

特集：多様な端末と大規模学習データが拓く新たな学習支援環境

自己評価と相互評価の差異を可視化するための 携帯端末用ルーブリック評価ツールの開発と試用

高橋 暁子*, 金西 計英*, 松浦 健二*, 吉田 博*, 和田 卓人**

Development and Practice of a Mobile Rubric Assessment for Visualizing Contrast between Self-Review and Peer-Review

Akiko TAKAHASHI*, Kazuhide KANENISHI*, Kenji MATSUURA*,
Hiroshi YOSHIDA*, Takuto WADA**

1. はじめに

グループワークなどの学習者中心の学習活動において、“自己評価”や“相互評価”，そして“自己評価と相互評価の併用”を組み込んだ教育実践が多数報告され，学習者の内省を促すことや学習動機の向上が期待されている⁽¹⁾．さらに，自己評価・相互評価活動では ICT の利用も増えている．たとえば，高等教育における 100 人程度の集合教育において，ICT を用いた多段階相互評価により，成果物の質的向上などが確認されている⁽²⁾．また，ICT を用いた継続的な相互評価によって評価能力が向上することも示唆されている⁽³⁾．このように，自己評価および相互評価活動を継続することで，目指していた能力や評価能力（メタ認知）が獲得される可能性がある一方，自己評価・相互評価に不慣れた学習者が評価活動の第一歩に踏み出しやすくする工夫も重要であろう．加えて，ICT の利用時には，固定的な PC に比して利用環境を選ばない携帯端末の利便性は論を待たない．そこで，自己評価と相互評価の差異を，携帯端末を用いて手軽に認識することから始めることが有効ではないかと考えた．

以上から，大学での学習者による主体的な活動を中心とした授業において，その後の深い学びの“きっか

け”を与えることをねらい，自己評価と相互評価の差異を可視化する携帯端末用ルーブリック評価ツールを開発した．本稿では第一段階として，動作確認・操作性検証とともに，ツールによって学習者が自己評価と相互評価の差異を認識しているかを確認した．

2. 設計と開発

2.1 基本設計

本ツールは，スマートフォンのブラウザソフトで動作する Web アプリケーションとして設計した．大学生のほとんどがスマートフォンを所持し，操作に慣れていることから，スマートフォンを用いることで，手軽な自己評価・相互評価活動が可能と考えた．

また，携帯端末型の汎用的な評価ツールとして，クリッカーを代表とするレスポンスアナライザーがある．レスポンスアナライザーは，1 人の教員がクラス全体の学生の反応を把握する目的で使用することが一般的である．一方，本ツールは 1 対多ではなく，学生 5 名程度の小グループ内で行う自己評価／相互評価活動の支援を指向した．

本ツールの対象者は大学生全般，特に 1 年生を想定し，評価手法としてルーブリック評価を採用するこ

*徳島大学 (Tokushima University)

**タワーズ・クエスト株式会社 (Towers Quest Inc.)

受付日：2015 年 6 月 8 日；再受付日：2015 年 9 月 5 日；採録日：2015 年 10 月 27 日

とした。これは、大学1年生は自己評価・相互評価に不慣れで、評価の観点をあらかじめ教員が設定したほうが評価しやすいと考えたためである。加えて、ルーブリック評価はパフォーマンス評価においてスタンダードな手法で、本ツールが想定する学習者による主体的な活動の成果を評価する手法として妥当と考えた。

2.2 開発

2.2.1 学習者向け機能

(1) 自己評価機能

学習者はスマートフォンのブラウザを用いて本ツールへアクセスする。ログイン後、“自己評価”を選択すると、自己評価画面が表示される(図1)。自己評価画面では、ルーブリックの観点の数だけ、スライダバーとコメント入力欄が表示される。スライダバーを左右にスライドすることで尺度が変更され、スライダバーの下に尺度の説明文が表示される。また、同時に上部のレーダーチャートの値も変更される。コメント入力欄には、「なぜそう思うか、理由を入力」という指示文を表示し、評価の根拠を考えさせることを意図した。最下部には、総評コメント欄が表示される。総評コメント欄には「総評を入力。上記にはない観点、改善案も歓迎します」という指示文を表示することで、学生自身が考えた観点による振り返りを促すことをねらった。

