

# 多肢選択式 Web テストによる 学習促進に関する考察

高林 友美<sup>1</sup>, 安間 文彦<sup>2</sup>

本稿では、現在でも多くのオンラインベースの教育で活用されている多肢選択式の Web テストについて、学習活動の教材としての側面を中心に議論する。多肢選択式の Web テストは、能力測定のためだけでなく、理解定着を意図した学習活動としても使われている。より学習活動を活性化するには、複数の知識を組み合わせた問題や知識を現実課題に応用する問題のような課題内容の設定と同時に、複数解答式問題や誤答選択問題といった手法も考慮することが望ましい。また、参考になる指標については、これまで積み重ねられているテストの標準が参考になる。こうした工夫の実例として、フルオンラインの高等教育機関で利用した場合について検討したところ、平均受験回数の増加などから、一部において学生の学習活動に対するエンゲージメントが高まった可能性が指摘された。

## 1. 研究の背景

e ラーニングの教材には、多肢選択式 Web テストが付き物であると言っても過言ではない。思考力・判断力・表現力を測るための記述式解答の自動採点や、学習のプロセスを記録した e ポートフォリオ評価の研究や実践が進んでいるが、こうした新しい手法が e ラーニングにおいて全面的に採用されることは少なく、複数の学習内容に対応し、正確な即時フィードバックが可能である多肢選択式 Web テストが使われることが一般的である。

しかし、多肢選択式のテストについては、当て推量の解答などによる不正確な能力評価になってしまうことへの議論が絶えない<sup>1)</sup>。また、実際に授業を運営した経験から、実際に文章を書くのではない、選択するだけのテストで能力を測定することに対して疑問を持つ教員も多いのではないだろうか。こうした議論に対して、どのようにすれば正確に能力を測定するテストが作成できるのかという、テストの標準を定める取組など多くの研究が積み重ねられており、現在では細目積み上げ式の客観テストによる信頼性と妥当性を担保する方法が確立されてきている<sup>2)</sup>。

では、いわゆるテストスタンダードに沿った問題であれば、オンラインベースの教育活

---

<sup>1</sup> サイバー大学 IT 総合学部・助教

<sup>2</sup> サイバー大学 IT 総合学部・教授

動において「良いテスト」と言えるのだろうか。学習單元ごとのまとめとして使用されることの多い多肢選択式 Web テストは、能力の正確な評価以外にも役割を負うことが期待されているからこそ、この形式でよいのだろうか、という疑問が出てくるのかもしれない。

本稿では、学習活動としての多肢選択式 Web テストのあり方に注目し、その改善について考察する。その際に参考になる指標について触れた上で、問いの形式だけでなく、テスト問題で問う内容を工夫することについて述べていく。多肢選択式では考えさせる問題ができないから、記述式回答によって学生の深い学びを促す、という考え方から、多肢選択式という eラーニングで効率的に活用できる形式を選んだ上で、どのように学びを促進できるのかを考えたい。

## 2. 学習活動としてのテスト

教育の現場で扱われるテストには、複数の目的がある。外国語教育における評価を専門とする渡部<sup>3)</sup>は「テストを受けている最中にも学習者は学んでいる (p. 2)」ことを強調し、テストを単に能力測定のみにするのではなく、学習活動として利用することを提案している。渡部は多肢選択式のテストについて教育実践例を挙げ、正答選択問題において誤った英文を複数読ませるよりも、誤答選択問題として少しでも多く正しい英文に触れられるようにすべきだと述べている。

また、テストに関連して、教育工学の視点から学習評価を論じまとめた永岡ら<sup>4)</sup>は、新しい時代に求められる学習評価の課題として、2012年時点で以下の6点を挙げている。

1. 学習と一体化した評価
2. 現実的な環境・課題での評価
3. 社会的な文脈での評価
4. ICT活用学習で蓄積された大量データの解析
5. 質的な評価の重視
6. 評価観点の多様化 (永岡ら、p. 2)

このうち、最初に述べられている「学習と一体化した評価」は、現在でも強く求められる課題であると考えられる。テストを行う目的として、その活動自体が学びとなるようにすることは、学習評価に関する議論からもその必要性が注目され続けていることが分かる。

学習という一連のプロセスにおいて、できるだけ実体験に近い形で知識・スキルを応用すること、練習とフィードバックの機会を与えることの重要性は既に古くから広く知られている。多肢選択式 Web テストをこうした機会の一つとし、オンライン教育における学習活動であると捉えるならば、正確に成果を測定するためだけでなく、学習内容の外化と内化を繰り返す深い学び<sup>5)</sup>に結びつくような Web テストの導入意図および期待される役割が見えてくるだろう。

