

間主観的な評価における適応型自己評価に関する研究 のグランドデザイン

眞坂 美江子¹ 宮崎 誠¹ 渡辺 博芳²

概要：ルーブリックやチェックリストを用いた間主観的評価において、継続的な自己評価は学習者に持続的な成長を促す効果が期待されるが、評価すべき指標の増加に伴う、評価疲れが問題となる。そこで、本研究では自己評価者の回答パターンに応じて最適な評価指標を提示することにより、より少ない評価指標の数で、すべての評価指標を用いた場合と同等の自己評価結果を得る「適応型自己評価」の開発を目指す。客観テストにおいては、項目応答理論(IRT)を用いた適応型テストが実用化されているが、間主観評価における適応型評価に関する研究は見あたらない。また、客観テストと異なり、適切な自己評価を行えない学習者の存在を考慮する必要がある。これについては自己評価の整合性に基づく助言を行うことを検討する。本論文では、帝京大学情報電子工学科における汎用的能力育成の取組を事例として、「適応型自己評価」の方法を開発する研究の構想と初期の検討状況について述べる。

キーワード：適応型自己評価、間主観的評価、ルーブリック、汎用的能力、

Grand Design of Research on Adaptive Self-Assessment for Inter-Subjective Assessments

MIEKO MASAKA^{†1} MAKOTO MIYAZAKI^{†1} HIROYOSHI WATANABE^{†2}

Abstract: In intersubjective assessment using rubrics and checklists, continuous self-assessment is expected to promote learners' continuous growth. However, there is a problem that learners become tired of the task of self-assessment when there are many indicators to be assessed. Therefore, this study aims to develop an "adaptive self-assessment" that provides the same self-assessment results as when all the indicators are used, but with fewer indicators, by presenting the most appropriate indicators according to the learner's response patterns. Although adaptive testing using item response theory (IRT) has been put to practical use in objective testing, there is no research on adaptive self-assessment in intersubjective assessment. In this paper, we present the design of a study to develop a method of "adaptive self-assessment" using the case study of the efforts of the Department of Information Electronics Engineering at Teikyo University to cultivate generic skills. The status of the initial study is also described.

Keywords: Adaptive self-assessment, Inter-subjective assessment, Rubric, Generic skills,

1. はじめに

従来は高等教育において知識・技能の修得が重視されていたが、近年はそれらに加えて修得した知識を活用した問題解決力、問題解決のための思考力・判断力・表現力等の修得も重視した教育が行われている。このような力は、「21世紀型スキル」、文部科学省の「学士力」、経済産業省の「社会人基礎力」などにおいて具体的な力が示されている。また、こうした力は「ソフトスキル」[1][2]と呼ばれることもある。専門的な知識や技能などの「ハードスキル」は教育や訓練によって修得され、定量的に評価しやすいのに対して、「ソフトスキル」は経験によって修得され、定量的な評価が難しいとされている。

帝京大学理工学部情報電子工学科においては、このようなソフトスキルを「汎用的能力」と位置づけ、JABEE[3]対応の教育プログラムの導入を機に、知識獲得と汎用的能力育成の両方を重視した教育カリキュラムを整備して教育実

践を進めている。本学科では、汎用的能力の育成のために、評価指標となるルーブリックとチェックリストを提案し、大学4年間の教育カリキュラムを通じた評価の枠組みを構築した[4]。この枠組みでは、提案したルーブリックとチェックリストを用いた学生による自己評価活動をベースとしていることから、汎用的能力自己評価システムを開発して活用している[5][6]。

この枠組みにおいて総括的評価に位置づけられるワークショップに参加した学生のアンケートからは「自分の今の能力を把握する良い機会となった」、「これから力を入れるべき内容が良く分かった」などの回答が得られ、一定の効果をあげている。一方で、「評価項目が多い」、「後半になるほど、疲れて評価がいい加減になってしまう」との意見も見られ、いわゆる「評価疲れ」が課題となっている。これは自己評価における評価項目数が多いことに起因しているため、より少ない評価項目ですべての評価項目を用いたときと同等の結果が得られれば、学生の負担を軽減できる

¹ 帝京大学 Teikyo University

² 電気通信大学 The University of Electro-Communications

のではないかと考えられる。

客観テストにおいては、項目反応理論(IRT)を用いて能力に応じた設問を提示することで、より少ない設問数で精度よく能力判定が可能な適応型テストが実用化されている。そこで、類似したアプローチによって適応的に自己評価を行うことが考えられる。ただし、ルーブリックやチェックリストを用いた評価は間主観的な評価であり、客観テストと異なり、より主観的な評価となることから、評価結果にばらつきが生じる可能性が大きい。また、自己評価を適切に行える学習者と適切には行えない学習者がいると考えられるので、このことも評価結果のばらつきの要因となり得る。

このような特性を持つ間主観的な自己評価を適応的に行う方法に関する研究は著者らの調査の範囲では見あたらない。適応型自己評価(Adaptive Self-Assessment)をキーワードとして文献を検索すると、適応型自己評価を表題に含む文献[7][8][9]が存在する。しかし、これらは学習者の自己学習における自己評価において、従来の適応型テストを効果的に用いる方法に関する研究であり、間主観的な評価指標を用いた自己評価とは異なる。

評価項目数あるいは質問項目数をできるだけ少なくして負担を軽減したいというニーズは、メンタルヘルスのためのストレスチェックなどにも存在する。奥村らは、教師のストレス自己評価において、共通項目への回答状況から、以降は高ストレス用か低ストレス用のいずれかに回答すればよいような2段階テストによって自己評価者の負担を軽減する方法について検討している[10]。このような適応型の自己評価を実現できると、学生の負担軽減につながる事が期待される。ただし、本研究が対象とする自己評価においては学習者が適切に自己評価をできているかという点も考慮する必要があるが、ストレスチェックのような診断では通常は考慮しない。

