に行う学習状況の聞き取りと学習方略のアドバイスの内容例を示している。3.1の「機能」は、機能2、3の区別を示している。「種類」は学習方略の種類を示している。「聞き取りとアドバイス」は、チャットボットの聞き取りとアドバイスの発話について、学生が選択した回答の実例を示している。

聞き取りとアドバイスの実行順序は、表3.1の上から順に実行される。

①自己モニタリングの聞き取りは、はじめに目標の達成見込みを聞き取り、学生が“どちらともいえない”を選択した場合の応答を返す、そして、課題を進めていく自信の聞き取りに進む。その後、学生が“まったくない”を選択した場合の応答を返している。すると、①自己モニタリングのアドバイスが提示される。

次に②目標設定に進む。ここでは聞き取りが無くアドバイスを提示している。その理由は①自己モニタリングの聞き取り段階で、既に②目標設定に必要な変数を得ていたため

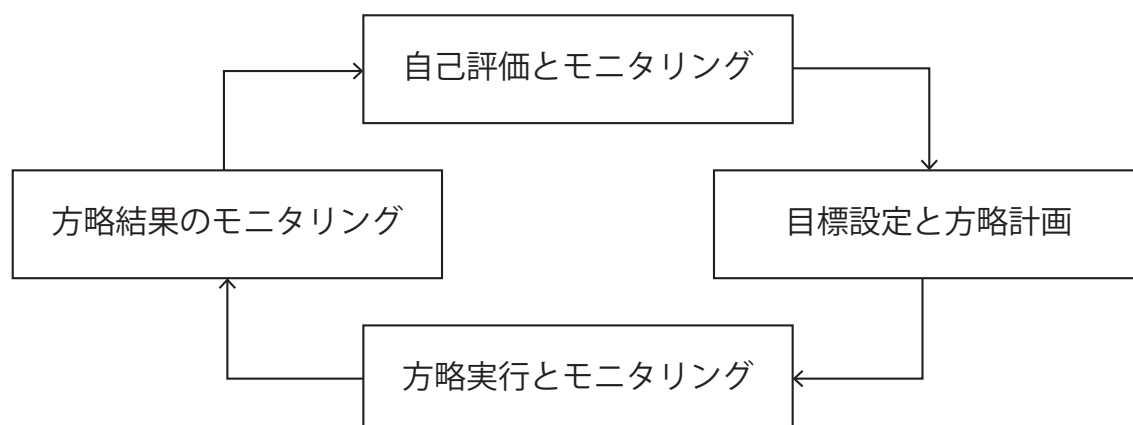


図3.3: 自己調整学習のサイクル・モデル [2] p.11, 図 1 を元に著者が手を加え作成した

ある。

その次に③自己教示に進み、聞き取りが行われている。以上のように、聞き取りとアドバイスが実行される。

表3.1の「種類」①から⑤は、焦点と定めた学習方略である。

表3.1の「種類」⑥方略アドバイスは、⑤課題方略に含まれる詳細なアドバイスを実現するために追加設定した。⑥は⑤で学生が選択する回答によっては実行されない場合がある。

学習方略の焦点 [1] は、自己調整学習の循環モデル [3] と自己調整学習のサイクル・モデル [2] (p.11) の学習方略から、自己調整学習の指導方法 [2] (p.28) にある例示 1 学習時間自己モニタリング書式の項目、および、初歩の自己調整学習者について目標を階層化させるという支援方法 [6] を参考にした。

✓ 計画支援：②目標設定

✓ 遂行支援：①自己モニタリング，③自己教示，④注意の焦点化，⑤課題方略，⑥方略アドバイス

3.3 対話が破綻しない自動的な対話機能（機能 1）

機能 1 の目的は、チャットボットと学生の対話が成り立つ最低限の条件として、機能 2, 3 に対話の破綻によるネガティブな影響がないことを確認することである。

破綻の無い自然な対話を実現するために、決定木の生成に人為的編集データを用いる。

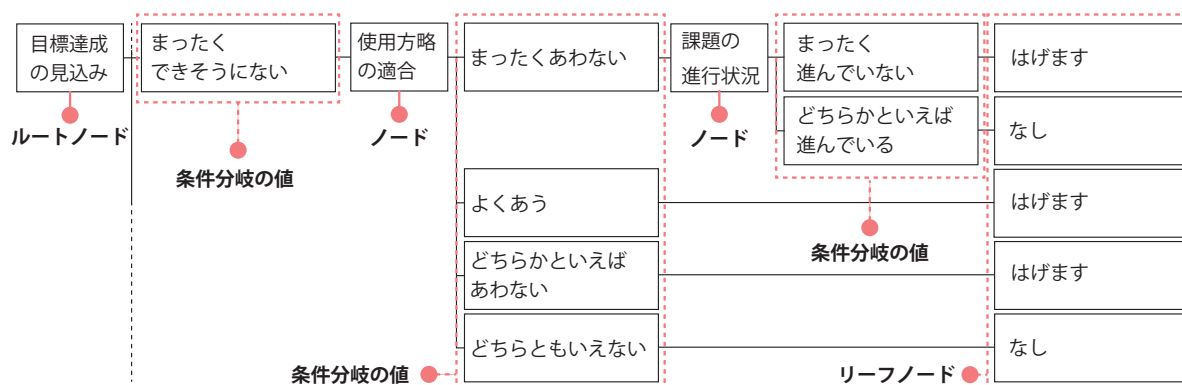


図3.4: 決定木ベースのチャットボット（一部抜粋）

図3.4は、自動対話を構成する決定木のノードについて一部を示した。各ノードにある条件分岐の値を選択肢として学生に提示し、学生が値を選択すると、分岐先の次ノードを辿る。ここまでの過程が機能 2 に該当する。リーフノード到達時には、機能 3 の自動的な学習方略のアドバイス提示機能が実行される。このように、チャットボットの対話の流れは、決定木のルートノードからリーフノードまでのノードの流れと対応する。

チャットボットが発話するテキストは、各ノードに設定されたテキストデータである。学生の回答選択時にチャットボットが返答するテキストは、条件分岐の値にあらかじめ設定されたテキストデータである。このように、決定木と設定されたテキストデータによって、チャットボットが聞取りを行うことで決定木の変数に値を得ながら、結論を導く対話を作ることができる。

しかし、聞取り項目と選択肢の出現順序は、決定木によって生成される。そのため、違和感が生じる場合があり、対話が破綻する要因となることがある。そこで、決定木の生成に使用する訓練データを人為的に編集することによって、対話が破綻する可能性を取り除く編集作業を行なった。

訓練データは、教員が想定した聞取り内容によるダミーの事例データを作成して用いた。

対話シナリオ（決定木とテキストデータ）の編集作業について説明する。編集作業の工程は、事例データ作成、決定木生成、テキストデータ設定の順に行われた。事例データ作成は、変数を設定し、値を入力する。決定木は事例データを用いて生成する。テキストデータの設定は、人手で作文されたテキストデータを図3.4のような決定木のルートノードからリーフノードまでの各ノードに設定した。

図3.5は、テキストデータの設定例を示している。決定木のノードと、条件分岐の各値に設定したテキストデータの例を示した。図3.5は、分析途中の研究成果 [1] の Fig. 5 に基づ

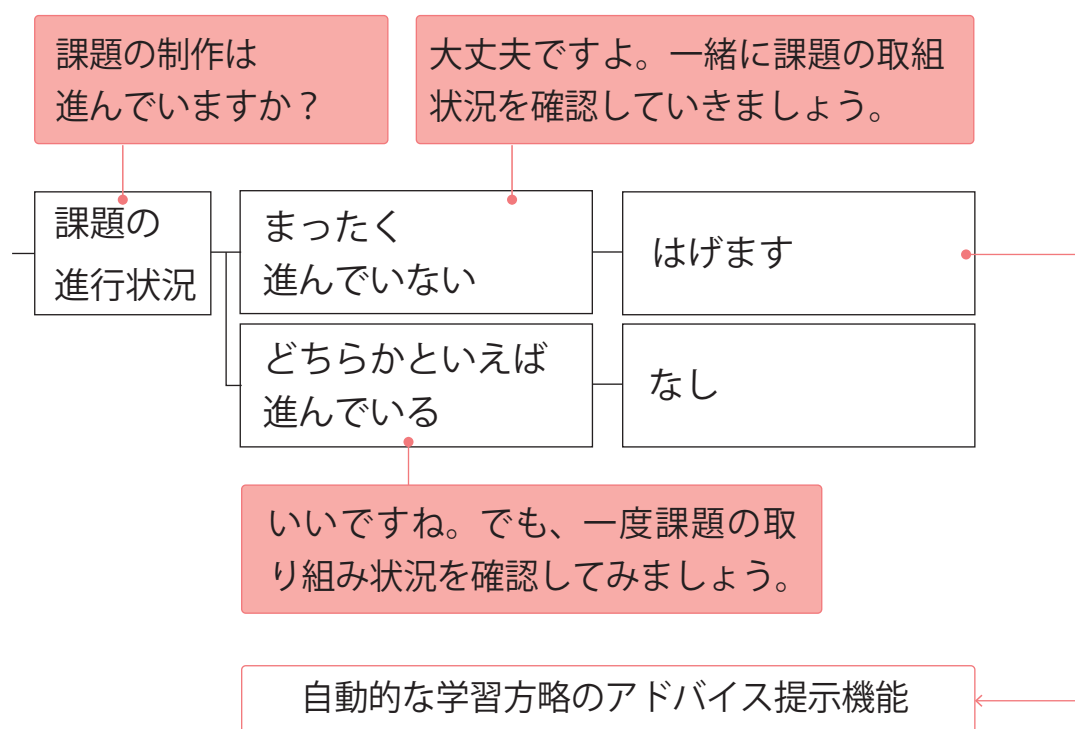


図3.5: テキストデータの設定例 [1] を著者が加工して使用

いて学習状況を聞取る対話部分を抜粋した図である。

このように、対話シナリオは、決定木に設定されたテキストデータで構成される。チャットボットとの自動対話を破綻させないためには、違和感のない対話の流れになるよう事例データとテキストデータを編集する必要がある。そこで、機能 2 と 3 において必要な問いかけと、選択肢の提示、選択した回答への返答、アドバイスの提示などを確認しながら、破綻しない流れとなるよう事例データ（訓練データ）に編集を重ねた。

3.4 自動的に学習状況の聞き取りを行う対話機能（機能 2）

機能 2 の目的は、学生が学習状況の振り返りを記述できるようにすることである。そのためには、学生の自己モニタリングを促すことが重要である。自己モニタリングを促すために、チャットボットは、学習状況の聞き取りを行う。うまく自己モニタリングを促すためには、学生個別の学習状況に応じた対話を行う必要がある。

そこで、本研究が焦点とした 6 つの学習方略①から⑥について、教員が学生に把握して

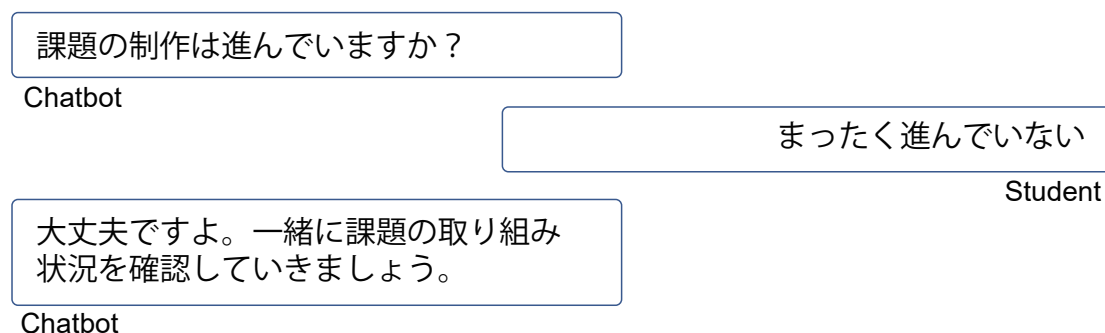


図3.6: 対話の流れの一部 イメージ例 [1] を著者が加工して使用

ほしい観点から学習状況の聞取りがされるように事例データを編集した。

表3.2は、作成した事例データの一部を抜粋した例を示している。表3.2は、分析途中の研究成果 [1] で示された TABLE I の目的変数名を“自己モニタリング”に修正した表である。

表3.2の事例データの変数は、教員が学生から聞き取る学習方略を想定して設定した。事例データに設定した値は、学生の回答を想定して作成したダミーの値である。目的変数に設定した値は、教員がどのような対応をとるのかを想定して作成したダミーの値である。

テキストデータは、人手による作文である。教員が学生の学習状況を聞取る場面での対話内容を想定した。

図3.6は、図3.5の部分がチャットログに表示されるイメージを示した。分析途中の研究成果 [1] の Fig. 6 の学習状況を聞取る対話部分を抜粋した図である。図3.5の決定木において、学生が“まったく進んでいない”と回答した場合の対話の流れについて、チャットログに表示されたイメージを示している。

なお、6つの学習方略ごとに、聞取りを行う対話がリーフノードに到達したとき、機能3によってアドバイスが提示される。

3.5 自動的な学習方略のアドバイス提示機能（機能 3）

機能3の目的は、学生が学習方略行動を工夫できるようにすることである。そのために、機能2で得た学生個別の学習状況に応じ、学習方略のアドバイスを自動的に選択して提示する機能を備えている必要がある。

学生個別の学習状況に応じたアドバイスを提示するために、決定木の各リーフノードに設定するテキストデータは、教員が学生に対して行う対応やアドバイス内容を想定した。これ

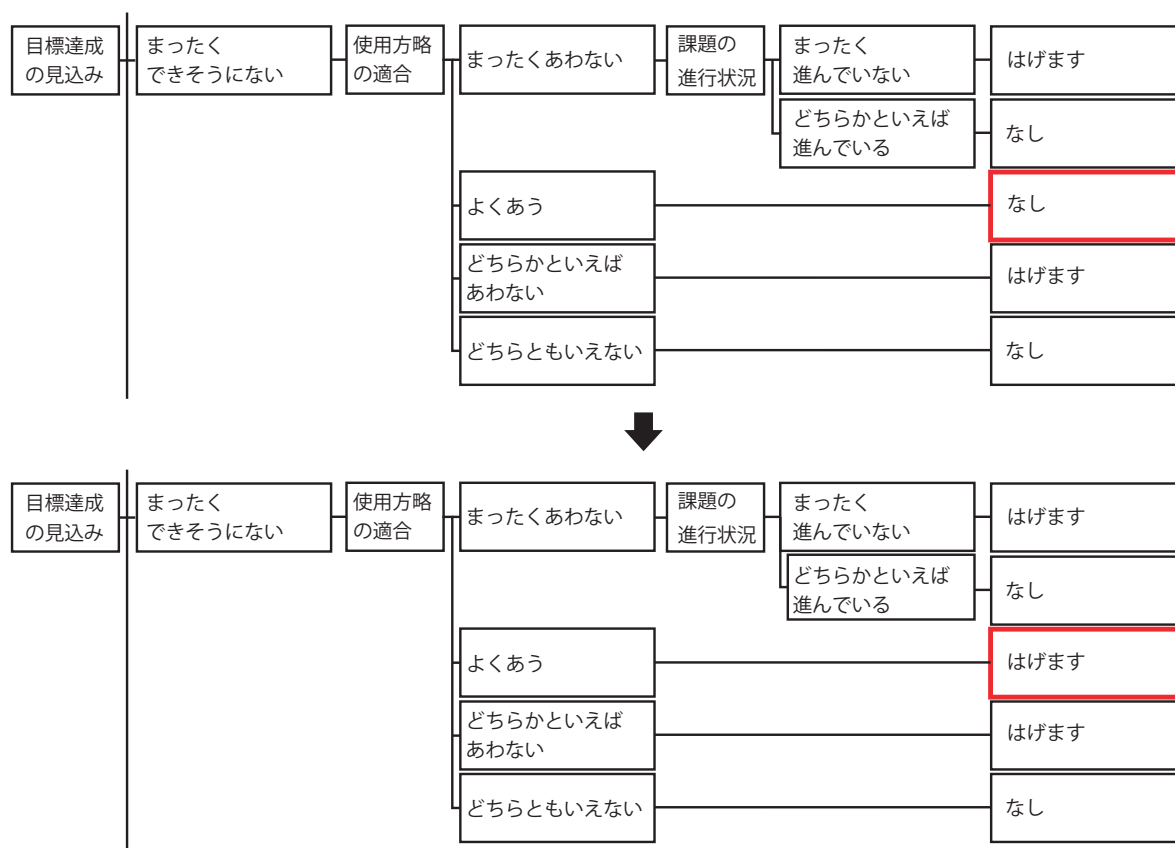


図3.7: 決定木の編集過程 [1]

により、機能 2 で得た学生個別の学習状況に応じてアドバイスが提示されるため、学生は学習方略の工夫を行うことができるようになる。

アドバイスの選択は、機能 2 と共通の決定木を用いる。アドバイスの選択は、6 つの学習方略ごとに生成された 6 つの決定木に基づき、必要な変数が満たされることで定まる。変数の値取得は、機能 2 によって得られた値が用いられる。

アドバイスとして提示されるテキストデータは、決定木のリーフノードに設定されたテキストデータが提示される。

6 つの学習方略についてアドバイスを判断・選択する仕組みについて説明する。表3.3と表3.4は、決定木を生成するための事例データの説明変数、目的変数を示している。図3.7は、アドバイスを最適化する編集工程を示している。図3.8は、生成された 6 つの決定木の組み合わせを示している。

表3.3は、チャットボットが行う学習状況の聞き取りとアドバイスを構成する決定木の生成に使用した事例データの構造を示している。表3.3は、分析途中の研究成果 [1] の TABLE II

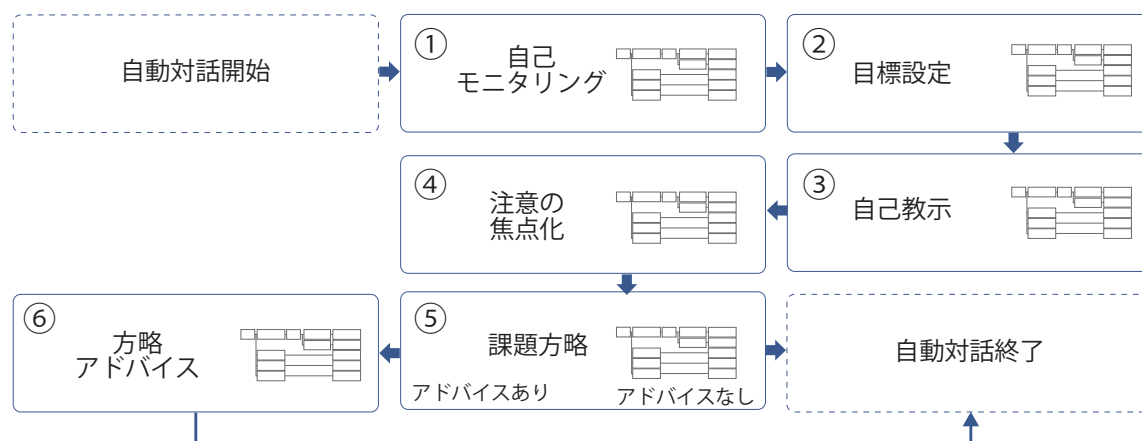


図3.8: 自動対話の流れを構成する決定木 [1] を著者が加工して使用

である。表3.4は、課題方略について詳細なアドバイスを加えるための決定木生成に使用した事例データの構造を示している。

表3.3と表3.4の説明変数の部分は、自動的に学習状況の聞取りを行う対話機能から得られた値が入る変数である。

表3.3と表3.4の目的変数と値は、教員が学生に行う対応やアドバイスを想定して設定したダミーの値である。

課題方略のアドバイスでは、決定木の役割を分割することで、デザイン学習の制作段階や状況に応じて、詳細な方略アドバイスをを行う。つまり、表3.3の課題方略は、アドバイスの必要を判断する。そして、必要と判断した場合、表3.4にある方略の工夫アドバイスの変数によって、より詳細なアドバイスを判断する。

以上の事例データから6つの決定木を生成した。ところが、アドバイスの判断が必ずしも学生の学習状況に適さないとと思われるケースも見られた。そこで、学生の学習状態に対応するように決定木の生成と事例データ、テキストデータの編集を繰り返し行った。図3.7は、決定木を編集することでアドバイスの判断を修正した過程の例を示している。図3.7は、分析途中の研究成果 [1] の Fig. 5 である。

図3.7の上部赤枠は、編集前の部分を示している。下部赤枠は、編集後の部分を示している。①自己モニタリング決定木の一部である。図3.7の上部の例は、編集前は目標達成が“まったくできそうでない”という状況に対して、学習の進め方が、よくあっていると回答した場合のアドバイスである。編集前はアドバイスを“なし”とした。しかし、目標達成ができそうにないと回答している学生に対して、課題の進め方があっていると感じているので

あれば、そのまま自信をもって、学習を進めるよう励ましたほうが良いとの考えから、“はげます”に変更した例である。

図3.8は、6つの学習方略について順にアドバイスの判断をしていくフローを示している。図3.8は、分析途中の研究成果 [1] の Fig. 7 に修正を加えた図である。アドバイスの選択順序は、図3.8の①から⑥に示した6つの決定木ごとのリーフノードによってなされる。

図3.8の6つの決定木（①から⑥）が提示するアドバイスについて概要を示している。6つの決定木は、6つの学習方略（①から⑥）に対応する。

- ① 自己モニタリングは、学生の進捗状況の聞き取り結果について、学習状況へのアドバイスや自己モニタリングの方略についてアドバイスを提示する。
- ② 目標設定は、学生の目標設定状況から、具体的かつ段階的な目標の設定についてアドバイスを提示する。
- ③ 自己教示は、学生の学習内容の理解状況から、方略の使用に必要な知識や理解の確認についてアドバイスを提示する。
- ④ 注意の焦点化は、学生の学習状況と結果に関係する自身の学習特性に注意を与え、学習に集中するための方略についてアドバイスを提示する。
- ⑤ 課題方略は、学生の取り組み内容と方略使用の状況から方略アドバイスの有無を判断する。
- ⑥ 方略アドバイスでは、課題方略でアドバイス必要と判断された場合、取り組み内容の詳細を聞き取り、課題方略についてアドバイスを提示する。

図3.9と図3.10は、実際のチャットボットと学生の対話例と決定木との対応関係を示している。聞き取りによって得た変数を用いて、自動的に学習方略を提示する対話の流れを示した。図3.9と図3.10は、6つの決定木のうち①から②の部分を示した。図3.10は、分析途中の研究成果 [1] の Fig. 8 に説明の修正とチャットログを追加した図である。

学生は、授業時間中の指示されたタイミングに、図3.10の“はじめる”ボタンをクリックする。すると、自動的に学習状況を聞き取る自動対話が始まる。図3.10の①は、聞き取りの開始を学生に知らせている。

図3.10の①-1 聞き取りは、図3.9の①-1 に設定されたテキストデータがチャットログに表示されている。図3.10の①-1 選択肢の提示は、聞き取りに学生が回答するための選択肢が提示される。

ここで提示される選択肢は、図3.9の①-1にある条件分岐の値である。図3.10の①-2 選択への応答は、①-2 学生の選択に対応し、図3.9の①-2 に設定されたテキストデータが表示される。

図3.10の①-3 アドバイスは、ここまでの聞取りによって図3.9の①自己モニタリングの変数①-1 と①-2 が得られたため、①-3 のリーフノードに到達したことで、①-3 に設定されたテキストデータが表示されたものである。

図3.9の①自己モニタリングから②目標設定への移行は、①がリーフノードに到達したため、図3.8のフローに沿って②に移行したことを示している。

図3.10の②-1 は、図3.9の②-1 に設定されたテキストデータが表示されている。図3.10の②-2 選択肢の提示から学生②-2 選択への応答は、図3.10の①-1 選択肢の提示から①-2 選択への応答と同様である。

図3.10の②-5 アドバイスは、図3.9の②-3 の聞取りと②-4 の学生の選択がなく②-5 に到達した。②-5 に設定されたアドバイスのテキストデータが表示されている。

図3.9に示す②-3 の聞取りと、②-4 の学生の選択がなく②-5 に到達した理由は、既に①-1 で①-2 の値を得ているため、図3.9 の②-3 に①-2 の値が与えられたためである。学生に同じ質問を繰り返さないためのしくみとして用いた。

3.6 実験設定

3.6.1 検証する仮説

本研究は、以下の 5 つの仮説を定義した。

仮説 1. 決定木と訓練データの人為的編集によって破綻の無い自然な対話の実現できる

なぜなら、人為的に編集を繰り返した事例データとテキストデータを用いたからである。その結果、機能 2 と 3 の最低限の条件として、破綻の無い自然な対話を実現する。

仮説 2-1. 振り返りの記述内容について：自己モニタリングの支援機能によって、実験群の学生が各項目に沿った記述ができるようになる

なぜなら、機能 2 に、教員が学生に学習方略項目ごとの学習状況を想定して、聞取りと回答を編集した事例データを訓練データとして用いたからである。その結果、学生は自身の学習の取り組み方について学習方略項目ごとの自己モニタリングが促され、その結果、振り返りの記録に学習方略項目に沿った記述が多くなる。

仮説 2-2. 達成度の振り返りについて：自己モニタリングの支援機能によって、実験群の学生の達成度評価は、教員の学習評価との乖離が軽減される

なぜなら、仮説 2-1 に示した機能 2 によって、自身の学習状況を客観的に自己評価する。その結果、振り返りの達成度自己評価と教員の学習評価との乖離が軽減される。

仮説 3-1. アドバイスを提示された実験群は、学習方略アンケートの結果が高まる

なぜなら、機能 3 に、教員が学生の学習状況に対して行う対応やアドバイスを想定し

て作成された訓練データを用いたからである。その結果、機能 2 で聞取った学習状況に応じて自動的にアドバイスが提示されるため、学生は学習方略の行動に工夫が促され、その結果、学習方略アンケートが高まる。

仮説 3-2. アドバイスを提示された実験群は、振り返りの記録にアドバイスを採用した記述が見られる

なぜなら、仮説 3-1 に示した機能 3 によって、学生が学習方略の行動に工夫を行うことができるからである。その結果、振り返りの記録にアドバイスを採用した記述が見られる。

仮説 1 が正しかった場合、機能 1 は、破綻しない対話を実現する目的 1 を達成したとする。仮説 2-1 と仮説 2-2 が正しかった場合、機能 2 は、客観的に自己評価された学習状況について振り返りの記述ができるようにする目的 2 を達成したとする。仮説 3-1 と仮説 3-2 が正しかった場合、機能 3 は、学生の自己申告と行動結果の両面から学習方略を工夫ができるようにする目的 3 を達成したとする。

3.6.2 実験方法

実験参加者は、最終的に 37 名の下承を得られた。10 名は後から追加で下承を得た。実験期間のはじめから実験参加に下承していたのは 27 名であった。後から追加となった 10 名は、はじめから下承していた 27 名と同様に、実験期間のはじめから実験タスクを実施していた。統制群は、被験者の学生番号奇数偶数で分けた 2 群とした。対照群は 18 名、実験群は 19 名となった。実験期間は、2020 年 10 月 27 日から 2020 年 12 月 15 日の期間で行った。

実験におけるデザイン学習の課題内容は、学生が個別に選んだ任意の問題について解決となるアイデアを可視化して提案することであった。学生は、アイデアの検討を重ねながら課題を進めるために、スケッチブックアプリの作図機能などを用いて、アイデアに含まれる要素を図形や文字で可視化して具体化した。

実験群と対照群は、図 3.11 の赤枠で示したタスクが異なる。図 3.11 は、分析途中の研究結果 [1] の Fig. 9 である。実験群は、デザイン学習直前に学習方略と自己効力感に関する実験アンケートの採取、学習中にチャットボットとの自動対話による学習方略のアドバイス、学習終了前に振り返り記録の入力とした。一方、対照群は、デザイン学習直前に学習方略と自己効力感に関する実験アンケートの採取、学習中に提示される学習方略の計画入力、学習終了前に学習の振り返り記録の入力とした。

実験群と対照群について提供した機能の違いを述べる。実験群には、チャットボットの機能 1 を伴う機能 2 の学習状況の聞き取りを行う対話機能と、機能 3 の学習方略のアドバイス提示機能を提供する。一方、対照群には、表 3.5 に示された質問文が提供される。

対照群の学生は、従来手法を模した入力フォームに学習計画を作文しておくことで、課題遂行中の自己モニタリングを促すが、学生が自ら教員に求めなければ何のフィードバックも得られないという設定である。つまりこの評価実験では、学生から質問する等の行動をしないと教員の支援が得られない状況と比較することで、提案手法の有効性を検証する。

実験タスクは、実験期間中の隔週ごとに行った。実験タスクの各タスクはチャットボット内に提示された“はじめる”ボタンを学生が操作することで開始された。チャットボットはタスク終了ごとに、次のタスクを開始するタイミングを指示した。学生はチャットボットの指示に従って各自の学習直前、学習途中、学習直後のタイミングで各タスクを実行した。

実験対象の授業は COVID-19 対策として、隔週で対照群、実験群が交互に対面、遠隔オンデマンドで実施される条件下で行われた。

実験タスクの実施回数は、当初は、隔週の授業時間と時間外学習を含めて 4 回の予定であったが、実際には学生が追加で行った 2 回の実験タスクも存在するため、計 6 回の実施となった。

学生に実験タスクの実施は強制しなかった。学生が実験タスクをやり忘れた場合は、振り返って実験タスクを実施することとした。実験タスクを複数回実施しなかった場合は、状況を振り返って実施することを許容した。1 回ごとの実験タスクの間隔は隔週を基本としたが、休講など授業実施日の都合により実験タスクの期間が 2 週となる場合があった。この場合も実験結果に含めた。学生の操作ミスなどにより実験タスクを複数回連続実施した場合は、学生の申告により該当データを削除した。

実験タスク 2 回目から 3 回目の間では、遠隔授業の時間に教員からのフィードバックが実施された。教員チャットチャンネルのチャット機能を利用して、学生ごとに 1 回のみ、学生のアイデア表現に対してコメントされた。フィードバック内容では、直接的に本研究の学習方略について触れないようにされた。遠隔授業の時間に実施されたため、実験結果への影響は、最小限に留まると考え実施した。教員チャットチャンネルでのフィードバックに関する実験結果への影響については、今後の課題とする。

図3.11で示した、対照群の学習方略の計画入力、および、実験群対照群両方の学習終了前の振り返り記録記入で使用された自動アンケートについて説明を加える。自動アンケートは、フィードバック機能を備えない。つまり、学生の回答内容に対して、チャットボットからの反応はない。すべての質問項目入力後の送信時にのみ、送信内容の受信を返す。自動アンケートで回収した回答については、教員のフィードバックもしなかった。つまり、統制群間の異なるタスクは、対照群の一律の質問項目に回答するタスクと、実験群のチャットボットの自動対話によって聞取りとアドバイスを受けるタスクである。

図3.12は、自動アンケートのイメージを示している。あらかじめ設定された固定の質問項

目が一問ずつ回答欄と共に表示され、学生は回答を文字入力した。学生の回答は保存されチャットログに表示された。

表3.5は、図3.11の対照群の学習方略の計画入力で使用された質問項目を示した。表3.6は、実験群および対照群の図3.11の振り返り記録入力の質問項目を示した。表3.5と表3.6の質問文は、学生の回答を促すことを意図して人手で作文した。

3.6.3 評価方法

仮説 1 の評価方法

破綻の無い自然な対話ができただどうかを評価した。そのために、図3.11の実験群に示されたチャットボットのタスクにおいて、毎回対話終了時に対話アンケートを採取した。対話や回答が不自然との回答は、対話が破綻していたとし、自然との回答は、最低限対話が破綻していなかったとする。

アンケートの採取は、チャットボットの自動アンケートを用いた。

対話アンケートは以下の 3 項目について 5 件法、あるいは自由記述で行った。

1. エージェントの発話の流れに不自然さを感じましたか？(5 件法, 1= “とても不自然さを感じた”, 5= “非常に自然な感じがした”)
2. あなたの回答に対するエージェントの発話に不自然さを感じましたか？(5 件法, 1= “とても不自然さを感じた”, 5= “非常に自然な感じがした” までの 5 件法)
3. 対話の流れで、特に不自然さを感じたところがあれば教えてください。(自由記述)

仮説 2-1 の評価方法

振り返り記録内容について、学生が各項目に沿った記述ができるようになったかどうかを評価した。そのために、図3.11の実験タスクにある「振り返り記録」の各記述項目の内容について実験群と対照群を比較した。

振り返り記録の 4 つの自由記述項目（目標，焦点，方略，集中方法）について、各項目に沿った記述が含まれているかどうかを筆者が主観的に判定した。

具体的には、以下の項目ごとの基準によって、適切な記述があると判定したスコアは 1 点、そうでない場合は 0 点とした。目標の基準は、「個別の学習状況に応じて到達点や目標を設定していたかどうかの記述がある」とした。焦点の基準は、「何のために何に着目していたかの記述がある」とした。方略の基準は、「どういう方法を工夫していたかの記述がある」とした。集中方法の基準は、「どういう学習環境で実施してどうであったかの記述がある」とした。一方、具体性に欠ける記述は 0 点とした。また、授業回を考慮して学生の学習状況とかけ離れている場合も 0 点とした。なお、統制群間の比較は、実験タスク回ごとの被

験者個別の各項目を判定したスコアを用いた。

仮説 2-2 の評価方法

学生の振り返りについて、実験群の学生の自己評価と、教員の学習評価との乖離が軽減されたかどうかを評価した。そのために、図3.11の実験タスクにある「振り返り記録」の達成度と教員の学習評価を用いて相関分析した。

学生の達成度は、学生自身が授業回ごとに定めた目標の達成度であり、教員の学習評価は、授業回ごとに授業担当教員（筆者らではない）が学生の学習報告を評価した点数である。

仮説 3-1 の評価方法

学生が学習方略の行動に工夫を行うことができたかどうかを定量的に評価した。そのために、チャットボットと学生の対話ログと実験アンケート（学習方略と自己効力感に関する実験アンケート）を分析した。

実験アンケートは、以下の質問項目を設定した。先行研究 [6] の評価で用いられたアンケート項目に基づいて、デザイン学習用の言葉に変更したものである。さらに、学習方略の利用に自己効力感が関与する [5] との指摘から、自己効力感についてのアンケート項目も含めた。

学習方略に関する実験アンケートは、12 項目を 5 件法（1= “まったくあてはまらない”，5= “非常によくあてはまる”）とした。自己効力感に関する実験アンケート項目は、自己調整学習 [19] の研究で用いられた自己効力感に関するアンケート項目をデザイン学習用に変更して使用した。

学習方略に関する実験アンケート項目

- SRL01** : 課題活動を行う際、段階的な目標設定を行う
- SRL02** : 課題を遂行する自信を持っている
- SRL03** : 自身の課題内容に興味があり、意欲的に取り組んでいる
- SRL04** : 課題活動を行う際、計画に基づいて実行する
- SRL05** : 自身の課題の現在の状況を絶えず把握している
- SRL06** : 自身の課題のプロセス経過を把握している
- SRL07** : 自身の課題について振り返りを行う
- SRL08** : 課題がうまくいかないと感じたとき、工夫している

自己効力感に関する実験アンケート項目

- SE01** : 課題がしっかりできると思う
- SE02** : 課題を進めていく自信がある

SE03 :これから先, この課題内容が得意であると思う

SE04 :この課題は, いい成績がとれるだろうと思う

なお, 先行研究 [6] は, 計画, 実行, 評価の各段階ごとに学習方略とアンケート項目を設定している. しかし, 第3章では, 実行段階で実験アンケートを行った. 本来であれば, 項目間の関係性を考慮して段階ごとに細かく実験アンケートを実施した調査を実施すべきであったが, 第3章では段階ごとに複数回調査を行うことが困難だった. そのため, 学生の実行段階の最初のタスクで実施された. したがって, 学習方略の使用状況をリアルタイムに反映したものとはいえない. しかし, 毎回の実行段階初期における学習方略に対する学生の認識や意識を反映したものといえる. 学習方略の認識や意識は, 意図的に学習方略をモニタリングし, 調整していく自己調整学習では, 重要な要素である.

仮説 3-2 の評価方法

学生が学習方略の行動に工夫を行うことができたかどうかを定性的に評価する. そのために「振り返りの記録」について, アドバイスを採用した記述が見られるかどうかを分析する.

振り返り記録に含まれる目標, 焦点, 方略, 集中方法の 4 つの自由記述項目について, 筆者が主観的にアドバイスを採用していると判定したアドバイス採用例を確認する. 確認できたアドバイス採用例について, アドバイス提示後に学習方略の行動に工夫があったと判定できる事例を確認する.

3.7 結果

仮説 1 は, 対話アンケート 59 件のうち, チャットボットの発話 39 件, および, チャットボットの回答 40 件について, チャットボットとの対話が自然であると回答されていた. 一方, 自然な流れであるとされた対話内容の偏りと, 8 件の回答に誤字による影響が示された.

仮説 2-1 は, 実験群の学生について振り返りの記述ができるようになったとする結果は得られなかった.

仮説 2-2 は, 学生の達成度自己評価は, 教員の学習評価との乖離がみられた.

仮説 3-1 は, 課題プロセスの経過を把握する方略 (SRL06) と学生が目標達成の見込みに「どちらかといえばできそうでない」と回答したときの対話パターン (I02) について, 順位相関分析と偏順位相関分析の結果, 対話パターン (I02) に限定したとき, SRL06 が増加した可能性はあるが, 機能 2 も関係していることが示唆された. しかし, SRL06 以外の学習方略については, 対話パターンと関係した増加を確認できなかった.

仮説 3-2 は, 目標を具体化, 段階化する対話パターン (I02) のアドバイスを採用して学習方略を変化させた可能性が 1 件にみられ, 6 件についてはアドバイスされる前からできて

いた可能性があった。一方、3件のアドバイスが提示されたにもかかわらず採用しなかった事例、23件のアドバイスが提示されず、該当する記述もなかった事例があった。

表3.7は、実験タスク回毎の実験タスク実施人数と実験アンケート実施人数を示した。

実験期間は、8週間のうち隔週で4週実施後、さらに追加で2回実施した。しかし、COVID-19の影響も大きく、期待していたとおりに実験データを回収することができなかった。実験期間中、授業への参加が困難になってしまった学生、受講者の半数ずつに隔週に交互で対面と遠隔で行う授業形態の混乱により欠席などの影響が生じた。その結果、実験参加者の実験タスクの実施や実験アンケートの回収ができず未回収が多数生じた。そのため、実験回ごとに回収できたデータ数が異なり、回収できたデータ数も少なくなった。そこで、実験結果に、実験タスク4回目の後、授業最終回の実験アンケートまでの期間に、実験タスクを複数回実施し、操作ミスであると申告をしていない学生のデータを含めた。つまり、実験結果には、5回目（4名）、6回目（1名）の実験タスクのデータも含まれる。

また、実験タスク3回目において実験群2名にチャットボットの対話が開始されない不具合が生じた。不具合の影響により、チャットボット、あるいは、学習計画を記入するタスクを行わずに振り返りの記録や実験アンケートだけを回答している事例が見られた。

3.7.1 対話が破綻しない自動的な対話機能（機能1）に関する実験結果

以下の実験結果から過半数の対話については破綻がなかったことがわかる。

図3.13は、対話アンケート回答数59件の内訳を示している。図3.13は、分析途中の研究成果[1]のFig. 10である。実験群による最大6回の実験タスクから59件の回答結果について説明する。本来60件のところ、1件は対話アンケートの不具合によって回答が得られなかった。

チャットボットの発話（図3.13の「発話の自然さ」）について、過半数が自然に感じたとは回答されていた。具体的には、39件66.1%が「どちらかといえば自然な感じがした」と「非常に自然な感じがした」、一方、「どちらかといえば不自然さを感じた」が10件16.9%、「とても不自然さを感じた」が0件、「どちらともいえない」が10件16.9%であった。

さらに、学生の回答に対してチャットボットが返す発話の自然さ（図3.13の「回答の自然さ」）についても、チャットボットの回答の過半数が自然に感じたとは回答されていた。具体的には、40件67.8%が「どちらかといえば自然な感じがした」と「非常に自然な感じがした」、一方、「どちらかといえば不自然さを感じた」の回答が11件18.6%、「とても不自然さを感じた」の回答が0件、「どちらともいえない」が8件13.6%であった。

発話の自然さと回答の自然さの2つの結果から、過半数の対話について、学生が自然さを感じており、最低限対話は破綻していなかったことが示唆された。

表3.8は、対話アンケート（5 = 非常に自然な感じがした，1 = 非常に不自然さを感じた）と対話パターン ID 出現の有無についての順位相関を示している（相関係数絶対値 >0.35 , $p<0.05$ を抽出）。表3.9は、対話パターンの一部を示している。

表3.8，表3.9から，発話や回答の流れが自然であると感じていたのは，学生が概ね課題内容を理解し，自信をもって，集中できていると回答した場合であることが読み取れる．なぜなら，正の相関を示したときの対話パターンは，ポジティブな選択肢であったためである．具体的には，H05「非常に自信がある」，J06「よく理解できている」，C01 は J06 に付随，L05 は「とてもよく集中している」，D01 は L05 に付随，H04 は自信について「どちらかといえばある」，I04 は達成見込みについて「どちらかといえばできそう」であった。

なお，対話アンケートの「不自然さ」「自然さ」の評価は，提示されたテキストデータが日本語として自然かどうかという意味であった可能性がある．なぜなら，対話パターン（J04）のテキストデータに誤字が含まれていたためである．6 名に 8 回，誤字を含んだ対話パターン（J04）が表示されていた．該当する 8 回の対話アンケートの結果は，以下のとおりであった。

| とても不自然さを感じた |: 0 件
| どちらかといえば不自然さを感じた |: 3 件
| どちらともいえない |: 1 件
| どちらかといえば自然な感じがした |: 3 件
| 非常に自然な感じがした |: 1 件

表3.10は，チャットボットとの対話について，特に不自然さを感じたところについて自由記述の回答を示している．ボットが同じ質問や回答を繰り返し示したこと，テキストデータの言葉が硬いこと，テキストデータの文が理解しづらいこと，違和感のあるテキストデータがあったことに対して指摘がされていた。

表3.10の「褒められるとやる気が出る」「選択肢の順番が毎回異なる」の回答は，不自然さについての回答ではないとした．なぜなら，該当する自然さ不自然さの回答データを確認したところ「どちらかといえば自然な感じがした」あるいは「非常に自然な感じがした」と回答されたときの記入であったためである。

3.7.2 自動的に学習状況の聞き取りを行う対話機能（機能 2）に関する実験結果

統制群間の比較のために，実験タスク 1 回から 4 回までを対象とした．実験タスクのチャットボット，または計画入力を実施せず，振り返り記録だけを実施していたデータについては除いた．実験群は 8 件を除いた 47 件，対照群は 7 件を除いた 38 件について，振り返り 4 項目のデータを用いた。

仮説 2-1 に関する実験結果

以下の実験結果から実験群は振り返りの各項目に該当する記述が増えるとした仮説が棄却されていることがわかる。図3.14は、振り返り記録の実験タスク回ごとのスコア平均を示している。実験群の平均値は、元から対照群よりも高かったことが読み取れる。実験群は対照群よりも振り返りの各項目に沿った記述が多くなるとする有意差は認められなかった ($p = 0.184 > 0.05$)。Brunner-Munzel 検定を用いた。

実験群：平均 = 0.5, $S = 0.5$, $n = 188$ (47 件 × 4 項目), 対照群：平均 = 0.428, $S = 0.495$, $n = 152$ (38 件 × 4 項目)。なお、3 回目で大きく下がった理由については考察で述べる。

以下は、「4.3.2 仮説 2-1 の評価方法」で述べた評価基準で判定した例である。

項目に沿っている例：“他人が見ても分かるような図を作ること”，“アイデアが浮かばないので、メモ書き程度に書く”

項目に沿っていない例：“デザインについて”，“よりよいアイデアを考える”

学習状況からかけ離れている例：“完成させる”，“課題の提出”，“最後までやり遂げる”

仮説 2-2 に関する実験結果

以下の実験結果から実験群は学生の達成度自己評価と教員の学習評価の乖離が軽減するとした仮説が棄却されていることがわかる。学生が自己評価した達成度と教員の学習評価を順位相関分析した結果、実験群は、対照群よりも弱い負の相関がみられた。実験群は、相関 = -0.184 , $p = 0.306 > 0.05$, $n = 43$ であり、対照群は、相関 = -0.07 , $p = 0.653 > 0.05$, $n = 33$ であった。なお、実験群 47 件のうち 4 件、対照群 38 件のうち 5 件は、教員の学習評価に必要な学習報告が未提出のため除いた。

3.7.3 自動的な学習方略のアドバイス提示機能（機能 3）に関する実験結果

仮説 3-1 に関する実験結果

以下の実験結果から対話パターン（I02）に限り、自身の課題のプロセス経過を把握する学習方略（SRL06）が増加した可能性があったことがわかる。

図3.15は、図3.16, 図3.17, 図3.18の分析方法を示している。図3.16, 図3.17, 図3.18は、分析途中の研究成果 [1] の Fig. 11, 12, 13 である。図3.15の条件 1 では、統制群間を比較し、有意に増えた実験アンケート項目（SRL, SE 項目）を抽出した。条件 2 では、実験群の最終回と初回を比較し、有意に増えた SRL, SE 項目を抽出した。条件 3 では、条件 1 と条件 2 を満たす SRL, SE 項目の増減とチャットボットの対話パターン ID の出現回数について相関分析した。

図3.16, 図3.17は, 授業最終回の実験アンケートに回答した 21 名 (実験群 10 名, 対象群 11 名) の結果を示している. なお, 最終回の実験アンケートを実施しておらず, 実験初回と最終で比較ができなかった 16 名のデータは除いた.

図3.16は, 条件 1 の結果を示している. 自身の課題プロセス把握方略の使用 (SRL06) を統制群間で比較した結果, 実験群は, 5% 有意水準にて有意な差 ($p = 0.034 < 0.05, n = 21$) をもって多かったことが読み取れる [1].

図3.17は, 条件 2 の結果を示している. 授業最終回の自身の課題プロセス把握方略の使用 (SRL06) は, 初回よりも 5% 有意水準にて有意な差 ($p = 0.016 < 0.05, n = 10$) をもって増えていたことが読み取れる [1].

図3.18は, 条件 3 の散布図を示している. チャットボットの対話パターン I02 の出現回数 (10 回) と自身の課題プロセス把握方略の使用 (SRL06) の増加に弱い正の相関 ($+0.3015, p = 0.023 < 0.05, n=54$) が読み取れる [1]. 図3.18は, 分析途中の研究成果 [1] の Fig. 13 のサンプルサイズを 54, 順位相関係数 0.3015, p 値を 0.023 に修正した図である. スピアマンの順位相関分析を用いた.

