

オンラインテストで出題される質問に関する予備的な検証

溝口 佳宏

帝京大学経済学部地域経済学科

概要

筆者は2018年度前期の担当科目の1つである「経済学概論」において、LMSのオンラインテスト機能を用いた試験を2回行った。また、それぞれの試験の前には練習問題をLMSのオンラインテスト機能を用いて(つまり、試験とほぼ同様の設定で)提供した。本稿の作成に際し、第1回試験(いわゆる中間試験に相当する試験)と、第1回試験用の練習問題で出題された質問について、LMSの項目分析機能を用いた検証を行った。本稿では、その結果について報告する。分析の結果として、当該科目のオンラインテストは、当該科目に関する知識を備えた学生と備えていない学生を適切に区別するテストとして機能しており、単位認定の手段として有効であるのが確かめられた。一方で、いわゆる2択の質問については、選択肢を増やす形での改訂が必要であることや、穴埋め(空欄補充)問題の問題文は、講義資料と酷似しないよう改訂することが必要であることも確かめられた。

1. はじめに

1.1 背景と目的

本稿では、筆者が担当している科目の1つである「経済学概論」において、LMSのオンラインテスト機能を用いて出題した第1回試験(中間試験にあたる試験)、この試験への対策として、(試験とほぼ同等の条件で)提供した練習問題の質について、若干の考察を試みる。

現在の学士課程教育では、学生が知識を獲得するだけでなく、知識の活用能力や生涯学び続ける力も培うことができるよう、カリキュラムを組むのが求められている。最近では知識の活用が重視されているものの、知識の獲得そのものが否定されている訳ではない。また、知識を活用するには、活用するための知識そのものを獲得しておく必要がある。それゆえ、各大学で実施されている学士課程教育のカリキュラムでは、知識獲得型科目が設置されている。

知識獲得型科目の履修を通じて学生に求められることは、その科目で取り上げられている知識を獲得することであるのは言うまでもない。そ

れゆえ、その科目を担当する教員は、履修している学生に対して、担当科目で取り上げる知識の獲得を促すような科目設計を行う必要がある。

現在の学士課程教育では単位制が採用されている。その点を踏まえると、知識獲得型科目を担当している教員にとって、その科目を履修している学生に知識の獲得を促す最も強力な働きかけ方は、言わずもがなではあるが「その科目で取り上げられている知識を獲得できている学生に、その科目の単位を与える」という方法である。そして、単位を与えるかどうか(つまり、その科目で取り上げられている知識を獲得できているかどうか)を担当教員が判断する際に、最も用いられる手段が試験である。

試験の形式については、いろいろな形式が存在するが、本稿で取り上げる試験形式は、LMSのオンラインテスト機能を活用する試験形式である。オンラインテストに関しては、いわゆる項目反応理論をベースとしたテスト理論があり、その理論を活用したオンラインテストの実践も多数行われている¹。そして、オンラインテストは、あらかじめサーバ

A preliminary verification on questions to be asked in online-testing

Yoshihiro Mizoguchi

Department of Regional Economics, Faculty of Economics, Teikyo University

¹ テスト理論に関して解説している文献としては、たとえば橋本[1]が挙げられる。また、実践例に関して解説している文献としては、たとえば赤倉[2]がある。

ーに蓄積されているテスト問題群(いわゆるテストバンク)からランダムに質問を選んで出題できる性質を持つことから、受験者の解答状況に応じて質問を出題する形式(いわゆる適応型テスト)を採用できる²。また、最近では、その機能を応用して、受講者の状況に応じて、あらかじめサーバーに蓄積されている教材から適切な教材を提示する学習支援(いわゆる適応型学習)も盛んになりつつある³。

LMS のオンラインテスト機能の活用を考える個々の大学教員にとって、最初に立ちはだかるハードルは、いわゆるテストバンクをあらかじめ作成する必要から生じる労力である。しかし、このハードルを乗り越えて、LMS のオンラインテスト機能の活用を始めると、当面の間は担当科目における試験に関連した労力は、実施した試験の採点作業に関わる労力を中心として大幅に削減される。それゆえ、LMS のオンラインテスト機能を活用している個々の大学教員にとって、試験に関わる労力の削減というメリットをできるだけ長く享受したいという気持ちは、ごく自然に生じるものと考えられる。

一方で、年月が経つにつれて、担当科目で取り扱う内容を多少なりとも改訂するのも、ごく自然と考えられる。その点を踏まえると、担当科目を履修している学生に対して、より適切な知識獲得を促すためには、いったん作成したテストバンクの改訂に、多少の労力が求められるにしても、取り組む必要があるだろう。