ただし、ドリルのような学習活動として活用されるのであっても、テストである以上能

力評価の役割を忘れることはできない。不正確な問題や、慣れた学生だけが持つテストスキル<sup>6)</sup>によって正解が導いてしまうような問題では、評価も学習活動も意図したものにならない。では、講義等で学んだことを実際に学生自身の手を動かしながら活用することを目的とした Web テストには、具体的にどのような工夫が求められるのだろうか。

### 3. テスト改善のための指標

学習活動のためのテスト改善を行う工夫において、重要になるのが改善のための指標である。改善すべき設問を見つけたり、改善されたのかどうかを比較したりするためには、これまでテストに関して開発され活用されてきた指標を確認しておくことが必要だろう。また、単独の指標のみを参照した結果、測定・評価もしくは学習の目的としている内容ではなくなってしまう場合もある。テストスキルがあれば容易に解ける問題ばかりになること、またはひっかけ問題など内容理解以外の面での認知的負荷を高めるものが増えてしまう危険もある。一つの問題・一つの指標にとらわれすぎずに、問題全体を確認しながら、複数の指標を確認して改善を進める必要があるだろう。

正確な能力を測定するためのテストとしては、正答率や平均点、学生の分布などが参考になる。選択肢に特化すると、全く選ばれていない、能力の見分けに役立たない選択肢がないかどうかを各選択率によって確認することも必要だろう。そのほか、テスト全体との相関を元にして、その問題がどの程度、能力の見分けに貢献しているのかを確認することも可能である（識別力<sup>7)</sup>）。このような指標についてはテスト標準に詳しいが、学習活動のために使われる e ラーニングやその他授業で使われるテストの改善を行う上では、こうした指標が意外で膝を打つような発見をもたらすことは少なく、むしろ教員が既に予感していたような欠点や、経験則による難度調整とほぼ同様の傾向を改めて提示するような場合が多いだろう。ただし、教員の肌感覚を裏付けして確かめ、エビデンスに基づく改善を行うという意味で、こうした指標を確認することが重要であると考えられる。

学習活動としてのテストのためには、それに加えてどれほど学生の関与が高まったのかを確認する指標が必要である。典型的な LMS では、受験の回数や受験にかけた時間が記録されている。どちらも単独では必ずしも深い学びを意味しないが、ある程度のクラスサイズの集団の状況を比較するためには十分な指標だと考えられる。個別に記録し、どの問題に時間がかかったか、どの選択肢を触って迷いが見られたのかを分析することも可能ではあるが、膨大すぎるデータを用いるよりは、回答の総時間から想定する方が現実的である。質的研究手法を活用するならば、テストを受験した学生にどの問題が考えさせられたか、どのような過程を経て問題を解いていったかについてインタビューしたり、ジャーナル記録を回収したりすることで得られるコメントも指標として有効であろう。

#### 4. 授業における実践例

本稿では学習活動としての多肢選択式 Web テストの例として、フルオンラインの高等教育機関である本学の1授業で使用されたものを取り上げたい。本学において Web テストは、インターネット上で完結する授業における質保証の取組のために活用されるものの一つであり、見直しや繰り返し視聴が可能である反面、受動的になりがちだと言われる講義動画の視聴に組み合わせることで、能動的活動を担保し、かつ理解度評価と出席確認を行うという役割を持つ<sup>8)</sup>。授業では必ず課題（レポート、ディベート・ディスカッション、または Web テスト）を設定することがガイドライン化されており、上記の役割を持つことから、課題として Web テストが多く使用されている。Web テストは基本的に8問構成（8点満点）で、最大5回まで受験可能である。ただし、受験後に表示されるのは総得点のみで、解答解説を閲覧した後は受験回数5回未満でも再受験することができない。なお、本事例の授業は専門応用講義科目（3年次以上配当に相当）であり、第1回～第15回全てに Web テストが設置されている。成績は期末試験だけでなく中間レポートと各回の Web テストの結果を総合して評価される。

この授業は毎学期開講されているが、2017年度後期に問題改訂を行い、以降は2019年度まで軽微な修正を続けながら使用されている。問題改訂では、出席確認の理解度評価が主であったところから、学生がより考えるように具体的な工夫として、1)複数回答問題を毎回1, 2問含めることで、テストスキルによる当て推量を減らした。また2)単純な語句選択ではなく、複数の知識を組み合わせることで解く正誤判定問題を増やした。さらに、3)単純な知識再認ではなく、設定された場面において知識の応用を行う問題も少数ながら加えた。図1は改善された問題の例であり、上述の1)～3)の全てを含んでいる。

**問7**

次のような2つのリレーションスキーマ 学生と学部を考えるとする。  
 リレーションスキーマ：  
 学生(学籍番号, 学部番号, 氏名, 電話番号)  
 学部(学部番号, 学部名)

この2つのリレーションスキーマの主キー、候補キー、スーパーキーについて述べた次の文章のうち、**誤っているものをすべて選べ。**  
 (※誤った選択肢をひとつでも選ぶと得点は入りません)