図3.19は, 課題プロセス把握方略の使用 (SRL06) の増加に弱い正の相関を示したアドバイスの提示 (対話パターン I02) について, 機能 2 で聞取りされる対話ログ (①自己モニタリング, ②目標設定, ③自己教示, ④注意の焦点化, ⑤課題方略) を交絡因子とした I02 出現数と SRL06 増減値の偏順位相関分析の関係を示している. 図3.19により, SRL06 の高まりは, 自動的なアドバイス (対話パターン I02) の提示によるものだけでなく, 機能 3 の I02 以外のアドバイス提示や機能 2 も関係していることが示唆された. 図3.19の結果は, 偏順位相関 $+0.207, n = 54, p = 0.182 > 0.05$ であった. さらに, 機能 2 の聞取りログに加え, 機能 3 の I02 以外のアドバイス提示を交絡因子として考慮した場合は, 偏順位相関 $+0.125, p = 0.732 > 0.05$ であった. スピアマンの偏順位相関分析を用いた.

分析に用いた 54 件について, データの抽出方法を説明する. 実験群のうち 1 名は, チャットボットの不具合によって実験タスクを実施できなかった. そこで, 実験アンケートを実施しており, 前回の実験タスクと実験タスク回ごとの実験アンケート項目の差分から増減値を得ることができた 54 名のデータを用いた. なお, 実験タスクを追加実施した 5 回目 (4 名), 6 回目 (1 名) の学生データが含まれている.

対話パターン I02 の出現数について算出方法を説明する. 図3.20は, 対話パターン I02 について, ノード「目標達成の見込み」から枝分かれする選択肢から学生が「どちらかといえどできそうにない」を選択したときに, 対話ログに記録されることを示している. 対話パターンの出現数は, ログに登録された対話パターンをカウントした. このことから, 算出対象のログ全体数と対話パターン I02 の出現数の整合性は, 学生がどの選択肢を選んだかによっ

て確認することができる。ただし、チャットボットの不具合などで同一実験タスクにおいて対話を繰り返したケースとノードをスキップしてしまったケースがあった。繰り返しによって再提示されたアドバイスは、少なからず影響を与えていた可能性があるものとして含めた。全体の件数（I01～I06）の合計は本来の 54 件に対して、図3.20は合計 56 件である。しかし、その内訳は、繰り返した件数 5 件と、ノードをスキップした 3 件があった。したがって、検算すると $56 \text{ 件} - \text{繰り返し } 5 \text{ 件} + \text{スキップ件数 } 3 \text{ 件} = 54 \text{ 件}$ となる。

なお、対話パターン I02 が影響した場合、学習方略を促したのか、あるいは学習方略の状況が対話パターン I02 と一致したのかという因果関係の方向は一意に定まる。図3.21に示した方略使用増減値の算出方法から、対話パターン I02 の方が時間的に先に起きているからである。また、図3.21の振り返り記録入力 C_i が A_{i+1} の SRL06 に影響している可能性も考えられたため、順位相関を調査したところ、影響は小さかったことが示唆された。なぜなら、統計的に有意ではなかったものの、 C_i に関する 8 項目と A_{i+1} の SRL06 の相関係数の絶対値が最大で $+0.246, p = 0.085 > 0.05$ であった。

表3.11は、図3.15の条件 1, 2 の結果を示している。表3.12は、図3.15の条件 3 から結果（対話パターン出現数 > 5 , 相関絶対値 $> 0.3, p < 0.05$ ）を示している。表3.13は、学習方略間の相関を示している。

表3.13からは、新しい仮説として SRL04（課題活動を行う際、計画に基づいて実行する）と SRL06 の弱い正の相関に着目した。表3.11の SRL04（課題活動を行う際、計画に基づいて実行する）は、条件 1：統制群間の比較について、本研究ではサンプルサイズが少ないために 5% 有意水準にて有意な差がなかったものの対照群平均と比較すると実験群平均に $+8.73$ ($n = 21, p = 0.034 < 0.05$) の差があった。表3.11の条件 2：実験群の初回と最終回の比較は、5% 有意水準にて有意な差の増加 ($+0.6, n = 10, p = 0.024 < 0.05$) が認められる。表3.12の条件 3：対話パターンとの相関は認められなかったが、表3.13に示された実験群の SRL 項目間の相関について、SRL06 と SRL04 に弱い正の相関 ($0.288, n = 54, p = 0.034 < 0.05$) があった。この結果は、SRL06 と I02 にも相関があったことから、SRL04 の増加は対話パターン I02 と間接的に関係する可能性に着目した。

その他、SE01（課題がしっかりできると思う）と SE02（課題を進めていく自信がある）については、条件 1 に有意な差が認められたものの条件 2 では実験群の結果に有意な差がなかった。また学習方略に関する実験アンケート項目（SRL01, 02, 03, 05, 07）、自己効力感に関する実験アンケート項目（SE03, 04）についても条件 1, 2 において有意な差のある結果が示されなかった。

仮説 3-2 に関する実験結果

以下の実験結果から対話パターン (I02) のアドバイスを採用したと示唆された事例が 1 件あり、6 件についてはアドバイスが提示される前からアドバイス内容を使用できていた可能性があったことがわかる。

実験群 54 件の振り返り記録について「目標」の記述を調べた。「目標」を対象としたのは、目標以外のアドバイスに該当すると判定できた記述を見つけられなかったためである。なお、「目標」に対応する対話パターン (I02) のアドバイス内容は「課題の目標をもっと具体的に、段階的に取り組んでいけるように設定するといいかもかもしれません。」であった。

表3.14は、振り返り記録の記述例を示している。アドバイスを該当する、あるいは、該当しないと筆者が主観に基づいて判定した例である。提示されたアドバイスに基づき、記述されている内容が具体的であるか段階的であるかを判定したことがわかる。

表3.15は、振り返り記録について、アドバイス提示の有無と該当する記述の判定についての集計結果を示している。一見すると、アドバイス提示がなく該当する記述がある 48% に対して、アドバイス提示があり該当する記述がある 70% に効果があったように見える。しかし、アドバイス提示前から該当する記述ができていた可能性の高いケースも含まれるため、これだけではアドバイスの有効性を断定できない。この点は今後の更なる検証が望まれる。

表3.16は、アドバイス提示によって、記述内容に変化があったと判定できる事例を示している。1 回目はアドバイス提示がなく、具体的な目標になっていないが、2 回目はアドバイス提示があり、1 回目の記述と比較して具体的になっていたことがわかる。なお、振り返りの記述内容について、教員は指導を行っていない。

3.8 考察

第3章の実験結果は、ある程度、積極的かつ肯定的にチャットボットを活用しようとする学生の傾向を示していた可能性がある。なぜなら、実験タスクの実施回ごとに、タスクを実行した学生数が減少していたが、その一方、追加で実験タスク実施が見られる積極的な学生も見られたためである。積極的かつ肯定的にチャットボットを活用したとする理由は、追加実施した学生は元々チャットボットの活用に興味強い学生であったと思われるためである。タスクを実行した学生数が減少した理由は、COVID-19 対策による隔週の遠隔授業などイレギュラーな授業方法によって、欠席等が増え、実験タスクの実施を徹底できなかったためである。

3.8.1 対話が破綻しない自動的な対話機能（機能 1）に関する考察

破綻しない対話を実現する目的 1 について、機能 1 は、過半数の対話について達成することができた。機能 1 の対話が破綻せずに成立することは、チャットボットに実装した機能 2、機能 3 の最低限の条件であった。そのため、過半数の対話については、対話が成立する最低限の条件は満たしており、ネガティブな影響はなかったと考えられる。

なお、編集する前の訓練データを用いた場合との定量的比較は行っていない。定量的には比較できていないが、対話の不自然な箇所を見て人為的な編集を行った結果、対話の自然さが改善していったという経緯から、定性的には人為的な編集の効果であるといえる。

また、対話アンケートの一部は、日本語としての自然さや不自然さを与えたかどうかの評価であった可能性がある。なぜなら、実験に用いたテキストデータに誤字が含まれていたことで、対話の流れの自然さよりも日本語としての自然さや不自然さの評価を意識させてしまっていたと考えられるからである。

分析途中の研究成果 [1] は、チャットボットの発話と学生の選択に対する返答が自然な流れであると感じさせることができていた可能性のみを示している。しかし、図3.13の内訳を分析したところ、発話の流れに自然さを感じさせていた対話パターンに偏りがみられた。そこで、表3.8、表3.9、表3.10から、対話パターンと対話の流れの自然さ、不自然さについて分析を加えた。

自然さを感じさせていたとする対話パターンでは、学生が課題をうまく進められている学習状況に対して、チャットボットの抽象度が高く肯定的な発話内容との関係があると考えられる。なぜなら、学生が自然であると感じ取れていたのは、チャットボットの抽象度が高い肯定的な発話を、学生が都合よく解釈して自身の学習状況に当てはめることができたからだと考えられる。たとえば、学生が自然であるとしたとき回答したときと正の相関がある対話パターン（H05, C01, J06, L05, D01, H04, I04）は学生がうまく課題を進められていると回答したときの対話であり、対話パターン（H05, C01, J06, L05）は、テキストの抽象度が高い。

一方、ネガティブな学習状況では、有意な相関ではないものの不自然と回答していた事例があった。なぜなら、ネガティブな学習状況に、抽象的なアドバイスや、簡潔ではないアドバイス、他者や教員に相談するなどの方法を示すなど、チャットボットが学生の状況に十分な対応できていなかったことが影響していたと考えられる。

表3.10に示された、対話の流れに不自然さを感じていたとする回答については、同じ質問を繰り返すのは、チャットボット不具合からの再起動による対話途中からの再開時において、同じ質問とアドバイスを繰り返す仕様に起因したのと考えられる。

テキストデータの言葉が硬いとの意見については、表示されたテキストデータが教員的な

言葉になっていたためであると考えられる。これは学生が想定したチャットボットの立場に対して、教員の言葉から感じられる立場の違いの差が不自然さを感じさせたものと考えられる。

以上の考察から、誤字を無くすことはもとより、学生自身が自身の状況を判断できない場合やネガティブな状況にも十分な対応ができるよう訓練用の事例データとテキストデータを改善する方法を今後の課題とした。

3.8.2 自動的に学習状況の聞き取りを行う対話機能（機能 2）に関する考察

学生が学習状況の振り返りを記述できるようにする目的 2 について、機能 2 は振り返り項目に沿った記述を増やすことができておらず、学生の自己評価も教員の学習評価との乖離を軽減することができなかった。

今後の改善のために、仮説 2-1、仮説 2-2 が棄却された理由について考察を行う。実験群の実験タスクでは、振り返りの項目について、自動的に学習状況の聞き取りを行う対話機能を実装した。一方、対照群では、チャットボットの代わりに、従来手法で用いられる入力フォームへの計画入力があった。学習方略の計画入力を記述するプロセスがあるとき、振り返りの際にも一定程度の記述が可能である。

対照群は、授業開始時に学習計画を記述し、授業終了時に振り返り記録を記述している。一方、実験群は、授業開始時にチャットボットが学習状況を聞き取りし、授業終了時に振り返り記録を記述している。

チャットボットによる学習状況の聞き取りでは、例えば「課題を進めていく自信はありますか？」の問いに対して「とてもよくできそう」から「まったくできそうにない」までの 5 段階の選択肢が提示されていた。そのため、学生は、選択肢を選ぶだけでは、学習計画を記述するほどに学習状況を具体的に客観視できなかったと考えられる。その結果、振り返りの項目にうまく記述できないことに加え、教員の学習評価との乖離についても軽減することができなかったと考えられる。

図3.14の 3 回目は統制群の差が大きく示された例であると思われる。3 回目は、授業課題を学生各自でリファインする初回であった。実験群の学生は自身の学習状況（課題状況）に基づいて何をリファインするべきかについて具体化することができなかった。そのため、振り返りの記録でもうまく記述することができなかったと考えられる。なお、4 回目は、対照群の平均値が下がり、実験群が上がっている。その理由は、対照群はリファイン作業が終わり、すべきことが希薄になったため記述が減った。一方、実験群は対照群より 1 回遅れで課題状況を理解し、すべきことを具体化できたため記述が増えたと考えられる。

対照群は学習計画を検討して吟味した後に、課題を遂行し、振り返り記録を記述してい

る．そのため、対照群は学習計画を記述するときに自身の学習状況（課題状況）の理解が進められ、振り返り記録を具体的に記述できたと思われる．このことから、チャットボットとの差は、記述による学習状況の吟味があったかどうかにあると考えられる．つまり、チャットボットが聞き取りをするだけでは学習状況を吟味させるには不十分であることがわかった．したがって、チャットボットが学習状況を聞き取るだけでなく、記述の支援もできる機能設計が今後の課題となる．

そこで、今後の課題として、学生が自身の学習状況を具体的かつ客観視して捉えられるようにする支援について、2つの改善案を検討した．1つ目は、学生が学習状況を客観視できるように聞き取りを行う訓練データを編集することであり、2つ目は、学習状況を吟味させるために、聞き取りに文字入力を組み合わせ、さらに記述を支援する方法である．

3.8.3 自動的な学習方略のアドバイス提示機能（機能3）に関する考察

仮説3-1についての考察

学生が学習方略行動を工夫できるようにする目的3について、仮説3-1は、学生が目標達成の見込みに「どちらかといえばできそうでない」と回答したときに限定すると、課題プロセスの経過を把握する方略を使用したと自己申告した割合を増やすことができた．

なお、実験タスクにおける振り返り記録の入力の影響を受けてないと仮定した場合である．振り返り記録の影響の有無については、振り返り記録とSRL06との相関が判明しなかったために影響が無いと仮定した．振り返り記録の影響を考慮した自動対話の評価方法について今後の課題とする．

しかし、機能3が焦点とした学習方略の大半については目的を達成できていないと考えられる．そこで、図3.3に示した4つの相互関係過程との関係性を用いて、詳細な考察を加える．以下に、実験アンケートの項目間が元々もつ関係について、図3.3に示した4つの相互関係過程の定義との関係を示した．

1. 自己評価とモニタリング

SRL02：課題を遂行する自信を持っている

SRL06：自身の課題のプロセス経過を把握している

2. 目標設定と方略計画

SRL01：課題活動を行う際、段階的な目標設定を行う

SRL08：課題がうまくいかないと感じたとき、工夫している

3. 方略実行とモニタリング

SRL05：自身の課題の現在の状況を絶えず把握している

SRL04：課題活動を行う際、計画に基づいて実行する

4. 方略結果のモニタリング

SRL07：自身の課題について振り返りを行う

SRL03：自身の課題内容に興味があり，意欲的に取り組んでいる

学生の自己申告に自身の課題プロセス経過の把握（SRL06）が増えた理由については，以下の 2 つが考えられる．

実行可能な学習計画の模索を試みた： 図3.3が示した自己評価とモニタリングから目標設定と方略計画につながる関係に着目した．対話パターン（I02）によって，目標の具体化と段階化アドバイスが提示された．そこで，学生は目標設定と方略計画を行うにあたり，次に目指す実行可能な段階を考えるために自己評価とモニタリングが起きた．このとき，学生が現在の学習状況に至るまでの自身の課題プロセス経過の把握方略（SRL06）を行おうとして増加がみられたと考えられる．

学習状況の聞き取り（機能 2）による影響があった： 学習方略の状況について聞き取りされたことが，自己評価とモニタリングに影響を与え，SRL06 の増加に役立った可能性が考えられる．なぜなら，偏順位相関分析により SRL06 の増加にチャットボットの聞き取り（機能 2）の影響があったことが示唆されたためである．

ただし，対話パターン（I02）のアドバイスを採用したとする「課題活動を行う際，段階的な目標設定を行う方略」（SRL01）の増加は確認できなかった．つまり，実際にはアドバイスを採用していないという可能性を示唆している．アドバイスの採用については仮説 3-2 にて考察する．

対話パターン（I02）に限定すると効果があった理由については，事例データで想定した学習状況が実際の学習状況とうまくマッチしたからだと考えられる．なぜなら，図3.20に示したように学生の回答によって提示されるアドバイスが異なるからである．つまり，対話パターン（I02）以外で効果がなかったのは，事例データに含めた学習状況の想定が不十分であったと考えられる．

その他の学習方略の増加が見られなかった理由は，以下の 2 つが考えられる．以下の 2 つの理由によって，学生は前提となる学習方略とモニタリングがうまくできておらず，アドバイスの提示があっても，学習方略をうまく工夫するに至らなかったと考えられる．

学習方略項目を個別に扱っていた： 学習方略項目を個別に扱い，相互に関連付けた前提となる学習方略のアドバイスを提示することはなかった．その結果，学生は，学習方略をうまく検討することができなかったと考えられる．

学習状況の聞き取り（機能 2）が不十分だった： 「学習状況の聞き取り（機能 2）による影響が

あった」とする考察に関係して、図3.3の学習方略項目の相互関係における前提となるモニタリングが十分に支援できていなかった影響があったと考えられる。

なお、表3.12の実験結果では、対話パターンと学習方略（SRL）・自己効力感（SE）の相関のみに着目したとき、対話パターン I02 と SRL06 以外にも正の相関があったことを示している。振り返り記録入力の影響を考慮する必要があるものの、チャットボットが学習方略に少なからず寄与していたことを示唆していると考えられることができる。

表3.13で着目した SRL06 と SRL04 の正の相関関係については、新たな仮説の探索において考察を加える。

仮説 3-2 についての考察

表3.16に示された 1 件の事例については、対話パターン（I02）のアドバイスの提示によって目標を具体化できたと考えられる。その理由は、1 回目はアドバイスがなく、学生は目標を抽象的に記述していたが、2 回目に目標を具体化、段階化するというアドバイスの提示があったため、比較的具体的な記述になっていたと考えられる。

表3.15から数字上は、70% の 7 件について効果があったように見える。しかし、母数が少ないことに加え、6 件についてはアドバイスが提示される前からアドバイスに該当する記述できていた可能性がある。

なお、仮説 3-2 は有効性を断定できない。なぜなら、仮説 3-2 に該当する事例が示唆されたのは 1 件のみであり、仮説にあてはまる事例がたまたまあった可能性が高いためである。母数を増やした実験によってさらに検証することを今後の課題とした。

仮説 3-2 に該当する事例が 1 件しかなかった理由と、3 件のアドバイスが提示されたにもかかわらず採用しなかった事例、そして、23 件のアドバイスが提示されず、該当する記述もなかった事例の理由を以下のとおり考察した。

学習状況に沿う具体的なアドバイスがなかった： 目標の具体化や段階化のために必要な前提となる知識や具体策について支援を行っていない。そのため、目標の具体化や段階化を実施することができなかつたと考えられる。たとえば、自分が取り組んでいる課題の作業や内容を分解して考えられるほど十分な理解ができておらず、目標の具体化や段階化を実現するに至らなかつた可能性がある。

なお、仮説 3-2 にも、機能 2 の影響があったものと考えられる。

そこで、学生個別の実効性や課題を捉え、関連する知識や具体策のアドバイスを学習方略と共に提示できる機能設計を今後の課題とした。

3.8.4 新たな仮説の探索

分析途中の研究成果 [1] で示されたように、第3章では機能 2 と機能 3 について、全体的な効果を明らかにすることができなかった。そこで、機能改善のために実験データについて焦点を絞り込んだ新しい仮説の探索を行い、今後の課題として新しい仮説を示した [1]。

対話シナリオにおいて、対話パターン I02 が SRL06 を促すとき、意図的に SRL06 と SRL04、または、SRL04 と SRL06 の関係性を学生に認知させる対話と途中経由する目標設定と方略計画、方略結果のモニタリングを支援する対話を加える。すると、SRL06 と SRL04 の両方の学習方略を促せるという新たな仮説が立てられる。なぜなら、表3.13で着目した SRL06 と SRL04 は、正の相関関係が示されていた。図3.3の SRL 項目が元々もつ項目間の関係性によると、自己評価とモニタリング（SRL06）が目標設定と方略計画を経由して、方略実行とモニタリング（SRL04）を増やす、あるいは、方略実行とモニタリング（SRL04）が方略結果のモニタリングを経由して、自己評価とモニタリング（SRL06）を増やすという関係を示しているからである。

これらの方略項目間の相互関係については、以下の 2 つの自己調整学習研究において指摘された関係性にあてはめて検討することができる。

適切な方略を選択したり遂行をモニターしたりすることができるようになって、自己調整が生起してくる [20, 21]

生徒の自己効力感、目標設定、自己モニタリング、自己評価、方略使用などの自己調整過程の実行を媒介して、学力にも影響を及ぼす [22, 21]

表3.17は、分析途中の研究成果 [1] の TABLE III を著者が修正した表である。

分析途中の研究成果 [1] では、次に改善できそうな仮説を示した実験アンケート項目として、弱い正の相関のある表3.17の 4 項目を挙げていた。

そこで、表3.17に示された 4 項目の支援に焦点を当て図3.3に示した 4 つの相互関係に基づいた以下の機能設計を行うことを課題とする。チャットボットが学生に、これから先、この課題内容が得意であると思わせる（SE03）ようにアドバイスを行うことで、図3.3の方略実行とモニタリング（SRL05）、方略結果のモニタリング（SRL07）、目標設定と方略計画（SRL08）の学習方略を促進することができる可能性がある、もしくは、チャットボットが方略実行とモニタリング（SRL05）の支援、方略結果のモニタリング（SRL07）の支援、あるいは、目標設定と方略計画（SRL08）の支援となるアドバイスを行うことで、学生にこの課題内容が得意であると思わせる（SE03）という自己効力感を高められる可能性を考えた。

3.9 結言

第3章で開発した学習支援システムのチャットボットは、自動的な聞き取りとアドバイス提示により、学生個別の自己調整学習を自動支援することを目的としている。実験結果からは、チャットボットが、一定の条件下でのみ、学生の学習方略の一部について、利用を促進する効果があることが示唆された。しかし、全ての学習方略に対して同様の効果が得られたわけではなく、特定の条件下でのみ効果が観察されていたことが明らかになった。

第3章では、デザイン学習を学ぶ学生の学習方略について、振り返りと工夫の支援を目的としたチャットボットの機能を試作して評価した。目的と機能の評価は、以下のように結論した。

目的1「破綻しない対話を実現する」は、機能1の破綻しない対話によって、過半数の対話では破綻の無い自然な対話を実現し、目的を達成していた。ただし、学習状況がポジティブでない場合の対話シナリオが十分ではなかった。そこで、誤字への対策を含め、対話シナリオの編集方法の改善について今後の課題とした。

目的2「学生が学習状況の振り返りを記述できるようにする」は、機能2の学習状況の聞き取りによって、従来手法（入力フォームに学習計画を書かせる）よりも自己モニタリングを促すことができず、目的を達成できなかった。改善点として学習状況の聞き取りを深掘できるような訓練データを増やすこと、そして、文字入力を組み合わせ、記述を支援する方法を今後の課題とした。

目的3「学生が学習方略行動を工夫できるようにする」は、機能3のアドバイス提示と機能2の学習状況の聞き取りの併用によって学生が目標達成の見込みに「どちらかといえばできそうでない」と回答したときに限定すると、課題プロセスの経過を把握する方略を使用したと自己申告した割合に増加がみられた。さらに、アドバイスが提示されたことで学習方略を工夫できていたと示唆された事例が1件あった。しかし、アドバイスが目標の具体化、段階化に限定され、母数が少ないことに加え、6件についてはアドバイスが提示される前から工夫できていた可能性があった。そのため、母数を増やした実験によってさらなる検証を加える必要があった。また、目標が具体化、段階化できていないにもかかわらずアドバイスの提示が無かった事例、アドバイスの提示があっても、うまく適用できていない事例があった。そこで、学生個別の実効性や課題を捉え、関連する知識や具体策のアドバイスを学習方略と共に提示できる機能設計を今後の課題とした。

第3章は、焦点とした大半の学習方略について効果を明らかにすることはできなかった。そこで、新たな仮説探索によって焦点を当てた学習方略と自己効力感の関係を分析した結果、以下の2つを今後の課題として定めた。1つ目は、対話パターン I02 と SRL06, SRL04 の2項目間について、図3.3に示した相互関係に基づいた対話シナリオ（訓練データ用事例

データとテキストデータ)の作成。2つ目は、SE03, SRL05, SRL07, SRL08 の 4 項目の支援に焦点を当て図3.3に示した相互関係に基づいた機能設計である。

なお、第3章の結果は、チャットボットの利用に積極的な学生に見られる傾向であった可能性がある。チャットボットの利用に興味を持ち、積極的に扱おうとする学生と、実験タスクを実施しなくなった学生の比較分析などは、今後の課題とした。

第3章の実験は、被験者数が少なくサンプルサイズも少ないこと、被験者の事前の学習方略を扱うスキルなどについても詳細にコントロールしていなかったこともあり、これまでに示した今後の課題の改善後、実験計画を改善し、よりも多くの被験者で検証を行う必要があることがわかった。

第3章の今後の課題に参考となる先行研究として、チャットボットによる振り返り支援機能の研究 [23] がある。具体的には、第3章の目標設定の支援が不十分であった点について、石川らの研究 [23] では、学習者の学習計画と学習状況を記録し、集約して表示するダッシュボードのような支援機能があった。これにより、学習者が学習計画において、より実行可能性の高い学習時間を設定できている可能性が示された。よって本研究でも同様のダッシュボード等を用いた振り返り支援機能の導入を検討したい。また、本研究における振り返り支援の影響評価する実験方法として、[23] のチャットボットの利用とダッシュボードの利用で統制群を分ける方法についても参考にできる。

その他、第3章の実験と評価方法について、多くの改善点が明らかになった。以下の点も今後の研究課題とする。学生が実験タスクを実施しなかったことの分析。学習方略の段階に応じて実験アンケートを回収するタイミングの改善。さらに、COVID-19 対策で行った授業方法など、チャットボット以外の影響について精査する方法である。

以上のことから、第3章では、あらかじめ定めた決定木と定形文を用いた自動的な聞き取りとアドバイス自動提示の手法について限界を認識した。したがって、表 2.1に示した学習支援システム研究の項目 1 において、課題遂行段階に対する学習スキルの自動助言については、一定条件下のみの限定的な効果に留まった。

特に、教員が想定した決定木と定型文だけでは、学生の学習状況との乖離がみられた。学生の学習状況との乖離を克服するため、第4章と第5章では、記述された学習状況との関連性を扱うことができる大規模言語モデルを用いた。第4章では、大規模言語モデルを用いて学生の記述内容を自動評価し、第5章では、記述内容に基づいたアドバイスの自動生成を導入する。

表3.1: 聞き取りとアドバイスの実例

種類と機能	聞き取りとアドバイス（学生の回答選択例）
	今日の課題達成目標について今の達成見込みを教えてください。 (どちらともいえない)
①自己モニタリング 聞き取り	課題の段取りがうまく定まっていなかったかもしれませんね。 課題を進めていく自信はありますか？ (まったくない) 大丈夫ですよ。課題の取り組み方に状況について私と確認しながらやり方を変えてみましょう。
①自己モニタリング アドバイス	何か、うまくいかない、うまく進められない原因があると思います。それがわかれば、そのことをクリアにする方法を探せばいいのです。教員にも遠慮なく相談しましょう。
②目標設定 アドバイス	今日達成したい課題目標が遠すぎるかもしれません。課題制作の完成とは別に、とりあえず、今日の時間にどこまでやっておくかを決めておく区切りがつき、メリハリのある取り組みや時間の使い方になるでしょう。もし、提出までに時間がない場合は、期限までの短い時間を区切って近い目標を定めると集中しやすいですよ。
③自己教示 聞き取り	課題の内容や進め方について、説明を資料の内容を把握、理解できましたか？ (どちらかといえば理解できている) いい傾向ですね。もし、課題の内容でわからない部分があれば、課題に対する自分のアイデアについて教員に相談してください。
③自己教示 アドバイス	まずは課題で問われていることをしっかりと理解することも大切な学習のプロセスです。まずは、課題で何を問われているのかを理解するために、課題について他の学習者や教員に相談しましょう。学生の皆さんが課題に取り組む大事な前提ですので教員もしっかりと対応して下さるでしょう。
④注意の焦点化 聞き取り	いまのところ私との対話以外では課題の取り組みには集中出来ていますでしょうか。 (どちらかといえば集中できている) まあまあ、集中できているところでしたね。
④注意の焦点化 アドバイス	自分が気を散らすことのない環境で課題に取り組んでいるようですね。適度に休憩を挟み、メリハリのある取り組みを続けていきましょう。
⑤課題方略 聞き取り	今やっている課題を進めるための方法や考え方は、自分や自分のアイデアにあった方法や考え方と思いますか？ (どちらかといえばあわない) まあまあ、集中できているところでしたね。
⑥方略アドバイス 聞き取り	今どんな方法で課題を進めているところですか？ (スケッチ操作) 今は、課題の何に取り組んでいるところでしょうか？ (人物の顔を描く)
⑥方略アドバイス アドバイス	顔を描く場合は顔の輪郭 → 耳 → 目 → 眉 → 鼻 → 口 → 髪型 → 服装の順に描いていくといいでしょう。インターネットの画像検索で「顔イラスト」などで検索し、できるだけシンプルな形状を選び、パーツごとに組み合わせると特徴を出せるかと思います。自分の想定イメージに近い特定の人物がいる場合は、画像を線でトレースする方法もあります。

表3.2: 事例データの一部 [1]

変数の種類	変数名 (一部抜粋)	事例データ値 (一部抜粋)		
説明変数	課題進行状態	よく進んでいる	まったく進んでいない	あまり進んでいない
	目標達成の見込み	とてもよくできそう	まったくできそうにない	どちらかといえばできそうにない
	使用方略の適合	よくあう	まったくあわない	どちらかといえばあわない
目的変数	自己モニタリング	継続を勧める	励ます	励ます

表3.3: 事例データの変数 [1]

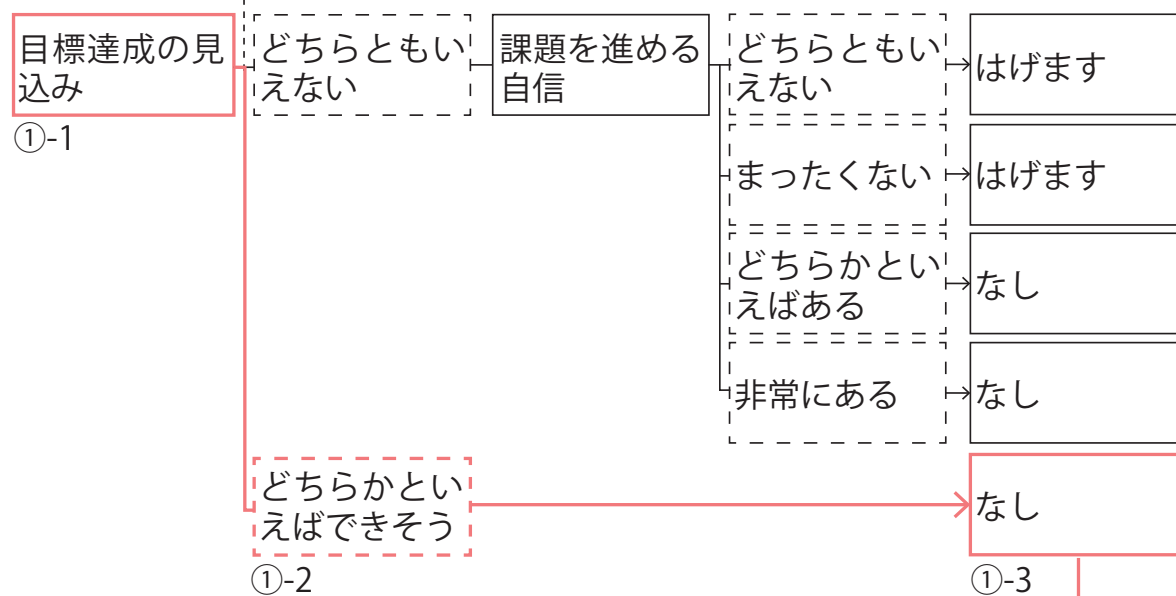
	変数名	事例データ値
説明変数	取り組み時間	数値 (分)
	課題進行状態	よく進んでいる - まったく進んでいない (5 件法)
	課題を進める自信	とてもある - まったくない (5 件法)
	目標達成の見込み	とてもよくできそう - まったくできそうにない (5 件法)
	使用方略の適合	よくあう - まったくあわない (5 件法)
	興味関心度	とてもよく興味関心がある - まったく興味関心がない (5 件法)
	負担感	とても負担を軽く感じる - まったく負担を感じない (5 件法)
	内容理解	よく理解できている - まったく理解できていない (5 件法)
	必要知識の充足	十分に足りていると思う - まったく足りていないと思う (5 件法)
	集中状況	とてもよく集中できている - まったく集中できていない (5 件法)
目的変数	自己モニタリング	褒める, 励ます, 現状維持, 様子見, 負担へのアドバイス, 方略の変更アドバイス, 原因検討, など
	目標設定	アドバイスをする, しない
	課題方略	アドバイスをする, しない
	自己教示	アドバイスをする, しない
	注意の焦点化	アドバイスをする, しない

表3.4: 方略の工夫アドバイスの変数

	変数名	事例データ値
説明変数	課題	わからない, スケッチ操作, シンボルを整える, など
	方略	わからない, スケッチの描き方を練習する, スケッチの描き方を練習する, など
	目的	課題を進めたい, スケッチを描けるようにしたい, スケッチを描けるようにしたい, など
	状況	よく進んでいる - まったく進んでいない (5 件法)
	助けてくれる人	教員, 学生, いない
目的変数	方略アドバイス	サービス構成, 人物設定の作成, スケッチ操作, 図形描画方法, 被援助要請, など

①自己モニタリング

(省略)



②目標設定

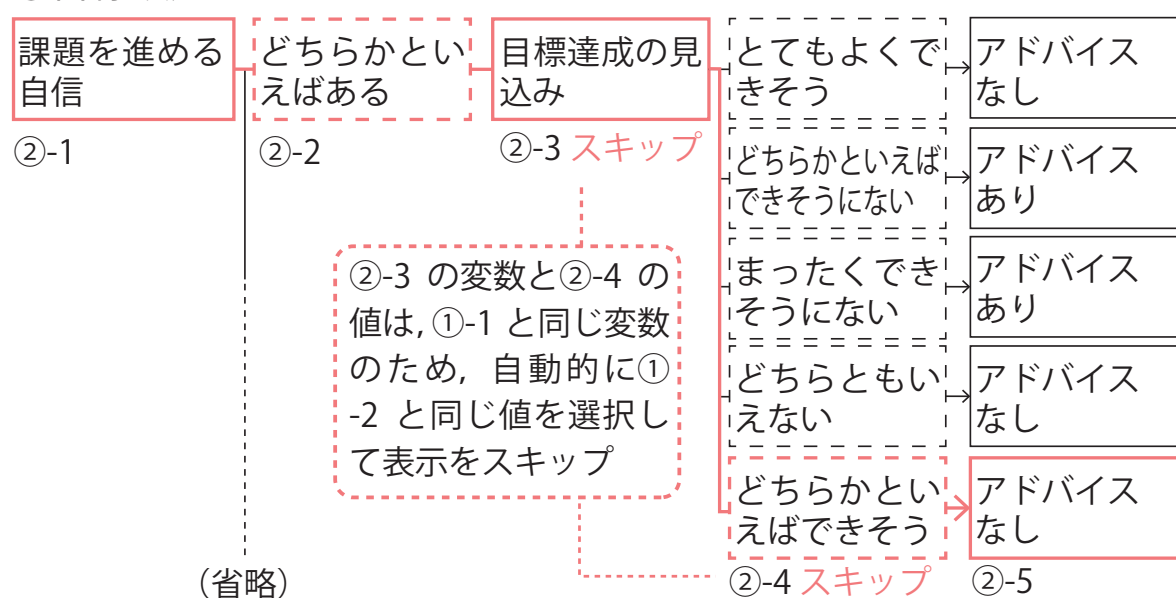


図3.9: 決定木の流れ一部

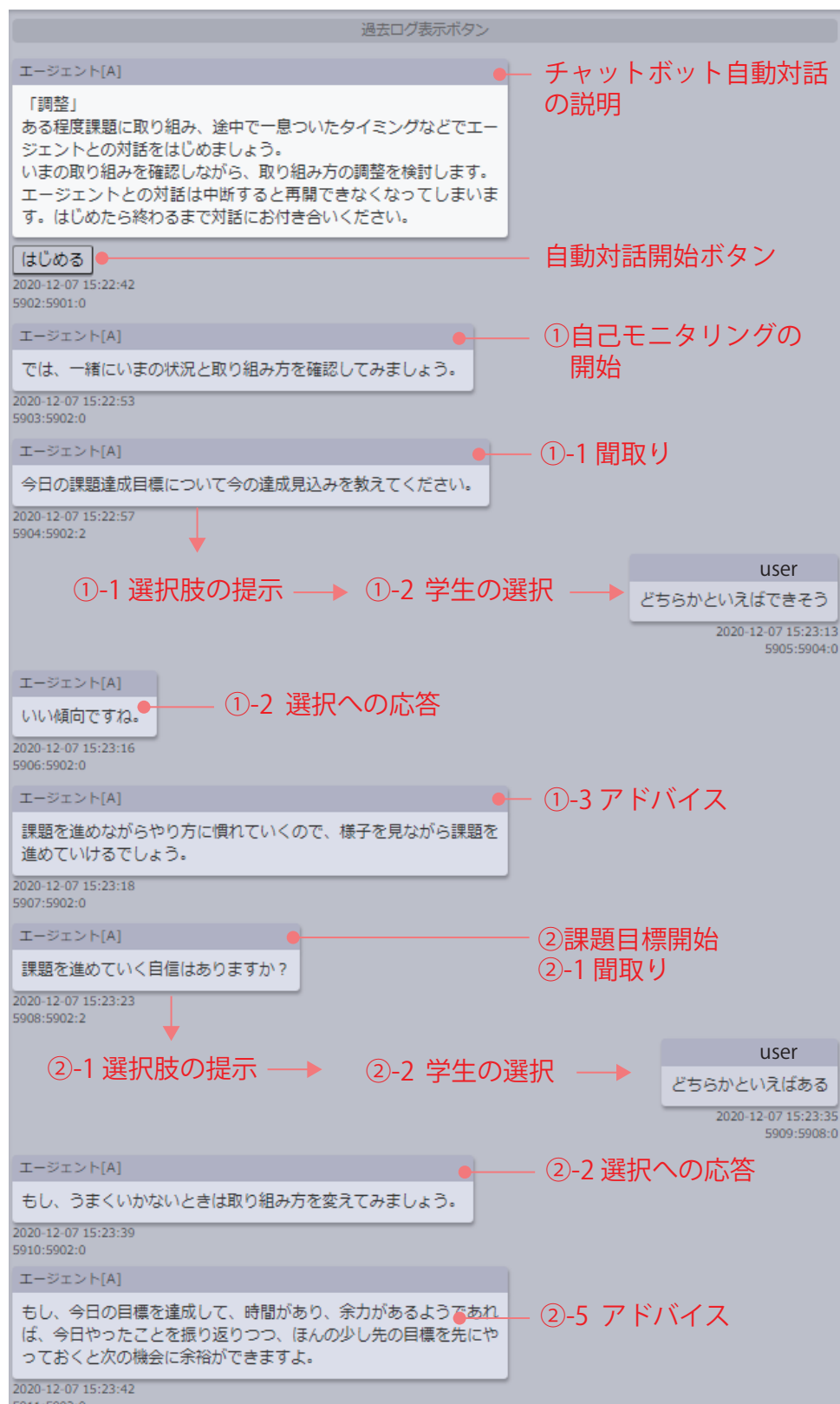


図3.10: チャットボットと学生の対話例 [1] を著者が加工して使用

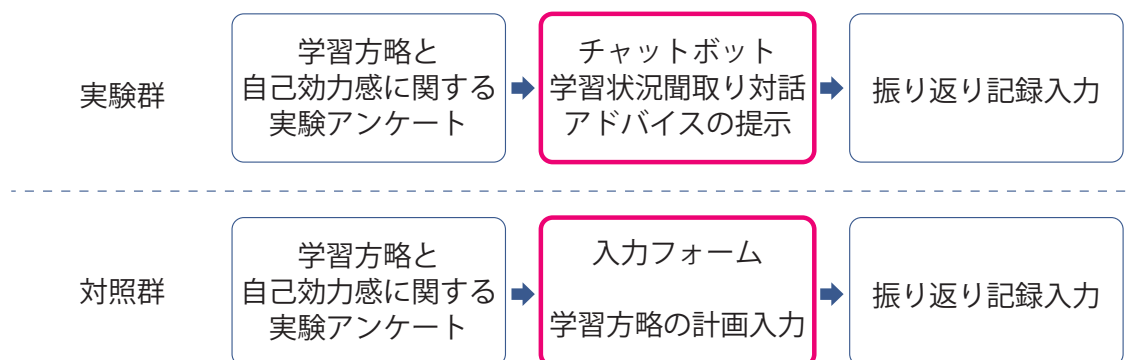


図3.11: 統制群ごとの実験タスク [1] を著者が加工して使用

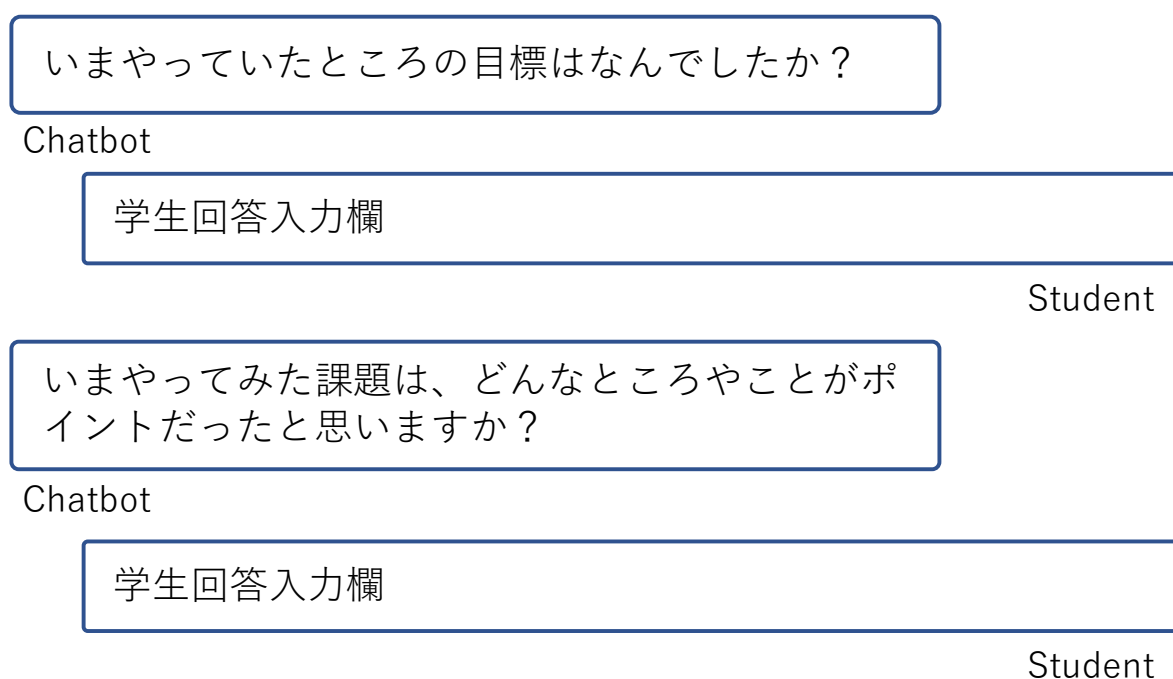


図3.12: 自動アンケート機能

表3.5: 学習方略の計画入力での質問項目

	入力項目	質問文
学習方略の計画	学習時間 (分：数値入力)	課題への取り組みにメリハリをつけると集中しやすくなりますよ。とりあえず、これから何分くらいはこの課題に取り組むことにしましょうか？分で教えてください。例えば1時間30分の場合はと入力してください。
	目標（文字入力）	いまから取り組む時間の中で到達できる段階的な目標をつけるとメリハリのある取り組みになりますよ。とりあえず、どこまでやっておきたいと思いますか？ここはやり終えておきたいとか、いくつやっておきたいかなど、なんとなくでも考えているところがあれば教えてください。
	焦点（文字入力）	課題のどういうところに着目するかを考えておくと、課題のポイントを捉えた結果に近づきますよ。今、考えているポイントやなんとなくでも気になっているところがあれば教えてください。
	方略（文字入力）	これから取り組むことや部分について、やり方や考え方、捉え方にある程度の目処をつけてから取り組み始めると、課題のやり方を色々と工夫しやすくなりますよ。今、なんとなくでも考えていることがあれば教えてください。
	集中方法（文字入力）	何か気が散るようなことに対して、あらかじめ対処方法を決めておくと集中しやすくなりますよ。何か気が散りそうなことと、それがもしそれが起きたとき、どうしますか？「〇〇が気になったとき□□しようと思う」のように描いておきましょう。
	自信（5件法） (1=まったくない, 5 = 非常にある)	どうでしょう。いま考えた課題の取り組み方について、今の自信のほどを教えてください。

表3.6: 振り返り記録入力での質問項目

振 り 返 り 記 録	学習時間 (分：数値入力)	今は何分くらいこの課題に取り組むことができましたか？分で入力をお願いします。1時間30分の場合は「90」と入力してください。
	目標（文字入力）	いまやっていたところの目標はなんでしたか？
	焦点（文字入力）	いまやってみた課題は、どんなところやことがポイントだったと思いますか？
	方略（文字入力）	いまはどんな方法や考え方などでやっていましたか？
	達成度 (%：数値入力)	自分で決めた目標の何パーセントぐらいできましたか？ (パーセントの値を数値入力、60%の場合「60」と入力してください。)
	負担感（5件法） (1=とても負担に感じた, 5 = 非常に負担は少なかった)	いまやってみた方法や考え方などは、大変でしたか？
	集中方法（文字入力）	課題をやっているときに何か気が散るようなことはありましたか？
	自信（5件法） (1=まったくない, 5 = 非常にある)	次回も課題を進めていく自信はありますか？

表3.7: 実験タスクと実験アンケート実施人数

回	実験タスク実施人数			実験アンケート実施人数		
	実験群	対照群	計	実験群	対照群	計
1	18	13	31	18	14	32
2	17	12	29	17	13	30
3	12	11	23	14	13	27
4	8	9	17	8	10	18
5	4	0	4	4	0	4
6	1	0	1	1	0	1
最終	-	-	-	10	11	21
計	60	45	105	72	61	133