本研究において、「適応型自己評価」とは、「自己評価者の評価指標への回答パターン等に応じて異なる評価指標を適応的に提示することにより、より少ない評価指標の数ですべての評価指標を用いた場合と同等の適切な自己評価結果を得る方法」と定義する。本論文は、帝京大学情報電子工学科での汎用的能力評価の枠組みを例として、適応型自己評価の方法を整理し、適応型自己評価を実現する際の課題、および適応型自己評価実現のための研究方法を示すことを目的とする。

2. 評価の枠組みと適応型自己評価導入の方針

2.1 汎用的能力評価の枠組み

帝京大学情報電子工学科で対象とする汎用的能力を表1に示す。表1で番号を付けた9項目を観点とする汎用的能力ルーブリックを定義し、英語コミュニケーション力を除く8項目についてチェックリストを定義した。つまり、

表 1 対象とする汎用的能力

1.情報リテラシー
2.思考力・問題解決力
コミュニケーション力
3.文章コミュニケーション力
4.口頭コミュニケーション力
5.数量的スキル
6.英語コミュニケーション力
主体的・継続的な実行力
7.主体的・継続的な学習力
8.実行力
9.チーム活動能力

汎用的能力ルーブリック

	0	1	2	3	4	
情報リテラシー						チェックリスト
思考力・問題解決力						チェックリスト
文章コミュニケーション						チェックリスト
:						:
チーム活動力						チェックリスト

図 1 評価指標の階層性

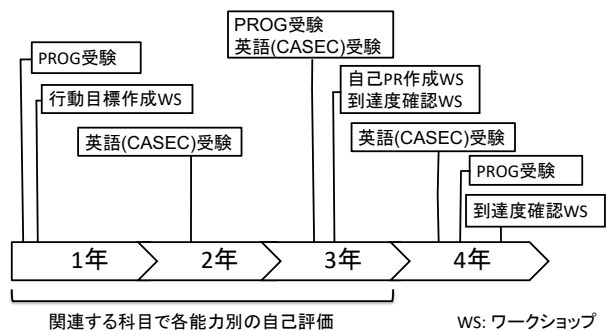


図 2 4年間の教育カリキュラムでの汎用的能力の評価の枠組み

図1に示すように、ルーブリックが上位でルーブリックの各観点が1つの能力を表し、それらに対して下位にその能力のチェックリストが存在するような階層構造になっている。

図2に帝京大学情報電子工学科における大学4年間の学習カリキュラムにおける汎用的能力評価の枠組みを示す。ジェネリックスキルの成長を支援するアセスメントプログラムPROGと英語コミュニケーションテストCASECも活用し、到達度確認ワークショップにおいて、汎用的能力を全体的に自己評価する。また、各能力については1年生から3年生の間に各能力の育成に重点をおいた科目内で継続的に自己評価を行う。関連する科目で行う各能力の自己評価とワークショップにおける全体的な自己評価では同じ評価指標を用いている。

各能力の評価においては、最初にその能力のチェックリストを用いて自己評価を行い、次にその結果を参照しながら、その能力に対応するルーブリック1つの観点の自己評価を行う。その上で、その能力に関して良い点と改善点を

記述する。

到達度確認ワークショップでは、8つの能力についてチェックリストとルーブリックを使った自己評価を行い、英語コミュニケーションについてルーブリックによる自己評価を行う。3年生の到達度確認ワークショップにおいて、従来の表計算ソフトを用いた自己評価(2020年度)と汎用的能力自己評価システムを用いた自己評価(2022年度・2023年度)について、自己評価に要する時間を図3に示す。システム導入により、大幅に時間を短縮することができたが、システム導入後も多くの学生が30分から45分、最も長い学生では1時間30分程度を要している。

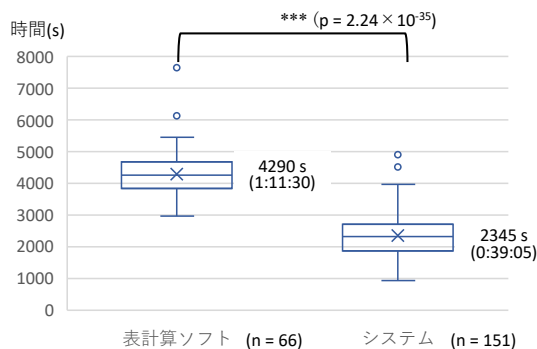


図3 自己評価に要する時間[6]

2.2 対象とする自己評価の特徴

前節で述べた汎用的能力評価の枠組みにおける自己評価は、次のような特徴を持つ。

- 評価指標には階層構造があり、上位のルーブリックの1つの観点に対して下位に位置づけられるチェックリストが定義されている。
- カリキュラムの中で、同じ評価指標を用いた自己評価が複数回行われる。
- 自己評価の評価指標(チェックリスト)は、各能力においてどのようなことができるようになるべきかを示しており、自己評価活動には評価指標の内容(クリアすべき目標)を知るといふ意義もある。
- すべての学生が適切な自己評価を行えるとは限らない。

2.3 適応型自己評価の導入方針

前節で述べた特徴から、適応型自己評価の導入方針について検討する。

まず、特徴(c)のように自己評価活動では評価指標の内容を理解することも重要であるので、ある評価指標を用いた自己評価活動を一度しか行わない場合に、適応型自己評価を行うのは不適切である。したがって、特徴(b)を持つ場合、つまり同じ評価指標を用いた自己評価が複数回行われる場合に限り導入し、複数回のうちの前半はすべての評価指標を用いた自己評価を行い、後半に適応型自己評価を導入するのが望ましい。

次に特徴(d)のすべての学生が適切な自己評価を行えるとは限らないことも考慮して、適応型自己評価を導入することが望ましい。特徴(a)を持つ場合は、下位の評価指標での自己評価結果と上位の評価指標での自己評価結果の整合性を検証し、学生へ助言を与えることが考えられる。また、下位の評価指標(項目)間に順序性や関連性が見いだせれば、評価項目間の整合性を検証して、学生へ助言を与えることも考えられる。このような自己評価結果の整合性に基づく助言を、すべての評価指標を用いて自己評価を行う段階で提示することで、学生が適切に自己評価を行う力が向上することが期待される。また、適応型自己評価においても、自己評価の結果の整合性を検証して、助言を与えられるようにすることは効果的であると考えられる。