(2) 相互評価機能

ログイン後、所属グループのメンバーの名前の一覧からメンバーの名前を選択すると相互評価画面が表示される。相互評価画面の操作は自己評価画面(図1)と同一である。

(3) 総合評価機能

自己評価を確定すると、ログイン後の画面には“総合評価”が表示されるようになる。総合評価画面(図2)には、自己評価の値と、他者からの評価の平均値がレーダーチャートとなって表示される。これにより、自分と他者の評価の違いを一目で把握できると考えられる。また、レーダーチャートの下には、観別に評価値とそのコメント(匿名)が表示される。

2.2.2 教員向け機能

教員はPCのブラウザソフトを用いて教員用ページ

にログインする。教員向けの機能としては、学生登録機能、ルーブリック評価表作成機能、評価結果一覧機能を実装した。主要機能はルーブリック評価表作成機能である。一つの授業で複数のルーブリック表を作成できる一方、他教員・他授業のルーブリック表の参照も可能とした。

2.2.3 開発・動作環境

開発は、学習者向けにはJavaScript、教員向けはRuby on Railsを用いた。データベースはMySQLである。端末環境は、2014年10月時点で一般に普及しているスマートフォン上のブラウザソフトを想定した。

3. 評価

3.1 目的

ツールが授業内のある時点での一斉利用に耐えうるか(動作確認)、利用者にとって使いやすいか(操作性)、利用者が自己評価と相互評価の差異を認識しているか(有用性)の確認を目的として実際の授業で試用した。

3.2 方法

サーバのログとアンケートの操作に関する設問から、本ツールの動作と操作性を確認する。また、アンケートに設けた「役に立つか」という設問と自由記述を踏まえて本ツールの有用性についても検討する。

ツールの評価にあたり、T大学の教養科目「情報メディアと教育」(履修者104名)の協力を得た。本科目は1グループ5人前後で構成されるグループ活動を主体とした授業を展開している。最終回の授業において、これまでのグループ活動について「協働力」(表1)の観点で振り返ってもらい、自己評価および相互評価を行う活動を約20分設けた。学生全員には評価手順書(ツールのURLを記載)、ルーブリック表、アンケートを配布し、“スマートフォン”と“紙”のどちらか好きなほうを選択して評価活動をしてほしい旨を伝えた。スマートフォンを選択した学生には、自身のスマートフォンのブラウザにURLを打ち込んでツールを利用した評価活動を、紙を選択した学生には配布したルーブリック表を用いた評価活動を実施して

①基本的なルール

②話す・聞く

③アイデアの発想と収束

尺度 (3 段階)