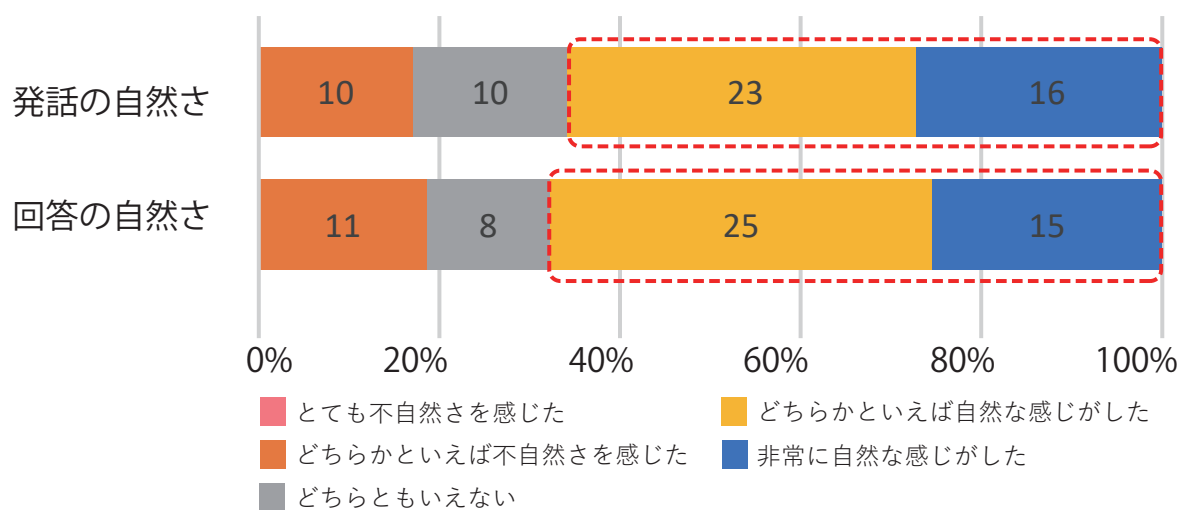


図3.13: チャットボットとの対話の流れの自然さ [1] を著者が加工して使用

表3.8: 対話アンケートと対話パターン ID の相関

項目	対話パターン ID	度数	順位相関	p 値
対話の流れ値	C01	59	0.4524	0.0003
	L05	59	0.4442	0.0004
	H05	59	0.3909	0.0022
	D01	59	0.3886	0.0024
	J06	59	0.3625	0.0048
回答への対応値	L05	59	0.5002	10^{-1} 未満
	J06	59	0.4653	0.0002
	H05	59	0.4556	0.0003
	C01	59	0.3896	0.0023
	D01	59	0.3803	0.003
	H04	59	0.3698	0.0039
	I04	59	0.3675	0.0042

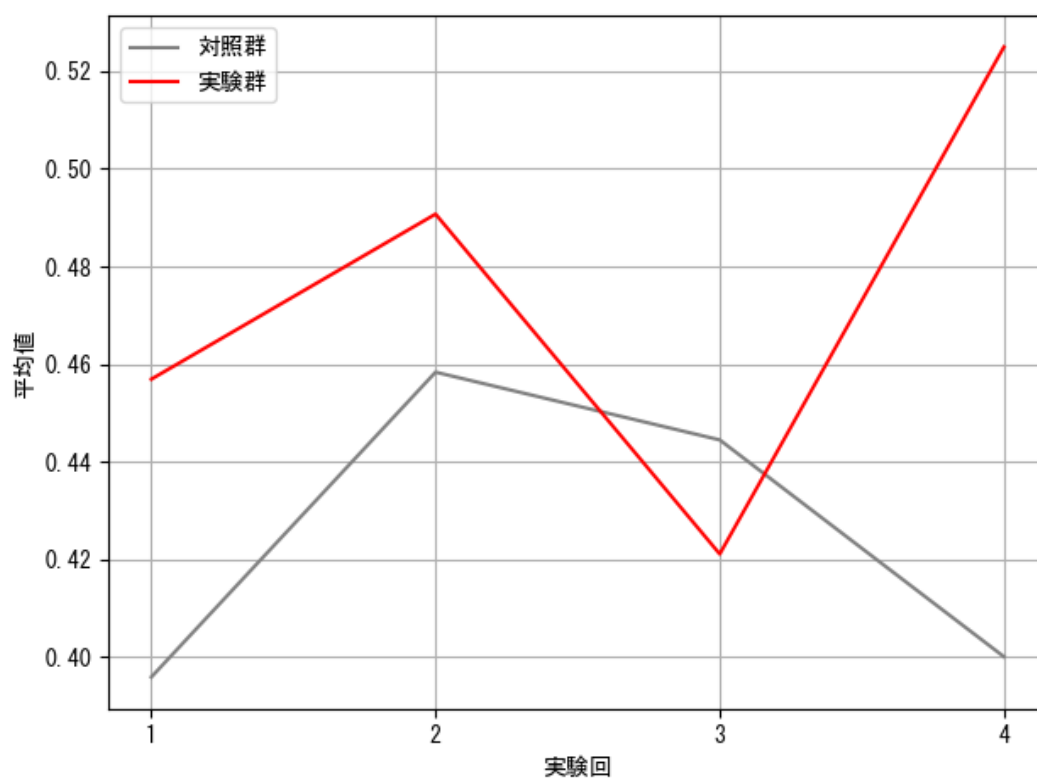
順位相関係数絶対値 >0.35 , $p<0.05$ 

図3.14: 振り返り記録のスコア平均

表3.9: 対話パターン（抜粋）

ID	チャットボットからの発話（問いかけ）	学生の回答
	チャットボットからの回答	
C01	(J06から続くリーフノードの対話テキスト) もし、課題についてわからないところや聞きたいことが出てきましたら、教員チャンネルなどで質問しましょう。	
H05	課題を進めていく自信はありますか？	非常にある
	いいですね。この調子で進めていきましょう。	
J04	課題の内容や進め方について、説明を資料の内容を把握、理解できましたか？	どちらともいえない
	課題を理解で来ているかどうか不安な場合は、他者を参考にしたり、教員に沿うd難（「沿うd難」＝「相談」の誤字）するなどをして、心配や不安を解消することが大切です。	
J06	課題の内容や進め方について、説明を資料の内容を把握、理解できましたか？	よく理解できている
	いい傾向ですね。もしわからないことがあるときは、教員に相談してください。	
L05	課題の取り組みには集中出来ていますでしょうか。とてもよく集中できている (自動)対話が(集中の)お邪魔したようすいません。もう少しお付き合いください。	

文中()は補足説明, 実験時には表示していない

表3.10: 特に不自然さを感じたところの集計

記述内容	件数
同じ回答が繰り返された	2
同じ質問が繰り返された	3
違和感のある回答があった	1
言葉がかたい	2
言葉が理解しづらい	1
選択肢の順序が毎回異なる	1
褒められるとやる気が出る	1

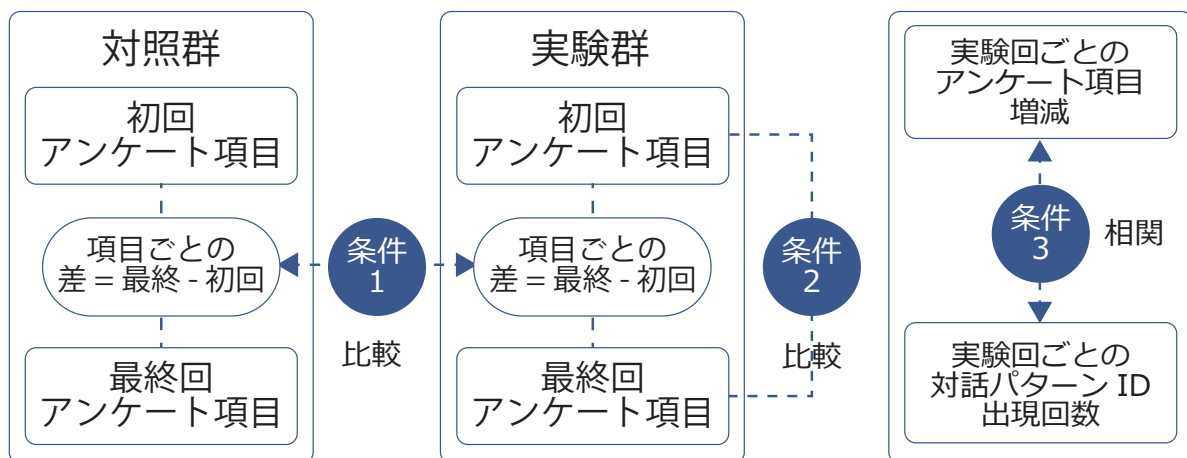


図3.15: 学習方略の分析方法

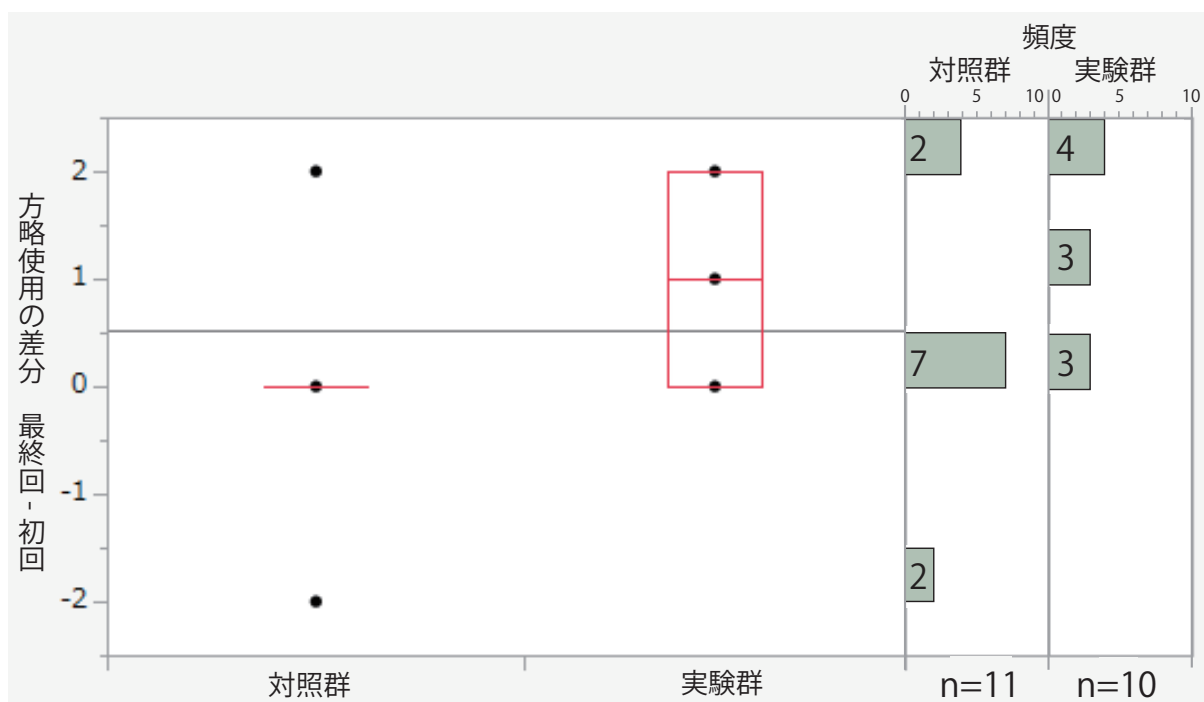


図3.16: 課題プロセス把握方略（SRL06）の統制群比較 [1]

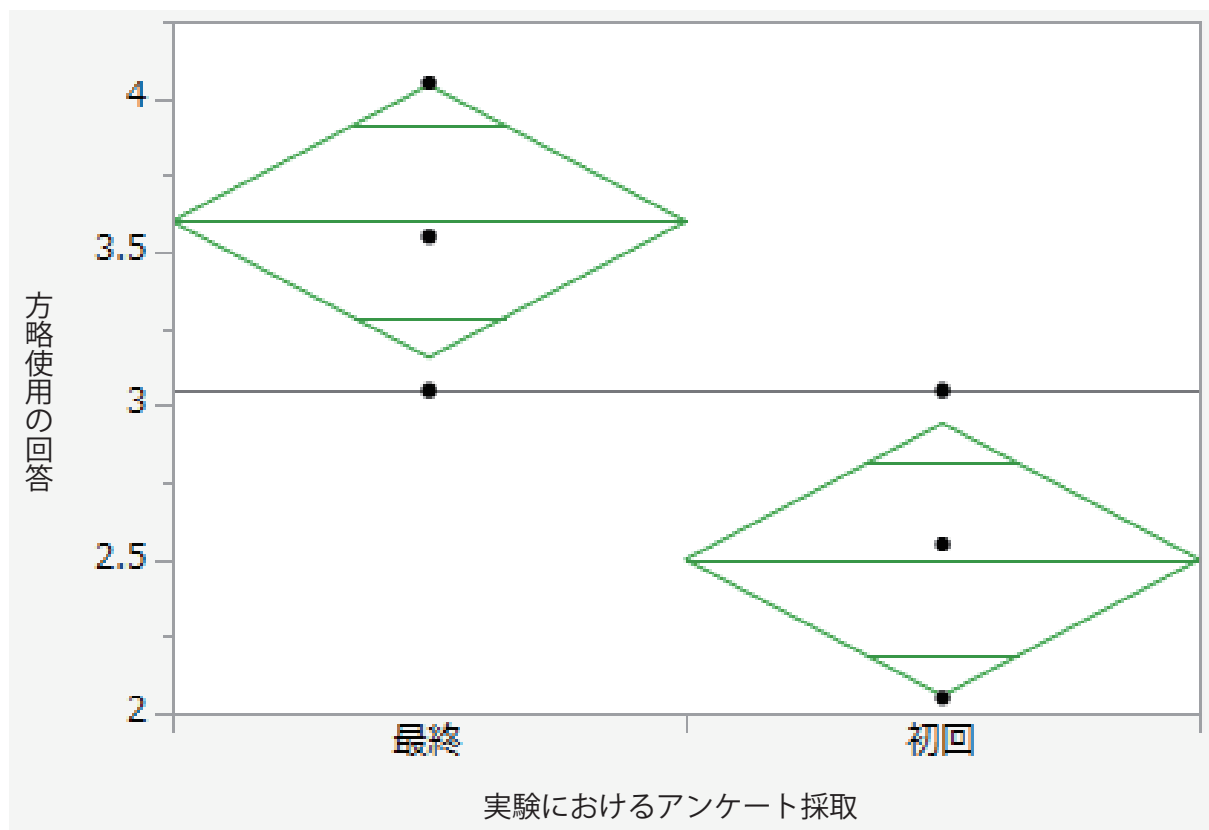


図3.17: 課題プロセス把握方略使用の初回 - 最終回比較 [1]

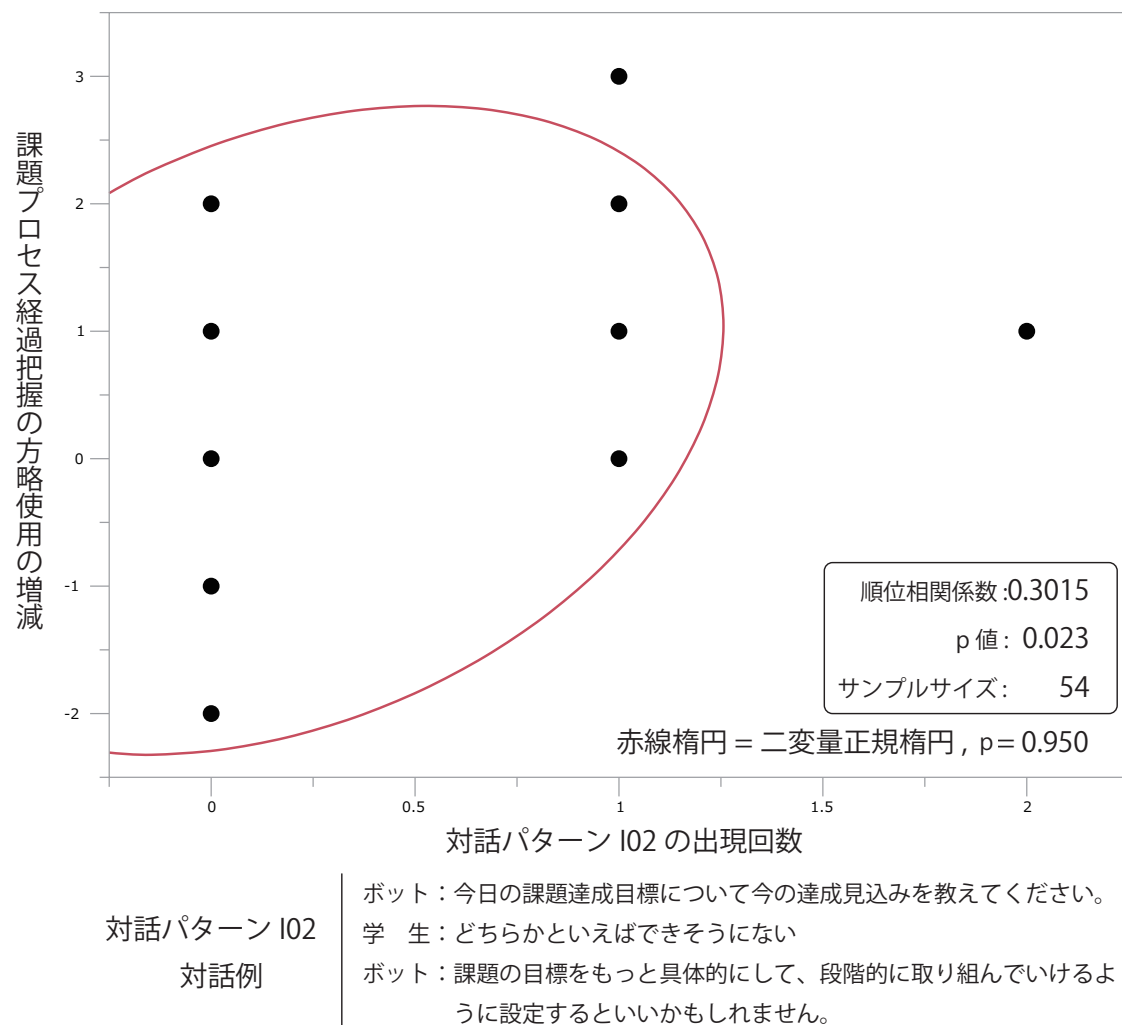


図3.18: 方略使用増減と対話パターン I02 出現回数の散布図 [1]

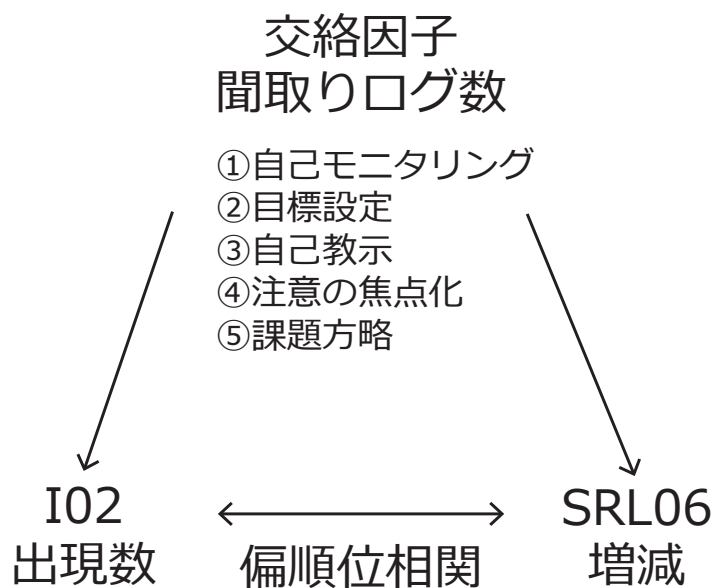


図3.19: 方略使用増減と対話パターン I02 と交絡因子

①自己モニタリング

ノード	選択肢	対話パターン	カウント
目標達成の見込み	まったくできそうにない	I 01	6
	どちらかといえばできそうにない	I 02	10
	どちらともいえない	I 03	14
	どちらかといえばできそう	I 04	24
	よくできそう	I 05	1
	とてもよくできそう	I 06	1

(不具合などで対話を2回実施した際の5件含む) 計56件
不具合で未表示になっていた3件

図3.20: 対話パターン I02 出現数の算出方法

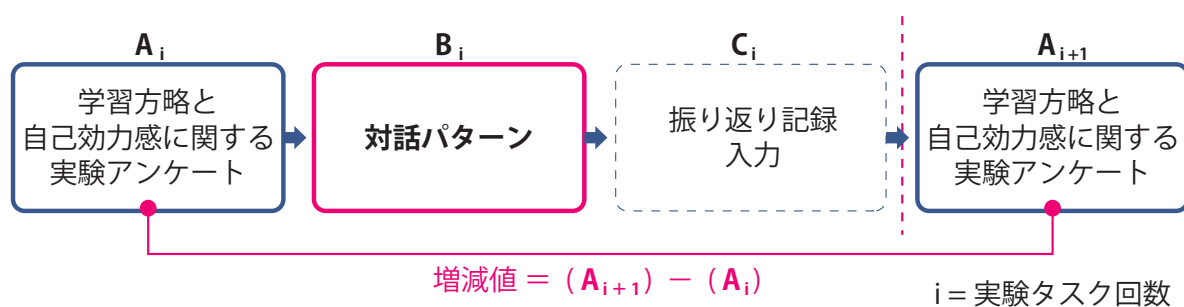


図3.21: 方略増減値の算出方法

表3.11: 条件 1: 統制群間比較と条件 2: 実験群最終初回比較

	条件1: 統制群間比較			条件2: 実験群初回最終回比較		
	対照群平均 (最終-初回)	実験群平均 (最終-初回)	p値(Prob> Z)	初回平均	最終回平均	p値(Prob> Z)
SRL01	0.455	0.3	0.586	2.7	3.0	0.394
SRL02	0.091	0.0	0.664	3.0	3.0	1.000
SRL03	0.727	0.3	0.768	3.4	3.7	0.560
SRL04	-0.273	0.6	0.223	2.9	3.5	0.024
SRL05	-0.273	0.1	0.625	3.6	3.7	0.798
SRL06	0.0	1.1	0.034	2.5	3.6	0.003
SRL07	0.0	0.5	0.274	2.4	2.9	0.138
SRL08	-0.273	0.2	0.168	3.4	3.6	0.343
SE01	-0.909	0.3	0.008	2.7	3.0	0.193
SE02	-0.727	0.3	0.013	3.4	3.1	0.279
SE03	0.0	0.5	0.631	2.4	2.9	0.299
SE04	0.091	-0.2	0.577	2.8	2.6	0.443

表3.12: 条件 3: 対話パターンとアンケート項目の相関

ID	出現数	アンケート	度数	順位相関	p 値
L02	7	SRL07	54	0.3387	0.0122
I02	10	SRL06	54	0.3105	0.0223
H03	9	SRL05	54	-0.3185	0.0189
ボットの問いかけ			ユーザー選択		
L02	いまのところ私との対話以外では課題の取り組みには集中出来ていますでしょうか。			どちらかといえば集中できていない	
I02	今日の課題達成目標について今の達成見込みを教えてください。			どちらかといえばできそうにない	
H03	課題を進めていく自信はありますか？			どちらかといえばできそうにない	

表3.13: 実験群の SRL 項目間の相関

変数	vs. 変数	度数	順位相関	p 値
SRL05	SRL06	54	0.5213	10^{-1} 未満
SRL02	SRL07	54	0.3235	0.017
SRL04	SRL06	54	0.2882	0.0346
SRL03	SRL06	54	0.2798	0.0404

表3.14: 振り返り記録の記述例

学 生	回	該当する記述例
D	2	エレメントの情報をわかりやすく編集する
E	1	不自然な点を見つけて対策を考えること
N	2	ペーパープロトタイプや UI の作成
学 生	回	該当しない記述例
D	1	大まかな完成
G	2	集中すること
L	2	よくわからない

表3.15: アドバイス有無と記述有無の集計

	該当する記述あり	該当する記述なし
アドバイス提示あり	7 件 (70%)	3 件 (30%)
アドバイス提示なし	21 件 (48%)	23 件 (52%)

表3.16: 記述内容に変化があったと判定できた事例

学 生	回	掲 示	記述内容
D	1	なし	大まかな完成
D	2	あり	エレメントの情報をわかりやすく編集する
D	3	あり	なるべく実在しそうな感じを出す

表3.17: 自己効力感と学習方略の相関表 [1] を著者が修正して使用

自己効力感	学習方略	度数	順位相関	p 値
SE03	SRL05	110	0.5169	10^{-1} 未満
SE03	SRL07	110	0.4406	10^{-1} 未満
SE03	SRL08	110	0.3211	0.0006
SE03	これから先、この課題内容が得意であると思う			
SRL05	自身の課題の現在の状況を絶えず把握している			
SRL07	自身の課題について振り返りを行う			
SRL08	課題がうまくいかないと感じたとき工夫している			

第 4 章

GPT-3 を用いた学習計画と振り返りの自動評価

4.1 序言学習計画と振り返り記述の評価，足場かけ，足場はずし

自己調整学習の予見と自己省察について，学習スキルの段階に沿った自動支援を実現するためには，学習者が記述した学習計画と振り返りの記述内容を捉えて，学習スキルの段階を自動評価する手法が必要である．そのために，本章では表 2.1の学習支援システム研究の項目 5 に示した，学生の記述内容から学習スキルの自動評価の実現を目指した．

自己調整学習の指導方法として，学習者に自身の学習状況を把握させ，学習方略の工夫を促す手段として，学習状況の振り返りと学習計画を記述させる方法がある．

このとき，学生個別の記述内容に応じた段階的な支援が重要である．自己調整学習の一般的な支援手法では，学習者の自己調整学習の熟達状況に応じて，足場かけや足場はずしを行う．足場はずしは，自己調整学習の習熟を促すために，学習方略を伝え行動を促す足場かけと，足場を段階的に外していく手法がある．

学生個別の記述内容に応じた段階的な自動支援を実現するためには，学習者が記述した学習計画と振り返りの記述内容を自動評価する手法と，同時に自動評価の品質を揃える手法が必要である．そこで，ルーブリックに基づいて教員が評価したトレーニングデータによって，ファインチューニングした大規模言語モデルを用いた自動評価手法を提案した．

実験では，学習者が記述した学習計画と振り返りの記述内容について，ルーブリックを用いて教員が評価した結果と，ファインチューニングした大規模言語モデルを用いた結果を比較した．

第4章は以下の 2 つを目的とする．

1. GPT-3 を用いて，具体性と躓き度について，教員の手動評価を模倣した自動評価ができる．

2. GPT-3 で自動評価した値を用いて足場はずしを自動的に設定できる。

以下に、前提となる学習支援における評価と足場かけ、足場はずしの関係と必要性を説明する。そして、自動評価を実装する開発中の学習支援システムと、第4章の目的である自動評価の必要性について説明を行う。

4.1.1 具体性と躓き度を評価する必要性

振り返りと計画の具体性、および、躓き度の評価をしなければならない理由は、学生の主体的な学習のために、どの程度の支援が必要なのかを判断するためである。

オンライン教育では、学習者が自宅などから参加して学習活動を行うほかに、参加しない時間における主体的な学習活動も重要となる。主体的な学習者は、自身の学習状況を常に把握して、自身の学習方略を自己調整しているとされる [24, 9]。つまり、オンライン教育などある程度主体的な学習が求められる学習活動では、学習者が自身の学習状況を把握して計画的に見通しを立て、学習活動を調整していくことが重要である [25]。

振り返りと計画が具体的に記述できていない学習者は、自身の学習状況の把握や学習の見通しがうまくできていない。さらには自身の躓きにも気が付いていないことがある。したがって、振り返りと計画が具体的に記述できるよう支援を行う必要がある [26, 1]。

4.1.2 足場かけと足場はずしの必要性

振り返りや計画の記述は、学習者が自身の状況を吟味を促し、学習の工夫を促す [25]。足場はずしは、主体的な吟味や工夫を修得する機会として必要である。しかし、うまく記述ができない学習者には、足場かけによる記述支援が必要である。さらに、教員による手動評価や機械学習による自動評価のためには、できる限り学習状況が記述された文章であることが望ましい。そのために、足場かけ、足場はずしを行う必要がある。

足場かけ 足場かけは、一定の学習状況を得るための記述支援措置とした。そのために、振り返りや計画について、ヒントや選択肢、定型文を用いて、ある程度の型にあてはめた記述をさせる。つまり、何をどのように記述すべきかわからない学習者からも学習状況を読み取りやすいテキストデータが得られる。しかし、型に当てはめようとするため学習者個別の状況に応じた主体的な工夫を阻害する恐れもある。そのために記述内容の状況に応じた足場はずしを必要とする。

足場はずし 足場はずしは、学習者個別の吟味と工夫を含んだ学習状況を得るための措置とした。そのために、自由記述欄を用いることで、学習者が自身の状況を吟味した記述させる。つまり、学習者個別の学習状況に応じて吟味と工夫を含んだテキストデータ

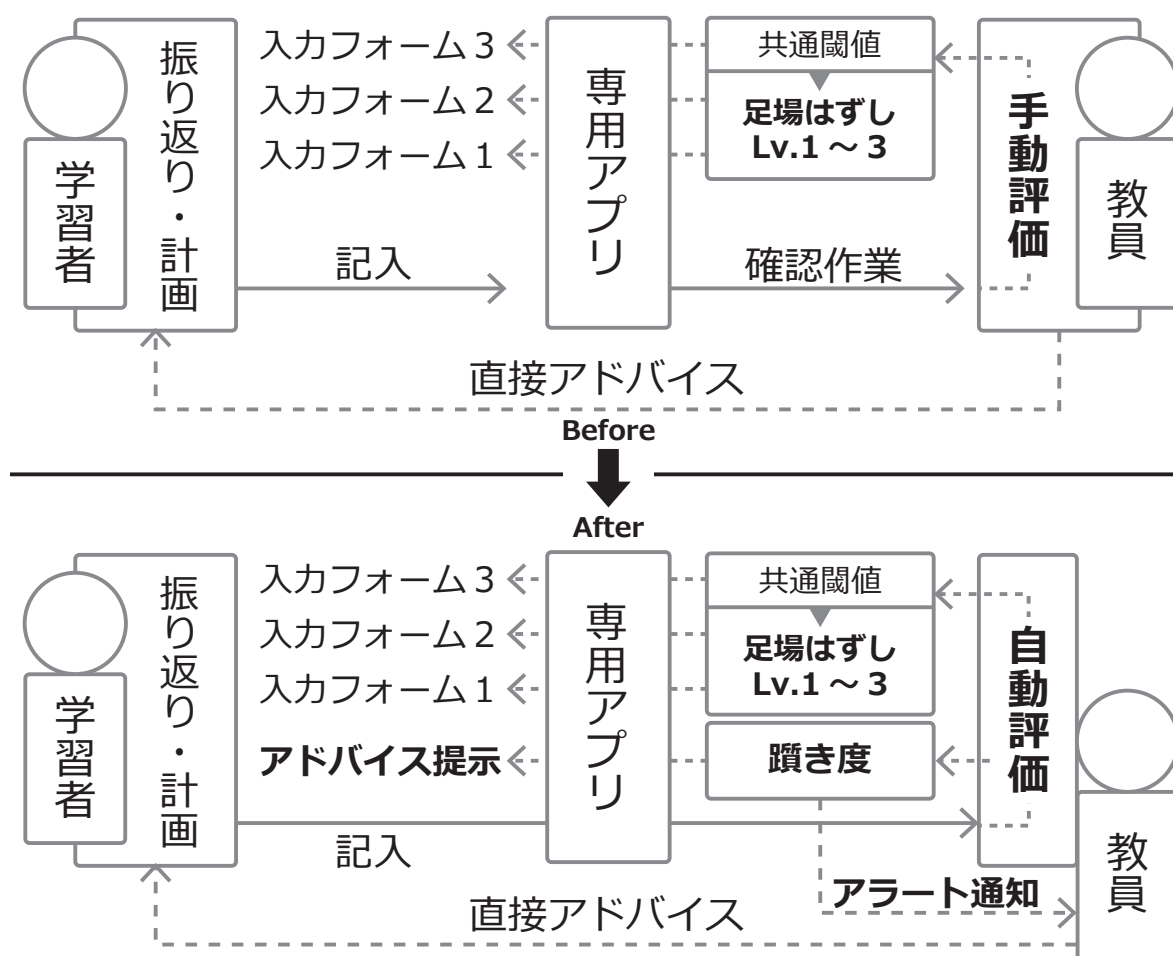


図4.1: 学習支援の概要

が得られる。しかし、型にあてはまらないテキストデータのため学習状況を読み取ることが比較的難しい。また、何をどのように記述すべきかわからない学習者については、教員が意図しないデータを含む可能性が高まる。そのために、記述内容の状況に応じた足場かけを必要とする。

なお、第4章では、足場かけ，足場はずしの妥当性や，主体的な記述ができるようになったかどうかの評価についてはできていない。今後の研究課題である。

これらの足場かけ，足場はずしは学習者個別の学習状況に応じて判断することが重要となる。

4.1.3 学習支援システムの概要

図 4.1は、著者らが開発している学習支援である。図 4.1の上部が提案手法実装前、下部が実装後を示す。著者らは、以前開発した学習支援 [1] を改善するため、学習者の振り返りと計画の記述の具体化、および記述に伴う学習状況把握から自己調整を支援するシステムを開発している。

具体的には、学習者の振り返りと計画の支援として、3つの足場はずしレベル (Lv.1-3) を設定した。

1. **Lv.1** 支援が手厚。ボットが記述内容をサポート。
2. **Lv.2** 支援が手薄。記述内容に関するヒント提示。
3. **Lv.3** 支援が無い。学習者が自力で記述。

図 4.2は、図 4.1に示した専用アプリの画面である。図 4.2の専用アプリは提案手法の実装前後で共通であり、第4章の実験に使用された。図 4.2の「TASK」が学習支援の実装部分である。専用アプリは、授業課題の制作環境と学習支援環境を統合した Web アプリである。「TASK」以外の「IMAGE」「CASE」「TEST」は、授業課題であるデジタルサイネージを制作するための機能やツールを備える。「TASK」には、学習支援として、振り返りや計画の記入、記入有無の表示、記述内容の再表示、アンケート回答、課題タスクを管理する機能を備える。

学習者は、振り返りや計画の記述を図 4.2に示した「TASK」内の赤枠点線内に提示される入力フォームに行く。入力フォームは3種類あり、授業回毎に学習者個別に設定された足場はずしレベル (Lv.1-3) に応じて選択され提示される。足場はずしレベルごとに提示される入力フォーム (1-3) の例は??節実験方法で示す。

振り返りや計画にされた記述内容は専用アプリ内に保存される。教員は、次の授業回の足場はずしレベルを設定するため専用アプリ内に保存された記述内容について目視と手作業で具体性の評価を行う。そして、教員は必要に応じて学習者にアドバイスをを行う。

4.1.4 自動評価の必要性

図 4.1の上部に示す手動評価による学習支援の運用について、大学などの毎週実施される授業では、学習者が多く、さらに複数の授業を担当する教員にとって多大な負担となった。そのために、学習支援のためのアドバイスについても緊急性が高い場合など最小限となり、提示するタイミングが遅れるなどの問題が生じた。そこで、学生個別の学習支援を充実しつつ教員の負担を増やさないために、足場はずしレベルの自動設定や提示するアドバイスの選択を自動化する方法が必要となった。



図4.2: 開発中の学習支援画面

図 4.1の下部は、自動化したときの学習支援を示す。学習者の振り返りと計画の記述内容から具体性と躰き度を自動評価する。自動評価によって以下 2 つの自動化を実現する。

1. 具体性の自動評価。足場はずしレベルの自動設定を行うために、学習者が記述した振り返りと計画から具体性を自動評価を行う必要がある。そのために、振り返りを細分化した 4 項目と計画を細分化した 4 項目について具体性を自動評価する。その合計値によって足場はずしレベルを設定する。
2. 躰き度の自動評価。今後の課題であるアドバイス自動選択と教員へのアラート自動通

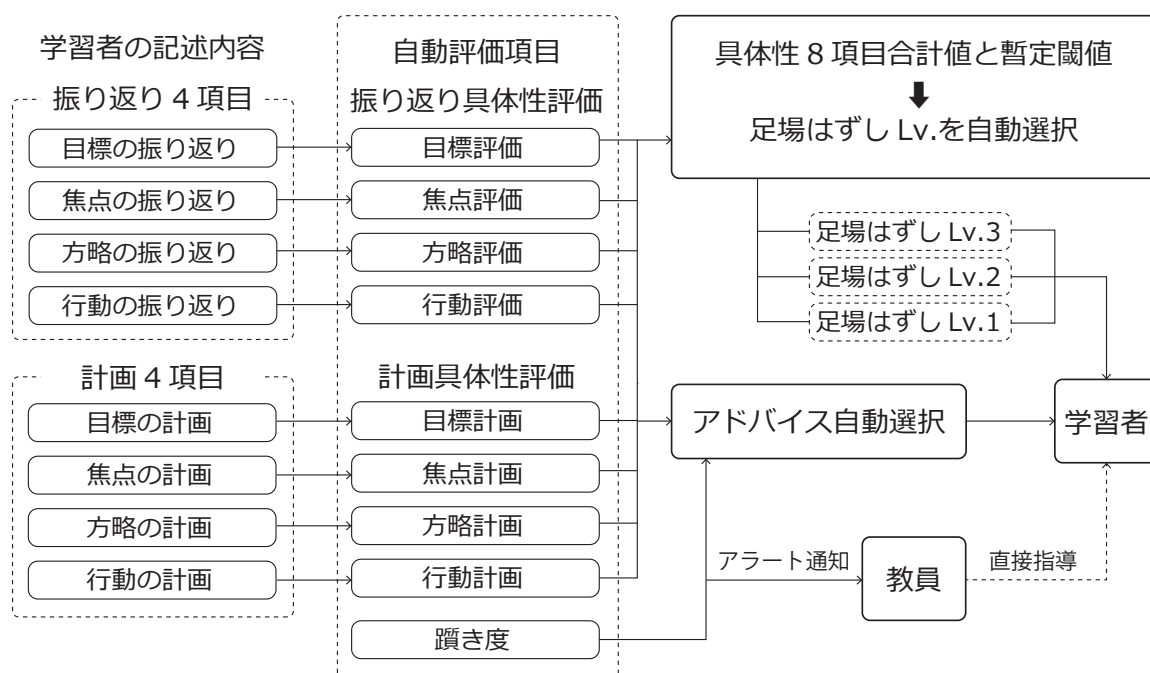


図4.3: 学習支援の自動評価したい部分

知を実現するために、躰き度を自動評価する必要がある。そのために、振り返り 4 項目と計画 4 項目の全体の記述内容について躰き度を自動評価する。

以上 2 つの自動化を実現することで、さらに入力フォームの提示と同時に学習者個別にアドバイスを自動提示することも実現できる。

図 4.3は、図 4.1の下部に示した自動評価の詳細を示す。学習者が記述する振り返りと計画の記述内容は、振り返り 4 項目、計画 4 項目の計 8 項目で構成される。図 4.3点線部に示した項目の自動評価を実現するため、学習の振り返りと計画の 8 項目の記述内容から具体性と躰き度を自動的に評価する手法が課題となる。

そこで、第4章では、OpenAI の大規模言語モデル GPT-3[27] の再学習を行うことで自動的な評価を実現する。自動的な評価を実現するための自動評価用モデルは、GPT-3 の事前学習モデル curie をトレーニングデータで再学習する。再学習した自動評価用モデルを用いることで、学習者の振り返りと計画の記述内容から具体性と躰き度を自動評価する手法を提案する。

第4章は、自動評価用モデルによる自動評価の有効性について、自動評価用モデルによる自動評価と教員による手動評価を比較した検証を行う。

先行研究では、先行研究における限界を示し、教員の評価を模倣した自動評価手法に

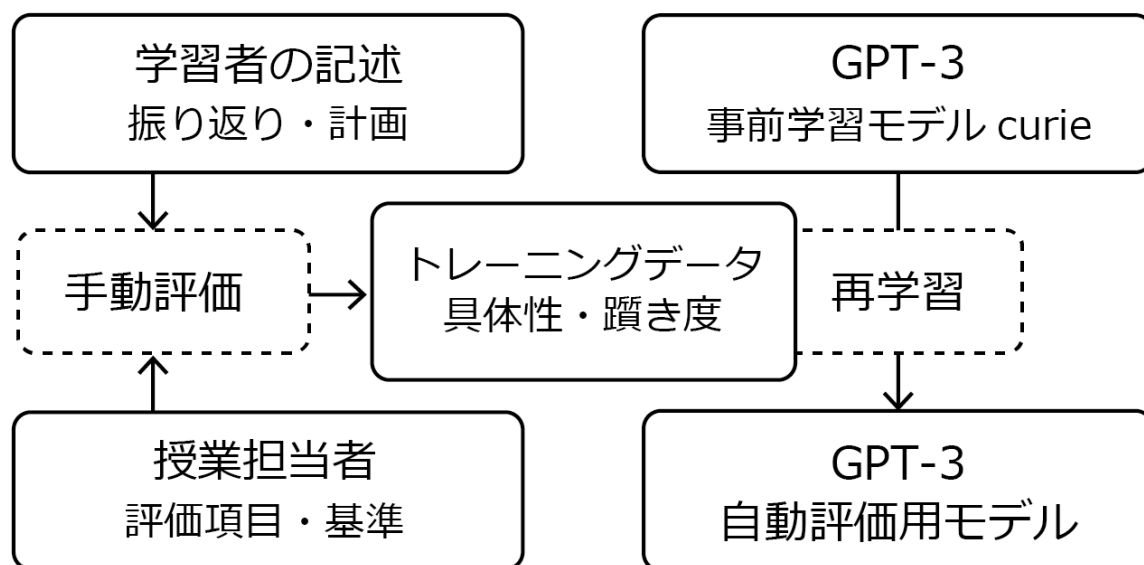


図4.4: 自動評価用モデルの作成

GPT-3 を用いる新規性を示す。提案手法では、自動評価モデルを得るための GPT-3 の再学習と、トレーニングデータの作成方法を示す。実験方法では、実際の授業からトレーニングデータ、および、テストデータを得た方法を示す。評価方法では、交差検証法を用いて自動評価した実験データを検証する方法を示す。実験結果では、具体性の自動評価に順位相関係数 $+0.815$ 、躓き度の自動評価に順位相関係数 $+0.876$ 、足場はずしレベルの自動設定に順位相関係数 $+0.804$ が得られたため、それらの詳細と、評価が一致した事例と誤差が生じた事例を示す。考察では、実験結果から第4章の本研究の目的 1 と 2 をある程度達成できており、実用に耐える性能と結論した考察と、考察から得られた今後の課題を示す。

4.2 振り返りの記述の具体性評価と躓き度評価

学習者の振り返りと計画の記述内容から具体性と躓き度を自動評価する。以下 2 つの目的を実現するために、大規模言語モデルの再学習を用いる手法を提案する。

1. 具体性と躓き度について、教員の手動評価を模倣した自動評価ができる。
2. 自動評価した値を用いて足場はずしを自動的に設定できる。

4.2.1 自動評価用モデルの作成手順

図 4.4は、自動評価用モデル作成の手順を示す。トレーニングデータは、学習者が記述した学習活動の振り返りと計画内容に対して授業担当教員が手作業で具体性と躓き度を評価

表4.1: 教員の手動による具体性評価の例（焦点振り返り）

焦点振り返り記述	評価
一番注目した所は、デシジョンツリーを作成する時、何回か設定した条件が出てこない現象が起こったので、どのように作成したら、設定した条件が出てくるか考えながら作成した部分が一番注目した。	5
効果があったと思う方法や考え方は、条件を可能な限り散らした。興味を惹くようなところは特になかった。課題をやってみた印象は、画像に合った条件を考えたこと。	4
様々なパターンを考えてデシジョンツリーの行を増やすことに注目した。	3
ユーザー目線の学習用データの入力	2
ディジョンツリー	1

した結果を用いる。

具体的には、トレーニングデータは、図 4.3の学習者の記述内容に示した振り返り 4 項目と計画 4 項目の計 8 項目について、図 4.3の自動評価で示す 9 項目（目標評価具体性、焦点評価具体性、方略評価具体性、行動評価具体性、目標計画具体性、焦点計画具体性、方略計画具体性、行動計画具体性、躓き度）について教員が手作業で評価した結果を用いる。

4.2.2 教員の手動評価によるトレーニングデータの作成

具体性 8 項目の評価は、記述の具体的なものを「5」とし、具体的でないものを「1」の 5 段階評価とした。躓き度の評価値は、躓きが大きいものを「5」とし、小さいものを「1」の 5 段階評価とした。具体的な評価基準は、4.3具体性評価のためのループリックに示した。なお、評価基準どおりに教員が評価できているかについて、評価の妥当性の検証はできていないため、今後の課題とする。

表 4.1に、振り返りの記述に対して教員が手動評価した一部例を示す。焦点の振り返り（焦点評価）に対する具体性評価 5 から 1 の例を示した。

表 4.1の各評価について、ループリックに照らした評価を示す。

評価 5 「設定した条件が出てこない」という結果を回避するために、データの作り方に関係があることに注意した記述が評価された。

評価 4 良い結果を得るために「条件を可能な限り散らした」「画像に合った条件を考えた」

表4.2: トレーニングデータの例

```

{ "prompt" "
振り返り__目標:「あまり進められなかった」,
振り返り__焦点:「深く考えすぎない」,
振り返り__方略:「うまくいきつつある」,
振り返り__行動計画:「空き時間に取り組んだ」,
計画__目標:「やれる時に一気にやる」,
計画__焦点:「効率」,
計画__方略:「考えすぎずにテキパキとこなす」,
計画__行動計画:「授業中や空き時間」
-> ",
"completion":
目標評価具体性 =1, 焦点評価具体性 =1,
方法評価具体性 =1, 行動評価具体性 =1,
目標計画具体性 =1, 焦点計画具体性 =1,
方法計画具体性 =1, 行動計画具体性 =2,
躓き度 =1
END "
}
```

が寄与していると注意した記述が評価された。

評価3 「様々なパターンを考え」ることに注意した記述が評価された。

評価2 「ユーザー目線」とすることに注意した記述が評価された。

評価1 記述内容が注意点として評価されなかった。躓き度1とした表2の事例は、振り返りと計画が具体的になっていない記述が評価された。

表4.2に、トレーニングデータの例を示す。「prompt」に振り返りと計画の記述、「completion」に教員の付けた評価を記述した。

表4.2の躓き度について、ループリックに照らした評価を示す。

躓き度1 振り返りと計画が具体的になっていない記述が評価された。

4.2.3 事前学習モデルの再学習

以上のトレーニングデータを用いて GPT-3 の事前学習モデルを再学習することで自動評価用モデルを得る。再学習は、GPT-3 の事前学習モデル curie を用い、バッチサイズは2、エポック数は6とした。

4.3 具体性評価のためのルーブリック

具体性評価のためのルーブリックは、学生の学習計画と振り返りの記述項目に対応する。学習計画と振り返りの記述項目は、自己調整学習における学習方略の目標設定、注意の焦点化、方略計画について対応する。なお、行動計画は学習方略の方略計画に含まれる。つまり、具体性評価のルーブリックは、学習方略の利用や調整の言及について具体性を評価するものとして作成した。なお、具体性評価のためのルーブリックの妥当性については検証されていない。

4.3.1 目標の振り返り具体性

評価 1 : 目標が何だったかについて書かれていない

評価 2 : 遠い目標、あるいは、抽象的な目標について書かれている

評価 3 : 具体的な目標について書かれている

評価 4 : 具体的な目標に対して、どのような過程を経たのかについて書かれている

評価 5 : 具体的な目標に対する過程でどこまで到達できたのか、できなかったのか、対応などについて書かれている

4.3.2 焦点の振り返り具体性

評価 1 : 注意点となったことについて書かれていない

評価 2 : 大まかな注意点について書かれている

評価 3 : 具体的な注意点について書かれている

評価 4 : 結果に寄与する注意点として書かれている

評価 5 : 注意点と成果や結果との関係などの説明が書かれている

4.3.3 方略の振り返り具体性

評価 1 : 使用した方法や考え方について書かれていない

評価 2 : 使用した方法や考え方について抽象的に書かれている

評価 3 : 具体的に使用した方法や考え方について書かれている

評価 4 : 使用した方法や考え方の適性について書かれている

評価 5 : 使用した方法や考え方の適性と対策などの説明が書かれている

4.3.4 行動計画の振り返り具体性

評価 1 : 取り組んだタイミングについて書かれていない

評価 2 : 取り組んだタイミングについて抽象的に書かれている

評価 3 : 取り組んだタイミングについて具体的に書かれている

評価 4 : 取り組んだタイミングの適性について書かれている

評価 5 : 取り組んだタイミングについて結果と対策などの説明が書かれている

4.3.5 目標の計画具体性

評価 1 : 目標について書かれていない

評価 2 : 遠い目標, あるいは, 抽象的な目標について書かれている

評価 3 : 具体的な目標について書かれている

評価 4 : 目標に対して段階化した近い目標が書かれている

評価 5 : 段階化した目標について, 理由や達成見込みなどの説明について書かれている

4.3.6 焦点の計画具体性

評価 1 : 注意点について書かれていない

評価 2 : 大まかな注意点について書かれている

評価 3 : 具体的な注意点について書かれている

評価 4 : 結果に寄与する注意点として書かれている

評価 5 : 注意点と見込まれる成果や結果との関係などの説明について書かれている

4.3.7 方略の計画具体性

評価 1 : 使用する方法や考え方について書かれていない

評価 2 : 使用する方法や考え方についてか抽象的に書かれている

評価 3 : 具体的に使用する方法や考え方が書かれている

評価 4 : 使用する方法や考え方の適性が書かれている

評価 5 : 使用する方法や考え方の適性と, うまくいかないときの対策などの説明について書かれている

4.3.8 行動計画の計画具体性

評価 1 : 取り組むタイミングについて書かれていない

評価 2 : 取り組むタイミングについて抽象的に書かれている

評価 3 : 取り組むタイミングについて具体的に書かれている

評価 4 : 取り組むタイミングの理由が書かれている

評価 5 : 取り組むタイミングが都合が悪いときの対策などの説明について書かれている

4.3.9 躓き度

躓き度 1 : 目標, 焦点, 方略, 行動計画から状況把握ができており, 計画による見通しが十分期待できる.