そこで本稿では、2018 年度前期に担当した「経済学概論」において、LMS のオンラインテスト機能を活用する形で出題した、第 1 回試験(中間試験にあたる試験)、この試験への対策として(第 1 回試験とほぼ同じ条件で)出題した練習問題の質について、LMS に装備されている項目分析機能を用いるのを通じて、若干の分析を行う。

1.2 研究の経緯と位置付け

本稿は、LMS を用いたオンラインテストに関する実践報告である。筆者は、帝京大学に着任した 2011 年度以降、担当科目の一部において、LMS を用いたオンラインテストを成績評価手段として使用するとともに、その実践報告を行ってきた。溝口[4]や溝口[5]では、LMS を用いたオンラインテストが主要な成績評価手段として、継続的に活用可能であることを示した。また、溝口[6]では、LMS におけるオンラインテスト機能が、授業外学習を促すことを通じて、知識獲得型科目で提供される知識の定着に有効であることを示した。

一方で、溝口[4]や溝口[5]は、いわゆる教場試験との比較を念頭におきながら、LMS のオンラインテスト機能を主要な成績評価手段として活用する可能性を論じたものである。オンラインテストで出題している問題の質については議論していない。

本稿の内容は、オンラインテストで出題している問題について、知識を獲得している学生と獲得していない学生を区別できるような問題になっているかという観点での評価を試みながら、オンラインテストの活用可能性について論じたものである。その意味で本稿は、LMS を用いたオンラインテストを主要な成績評価手段として利用することに関する分析を深化させるものと位置付けられる。

1.3 本稿の以降の構成

本稿の以降の構成は次の通りである。第 2 章では、本稿で取り上げる 2018 年度前期「経済学概論」において、どのような試験を実施したのかを記す。第 3 章では、LMS に備えられている項目分析機能について記す。第 4 章では項目分析機能を用いた結果について記す。第 5 章では、結果の考察を行う。第 6 章では結論と今後の課題を記す。

² 赤倉[2]でも言及されているが、典型例の 1 つは TOEFL iBT である。

³ この点に触れた文献としては、例えば田中[3]がある。

2. 実施した試験の概要

本章では、本稿で取り上げる 2018 年度前期「経済学概論」において実施した試験(オンラインテスト)の概要について記す。

「経済学概論」は地域経済学科の 1 年生に配当されている 2 単位科目である。また、地域経済学科に所属する学生にとっての必修科目の 1 つである。この科目では、ミクロ経済学の入門にあたる内容と、マクロ経済学の入門にあたる内容を取り扱う。単位認定は LMS のオンラインテスト機能を活用する形で実施される試験の得点でなされる。ミクロ経済学の入門が終了した時点で実施する第 1 回試験、第 15 回授業が終了したのちに実施する第 2 回試験のそれぞれで、100 点満点中 60 点以上を取得することが、単位を認定されるための条件である。

第 1 回試験は、2018 年 7 月 6 日の 13 時から 7 月 10 日の 18 時までの期間で実施した。この期間中に、履修学生は最大 4 回までオンラインテストを受験することができる。1 回の受験では、15 分以内に、いわゆるテストバンクからランダムに出題される 10 題に解答する必要がある。1 題の配点は 10 点である。テストバンクに収録されている質問は、空欄補充問題、正誤問題、計算問題である。空欄補充問題と計算問題については、複数の解答欄が設定されている質問がある。そのような質問については、正答した比率に応じて部分点が与えられる。単位を取得するには、先に記したように 100 点満点で 60 点以上を取得する必要がある。なお、オンラインテストを複数回受験した場合には、最高得点で判断がなされる。受験者に対するフィードバックは、答案提出の動作を行った後すぐに、正答した質問、誤答した質問、正答した質問について取得した得点、合計得点が示される形で行われる。

先に記したように、「経済学概論」は 1 年生に配当されている科目であり、必修科目の 1 つである。大学に入学するまでに、いわゆる Computer Based Test (CBT)を受験した経験がない学生も

多く存在すると思われる⁴。それゆえ、オンラインテストになれてもらう意味も込めて、第 1 回試験および第 2 回試験の前に、練習問題をオンラインテスト形式で提供している。2018 年度前期の「経済学概論」においては、第 1 回試験実施前の 2018 年 6 月 28 日の 13 時から 7 月 4 日の 22 時まで、第 1 回試験に対応する練習問題を、LMS のオンラインテスト機能を用いて提供した。提供期間中は回数無制限で受験できることと、1 回の受験で最大 20 分まで取り組むのができることの 2 点については、第 1 回試験の設定と異なるが、その 2 点以外は第 1 回試験の設定と同一である。