以上をまとめると、適応型自己評価の導入方針は次のようになる。

- 同じ評価指標を用いた自己評価が複数回行われる場合に限り導入し、複数回のうちの後半に適応型自己評価を導入する。
- 複数回のうちの前半には自己評価結果の整合性に基づく助言を与え、後半の適応型自己評価においても結果の整合性に基づく助言を与えることが望ましい。

帝京大学情報電子工学科の枠組みで考えると、複数回のうちの前半は各能力に関連した科目内で行う能力別の自己評価、後半は3年次と4年次に実施する到達度確認ワークショップに対応付けるのがよいと思われる。

表2 適応的自己評価で用いるデータ

Ra	ルーブリックの当該能力の回答結果(上位の評価指標の回答結果). 1回の自己評価に1つ.
C	チェックリスト(チェック項目の質問リスト). 能力ごとに個数が異なる. i 番目のチェック項目を C_i で表す.
Ca	チェックリストに対する回答結果. 個数はチェックリストと同じ. i 番目のチェック項目への回答を Ca_i で表す.
Pra	ルーブリックの当該能力の過去の回答結果. その自己評価者が過去に当該能力の自己評価を実施した回数分だけ存在. j 回目の自己評価の時の回答結果を Pra_j で表す.
Pca	チェックリストに対する過去の回答結果. その自己評価者が過去に当該能力の自己評価を実施した回数分, チェック項目の数だけ存在. j 回目の自己評価における i 番目のチェック項目の回答を Pca_{ij} で表す.

3. 適応型自己評価の処理

3.1 処理で利用するデータ

ある1人の自己評価者に対する適応型自己評価の処理で用いるデータを表2に示す。ある能力の1回の自己評価では、Cのチェックリストについて回答し、チェックリスト

表3 適応型自己評価の処理アルゴリズム

1. チェックリストCに質問していない項目がある間繰り返す
 - 1.1 チェックリストCから質問していない項目 C_i を選択
 - 1.2 項目 C_i を質問し, Ca_i に回答を得る
 - 1.3 Pra, Pca, 回答済 Ca から, Ra と Ca の残りを推定
 - 1.4 もし 信頼度が 閾値 R_i 以上で推定できたならば
推定結果を Ra, Ca に保存して繰り返しを抜ける
2. もし 推定結果を得ていないならば
 - 2.1 総合評価を質問し, Ra に回答を得る

表4 助言付自己評価のアルゴリズム

1. チェックリストCに対する回答を Ca に得る
2. 総合評価回答を Ra に得る
3. Ra と Ca に対して整合性を検査
4. もし整合性に問題があれば
 - 4.1 整合検査結果に基づき助言を表示する
 - 4.2 自己評価結果の再検討を促し, Ca と Ra に結果を得る

表5 助言付適応型自己評価のアルゴリズム

1. チェックリストCに質問していない項目が有る限り繰り返す
 - 1.1 チェックリストCから質問していない項目 C_i を選択
 - 1.2 項目 C_i を質問し, Ca_i に回答を得る
 - 1.3 Pra, Pca, Ca から, Ra と Ca の残りを推定
 - 1.4 もし 信頼度が R_i 以上で推定できたならば
推定結果を Ra, Ca に保存して繰り返しを抜ける
2. もし 推定結果を得ていないならば
 - 2.1 総合評価を質問し, Ra に回答を得る
 - 2.2 Ra と Ca に対して整合性を検査
 - 2.3 もし整合性に問題があれば
 - 2.3.1 整合検査結果に基づき助言を表示する
 - 2.3.2 自己評価結果の再検討を促し, Ra と Ca に結果を得る

の自己評価結果は Ca に保存される。その能力の総合的な評価結果(図1のルーブリックの1行分)の自己評価結果は Ra に保存される。過去の自己評価における総合評価は Pra に、チェックリストの結果は Pca に保存されている。

3.2 適応型自己評価のアルゴリズム

適応型自己評価の処理アルゴリズムの例を表3に示す。表3はチェックリストのチェック項目を1つずつ質問し、回答を得る度に残りの結果を推定する流れになっている。1.3の推定では、推定結果の信頼度を算出するものとし、1.4の R_i は推定結果を採用する信頼度の閾値を表す。

自己評価者の評価指標への回答パターン等に応じて異なる評価指標を適応的に提示する方法は他にもあり得るが、本研究では一旦、表3のアルゴリズムを想定して検討する。

3.3 助言付自己評価のアルゴリズム

すべてのチェック項目を用いて行う自己評価、つまり適応型でない自己評価において、自己評価結果の整合性に基づく助言を行う処理アルゴリズムの例を表4に示す。表4は、ひととおり自己評価が終了してから整合性を検査して、その結果に基づいて助言を表示する。他に、チェック項目の回答を1つずつ得る度に整合性を検査することも考えられるが、2.3節で述べた複数回の自己評価の前半では、自己評価の途中に介入するのは適切ではないと考えた。表4の3の整合性を検査する際には、Ra と Ca 以外に、過去の自己評価結果である Pra と Pca を用いることも考えられる。

3.4 助言付適応型自己評価のアルゴリズム

自己評価結果の整合性検査による助言を含む適応型自己評価の処理アルゴリズムの例を表5に示す。3.2節と3.3節で述べたアルゴリズムを合成した処理になっている。

4. 研究課題と研究体制

4.1 研究課題

3章で述べた自己評価処理から、本研究の研究課題として、自己評価結果の整合性に基づく助言と適応的自己評価の2つが挙げられる。

4.1.1 自己評価結果の整合性に基づく助言

自己評価結果の整合性に基づく助言については、自己評価結果の整合性をどのように検証するかが重要な課題となる。自己評価結果の整合性の検証方法が定めれば、検証結果に基づいてどのような助言を行うかも検討しやすくなる。また、整合性がないと判定された自己評価結果については、もう1つの課題である適応的自己評価における結果の推定法を検討する際にそれらのデータを除外することも考えられる。