躓き度 2 : 状況把握がほぼできており, 計画による見通しがある程度期待できる.

躓き度 3 : 状況把握ができていないものの, 計画による見通しはある程度できている.

躓き度 4 : 状況把握がある程度できているものの計画による見通しがあまり期待できない.

躓き度 5 : 振り返りで状況把握ができていない. 学習が進んでいない旨が書かれており, 計画による見通しに期待できていない.

4.4 自動評価のために GPT-3 を用いた理由

以下の 3 点は, 具体性や躓き度の評価に GPT-3 を用いる理由である.

1. ドメイン固有ではない予測困難な語彙にも対応できる. GPT-3 は大規模なコーパスで事前学習されているため, 学生が用いる予測不能な語彙にも対応可能である.
2. 再学習 (fine-tuning) が容易である. 例えば BERT の場合, 分類問題か回帰問題かによって再学習の方法が変わってくるが, GPT-3 であればどんな問題であっても出力をテキスト形式で扱うため, 訓練データの設計が容易である.
3. GPT-3 の意味解釈の性能の良さ. 具体性や躓き度の評価は, 振り返りや計画に記述されているテキストから, 学習に関する重要な点と, それらの関係性を捉えて解釈する必要がある. GPT-3 は, そのような意味解釈を高精度に行える.

4.5 実験

再学習に用いるトレーニングデータ, および, 評価に用いるテストデータを実際の授業から得るために実験を行った. 以下にデータを得るために実施した実験方法について示す.

4.5.1 実験から評価までの手順

1. トレーニングデータの作成. 実験対象の実際の授業から学習活動の振り返りと計画を取得. 記述内容について教員が手作業による具体性と躓き度の評価を行う.
2. 自動評価用モデルの作成. トレーニングデータを用いて自動評価用モデルを作成.
3. 交差検証. 自動評価用モデルを適用した自動評価の結果と教員による手動評価を検証.

4.5.2 実験対象

対象授業 デジタルデザイン表現

被験者 大学生 156 名

演習課題 デシジョンツリーで動くデジタルサイネージの作成

対象授業回 10 回目, 12 回目

振り返りと計画の提出数 253 件 (10 回目 128 件, 12 回目 125 件) 未提出は含まない

毎回の授業において 156 名の受講者に振り返りと計画の記入を求めた。第4章では, 10 回目と 12 回目の 253 件の学習活動の振り返りと計画の内容を実験対象および実験用のトレーニングデータとして使用した。

10, 12 回目を対象にした理由を述べる。10 回目は, 足場はずしレベル (Lv.) ごとの入力フォームの調整を終え, 全員を足場はずし Lv.3 に統一したタイミングであった。そのため教員の手作業による評価が安定すると考えた。11 回目は, 担当教員である著者が手作業で評価を行う前に, 著者が練習で試した自動評価の結果に触れてしまっていたため, あらかじめ自動評価の結果を知っている著者が手動評価を行うと, 自動評価の結果を意識せずに本来の手動評価を行うことが困難であるため使用しなかった。13 回目は, 11 回目と同じ理由で使用しなかった。

以下に, 12 回目の記述に使用された足場はずし Lv. の件数を示す。足場はずし Lv.3 : 16 件, Lv.2 : 63 件, Lv.1 : 46 件。

4.5.3 振り返りと計画のサイクルと足場はずしレベルの設定

図 4.5に授業回ごとに実施した振り返りと計画のサイクルを示す。例えば 10 回目の振り返りは, 9 回目に建てた計画に基づいて記述する。そして, 10 回目の計画は, 振り返りと 10 回目の授業内容に基づいて記述する。なお, 振り返りと計画のサイクルにおいて提示される入力フォームの種類については, 例えば, 10 回目に記述するときは, 9 回目の記述内容の具体性評価に基づいた足場はずしレベルによって定まる。

以下に 10, 12 回目に関連する可能性のある 9 回目から 13 回目までの授業内容を簡潔に示す。振り返りと計画の記述内容は授業進行に沿って記述される事が想定される。例えば, 10, 12 回目の振り返りは 9, 11 回目に記述した学習計画に対する自己評価であり, 10, 12 回目の計画は 11, 13 回目までの学習計画となる。

9 回目 デシジョンツリーの調整

10 回目 キャッチコピーの考案

11 回目 キャッチコピー表示とレイアウト

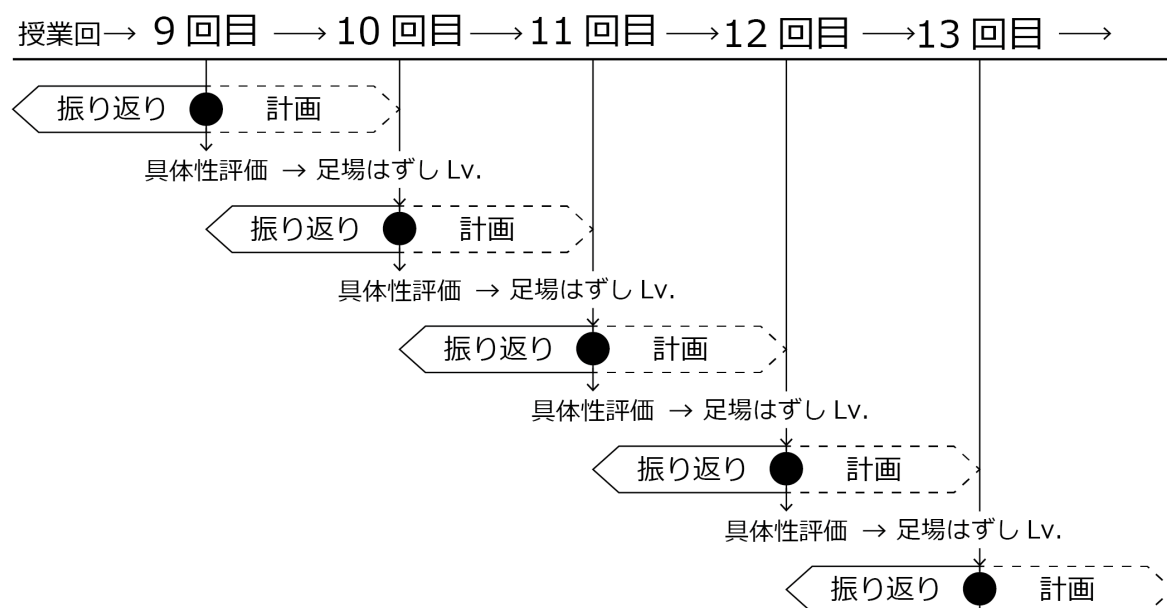


図4.5: 授業回ごとの振り返りと計画のサイクル

12 回目 デジタルサイネージの動作設定

13 回目 画像表示と切り替えアニメーション

足場はずしレベルの設定は、図 4.3の学習者の記述内容に示された振り返り 4 項目と計画 4 項目について各項目の具体性評価値の合計と暫定閾値（Lv.1: 合計値< 15, Lv.2: 合計値< 25, Lv.3: 合計値 \geq 26）によって設定される。なお、暫定閾値の適性検証は今後の研究課題である。

4.5.4 入力フォームと足場はずしレベル

図 4.6に足場はずしレベル（Lv.）ごとの入力フォームの一部を示す。図 4.6は、上から順に足場はずし Lv.3, 2, 1 と並ぶ。全ての入力フォームは、図 4.3に示した振り返り 4 項目（目標の振り返り、焦点の振り返り、方略の振り返り、行動の振り返り）と計画 4 つの項目（目標計画、焦点計画、方略計画、行動計画）について学習者が記述する。

以下に各項目の説明を示す。

1. 目標の振り返り（目標評価） 前回計画した目標の妥当性と、どこまで到達したか、何ができなかったのかについて自己評価する。
2. 焦点の振り返り（焦点評価） 前回計画した焦点の妥当性と、取り組みから得られた結果と焦点を関係付けて自己評価する。

3. 方略の振り返り（方略評価） 前回計画した方法の妥当性と、得られた結果と方法に関係付けて自己評価する.
4. 行動の振り返り（行動評価） 前回計画した行動の妥当性と、得られた結果と行動に関係付けて自己評価する.
5. 目標計画 次回までの到達目標について、前回の自己評価と現状を踏まえて段階的に計画する.
6. 焦点計画 課題から着目したポイントを明確にし、自身の興味関心などと関連付けて計画する.
7. 方略計画 課題の方法や考え方、試みについて、課題内容や焦点、自身の得意不得意などと関連付けて計画する.
8. 行動計画 課題をいつ実施するのか、次回までの自身の予定と関連付けて計画する.

以下に、図 4.6の入力フォームの違いを示す.

足場はずし Lv.3 の入力フォーム 学習者がほぼ自力で記述する. 振り返り 4 項目と計画 4 つの項目の問いと入力欄が提示される.

足場はずし Lv.2 の入力フォーム 記述内容を示唆する支援がある. Lv.3 の入力欄に加え、振り返りには前回の記述内容、計画には授業内容に沿ったヒントが提示される.

足場はずし Lv.1 の入力フォーム 記述内容の定形文を示す支援がある. ボットとの対話で、学習者が選択肢を選び、選択肢に応じた定形文があらかじめ示された入力フォームが提示される.

10 回目は、全員が足場はずし Lv.3 のフォームで記入され、12 回目は、11 回目の記述内容によって学習者個別に足場はずしレベルが設定され、学習者個別に足場はずしレベルに応じた種類の入力フォームで記述された.

4.6 評価方法

目的 1 : 教員が手作業で評価した手動評価を模倣することができているか、および、目的 2 : 自動評価した値を用いて足場はずしを自動的に設定できるかどうかについて、実験で得られたデータを用いた交差検証法によって評価する.

以下に示す (1) から (3) の 3 つの自動評価は、本研究の目的 2 つに対応する. 3 つの自動評価について、交差検証法を用いた検証を行う. これにより、本研究の 2 つの目的について評価する. なお、提案手法の評価軸は、自動評価と教員による手動評価の相関係数であり、交差検証法 (Cross-validation) で性能評価するのが一般的である.

表4.3: 手動評価された評価項目と評価値の度数分布

評価項目	1	2	3	4	5
目標評価具体性	69	57	82	41	4
焦点評価具体性	74	61	73	39	6
方略評価具体性	143	59	32	18	1
行動評価具体性	42	124	69	15	3
目標計画具体性	60	58	75	52	8
焦点計画具体性	58	68	79	43	5
方略計画具体性	55	63	97	34	4
行動計画具体性	35	126	67	3	2
躓き度	87	49	46	37	34

1. 具体性の自動評価. 目的 1: 学習者が記述した振り返りと計画の記述内容について具体性を自動評価したときの性能を評価する. 振り返り 4 項目と計画 4 項目の計 8 項目を対象に項目ごとの記述内容を自動評価する.
2. 躓き度の自動評価. 目的 1: 振り返り 4 項目と計画 4 項目の全 8 項目の記述内容を総合して躓き度を自動評価したときの性能を評価する. なお, 本研究では躓き度のみを評価する. なぜなら, 躓き度によるアドバイス自動選択と教員へのアラート自動通知については開発中であり実装されていないためである.
3. 足場はずしレベルの自動設定. 目的 2: 自動評価した具体性 8 項目の合計値を暫定閾値 (手動評価と自動評価共通) に代入して足場はずしレベルを設定する. 足場はずしレベルの自動設定について性能評価を行う.

表 4.3は, 実験用トレーニングデータに含まれた手動評価による評価項目と評価値の度数分布を示す. 10, 12 回目の手動評価のみを使用したため, 各項目とも評価値 5 の件数が少ない.

表 4.4に交差検証用のグループ分けを示す. グループ分けは, 手動評価の足場はずしレベルの分布が大きく偏らないようにした. データの先頭から順に ABCDE の各グループに Lv. 別に割り振った.

交差検証を行うため, 各グループのデータを除外した 5 つの実験用トレーニングデータを作成し, そこから 5 つの交差検証用の自動評価用モデルを再学習した. 具体的には, A グループ 51 件を除いた自動評価用モデル「Excluding A」, B グループ 51 件を除いた自動評価用モデル「Excluding B」, C グループ 51 件を除いた自動評価用モデル「Excluding C」, D グループ 50 件を除いた自動評価用モデル「Excluding D」, E グループ 50 件を除いた自

表4.4: 交差検証用グループの設定

Groups	Total number	Lv.1	Lv.2	Lv.3
A	51	15	27	9
B	51	15	27	9
C	51	15	26	10
D	50	14	27	9
E	50	14	27	9
Total	253	73	135	45

動評価用モデル「Excluding E」を再学習した。再学習時のパラメータは、事前学習モデル curie を用い、再学習時のバッチサイズは 2、エポック数は 6 とした。また、再学習によって得られた自動評価用の GPT-3 モデルを適用して自動評価する際、ランダム性を表すパラメータ temperature は 0 に設定した。

交差検証用の自動評価結果を得るため、交差検証用の自動評価用モデルで対応するグループの自動評価を行った。具体的には、「Excluding A」モデルで A グループ 51 件、「Excluding B」モデルで B グループ 51 件、「Excluding C」モデルで C グループ 51 件、「Excluding D」モデルで D グループ 50 件、「Excluding E」モデルで E グループ 50 件を自動評価した。

足場はずしレベルの自動設定は、自動評価された具体性の合計値と暫定閾値（手動評価と同じ閾値）によって設定された。

4.7 結果

第4章の 2 つの目的のために以下の 3 つを交差検証法によって評価した結果を示す。自動評価の性能を示すために、自動評価と手動評価について、ノンパラメトリック（スピアマン）の順位相関分析を行った。

1. 具体性の自動評価（目的 1）：順位相関係数 +0.815
2. 躓き度の自動評価（目的 1）：順位相関係数 +0.876
3. 足場はずしレベルの自動設定（目的 2）：順位相関係数 +0.804

以下に、自動評価の値と手動評価の値の散布図を示す。ただし、評価値が離散的な整数値であるため、そのまま散布図を描くと単なる格子点となり、標本の密度を読み取ることができない。そこで、自動評価値、手動評価値ともに $[-0.05, +0.05]$ の一様乱数を加算して表示している。

なお、以下に示す実験結果の順序は、具体性の自動評価、足場はずし Lv. の自動評価、躓

き度の自動評価とした。足場はずし Lv. は、具体性の自動評価の値によって設定されるためである。

4.7.1 具体性の自動評価

図 4.7からは、振り返りと計画の計 8 項目全体の具体性評価について手動評価と自動評価の順位相関係数 $+0.815$ ($p<0.0001$, $n=2024$) が示された。

振り返り 4 項目と計画 4 項目の項目別の結果

図 4.8は、振り返り 4 項目と計画 4 項目ごとの具体性評価について手動評価と自動評価の関係を示す。

目標評価具体性（図 4.8, a）の順位相関係数 $+0.855$ ($p<0.0001$, $n=253$) が最も相関が強く、焦点評価具体性（図 4.8, b）の順位相関係数 $+0.854$ ($p<0.0001$, $n=253$)、方略評価具体性（図 4.8, c）の順位相関係数 $+0.777$ ($p<0.0001$, $n=253$)、目標計画具体性（図 4.8, e）の順位相関係数 $+0.841$ ($p<0.0001$, $n=253$) が正の相関を示した。

一方、行動評価具体性（図 4.8, d）の順位相関係数 $+0.647$ ($p<0.0001$, $n=253$) は 8 項目の中で最も相関が弱く、行動計画の 4 項目内では、行動計画具体性（図 4.8, h）の順位相関係数 $+0.708$ ($p<0.0001$, $n=253$) が最も相関が弱い。

表 4.5は、具体性評価項目ごとの誤差（自動評価値 - 手動評価値）の件数を示す。全ての評価項目について誤差 ± 1 は一定数あったことが読み取れる。さらに、誤差 ± 2 に注目したとき、振り返りでは、目標評価具体性と焦点評価具体性について、自動評価が手動評価よりも高く評価する場合があった。方略評価具体性と行動評価具体性誤差では、手動評価よりも低く評価した事例が 3 件、高く評価した事例が 2 件あった。計画に関する 4 項目では、手動評価よりも具体性を 2 高く、もしくは 2 低く評価した事例が見られた。

表 4.6は、手動評価と自動評価の具体性評価が一致した記述事例の一部を示す。記述内容欄の括弧内に足場はずし Lv を示した。Lv.3 の事例は、自動評価モデルが、十分に記述内容があるため、自動評価による単語や文の重要な点と、それらの関係性の解釈がうまくできた可能性を示す。Lv.2 では、加えてヒントによる語彙の類似性によってうまくできた可能性を示す。Lv.1 の事例は、定形文による類似性によってうまくできた可能性を示す。

表 4.7は、手動評価と自動評価の具体性評価に誤差があった記述事例の一部を示す。これらの事例は、比較的文字数が少ない事例では、文や単語の重要な点、それらの関係性をうまく捉えられず誤差が生じた可能性を示す。誤字や文に違和感のあるものは、自動評価は教員が行うような文章の補完を行わなかった可能性を示す。授業内容との関係の認知期の違いなどの要因があった可能性を示す。事例ごとの分析は考察で示す。

表 4.8の記述内容は、想定外の記述内容にもかかわらず、比較的うまく自動評価された事

表4.5: 項目ごとの具体性評価誤差

Evaluation items	-2	-1	0	1	2
目標評価具体性	0	41	168	43	1
焦点評価具体性	0	48	162	42	1
方略評価具体性	3	31	190	27	2
行動評価具体性	3	48	163	37	2
目標計画具体性	1	48	155	48	1
焦点計画具体性	2	49	149	51	2
方略計画具体性	1	53	136	62	1
行動計画具体性	2	46	161	44	0
Total	12	371	2294	354	10

例を示す。課題内容と直接関係性が少ない記述が含まれるものの、学習者の工夫が含まれた記述内容である。手動評価とほぼ同等の自動評価がされた。

足場はずしレベルの自動設定

足場はずしレベルについて実験結果を示す。足場はずしレベルは、手動評価においても自動評価においても具体性評価値の合計を暫定閾値で区分することで自動設定された。図 4.9からは、手動評価と自動評価の結果に強い正の相関 $+0.804$ ($p < 0.0001$, $n = 253$) が読み取れる。図 4.9の散布図は、縦軸が手動評価、横軸が自動評価による足場はずしレベル設定を示す。表 4.9は、手動評価と自動評価によって自動選択された足場はずしレベルごとの件数を示した。足場はずしレベルごとの選択数に僅かな誤差が認められる。表 4.10は、足場はずしレベルの誤差（自動評価の Lv. - 手動評価の Lv.）の詳細を示す。自動評価が手動評価より低いレベルを設定した件数が 22 件あり、高いレベルを設定した件数が 23 件あった。

4.7.2 躓き度の自動評価

図 4.10は、躓き度の評価について手動評価と自動評価の散布図を示す。縦軸が手動評価、横軸が自動評価、評価は躓き度が強いほど 5、弱いほど 1 を示す。手動評価と自動評価の結果に強い正の相関 $+0.876$ ($p < 0.0001$, $n = 253$) が示された。足場はずしレベルごとに分析すると、Lv.3 は正の相関 $+0.666$ ($p < 0.0001$, $n = 46$)、Lv.2 は強い正の相関 $+0.8$ ($p < 0.0001$, $n = 134$)、Lv.1 は正の相関 $+0.792$ ($p < 0.0001$, $n = 73$) が示された。

表 4.11は、躓き度評価値の誤差について詳細を示す。92 件の誤差 ± 1 と 7 件の誤差 ± 2 があった。手動評価の足場はずしレベル別で見ると Lv.2（ある程度の具体性がある記述）に誤差の件数が多い傾向が読み取れる。

表 4.12は、手動と自動による躰き度評価が一致した記述例として、手動評価、自動評価とも躰き度 1（躰きがほとんど見られない）の例を示す。この事例は、各項目について全体的に十分な文と単語が含まれていることから、自動評価は文や単語の重要な点、それらの関係性をうまく捉えることができていた可能性を示している。

表 4.13は、手動と自動による躰き度評価に誤差があった記述例として、手動評価の躰き度 4（だいたい躰きが見られる）、自動評価の躰き度 2（やや躰きがある）の誤差の例を示す。この事例は、各項目について文字数と文と単語の偏りがあることから、自動評価は文や単語の重要な点、それらの関係性をうまく捉えることができなかったために誤差が生じた可能性を示している。

4.7.3 足場はずしと記述内容

12 回目の記述内容から、足場はずし Lv. によって以下 3 つのテストデータの質の違いが読み取れる。10 回目の 128 件は Lv.3 のみであり、12 回目の 125 件は Lv.3 : 16 件, Lv.2 : 63 件, Lv.1 : 46 件が混在する。

1. Lv.3 は、学習自身が吟味して工夫された記述されている。ただし、うまく記述できない学習者に短い文かつ抽象的な記述が見られた。
2. Lv.2 は、提示されたヒントを参考に記述されている。
3. Lv.1 は、学習者が選ぶ選択肢に応じた定形文に沿って記述されている。

なお、Lv.1 の 46 件について、33 件に定形文に沿った記述が見られた。一方、残りの 13 件は定形文をあえて削除して書き直したと見られる記述があった。

また、うまく記述できない学習者については、短い文かつ抽象的な記述が見られた。具体的には「コツコツとやる」「ない」「思うままにやる」「空いてる時間」「特にない」などであった。これらは、Lv.3, Lv.2, および、Lv.1 の定形文をあえて削除した記述に見られた。

以下は、12 回目の記述内容の事例として計画の一部について、足場はずし Lv. ごとの事例を示す。計画の事例を示した理由は、Lv.2 において、ヒントの影響が分かりやすいためである。

Lv.3 の事例は、自由記述欄に記入された事例を示した。

Lv.2 の事例は、ヒントを活かして記述された事例を示した。提示されたヒントは「例えば、まずは、デシジョンツリーの遷移を設定して、動作を確認してから、言葉の表示を設定していくとかかな。どうかな。自分なりのステップを作ってみよう。」であった。したがって、「ツリーを作成して」「言葉の位置を調整」の記述によって、ヒントを活かしていることがわかる事例である。

Lv. 1 は、定形文を編集して記述された事例を示した。「計画する目標地点」についての難易度、「自分の目標地点」について、自分が到達できると思う目標、「一番近い目標地点」について、穴埋め的に記述された事例である。

目標の計画

Lv.3 実行テストを繰り返し行い、改善点を見つける。デシジョンツリーの修正を行う。文字設定を行う。

Lv.2 ツリーを作成して、キャッチフレーズを紙に書き出してみたら、画像と言葉の位置を調整していく。

Lv.1 計画する目標地点は、困難。自分の目標地点は、初期設定までにある程度完成させる。一番近い目標地点は、動作確認までにある程度完成させる。

4.8 考察

実験結果が示した自動評価と手動評価の比較において、以下 3 つの自動評価の結果が示された。

1. 具体性の自動評価（目的 1）。順位相関係数 +0.815
2. 躓き度の自動評価（目的 1）。順位相関係数 +0.876
3. 足場はずしレベルの自動設定（目的 2）。順位相関係数 +0.804

したがって、提案手法による自動評価用モデルは、目的 1：教員の手動評価を模倣して具体性と躓き度の自動評価が出来ていた可能性と、目的 2：自動評価で得た値を用いて足場はずしの自動設定が出来ていた可能性が示された。このことから、十分に実用に耐える性能と考えられる。以下に、実験結果について考察を加える。

4.8.1 自動評価の妥当性

まず、実験結果の全般において自動評価と手動評価に正の相関が見られた理由について考察を述べる。

自動評価は、教員をある程度模倣した評価ができていたと考えられる。なぜなら、具体性の自動評価、躓き度の自動評価、足場はずしレベルの自動設定の 3 つとも、手動評価との順位相関係数に強い正の相関を示していたためである。また、表 4.6 は、4.3 のループリックにある程度沿って教員が評価できていたと思われる事例であり、これらの評価が自動評価と手動評価で一致していたからである。ただし、教員の評価がループリックどおりにできていたかどうかの検証はできておらず、今後の課題としている。

1. 標本が授業内容に依存していた. 表 4.6に示した 1. 目標評価, 2. 焦点評価, 3. 方略評価, 5. 目標計画, 6. 焦点計画, 7. 方略計画の記述内容は, 学習者の個人差よりも授業内容に依存した内容になりやすいと考えられる. そのため, 交差検証用のトレーニングデータに他グループの授業内容に関する十分な標本が含まれたため, 性能が高くなったと考えられる.
2. 標本の範囲が定まっていた. 交差検証用に用いたトレーニングデータは, 自動評価の対象となる授業回 10, 12 回のみであった. そのため, 標本の範囲が限定され評価値を付けやすくなり性能が高くなったと考えられる. なお, 他の授業回や全ての授業回を用いた場合にも同様の結果が得られるのかについてはさらに検証を加える必要がある. また, 10 回, 12 回目以外のトレーニングデータに含まれていない授業回について検証するなどトピック依存性の検証について今後の研究課題とする.
3. 足場かけによる記述支援の影響. 自動評価の精度を高めた要因として, 12 回目のテストデータに含まれた足場かけの影響が考えられる. なぜなら, 12 回目のテストデータは, 足場かけとして記述支援を受けたデータを含むからである. 12 回目では, 選択肢によって提示される定形文を用いた記述支援 (足場はずし Lv.1) が 46 件あった. そのうち 33 件に, ある程度定型文にあてはまるテキストデータが見られた. 表 4.6の「4. 行動評価」と「8. 行動計画」は, 定型文によって, 手動評価と自動評価が一致したと見られる事例である.

さらに, 自由記述欄にヒントを提示した記述支援 (足場はずし Lv.2) も 63 件あった. 与えられたヒントによって, 記述内に使用される語彙の範囲がある程度限られた可能性が考えられる. 表 4.6の「6. 焦点計画」は, ヒントによって, 手動評価と自動評価が一致したと見られる事例である.

一方, 10 回目のテストデータは, 記述支援なしのデータであった. そのため, 12 回目のデータよりもうまく記述できない学習者のデータを多く含んでいた可能性が考えられる. 10 回目では, うまく記述できない学習者に短い文で抽象的な記述が見られた. したがって, 足場かけのあった 12 回目のテストデータは, 足場かけがなかった 10 回目よりも自動評価の精度を高める要因が多く含まれていたと考えられる.

なお, 自動評価の性能評価について, 記述時の足場はずしレベルごとの検証は行っていない. なぜなら, 実際の運用では, 足場はずしレベルが混在することを想定したからである. ただし, 当初の想定よりも, 記述時の足場はずしレベルの違いによるデータへの質的影響があると考えられたため, 足場はずしレベルごとの検証は今後の課題とする.

4.8.2 自動評価の誤り例とその分析

表 4.13のように、自動評価と手動評価の躓き度に差異が生じる事例も確認された。そこで、誤差の事例について考察を加える。

具体性の自動評価における誤差の分析

具体性評価項目ごとの相関は、自動評価と手動評価は全体的に高い相関を示している。しかし、行動振り返り（行動評価）と行動計画の具体性評価は、他の具体性評価項目に比べて相関が低い傾向が見られた。その原因は以下の2つと考えられる。

1. 行動の個人差。上述したように、目標、焦点、方略の記述内容は授業内容に依存する。しかし、行動は授業内容よりも学習者個別の都合に依存するため、より多くのトレーニングデータが必要と考えられる。トレーニングデータに十分な多様性が無く性能が低くなったと考えられる。
2. 高評価の標本の少なさ。表 4.3に示されたトレーニングデータに用いられた手動評価の件数を確認すると、行動について高い評価を付けている標本数が少ない。そのため、標本数が少なかった高い評価よりも、低い評価値を付けやすくなり、性能が低くなったと考えられる。

表 4.7に示された誤差の理由について、ループリックに照らした考察を加える。以下の考察によって、授業内で用いられた言葉に対する認識や解釈の差、および、記述内容を補完した解釈から誤差が生じていた可能性が考えられる。

1. 目標評価 「公欠」「休んでしまった」「取り返したい」が学習に関係する内容かどうかの認識の違いによって誤差が生じたと考えられる。教員は、記述内容が学習内容に直接関係しないと判断したと考えられる。一方、自動評価は、何らかの振り返りで改善点が示されていると判断した可能性が考えられる。
2. 焦点評価 「ディジョンツリー」（デシジョンツリーの間違い）について、自動評価は、学習に関する語彙であると誤認した可能性が考えられる。さらに、単語のみで評価3としたのは、トレーニングデータに単語のみで評価していた事例が含まれていた可能性が考えられる。一方、教員は記述に誤字があることに加え、単語を示すのみでは振り返りとしては不足と判断したと考えられる。
3. 方略評価 記述の冒頭と、その後の記述の関係をどう認識して解釈したのかによって誤差が生じたと考えられる。自動評価は冒頭の適性を示した部分と、その後に続く具体的な部分との関係についてうまく評価できなかった可能性が考えられる。一方、教員は、文章の前後関係を補完して解釈したのち評価したと考えられる。

4. 行動評価 「テスト」が最終的な機会であることの認識の差によって誤差が生じたと考えられる。自動評価は（何度も実施されるような）「テスト」の機会に行うとした評価になったと考えられる。一方、教員は最終的な機会であることを認識しているため、行動計画の振り返りの記述として不十分であるとした可能性が考えられる。
5. 目標計画 「Excel」と「デシジョンツリー作成」の関係の認知の違いで誤差が生じたと考えられる。教員は、「Excel」と「デシジョンツリー作成」の2つが授業においては一体的な工程であることを知っている。一方、自動評価は、授業における2つの関係を認識していない。そのため、異なる2つの工程を段階的に進めることが記述されているとして評価した可能性が考えられる。
6. 焦点計画 「著作権」の扱いが異なることで誤差が生じたと考えられる。教員は、授業で扱った内容を含めて単語のみで具体的であるとした可能性が考えられる。一方、自動評価は、授業で扱った注意点であることを認識していない。そのため、用語のみでは、注意点について具体的ではないと判断した可能性が考えられる。
7. 方略計画 自動評価は、誤字を含め何らかの方法や考え方が記述されていると誤認した可能性が考えられる。一方、教員は、誤字を補完しつつも、課題を進めるための方法についての記述ではないと判断した評価と考えられる。
8. 行動計画 教員が文を補完して解釈し評価していたことで誤差が生じた可能性が考えられる。教員は「バイト終わりにする」と「授業の空きコマ」を分けて評価したと考えられるが、自動評価は、1つの具体的なタイミングがあるのみと評価した可能性が考えられる。

躓き度の自動評価における誤差の分析

自動評価された躓き度において、手動評価との誤差が比較的大きかった記述例について考察を加える。

表 4.13の誤差のあった記述例では、学習者は計画に書いたことが実施できていないという経緯があった。しかし、振り返りでは、出来ているように記述していた。そのため、自動評価では躓き度 2 と評価された。一方、手動評価した教員は、上記の経緯を把握していたため躓き度 4 とした可能性がある。つまり教員による手動評価では、学習者の経緯を加味して評価が行われたと考えられる。

このことから、教員の手動評価を模倣するためには、課題の制作状況や専用アプリの操作ログなどの経緯を含めたトレーニングデータが必要となる。このような学習者の状況や経緯を含めた評価手法は今後の研究課題とする。

表 4.13について、ループリックに照らした考察を加える。以下の考察によって、自動評価は、教員とは異なる記述項目を重視して判断していた可能性が考えられる。つまり、躓き

度の自動評価は、記述された項目の全てが十分に具体的に記述されていない場合、少ない記述内容だけで、判断できるほど十分なトレーニングデータが無かった可能性が考えられる。

手動評価 振り返りでは、目標評価はある程度自己把握できているものの、焦点評価に結果との関連付けた振り返りがなく、方略評価、行動評価が具体的でない、計画では、目標計画は段階的に示されているものの、焦点計画、方略計画、行動計画が具体的ではないことから見通しに期待ができないと判断されたものと思われる。

自動評価 目標評価とある程度詳細な記述があり、あまりうまく進められていない記述があるものの、目標計画にある程度の過程が示されており、焦点計画、方略計画、行動計画にある程度の記述があることによって手動評価ほど躓いているとは判断されなかった可能性が考えられる。

4.8.3 想定外の記述内容について

一方、表 4.8のように想定していない記述内容について、うまく評価されていた事例が見られた。事例は、授業課題のデジタルサイネージに、学習者自身の興味関心のある 3DCG の映像を組み合わせた内容を記述していた。トレーニングデータに該当事例と似たような記述が含まれていないにもかかわらず、自動評価は手動評価とほぼ同様にうまく評価していた。

このように、授業内容に自身の興味関心を組み合わせる工夫は、自己調整学習において焦点の調整方略として歓迎すべき内容である。

また、この事例によって GPT-3 を用いたことの妥当性が示されたと考える。なぜなら、「3DCG でのロゴアニメーションを作って YouTube にアップロードしたものを設定した。」は、「3DCG」「ロゴアニメーション」「YouTube」などの想定外の語彙を含む記述であっても、その文脈から複数の過程の記述であると判断された可能性が考えられるからである。

ただし、この事例から、記入項目や授業内容にまったく関係しない内容を具体的に記述した場合についても、具体性を高く評価する可能性が考えられる。

なお、学習者個別の特性や興味関心に伴う焦点の調整方略と授業内容や記入項目との整合性の課題については今後の研究課題とする。

4.8.4 教員の過度な依存への対策

第4章の自動評価は、十分な性能が示された。しかし、自動評価に教員が過度に依存することは危険が伴う。なぜなら、上述の躓き度誤差分析から自動評価が学習者の躓きを見逃す事例が明らかになっている。そのため自動評価への過度な依存は、学習者の躓きを見逃す恐れがある。そこで、対策案としては、例えば自動評価と教員による手動評価を授業回毎に交互に運用する仕組みとすることによって、教員の負担を一定程度軽減しながら、自動評価へ

の過度な依存を防ぐ対策が考えられる。

4.8.5 自動評価の誤差に対する学習者の許容範囲

具体性の自動評価の誤差は、足場はずしレベルの自動設定に影響して学習者が記述する内容に影響を与えられられるが、誤差による影響は学習者の許容範囲に収まると考えられる。なぜなら、以下 2 つの措置による。

1. 足場はずし Lv. を 3 段階設けている。そのため、具体性評価 8 項目の合計値から設定される足場はずし Lv. を高く見積もりされた場合に極端な足場はずしとなる事例を少なくできる。
2. 足場かけにおける記述支援でも記述内容を自由に編集できる機会を設けている。そのため、足場はずし Lv. を低く見積もりされた場合に学習の工夫を記述する機会を得られる。

ただし、誤差が学習者に以下のような影響を与える可能性もある。

1. 足場はずし Lv. が低く見積もられた場合、足場かけの記述支援が提示するヒントや定形文によってバイアスを受ける。そのため、学習者が主体的に個別の状況を吟味して工夫する機会を失う可能性が考えられる。
2. 足場はずし Lv. が高く見積もられた場合、学習者は記述すべき内容がわからず、振り返りや計画をうまく記述できなくなる可能性が考えられる。

表 4.11 に示したような躓き度の誤差について、躓き度が低く見積もられた場合においては、学習者の許容範囲に収まらないと考えられる。なぜなら、教員などから学習支援を受ける機会を失う可能性が考えられる。そのため、今後の課題として、教員による確認を要とする仕組みの構築が必要となる。一方、躓き度が高く見積もられた場合については、学習者の許容範囲に収まると考えられる。なぜなら、不要な学習支援を受ける負担が若干あるものの、学習としては無駄にはならないものと考えられる。

4.9 結言

第4章が提案した2つの目的について、自動評価手法は、教員の手動評価をある程度模倣した自動評価ができ、十分に実用に耐える性能であったと考えられる。

1. GPT-3 を用いることで、具体性と躰き度について、教員の手動評価を模倣した自動評価ができる。
2. GPT-3 で自動評価した値を用いて足場はずしを自動的に設定できる。

ただし、第4章は実際の授業の一部から得られたトレーニングデータをそのまま使用した。そのために、標本数の少ない高評価値をうまく自動評価できない事例が生じていた。このことから、実際の授業データをトレーニングデータとして用いる場合、事前に評価値の分布を分析し、標本に偏りがある場合にはダミーの記述内容や評価を加えるなどの補完する措置が考えられる。

また、躰き度の自動評価では、教員の手動評価と異なり学習者の状況や経緯を考慮しなかった。そのために、学習者の躰き度を低く評価した事例が確認された。このことから、課題の制作状況や専用アプリの操作ログなどの経緯を含めたトレーニングデータを含めるなどの今後の研究課題が明らかになった。

さらに、足場かけによる記述支援が、自動評価の精度に影響を与えていたことがわかった。特に潜在的な支援の必要性を見つけ出して学習者を支援するためには、精度よく自動評価できることが重要である。そのために、うまく記述できない学習者が学習状況を記述できる支援が重要である。そこで、今後の課題として、足場かけを改善する。具体的には学習状況の詳細な聞取りなどの自動対話の内容を改善する。

なお、自動評価への教員の過度な依存は、学習者の躰きを見逃す恐れがある。そのため、教員が自動評価に依存しない対策が必要である。自動評価は一定の性能を示したが、上述で示したようにうまく評価できない事例もある。

そもそも、各学生への個別最適化した学習支援を行うにはテキスト情報だけでは不十分であり、学習者の特性や興味関心などを考慮する必要がある。

本章の提案手法は、それらの要素を考慮できていない。そのため、個別最適化のためには人間の教員による評価と併用する必要がある。そこで、例えば自動評価と教員の手動評価を交互に運用する仕組みにするなどの対策案が考えられる。また、自動評価への過度な依存を防止するためのインストラクションが不可欠となろう。

第4章が提案した自動評価手法は、実際の授業から蓄積したデータを活用することで評価を自動化する手法である。つまり、従来の学習支援において教員が行わなければならなかった記述内容の評価部分について、教員が行っていた評価を模倣して自動化する提案である。

そのため、教員が個別に学習支援に使用してきたアドバイスや足場かけ、足場はずしの手法と親和性が高いと考えられる。

なお、第4章について、新たな研究課題として次のことが明らかになった。

1. 躓き度について学習者の状況や経緯を含めた評価手法。
2. 焦点の調整方略による想定外の記述と無関係の記述をうまく自動評価する方法。
3. 自動評価に対する教員の過度な依存への対策。
4. トレーニングデータに含む授業回、含まない授業回に対する自動評価の検証。
5. 振り返りと計画記入時の入力フォームの種類と自動評価の検証。

これらの研究課題についてさらなる検討と検証を行い、学習支援の自動化に向けた改善点を明らかにする。

本章では、学習者が記述した学習計画と振り返りの記述内容を自動評価するための手法として、ループリックを用いたデータでファインチューニングした大規模言語モデルを用いた。提案手法を評価した結果、具体性と躓き度の評価について、教員の手動評価を模倣した自動評価と、教員の手動評価を模倣して足場はずしの自動的は、十分に実用に耐える性能であったことがわかった。したがって、表 2.1 に示した学習支援システム研究の項目 5 において、学生の記述内容から学習スキルの自動評価の実現に貢献したと考えられる。

本章で確認された自動評価の有効性を踏まえ、第5章では、学生個別の学習計画と振り返りに基づく具体的なアドバイスを自動生成する手法を提案する。

振り返りをします。前回以降の課題取り組みについて、以下の項目を自己評価をしましょう。

1. 前回の計画に対してどこまでできましたか
2. 課題に取り組んで一番注目したことは
3. 計画した方法や考え方はうまくいきましたか
4. いつ、どんな時に課題に取り組みましたか

次へ

足場はずし Lv.3 入力フォームの一部

今回の課題について次回（授業前）までの計画をしましょう。

5. 今回の課題はどんな段階に分けてやると思いますか？
例えば、まずは画像の表示設定や切り替え効果を試してから、Excelを30行まで進めるとかな。
どうかな。自分なりのステップを作ってみよう。
6. 今回の課題では何に一番注目できそう？
例えば、学習用データの条件設定と、入力するデータによって、AIがどう判断を変えていくのか？とかかな。
どう？自分の主観的なことでいいよ。

次へ

足場はずし Lv.2 入力フォームの一部

オートTA: 次は2つ目の記入項目ね。
課題をやってみたと思うんだけど、
そのとき、やってみた方法や考え方で、課題を進めるのに役立ったことや効果が
あったと思えることは何かある？

あった なかった

ユーザー: なかった

オートTA: そうか。問題ない。そういうときもある。

オートTA: やってみた方法や考え方で、もっと試したいことや、もっと知っておきたいこと、や
らなさやと思うことは何かあった？

あった なかった

ユーザー: あった

オートTA: いいね。どんなところをもっと試したり、知りたいと思ったの？

送信

オートTA: 2つ目の記入項目、前回の課題で一番注目したことについて追加記入して
送信してね。
送信したら、次の記入項目だよ。

確認送信

足場はずし Lv.1 入力フォームの一部

図4.6: 足場はずしレベルごとの入力フォーム（上から Lv.3, 2, 1）

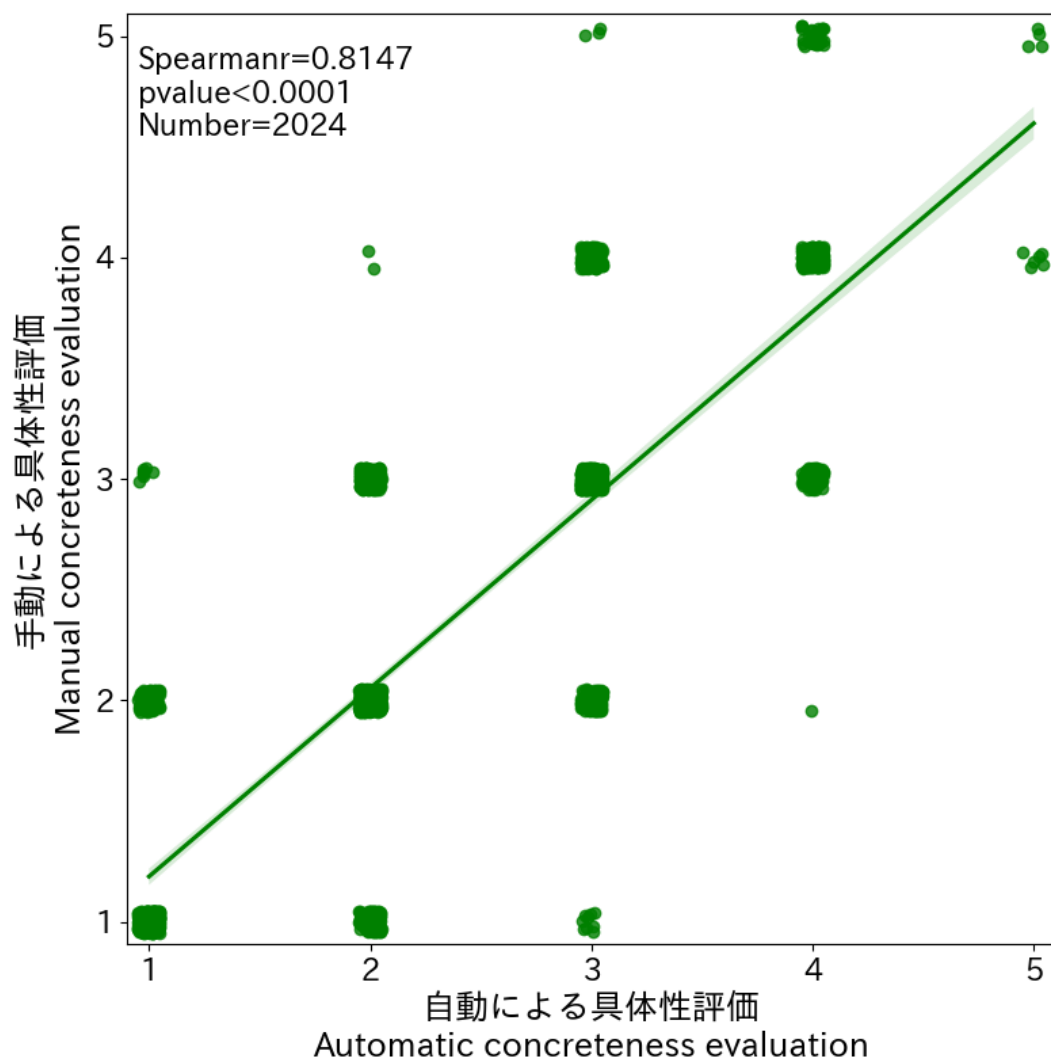


図4.7: 全ての具体性評価に関する手動評価と自動評価

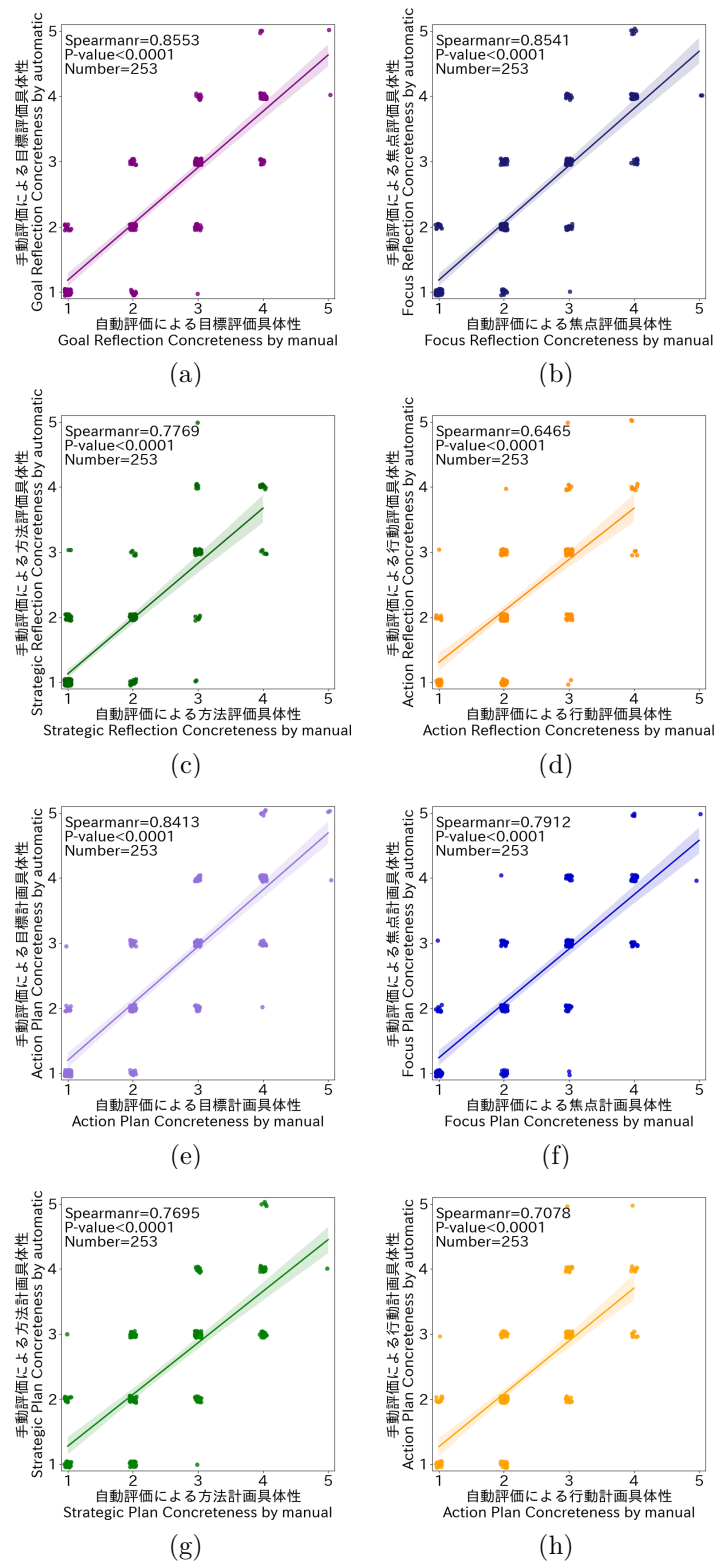


図4.8: 項目ごとの具体性評価に関する手動評価と自動評価

表4.6: 手動と自動の具体性評価が一致した記述内容

種類	記述内容（足場はずしレベル）	評価
1. 目標評価	前回の計画に対して順調に進める ことことができ、今回はディジ ョンツリー上で文字設定とルール 遷移設定を終わらせることがで きた。(Lv.3)	4
2. 焦点評価	本当にその選択肢はターゲットユ ーザーを想定して作成できている かを考えることが一番注目したこ とである (Lv.3)	4
3. 方略評価	条件について考えることでスム ーズに学習用データの作成を進める ことが出来た。(Lv.3)	4
4. 行動評価	計画したように行動できた。計画 した行動で目標の課題を終えた。 計画した行動は自分のスケジュール にあっていたと思う。(Lv.1)	3
5. 目標計画	まずは画像の表示設定を行ってい きたいです。画像の表示設定が終 われば次の課題、と段階を踏んで やっていきたいです。(Lv.3)	4
6. 焦点計画	言葉の表示設定をブラッシュアップ していくことに注目していきたい (Lv.2)	3
7. 方略計画	同じような素材でもそれぞれ違う 角度から写真を比べた時に出てく る条件があるためそれをたくさん 出し、データに反映していく (Lv.3)	4
8. 行動計画	いつどこでやるかのイメージがで きている。いつどこでやるかを調 整しておく必要は今のところなさ そう。今回の課題は負担がありそ うに思う。(Lv.1)	2