- ・ (C) 努力しましょう
- ・ (B) まあいい感じですよ
- ・ (A) すばらしい

①基本的なルール

「ゴールの共有」、「時間の管理」、「役割の自覚」のすべてを守っている。

なぜそう思うか、理由を入力

②話す・聞く

要点を明確にして話し、確認・要約をしなが
ら聞いている場面がある。

なぜそう思うか、理由を入力

図 1 自己評価画面

①基本的なルール

②話す・聞く

③アイデアの発想と収束

凡例

- ・ 黄色: 自己評価
- ・ 青色: 他者からの評価の平均

①基本的なルール

自己評価

(B) まあいい感じです - 「ゴールの共有」、「時間の管理」、「役割の自覚」のうち、1個または2個を守っている。

コメント: レジュメや、発表資料作りに積極的に参加できたから。

他者評価

(A) すばらしい - 「ゴールの共有」、「時間の管理」、「役割の自覚」のすべてを守っている。

図 2 総合評価画面

表1 協働力のルーブリック

		尺度		
		(A)すばらしい	(B)まあいい感じです	(C)努力しましょう
観点	①基本的なルール	「ゴールの共有」、「時間の管理」、「役割の自覚」のすべてを守っている。	「ゴールの共有」、「時間の管理」、「役割の自覚」のうち、1個または2個を守っている。	「ゴールの共有」、「時間の管理」、「役割の自覚」のすべてが守られていない。
	②話す・聞く	常に要点を明確にして話し、確認・要約をしながら聞いている。	要点を明確にして話し、確認・要約をしながら聞いている場面がある。	要点を明確にして話すこと、確認・要約をしながら聞くことを全くしていない。
	③アイディアの発想と収束	アイディアの発想・共有と、収束・統合の二段階を明確に分けて課題に取り組んでいる。	アイディアの発想・共有と、収束・統合の二段階を分けて課題に取り組もうとしているが、区別が曖昧になっている。	アイディアの発想・共有と、収束・統合の二段階を全く意識していない。

表2 アンケートの回答結果 (5件法)

操作性	このツールの操作は簡単でしたか？(N=74)	平均 3.85 (SD=1.15)
	画面上の指示や説明はわかりやすかったですか？(N=75)	平均 3.8 (SD=1.12)
有用性	総合評価画面のグラフやコメントの表示はわかりやすかったですか？(N=75)	平均 3.92 (SD=1.00)
	総合的に考えて、本ツールは自己評価や相互評価の役に立ちそうですか？(N=75)	平均 3.64 (SD=0.94)

もらった。自己評価とグループメンバー全員に対する評価が終わった学生からアンケートを記述・提出してもらった。

3.3 結果

3.3.1 ログ

本評価時の授業に出席した学生はおおよそ 95 名であった(目視による集計のため正確には不明)。サーバのログによると、本ツールを利用した学生は 80 名で、1 人あたりの利用時間の平均は 9 分 29 秒であった。サーバのレスポンスタイムは、アクセス全体の 99.994%が 100ms 以内であった。

3.3.2 アンケート

ツール利用者 80 名中、アンケート回答者は 75 名であった。用いた端末に対する回答者は 74 名で、そのうち 46 名が iPhone、28 名が Android 系端末であった。通信環境は、設問を設けていないので不明である。

表 2 に示すように、ツール利用者を対象とした「このツールの操作は簡単でしたか？」といった操作に関する二つの設問(5 件法; 5 が最も良い)に対する回答は、すべて平均 3.8 以上となった。

また、有用性の設問に関しては、「総合評価画面のグラフやコメントの表示はわかりやすかったですか？」が平均 3.92、「総合的に考えて、本ツールは自己評価や相互評価の役に立ちそうですか？」が平均 3.64 であった。「役に立った」(4 または 5)と回答した理由の自由記述は 25 件あり、最も多かったのは「全体的に使い勝手が良かったです」といった操作の容易さに関する意見(10 件)であった。一方で「役に立たない」(1 または 2)と回答した学生は 8 名おり、そのうち 5 名が理由を記述し、うち 3 名が「班員の評価をしてどうするんですか？」など、評価活動そのものに対する否定的な意見であった。

良いと思う点の自由記述は 47 名分あり、1 人で複数の意見を述べているデータに関しては文章で分割したところ 50 件のコメントが得られた。「グラフ化されてわかりやすい」という意見が最も多く(14 件)、次いで「気軽にできる」などの手軽さを指摘するコメントが 11 件あった。さらに、「他者からどのような評価をされているのか、コメントもまじえて知ることができる」といった他者からの評価が有効だというコメントも 9 件と多かった。特徴的なコメントとして、グラフに言及していないが「自分が考えている自分と

ほかの人が考えている自分を比べることができて参考になる」という他者との差異を認識することを有益と考えているコメントが2件あった。また、「自分がどのようにこの授業に取り組んだのか振り返ることができる」というリフレクションを示唆するコメントも3件あった。

一方で改善点の自由記述で最も多かったのは、ツールそのものではなくブラウザにURLを打ち込むことで、「QRコードにしてほしい」という意見が12件あった。次いで「五段階くらいあったほうがいい」という評価尺度に関する意見が5件あった。