3. 項目分析機能の概要

現在、帝京大学で使用されている LMS は Blackboard 社が開発・販売している Blackboard Learn R9.1 である。この LMS には、オンラインテスト機能を用いて出題されている質問のそれぞれについて、いわゆる良い質問であるのかどうかを、オンラインテストの受験データを用いて判定する機能がある。本章では、その機能の概要について、Blackboard 社のヘルプページを参照しつつ記す。

本稿で問題となる統計数値は、識別力と難易度である。これらの数値は、個々の受験者が最後に提出した答案をデータとして用いながら算出される。先に示したように、本稿で分析の対象となっている 2018 年度前期「経済学概論」の第 1 回試験では、最大 4 回まで受験できることになっている。それゆえ、ある学生が 4 回受験した場合には、4 回目の数値がデータとして用いられることになる。

識別値は、科目についての知識がある学生とない学生を、どれだけうまく区分しているかを表した値である。Pearson の相関係数を使用して計算されているため、値の範囲は-1 から+1 の範囲をとる。識別値が 0.3 を上回る質問を良い質問、識別

⁴ 今後は大学入試において、英語外部試験の導入が進むと思われる。その点を踏まえると、いわゆる CBT の受験経験を持つ大学新生も増えると思われる。本稿で取り上げているような、LMS のオンラインテスト機能を用いた練習問題の出題は、履修学生の授業外学習を促す側面を強める形で運営する必要が高まるのかもしれない。

値が0.1から0.3の間にある質問を普通の質問、識別値が0.1を下回る質問を悪い質問と位置付けている。そして、悪い質問については、質問を見直すように促すフラグがつく。なお、難易度が100%の場合(つまり、受験した全員が正解した場合)や、すべての受験者の得点と同じであった場合には、識別力の値は計算されない。

難易度は、質問に正答した学生のパーセンテージを表した値である。パーセンテージであるため、値は0%から100%の間をとる。値が30%未満の場合は、難しい質問と位置づけられる。値が30%から80%の間をとる場合は、中程度の質問と位置付けられる。値が80%以上の値をとる場合は、簡単な質問と位置付けられる。そして、難しい質問と簡単な質問については、質問の見直しを促すフラグがつく。

4. 項目分析の結果

本章では、2018年度前期に開講した「経済学概論」において実施した第1回試験と、第1回試験の対策用として提供した練習問題について、項目分析を行った結果を紹介し、得られた結果に対する考察を行う。

第1回試験については、テストバンクに収録された38個の質問から、個々の受験者に対してランダムに10題の質問が出題され、115人が受験した。38個の質問のうち、識別力については36個の質問が良い質問、2つの質問が計算不能との結果であった。難易度については、26個の質問が簡単な質問、11個の質問が標準の質問、1個の質問が難しい質問との結果であった。識別力が計算不能と判定された2つの質問のうち、1つは複数穴埋め問題であり、この問題は16人に対して出題されて正解者が0人であったために難易度は0%と算出され、難しい質問と判定されている。もう1つは正誤問題であり、33人に対して出題されて全員が正解しているために、難易度は100%と算出され、簡単な質問と判定された。

第1回試験に向けた練習問題については、テストバンクに収録された40個の質問から、個々の受験者に対して10題の質問が出題され、88人が受験した。40個の質問のうち、識別力については39個の質問が良い質問、1個の質問が計算不能との結果であった。難易度については、5個の質問が簡単な質問、33個の質問が標準の質問、2個の質問が難しい質問と判定されている。識別力が計算不能とされた1個の質問は、複数穴埋め問題であり、この問題は14人に対して出題されて正解者が0人であったために難易度は0%と算出され、難しい質問と判定されている。

5. 結果の考察

第4章で記したように、識別力については、計算不能とされた2つの質問を除いて、良い質問との判定であった。よって、「経済学概論」における試験で使用されたテストバンクの質問は、科目に関する知識のある学生とない学生を適切に区分できており、これらの質問を用いたオンラインテストを単位認定の手段として用いることは適切といえるだろう。