表4.7: 手動と自動の具体性評価に誤差のあった記述内容

種類	記述内容（足場はずしレベル）	手動	自動
1. 目標評価	公欠で休んでしまったので、取り返していきたい。(Lv.3)	1	3
2. 焦点評価	ディジョンツリー (Lv.3)	1	3
3. 方略評価	あまりうまくいかなかった。実行テストを行っても、言葉と出てくる文字の不一致がでてしまったため、何度もテストを行い修正を重ねるべきだと思う。(Lv.3)	5	3
4. 行動評価	テストでちゃんとできるようにする。(Lv.3)	1	3
5. 目標計画	Excel を終え、デッシジョンツリー作成に入る (Lv.3)	2	4
6. 焦点計画	著作権 (Lv.3)	3	1
7. 方略計画	以下に物を売るか (Lv.3)	1	3
8. 行動計画	毎日する事が大事である。学校、バイト終わりにする授業の空きコマなどを使いやってみる。(Lv.3)	5	3

表4.8: 想定外の記述内容への自動評価事例

種類	記述内容	手動	自動
目標評価	3DCG でのロゴアニメーションを作って YouTube にアップロードしたものを設定した。	4	4
焦点評価	平面だったロゴを綺麗に立体化するにはどうするか。	4	3

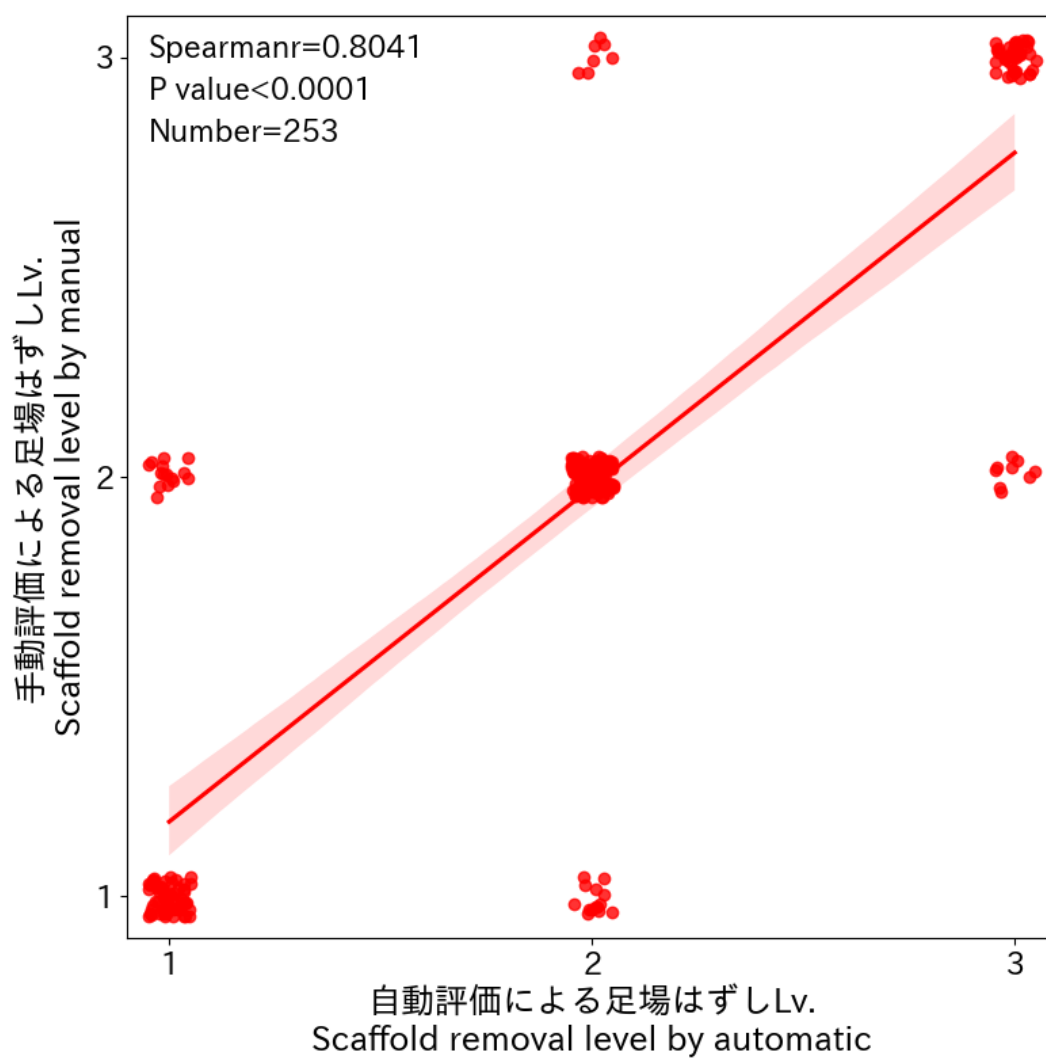


図4.9: 足場はずし Lv に関する手動評価と自動評価

表4.9: 手動評価と自動評価に関する足場はずし Lv. 件数

Method of evaluation	Lv.1	Lv.2	Lv.3
手動 Manual	73	134	46
自動 Automatic	74	131	48

表4.10: 手動評価と自動評価に関する足場はずし Lv. 誤差

Lv. by automatic - Lv. by manual	Count
-1	22
0	208
1	23

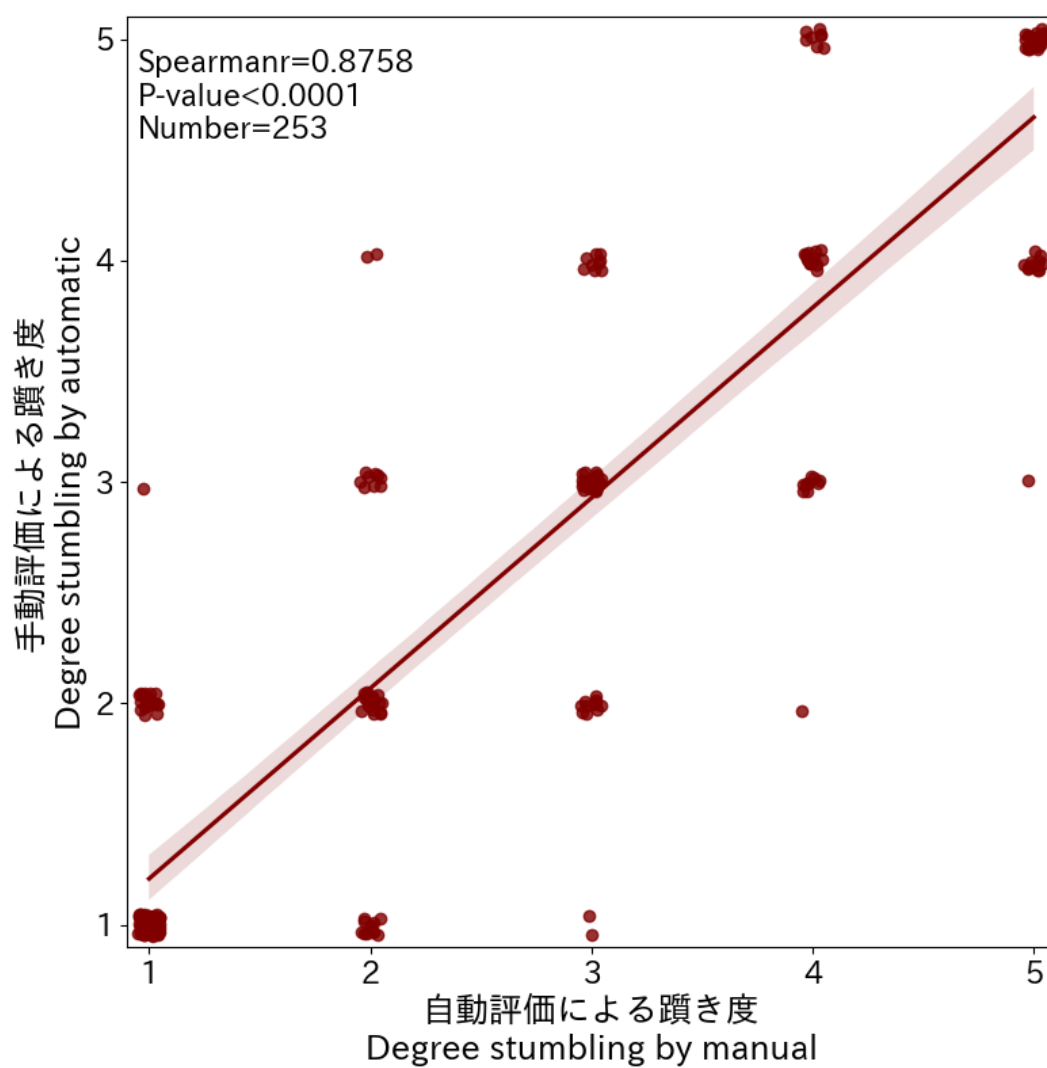


図4.10: 躓き度に関する手動評価と自動評価

表4.11: 躰き度評価の誤差

Difference	All Lv.	Lv.1	Lv.2	Lv.3
-2	3	0	3	0
-1	46	10	30	6
0	154	43	76	35
1	46	18	23	5
2	4	2	2	0

表4.12: 手動と自動による躰き度評価が一致した記述例

記入項目	記述内容
目標評価	実際にディジションツリーが稼働するかや、条件の設定をした。
焦点評価	条件の仕組みを工夫することに1番注目した。
方略評価	かなりうまくできたので同じ方法、考え方で次回も乗り切りたい
行動評価	授業前やバイトの休憩時間
目標計画	今回の課題はかなり複雑な内容なので3つの段階ごとに分けて徐々に単純化していくやり方でいきたい。
焦点計画	今回の課題はかなり実践的な内容なので本番をイメージしながら調整していくと良いと思った。
方略計画	ただ取り組むのではなく、第三者目線なども考えて多面的に情報を汲み取る考え方で取り組みたい。
行動計画	授業の始まる前や休日、バイトの休憩時間

表4.13: 手動と自動による躰き度評価に誤差（手動 4，自動 2）のあった記述例

記入項目	記述内容
目標評価	課題用データをアップロードしようと思ったが、エクセルの行数が足りないことや、写真を変更しなければいけないことに気づいたので、あまり進むことはできなかった
焦点評価	どのように別のルートで画像を表示させるか
方略評価	あまりうまくいかなかった
行動評価	月・水・金の授業後
目標計画	まずは一週間の前半でデータのアップロードまで終了させ、後半で細かな調整を行う
焦点計画	どれだけ自分のイメージと近い演出にできるか
方略計画	授業資料と照らし合わせ、適切な方法で進める
行動計画	月・水・金の授業後

第 5 章

GPT-4 を用いた自動評価に基づく学生個別アドバイスの自動生成

5.1 序言 自動評価に基づく学生個別アドバイス自動生成

自己調整学習の予見と自己省察について，自動支援を実現するためには，学習スキルに沿いつつ，学生個別の予見と自己省察の内容に追従して具体的なアドバイスを自動生成する手法が必要である．従来手法における定量的かつ定型的なアドバイス提示では，学習時間，小テスト，文字数，学習行動数などの傾向に基づく分析によって客観的な学習傾向を捉えることができる．そのため，学習支援に最適な学習方法の状況や傾向を客観的に捉えることができる．しかし，答えや解法が定まらない学習においては，学生個別に異なる学習目的，焦点，方法などの学習実態に沿った支援が重要になる．

そこで，本章では，表 2.1 の項目 6 に示した，学生個別の記述内容に追従した助言自動生成手法の実現を目的とした．そのために，学生が記述した学習計画と振り返りの内容に追従しつつ，第 4 章 4.3 節に示したループリックにも沿った具体的な改善アドバイスを自動生成するための手法を提案する．

提案手法では，アドバイスを学生の記述内容に追従させるために，大規模言語モデル GPT-4 のゼロショット学習を活用する．また，大規模言語モデルのゼロショット学習を用いて目的のアドバイスをうまく生成するためには，適切な入力プロンプトが必要になる．本研究の入力プロンプトでは，ループリック評価，アドバイス方針，学生の記述内容，学生個別のテーマと学生のガントチャートにあるタスク情報で構成する．そのために，アドバイス生成ごとに学生個別の入力プロンプトを構成する機構が必要になる．そこで，学生個別の入力プロンプトの作成に必要な学生個別のデータの収集と，入力プロンプトを自動的に構成，そして，入力プロンプトを大規模言語モデルに入力して助言を自動生成する機構を開発し，学習支援システムに実装した．

本章の課題は，従来手法 [17] の課題となった学生の記述内容が不十分な場合でも，学生

の学習状況と学習内容に沿った具体的なアドバイスを自動生成することである。そのために、入力プロンプトに学生の記述内容だけでなく、ルーブリック評価、学生の記述内容、学生個別のテーマと学生のガントチャートにあるタスク情報を併用している。

さらに、本章では、第4章で確認された自動評価の有効性を踏まえ、足場かけと足場はずしの概念を発展させた。具体的には、学習スキルの段階により細かくアドバイスの段階を調整する。なお、第4章では、記述項目ごとの評価値の合計に閾値を設けて、全ての記述項目を対象に一律の足場かけと足場はずしを行っていた。これを第5章では、記述項目ごとのルーブリック評価に沿って、アドバイスの段階を調整するものとした。そのために、入力プロンプトには、ルーブリック評価ごとのアドバイス方針を追加している。

本章において、ルーブリックに沿ったアドバイスを学生の記述内容や学習内容に追従させる方法を採用した理由は2つある。

1つ目は、学生の学習計画と振り返りのルーブリック評価に基づき、学習計画と振り返りの入力方法を3段階に分けて段階的に変更する足場かけ、足場はずし手法を試作し、試用したところ、足場かけの効果が見られなかったためである。具体的には、入力方法をルーブリック評価が低い順から、選択肢の提示、記述内容の雛形文提示、自由記述形式と切り替える手法を用いたところ、選択肢の提示だけでは、学習状況と学習方略の状況を抽象的にしか入力できず、雛形文提示では、雛形に含まれる学習方略を自身の学習状況に適用した内容を記述することが困難であった。

2つ目は、学習者自身が学習状況をうまく把握できていない場合、定型文のアドバイスだけでは、提示された学習方略と自身の学習状況を結び付けることが困難なためである。

以上の2つの理由によって、アドバイスを学生の記述内容に追従するように反映させる手法が必要であると考えた。

学生の記述内容は、目標設定、注意の焦点化、方略計画の利用と調整という3つの学習方略に対応する6つの項目に分類されている。

1. 目標の計画
2. 焦点の計画
3. 方略の計画
4. 目標の振り返り
5. 焦点の振り返り
6. 方略の振り返り

図 5.1は、ルーブリック評価を1段階向上させる足場かけと足場はずしの概念図である。ルーブリック価値ごとに対応したアドバイス方針を用いて、図 5.1に示したように、評価ご

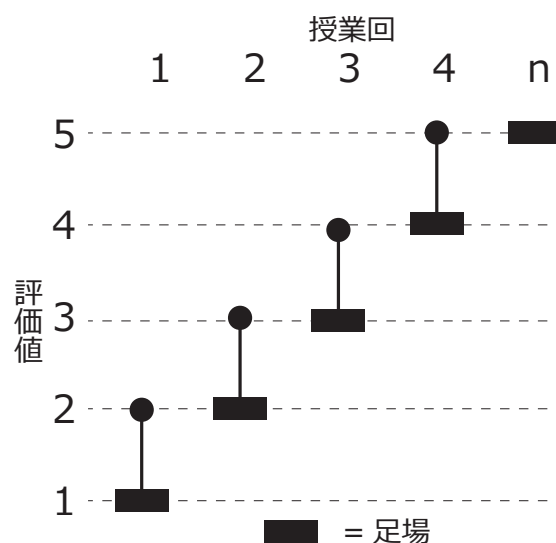


図5.1: 評価を1段階向上させる足場かけと足場はずし

とに，ループリック評価を1段階向上させるアドバイスを自動生成する．つまり，ループリック評価に応じて提示されるアドバイスを足場かけとする．

5.1.1 自動的なアドバイス生成による自動学習支援システム

本章では，学生の学習計画や振り返りを支援するために，学生の学習計画と振り返りの記述内容に追従する改善アドバイスによって，自動的に支援を行う仕組みの実現を目指している．自動的なアドバイス生成による自動学習支援を実現するためには，以下3つの機構を組み合わせたシステムが必要となる．

1. アドバイスの生成に必要な学生の学習情報を自動収集して記録する機構
2. 自動生成に必要なシステムメッセージとユーザーメッセージを作成する機構
3. 生成したアドバイスを学生に自動提示する機構

図5.2は，3つの機構による自動学習支援システムを示す．本章の自動生成に使用するゼロショット学習に inputs するプロンプトは，メッセージのリストである．メッセージのリストは，システムメッセージとユーザーメッセージで構成される．本章のメッセージリストは，学生個別のアドバイス生成ごとに作成する必要がある．

図5.2は，本章で開発して用いた学習支援システムの機構と，教員と学生とチャットボットとの関係を示している．学習支援システムに実装したチャットボットは，学生を支援対応

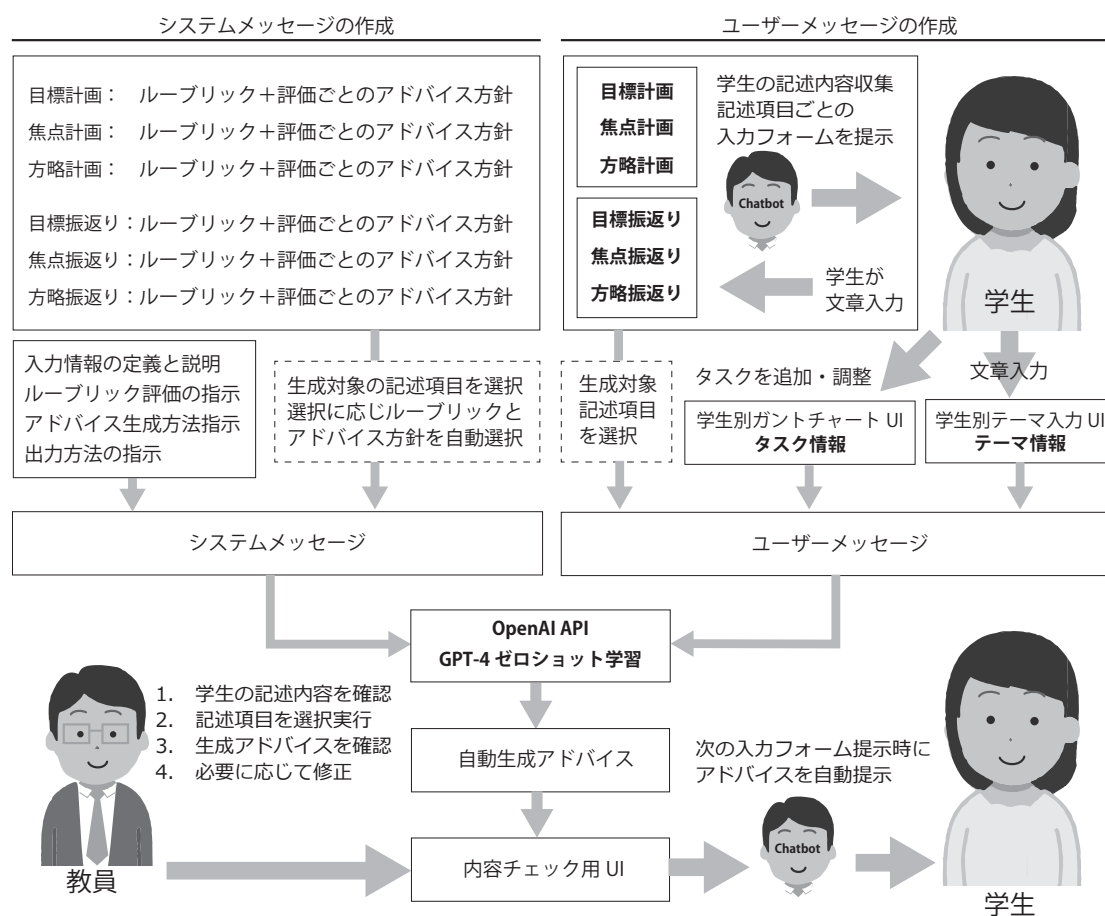


図5.2: 自動的なアドバイス生成による自動学習支援システム

するエージェントとなる。

チャットボットは、対話的に入力フォームを自動提示してユーザーメッセージの作成に必要な記述内容を収集する。また、ユーザーメッセージの作成に必要なタスク進捗情報についても対話的に入力フォームを提示して更新を促す。さらに、記述内容にアドバイスがフィードバックされた場合、学生に提示する。

学生は、チャットボットに提示された入力フォームへの入力と、タスク情報とテーマ情報の入力や更新、フィードバックされたアドバイスの確認を行う。

教員は、学生が入力した内容の確認とアドバイス自動生成の実行指示、生成されたアドバイスの確認と修正を行う。なお、アドバイスの修正については、本研究の実験では行っていない。

学習支援システムは、教員の指示によって、システムメッセージとユーザーメッセージを

自動的に作成して、大規模言語モデルに入力し、アドバイスを自動生成する。

システムメッセージとユーザーメッセージの作成について詳細を説明する。

システムメッセージの作成

システムメッセージの作成では、アドバイス生成についてのエージェントの役割と、ユーザーメッセージで入力される情報の定義と説明、ループリック評価の指示、アドバイス生成の指示、出力方法の指示に加え、記述項目ごとにループリックとループリック評価値ごとのアドバイス方針を自動選択して与える。

図 5.2 のシステムメッセージの作成は、ループリックとループリック評価値ごとのアドバイス方針のセットが記述項目に対応して自動選択される。ループリックとループリック評価値ごとのアドバイス方針のセットを自動選択して与えている理由は、記述項目ごとに生成するアドバイスの品質を高めるためである。なぜならば、ここで全てのループリックとアドバイス方針を一度に含めると、混同したアドバイスが生成されるなど、生成結果の品質が安定しないためである。

なお、図 5.2 のループリックは、第4章4.3節のループリックを簡略化したものである。また、図 5.2 の評価値ごとのアドバイス方針は、付録 A に示すアドバイスの雛形文を参考に作成したものである。

ユーザーメッセージの作成

ユーザーメッセージの作成では、学生から必要な情報を収集するための機構と、収集した情報を用いてユーザーメッセージを作成する 2 つの機構を備える必要がある。

はじめに、ユーザーメッセージの作成に必要な情報を収集する機構について説明する。

付録 D に示した対話画面は、図 5.2 のユーザーメッセージ作成にあるチャットボットの対話例である。チャットボットは、学生に学習計画と振り返りをテキストで聞き取る入力フォームを提示する機構を備えており、フォームに入力されたテキストデータは、データベースに保管される。記述項目ごとの学生の記述を効率よく収集するため、授業回ごとにチャットボットが対話的に記述項目ごとの入力フォームを提示する。学生は、チャットボットによって提示された入力フォームに、記述内容に応じた内容を文章で入力する。

図 5.3 に示した入力フォームは、図 5.2 の学生別テーマ入力 UI である。赤点線枠 A は、学生が個別のテーマを入力するテキストフォームである。

図 5.4 に示したガントチャートの画面は、図 5.2 の学生別ガントチャート UI である。赤点線枠 B の上部は、日付がある横スクロールのカレンダーになっている。

赤点線枠 B の 2 行目は、学習計画と振り返りが記入されていることを示す記入マークが表示される。ここで、学生や教員が記入マークの 1 行目を選択すると学習計画、2 行目は振

クラス名

卒業研究2023

プロジェクト名

自己肯定感診断のpersona

プロジェクトの概要

タイトル：ネガティブ思考をポジティブに変換するサービスの提案

取り組む課題

- 嫌な事、悩み事などネガティブに考えてしまう事をポジティブに考えられるサービスを作る事

作るもの

- ターゲットユーザーの嫌な事、悩み事に対し、ポジティブな言葉を返答するサービス

出来てから効果を知る為に検証する事

- サービスを使う前、使った後にアンケートを答えてもらい、サービスの課題を見つける

図5.3: 学生個別のテーマ入力画面

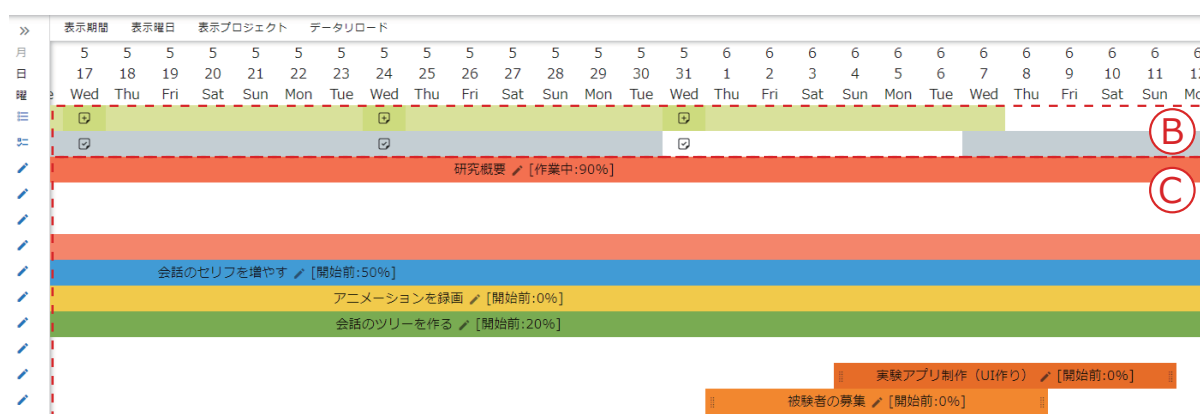


図5.4: 学生個別のタスク進捗管理画面

り返りを確認するダイアログが表示される。

赤点線枠 C は、タスク進捗管理用のガントチャートになっている。学生は、タスクの追加、進捗状況の更新ができる。既存するタスクの進捗状況については、付録 D の自動的な聞取りによっても更新される。

タスクのバーには、開始前や作業中などのステータス項目も見られるが、今回はタスクのステータス情報を使用していない。学生がステータス情報をあまり活用していなかったためである。

図 5.5 は、振り返り内容の確認画面の一部である。学習計画内容の確認画面も基本的な構

造は同じであるが、項目名が「計画」に変更された画面となる。

赤点線枠 D は、学習計画や振り返りの対象となる期間が設定される。付録 D の自動的な聞取りによって入力される。

赤点線枠 E は、学生が記述した記入項目の 1 つである。付録 D の自動的な聞取りによって、学生が入力したテキストが反映される。振り返りの記述項目は、3 項目（目標、焦点、方略）ある。学習計画の場合も 3 項目（目標、焦点、方略）である。つまり、学習計画と振り返りで 6 つの記述項目がある。

赤点線枠 F は、教員がアドバイスを記入する部分である。記入項目ごとにアドバイスを作成し、保存すると次回授業時の自動的な対話の際に学生に提示される。「GENERATE」ボタンは、アドバイスの自動生成を実行して、フィードバック欄に生成された結果を表示する。自動生成後に教員が目視で確認し、必要に応じた修正を想定している。なお、本章の実験において、自動生成されたアドバイスには一切の修正を加えていない。

以下に説明する画面の項目は、関連研究 [29], [17] で使用した項目であり、第5章では使用していない。前回の目標計画の「OPEN」、「CLOSE」ボタンは、前回の学習計画の目標項目を教員が参照できるように設けた。星マークは、学生の記述内容の具体度を教員が手動で評価するために設けた。「TEMPLATE」ボタンは、アドバイスを教員が人手で作文するとき、ループリックの評価値ごとのアドバイスの雛形文を教員が参考にするために設けられた機構である。

5.2 アドバイス自動生成のための GPT-4 とプロンプトの設計

次に、収集した情報を用いてユーザーメッセージを作成する機構について説明する。

ユーザーメッセージの作成では、選択された記述項目ごとの学生の記述内容とテーマ情報、タスク名とその進捗情報の 3 つの情報を用いて構成される。

選択された記述項目とは、アドバイス生成の対象とする記述項目を教員が都度手動で選択した項目である。選択された記述項目の記述内容は、その全文がユーザーメッセージに与えられる。この時選択された項目が振り返り項目の場合は、振り返り期間に該当する前回の学習計画の同記述内容も自動的にユーザーメッセージに加えられる。

テーマ情報は、入力されている全文がユーザーメッセージに与えられる。テーマ情報を用いた理由は、テーマは学生の最終目標や方向性や方針などの長期的な情報を含む。そのため、学生の毎回の学習計画や振り返りの抽象度が高い場合でも、長期的な目標や方針、方向性を用いたアドバイスの生成に期待できるためである。

タスク情報とは、記述項目に対象期間内に存在するタスクを対象として、タスクの名称と進捗情報がユーザーメッセージに与えられる。タスク情報を用いた理由は、タスクは学生が

振り返り

(D)

前回から昨日までの振り返りをしましょう

振り返りの対象期間はじまり
📅 2023-05-10

振り返りの対象期間おわり
📅 2023-05-16

(E)

目標の振り返り
アバターを完成させること (優先)
会話のセリフを増やすこと

(F)

前回の目標計画

OPENCLOSE

★ ☆ ☆ ☆ ☆

目標の振り返りフィードバック

🗨️ ちょっと待って、君の目標の振り返りが足りないよ。アバターを完成させることと会話のセリフを増やすことが目標だったんだよね？ それに対する進捗状況や自己評価を書いてみよう。例えば、「アバター作りは90%まで進んで、あとは細部の調整が残っている。セリフ出しは50%で、まだまだ追加するセリフが必要だ」とかさ。

TEMPLATEGENERATE

図5.5: 学習計画, 振り返り確認画面

スケジュールした行動計画や進捗情報など、テーマと比較して短期的な情報となる。そのため、学生の毎回の学生計画や振り返りの抽象度が高い場合でも、短期的な行動計画と進捗状況を用いたアドバイスの生成に期待できるためである。

初回授業時にタスクを入力させ、毎回の授業において学生が自身の必要に応じてタスクの追加や進捗状況の更新、日程の調整を行った。初期設定では、学習計画の場合は記述時から1週間後までのタスク、振り返りの場合は記述時から1週間前までの期間に存在するタスクが対象となる。なお、初期設定の対象期間は学生が調整できるため一定ではない。

生成したアドバイスを学生に自動提示する機構

自動生成したアドバイスの提示は、図 5.2のチャットボットが学生に提示する機構を備える。

ただし、自動生成したアドバイスは必ずしも適切な内容とは限らない。そのため、事前に教員による確認を要する機構が必要である。教員は、図 5.2の内容チェック用の UI を用いて、学生個別の記述内容を確認する。確認後、記述項目ごとにアドバイスの自動生成を実行する。生成されたアドバイスは、再度内容チェック用の UI を用いて確認後、必要に応じて修正を加える。確認済みの自動生成アドバイスは、次の学生の記述内容収集時に、チャットボットから学生に提示される。なお、アドバイスの修正については、本研究の実験では行っていない。

5.3 アドバイス自動生成の予備実験

学生の記述内容とループリックを用いた従来の手法 [17] では、学習計画や振り返りの記述が不十分な場合、具体的なアドバイスを生成することに課題があった。

従来手法 [17] で明らかになったトレーニングデータの不足と偏り課題に対応するため、アドバイス自動生成に OpenAI の大規模言語モデルのゼロショット学習を用いる方法に着目した。なぜなら、ゼロショット学習では、ファインチューニングに用いるトレーニングデータに潜在する偏りの影響を受けることなく、入力プロンプトに基づいて、アドバイスを生成することを期待できたためである。

そこで、予備実験では、gpt-3.5-turbo のゼロショット学習を用いたアドバイス自動生成について以下の 2 点を確認した。

ループリックに沿ったアドバイスが生成できたかどうか 自動生成したアドバイスは、教員手書きのアドバイスに対してどの程度アドバイスの雛形文に沿った内容になっているか

学生にとって具体的なアドバイスが生成できたかどうか 自動生成したアドバイスは、学生に自分の記述内容にあい、改善点がわかりやすく、改善できると思わせることができるか

5.3.1 予備実験の計画

予備実験では、gpt-3.5-turbo を用いたアドバイスと、教員が人手で作文したアドバイス、そして、GPT-3 ファインチューニングモデルを用いたアドバイス [17] の 3 種類のアドバイスについて、学生によって回答された評価を比較した。

アドバイス評価を回答した被験者は、アドバイス生成の元となる記述内容を記述した学生である。また、被験者に 3 種類のアドバイスの作成方法は開示せず、被験者への提示順もランダムとした。

予備実験の評価に使用したデータと自動生成アドバイスについて説明する。予備実験は、学生の記述内容は、従来手法 [17] で使用したデータを用いた。予備実験で用いた gpt-3.5-turbo は、従来手法 [17] のようにファインチューニングを実施しない。

gpt-3.5-turbo を用いたアドバイスの自動生成では、入力プロンプトで指示を与える。gpt-3.5-turbo に入力したプロンプトは、メッセージのリストである。本稿では、C.1 に示すシステムメッセージ 1 つと、C.2 に示すユーザーメッセージ 1 つから成るリストをプロンプトとして用いた。

システムメッセージは、アドバイスを生成する方法と扱う項目についての定義とループリック、アドバイス方針、そして、生成するシステムの役割について指示を与えている。ユーザーメッセージでは、アドバイスの生成するための学生の記述内容を渡している。

学生の記述内容は、実際の授業において学生が入力したテキストデータを用いた。学生の記述内容は、図 5.2 に示した学習支援システムの予備実験段階を用いて、学生の記述内容収集方法部分によって入力された。なお、予備実験段階の学習支援システムは、図 5.2 に示した学習支援システムよりも、行動計画の計画と振り返りの記述項目が 2 つ多く、8 つの記述項目であった。8 つの項目とは、目標計画、焦点計画、方略計画、学習行動計画、目標振り返り、焦点振り返り、方略振り返り、学習行動振り返りである。そのため、授業回（最大 15 回）ごとに学生によって入力された 8 つの項目のテキストデータを用いた。

アドバイスを自動生成する際は、8 つの項目ごとに、gpt-3.5-turbo を用いて、temperature=0 で自動生成を実行した。

予備実験では、以下の 3 種類のアドバイスについて、学生の評価を比較した。

- gpt-3.5-turbo を用いて自動生成したアドバイス
- GPT-3 ファインチューニングモデルを用いたアドバイス
- 教員手書きのアドバイス

なお、教員手書きのアドバイスは、学生の記述内容を確認し、ループリックにしたがって具体生の評価を行った後に、アドバイスの雛形文を参考にして作文された。アドバイスの雛形文とは、ループリックに基づいてあらかじめ教員が作文したアドバイスの雛形であり、ループリックに沿っているテキストであると仮定したものである。

アドバイスの雛形文は、教員が、ループリックに沿ったアドバイスを作文できるようにアドバイスの雛形文 [17] を事前に用意した。付録付録 A にアドバイスの雛形文を示した。

ループリックに沿ったアドバイスかどうかの評価では、雛形文と自動生成したアドバイスの Embedding 間のコサイン類似度と、雛形文と教員手書きのアドバイスの Embedding 間のコサイン類似度を比較して評価した。このとき Embedding には、OpenAI の埋め込みモデル text-embedding-ada-002 を用いた。また、生成されたアドバイスについてループリックに沿った内容ができていたかどうかについて確認した。

比較評価には、1つの学生の記述項目に対して、3種類のアドバイスが揃った 1433 件を用いた。3種類のアドバイスが揃っていない場合は欠損値として除外した。

学生の評価では、被験者は、自分の記述した内容にあっており、改善点がわかりやすく、改善できると思わせることができるかどうかを学生に回答してもらうアドバイスの評価を行った。アドバイスの評価の回答は、オンラインアンケートを通じて収集された。オンラインアンケートでは、被験者自身の記述内容の提示と、その記述内容に対応する、3種類のアドバイスがランダムな順序で提示された。

3種類のアドバイスについて、以下の3つの項目（適合性、理解性、実行性）について被験者の回答を得た。

適合性 アドバイスの内容は私の記述内容にあっている

理解性 記述内容のどこをどのように改善すべきかわかった

実行性 このアドバイスに従えば記述内容をすぐに改善できるとおもう

得られた回答の評価には、3種類のアドバイスについて、3つの項目の回答が揃っていた 598 件を用いた。3つの項目の回答が揃っていない場合は欠損値として除外した。

5.3.2 予備実験の結果

自動生成したアドバイスの件数は、1433 件得られた。なお、アドバイスがうまく生成されなかった 36 件はあらかじめ除外している。

図 5.6は、使用した学生の記述内容のループリック評価について評価別の頻度を示している。図 5.6からは、約 9 割がループリック評価 3 以下に偏っていたことが読み取れる。つまり、評価 3 以下の学生が多いため、評価 3 以下に対応することが重要であることが読み取れる。図 5.7は、被験者の回答が得られた 598 件について、学生の記述内容に対するループリック評価の頻度を示している。図 5.7からは、表 5.6と同様にループリック評価 3 以下に偏っていた傾向が示されている。つまり、被験者の回答が得られた 598 件について、元となる記述内容のループリック評価は、全体と同様の傾向であったことが読み取れる。

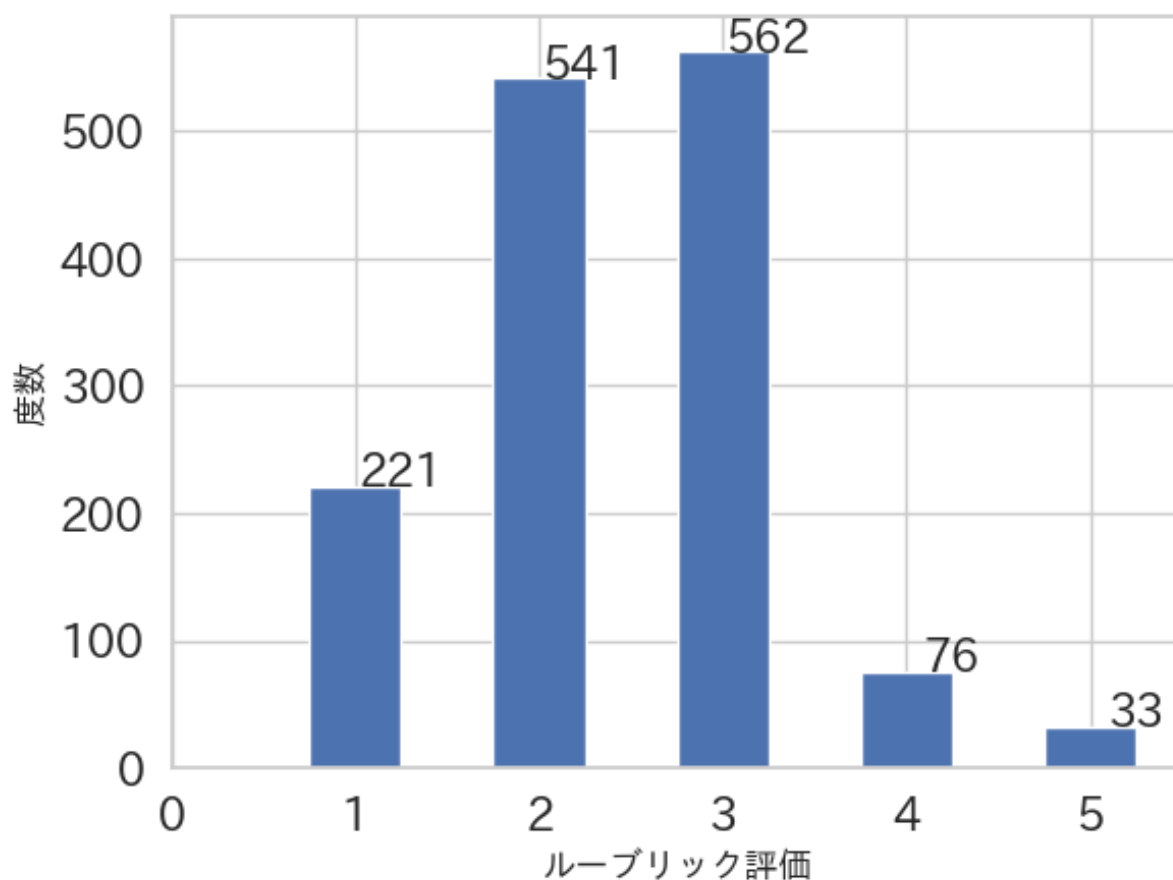


図5.6: 予備実験：ループリック評価別の頻度

予備実験：ループリックに沿ったアドバイスかどうかの結果

図 5.8は, gpt-3.5-turbo のゼロショット学習によって生成されたアドバイスが, ループリックに沿ったアドバイスになっているかを確認した方法を示している. 具体的には, gpt-3.5-turbo のゼロショット学習によって生成されたアドバイスと雛形文の Embedding 間のコサイン類似度平均を, 教員手書きのアドバイスと雛形文の Embedding 間のコサイン類似度平均と比較した.