図1で示した評価指標の階層性から、自己評価結果の整合性は次の2つが考えられる。

- ・チェックリスト内のチェック項目間の整合性
- ・チェックリストの評価結果とルーブリックの評価結果(総合評価)の間の整合性

これらの整合性の検証方法を検討し、検証結果に基づいた助言の内容を検討する予定である。

4.1.2 適応型自己評価

適応型自己評価においては、未回答項目の結果の推定が重要な課題となる。表3のアルゴリズムからは、次に質問する項目の選択方法も課題となるが、これは未回答項目の結果の推定方法に依存すると考えられる。同様に推定結果の信頼度をどのように計算するか、信頼度の閾値 R_i をどのように決定するかという問題もあるが、これらも回答結果の推定法に依存する。

表3のアルゴリズムで1.3の処理が未回答項目の結果推定に該当するが、この処理を実現しようとすると、選択された質問項目やそれに対する自己評価者の回答を含めてシミュレーションする必要が生じる。そこで、次のように、段階的な推定課題を複数設定する。

- (a) 1回の自己評価データを用いて、チェックリストの回答結果からルーブリックの回答結果を推定する。
- (b) 1回の自己評価データを用いて、チェックリストの項目数のうちの1項目の結果を未回答として、残りの回答結果を用いてその結果を推定する。
- (c) (b)において未回答の項目数をどれだけ増やせるかを検証する。
- (d) 2回分の自己評価データを用いて、上の(a)から(b)について検討する。

その後、3回分の自己評価データを用いるなどして、最終的には表3の1.3の処理が行えるような方法を検討する。ただし、収集できている自己評価データ数を考慮して、当面は上記の(a)から(d)の課題について検討する予定である。

4.2 研究体制

本研究は複数の研究機関に属する研究者によって進める。具体的には帝京大学と電気通信大学である。研究の対象となるデータは帝京大学で運用されている汎用的能力自己評価システムに蓄積されている。そこで、図4に示す体制で研究を進める。

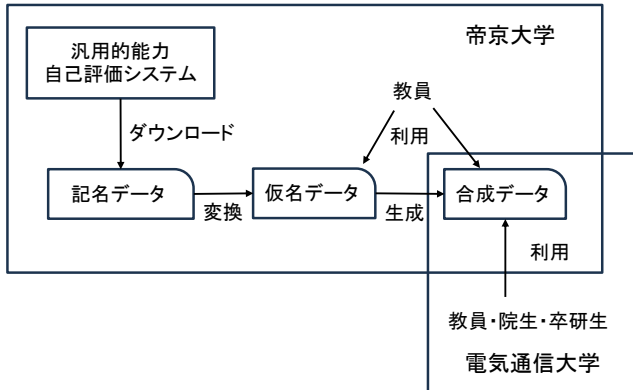


図4 研究体制と利用するデータ

まず、汎用的能力自己評価システムからダウンロードした記名データを仮名データに変換する。次にこの仮名データから合成データを生成する。合成データは仮名データをもとに統計的な特性を保持した人工のデータである。したがって、合成データには個人情報を含まれず、個人情報を推定することもできない。電気通信大学に所属する研究者は合成データのみを扱う。一方、帝京大学に所属する研究者は合成データと仮名データの両方を扱って検討を進める。

電気通信大学側で開発した手法やプログラムは、合成データを用いて検討を進めるが、最終的には、それらの手法やプログラムを帝京大学側に提供し、帝京大学の研究者が仮名データを使って有用性を検証する予定である。

5. 対象となるデータ

5.1 データ形式

まず、汎用的能力自己評価システムからダウンロードしたデータから、表6に示す形式の仮名データを作成する。これは、1名のある能力についての1回分の自己評価を表すデータである。person-idは同じ個人が同一のIDとなるようにして仮名化してある。index-idは評価をする能力、つまり使われる評価指標を識別するIDであり、その値は表7に示す通りである。activity-idは1回のワークショップ、ある授業での1回の自己評価活動など、活動を識別するIDである。自己評価では各能力の自己評価事に良い点と改善点を記述することになっているが、それに記述した文字数が

表6 仮名データの形式

person-id	個人を識別する ID.
index-id	評価指標を識別する ID
date	自己評価を実施した月日
activity-id	自己評価活動を識別する ID
activity-type	活動のタイプ w すべての項目を評価するワークショップ、 i 各科目で一部の項目を評価する活動
assessment-result	ループリックの評価結果(総合評価) 0~4 (0が低く、4が高い)
reflection-length	振り返りに記述した文字数
c-1	1番目のチェック項目の結果 0~2 (0が低く、2が高い)
:	以下、チェック項目数分存在

表7 評価指標

評価指標番号	能力	項目数
1	情報リテラシー	13
2	思考力・問題解決力	18
3	文章コミュニケーション	35
4	口頭コミュニケーション	26
5	数量的スキル	13
6	主体的・継続的な学習力	8
7	実行力	11
8	チーム活動力	10

reflection-lengthである。最後にc-1から表7に示す項目数だけチェック項目の結果が存在する。チェック項目の結果は次の値を取る。

- 0 あまりできていない
- 1 ある程度できている
- 2 できている

合成データは表6と同じ形式であるが、dateとactivity-idは削除する。

5.2 合成データの生成方法

合成データは、実在するデータと同じ構造で異なる値を持つデータであり、実在する個人のデータを直接用いないことで、プライバシーを保護できる。合成データの生成方法は、複数提案されている[11][12]。本研究においては、個人個人の回答の特徴を保持することが重要であることから、仮名データをベースに、共分散行列からデータを生成する手法を用いた。統計分析ソフトHAD[13]は、ベースとなるデータと全く同じ平均、共分散行列を持つデータを生成する機能を持つので、この機能を利用することとした。ただし、ここでHADによって生成された合成データは、実数値となるので、評価結果として取り得る整数値になるように調整する。4.1.2節で述べた(a)~(c)を検討するための合成データの具体的な手順は、次の通りである。