リレーションスキーマ学生の主キーとして{学籍番号}を選ぶ

リレーションスキーマ学生の候補キーは{学籍番号}、{電話番号}である

リレーションスキーマ学部の主キーとして{学部名}を選ぶ

{学籍番号, 電話番号}や{学籍番号, 氏名}はリレーションスキーマ学生のスーパーキーであるが、{氏名, 学部番号}はスーパーキーではない

図1 改訂テスト問題例

## 多肢選択式 Web テストによる学習促進に関する考察

以上の工夫を加えることで実際に学習活動が活発化されたかについて、2017 年度前期（第 1 回出席者 68 名）と、2019 年度後期（第 1 回出席者 77 名）の回答ログを比較した。ガイダンスを含む第 1 回を除き、授業全体から見ると基礎的な内容である第 2 回から第 4 回で比べた結果、第 3・4 回の改訂版の Web テストにおいて平均受験回数が増加していた（表 1）。

表 1 各回 Web テスト平均受験回数の比較

	2017 回数	2019 回数	<i>t</i>	<i>df</i>	<i>p</i>
第 2 回	1.62	1.82	1.11	140	n. s.
第 3 回	1.51	2.43	4.25	138	<.01
第 4 回	1.54	2.56	4.67	134	<.01

単元自体の難易度は変化していない状態だが、単純な知識再認テストではなくなったことで、学生が迷ったり振り返ったりしながらテストを受験したと考えられる。ただし、複数回受験する学生もいれば、得点が低くても 1 度しか受験しない学生もいることから、本実践例の教育的効果をより詳しく見るためには、受講生全体の平均値の比較だけでなく、層を仮定した上で統計分析が必要となる。次節において本稿全体をまとめ、この点を含めた課題と展望を議論する。

## 5. 課題と展望

本稿ではオンライン教育で多く用いられている多肢選択式 Web テストの受験過程が学びの機会となりうることについて議論した。具体的に学習活動としての Web テストの例を試行的に分析すると、一部の学生に限られる可能性はあるものの、多肢選択式のままであっても、より手間をかけてテストに取り組んだ可能性が見えてきた。

具体例の分析については、受験回数という指標のみでは学生の活動を捕捉しきれない部分があるため、引き続き検討が必要である。また、科目レベルの異なる授業における学生の回答ログやその他のデータも分析していく必要がある。

また、当て推量と似たテストの正確な能力評価の阻害要因として、受動的かつ消極的なテスト受験が考えられる。4 節で紹介した、1 度目の挑戦で点数が低かったにもかかわらず再チャレンジをしない集団には、こうしたテスト受験の姿勢の者も見られるかもしれない。テストの内容や回答方式に工夫が加えられていたとしても、学生が主体的に課題に取り組むようなコース設計・プログラム設計になっていなければ、学生の思考力を深くすることは難しい。ミクロな学習活動としてのテストを充実させ、深い関与や学生の思考を推し進める作問を考え続けるだけでなく、マクロデザインについても考慮していく必要があるだろう。

注および参考文献

- 1) 小柏香穂理・浜本義彦・王躍・刈谷丈治・小河原加久治「形成的評価を支援する可視化ツールの開発：Moodleの小テストモジュールを活用して」『情報処理学会研究報告』第22号、2013、pp. 1-4.
- 2) 日本テスト学会『見直そう、テストを支える基本の技術と教育』金子書房、2010.
- 3) 渡部良典「学習者を育てるためにテストを使う」『テストが到達目標と指導に与える影響セミナーレポート』ブリティッシュ・カウンシル、2018、pp. 2-9.
- 4) 永岡慶三・植野真臣・山内祐平『教育工学における学習評価』ミネルヴァ書房、2012.
- 5) 松下佳代『ディープ・アクティブラーニング』勁草書房、2015.
- 6) 問われている内容についての知識理解ではなく、テスト問題を解くための方略に則って正答を導くことが出来る能力を示す。テストの適応の議論を参照されたい。
- 7) よく知られる手法としては、項目の正解/不正解をダミー変数として、テスト総得点との相関を計測する I-T 相関がある。その問題 (Item) を正解することと、テスト全体 (Total) の点数が高いことにある程度の相関が見られれば、その問題が測りたい能力を見分ける力が高いことが示される。こうした項目分析は現在の大規模テストでは項目反応理論 (Item-Response Theory) で整理されており、本質的には eラーニングとの親和性が高いはずだろう。
- 8) 小野邦彦『『メディアを利用する授業』の質保証の取組み』『eラーニング研究』第1号、2010、pp. 29-35.

本稿は、2020年2月29日に日本教育工学会2020年春季全国大会にて著者が発表を行った「主体的な学びを促進する多肢選択式 Web テストに向けて」の内容をまとめたものである。