表 5.1からは, gpt-3.5-turbo のゼロショット学習によって自動生成したアドバイスと, 雛形文の意味的な近さの平均 (コサイン類似度平均 = 0.876, std=0.032, n=1433) は, 教員手書きのアドバイスと雛形文の意味的な近さの平均 (コサイン類似度平均 = 0.874, std=0.029, n=1433) と同程度であったことが示された. なお, 従来手法 [17] の GPT-3 davinci FT モデルと比較しても同程度の意味的な近さを示している. したがって, gpt-3.5-turbo のゼロショット学習によって自動生成したアドバイスと雛形文の意味的な近さは, 教員手書きのア

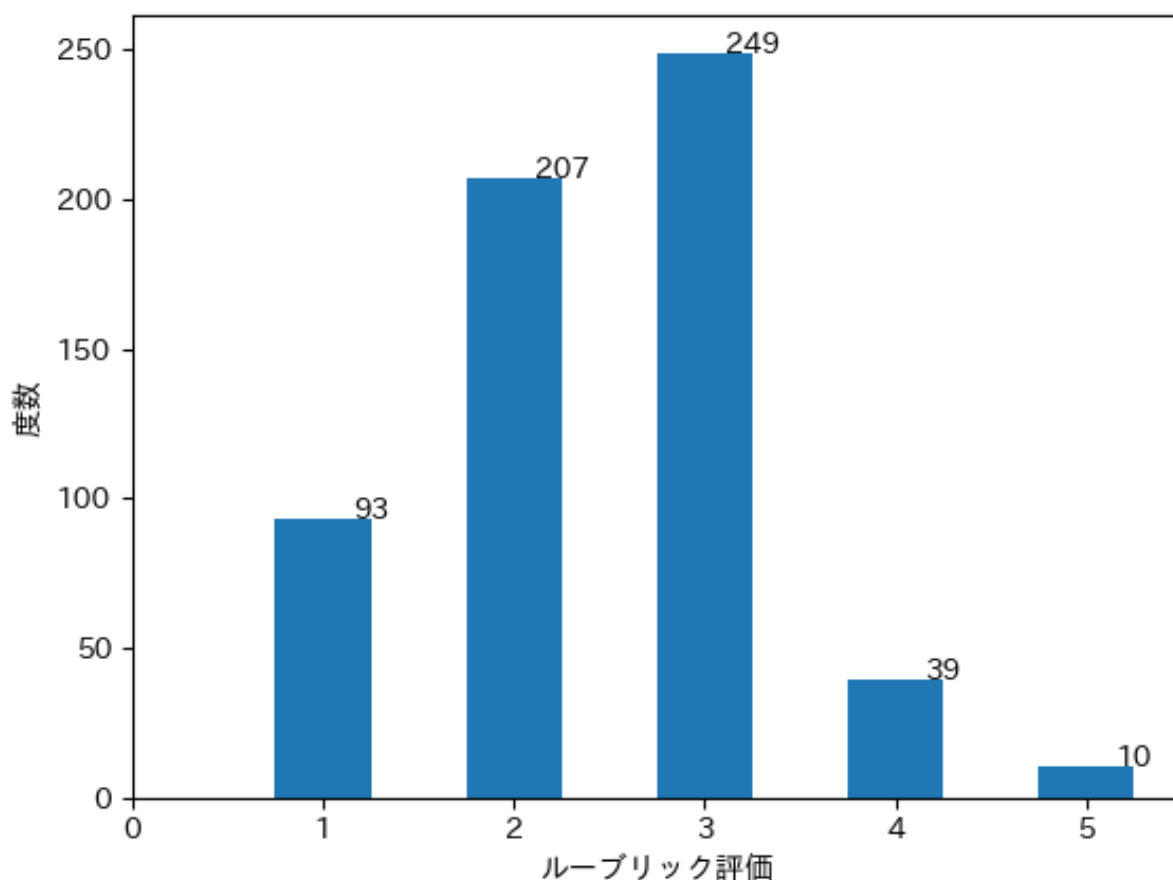


図5.7: 予備実験：回答者のループリック評価別の頻度

ドバイスと雛形文と同程度にできていた可能性がある。

さらに、生成されたアドバイス内容を確認してループリック評価に即したアドバイスになっていたかを確認した。表 5.2からは、概ね目標計画のループリック具体性評価に沿って、評価を1段階高めるアドバイスが生成されていたことを確認した。表 5.2のアドバイスは、以下のアドバイスをしている

評価1のアドバイス 目標としている記述が見つからないことから、一般的な目標を挙げたアドバイスをしている。

評価2のアドバイス 目標が遠い、あるいは、抽象的であることから、具体的な目標になるよう、現状の問題や解決手法を確認するようアドバイスしている。

評価3のアドバイス 目標を達成するための、過程を踏まえて目標が記述できるよう、学生の記述した「課題を進める」に関連する、課題遂行のステップを事例に挙げてアドバ

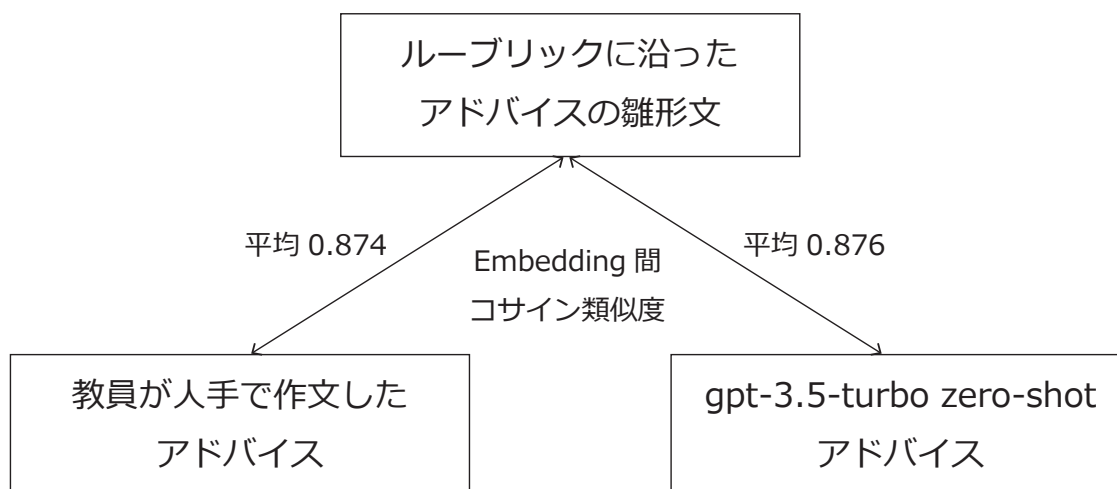


図5.8: アドバイスと雛形文との Embedding 間のコサイン類似度平均の比較

表5.1: 予備実験：アドバイスと雛形文のコサイン類似度

アドバイス生成方法	平均	SD	n
gpt-3.5-turbo zero-shot learning	0.876	0.032	1433
GPT-3 davinci FT モデル	0.872	0.027	1433
教員が人手作文	0.874	0.029	1433

イスしている。

評価 4 のアドバイス 目標達成の条件や見込みを具体的に記述できるよう，学生の記述した「事例作成のためのデータを集める」に関連する，データの扱いに関する条件の検討を促すアドバイスをしている。

評価 5 のアドバイス 目標と達成の条件や見込みを具体的に記述をピックアップしつつ，内容のより具体化を促すアドバイスをしている。

一方，目標を具体化する支援としては，学生が取り組む内容に関連する目標の例を提示したい。具体性評価 1,2,3 のアドバイスを見ると，抽象度の高い記述内容に対して，アドバイス内容が抽象化されており，学生が取り組んでいる内容に関連付けたアドバイスができていない。例えば，評価 1 のアドバイスは，誰にでも共通する「今週中に課題の提出を完了す

表5.2: 予備実験：目標計画ルーブリック評価ごとの自動生成アドバイス

評価	目標計画に対する自動生成アドバイス
1	具体的な目標を定めて、それを達成するための取り組みを考えて記述してください。例えば、「今週中に課題の提出を完了する」など、具体的な目標を定めてください。
2	目標をもう少し具体的に定めてみましょう。例えば、どのような問題があるのか、それを解決するためにどのような手法を使うのか、などを考えてみると良いでしょう。
3	課題を進めるために、まずはどのようなステップを踏む必要があるかを考え、それを段階的に記述してみましょう。例えば、課題の分析、情報収集、仮説の立て方など、具体的なステップを考えて記述することで、目標が明確になります。
4	今後は、自分が作成した事例をもとに、どのような条件で目標を達成するのか、達成見込みはどの程度かを記述することで、より具体的な目標の計画を立てることができます。例えば、作成した事例をもとに、どのようなデータを集める必要があるのか、そのデータをどのように分析するのか、分析結果をどのように活用するのかなど、具体的な条件を考えて記述してみましょう。
5	睡眠データを取得することで、ユーザーに適した音楽を研究し、質の高い睡眠をするために必要な要素を調べることができると考えられます。今後も睡眠データを取得し、データを分析することで、より具体的な成果を得ることができるでしょう。また、他の人からもデータを取得することで、より多角的な分析が可能になります。

る」という事例を挙げて、目標を具体化するように促している。

予備実験：学生にとって具体的なアドバイスかどうか

学生評価を回答した学生は、振り返りや学習計画を記入した学生本人 9 名である。教員手書きアドバイスの評価と gpt-3.5-turbo 自動生成アドバイスの評価の両方に 598 件の回答を得た。なお、片方のアドバイスのみに回答されていた 1 件は欠損値として除外している。

表 5.3からは、gpt-3.5-turbo を用いた自動生成アドバイスは、教員手書きのアドバイスよりも、適合性、理解性、実行性について、学生に自身の記述内容にあっており、改善箇所がわかりやすく、改善を実行できると思わせることができていることが読み取れる。適合性については、gpt-3.5-turbo の平均 5.56 (SD0.93) は、教員手書き平均 5.1 (SD1.1) よりも有意に高かった ($p = 1.22 \times 10^{-14} < .05$)。理解性については、gpt-3.5-turbo の平均 5.36 (SD1.1) は、教員手書き平均 5.05 (SD1.17) よりも有意に高かった ($p = 7.62 \times 10^{-8} < .05$)。実行性については、gpt-3.5-turbo の平均 5.23 (SD1.1) は、教員手書き平均 4.98 (SD1.24) よりも有意に高かった ($p = 0.0002 < .05$)。

また、表 5.4からは、gpt-3.5-turbo を用いた自動生成アドバイスは、従来手法 [17] である GPT-3 の davinci をファインチューニングしたモデルを用いたアドバイスよりも、適合

表5.3: 予備実験：教員手書きアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較

評価項目	教員手書き			gpt-3.5-turbo			比較
	平均	SD	n	平均	SD	n	
適合性	5.1	1.1	598	5.56	0.93	598	$1.22 * 10^{-14}$
理解性	5.05	1.17	598	5.36	1.1	598	$7.62 * 10^{-8}$
実行性	4.98	1.24	598	5.23	1.19	598	0.0002

表5.4: 予備実験：GPT-3 のアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較

評価項目	GPT-3 davinci FT			gpt-3.5-turbo			比較
	平均	SD	n	平均	SD	n	
適合性	5.2	0.98	598	5.56	0.93	598	$4.77 * 10^{-11}$
理解性	5.02	1.15	598	5.36	1.10	598	$8.23 * 10^{-9}$
実行性	5.01	1.16	598	5.23	1.19	598	$7.86 * 10^{-5}$

性、理解性、実行性について、学生に自身の記述内容に合っており、改善箇所がわかりやすく、改善を実行できると思わせることができていることが読み取れる。適合性については、gpt-3.5-turbo の平均 5.56 (SD0.93) は、従来手法平均 5.2 (SD0.98) よりも有意に高かった ($p = 4.77 * 10^{-11} < .05$)。理解性については、gpt-3.5-turbo の平均 5.36 (SD1.1) は、従来手法平均 5.02 (SD1.15) よりも有意に高かった ($p = 8.23 * 10^{-9} < .05$)。実行性については、gpt-3.5-turbo の平均 5.23 (SD1.1) は、従来手法平均 5.0 (SD1.16) よりも有意に高かった ($p = 7.86 * 10^{-5} < .05$)。

表 5.1では具体度評価が低いデータに約 9 割りの偏りが確認されたため、具体度評価の比較的高いデータ（評価 4, 5）の評価を確認した。表 5.5と表 5.6からは、学生の記述の具体度が高い場合には、統計的な有意差はなかった。したがって、学生の具体度評価が比較的高い場合については、アドバイスの学生評価を明らかにできていない。

表5.5: 予備実験：（具体度が高いデータ）教員手書きアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較

評価項目	教員手書き			gpt-3.5-turbo			比較
	平均	SD	n	平均	SD	n	$p < .05$
適合性	5.8	1.31	49	5.61	1.13	49	0.248
理解性	5.76	1.28	49	5.35	1.42	49	0.140
実行性	5.8	1.38	49	5.35	1.48	49	0.095

表5.6: 予備実験：（具体度が高いデータ）GPT-3 のアドバイスと gpt-3.5-turbo のアドバイスの学生評価の平均比較

評価項目	GPT-3 davinci FT			gpt-3.5-turbo			比較
	平均	SD	n	平均	SD	n	$p < .05$
適合性	5.80	1.02	49	5.61	1.13	49	0.418
理解性	5.57	1.22	49	5.35	1.42	49	0.503
実行性	5.53	1.26	49	5.34	1.48	49	0.644

予備実験のまとめ

予備実験の結果は、gpt-3.5-turbo を用いた自動生成したアドバイスは、教員手書きのアドバイスとほぼ相似して雛形文に沿った意味内容を生成できていた可能性がある。また、教員手書きのアドバイスよりも、学生自身の記述内容に合っており、改善箇所がわかりやすく、改善を実行できると思わせることができていたことを示していた。ただし、学生の記述内容が不十分な場合に具体的なアドバイスを生成することができていなかった。なお、予備実験の結果は、学生の具体度評価の低い場合に限り、学生の具体度評価が比較的高い場合については明らかにできていない。

5.4 学生個別のテーマとタスク情報を追加したアドバイス自動生成の評価

従来手法 [17] の問題点は 2 つ挙げられる。従来手法 [17] の問題点 (1) として、具体的なアドバイスを生成するための材料がなかった。そこで、学生の記述内容とループリックに加

えて、以下の 2 つのデータを併用した。

1. 学生の取り組みテーマ
2. 学生のガントチャートのタスク進捗状況

従来手法 [17] の問題点 (2) として、記述内容の具体度が低い場合に対応するトレーニングデータが不十分だったという問題があった。

そこで、予備実験の結果を踏まえ、OpenAI の大規模言語モデル GPT-4[28] を用いた。予備実験の結果は、ゼロショット学習の入力プロンプトを工夫することにより、学生の記述内容と関連性の高いアドバイスを生成できる可能性が示されていた。そこで、予備実験で用いた gpt-3.5-turbo よりも、高性能なモデルである GPT-4 を用いて、学生の記述内容が不十分な場合でも、学生個別のテーマやタスク進捗情報を併用することで、具体的なアドバイスを生成する手法について評価を行った。

学生個別の学習計画と振り返り、そして、個別のテーマ入力とタスク進捗の入力には、予備実験と同様に 付録 B に示した学習支援システムを用いた。

従来手法 [17] の問題は以下の 2 つである。

1. 記述内容の具体度が低い場合に具体的なアドバイスを生成するための材料がなかった
2. 記述内容の具体度が低い場合に対応するトレーニングデータが不十分だった

節5.3の予備実験では、gpt-3.5-turbo のゼロショット学習について、従来手法 [17] よりも平均的に高い評価結果が得られていた。しかし、予備実験では、記述内容の具体度が低い場合の検証は行っていない。

そこで、提案手法では、2 つの問題に対応するために以下の 3 つを工夫した。

1. 従来の学生の記述内容に加えて、学生の取り組みテーマ、そして、学生のガントチャートのタスク進捗状況を併用した
2. トレーニングデータに依存しないゼロショット学習を用いた。そのための入力プロンプトを設計した
3. より高性能な GPT-4 を用いた

記述内容の具体度が低い場合に具体的なアドバイスを生成するための材料がなかった問題に対して、学生の記述内容や学生の取り組みテーマ、そして、学生のガントチャートのタスク進捗状況を併用する。

記述内容の具体度が低い場合に対応するトレーニングデータが不十分だった問題について、トレーニングデータに依存しないゼロショット学習を用いた。なぜなら、節5.3の予備

実験では、具体度が低い場合の検証を行っていないため、本実験で詳しく検証する必要があるが、全体的な平均では、高い評価テーマとタスク情報を併用する工夫とトレーニングデータに依存しないゼロショット学習を組み合わせることで、記述内容の具体度が低い場合に対応できると期待したためである。

節5.3の予備実験では、gpt-3.5-turbo のゼロショット学習を用いたが、提案手法はより高性能な GPT-4 のゼロショット学習を用いた。

GPT-4 のゼロショット学習を用いてアドバイスを自動生成するために、入力プロンプトとして、システムメッセージ (role=system) と、ユーザーメッセージ (role=user) を与えた。

付録 E に、入力プロンプトの構成 (E.1), システムメッセージの実例 (E.2), ユーザーメッセージの実例 (E.3) を示した。

システムメッセージは、アドバイスを生成する方法と扱う項目についての定義、そして、生成するシステムの役割について指示を与えている。ユーザーメッセージでは、アドバイスの生成するための学生の記述内容とテーマ内容、そして、タスク進捗のテキストを渡している。

提案手法における第5章のコアとなる工夫は、学生のテーマとタスク進捗状況を併用する部分と、トレーニングデータを用いることなく GPT-4 を用いて、学生の記述内容の具体度を評価するループリックに基づくアドバイスの自動生成を行っている部分の2つである。2つの工夫について、具体的には E.1 に示したシステムメッセージの構成にある (2) ユーザーメッセージで渡すデータの書式と (5) アドバイスの作成方法に含まれている。

5.5 実験

記述内容を具体的に書くことができていない学生に、記述内容を改善させることができる具体的なアドバイスとは、以下の3つを備えるアドバイスとした。

適合性 学生の記述内容やテーマ、タスクに対して、生成されたアドバイスの内容が追従している

理解性 生成されたアドバイスの内容や説明が、学生にとってわかりやすい

実行性 学生が自分でも書き直して実行できると思える

そこで、自動生成されたアドバイスについて、以下3つの評価項目について、学生の評価を得た。

適合性 アドバイスの内容は私の記述内容にあっている

理解性 記述内容のどこをどのように改善すべきかわかった

実行性 このアドバイスに従えば記述内容をすぐに改善できるとおもう

3つの評価項目は、7件法（1＝非常にそう思わない、2＝かなりそう思わない、3＝ややそう思わない、4＝どちらでもない、5＝ややそう思う、6＝かなりそう思う、7＝非常にそう思う）を用いて学生によるアドバイス評価の回答を得た。

なお、第5章では、学生が実際に記述内容を改善したかどうかについては評価していない。学生が実際に記述内容を改善したデータを用いた評価については、今後の課題とする。

5.5.1 実験手法

実験では、「テーマとタスク情報」、「テーマ情報のみ」、「タスク情報のみ」の3パターンの与える情報を変えて生成した3種類のアドバイスについて、学生による評価を比較した。

テーマ＋タスク 学生の記述内容と学生個別のテーマとタスク進捗情報を組み合わせた

テーマ 学生の記述内容と学生個別のテーマ情報を組み合わせた

タスク 学生の記述内容とタスク進捗情報を組み合わせた

アドバイス生成の元となる学生の記述内容は、個別のテーマのデザイン学習に取り組む授業に参加した3年生と4年生26名の1509件を用いた。学生の記述内容とテーマ、タスク進捗情報は、図5.2に示した学習支援システムを用いて、実際の授業において学生が入力した新しいテキストデータを用いた。なお、本実験では、予備実験とは異なり、学生の記述する項目は6つとした。これは、学生の入力負担軽減のためである。6つの項目とは、目標計画、焦点計画、方略計画、目標振り返り、焦点振り返り、方略振り返りである。授業回（最大15回）ごとに学生によって入力された6つの項目のテキストデータを用いた。

アドバイスの自動生成には、OpenAIが提供するGPT-4モデルのバージョンgpt-4-0314を使用した。temperatureを0に設定し、5.2節のプロンプトを用いた。

なお、学生の記入がない授業回や記入項目については、アドバイスを生成しておらず、アンケートにも含まれていない。

アンケート調査は、元になった学生の記述内容を記述した26名の学生を対象に、オンラインアンケートで実施し、回答を収集した。

オンラインアンケートでは、アドバイス生成元になった記述内容を記述した被験者に3種類のアドバイスを提示した。このとき提示した3種類のアドバイスの生成方法は被験者に開示していない。また、表示順序もランダムで表示された。

提示された3種類のアドバイスには、1種類のアドバイスごとに、適合性、理解性、実行性の3つの質問項目が表示され、被験者が3つの質問項目について評価を回答した。

表5.7: 実験結果に用いた回答件数の内訳

アドバイス種類	適合性	理解性	実行性	計
テーマ+タスク	256	251	250	757
テーマ	256	251	250	757
タスク	256	251	250	757
計	768	753	750	2271

アドバイスの提示とアンケート調査は、成績への影響を考慮し、授業最終回以後に実施し、2週間の回答期間を設けてまとめて回答してもらった。

5.6 結果

本章の実験では、自動生成されたアドバイスが学生にとって具体的かつ理解しやすいものであるかを評価した。これは、序言で述べた研究の目的である、学生の記述内容の具体度を高めるアドバイスの自動生成の有効性を検証するためである。以下に示す結果は、この目的に対する手法の効果を示すものである。

表 5.7に実験結果に用いた回答件数を示した。

実験結果として、13名の2271件の回答を用いた。回答期間終了後、被験者15名から3224件の回答が得られたものの、比較条件を揃えるため以下の回答は除外した。

- テーマやタスク進捗情報の記入がされていない記述内容に対する回答の場合
- 同一記述内容に対して3種類のアドバイスへの回答が、1つでも欠けている場合

図 5.9は、全体の学生26名の記述内容についてルーブリック評価別の頻度を示している。図 5.10は、回答した学生13名の記述内容についてルーブリック評価別の頻度を示している。図 5.9と図 5.10からは、ルーブリック評価が3以下が9割以上であったことが読み取れる。これにより、予備実験と同様にルーブリック評価が3以下への対応の重要性が示されている。一方、図 5.9の全体と図 5.10の回答者の記述内容のルーブリック評価の分布傾向を比較すると、評価1から3の分布傾向が異なる。これは、回答した学生は、全体の傾向よりも若干やる気のある学生であったことを示唆している。なお、本実験ではルーブリック評価の妥当性については検証されていないため、ルーブリック評価の妥当性については今後の課題とする。

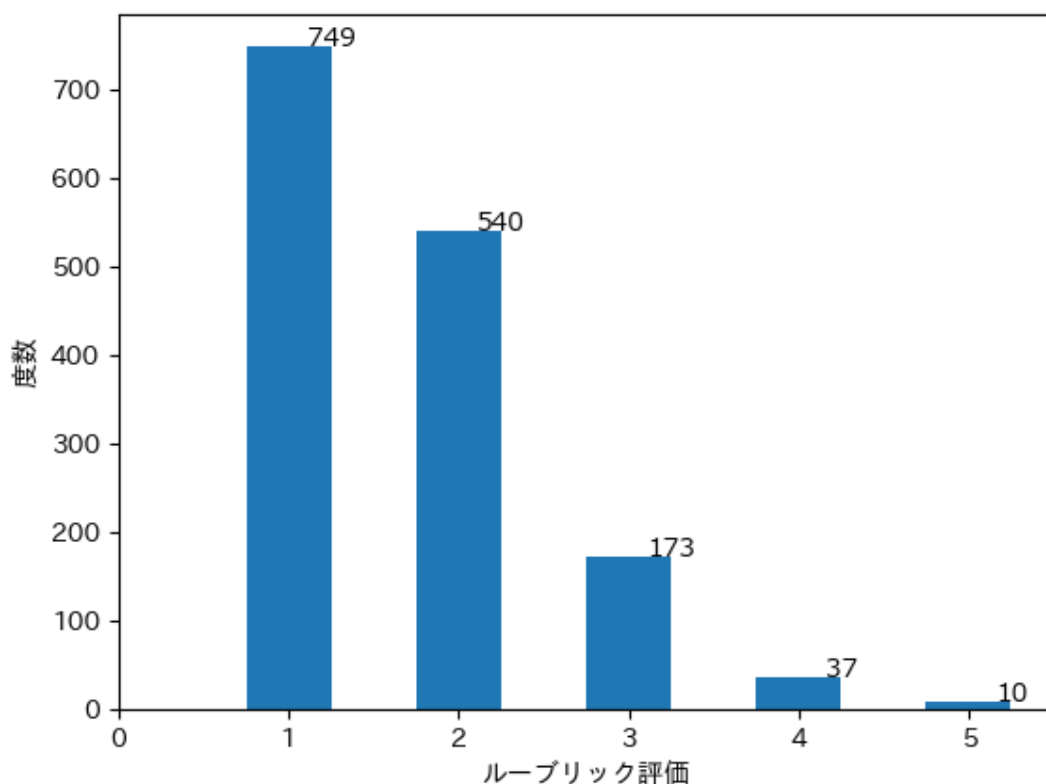


図5.9: 全体のループリック評価別頻度

5.6.1 アドバイス評価の結果：全体

図 5.11と表 5.8は、3 種類のアドバイス（テーマ+タスク、テーマ、タスク）について、適合性、理解性、実行性の評価を比較した結果を示す。なお、3 つの質問項目（適合性、理解性、実行性）間で回答数が異なるのは、質問項目への回答漏れのためである。

3 種類のアドバイスの評価は、3 種類ともに平均 5 以上あり、適合性、理解性、実行性について、学生の記述内容にやや沿っており、ややわかりやすく理解でき、やや改善を実行できる内容であると学生に思わせることができていたことが読み取れる。なお、理解性の評価は、「テーマ」だけを加えたほうが「テーマ+タスク」よりも学生が理解しやすいアドバイスを生成できていた可能性が示された。なぜなら、「テーマ」（平均 5.39, $SD0.86$ ）の平均値が「テーマ+タスク」（平均 5.2, $SD0.87$ ）よりも有意な差 ($p = 0.033 < 0.05$) で平均値が +0.19 高かったからである。

表 5.9は、有意な差が示された理解性の評価について、6 つの記述項目ごとの評価を比

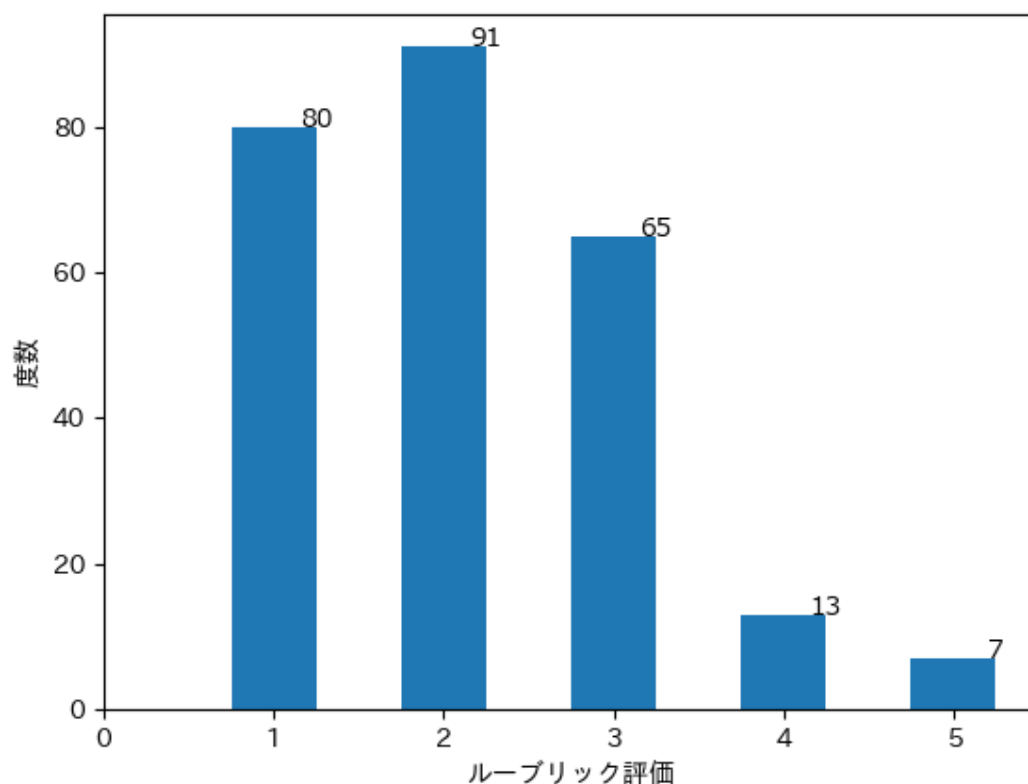


図5.10: 回答者のループリック評価別頻度

較した結果を示している．表 5.9は，特に目標計画についてのアドバイスの理解性評価は「テーマ」だけを加えたほうが「テーマ+タスク」よりも評価が高かったことが示されている．なぜなら，「テーマ」（平均 5.47, $SD0.75$ ）の平均値が「テーマ+タスク」（平均 4.96, $SD0.81$ ）よりも有意な差 ($p = 0.014 < 0.05$) で平均値が +0.51 高かったからである．

なお，表 5.8と表 5.9の比較の列（vs. テーマと vs. タスク）は，scipy モジュールのクラスカル・ウォリス検定を用いて，3 種類のアドバイス間に有意な差があった場合について，scikit_posthocs モジュールのスティール・ドゥワス検定を用いた多重比較の結果， $p < 0.05$ （信頼区間 95%）を示している．

5.6.2 アドバイス評価の結果：記述内容の具体性が低い場合

第5章の焦点は，学生の記述内容の具体度が低い場合でも，学生にとって具体的なアドバイスを自動生成できていたかを検証することである．そのために，学生の記述内容の具体度

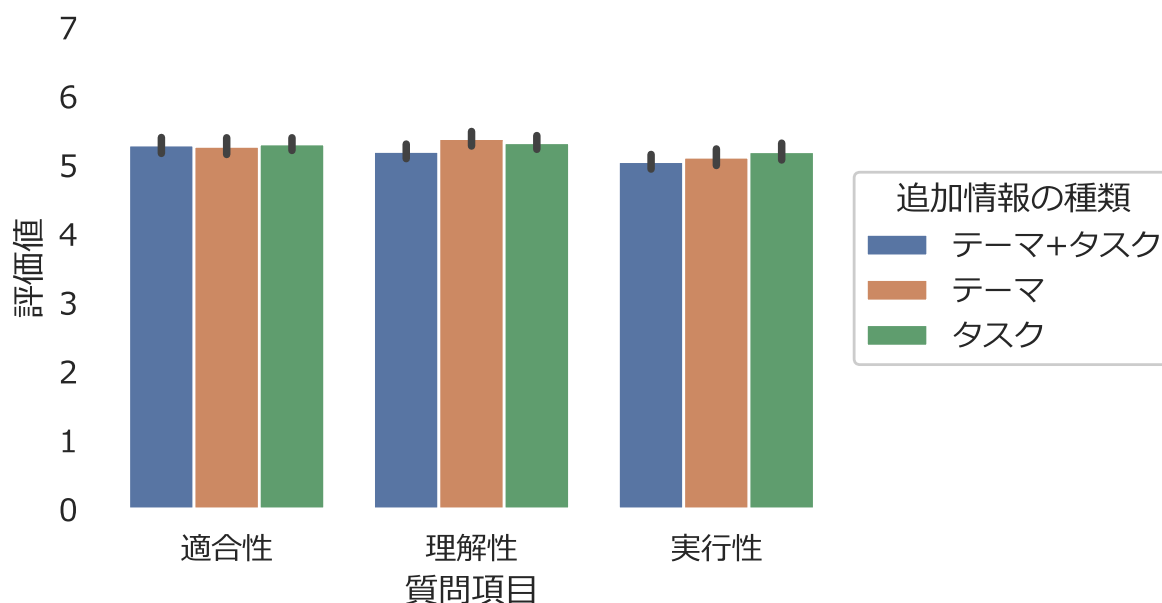


図5.11: 全体のアドバイス種類別平均値の比較

が3以下を抽出したときの、3種類のアドバイス（テーマ+タスク，テーマ，タスク）の適合性，理解性，実行性の評価について結果を示す。

図 5.12と表 5.10からは，学生の記述内容の具体性が低い場合でも，3種類のアドバイスの評価は，3種類ともに平均5以上あったことが読み取れる。

なお，具体性が低い場合でも，理解性については，「テーマ」だけを併用した場合，「テーマ+タスク」よりも学生が理解しやすいアドバイスを生成できていた可能性が示された．なぜなら，「テーマ」（平均 5.4, $SD0.84$ ）の平均値が「テーマ+タスク」（平均 5.16, $SD0.87$ ）よりも有意な差 ($p = 0.008 < 0.05$) で平均値が +0.24 高かったからである。

また，表 5.11は，有意な差が示された理解性の評価について，6つの記述項目ごとの評価を比較した結果を示している．表 5.11は，特に目標計画についてのアドバイスの理解性評価は「テーマ」だけを加えたほうが「テーマ+タスク」よりも評価が高かったことが示されている．なぜなら，「テーマ」（平均 5.47, $SD0.75$ ）の平均値が「テーマ+タスク」（平均 4.96, $SD0.81$ ）よりも有意な差 ($p = 0.014 < 0.05$) で平均値が +0.51 高かったからである。

5.6.3 アドバイス評価の結果：記述内容の具体性が高い場合

学生の記述内容の具体性が低い場合と比較するために，学生の記述内容の具体性が高い場合のアドバイス評価の結果を示す．そのために，学生の記述内容の具体度が4以上を抽出したときの，3種類のアドバイス（テーマ+タスク，テーマ，タスク）の適合性，理解性，実

表5.8: 全体のアドバイス種類別平均値と有意な差

質問項目	評価値			比較 ($p < .05$)	
	平均	SD	n	vs. テーマ	vs. タスク
適合性					
テーマ+タスク	5.29	0.91	256	p=0.9	p=0.9
テーマ	5.27	0.93	256		p=0.807
タスク	5.31	0.78	256	p=0.807	
理解性	平均	SD	n	vs. テーマ	vs. タスク
テーマ+タスク	5.2	0.87	251	p=0.033	p=0.217
テーマ	5.39	0.86	251		p=0.648
タスク	5.33	0.8	251	p=0.648	
実行性	平均	SD	n	vs. テーマ	vs. タスク
テーマ+タスク	5.05	0.87	250	p=0.701	p=0.23
テーマ	5.12	0.95	250		p=0.668
タスク	5.2	0.94	250	p=0.668	

行性の評価について結果を示す。

図 5.13と表 5.12は、学生の記述内容の具体性が高い場合においても、適合性、理解性、実行性の評価が平均 5 以上あったことを示している。3 種類のアドバイス（テーマ+タスク、テーマ、タスク）の評価に有意な差は見られなかった。ただし、具体性が高い場合のサンプル数が 20 件と少ないために記述内容の具体性が高い場合を検証するには十分な結果が得られていない可能性が高い。

学生記述の具体性の低い例に着目した場合

次の仮説を設定した、テーマとタスク情報を併用することで、学生記述の具体性が低い場合でも学生個別の具体的なアドバイスが自動生成できる。その場合、学生記述の具体性の高い場合と低い場合でアドバイスの学生評価に大きな差が出ないはずである。なぜなら、テーマとタスク情報を併用すると、学生記述の具体性が低い場合であっても、学生記述を補ってアドバイスを自動生成できるため、学生評価も低くならないと考えたためである。

この仮説を検証するために、学生記述の具体性の高い場合と低い場合によって自動生成したアドバイスの学生評価に大きな差がなかったかどうかを確認した。

具体的には、3 種類のアドバイスの学生評価項目（適合性、理解性、実行性）について、

表5.9: 記述項目別のアドバイス種類別平均値と有意な差

記述項目	質問項目	評価値			比較 ($p < .05$)	
目標計画	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	4.96	0.81	47	p=0.014	p=0.088
	テーマ	5.47	0.75	47		p=0.875
	タスク	5.36	0.86	47		p=0.875
焦点計画	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.36	1.07	47	p=0.667	p=0.332
	テーマ	5.15	1.08	47		p=0.863
	タスク	5.12	0.77	47	p=0.863	
方略計画	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.15	0.87	46	p=0.565	p=0.366
	テーマ	5.39	0.86	46		p=0.9
	タスク	5.41	0.8	46	p=0.9	
目標振り返り	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.33	0.8	40	p=0.242	p=0.817
	テーマ	5.63	0.81	40		p=0.507
	タスク	5.43	0.75	40	p=0.507	
焦点振り返り	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.08	0.72	37	p=0.118	p=0.503
	テーマ	5.46	0.8	37		p=0.707
	タスク	5.3	0.91	37	p=0.707	
方略振り返り	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.35	0.84	34	p=0.9	p=0.9
	テーマ	5.24	0.74	34		p=0.778
	タスク	5.35	0.65	34	p=0.778	

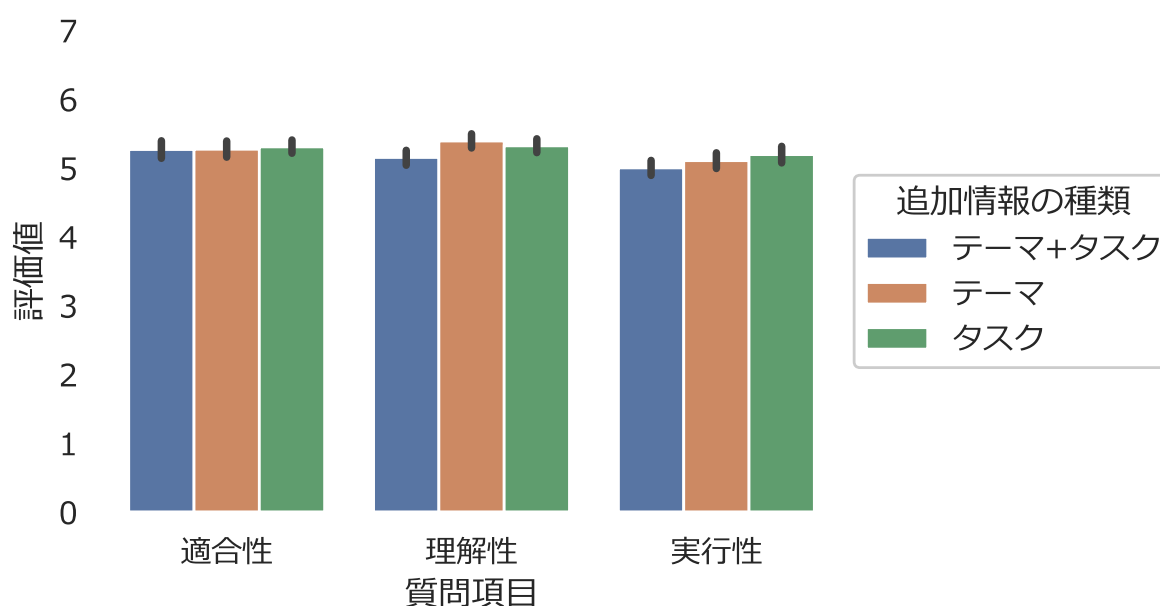


図5.12: 記述内容の具体性が低い場合のアドバイス種類別平均値の比較

表 5.12の具体性の高い場合の平均値と、表 5.11の具体性の低い場合の平均値を比較した。図 5.14と図 5.15の結果は、仮説に反して、学生記述の具体性が低い場合、テーマ+タスクを併用すると、実行性と理解性の評価が下がってしまうことを示唆している。図 5.15と図 5.15は、最も大きな差があった項目として着目したテーマ+タスクを併用したアドバイスの実行性と理解性である。

図 5.14は、テーマ+タスクを併用したアドバイスの実行性評価について記述内容具体性高低を比較した結果を示した。図 5.14からは、記述内容具体性が低くなると実行性の平均値が有意な差 ($p = 0.002 < 0.05$) で 0.6 下がっていたことが読み取れる。

図 5.15は、「テーマ+タスク」を併用したアドバイスの理解性評価について記述内容具体性高低を比較した結果を示した。図 5.15からは、記述内容具体性が低くなると理解性の平均値が有意な差は無かった ($p = 0.063 > 0.05$) ものの実行性と同じく 0.6 下がったことが読み取れる。

なお、「テーマとタスク」を併用した場合のその他の質問項目と「テーマ」単独、さらに、「タスク」単独の場合の差は、いずれも 0.6 未満であり、有意な差についても示されなかった。

表5.10: 記述内容の具体性が低い場合のアドバイス種類別平均値

質問項目	評価値			比較 ($p < .05$)	
	平均	SD	n	vs. テーマ	vs. タスク
適合性					
テーマ+タスク	5.27	0.92	236	p=0.9	p=0.9
テーマ	5.28	0.92	236		p=0.897
タスク	5.3	0.79	236	p=0.897	
理解性					
テーマ+タスク	5.16	0.87	231	p=0.008	p=0.12
テーマ	5.4	0.84	231		p=0.561
タスク	5.32	0.82	231	p=0.561	
実行性					
テーマ+タスク	5.0	0.85	230	p=0.455	p=0.16
テーマ	5.11	0.94	230		p=0.792
タスク	5.17	0.95	230	p=0.792	

5.6.4 自動生成されたアドバイスの事例

評価の高かった事例

具体度の低い学生の記述内容に対して、併用する情報「テーマ+タスク」で生成されたアドバイスについて、評価が高かった事例を示す。学生の記述内容について、テーマとタスク進捗情報を組み合わせた「テーマ+タスク」、テーマのみの「テーマ」、タスクのみの「タスク」という3種類のアドバイスを示している。「テーマ+タスク」、「テーマ」、「タスク」項目の右側に学生の評価値を示した。

.....
アドバイス生成事例 1

記述項目 焦点の計画

学生のテーマ 私は安全運転 AI を作成しておりこの AI の活用場所を教習所にしました。活用方法として運転初心者の方たちにしたら何が当たり前なのか、してはいけないことなのかがまだ知識不足です。そのためにシミュレーションの際に「スピードが出すぎじゃない?」や「今標識しっかり確認していた?」などの声掛けとともに運転後いくつかのポイントで点数を付け採点をしていき、卒業までに安全運転が自然にできるよ

表5.11: 記述内容の具体性が低い場合の記述項目別のアドバイス種類別平均値と有意な差

記述項目	質問項目	評価値			比較 ($p < .05$)	
目標計画	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	4.91	0.79	45	p=0.01	p=0.082
	テーマ	5.44	0.76	45		p=0.852
	タスク	5.33	0.9	45	p=0.852	
焦点計画	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.22	1.13	32	p=0.9	p=0.415
	テーマ	5.16	1.05	32		p=0.622
	タスク	5.0	0.84	32	p=0.622	
方略計画	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.13	0.87	45	p=0.561	p=0.358
	テーマ	5.38	0.86	45		p=0.9
	タスク	5.4	0.81	45	p=0.9	
目標振り返り	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.33	0.8	40	p=0.242	p=0.817
	テーマ	5.63	0.81	40		p=0.507
	タスク	5.43	0.75	40	p=0.507	
焦点振り返り	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.06	0.73	35	p=0.075	p=0.397
	テーマ	5.49	0.82	37		p=0.707
	タスク	5.31	0.93	37	p=0.707	
方略振り返り	理解性	平均	SD	n	vs. テーマ	vs. タスク
	テーマ+タスク	5.35	0.85	34	p=0.9	p=0.9
	テーマ	5.24	0.74	34		p=0.778
	タスク	5.35	0.65	34	p=0.778	

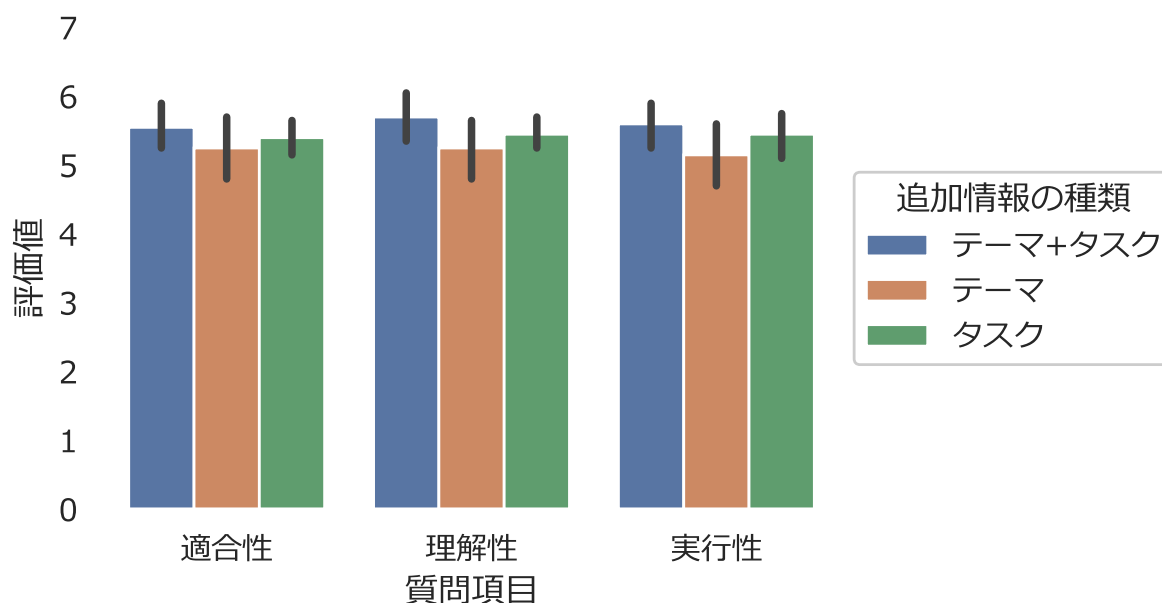


図5.13: 記述内容の具体性が高い場合のアドバイス種類別平均値の比較

うに作成を目指しています。

学生のタスク進捗 「実際にハンドルやアクセル，ブレーキを使う」の進捗状況 =0%，「事例作成」の進捗状況 =85%，「動作作成」の進捗状況 =60%

学生の記述内容 事例作成自体はかなり進んできているので次につながることを見つけて進めていきたいです。

具体度評価 2

テーマ+タスク 適合性 7，理解性 7，実行性 6

次につながることを見つける，っていうのは具体的に何を指してるのかな？例えば，「事例作成が進んできたから，次はそれを元に AI が適切な声掛けをする動作作成に取り組む」っていう具体的な内容にすると，焦点と成果の関連性が見えてくるよ！

テーマ 適合性 6，理解性 5，実行性 6

次につながることを見つける，っていうのはすごく大事だね。でも，それが具体的にどういうことなのか，そしてそれがどうやって安全運転 AI の完成につながるのかをもっと詳しく書いてみてはどうか？例えば，「次には，AI が適切な声掛けをするためのアルゴリズムを作成し，それが初心者の安全運転にどう寄与するかを考える」みたいな感じでさ。

タスク 適合性 7，理解性 6，実行性 6

次につながることを見つけるのはいいね！でも，それがどういう成果につながるのか

表5.12: 記述内容の具体性が高い場合のアドバイス種類別平均値

質問項目	評価値			比較 ($p < .05$)	
	平均	SD	n	vs. テーマ	vs. タスク
適合性					
テーマ+タスク	5.55	0.76	20	p=0.647	p=0.9
テーマ	5.25	1.12	20		p=0.838
タスク	5.4	0.6	20	p=0.838	
理解性					
テーマ+タスク	5.7	0.8	20	p=0.401	p=0.493
テーマ	5.25	1.07	20		p=0.9
タスク	5.45	0.51	20	p=0.9	
実行性					
テーマ+タスク	5.6	0.82	20	p=0.365	p=0.774
テーマ	5.15	1.09	20		p=0.679
タスク	5.45	0.76	20	p=0.679	

がもう少し具体的だといいな。例えば、「事例作成が85%進んでるから、次はそれを元に”動作作成”に取り組む。これにより、全体の進行をスムーズにする」みたいに書いてみてはどうか？

「テーマ+タスク」のアドバイスでは、学生のテーマ内容を「AIが適切な声掛けをする」と解釈していた。テーマの解釈とタスクにある「動作作成」が事例作成よりも進捗状況が低いことを用いて、学生のテーマと状況に合った具体的な改善例を提示できていたことが読み取れる。一方、「テーマ」のアドバイスの改善例では、学生のテーマ内容を解釈した改善例を提示できていた。しかし、学生が取り組んでいるタスク状況を考慮した内容にはなっていない。また、「タスク」のアドバイスの改善例では、学生のテーマとは関係なく個々のタスクの進捗を進めることに焦点していたことがわかる。

評価の低かった事例

具体度の低い学生の記述内容に対して、併用する情報「テーマ+タスク」で生成されたアドバイスについて、評価が低かった事例を示す。なお、記述項目が「振り返り」であるため、

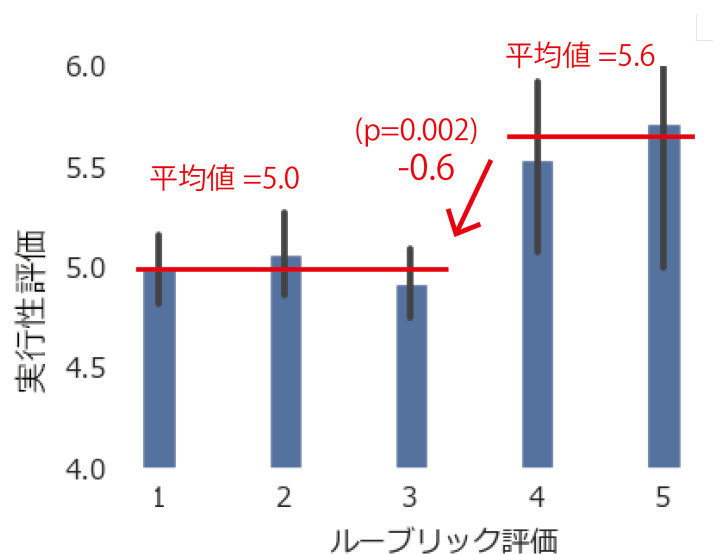


図5.14: 実行性評価について記述内容具体性高低の比較

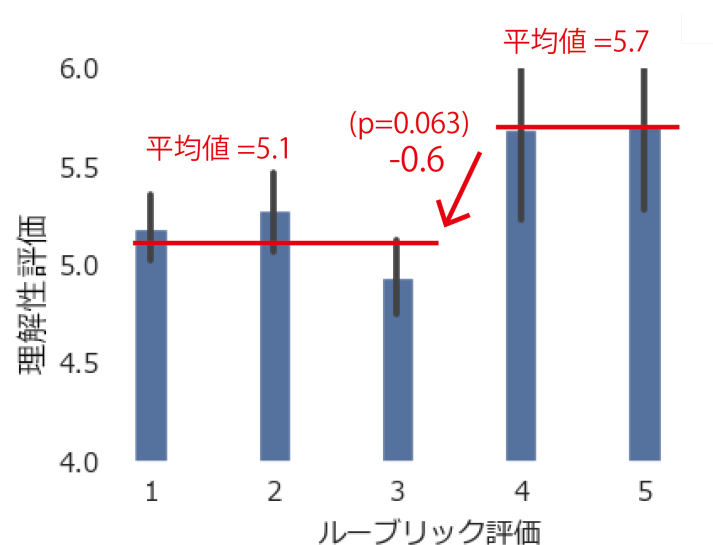


図5.15: 理解性評価について記述内容具体性高低の比較

「前回の記述内容」もアドバイス生成時のプロンプトに含まれている。

.....