3.4 考察

3.4.1 動作

ほぼ同時に80名が利用できたことから、一般的な携帯端末を用いて、本ツールは問題なく動作するとと言える。

3.4.2 操作性

アンケートの操作性に関する不達の設定がすべて平均3.8以上で、役に立ったと考える理由の自由記述で最も多かったのが「操作の容易さ」であることなどから、本ツールの操作性に大きな問題はなく、手軽な評価活動が実現できたと考える。

しかしながら、ログインに至るまでを配慮しないと使い勝手が悪いという印象を与え、ツールの利用に至らない可能性があることがわかった。本質ではないが、QRコードを付記した紙を配布するなどの必要がある。

3.4.3 有用性

アンケートの「総合評価画面のグラフやコメントの表示はわかりやすかったですか？」の平均が3.92で、良いと思う点の自由記述で最も多かったのがグラフによる評価の視覚化（14件）であり、他者からの評価を有効と捉えるコメントが9件、差異の認識を有益と考えているコメントが2件あったことから、グラフや相互コメントの表示によって、「自己評価と相互評価の差異を認識すること」ができていと考えられる。

3.4.4 まとめと今後の課題

今回試用したルーブリックには「役割の自覚」など自己評価を念頭に置いた内容も含まれており、表出さ

れにくい学習者個人の意識レベルを他者が評価することは困難であることから、相互評価における他者からの評価結果はあまり信頼できるものではない。しかし、「動作確認」、「操作性」、「有用性」（自己評価と相互評価の差異の認識）という本論文の評価目的には、他者評価の信頼性は直接的に関係しないため、今回の評価の目的は達成できたと考える。

一方で、リフレクションを示唆するコメントが3件あったものの、差異の認識や他者からのコメントが深い学びのきっかけになっているかは今後明らかにすべき課題である。今後、本ツールを用いた授業実践を通じて、入力内容や授業の成果物の質などを分析し、ツールを用いることの妥当性や、ツールの利用が評価能力（メタ認知）に与える影響を検証したい。

また、「班員の評価をしてどうするんですか」などの評価活動そのものに対する抵抗感や理解不足を示唆するコメントがあることに留意したい。評価活動そのものの指導も含めて、本ツールを有効に活用する授業方法を検討していく必要がある。その一方で、評価力不足の学生がいることを前提としたツールの改訂も必要であろう。たとえば1人あたりの利用時間平均9分29秒の中で、自己評価と4～5名の相互評価および評価結果の確認を行っていることから、自己評価および他者1人分の評価は1～2分で済ませていると推察できる。加えて「五段階くらいあったほうがいい」というコメントから、短時間評価の背景には、スライダバーをスライドすることで尺度を変更し、評価基準の記述内容をよく読まずに安易な数値評価にとどまっている学習者がいる可能性がある。この場合、評価尺度の入力インターフェースをスライダバーから選択肢を読ませてタップする形に変更することが考えられる。

4. おわりに

本研究では、学習者自身の自己評価と学習者間の相互評価活動を支援するツールを開発した。評価の結果、動作や操作性に大きな問題はなく、学習者はレーダーチャートなどによって自己評価と相互評価の差異を認識していることが示唆された。今後、本ツールを用いた授業実践を通じてツールの妥当性やメタ認知に与える影響を検証したい。

参 考 文 献

- (1) 藤原康宏, 加藤 浩: “4.2 相互評価”, 永岡慶三, 植野真臣, 山内佑平 (編著) “教育工学における学習評価”, ミネルヴァ書房, 東京, pp. 143-158 (2012)
- (2) 布施 泉, 岡部成玄: “多段階相互評価法による学習の実践と効果”, 日本教育工学会論文誌, Vol. 33, No. 3, pp. 287-298 (2010)
- (3) 藤原康宏, 大西 仁, 加藤 浩: “継続的な学習者間評価を導入した情報教育の実践”, 情報処理学会論文誌, Vol. 49, No. 10, pp. 3428-3438 (2008)