- (1) 表6の情報を1行で記述したcsvファイル(仮名データ

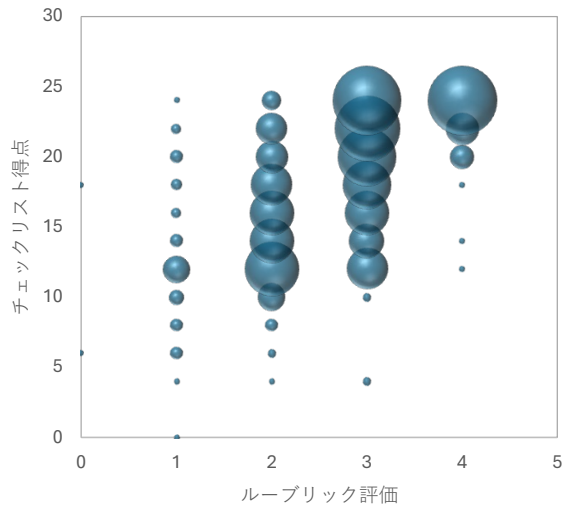
分類	チェックリスト	情報機器の活用					情報の収集					情報の分析				
		c-1	c-2	c-3	c-4	c-5	c-6	c-7	c-8	c-9	c-10	c-11	c-12	c-13		
情報機器の活用	c-1	1.00														
	c-2	0.19	1.00													
	c-3	0.12	0.25	1.00												
	c-4	0.08	0.33	0.16	1.00											
	c-5	0.10	0.12	0.03	0.14	1.00										
情報の収集	c-6	0.06	0.04	0.01	0.03	0.09	1.00									
	c-7	0.01	0.05	0.04	0.04	0.05	0.20	1.00								
	c-8	0.00	0.01	0.01	0.03	0.05	0.19	0.20	1.00							
	c-9	0.00	0.04	0.02	0.07	0.06	0.19	0.16	0.13	1.00						
	c-10	0.02	0.05	0.00	0.03	0.03	0.08	0.08	0.13	1.00						
情報の分析	c-11	0.03	0.00	0.06	0.00	0.00	0.09	0.09	0.07	0.04	0.19	1.00				
	c-12	-0.01	0.06	0.01	0.07	0.03	-0.03	-0.02	0.08	0.16	0.09	0.17	1.00			
	c-13	-0.01	-0.03	0.09	-0.02	0.03	0.03	-0.01	0.11	0.08	0.09	0.18	0.25	1.00		

(a) 情報リテラシー(n=1432)

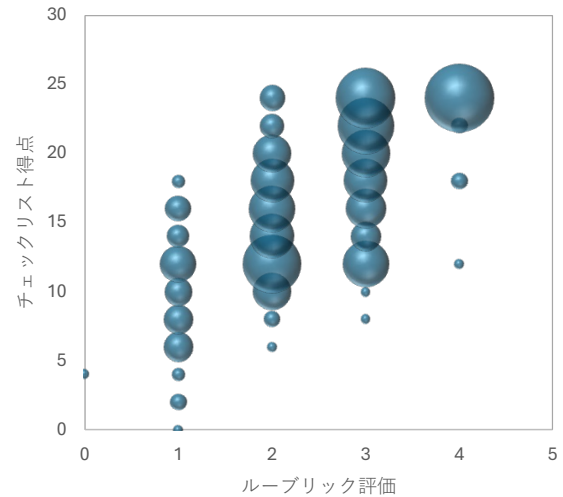
分類	チェックリスト	数学的表現				グラフ					表			
		c-1	c-2	c-3	c-4	c-5	c-6	c-7	c-8	c-9	c-10	c-11	c-12	c-13
数学的表現	c-1	1.00												
	c-2	0.32	1.00											
	c-3	0.22	0.40	1.00										
	c-4	0.10	0.22	0.19	1.00									
グラフ	c-5	0.21	-0.01	0.02	-0.05	1.00								
	c-6	0.11	0.00	0.00	0.07	0.20	1.00							
	c-7	0.02	0.03	0.02	0.03	0.14	0.28	1.00						
	c-8	0.01	-0.05	0.04	0.12	0.14	0.18	0.22	1.00					
	c-9	-0.12	0.06	0.11	0.22	0.02	0.11	0.04	0.14	1.00				
表	c-10	0.04	-0.11	0.08	0.00	0.13	0.01	0.09	0.07	-0.05	1.00			
	c-11	0.01	-0.03	0.05	-0.01	0.08	0.01	0.02	-0.06	0.10	0.27	1.00		
	c-12	-0.01	0.03	-0.03	-0.12	0.00	0.10	0.14	-0.01	0.06	0.18	0.36	1.00	
	c-13	-0.03	0.06	0.04	0.11	-0.03	-0.03	-0.01	0.16	0.04	0.21	0.12	0.25	1.00

(c) 数量的スキル(n=512)