なお、識別力が計算不能とされた2つの質問について詳細を確認したところ、1つの質問については、解答の設定範囲が狭すぎたのが原因と考えられる。具体的には、解答として設定されているのは「意思決定」であるが、この質問が出題された受験者は「意志決定」と解答したために、不正解になっていると考えられる。この点については、両方を正解として設定し直すことが必要と考えられる。もう1つの質問については、出題された受験者のすべてが正解しているために、識別力が計算不能になっている質問であった。これは、あまりに質問が簡単すぎると判断できるので、削除を含めた改訂が必要と考えられる。

一方、難易度については、簡単な質問と判定され、質問の見直しを促されるケースが多くなっている。まず、練習問題において簡単な質問と判定された5つの質問について詳細を確かめると、正誤問題が4つ、多肢選択問題が1つであった。多

肢選択問題については、選択肢が2つであったため、実質的には正誤問題と同じである。この点を踏まえると、正誤問題については、単純に「正」「誤」という選択肢ではない形にして、実質的には多肢選択問題にする必要がある。また、多肢選択問題は少なくとも選択肢を3つ以上設定する必要がある。一方で、これらの質問の識別値は0.63から0.77の値を示した。難易度を高めるような質問の改訂が、識別値にどのような影響をもたらすかは、今後もデータの収集を通じて検討する必要がある。

第4章で示したように、簡単な質問の数は練習問題では5個であったが、第1回試験では26個に増加した。増加分21個のうち1個を除いて、練習問題の際に出題されており、その際は標準の質問もしくは難しい質問に位置づけられている。その20個の質問のうち、正誤問題は6個、多肢選択問題は4個、複数穴埋め問題は10個である。

正誤問題6個について、識別値の変化を確かめたところ、練習問題から第1回試験に向けて、識別値が上昇した質問が4個、識別値が減少した質問が1個であった。残りの1個については、すべての受験生が正解しているため算出されていない。多肢選択問題4個について、識別値の変化を確かめたところ、練習問題から第1回試験に向けて、識別値が上昇した質問が2個、識別値が減少した質問が2個であった。多肢選択問題は実質的に正誤問題と同じと考えると、いわゆる2択の質問10個についてみた場合に、識別値の変化を確かめたところ、練習問題から第1回試験に向けて、識別値が上昇した質問が6個、識別値が減少した質問が3個という結果が導出された。

この結果の解釈としては、練習問題に取り組んでいるときは、単位取得に直接結びついていないこともあって、一見簡単に見える問題に対しては、受験生は全体として慎重に取り組んでいない一方、第1回試験という単位取得に直結するオンラインテストになると、2択とはいえ質問を

慎重に確かめて解答する受験生が多くなり、結果として、知識のある学生は正答し、そうでない学生は不正解になる、といった解釈が成り立ちうる。その解釈が正しいならば、練習問題の段階から問題に慎重に取り組んでもらうためにも、先に記したように、正誤問題は実質的には多肢選択問題に改定する必要があるし、選択肢は3つ以上設定する必要がある。

複数穴埋め問題10個の質問について、識別値の変化を確かめたところ、練習問題から第1回試験に向けて、識別値が上昇した質問が4個、識別値が減少した質問が6個であった。また、識別値の上昇幅は0.03から0.13の範囲であるのに対し、識別値の減少幅は0.05から0.31の範囲を示した。

BlackboardのLMSでの識別力は、先に記したように、識別値が0.1を超えていれば良い質問と判定する。一方で、上記の結果は、複数穴埋め問題の一部は状況によって識別値を大きく減らす、つまり、科目に関する知識のある学生とない学生を適切に区別する力を大きく減らす質問になっていることを示している。識別値の値をもとにしたLMSの判定を過信しない必要性を示しているように思われる。このような結果になった原因として1つ考えられるのは、一部の質問の文章が講義資料に記載されている文章とあまりに酷似していることであろう。それゆえ、練習問題には取り組んでいないものの、第1回試験は単位取得のために受験している学生は、(監視の目のない)オンラインテストである点を利用して、講義資料を参照しつつ出題された質問に解答し、正解して得点を得たという予測が成り立ちうる⁵。この予測が成り立つと考えられる質問については、問題文の改訂が必要であろう。

第1回試験において、標準の質問と判定されているのは10個であり、1個を除いて、練習問題でも出題されている。練習問題と第1回試験の双方で出題されている9個のうち、数式計算問題は2