アドバイス生成事例 2

記述項目 目標の振り返り

学生のテーマ タイトル：ネガティブ思考をポジティブに変換するサービスの提案取り組

む課題・嫌な事、悩み事などネガティブに考えてしまう事をポジティブに考えられるサービスを作る事。作るもの・ターゲットユーザーの嫌な事、悩み事に対し、ポジティブな言葉を返答するサービス出来てから効果を知る為に検証する事・サービスを使う前、使った後にアンケートを答えてもらい、サービスの課題を見つける

学生のタスク進捗 「アンケート作成」の進捗状況 =100%, 「研究概要」の進捗状況 =100%, 「ポジティブな言葉、ブラッシュアップ」の進捗状況 =90%, 「実証試験の準備」の進捗状況 =80%, 「チャットについて」の進捗状況 =90%, 「研究課題の設定」の進捗状況 =100%

前回の計画内容 アンケート作成, 実証試験の内容をまとめる

学生の記述内容 アンケート作成, 実証試験の内容をまとめる事

具体度評価 1

テーマ+タスク 適合性 4, 理解性 4, 実行性 4

ちょっと待って、君の振り返り、自己評価や進捗状況が見えないよ。例えば、「アンケート作成は計画通り 100% 達成できた。しかし、ポジティブな言葉のブラッシュアップは 90% で、まだ改善の余地がある」とか、具体的に書いてみてはどう？

テーマ 適合性 5, 理解性 5, 実行性 5

アンケート作成や実証試験の内容をまとめるという目標に対して、どの程度進んでいるのか、またそれが自分の予想とどう違うのかを書いてみよう。例えば、「アンケートは全 10 問作成したが、まだ実証試験の内容をまとめることができていない。予想よりもアンケート作成に時間がかかってしまった」というように具体的に書くといいね。

タスク 適合性 5, 理解性 5, 実行性 5

ちょっと待って、君の振り返り、具体的な進捗状況が見えないよ。例えば、「アンケート作成」は 100% 達成したんだよね？それなら、「アンケート作成は計画通りに進み、全ての項目を完成させた」とか書いてみてはどう？

.....

「テーマ+タスク」のアドバイスでは、タスク状況を用いた改善案を示していた。しかし、テーマを考慮した内容にはなっていない。一方、「テーマ」のアドバイスでは、テーマの内容ではなく、テーマに記述された方略やタスクに焦点を当てた改善案が示されていた。また、「タスク」のアドバイスでは、タスク進捗を用いて、その説明の記述方法について改善案が示されていた。

理解性の評価の低かった事例

具体度の低い学生の記述内容に対して、併用する情報「テーマ+タスク」で生成されたアドバイスについて、理解性の評価が低かった事例を示す。なお、記述項目が「振り返り」であるため、「前回の記述内容」もアドバイス生成時のプロンプトに含まれている。

.....
アドバイス生成事例 3

記述項目 目標の振り返り

学生のテーマ (アドバイスを生成事例 1 と同じのため表記を省略した)

学生のタスク進捗 「地図のアプリを見て良い道を見つける」の進捗状況 = 0%, 「実際にハンドルやアクセル, ブレーキを使う」の進捗状況 = 0%, 「事例作成」の進捗状況 = 85%, 「動作作成」の進捗状況 = 60%

前回の計画内容 もう少しで前期も終わり提出する期間に近づいているのでもう一度進み状況を計画していたものと照らし合わせて振り返っていきたいです。

学生の記述内容 少し計画通りに進んでいないところがあるので計画を立て直し再度進めていききたいです。

具体度評価 2

テーマ+タスク 適合性 6, 理解性 4, 実行性 5

ちょっと具体的にどの部分が計画通りに進んでないのか, どのタスクが遅れているのかを書いてみよう。例えば, 「事例作成」は 85% 進んでるけど, 「実際にハンドルやアクセル, ブレーキを使う」や「地図のアプリを見て良い道を見つける」はまだ 0% だよ。これが計画通りに進んでない理由かな? それとも他に理由があるのかな? もう少し具体的に書いてみて!

テーマ 適合性 4, 理解性 7, 実行性 5

ちょっと具体的な進捗状況が見えてこないね。例えば, 「AI の声掛け機能は完成したけど, 点数採点機能はまだ途中」みたいな感じで, どの部分が計画通りに進んでいて, どの部分が遅れているのかを書いてみよう。それによって, どの部分に焦点を当てて計画を立て直すべきかが明確になるよ。

タスク 適合性 6, 理解性 6, 実行性 5

ちょっと具体的にどの部分が計画通りに進んでないのか, どのタスクが遅れているのかを書いてみよう。例えば, 「ハンドルやアクセル, ブレーキを使う」タスクが 0% なので, ここが計画通りに進んでいないと書くとかさ。そうすると, どこに問題があるのかが明確になるよ。

「テーマ+タスク」のアドバイスでは、問題点の指摘のみで、改善案の提示がなかった。一方、「テーマ」のアドバイスでは、テーマの内容に基づいた改善案が示されていた。また、「タスク」のアドバイスでは、タスク進捗を用いて、タスク状況に基づいた改善案が示されていた。

5.7 考察

第5章は、従来手法 [17] の課題であった学生の記述内容が不十分な場合でも、学生の学習内容に沿った具体的なアドバイスを生成するために、学生の記述内容にテーマとタスク情報を併用した入力プロンプトを GPT-4 のゼロショット学習に用いてアドバイスの自動生成を行った。

表 5.10 に示した学生の記述内容のルーブリック評価は、3 以下の評価が 9 割以上であり、具体性が低い記述への対応の重要性が示されていると考えられる。さらに、図 5.11 と表 5.8 の比較から、実験結果は全体よりも比較的やる気のある学生の評価であったことを考慮すべきである。ただし、本実験におけるルーブリック評価の妥当性については、未検証である。そのため、ルーブリック評価の妥当性の検証については今後の課題とする。本研究の考察と結論は、ルーブリック評価に一定の妥当性があるものと仮定して述べる。なお、実際の具体性とルーブリック評価の差異による影響については、実際よりもルーブリック評価が高く評価された場合は、学生にとって難度の高いアドバイスとなるため、理解性や実効性の学生評価が低くなる影響が考えられる。その一方、実際よりもルーブリック評価が低く評価された場合は、学生にとって難度の低いアドバイスとなるため、理解性や実効性の学生評価が高くなる影響が考えられる。したがって、学生の評価についてはルーブリック評価の差異による若干の誤差が影響していた可能性がある。

学生の評価結果は以下のとおりである。自動生成したアドバイス（テーマとタスク情報を併用した場合）について、学生の評価を得た結果、具体度が低い学生の記述内容に対して以下の 3 つの評価項目について学生の評価結果が示された。

適合性 平均 5.27

理解性 平均 5.16

実行性 平均 5.00

したがって、提案手法による自動生成したアドバイスは、具体度が低い学生の記述内容に対して、自分が記述した内容にやや沿っていると思わせることができていた可能性があり、改善点についてもやや分かりやすい内容と思わせることができていた可能性がある。

ただし、理解性については、テーマ単独で用いると、テーマとタスク情報を併用した場合よりも学生に分かりやすい内容と思わせることができていた。なぜなら、理解性についてテーマ単独とテーマとタスク情報の併用を比較すると平均値が +0.24、有意な差 ($p = 0.008 < 0.05$) で高くなっていたためである。理解度についての詳細な考察は、第5.7.2節で述べる。

アンケートは授業最終回後の 2 週間にわたって実施され、リアルタイムのフィードバックではなかった。しかし、学生は自身の記述内容を参照しながらアドバイスを評価したため、リアルタイムでないことの影響は少なかったと考えられる。

実験結果については、以上の条件を前提とした考察を加える。

5.7.1 アドバイスの適合性についての考察

テーマとタスク情報を併用して自動生成したアドバイスの適合性について、学生の評価は、平均 5.27 であり、アドバイスの内容は、学生に自分の内容にやや沿っていると思わせることができていたと考えられる。

なお、学生の適合性の評価については、学生自身が記述した内容のみならず、テーマやタスクとの適合も含めた適合性が回答されていたと考えられる。なぜなら、学生は自身の記述内容が抽象的であっても、テーマやタスクの内容を含んだアドバイスに対して、適合性の評価を平均して「ややそう思う」と評価していたためである。したがって、適合性がややあると評価されていた「テーマ+タスク」のアドバイスは、学生が具体的に書くことができていなかったテーマやタスクの情報を用いたことで、学生に適合性を感じさせることのできるアドバイスが生成できていたと捉えることができる。

これらのテーマとタスクへの適合性の評価の事例については、第5.6.4節の評価の高かったアドバイスに示されている。学生の記述内容が抽象的であるにもかかわらず、「テーマ+タスク」のアドバイスには、学生のテーマとタスク状況の両方の内容を含むことが確認された。学生は、この「テーマ+タスク」のアドバイスの適合性評価について「非常にそう思う」と評価していた。

5.7.2 アドバイスの理解性についての考察

テーマとタスク情報を併用して自動生成したアドバイスの適合性について、学生の評価は、平均 5.16 であった。したがって、アドバイスの内容の改善点は、学生にややわかりやすいと思わせることができていたと考えられる。

なぜならば、具体的に何を目標に、何処を捉え、何をどうすべきかわからない学生は、自分のテーマに基づく計画や方略を漠然と考えいる。そこに、自動生成されたアドバイスで

は、個別のテーマに基づいた具体的な目標が提示されたため、学生は自分のテーマについて何をどうしたらいいのかについて分かったと感じさせることができたと考えられる。例えば、第5.6.4節の「テーマ」のアドバイスにあるように、あまり具体的ではない学生の記述に対して、テーマに基づいた具体的な方略についてアドバイスを提示していた。このとき「テーマ」のアドバイスは理解性を7と評価された。

しかし、理解性について、「テーマ」を単独で用いたアドバイスは、「テーマ+タスク」を用いたアドバイスのよりも有意な差で高く、特に目標計画へのアドバイスについて、有意な差で高かったことが示されている。「テーマ」単独のアドバイスの平均が「テーマ+タスク」よりも高くなった理由は、学生個別のテーマに関連する目標や焦点、方略を簡潔にアドバイスすることができていたためである。一方、「テーマ+タスク」を用いたアドバイスの理解性が有意な差で低かった理由は、テーマに基づく内容の判読性が下がったためと考えられる。なぜならば、アドバイスの種類がテーマとタスクの2つに増えたため、テーマに基づく内容の判読性が下がったためであると考えられる。

また、特に目標計画への「テーマ」を単独で用いたアドバイスについて、理解性の評価に有意な差で高かった理由は、学生が自身の目標計画を行うためのアドバイスにおいて、学生に理解されるために重要な内容が、学生のテーマに基づく内容であったためと考えられる。

あるいは、「テーマ+タスク」を用いたアドバイスについて、理解性の評価に有意な差で低かった理由は、テーマに基づく内容が読み取れなかったためと考えられる。

例えば、第5.6.4節の「テーマ+タスク」のアドバイスは、テーマに対するアドバイスと、タスクに対するアドバイスを分けて提示すべきところを、テーマについての内容が読み取れない。そのため、自分のテーマに沿った内容かどうかがわかりにくいと理解の評価が下がったと考えられる。このとき、「テーマ+タスク」のアドバイスでは理解性を4と評価されていた。

このように、「テーマ+タスク」を用いたアドバイスの判読性が下がった原因としてシステムメッセージに含まれた以下の2つの指示が影響していた可能性がある。

- 「目標の計画」に対して学生の記述内容とテーマとタスク進捗状況の中から関連が最も強い内容を用いた改善例を示してください
- アドバイスを200文字以内の文章で作文してください

以上の2つの指示の影響によって、200文字の制約の中で、テーマとタスクの両方を簡潔に理解できるようなアドバイスの生成できなかったと考えられる。したがって、今後の課題として、システムメッセージで、2つの指示を変更し、テーマに対するアドバイスと、タスクに対するアドバイスを分けて提示する指示を追加することで改善される可能性がある。

5.7.3 アドバイスの実行性について考察

テーマとタスク情報を併用して自動生成したアドバイスの実行性について、学生の評価は、平均 5.00 であった。そして、3 つのアドバイスについて平均値を比較したとき、有意な差は無かった。したがって、3 つのアドバイスに示された改善点は、学生にやや改善を実行できそうな内容であると感じさせることができていると考えられる。最低限、実行可能なアドバイスは生成ができていたと考えられる。

実行性について、有意な差が無かった理由として、以下の 2 つが挙げられる。

- 学生自身が実行可能な範囲で記述したテーマ、タスク、学生の記述内容に基づいていたため
- 具体性が低い場合のアドバイスは、アドバイスに提示された改善案がそのまま実行可能な内容であったため

例えば、第5.6.4節のアドバイスは、具体的な改善案が提示されており、改善が容易な内容であった。第5.6.4節の 3 つのアドバイスの実行性評価は、3 つとも 6 であった。

学生記述の具体性の低い例に着目した場合の考察

仮説としてテーマとタスク情報を併用することで、学生記述の具体性が低い場合でも学生個別の具体的なアドバイスが自動生成できた場合、学生記述の具体性の高い場合と低い場合でアドバイスの学生評価に大きな差が出ないと考えていた。ところが検証した結果は、仮説に反して、実行性について有意な差で下がっていた。さらに、有意差はなかったものの、理解性についても下がっていたことが示された。つまり、テーマとタスクを併用した場合、学生記述の具体性の低くなると、アドバイスの内容がうまく理解できず、実行性が下がったと考えられる。

この要因については、アドバイスの種類がテーマとタスクの 2 つに増えたため、内容の判読性が下がったためであると考えられる。この要因は、第5.7.2節の考察と共通している。

5.7.4 学生評価の低かったアドバイスについての考察

第5.6.4節の学生評価の低かったアドバイスの事例について、適合性の評価が低かった理由を考察する。

第5.6.4節の「テーマ+タスク」のアドバイスについて、学生の記述内容には「アンケートの作成と実証試験の内容」を実施していたとある。しかし、提示されたアドバイスには、「ポジティブな言葉のブラッシュアップ」について振り返りをするよう提案されていた。そのため、学生は自分が行っていないことを振り返りするように提示されたために適合性、実

行性を低く評価したと考えられる。

なお、第5.6.4節の「テーマ」のアドバイスについては、テーマというよりも方法やタスクに焦点を当てた改善案であった。その理由は、学生が記述していたテーマ内容に、テーマを進めていくための方法やタスクについての記述が比較的多く含まれていたためであると考えられる。このことから、今後の課題として、テーマやタスクの記述について、事前にある程度方針を定めて、十分な指導や支援策を検討する必要があると考えられる。

第5.6.4節の事例では、具体的な改善例の提示が無かった事例も見られた。第5.6.4節の「テーマ+タスク」のアドバイスは、単に学習活動が進まなかった要因を追求するような内容になってしまっていた。そのため、具体的な改善例の提示が無く、改善点がわかりにくいと評価されていた場合があったことがわかった。その原因は、付録 E.2に示したプロンプトにある 200 文字に制限した指示の影響下で、学生の記述内容の問題点を指摘する指示が強く効き、改善案を生成する指示が効かなかったためであると考えられる。このことから、今後の課題として、アドバイスの文字数制限の見直しや、記述内容の問題点を指摘する指示の方法を再検討する必要がある。また、改善案のわかりやすさについて、学生が求める質と教員が求める質が異なるため、教員による評価についても検討する。さらに、評価を回答した学生について、記述内容の具体性の分布とその影響については検証できていない。そのため、記述内容の具体性評価の分布と評価結果への影響について分析と検討を加える必要がある。

5.8 結言

本章の目的は、抽象度が高い学生の学習計画と振り返りの記述内容に対しても、わかりやすい具体的な改善アドバイスを自動生成することである。

第5章が提案する学生個別のテーマと学生のタスク進捗情報を併用して自動生成したアドバイスは、学習計画や振り返りの記述内容が不十分な学生に対して、学生個別のテーマと学生のタスク進捗情報にやや沿っており、改善点がややわかりやすく、改善案をやや実行しやすいと学生に思わせることができていた。

しかし、テーマを単独で用いて自動生成したアドバイスの方が、テーマとタスクを併用するよりも学生の評価は高くなることが分かった。さらに、仮説に反して、テーマとタスク情報を併用すると学生記述の具体性が低い場合に、学生がアドバイスを理解し難くなり、実行性の低下につながることがわかった。加えて、学生の評価が低いアドバイスの内容を抽出して確認したところ、アドバイスがうまく自動生成できていない事例があったことが確認された。なお、これらの結果は、全体の傾向よりもややモチベーションの高い傾向にある学生の評価である。

本研究の提案手法は評価項目として設定した3つについて以下の結論とした。

適合性 学生にとって、やや自分が記述した内容に沿っていると思わせることができる。

理解性 学生にとって、やや改善点について理解ができる内容であると思わせることができる。

実行性 学生にとって、やや改善を実行することができる内容であると思わせることができる。

ただし、理解性については、テーマ単独で用いたアドバイスがテーマとタスクを併用したアドバイスよりも高く評価された。なぜなら、学生が自身のテーマについて学習行動を具体化できていない場合に、テーマに基づく学習方略を改善案として具体的に提示できていたためである。一方、テーマとタスク情報を併用するとテーマ単独よりも理解性や実行性が低くなるのは、テーマとタスクをうまく切り分けたアドバイスができていなかったためである。

また、テーマとタスク進捗状況が偏ってアドバイスされる事例も見られた。その原因は、システムメッセージの設計と、学生のテーマの記述内容にある可能性があった。

また、学生のテーマの記述内容と関連して、学生のタスクの記述方法についても似たような影響を与えていた可能性が考えられる。

このことから、今後の課題として、テーマに対するアドバイスと、タスクに対するアドバイスを分けて提示する指示を追加するシステムメッセージの改良と、学生のテーマやタスクの支援方法についても検討する余地があることがわかった。

本章の提案手法は、学生に対して、やや受け入れられる改善アドバイスを生成できる可能性が示されたが、改善の余地があることがわかった。したがって、本章は、表 2.1 に示した学習支援システム研究の項目 6 において、学生個別の記述内容に追従した助言自動生成の実現に対して一定の貢献をすることはできた。

しかし、実用するためには検証が不十分な点も見られた。アドバイス自動生成手法の有効性を検証するためには、学生が実際に記述内容を改善したかどうか、およびその改善が学習方略の利用や調整にどのような影響を与えたかについての実証的なデータが必要である。また、学生のアンケート評価だけでは、学習の効果や深度が適切に測れていない可能性があるため、教員によるアドバイス内容の評価も検討する。したがって、今後の研究では、実際の学習環境におけるアドバイスの実施とその効果に関する実証的な評価を行うことが求められる。

なお、実用するためには、教員による人手の確認と補完を併用する必要がある。なぜなら、学生が、学習以外の就職活動や教職実習への不安や不調、PC の不具合を関連付けて記述した事例もあった。提案手法では、学習活動に関連しない事柄にも対応し、励ましの言葉を含めつつ、学習のテーマやタスクの進行を促すアドバイスを組み込む事例が確認された。これらの学生からの学習活動躓きのサインを捉え、適切なサポートが受けられるように導くには、教員による人手の支援や別の仕組みが必要になると考えている。

第 6 章

本研究の包括的な考察

本章では、本研究の目的について包括的な考察を行い、表 2.1 に示した本研究の学習支援システム研究における貢献について言及する。

デザイン教育などに見られる学生個別に答えや遂行過程の定まらない学生個別の学習活動を支援するとき、何を目標に、何に着目し、どのような方法で取り組むのかという学生個別の学習実態に基づいた自動支援手法を確立することである。そのため、本研究は、学生個別の学習状況や学習内容の学習実態を直接扱うことができる定性的かつ学生個別の自動助言の実現を目的とした。

本研究は以下 3 つの手法を提案した。

- 項目 1 課題遂行段階に対する学習スキルの助言（第3章）
- 項目 5 学生の記述内容から学習スキルの自動評価（第4章）
- 項目 6 学生個別の記述内容に追従した助言自動生成（第5章）

第6章では、本研究で提案された 3 つの定性的かつ学生個別の自動助言手法が、自己調整学習の何にどこまでどのように対応したと言えるのかについて考察し、そして、表 2.1 に示した学習支援システムの研究にどのような貢献をしたと言えるのかについて考察を加えた。

図 6.1 は、本研究の提案手法と自己調整学習の循環モデル [3] の関係を示している。図 6.1 の ● 印は、本研究の提案手法が対応を目指した項目である。● 印の項目では、学生個別に異なる学習状態や学習内容を捉える定性的かつ学生個別のアプローチが重要になる。しかし、定量的かつ定形的なアプローチによる自動支援だけでは、学生が個別に取り組む学習の状況や内容を直接に扱うような自動支援は困難である。

以下に、図 6.1 について、定性的なアプローチが重要かつ提案手法を適用した ● 印項目についての考察、そして、今後の課題とする ○ 印項目について考察を述べる。

第3章では、図 6.1 の遂行コントロール（課題遂行段階）に対応する支援手法を提案した。課題遂行段階における学習スキルの自動助言を実現するために、定形文と決定木を用いた自

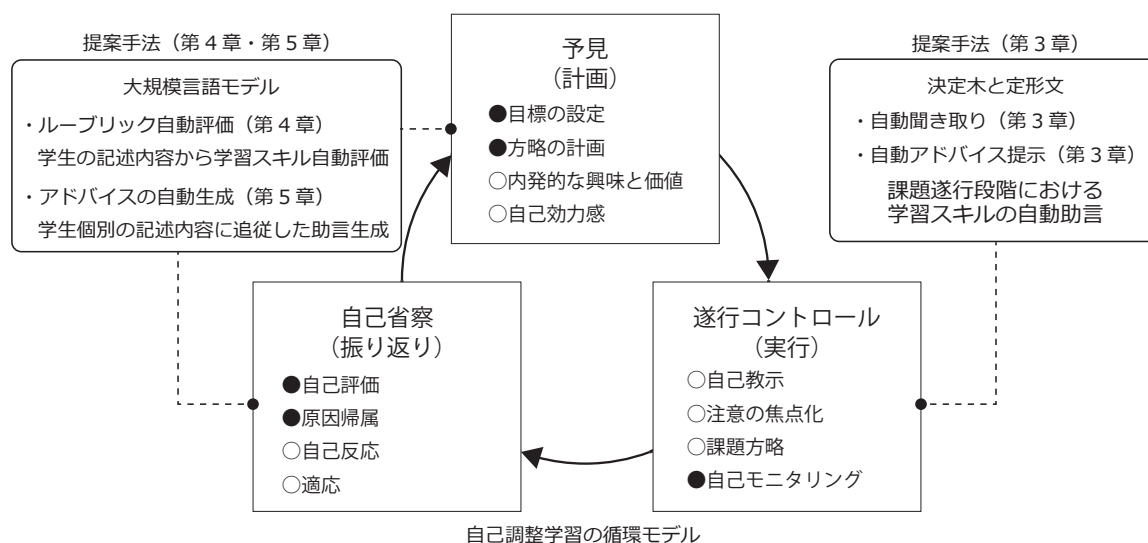


図6.1: 自己調整学習の循環モデル [3]

動的な聞き取りとアドバイス提示の手法を採用した。なぜならば、遂行コントロールは学習活動中のプロセスであり、学生の負担を軽減することが重要である。そこで、第4章と第5章で採用された詳細な記述活動を学生に要求することなく、自動的な学習状況の聞き取りだけでアドバイス提示を可能とする手法が求められたためである。

実験結果は、システムが自動的に学生の目標達成の見込みを聞き取った時に、学生が「どちらかといえばできそうでない」と回答したケースに焦点を当てたところ、学生自身による課題プロセスの経過把握が増加することが明らかになった。

具体的には、自動的な聞き取りによって、目標達成の見込みに「どちらかといえばできそうでない」と学生が回答した場合に対して提示された目標の具体化と段階化のアドバイスが、課題プロセスの把握を促したと考えられる。

つまり、課題プロセスの経過把握が増加したのは、学生が現在の目標達成の見込みについて回答する際に、自己モニタリングを通じて自身の学習状況を再確認したためであると考えられる。

さらに、このときに提示された目標の具体化と段階化アドバイスにより、学生は次の実行可能な段階を考える際に、自身の学習状況と課題プロセスの経過を把握し、それに基づいて自己評価を行うことができたと考えられる。

以上のことから、図 6.1の遂行コントロールに適用した学習スキルの自動助言として、一定の条件下でのみ自己モニタリングの一部が促進された可能性が示されたと考えられる。

一方、遂行コントロールの自己教示、注意の焦点化、課題方略の支援は不十分であった。これは、教員が編集した決定木と定形文による聞き取りとアドバイスが、実際の学生の学習状況と学習方略との間に乖離があったためであると考えられる。

したがって、第3章は、表 2.1に示した学習支援システム研究の項目 1：課題遂行段階における学習スキル助言の実現についての貢献は、一定条件下において限定的な可能性を示したが、本研究が目的とする定性的なアプローチによる自動支援としては、不十分であったと考えられる。

このように、第3章の自動的な聞き取りとアドバイスの提示手法について改善の余地が多くあった。そこで、今後の課題として、自動聞き取りとアドバイスに大規模言語モデルを用いる手法を検討することとした。

本研究の第4章では、図 6.1に示される予見と自己省察に対する自動評価手法を提案している。学生個別の記述内容から学習スキルの自動評価を実現するために、GPT-3 を用いた学習計画と振り返りの自動評価を用いて、予見と自己省察における学生の記述内容から学習スキルの自動評価を行い性能を評価した。

第4章の自動評価は、第5章で提案するアドバイス自動生成手法の基盤となるものであり、学生の学習計画と振り返りの記述内容における学習スキルの段階を評価しておくことで、より適切なアドバイスを生成するための前提条件を整える役割を果たしている。このように、第4章で提案された自動評価手法は、学生の自己調整学習を支援するためのアドバイス生成において重要な機能を担っており、第5章のアドバイス生成手法と密接に連携している。

第4章の自動評価手法は、学習の計画（予見）と振り返り（自己省察）に関する記述内容を、ファインチューニングした大規模言語モデルを用いて自動評価を行う。

具体的には、学習計画の記述内容と、学習の振り返りの記述内容について、あらかじめ定めたループリックに基づいた自動評価を行う。

ループリックに基づいた自動評価のためには、学生の自由記述のテキストデータとループリックの内容の関連を自動的に評価する必要がある。そのため、教員が人手で評価した結果と対応する学生の記述内容を用いて、大規模言語モデル GPT-3 をファインチューニングした。

実験の結果、ループリックに基づいて学生の記述内容を自動評価する手法は、十分に実用可能な性能を示した。その理由は、学生の記述内容に対する教員のループリック手動評価をファインチューニングしたことで、学生の記述内容とループリック評価との関係性の特徴を備えた回帰分析ができるモデルが得られた結果であると考えられる。

したがって、第4章は、表 2.1に示した学習支援システム研究の項目 5 において、学生の記述内容から学習スキルの自動評価の実現について一定の貢献ができたと考えられる。従来

手法では、学生の記述内容を分類する手法 [15] やルーブリックに基づくレポートの自動採点手法 [30] は検証されてきたが、学習計画や振り返りの自由記述から学習スキル具体性を自動評価する手法については検証されていなかった。

ただし、トレーニングデータの偏りが評価精度に影響を与える可能性があるため、この点の改善が必要である。また、トレーニングデータに含まれない記述内容の評価には不確実性が残ることから、この問題にも対処する必要がある。そのために、今後の課題として、ファインチューニング不要のゼロショット学習を活用し、ルーブリックに基づく自動評価手法の開発を進めることとした。

本研究の第5章では、図 6.1に示される予見と自己省察のプロセスにアドバイスを自動生成する手法を提案している。学生に提示するアドバイスは、予見と自己省察の学習方略について、学習方略の利用や調整の具体化を促すアドバイスを提示することが重要である。特に、学習状況の把握や方略の計画について具体的に記述することができていない学生に受け入れられやすい具体的なアドバイスになっていることが重要である。

従来手法 [17] では、教員の手動作文と同程度にルーブリックに沿ったアドバイスの雛形文に意味的に近いアドバイスが生成できていた可能性が示された。しかし、学生の記述内容が不十分な場合に、具体的なアドバイスを自動生成することができていないことがわかった。そこで、学生の記述内容が不十分な場合でも、学習方略の利用や調整を具体化を促すアドバイスを自動生成するために、GPT-4 のゼロショット学習用に入力プロンプトを設計した。具体的には、ルーブリックの自動評価とルーブリックごとのアドバイス方針を含み、学生個別の記述内容と学生個別のテーマとガントチャートのタスク情報を併用してアドバイスを生成するシステムメッセージとユーザーメッセージを設計した。

提案手法の評価は、自動生成したアドバイスが学生に受け入れられるかどうかの確認を行った。

実験の結果、学生個別のテーマ、ガントチャートのタスク情報を併用して自動生成されたアドバイスは、学生に自身の取り組む内容にやや沿っており、改善点がややわかりやすく、改善案をやや実行しやすいと学生に思わせることができていたと考えられる。その理由として、以下の2つの点が考えられる。

1. ユーザーメッセージで与えた学生の個別テーマやタスク進捗状況に基づいて学生の記述内容を補完するアドバイスができていた
2. プロンプトのシステムメッセージで与えたルーブリックとアドバイス方針、そして自動評価結果に基づいて、学生の学習状況の段階に応じたアドバイスが自動生成できていた

ただし、理解性については、テーマを単独で用いた場合は、テーマとタスクを併用する場合よりも学生の高い評価が得られていた。特に記述項目別のアドバイスを比較した場合は、目標計画へのアドバイスについてのみ、テーマを単独で用いた場合に高い評価が得られていた。このことから、以下の3つことが挙げられる。

1. 目標計画へのアドバイスは、テーマを単独で用いた方が学生に理解し易いアドバイスを生成できていた可能性がある。
2. テーマとタスクを併用した場合に理解性が低かったのは、アドバイスの種類がテーマとタスクの2つに増えたために、テーマに関する内容の判読性が下がったためと考えられる。
3. 入力プロンプトにテーマに対するアドバイスと、タスクに対するアドバイスを分けて提示する指示を追加することで改善される可能性がある。

なお、第5章のアドバイスを自動生成する提案手法は、全体的な学生の評価については、平均 5.0 以上であった。このことから、学生の学習計画と振り返りの記述内容に対して、自動生成されたアドバイスは、学生に受け入れられて学生の記述内容の改善にやや寄与するが、まだ改善の余地があると考えられる。

したがって、第5章は、表 2.1 に示した学習支援システム研究の項目 6 において、学生個別の記述内容に追従した助言自動生成の実現については一定の可能性を示すことができたと考えられる。なぜならば、これまでの学習支援システムでは、困難であった学生の記述内容に追従しつつ学習方略を改善するアドバイスを自動生成することができ、生成されたアドバイスは学生から一定の評価を得ることができたためである。

第4章の学生の記述内容から学習スキルの自動評価と第5章の学生個別の記述内容に追従した助言自動生成によって、自己調整学習の予見と自己省察の自動支援手法について、定性的なアプローチについて一定の可能性を示すことができた。

自己調整学習の自動支援については、図 6.1 の予見では、目標設定、方略の計画に対応した自動支援手法について可能性を示したと考えられる。なぜならば、自動生成されたアドバイスは、学生の記述した目標の計画と方略の計画について、学生の学習状況の段階に応じた方略の利用や調整の記述を促す可能性を示したためである。また、図 6.1 の自己省察についても、自己評価、原因帰属に対応した自動支援手法の可能性を示したと考えられる。なぜならば、自動生成されたアドバイスは、学生の記述した目標の振り返りと焦点の振り返り、そして、方略の振り返りについて、学生の学習状況の段階に応じた計画した方略の利用や工夫の評価と、問題点を方略の利用や工夫に帰属させる記述を促す可能性を示したためである。

ただし、第5章に示した、項目 6 の学生個別の記述内容に追従した助言自動生成の実現に

については、アドバイス自動生成手法の入力プロンプトについて改善の余地がある。また、学生が実際に記述内容の改善を実行するかどうか、さらには、実際に改善を実行したときに学習方略の利用や調整が増加するのかどうかまでは明らかにしていない。アドバイス自動生成手法の有効性を検証するためには、学生が実際に記述内容を改善したかどうか、およびその改善が学習方略の利用や調整にどのような影響を与えたかについての実証的なデータが必要である。したがって、今後の研究では、実際の学習環境におけるアドバイスの実施とその効果に関する実証的な評価を行うことが求められる。

その他、第5章のアドバイスの自動生成の手法は、学習状況と関係性の少ない記述内容を関連性のある内容と見なしてアドバイスを生成してしまう事例も見られている。具体的には、学生の記述内容に対してテーマやタスク進捗状況の内容が偏ってアドバイスされる事例が見られた。

そのために、今後の課題として、以下の4つの点について検討を進める。

1. 入力プロンプトにテーマに対するアドバイスと、タスクに対するアドバイスを分けて提示する方法
2. 学習状況と関係性の少ない記述内容への対応方法
3. 学生が実際に記述内容の改善を実行するかどうかの検証
4. アドバイスに従って改善を実行したときに学習方略の利用や調整が増加するかどうかの検証

以上、本研究の提案手法が図 6.1の●印の項目の支援にどこまで対応したのかについてを検討を加えた。

一方、図 6.1の○印に示した、遂行コントロールの自己教示、課題方略、注意の焦点化、自己省察の自己反応、適応、予見の内発的な興味と価値、自己効力感への支援方法については、本研究では明らかにできていない。そのため、今後の課題として、これらの対応できていない項目について、学生の記述内容や聞き取り結果を自動評価をする手法と評価結果に基づいたアドバイスを生成する手法について以下に検討を行う。

予見の内発的な興味と価値については、第5章で扱った学生個別のテーマに対する興味と価値を問い、回答文をあらかじめ定めた専用のループリックで評価した後に、必要に応じてアドバイスを自動生成して提示する支援手法を検討する。

自己効力感については、第3章で得られた目標設定と自己モニタリング、そして、振り返りによる課題遂行プロセス把握の関係という新しい仮設に基づき、第4章で扱った躓き度の自動評価と、学習計画の記述内容を用いて、必要に応じてアドバイスを自動生成して提示する支援手法を検討する。

遂行コントロールの注意の焦点化と課題方略，自己教示については，焦点計画の記述内容について，課題遂行途中にて自動的な聞き取りとアドバイスを自動生成して対話する支援手法を検討する。

自己省察の自己反応，適応については，第4章で扱った躓き度の自動評価と，学習振り返りの記述内容を用いて，必要に応じてアドバイスを自動生成して提示する支援手法を検討する。

以上が今後の課題となる図 6.1 の ○ 印の項目についての検討である。

次に，本研究が提案した支援手法には以下の限界があると考えられる。したがって，提案手法を実用する場合には，学生に提示する事前に教員の確認と，必要に応じて教員が編集する機構を要すると考えられる。

学生のこれまでの学習経緯や教員の指導方針などを考慮していない

振り返りの記述項目では，直前の計画の記述は入力プロンプトに含むものの，それ以前の学習経緯は含まない。そのため，これまでの学習経緯や教員の指導方針などを考慮しないアドバイスが自動生成されることがある。今後の課題として，入力プロンプトに，これまでの経緯や教員の指導内容を要約して取り入れることを検討する。

学生の記述項目の間違いを判断できない

学生が目標の計画に，誤って焦点の計画内容を記述するような記述間違いがあっても，記述内容を目標の計画としてアドバイスを生成する。今後の課題として，システムメッセージに，記述項目の間違いを判定する手続きとその結果に基づく対応を含めることを検討する。

学生の学習と直接的に関係のない内容をこじつけてしまう

体調不良や就職活動など，学習内容と直接的に関係のない内容を，学習内容と関係付けてアドバイスを自動生成する。心身の不調や就職活動の悩みなどデリケートな内容には対応できない。今後の課題として，システムメッセージに，学習との関係性を評価する手続きと評価に応じた対応を含めることを検討する。

学生の嘘や取り繕った記述に気づけない

学生の記述内容や学生が管理するガントチャートの情報に基づいてアドバイスを自動生成している。そのため，記述内容やガントチャートの情報が実際と異なっても気づくことができない。今後の課題として，ユーザーメッセージに経緯情報と実績データを含めることで嘘や取り繕いを評価する方法を検討する。

第 7 章

結論

7.1 本研究のまとめ

本研究の目的は、デザイン教育などに見られる学生個別に答えや遂行過程の定まらない学生個別の学習活動を支援するとき、何を目標に、何に着目し、どんな方法で取り組むのかという学生個別の学習実態に基づいた自動支援手法を確立することである。

そのために、本研究は、表 2.1 に示した以下 3 つの定性的かつ学生個別なアプローチによる自動助言手法を提案した。

項目 1 課題遂行段階に対する学習スキルの助言（第3章）

項目 5 学生の記述内容から学習スキルの自動評価（第4章）

項目 6 学生個別の記述内容に追従した助言自動生成（第5章）

本研究は、何を目標に、何に着目し、どんな方法の学習なのかという実態を捉えた自動支援の実現のために、定性的かつ学生個別なアプローチによる自動助言手法を明らかにするものである。

そして、デザイン教育のような問題解決のために様々な手法を試行錯誤する、答えや解法が定まっていない探索型の学習活動を対象として、自己調整学習の循環プロセスに基づいて自動的に支援する方法を確立することを目指した。

本論文は著者が行ってきた研究をまとめたものである。以下に本研究のまとめとして第3章、第4章、第5章の成果を要約し、最後に今後の課題と展望を述べる。

学生が主体的に学習活動を遂行できるよう支援するためには、自己調整学習の循環プロセスに基づいた予見、遂行、自己省察の各プロセスにおいて、目標設定、方略の計画、内発的な興味と価値による動機づけ、自己効力感、自己モニタリング、自己教示、注意の焦点化、課題方略、自己評価、原因帰属、自己反応、適用の調整を促進する支援が必要である。これらの調整を支援するには、定量的なアプローチによる学習傾向を捉えるだけでなく、調整すべき学習内容を捉える必要があった。

そこで、本研究では以下の自己調整学習の循環プロセスに基づく自動支援手法について、定性的かつ学生個別のアプローチを実現するための手法を提案した。以下の遂行コントロール、自己省察、予見の3つは、特に定性的かつ学生個別の支援が重要となる。

遂行コントロール

課題遂行段階における自動助言による、学習状況に基づいた方略の利用と調整支援（第3章）

自己省察と予見

学生の記述内容から学習スキルの自動評価（第4章）と学生個別の記述内容に追従した助言自動生成（第5章）による、学習振り返り記述内容の方略の利用と調整支援

本研究が提案した手法は、以下の自己調整学習の各項目に対する自動的な支援を目指したものである。

遂行コントロールの自動的な支援

自己モニタリング（第3章）

自己省察の自動的な支援

自己評価，原因帰属（第4章，第5章）

予見の自動的な支援

目標の設定，方略の計画（第4章，第5章）

第3章では、デザイン教育における学習活動中の自己モニタリングと、適切な学習方略の利用と調整を促進するために、表 2.1の項目 1：課題遂行段階に対する学習スキルを自動助言する手法を提案した。具体的には、決定木と定形文を組み合わせた自動対話システムを用いて、学生の学習状況を聞き取り、その結果に基づいて適切なアドバイスを自動的に提示する手法である。この手法が学生の自己調整学習に与える影響についての評価結果を報告した。

第3章の実験結果は、学生が目標達成の見込みについて「どちらかといえばできそうでない」と回答した場合に限り、課題プロセスの経過を把握する学習方略を使用したと自己申告の割合が増加したことがわかった。つまり、学生は問われた現在の目標達成の見込みについて回答を検討するために、自身の学習状況を再確認する自己モニタリングがあったと考えられる。これは、自動対話システムが学習活動中の遂行コントロールの一部を限定的な条件においてのみ支援していた可能性を示唆していた。

一方、他の多くの学習方略に対する効果は明らかにすることができなかった。理由としては、事前に設定された決定木と定形文が実際の学生の学習状況と乖離していたため、学習状況に応じた具体的な助言を提供できていなかったことが原因であると考えられる。

この結果を踏まえ、新たな仮説を探索した結果、自己効力感と目標設定、方略計画、自己モニタリングの関係性に焦点を当てることの重要性を確認した。今後は、この関係性をさらに探求し、学習方略の利用と調整を促進するための新しいアプローチを開発することが求められた。

第3章の結果から、あらかじめ定めた決定木と定形文を用いた自動的な聞き取りとアドバイス提示の限界を認識した。第3章での自動的な聞き取りとアドバイスの提示による支援方法の評価を踏まえ、第4章では、より詳細な学習状況の分析と個別化されたアドバイスの提供を目指すために、大規模言語モデルを用いた自動評価へと研究の焦点を移行した。

以上のことから、第3章にて提案した図 6.1の項目 1：課題遂行段階に対する学習スキルの自動助言は、本研究の定性的かつ学生個別のアプローチとしては極めて限定的な効果であったと結論した。

そこで、第4章では、図 6.1の項目 5：学生の記述内容から学習スキルの自動評価の手法を提案し、5章では、図 6.1の項目 6：学生個別の記述内容に追従した助言自動生成する手法を提案した。

まず、第4章では、図 6.1の項目 5：学生の記述内容から学習スキルの自動評価の手法を提案した。学生の学習計画と振り返りの記述内容における学習スキルの段階を自動的に評価するために、ループリックに基づいたデータを使用してファインチューニングされた大規模言語モデルを生成し、その自動評価モデルによる自動評価の性能を検証した。検証では、どの程度ループリックに沿った評価ができたのか、また、その結果が教員の手動評価とどのように比較されるのかについての評価結果を報告した。

第4章の結果は、具体性と躓き度に関して、教員の手動評価に匹敵する性能を持つ自動評価が可能であり、その評価結果を用いて自動的に足場をはずす設定ができることを示した。

この評価結果は、学生の記述内容に対する教員のループリック手動評価をファインチューニングしたことで、学生の記述内容とループリック評価結果との関係性の特徴を備えた回帰分析ができるモデルが得られたためであると考えられる。

以上のことから、第4章は、表 2.1に示した学習支援システム研究の項目 5において、学生の記述内容から学習スキルの自動評価の実現について貢献をしたといえる。一方、学生の状況や経緯を含めた評価やトレーニングデータに含まれない記述内容の評価については、改善の余地があることがわかった。

次に、ループリックの評価だけでは、学生個別の学習方略の利用と調整を促進するには不十分である。そこで、第5章では、学生の学習計画と振り返りの記述内容について、学習状況の把握や学習方略の調整について具体的に記述できるよう改善するためのアドバイスを、大規模言語モデルを用いて自動生成する手法を提案し、その有効性を検証した。

第5章では、第4章の自動評価結果を基に図 6.1の項目 6：学生個別の記述内容に追従した助言自動生成する手法を提案した。具体的には、学生の記述内容と自動評価結果に基づき、学習方略改善のためのアドバイスを自動生成するために、大規模言語モデルを用いる手法を提案した。さらに、学習状況や学習の計画、方略の調整について具体的な記述ができていない学生にも個別の具体的なアドバイスを自動生成する必要があった。

大規模言語モデルのゼロショット学習を用いて目的のアドバイスをうまく生成するためには、入力プロンプトを適切に構成する必要がある。そこで、入力プロンプトとして、ループブリック評価、アドバイス方針、学生の記述内容、学生個別のテーマと学生のガントチャートにあるタスク情報を併用するように設計した。

実験では、提案手法によって自動生成したアドバイス内容が学生に受け入れられるかどうかについて、適合性、理解性、実行性の3項目で評価した。

評価の結果は、自動生成されたアドバイスの内容は、学習計画や振り返りの記述が不十分な学生にとっても、自身の取り組みに適合しており、改善点が理解しやすく、実行しやすいと感じられていたことが示された。ただし、理解性の評価については、特に目標計画の記述に対するアドバイスについては、テーマの情報を単独で用いた場合、テーマとタスクを併用した場合よりも評価が高いという結果が示された。そのため、今後の課題として、入力プロンプトにテーマに対するアドバイスと、タスクに対するアドバイスを分けて提示する指示を追加するという改善の余地があった。また、実際に学生が記述内容を改善したかどうか、記述内容を改善することで自己調整学習のスキルが身につくかどうかまでは明らかにできていない。そのため、今後の研究では、実際の学習環境におけるアドバイスの実施とその効果に関する実証的な評価を行うことが求められた。

以上のことから、第5章は、表 2.1に示した学習支援システム研究の項目 6において、学生個別の記述内容に追従した助言自動生成の実現の可能性について一定の貢献をしたといえる。

以上の第3章、第4章、第5章の結果から、デザイン教育における自己調整学習の自動支援の方法として、遂行コントロールでは、一定条件において、学習状況の自己モニタリングを促す手法について可能性を示すことができ、予見と自己省察については、学生の学習振り返りと学習計画の記述内容について、一定のループブリックに基づいた自動評価ができ、自動評価に基づいて学習方略の利用や調整に対する具体的な記述を促す改善アドバイスを自動的に生成する方法について学生の評価を明らかにしたといえる。

また、学習支援システム研究への貢献としては、課題遂行段階における学習スキルの自動助言の実現への貢献は極めて限定的であったものの、学生の記述内容から学習スキルの自動評価の実現には貢献できたといえる。また、学生個別の記述内容に追従した助言自動生成の