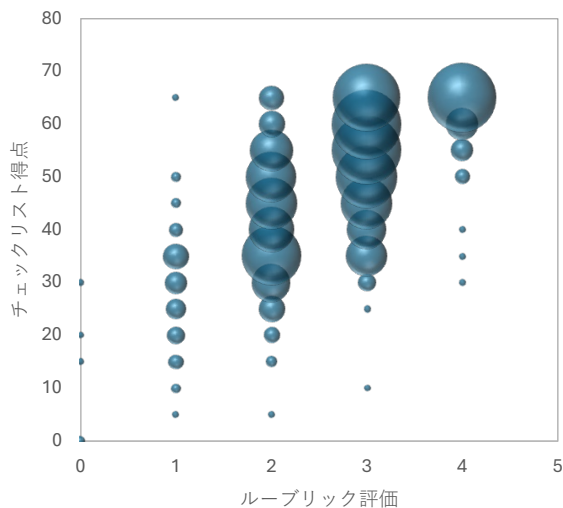
分類	チェックリスト	構成上内容										意味										形式										資料の参照										要件・状況																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
		c-1	c-2	c-3	c-4	c-5	c-6	c-7	c-8	c-9	c-10	c-11	c-12	c-13	c-14	c-15	c-16	c-17	c-18	c-19	c-20	c-21	c-22	c-23	c-24	c-25	c-26	c-27	c-28	c-29	c-30	c-31	c-32	c-33	c-34	c-35	c-36	c-37	c-38	c-39	c-40	c-41	c-42	c-43	c-44	c-45	c-46	c-47	c-48	c-49	c-50	c-51	c-52	c-53	c-54	c-55	c-56	c-57	c-58	c-59	c-60	c-61	c-62	c-63	c-64	c-65	c-66	c-67	c-68	c-69	c-70	c-71	c-72	c-73	c-74	c-75	c-76	c-77	c-78	c-79	c-80	c-81	c-82	c-83	c-84	c-85	c-86	c-87	c-88	c-89	c-90	c-91	c-92	c-93	c-94	c-95	c-96	c-97	c-98	c-99	c-100	c-101	c-102	c-103	c-104	c-105	c-106	c-107	c-108	c-109	c-110	c-111	c-112	c-113	c-114	c-115	c-116	c-117	c-118	c-119	c-120	c-121	c-122	c-123	c-124	c-125	c-126	c-127	c-128	c-129	c-130	c-131	c-132	c-133	c-134	c-135	c-136	c-137	c-138	c-139	c-140	c-141	c-142	c-143	c-144	c-145	c-146	c-147	c-148	c-149	c-150	c-151	c-152	c-153	c-154	c-155	c-156	c-157	c-158	c-159	c-160	c-161	c-162	c-163	c-164	c-165	c-166	c-167	c-168	c-169	c-170	c-171	c-172	c-173	c-174	c-175	c-176	c-177	c-178	c-179	c-180	c-181	c-182	c-183	c-184	c-185	c-186	c-187	c-188	c-189	c-190	c-191	c-192	c-193	c-194	c-195	c-196	c-197	c-198	c-199	c-200	c-201	c-202	c-203	c-204	c-205	c-206	c-207	c-208	c-209	c-210	c-211	c-212	c-213	c-214	c-215	c-216	c-217	c-218	c-219	c-220	c-221	c-222	c-223	c-224	c-225	c-226	c-227	c-228	c-229	c-230	c-231	c-232	c-233	c-234	c-235	c-236	c-237	c-238	c-239	c-240	c-241	c-242	c-243	c-244	c-245	c-246	c-247	c-248	c-249	c-250	c-251	c-252	c-253	c-254	c-255	c-256	c-257	c-258	c-259	c-260	c-261	c-262	c-263	c-264	c-265	c-266	c-267	c-268	c-269	c-270	c-271	c-272	c-273	c-274	c-275	c-276	c-277	c-278	c-279	c-280	c-281	c-282	c-283	c-284	c-285	c-286	c-287	c-288	c-289	c-290	c-291	c-292	c-293	c-294	c-295	c-296	c-297	c-298	c-299	c-300	c-301	c-302	c-303	c-304	c-305	c-306	c-307	c-308	c-309	c-310	c-311	c-312	c-313	c-314	c-315	c-316	c-317	c-318	c-319	c-320	c-321	c-322	c-323	c-324	c-325	c-326	c-327	c-328	c-329	c-330	c-331	c-332	c-333	c-334	c-335	c-336	c-337	c-338	c-339	c-340	c-341	c-342	c-343	c-344	c-345	c-346	c-347	c-348	c-349	c-350	c-351	c-352	c-353	c-354	c-355	c-356	c-357	c-358	c-359	c-360	c-361	c-362	c-363	c-364	c-365	c-366	c-367	c-368	c-369	c-370	c-371	c-372	c-373	c-374	c-375	c-376	c-377	c-378	c-379	c-380	c-381	c-382	c-383	c-384	c-385	c-386	c-387	c-388	c-389	c-390	c-391	c-392	c-393	c-394	c-395	c-396	c-397	c-398	c-399	c-400	c-401	c-402	c-403	c-404	c-405	c-406	c-407	c-408	c-409	c-410	c-411	c-412	c-413	c-414	c-415	c-416	c-417	c-418	c-419	c-420	c-421	c-422	c-423	c-424	c-425	c-426	c-427	c-428	c-429	c-430	c-431	c-432	c-433	c-434	c-435	c-436	c-437	c-438	c-439	c-440	c-441	c-442	c-443	c-444	c-445	c-446	c-447	c-448	c-449	c-450	c-451	c-452	c-453	c-454	c-455	c-456	c-457	c-458	c-459	c-460	c-461	c-462	c-463	c-464	c-465	c-466	c-467	c-468	c-469	c-470	c-471	c-472	c-473	c-474	c-475	c-476	c-477	c-478	c-479	c-480	c-481	c-482	c-483	c-484	c-485	c-486	c-487	c-488	c-489	c-490	c-491	c-492	c-493	c-494	c-495	c-496	c-497	c-498	c-499	c-500	c-501	c-502	c-503	c-504	c-505	c-506	c-507	c-508	c-509	c-510	c-511	c-512	c-513	c-514	c-515	c-516	c-517	c-518	c-519	c-520	c-521	c-522	c-523	c-524	c-525	c-526	c-527	c-528	c-529	c-530	c-531	c-532	c-533	c-534	c-535	c-536	c-537	c-538	c-539	c-540	c-541	c-542	c-543	c-544	c-545	c-546	c-547	c-548	c-549	c-550	c-551	c-552	c-553	c-554	c-555	c-556	c-557	c-558	c-559	c-560	c-561	c-562	c-563	c-564	c-565	c-566	c-567	c-568	c-569	c-570	c-571	c-572	c-573	c-574	c-575	c-576	c-577	c-578	c-579	c-580	c-581	c-582	c-583	c-584	c-585	c-586	c-587	c-588	c-589	c-590	c-591	c-592	c-593	c-594	c-595	c-596	c-597	c-598	c-599	c-600	c-601	c-602	c-603	c-604	c-605	c-606	c-607	c-608	c-609	c-610	c-611	c-612	c-613	c-614	c-615	c-616	c-617	c-618	c-619	c-620	c-621	c-622	c-623	c-624	c-625	c-626	c-627	c-628	c-629	c-630	c-631	c-632	c-633	c-634	c-635	c-636	c-637	c-638	c-639	c-640	c-641	c-642	c-643	c-644	c-645	c-646	c-647	c-648	c-649	c-650	c-651	c-652	c-653	c-654	c-655	c-656	c-657	c-658	c-659	c-660	c-661	c-662	c-663	c-664	c-665	c-666	c-667	c-668	c-669	c-670	c-671	c-672	c-673	c-674	c-675	c-676	c-677	c-678	c-679	c-680	c-681	c-682	c-683	c-684	c-685	c-686	c-687	c-688	c-689	c-690	c-691	c-692	c-693	c-694	c-695	c-696	c-697	c-698	c-699	c-700	c-701	c-702	c-703	c-704	c-705	c-706	c-707	c-708	c-709	c-710	c-711	c-712	c-713	c-714	c-715	c-716	c-717	c-718	c-719	c-720	c-721	c-722	c-723	c-724	c-725	c-726	c-727	c-728	c-729	c-730	c-731	c-732	c-733	c-734	c-735	c-736	c-737	c-738	c-739	c-740	c-741	c-742	c-743	c-744	c-745	c-746	c-747	c-748	c-749	c-750	c-751	c-752	c-753	c-754	c-755	c-756	c-757	c-758	c-759	c-760	c-761	c-762	c-763	c-764	c-765	c-766	c-767	c-768	c-769	c-770	c-771	c-772	c-773	c-774	c-775	c-776	c-777	c-778	c-779	c-780	c-781	c-782	c-783	c-784	c-785	c-786	c-787	c-788	c-789	c-790	c-791	c-792	c-793	c-794	c-795	c-796	c-797	c-798	c-799	c-800	c-801	c-802	c-803	c-804	c-805	c-806	c-807	c-808	c-809	c-810	c-811	c-812	c-813	c-814	c-815	c-816	c-817	c-818	c-819	c-820	c-821	c-822	c-823	c-824	c-825	c-826	c-827	c-828	c-829	c-830	c-831	c-832	c-833	c-834	c-835	c-836	c-837	c-838	c-839	c-840	c-841	c-842	c-843	c-844	c-845	c-846	c-847	c-848	c-849	c-850	c-851	c-852	c-853	c-854	c-855	c-856	c-857	c-858	c-859	c-860	c-861	c-862	c-863	c-864	c-865	c-866	c-867	c-868	c-869	c-870	c-871	c-872	c-873	c-874	c-875	c-876	c-877	c-878	c-879	c-880	c-881	c-882	c-883	c-884	c-885	c-886	c-887	c-888	c-889	c-890	c-891	c-892	c-893	c-894	c-895	c-896	c-897	c-898	c-899	c-900	c-901	c-902	c-903	c-904	c-905	c-906	c-907	c-908	c-909	c-910	c-911	c-912	c-913	c-914	c-915	c-916	c-917	c-918	c-919	c-920	c-921	c-922	c-923	c-924	c-925	c-926	c-927	c-928	c-929	c-930	c-931	c-932	c-933	c-934	c-935	c-936	c-937	c-938	c-939	c-940	c-941	c-942	c-943	c-944	c-945	c-946	c-947	c-948	c-949	c-950	c-951	c-952	c-953	c-954	c-955	c-956	c-957	c-958	c-959	c-960	c-961	c-962	c-963	c-964	c-965	c-966	c-967	c-968	c-969	c-970	c-971	c-972	c-973	c-974	c-975	c-976