⁵ この予測と本稿の分析結果との関係について検証することは、今後の課題の1つと考えられる。

個、複数穴埋め問題が 7 個である。また、この 9 個の質問は練習問題の時にも標準の質問と判定されている。この 9 個の質問について、識別値の変化を確かめたところ、練習問題から第 1 回試験に向けて、識別値が上昇した質問が 5 個、識別値が減少した質問が 4 個であった。識別値の上昇幅は 0.03 から 0.23 の範囲であるのに対し、識別値の減少幅は 0.14 から 0.25 の範囲である。数式計算問題は 2 個とも識別値が低下している。

この結果の解釈としては、練習問題に取り組んでいるときは、単位取得に直接結びついていないこともあって、数式計算問題という一見難しく見える質問に対しては、知識が十分でない学生を中心に解答をあきらめる受験生が多い一方、第 1 回試験という単位取得に直結するオンラインテストになると、出題される 10 題のうち数式計算問題がどの程度出題されるか分からないために、解答に取り組まざるを得ないと考える受験生が多くなったという解釈が成り立ちうる。

6. まとめと今後の課題

本稿では、筆者が 2018 年度前期に担当した「経済学概論」において、LMS のオンラインテスト機能を用いて実施した第 1 回試験(中間試験にあたる試験)と、その試験への対策として出題した練習問題の質を、LMS の項目分析機能を用いることで導出される識別力と難易度に注目しながら分析した。分析の結果、識別力、つまり、科目に関する知識のある学生とない学生を適切に区別する力については、おおむね問題なく、オンラインテストを単位認定の手段として用いることが適切であるのが確かめられた。一方で、選択肢が 2 つの質問(いわゆる 2 択の質問)は難易度の関係から、選択肢を 3 つ以上設定するように改訂の必要があることや、穴埋め問題については、講義資料の文章と酷似していない問題文で作成する必要があることが確かめられた。

筆者にとって今後の課題は、差し当たり、先に記したような結論を踏まえ、テストバンクに収録さ

れている質問の改訂に取り組むことだろう。本稿の分析を踏まえた改訂においては、簡単な質問から標準の質問に改訂するといった、難易度を意識した改訂が中心になるだろう。しかし、そのような改訂を施した質問が識別力を保った質問として機能するかどうかは、今後のオンラインテストの実施を通じて得られるデータを用いて、検証する必要がある。そして、その検証を通じて、オンラインテストの形式で出題される質問における、識別力と難易度の関係について、知見を積み増してゆきたい。

LMS のオンラインテスト機能は、授業外学習を促すのを通じて知識の定着に寄与する機能でもあるのを第 1 章において言及した。本稿での分析を踏まえて、オンラインテストで出題している問題の改訂に取り組み、改訂した問題をオンラインテストで出題した際に、知識の定着により寄与する結果に結びついているか検証することも、筆者が今後取り組む必要のある課題であろう。

謝辞 本稿の作成過程で、渡辺博芳先生からコメントを受けた。そのコメントに対応することを通じて、本稿の質が向上した。また、LT 開発室室員からは、コメントや原稿フォーマットのサポートを受けた。さらには、ICT を大学での教育、特に学士課程での教育に活用してゆくことに、熱心に取り組んでおられる帝京大学宇都宮キャンパスの教職員の方々から、いろいろな刺激を受けており、その刺激は本稿を執筆する原動力の 1 つになっている。記して感謝する。

参考文献

- [1] 橋本貴充, “テスト理論”, 日本教育工学会監修 永岡慶三・植野真臣・山内祐平 編著 『教育工学における学習評価』, 第 2 章「学習評価の基礎」, pp.18-36, 2012
- [2] 赤倉貴子, “コンピュータ利用テストの実用”, 日本教育工学会監修 永岡慶三・植野真臣・山内祐平 編著 『教育工学における学習

評価』, 第 3 章「コンピュータ利用テスト」,
pp.69-91, 2012

[3] 田中晃, “アダプティブ・ラーニング(適応
学習)の潮流ー教育ビッグデータを活用した
学習の個別最適化ー”, 情報処理, 58 卷 3 号,
pp.184-185, 2017

[4] 溝口佳宏, “WebCT のオンラインテスト
機能の活用に関する実践報告”, 帝京大学ラー
ニングテクノロジー開発室年報, 第 9 卷,
pp.81-88, 2012

[5] 溝口佳宏, “オンラインテストは期末試験
として継続利用できるか?”, 帝京大学ラー
ニングテクノロジー開発室年報, 第 10 卷,
pp.93-100, 2013

[6] 溝口佳宏, “オンラインテストはセルフラ
ーニングを支援する手段なのか?”, 帝京大学
ラーニングテクノロジー開発室年報, 第 14
卷, pp.87-91, 2017