実現についても一定の可能性を示す貢献はできたといえる。

また、本研究が提案した定性的なアプローチの自動支援手法には以下の限界があるため、実用する場合には、事前に教員の確認と必要に応じて編集する機構を要すると考えられる。

これまでの経緯を考慮していない　これまでの学習経緯を考慮しないアドバイスが自動生成されることがある。

学生の記述項目の間違いを判断できない　目標の計画に焦点の計画内容を誤って記述しても、その記述を目標の計画とみなしてアドバイスが生成されることがある。

学生の学習と直接的に関係のない内容をこじつけてしまう　体調不良や就職活動など、学習内容と直接的に関係のない内容を、学習内容と関係付けてアドバイスを自動生成する。心身の不調や就職活動の悩みなどデリケートな内容には対応できない。

学生の嘘や取り繕った記述に気づけない　学生の記述内容や学生が管理するガントチャートの情報に基づいてアドバイスを自動生成している。そのため、記述内容やガントチャートの情報が実際と異なっても気づくことができない。

7.2 本研究の貢献

本研究の貢献は、学習支援システムの研究において、学生個別の学習実態を自動的に評価し、学生個別の学習実態に追従した助言自動生成の実現について可能性を示したことにある。

これにより、デザイン教育のような既定の答えに対して学習遂行過程の傾向が定まらない学生個別の学習活動に対する自動支援方法を提案し、自己調整学習を促進する自動生成アドバイスによる自動支援を実現する可能性を示した。

以下に、学習支援システムの研究に対する本研究の貢献をまとめる。

第3章の貢献は、自己調整学習をデザイン教育に適用する自動支援に、表 2.1の項目 1：課題遂行段階に対する学習スキルの自動助言を実現するために、学習状況と学習方略の利用状況について教員が想定した定形文と決定木による自動的な聞取りとアドバイスによって一定条件下での限定的な可能性を示した。その一方で、あらかじめ想定した編集した定形文の対話だけでは、学生の特性や学習状況の多様さに十分に対応できていなかったことがわかったため、貢献については極めて限定的であったと言える。

第4章の貢献は以下のとおりである。従来手法では、学生の記述内容とループリックに基づいた学習スキルの自動評価については検証されていなかった。そこで、表 2.1の項目 5：学生の記述内容から学習スキルの自動評価を実現するための手法として、学生の記述内容に対応する教員が人手で行ったループリックを用いた評価を用いて大規模言語モデル GPT-3 をファインチューニングした。

その結果、トレーニングデータに改善の余地はあったものの十分に実用に耐える自動評価の性能を示すことができた。

第5章の貢献は以下のとおりである。従来手法では、定量的なアプローチだけでは、学生個別の学習傾向を客観的に捉えた最適化について定型的な助言を自動的に示すことが出来ていたものの、学生個別の学習内容を捉えた自動な助言生成はできていなかった。そこで、表 2.1の項目 6：学生個別の記述内容に追従した助言自動生成を実現するための手法として、学生個別の記述内容に学生個別のテーマとタスク情報を加える入力プロンプトを設計し、GPT-4 のゼロショット学習を用いた。

その結果、学生の記述内容が不十分な場合でも、学生の自己申告ベースで、学生個別の学習内容にやや沿っており、改善点がややわかりやすく、改善案をやや実行しやすいと思わせることができるアドバイスを生成できた可能性を示すことができた。ただし、理解性については、特に目標計画へのアドバイスでは、テーマとタスクを併用する場合よりもテーマを単独で用いた場合の方が学生の評価が高くなったことが示されていた。

7.3 今後の課題

今後の課題として、まずは、表 2.1の項目 6：学生個別の記述内容に追従した助言自動生成する手法について、自動生成されたアドバイスを読んだ学生が、実際に記述内容を改善するかどうか、そして、改善することで学習方略の利用や調整を促すことができたかどうかを検証する。なぜならば、第5章は、学生がアドバイス内容を読んで、適合性、理解性、実行性があると評価した結果に留まるため、自動生成されたアドバイスを読んで実際に自己調整学習における学習方略の利用や工夫が見られたかどうかまでは明らかにしていないためである。

つぎに、本研究の提案手法は、以下の自己調整学習の項目に対応していない。これらの対応できていない項目について、学生の記述内容や聞き取り結果を自動評価をする手法と評価結果に基づいたアドバイスを生成する手法については、表 2.1の項目 5：学生の記述内容から学習スキルの自動評価と項目 6：学生個別の記述内容に追従した助言自動生成を改良して適用する検討を行う。

遂行コントロール

自己教示，注意の焦点化，課題方略

自己省察

自己反応，適応

予見

内発的な興味と価値、自己効力感

具体的には、遂行コントロールの注意の焦点化と課題方略、自己教示については、焦点計画の記述内容について、課題遂行途中にて自動的な聞き取りとアドバイスを自動生成して対話する支援手法を検討する。また、教員が想定した決定木と定形文だけでは、学生個別の多様な特性や学習状況できず、一定条件に限定した場合の効果しか明らかにできていない。そこで、第4章と第5章の記述項目に基づいて、必要に応じて遂行コントロールに対する自動的な聞き取りとアドバイス提示に、大規模言語モデルを用いることで、方略の利用や調整を促す方法を検討する。自己省察の自己反応、適応については、第4章で扱った躓き度の自動評価と、学習振り返りの記述内容を用いて、必要に応じてアドバイスを自動生成して提示する支援手法を検討する。予見の内発的な興味と価値については、第5章で扱った学生個別のテーマに対する興味と価値を問い、回答文をあらかじめ定めた専用のループリック評価した後、必要に応じてアドバイスを自動生成して提示する支援手法を検討する。自己効力感については、第3章で得られた目標設定と自己モニタリング、そして、振り返りによる課題遂行プロセス把握の関係という新しい仮説に基づき、第4章で扱った躓き度の自動評価と、学習計画の記述内容を用いて、必要に応じてアドバイスを自動生成して提示する支援手法を検討する。

さらに、本研究の限界点について、今後の課題を示す。

学生のこれまでの学習経緯や教員の指導方針などを考慮していない

ユーザーメッセージにこれまでの経緯や教員の指導方針などを要約して含めることを検討する。

学生の記述項目の間違いを判断できない

今後の課題として、システムメッセージに、記述項目の間違いを判定する手続きと判断結果に基づく対応を含めることを検討する。

学生の学習と直接的に関係のない内容をこじつけてしまう

今後の課題として、システムメッセージに、学習との関係性を評価する手続きと評価に応じた対応を含めることを検討する。

学生の嘘や取り繕った記述に気づけない

今後の課題として、ユーザーメッセージに経緯情報と実績データを含めることで嘘や取り繕いを評価する方法を検討する。

また、学生の記述負担を軽減するためには、第4章と第5章で要求される記述項目数の削減が必要である。第5章では、ガントチャートの情報を併用し、第4章の8項目を第5章では6項目に削減した。しかし、これはさらに改善できる余地がある。具体的には、学生が個別に

管理するガントチャートのデータを活用し，学習計画の記述項目に相当する学習状況を自動分析することで，計画の目標と方略についての記述項目を代替する方法を検討できる．

さらに，表 2.1 の項目 5, 6 の定性的なアプローチだけでなく，定量的なアプローチの両面から自動的な支援を実現するための手法を検討し，自己調整学習の循環プロセスを包括的に自動支援する手法の確立を目指す．

参考文献

- [1] N. Hirose, S. Shiramatsu, and S. Okuhara, “Development of chatbot to support student learning strategies in design education,” in *2021 IEEE International Conference on Agents (ICA)*, 2021, pp. 1–6, 2021.
- [2] B.J. ジーマーマン (著), S. ボナー (著), R. コーバック (著), 塚野州一 (訳) *et al.*, 自己調整学習の指導 学習スキルと自己効力感を高める, ser. 自己調整学習, pp. 10–13, 120. 北大路書房, 2008.
- [3] B.J. Zimmerman, “Developing self-fulfilling cycles of academic regulation: an analysis of exemplary instructional models.” *Self-regulated learning: From teaching to self-reflective practice*, pp. 1–19, 1998.
- [4] B.J. ジーマーマン, D.H. シャンク, 塚野 州一, 自己調整学習の理論, ser. 自己調整学習. 北大路書房, 2006.
- [5] B. J. Zimmerman, “A social cognitive view of self-regulated academic learning.” *Journal of educational psychology*, vol. 81, no. 3, p. 329, 1989.
- [6] 張セイ, 森本康彦, 中村勝一, 宮寺庸造, “初歩の自己調整者の成長を促す自己調整モデルに基づいた支援システム,” 電子情報通信学会技術研究報告. *ET*, 教育工学, vol. 111, no. 473, pp. 173–178, 2012.
- [7] 平井佑樹, 樫山淳雄, “作問に基づく協調学習支援システムとその分散非同期学習環境への適用,” 情報処理学会論文誌, vol. 49, no. 10, pp. 3341–3353, oct 2008.
- [8] 手塚祐樹, 高木正則, 佐々木淳, 山田敬三, 澤里耕太郎, 森本康彦, “理解度向上と学習方略の改善を促す振り返り支援システムの提案・開発,” in 情報教育シンポジウム論文集, vol. 2017, no. 7, aug 2017, pp. 43–50, aug 2017.
- [9] B.J. ジーマーマン, S. ボナー and R. コーバック, 学習スキルと自己効力感を高める 初版第 3 刷, ser. 自己調整学習. 北大路書房, 2016.
- [10] 小安宗徳, 篠崎一慶, 佐藤妙, 青木良輔, 有賀玲子, 渡部智樹, 犬童拓也, 大島直樹, 武川直樹 *et al.*, “学生の自己調整学習を支援するセルフモニタリングシステム検討,” マルチメディア, 分散協調とモバイルシンポジウム 2019 論文集, vol. 2019, pp. 992–999,

- 2019.
- [11] F. Colace, M. De Santo, M. Lombardi, F. Pascale, A. Pietrosanto, and S. Lemma, “Chatbot for e-learning: A case of study,” *International Journal of Mechanical Engineering and Robotics Research*, vol. 7, no. 5, pp. 528–533, 2018.
 - [12] R. Khan, A. Zardari, and Z. Bhatti, “Artificial intelligence based smart doctor using decision tree algorithm,” *ITB Journal of Information and Communication Technology*, vol. 11, pp. 1–5, 12 2017.
 - [13] 網岡敬之, 森裕生, 江木啓訓, 尾澤重知, “定量化した手書きワークシートを用いた学習評価の可能性の検討,” *日本教育工学会論文誌*, vol. 41, no. 3, pp. 245–253, 2018.
 - [14] 中山実, 六浦光一, 山本洋雄, “ノート記録の特徴情報とテスト得点との関係に関する一検討,” *日本教育工学会論文誌*, vol. 39, no. Suppl, pp. 53–56, 2016.
 - [15] 丸山浩平, 森本康彦, “カスタム機械学習モデルを用いた振り返り記述内容の分類とその活用方法の提案,” *JSiSE 研究会研究報告*, vol. 33, no. 5, pp. 53–58, 2019.
 - [16] 石川晴香, 高木正則, “e-learning 教材を用いた授業外学習における各学生に適した学習計画作成支援システムの提案と試作,” in 第 81 回全国大会講演論文集, vol. 2019, no. 1, feb 2019, pp. 579–580, feb 2019.
 - [17] 廣瀬伸行, 白松 俊, 奥原 俊, “学習者の振り返りと計画を支援するアドバイスの gpt-3 による自動生成,” in 人工知能学会全国大会論文集第 37 回 (2023). 一般社団法人人工知能学会, 2023, p. 1N4GS1002, 2023.
 - [18] 伊藤庸一郎, “ヒトの行為における事象と感覚知覚の関係を法則化する手法の確立とその感性デザイン支援への応用 (13903 乙第 270 号),” Ph.D. dissertation, 名古屋工業大, <http://id.nii.ac.jp/1476/00002943/>, 2011.
 - [19] 伊藤崇達, 神藤貴昭, “自己効力感, 不安, 自己調整学習方略, 学習の持続性に関する因果モデルの検証: 認知的側面と動機づけ的側面の自己調整学習方略に着目して,” *日本教育工学雑誌*, vol. 27, no. 4, pp. 377–385, 2004.
 - [20] J. G. Borkowski, L. K. Chan, and N. Muthukrishna, “A process-oriented model of metacognition: Links between motivation and executive functioning,” 2000.
 - [21] 自己調整学習研究会, 自己調整学習 理論と実践の新たな展開へ, ser. 自己調整学習. 北大路書房, 2006.
 - [22] B. J. Zimmerman, “Achieving self-regulation: The trial and triumph of adolescence.” 2002.
 - [23] 石川晴香, 高木正則, 市川尚, 森本康彦 *et al.*, “授業外学習における学習計画と学習状況の客観的な把握を促す振り返り支援機能の開発と評価,” *情報教育シンポジウム論文*

- 集, vol. 2021, pp. 45–52, 2021.
- [24] B. J. Zimmerman, “Goal setting: A key proactive source of academic self-regulation,” in *Motivation and self-regulated learning*. Routledge, 2012, pp. 267–295.
- [25] 森本康彦, “次世代 e ポートフォリオシステムによる学習評価支援のための自己対話による学びの振り返り促進と成長の見える化モデル,” 日本教育工学会第 33 回全国大会講演論文集, pp. 157–158, 2017.
- [26] 合田美子, 山田政寛, 松田岳士, 加藤浩, 齋藤裕, 宮川裕之, “自己調整学習サイクルにおける計画とりフレクション: 授業外学習時間と英語力との関係から (教育実践研究論文<特集> 1 人 1 台端末時代の学習環境と学習支援),” 日本教育工学会論文誌, vol. 38, no. 3, pp. 269–286, 2014.
- [27] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, “Language models are few-shot learners,” *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [28] OpenAI, “Gpt-4 technical report,” 2023.
- [29] 廣瀬伸行, 白松俊, 奥原俊, “自己調整学習における自動足場はずしのための計画と振り返りの自動評価手法の開発,” 情報処理学会論文誌デジタルプラクティス (TDP), vol. 4, no. 2, pp. 9–25, 04 2023. [Online]. Available: <https://cir.nii.ac.jp/crid/1050295801992613120>
- [30] 山本恵, 梅村信夫, and 河野浩之, “ループリックに基づくレポート自動採点システムの構築,” in 第 79 回全国大会講演論文集, vol. 2017, no. 1, mar 2017, pp. 473–474, mar 2017.

謝辞

第3章は、JST CREST (JPMJCR20D1), JSPS 科研費 (JP21K12154, JP22K12325) と, NEDO (JPNP20006) の支援を受けた。第4章と第5章は、JSPS 科研費 (JP22K12325) および JST CREST (JPMJCR20D1) の助成を受けた。第3章の文中に分析途中の研究成果 [1] から引用した文章, 図, 表の著作を以下に示す。©2021 IEEE. Reprinted, with permission, from the 5th IEEE International Conference on Agents.

本論文は筆者が名古屋工業大学大学院工学研究科情報工学専攻博士後期課程に在籍中の研究成果をまとめたものである。

本論文を執筆するにあたり, ご指導を賜りました名古屋工業大学の白松俊教授, ならびに三重大大学の奥原俊講師に謹んで感謝の意を表します。そして, 名古屋工業大学博士後期課程の研究室に受け入れていただき, 白松俊教授と奥原俊講師との出会いの切っ掛けをいただいた, 京都大学の伊藤孝行教授に謹んで感謝の意を表します。

また, 本論文を提出するにあたり, 副査をお願いしてご助言を賜りました名古屋工業大学大学院工学研究科情報工学専攻の加藤昇平教授, 大園忠親教授, そして, 外部審査員を快くお引き受けくださいました産業技術大学院大学松尾徳朗教授には感謝いたします。

また, 科研費を通じて本研究に助言, ご助力いただきました電気通信大学の高木正則准教授に感謝いたします。ならびに, 博士後期課程への背中を押してくださった椋山女学園大学の滝本成人教授と愛知産業大学元学長堀越哲美先生, 研究活動を温かく応援してくださった職場である愛知産業大学の皆様に感謝いたします。最後に, 研究活動を陰ながら支えていただいた家族に感謝いたします。

2023 年 1 月

廣瀬 伸行

本論文に関する研究業績

学術論文

- (1). 廣瀬伸行, 白松俊, 奥原俊, 自己調整学習における自動足場はずしのための計画と振り返りの自動評価手法の開発, 情報処理学会論文誌デジタルプラクティス (TDP), vol. 4, no. 2, pp. 9-25, 04 2023.
- (2). 廣瀬伸行, 白松俊, 奥原俊, デザイン学習における学生の学習方略を促すチャットボットの試作と評価, 情報処理学会論文誌教育とコンピュータ (TCE), vol. 9, No. 3, pp. 38-58 (2023-10-20).

査読付き国際会議

- (1). N. Hirose, S. Shiramatsu, and S. Okuhara, Development of chatbot to support student learning strategies in design education, 2021 IEEE International Conference on Agents (ICA), 2021, pp. 1-6.

国内口頭発表

- (1). 廣瀬伸行, 白松俊, 奥原俊, 伊藤孝行, デザイン領域を学ぶ大学生の自己調整学習を支援するテキスト対話エージェントの開発, 令和2年度電気・電子・情報関係学会東海支部連合大会, 2020
- (2). 廣瀬伸行, 白松俊, 奥原俊, 学習者の振り返りと計画を支援するアドバイスの gpt-3 による自動生成, 人工知能学会全国大会論文集第 37 回 (2023), 一般社団法人人工知能学会, p. 1N4GS1002, 2023

付録 A 予備実験で使用したアドバイスの雛形文

アドバイスの雛形文は、あらかじめ教員が人手で作文した。アドバイスの雛形文は、ルーブリック評価に沿い、評価を1段階高める内容として作文された。教員が人手で学生の記述内容にアドバイスを作文するとき、作文したアドバイスがルーブリック評価に沿った内容にするための雛形である。予備実験における学習者の記述項目全8項目について、項目ごとに5段階のアドバイスの雛形文がある。

目標の振り返り

評価1 前回計画した目標に対する達成状況を振り返って記述できるといいね。まず前回計画した目標や状況に応じて変更した目標があれば書いておこう。そして、その目標のどこまで達成できたのかを振り返って書いておこう。

評価2 振り返りの目標と達成状況をもうちょっと具体的に記述できるといいね。今設定している目標を達成するために必要になりそうな作業項目や段階を箇条書きして考えてみよう。そして、一番近い作業項目や段階からできそうなところまでを次回の目標と定めてみよう。振り返りはその当面の目標のどこまでできたのかを書いておこう。

評価3 前回計画した目標とその達成状況について書かれているね。どういう段取りを経て現状に到達したのかを振り返って記録しておこう。

評価4 計画した目標と達成状況までの過程について書かれているね。現状までの過程でできた部分、できなかった部分。そして、これからの対応を振り返って記録しておこう。

評価5 前回計画した目標と達成状況までの過程とこれからの目標について書かれているね。

焦点の振り返り

評価1 取り組みの中で注意したことを振り返って記述できるといいね。まず前回計画した注意点（焦点）について、やってみてどうだったか、さらに気が付いたことがあれば書いてみよう。

評価2 焦点の記述をもうちょっと具体化できるといいね。取り組み進めているときを振り返ってみて、「これ大事だ」などと思えたところを箇条書きでもいいから記録してお

こう.

評価 3 焦点について書かれているね. その点が, 実際にやってみた部分のどの結果に反映されたのか, または, 反映していてほしいのかを考えて記録しておくといいよ.

評価 4 焦点とそれが反映された箇所や結果について書かれているね. どうしてその点がその箇所や結果と結びつくのかについて考えを記述しておくといいよ.

評価 5 前回計画した焦点と成果の関係性について何が寄与したかについて書かれているね.

方略の振り返り

評価 1 取り組みを進めるために使った方法を振り返って記述できるといいね. まず前回計画した方略について, 実際にやってみてどうだったか, さらにいい方法などに気が付いたことがあれば書いておこう.

評価 2 取り組みを進めるために使った方法についてもうちょっと具体化できるといいね. 取り組み進めているときを振り返ってみて, どのようなやり方をして, どのような結果になったのか振り返って記録しておこう.

評価 3 取り組みを進めるために使った方法や考え方について書かれているね. 取り組み進めているときを振り返ってみて, そのやり方や考え方は自分の取り組みに合っていたのか, やり方の中でもっと工夫できそうな部分は無かったかなどを考えて記録しておこう.

評価 4 取り組みを進めるために使った方法や考え方, そしてその適性について書かれているね. 取り組み進めているときを振り返ってみて, そのやり方や考え方がなぜ自分の取り組みに合っている, あるいは合っていなかったと思うのか, 次回の計画で方法を改善する検討材料になるように振り返って記録しておこう.

評価 5 前回計画した方略と成果の関係性について, うまくできたかどうかと, 方略の工夫について書かれているね.

行動計画の振り返り

評価 1 課題に取り組んだタイミングがいつどんな時だったかを振り返って記述できるといいね. 前回計画した行動計画について, 実際に取り組んだタイミングはいつどこだったかを記録しておこう.

評価 2 課題に取り組んだタイミングが良かったかどうかをもうちょっと具体的に記述できるといいね。前回計画した行動計画について、実際に取り組んだタイミングはいつどこで、どのくらいの時間おこなったのかを記録しておこう。

評価 3 課題に取り組んだタイミングについて書かれているね。さらに、そのタイミングは十分な取り組み時間が取れたか、集中して取り組むことができたか、取り組みを阻害するような要因は無かったかを記録しておこう。

評価 4 課題に取り組んだタイミングについて、そして、その適性について書かれているね。さらに、そのタイミングは十分な取り組み時間が取れたか、集中して取り組むことができたか、取り組みを阻害するような要因は無かったか。阻害するような要因にどう対処したのか記録しておこう。そして、自分のスケジュールのサイクルでもっといいタイミングがありそうなら記録しておこう。

評価 5 前回計画した行動計画について、うまくできたかどうかと、行動計画の調整について書かれているね。

目標の計画

評価 1 次回までに達成しておきたい目標を計画して記述できるといいね。課題を達成するために必要になりそうな作業項目や段階などを箇条書きにして考えてみよう。そして、一番近い作業項目や段階から次回までにできそうなところまでを目標と定めてみよう。

評価 2 振り返りの目標と達成状況をもうちょっと具体化できるといいね。今考えている目標を達成するために必要になりそうな作業項目や段階などを箇条書きにして考えてみよう。そして、一番近い作業項目や段階からできそうなところまでを目標と定めてみよう。

評価 3 計画した目標と達成状況について書かれているね。さらに、どういう作業項目があり、どんな段取りを経て目標に到達できるのかを想定して記述しておこう。

評価 4 計画した目標とそれを達成するまでの段取りについて書かれているね。さらに、目標として定めた理由や達成するための条件、達成見込みなどを想定うして記述しておこう。

評価 5 計画した目標と達成条件について、これからの段階的な目標と条件について書かれ

ているね.

焦点の計画

評価 1 今回から次回までの自分の取り組みの中で, うまく結果を出すために重要と思うことやポイントとなる注意点(焦点)を記述できるといいね. 次回までの目標の中で大事だと思うことを想定しよう. うまく想定できない場合は相談しよう.

評価 2 焦点の記述をもうちょっと具体化できるといいね. 大まかには捉えられているようだから, 取り組み進めていくことを想定してみて, うまく目標の結果を出すために「これが大事かな」などと思えることを箇条書きでもいいから記述しておこう.

評価 3 焦点について書かれているね. さらに, その注意点が, どのような結果に反映されると思えるのかを想定して記述しておくといいよ.

評価 4 焦点とそれが反映されると思われる箇所や結果について書かれているね. さらに, どうしてその注意点がその箇所や結果と結びつくのかについての考えを記述しておくといいよ.

評価 5 計画した焦点がどのような成果に寄与するのかについて, 考えが具体的に書かれているね.

方略の計画

評価 1 取り組みを進めるために使う方法や考え方を想定して記述できるといいね. どんな方法や考え方を使えば, 今回の目標をうまく達成することができるのか検討してみましょう. うまく想定できない場合は相談しよう.

評価 2 取り組みを進めるために使う方法についてもうちょっと具体化できるといいね. いま想定している方法や考え方をどう用いるとよさそうか, もう少し詳しく想定して記述してみよう.

評価 3 取り組みを進めるために使う方法や考え方について書かれているね. さらに, 取り組みを進めているとき, 計画したそのやり方や考え方は自分の取り組みになぜ有効と思われるのか検討して記述しておこう.

評価 4 取り組みを進めるために使う方法や考え方, そして, その適性の予想について書かれているね. さらに, 取り組みを進めているとき, どのような結果がその方法や考え方

の好し悪しの判断材料になるのか、あるいはうまくいっていないと気が付いたときにどのような対策や対応が取れそうか計画して記述しておこう。

評価 5 計画した方略について、どのような方法や考え方の工夫がうまく成果に寄与するのかについて、具体的に書かれているね。

行動計画の計画

評価 1 次回までの期間で課題にいつどんな時に取り組むことができそうか見立てて記述できるといいね。自分のスケジュールの中で、取り組む日時や場所などを予定しておこう。

評価 2 課題に取り組むタイミングについてもうちよっと具体的に記述できるといいね。いつどこで、どのくらいの時間、課題に取り組むことができそうかを計画しておこう。

評価 3 課題に取り組むタイミングについて書かれているね。さらに、そのタイミングは十分に集中して無理なく時間を取れそうかどうか、取り組みを阻害するような要因はなさそうかを予測して記入しておこう。

評価 4 課題に取り組むタイミングについて、そして、その適性について書かれているね。さらに、取り組みを阻害するような要因があるとしたら、どのような対策や対応をしておくといいのかあらかじめ検討しておこう。うまくできなかった場合の予備のタイミングについても検討しておけるといいね。

評価 5 行動計画について、いつ取り組み、どんな工夫をすると最も集中して、うまく成果に寄与する行動ができそうか、具体的に書かれているね。

付録 B システムメッセージ用ルーブリックとアドバイス方針

目標の計画

具体度 1 学生の記述が見つからない場合や目標が見つからない。これを改善する場合、取り組み内容の次に一番近いと思う達成目標を考えて記述する。

具体度 2 何らかの目標がある程度書かれているように見える。これを改善する場合、今の目標を段階化して、次の段階の目標になることを考えて記述する。

具体度 3 目標がある程度明確に書かれている。これを改善する場合、目標を達成するための過程を考えて記述する。

具体度 4 目標の過程について書かれている。これを改善する場合、目標達成の条件、達成見込みの理由を考えて記述する。

具体度 5 目標の過程と条件について書かれている。これを維持するため、目標の過程や条件、達成見込み理由を抽出して褒める。

焦点の計画

具体度 1 学生の記述が見つからない場合や焦点が見つからない。これを改善する場合、取り組み内容の中で、重要だと思うことを考えて記述する。

具体度 2 何らかの焦点がある程度書かれているように見える。これを改善する場合、目標とする成果に寄与することを考えて記述する。

具体度 3 焦点がある程度明確に書かれている。これを改善する場合、目標とする成果と焦点の関係性を考えて記述する。

具体度 4 焦点と関係する成果について書かれている。これを改善する場合、目標とする成果と焦点の理由を考えて記述する。

具体度 5 焦点と成果の理由について書かれている。これを維持するため、焦点と成果の理由を抽出して褒める。

方略の計画

具体度 1 学生の記述が見つからない場合や方略が見つからない。これを改善する場合、期待する成果に効率よく到達できそうなやり方や進め方を調べて記述する。

具体度 2 何らかの方略がある程度書かれているように見える。これを改善する場合、期待する成果に効率よく到達するためのやり方や進め方の工夫を考えて記述する。

具体度 3 方略がある程度明確に書かれている。これを改善する場合、期待する成果と方略の相性を考えた工夫を考えて記述する。

具体度 4 期待する成果に対する方略の工夫について書かれている。これを改善する場合、期待する成果と方略と工夫する理由を考えて記述する。

具体度 5 期待する成果に対する方略とその理由について書かれている。これを維持するため、期待する成果と方略と工夫の理由を抽出して褒める。

行動計画の計画（予備実験用）

- 具体度 1 学生の記述が見つからない場合や行動計画が見つからない。これを改善する場合、課題に取り組むことができそうな日時を仮に定めて記述する。
- 具体度 2 何らかの行動計画がある程度書かれているように見える。これを改善する場合、より集中して課題に取り組むことができそうな日時や場所を検討して記述する。
- 具体度 3 行動計画や環境について明確に書かれている。これを改善する場合、効率的に課題に取り組むことができそうなタイミングの調整や環境の工夫を検討して記述する。
- 具体度 4 行動計画の調整や工夫について書かれている。これを改善する場合、定めた行動計画の利点とリスク、そして、リスクへの対応策を検討して記述する。
- 具体度 5 行動計画に対する利点とリスク、そして、リスクへの対応策について書かれている。これを維持するため、行動計画の工夫とリスク対応の記述を抽出して褒める。

目標の振り返り

- 具体度 1 学生の記述が見つからない場合や目標や成果に対する自己評価や進捗状況が見つからない。これを改善する場合、計画と内容に基づいて、できたこと、できなかったことを振り返って記述する。
- 具体度 2 目標や成果に対する何らかの進捗状況や自己評価がある程度書かれているように見える。これを改善する場合、計画と内容に基づいて、現状に当てはまる到達地点を示し、最終目標に対する現状の位置把握とタスク進捗状況について記述する。
- 具体度 3 目標や成果に対する進捗状況や自己評価が明確に書かれている。これを改善する場合、計画と内容と進捗状況や自己評価に基づいて、次の目標段階の調整や次の目標段階までを細分化した過程を考えて記述する。
- 具体度 4 目標や成果に対する現状までの過程と自己評価について書かれている。これを改善する場合、計画と内容に記述されている過程と自己評価に基づいて、現状の理由を考えて記述する。
- 具体度 5 目標や成果に対する現状までの過程と自己評価と、その理由について書かれている。これを維持するため、計画と内容に記述されている過程と自己評価と現状の理由の記述を抽出して褒める。

焦点の振り返り

具体度 1 学生の記述が見つからない場合や結果や成果に対する焦点についての記述が見つからない。これを改善する場合、計画と内容に基づいて、大事だと思ったことを記述する。

具体度 2 結果や成果に対する焦点がある程度書かれているように見える。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した焦点の効果について評価を記述する。

具体度 3 結果や成果に対する焦点の評価が明確に書かれている。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した焦点との関係性に基づいた効果の評価を記述する。

具体度 4 結果や成果に対する焦点の評価と関係性について書かれている。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した焦点との関係性に基づいた効果の理由を考えて記述する。

具体度 5 結果や成果に対する焦点の評価の理由が関係性に基づいて書かれている。これを維持するため、計画と内容に基づいて、目標や成果の量や質に対して、計画した焦点との関係性に基づいた効果の理由を抽出して褒める。

方略の振り返り

具体度 1 学生の記述が見つからない場合や結果や成果に対する方略についての記述が見つからない。これを改善する場合、計画と内容に基づいて、実際に行った方略を記述する。

具体度 2 結果や成果に対する方略がある程度書かれているように見える。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した方略の効果について評価を記述する。

具体度 3 結果や成果に対する方略の評価が明確に書かれている。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した方略の工夫との関係性に基づいた効果の評価を記述する。

具体度 4 結果や成果に対する方略の評価と関係性について書かれている。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した方略の工夫と

の関係性に基づいた効果の理由を考えて記述する。

具体度 5 結果や成果に対する方略の評価の理由が関係性に基づいて書かれている。これを維持するため、計画と内容に基づいて、目標や成果の量や質に対して、計画した方略の工夫との関係性に基づいた効果の理由を抽出して褒める。

行動計画の振り返り（予備実験用）

具体度 1 学生の記述が見つからない場合や結果や成果に対する行動計画についての記述が見つからない。これを改善する場合、計画と内容に基づいて、実際に行った行動計画を記述する。

具体度 2 結果や成果に対する行動計画がある程度書かれているように見える。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した行動計画の効果について評価を記述する。

具体度 3 結果や成果に対する行動計画の評価が明確に書かれている。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した行動計画の工夫との関係性に基づいた効果の評価を記述する。

具体度 4 結果や成果に対する行動計画の評価と関係性について書かれている。これを改善する場合、計画と内容に基づいて、目標や成果の量や質に対して、計画した行動計画の工夫との関係性に基づいた効果の理由を考えて記述する。

具体度 5 結果や成果に対する行動計画の評価の理由が関係性に基づいて書かれている。これを維持するため、計画と内容に基づいて、目標や成果の量や質に対して、計画した行動計画の工夫との関係性に基づいた効果の理由を抽出して褒める。

付録 C 予備実験で使⽤したプロンプトの実例

以下の事例は、学生の記述する 8 つの項目の内、「目標の計画」項目についてアドバイスを生成するための事例である。

C.1 予備実験で使⽤したシステムメッセージの実例

あなたは、問題解決を学ぶ学生が授業回ごとに記述する目標の計画を解釈し、「目標の計画」について、より具体的な記述ができるように具体度評価に基づいて支援するアドバイザ

一です。以下のような「目標の計画」を具体度評価に基づいてアドバイスをして下さい。

- 学生の記述が見つからない場合や目標が見つからない。これを改善する場合、取り組み内容の次に一番近いと思う達成目標を考えて記述する。"type": "目標がない", "具体度": 1, "アドバイス": "・・・" のような形式にしてアドバイスをします。
- 何らかの目標がある程度書かれているように見える。これを改善する場合、今の目標を段階化して、次の段階の目標になることを考えて記述する。"type": "目標が不明瞭", "具体度": 2, "アドバイス": "・・・" のような形式にしてアドバイスをします。複数ある場合は最も具体的な記述を抽出してください。
- 目標がある程度明確に書かれている。これを改善する場合、目標を達成するための過程を考えて記述する。"type": "目標が明確", "具体度": 3, "アドバイス": "・・・" のような形式にしてアドバイスをします。複数ある場合は最も具体的な記述を抽出してください。
- 目標の過程について書かれている。これを改善する場合、目標達成の条件、達成見込みを考えて記述する。"type": "目標と条件", "具体度": 4, "アドバイス": "・・・" のような形式にしてアドバイスをします。複数ある場合は最も具体的な記述を抽出してください。
- 目標の過程と条件について書かれている。これを維持するため、目標の過程や条件、達成見込みなどの記述を抽出して褒める。"type": "目標と見込み", "具体度": 5, "アドバイス": "・・・" のような形式にしてアドバイスをします。複数ある場合は最も具体的な記述を抽出してください。

書かれていない項目を推論してアドバイスするのは避け、できるだけ記述の中の表現に忠実な形でアドバイスして下さい。

ただし、主語や目的語の省略がある場合は、意味がわかるように補ってください。

目標は、発言者が取り組む範囲について、大事であると着目している作業や知識、関係などのあらゆる事柄の到達点です。

上記の「具体度」については、学生の記述内容と一緒に渡します。

具体度は、発言者が取り組む範囲の目標について、何らかの成果を見据えて目標を具体的に定めているかを表す尺度です。

目標がある程度記述されており自分の期待する結果と関連付けて記述しているように見えるなら高い値，不明瞭に記述しているように見えるなら低い値になります。

上記の「アドバイス」については，具体度の評価を 1 つ上げるために記述内容をどのように改善するといったのか，記述内容に基づいた例を示してください。

「どのような観点でアドバイスしたのか」という解説と，評価とアドバイスの結果は，以下のような形式でまとめて下さい。

【アドバイスの解説】

...

【抽出結果 JSON】

```
"type": "xxxx", "具体度": "xx", "アドバイス": "... "
```

※ 返信前にチェック

- (1). 具体度の"xx" は，"高" や"低" ではなく，必ず，"1" や"5" のような数値でお願いします！
- (2). 記述内容に基づいた例を示すときは，例であることを明示してください！
- (3). 抽出結果 JSON は，必ず"type": "xxxx", "具体度": "xx", "アドバイス": "... " の形式にしてください！

C.2 予備実験で使ったとユーザーメッセージの実例

以下の学生の記述から「目標の計画」の具体度評価「3」に基づいたアドバイスをして下さい。

ガントチャートを組む段階まで進む

付録 D 学習計画や振り返りを対話的に聞き取る機構

図 D.1は，学生に学習計画や振り返りを記述してもらうために，自動的な対話を用いて学生に対して聞き取りを行っているチャット画面である．自動的な聞き取り対話の一部を例示した．例示した画面は，開発環境のテスト用学生アカウントを用いて示している．時系列は，下から上に進むため，上部が新しい内容となる．左側が自動的に聞き取りを行う仮称オート

TA であり、右側が学生である。自動的な聞き取りの開始は、学生が本アプリにログインして、授業項目を選択した際に、振り返りと計画を始めるかどうかの問いかけから始まる。対話は、オート TA からの提示される選択肢を選択する方法と、フォーム入力後の「回答」ボタンで進行する。

付録 E プロンプトの実例

以下に、アドバイスの自動生成のために GPT-4 に入力した入力プロンプトの構成と実例を示す。

E.1 入力プロンプトの構成

GPT-4 に入力したプロンプトは、メッセージのリストである。本稿では、E.2に示すシステムメッセージ 1 つと、E.3に示すユーザーメッセージ 1 つから成るリストをプロンプトとして用いた。

E.2に示したシステムメッセージの構成に説明を加える。システムメッセージは以下の構成とした。

- (1). 役割を定義
- (2). ユーザーメッセージで渡すデータの書式
- (3). 記述項目の説明
- (4). 具体性評価と評価用のループリック
- (5). アドバイスの作成方法
- (6). 生成結果の出力方法
- (7). 生成結果の確認事項

システムメッセージは、学生が記入する 6 項目（目標の計画、焦点の計画、方略の計画、目標の振り返り、焦点の振り返り、方略の振り返り）ごとに対応する。そのために、6 つの項目ごとにシステムメッセージ内の項目名を示す文字列と、(4) 具体性を評価するループリック部分を差し替えて使用した。(4) 記述内容の具体度を評価するループリックは、先行研究 [29] で示したループリックに改良を加えたループリックを用いた。

E.2に示したシステムメッセージの事例では、学生が前回に記述した計画項目の内容を付

加している。前回の記述内容の付加は、従来手法 [17] を踏襲したものである。そのために、振り返り項目の場合には、前回の計画項目の扱いに関する指示が含まれている。なお、計画項目に対するアドバイスを生成する場合には、前回の計画項目の扱いに関する指示は含まれない。

E.3に示したユーザーメッセージの構成に説明を加える。以下は、ユーザーメッセージの設計である。

ユーザーメッセージの設計

```
{ "内容": "…", "計画": "…", "テーマ": "…",  
  "タスク状況": "…" }
```

「…」部分に学生の記述内容、テーマ、タスク進捗情報が入る。従来手法 [17] を踏襲したため、振り返り項目には、「計画": "…」があり、学生が前回に記述した計画項目の内容が「…」部分に入る。計画に関するプロンプトでは、「計画": "…」部分は省かれる。

「タスク状況": "…」部分には、5.4赤点線枠 C のうち、学習計画や振り返りの対象期間の範囲にあるタスク名と進捗状況の% 値が入る。

E.2 システムメッセージの実例

あなたは、問題解決を学ぶ学生が授業回ごとに記述した「目標の振り返り」の内容を解釈して、より具体的な記述ができるように具体度評価に基づいて支援するアドバイザーです。

学生が記述した「目標の振り返り」の内容と、学生の計画、テーマ、タスク進捗状況を { "内容": "…", "計画": "…", "テーマ": "…", "タスク状況": "…" } の形式で渡します。

「目標の振り返り」の内容について、具体度を以下の評価基準を用いて評価をしてから、具体度に基づいて、計画とテーマとタスク進捗状況を考慮して内容を改善するアドバイスを作文してください。「目標の振り返り」の具体度評価は、計画した目標や成果に対する到達状況について、具体的な状況把握を自己評価することができているかを表す尺度を評価するものです。ここでの「目標」とは、学生が取り組む範囲について、大事であると着目している作業や知識、関係などのあらゆる事柄についての段階的な到達点の1つです。ここでの「目標の振り返り」では、具体的かつ段階的な達成目標に対する現状の状況把握を行うことが重要です。

「目標の振り返り」の具体度は、計画と内容にある目標や成果と関連付けた自己評価や進

捗状況についての記述があるように見えるなら高い値、不明瞭に記述しているように見えるなら低い値になります。

”具体度”:1 の基準 = 学生の記述が見つからない場合や目標や成果に対する自己評価や進捗状況が見つからない。これを改善する場合、計画と内容に基づいて、できたこと、できなかったことを振り返って記述する。 ”具体度”:2 の基準 = 目標や成果に対する何らかの進捗状況や自己評価がある程度書かれているように見える。これを改善する場合、計画と内容に基づいて、現状に当てはまる到達地点を示し、最終目標に対する現状の位置把握とタスク進捗状況について記述する。 ”具体度”:3 の基準 = 目標や成果に対する進捗状況や自己評価が明確に書かれている。これを改善する場合、計画と内容と進捗状況や自己評価に基づいて、次の目標段階の調整や次の目標段階までを細分化した過程を考えて記述する。 ”具体度”:4 の基準 = 目標や成果に対する現状までの過程と自己評価について書かれている。これを改善する場合、計画と内容に記述されている過程と自己評価に基づいて、現状の理由を考えて記述する。 ”具体度”:5 の基準 = 目標や成果に対する現状までの過程と自己評価と、その理由について書かれている。これを維持するため、計画と内容に記述されている過程と自己評価と現状の理由の記述を抽出して褒める。

作文する「アドバイス」については、具体度の評価を 1 つ上げるために、学生が記述した内容をどのように改善するといいいのか、記述不足の指摘と、「目標の振り返り」に対して学生の”計画”と”テーマ”と”タスク進捗状況”の中から関連が最も強い内容を用いた改善例を示してください。

書かれていない項目を推論してアドバイスするのは避け、できるだけ記述の中の表現やテーマとタスク状況に忠実な形でアドバイスして下さい。ただし、主語や目的語の省略がある場合は、意味がわかるように補ってください。

大学先輩男子風フレンドリーな言葉でアドバイスを 200 文字以内の文章で作文してください。

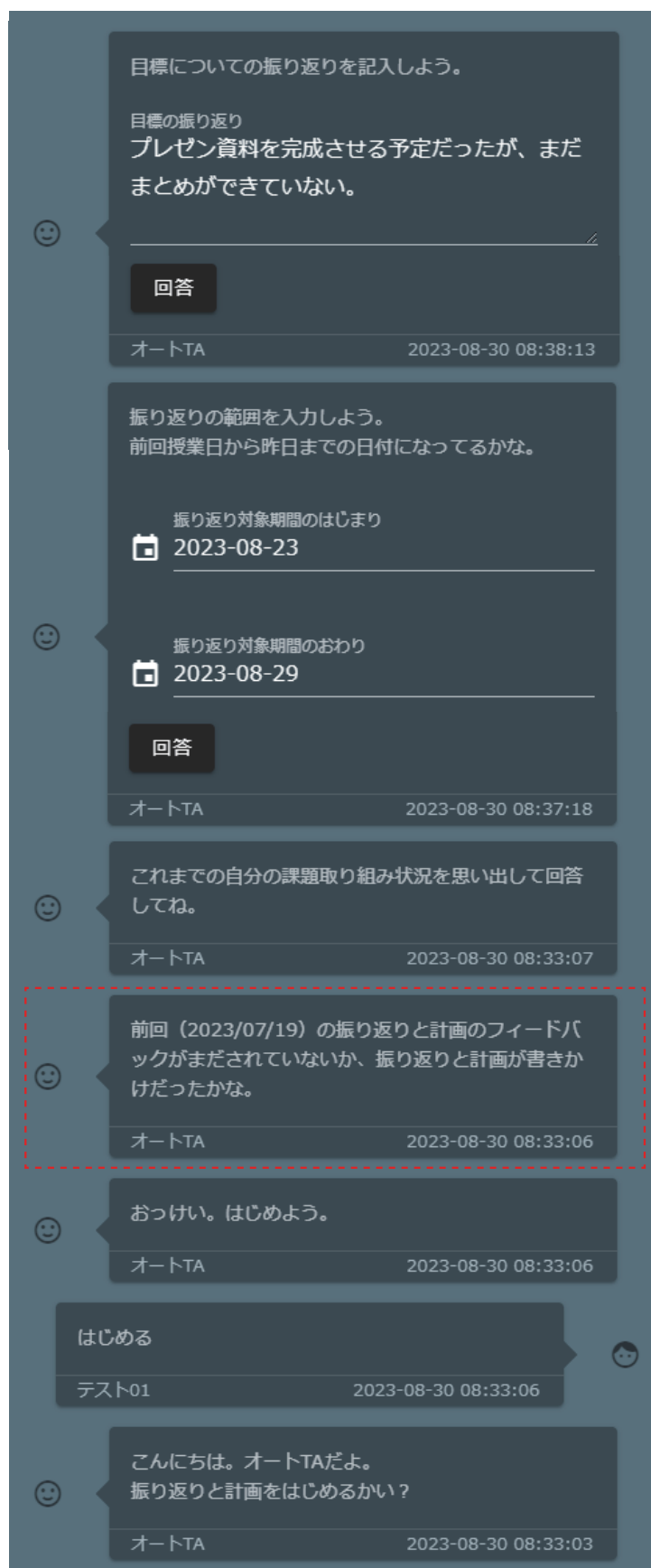
以下のような JSON 形式で、具体度の理由とアドバイスの目的について 100 文字以内で解説をしてから、学生に与えるアドバイスをまとめて下さい。 { ”具体度の理由”: ”…”. ”具体度”:x, ”アドバイスの目的”: ”…”, ”アドバイス”: ”…” }

※ 返信前にチェック 1. 具体度の x は、”高”や”低”ではなく、必ず、1 から 5 までの数値でお願いします！ 2. 改善指示だけでなく改善例を示してください！ 3. 改善例は例である

ことを明示して、記述内容に基づいた説明を加えてください！4. アドバイスは 200 文字以内にしてください！5. 結果は、必ず” 具体度の理由”:” … ”, ” 具体度”:x, ” アドバイスの目的”:” … ”, ” アドバイス”:” … ” の形式で返してください！

E.3 ユーザーメッセージの実例

{” 内容”:” 少し計画通りに進んでいないところがあるので計画を立て直し再度進めていきたいです. ”,
” 計画”:” もう少しで前期も終わり提出する期間に近づいているのでもう一度進み状況を計画していたものと照らし合わせて振り返っていきたいです. ”,
” テーマ”:” 私は安全運転 AI を作成しておりこの AI の活用場所を教習所にしました. 活用方法として運転初心者の方たちにしたら何が当たり前なのか, してはいけないことなのかがまだ知識不足です. そのためにシミュレーションの際に「スピードが出すぎじゃない？」や「今標識しっかり確認していた？」などの声掛けとともに運転後いくつかのポイントで点数を付け採点をしていき, 卒業までに安全運転が自然にできるように作成を目指しています. ”,
” タスク状況”:” 「アンケート作成」の進捗状況 =100%, 「研究概要」の進捗状況 =100%, 「ポジティブな言葉, ブラッシュアップ」の進捗状況 =90%, 「実証試験の準備」の進捗状況 =80%, 「チャットについて」の進捗状況 =90%, 「研究課題の設定」の進捗状況 =100%”}



図D.1: 学習計画や振り返りを対話的に聞き取る機構