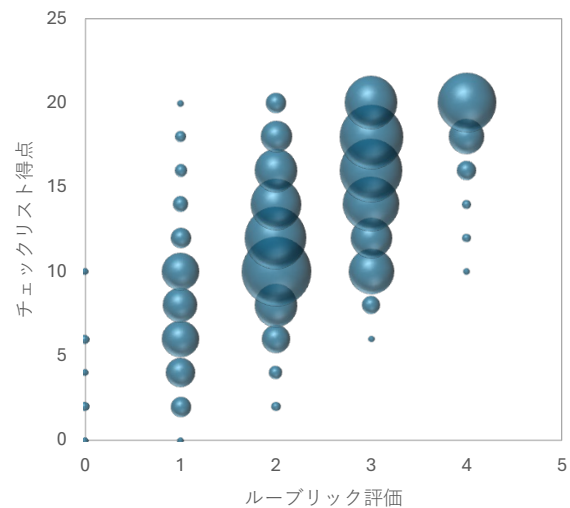
(a) 情報リテラシー(n=1432, チェック項目数=13)



(c) 数量的スキル(n=512, チェック項目数=13)



(b) 文章コミュニケーション(n=1317, チェック項目数=35)



(d) 実行力(n=1372, チェック項目数=11)

図6 ルーブリック評価結果とチェックリスト合計得点の関係

している。図5(a)~(d)を概観すると、偏相関係数は、特定のチェック項目間で突出して高いものは見られず、全体的に弱い相関を表す係数であることが分かる。一方で、網掛けされている位置は、特定の箇所に固まる傾向が見られている。

数量的スキルのチェックリストは、「数学的表現」、「グラフ」、「表」に関するチェック項目に分類されている。図5(c) 数量的スキルに着目すると、同一分類に含まれるチェック項目間に、網掛けされた位置が固まる傾向が見られている。この傾向は、情報リテラシー、実行力においても見受けられる。一方、文章コミュニケーションにおいては、網掛け位置にある程度の塊はみられるが、分類ごとの顕著な傾向とはなっていない。以上の結果から、チェック項目評価間の推定可能性に関して、ある1つのチェック項目評価結果から別のチェック項目評価結果を推定することは難し

いと考えられる。個々の偏相関係数は大きな値ではないが、分類ごとのチェック項目評価結果間で多変量的に推計をすることで、一部のチェック項目の結果から、他のチェック項目結果を推定できる可能性があることが確認できた。ただし、この分類は、評価指標作成者が主観に基づいて決められたものであり、これらの分類がそのまま使用できるとは限らず、分析結果を踏まえて、見直しの必要がある。

6.2 ルーブリック評価結果とチェックリスト評価間に関する検討

図6は、ルーブリックの評価結果とチェックリスト合計得点の関係を示したバブルチャートである。それぞれ横軸がルーブリックを用いた対象となる能力の総合評価で、0~4の値をとる。縦軸は対象となる能力のチェックリストの各項目の評価結果(0~2)の値を単純に合計した値である。チェック項目の数が能力によって異なるので、図6の縦軸

も能力によって異なる。バブルはその点に該当する人数に応じた大きさである。図6のnは各能力のデータ数を表す。評価を行う科目数は能力ごとに異なり、科目によっては事前事後で評価しているケースもあるので、能力によって行われた自己評価の回数が異なる。

図6を見ると、どの能力も概ねチェックリストの合計点がより高い方がルーブリックによる評価も高い傾向があることがわかる。一方で、チェックリストの合計点が高いにも関わらず、ルーブリックによる評価が低い学習者や逆にチェックリストの合計点が低いにも関わらず、ルーブリックによる評価が高い学習者が存在することもわかる。このように、学習者は適切な自己評価ができていない可能性があるため、自己評価後に助言を与えて、再評価を促すことが考えられる。このような観点から、4.1.1節で述べた「チェックリストの評価結果とルーブリックの評価結果(総合評価)の間の整合性」に基づく助言である。今後、具体的に、各データに対して整合性の有無を切り分ける方法を検討したい。

チェックリストの評価結果とルーブリックの評価結果の整合性の有無を判定し、整合性があると判定されたデータのみを用いると、チェックリストの合計点がより高い方がルーブリックによる評価も高い傾向がより明確になると考えられる。今後、各チェック項目と総合評価との関係も調査しつつ、チェックリストによる評価結果からルーブリックによる評価結果を推定する方法について検討したい。

7. おわりに

本稿では、ルーブリックやチェックリストを用いた間主観的な自己評価における適応型自己評価に関する研究へのアプローチと初期の検討結果について述べた。具体的には、本研究における適応型自己評価を定義し、自己評価の特性から自己評価者への助言と合わせて検討する必要性を示した上で、その導入方針と処理プロセスを明確にした。また、帝京大学情報電子工学科の汎用的能力育成の取組を対象として、研究で用いるデータ、研究課題、研究方法と体制を示し、現在までの検討状況について述べた。

今後は、研究課題に挙げた自己評価結果の整合性に基づく助言と適応型自己評価のための推定課題について、具体的に検討を進める予定である。特に推定課題については、自己評価結果の整合性に基づいて対象データを絞り込み、機械学習などの方法によって推定を試みたい。

また、現時点では、最終的には仮名データによって検証を行うため、一旦、5.2節での述べた方法で作成した合成データを用いる予定である。ただし、より妥当な合成データを作成できれば、より効率的に研究を進められると考えら

れるので、合成データの妥当性の検証方法と合成データ作成方法の改善についても検討をしたい。

謝辞

本研究は JSPS 科研費 24K06217、および帝京大学先端総研 インキュベーション助成金 23-98 の助成を受けている。

参考文献

- [1] Schulz, B. : The Importance of Soft Skills: Education beyond academic knowledge, *Journal of Language and Communication*, 2(1), pp.146-154 (2008).
- [2] Lamri, J. and Lubart, T. : Reconciling Hard Skills and Soft Skills in a Common Framework: The Generic Skills Component Approach, *Journal of Intelligence* 11(6) 107 (2023).
- [3] JABEE 日本技術者教育認定機構. <https://jabee.org>, (参照 2024-05-12)
- [4] 渡辺博芳, 荒井正之, 佐々木 茂, 盛 拓生, 古川文人, 水谷 晃三, 眞坂美江子, 塩野目剛亮, 高井久美子, 有本泰子 : 汎用的能力評価のためのルーブリックとチェックリストの提案, *情報教育シンポジウム SSS2019 講演論文集*, pp.30-37 (2019).
- [5] Miyazaki, M., Watanabe, H., Masaka, M. and Takai, K.: Developing a Generic Skill Assessment System Using Rubric and Checklists, *Proc. 29th International Conference on Computers in Education (ICCE2021)*, pp.191-200 (2021).
- [6] 渡辺博芳, 宮崎 誠, 眞坂美江子, 高井久美子 : ルーブリックとチェックリストを用いた汎用的能力のための自己評価システムの実用に基づく評価, *情報処理学会第 86 回全国大会*, 7F-01 (2024).
- [7] Marinagi, C. : Web-based adaptive self-assessment in Higher Education, *Education in a technological world: communicating current and emerging research and technological efforts*. pp.607-614 (2014).
- [8] Economides, A. A. and Moridis, C. N. : Adaptive self-assessment trying to reduce fear., *Proceedings of the First International Conference on Advances in Computer-Human Interaction* , pp. 158-163 (2008).
- [9] Hashim, H., Salam, S., Mohamad, S. N. M., and Sazali, N. S. S. : The designing of adaptive self-assessment activities in second language learning using massive open online courses (MOOCs). *International Journal of Advanced Computer Science and Applications*, 9(9), pp.276-282 (2018).
- [10] 奥村太一, 森 慶輔, 宮下敏恵, 西村昭徳 : 教師用コンピュータ適応型ストレス自己評価システムの構築に関する予備的研究, *上越教育大学研究紀要*, Vol.36, No.1, pp.41-51 (2016).
- [11] 岡田莉奈他, 統計値を用いたプライバシー保護擬似データ生成手法, *コンピュータセキュリティシンポジウム*, 2017
- [12] Jim Young et al., Using Bayesian Networks to Create Synthetic Data, *Journal of Official Statistics*, Vol. 25, No. 4, 2009, pp. 549-567
- [13] 統計分析ソフト HAD <https://norimune.net/had> (参照 2024-05-25)
- [14] 眞坂美江子, 渡辺博芳, 宮崎 誠 : ルーブリックとチェックリストを用いた間主観的評価における適応型自己評価のための問題定義, *情報処理学会第 86 回全国大会*, 7F-02 (2024).