

オンラインテストで出題する質問と学習目標との関連についての予備的な 分析

溝口 佳宏

帝京大学経済学部地域経済学科

概要

筆者の担当科目において実施しているオンラインテストで出題する質問が、履修者の学習活動を確認できる質問になっているかどうか予備的な検証を試みた。分析では 2018 年度前期に担当した「経済学概論」で実施したオンラインテストの受験データを基にした。このデータを LMS の項目分析機能を用いて処理するのを通じて導出された数値を参照し、ガニエの学習成果の 5 分類に照らし合わせる形で検証を試みた。その結果、複数穴埋め問題は、教員側が期待している学習行動と履修者が実際に取っている学習行動が一致する可能性が相対的に高く、学習成果を測る上で適切な問題形式である可能性が相対的に高いのが確かめられた。一方で、数式計算問題は、本来の学習課題が試験条件によって歪められている可能性があり、学習成果を測る上で適切な問題形式とは必ずしもいえないのが確かめられた。

1. はじめに

筆者が担当している科目の 1 つである「経済学概論」では、単位取得の条件としてのオンラインテストを 2 回実施し、履修学生に対して受験を義務付けている。本稿では、そのオンラインテストにおいて出題している質問が、「経済学概論」の履修を通じて知識を獲得しているかどうかを確かめる質問であるだけでなく、履修の際に期待される学習行動を経ているかどうかを確認できる質問になっているかどうか、予備的な分析を試みる。

本稿で分析対象とする「経済学概論」という科目は、いわゆる知識獲得型科目と位置付けられる。知識獲得型科目で目指すことは、言うまでもなく、その科目で取り上げられている事柄に関する知識を、履修学生が獲得することである。そして、履修学生が知識を獲得できているかどうかを確かめる方法として、「経済学概論」では LMS の

オンラインテスト機能を利用した試験を用いている。

LMS のオンラインテスト機能を利用した試験が、履修学生の知識獲得の程度を確かめる適切な手段であるかどうか確かめるには、オンラインテストの受験データを用いた分析が有効と考えられる。分析する際のアプローチの 1 つは、出題している質問それぞれについて識別力を計算し、その値を見るのを通じて、科目についての知識がある学生とない学生を判別する質問かどうかを判別するというアプローチである。帝京大学で使用されている LMS (Blackboard Learn 9.1) には項目分析機能が備えられていることから、識別力の計算は容易である¹。また、同時に困難度も計算してくれる²。溝口[1] は、2018 年度前期に担当した「経済

A preliminary analysis on the relationship between questions in online-testing and learning objectives
Yoshihiro Mizoguchi
Department of Regional Economics, Faculty of Economics, Teikyo University

¹ Blackboard Learn 9.1 の項目分析機能では「識別値」と表記されている。なお、この値は Pearson の相関係数をもとに算出されており、古典的テスト理論における項目識別力 (Item-Total Correlation: I-T 相関) に相当する。

² 古典的テスト理論との関連をふまえ、「困難度」と記しているが、Blackboard Learn 9.1 の項目分析機能では「難易度」と表記されている。古典的テスト理論では、「正答率」や「通過率」といった用語で表されることもある。筆者としては、「正答率」という用語が、定義 (辞書的な意味) と人々の持つイメージとの乖離が少ないと思われ、混乱しない用語と考える。

学概論」において実施したオンラインテストで出題した質問について、上記の機能を用いた分析を行った結果、実施したオンラインテストは当該科目で取り扱った知識を備えた学生と備えていない学生を判別できており、単位認定の手段として有効であることを示している。

一方で、上記のアプローチにより得られる結果は、オンラインテストの受験者それぞれが、どのような過程で知識を獲得しているのかを直接示してくれる訳ではない。言い換えると、教員が期待している方法で、担当科目の履修者が知識を獲得しているかどうかを直接示している訳ではない。たとえば、空欄補充問題を出題する際に出題者が期待していることは、受験者があらかじめ科目で取り扱った知識を記憶するという準備を経たうえで、その問題に正答してもらうことである。ところが、実際に受験者がしていることは、その問題が出題された時に初めて（もしくは慌てて）、その問題の答えが書かれている場所を探索し、その探索結果をもとに（もしくは鵜呑みにして）解答することかもしれない。この場合は、その受験者がその問題に正答していたとしても、その受験者が知識を獲得しているとは言えないかもしれない。

それゆえ、オンラインテストを活用することの有効性を示すには、溝口[1]で示した結果だけではなく、教育の効果・効率・魅力を高めるための方法論であるインストラクショナルデザインの観点からの検証をする必要があるかもしれない。よって本稿では、溝口[1]で示した識別力および困難度、つまり、2018年度に担当した「経済学概論」において実施したオンラインテストで出題した質問について導出した識別力と困難度を引き続き用いつつ、インストラクショナルデザインの観点から上記の2つの値を解釈するのを通じて、筆者の担当科目の1つである「経済学概論」において実施しているオンラインテストの予備的な検証を試みる。

本稿で取り上げているオンラインテストは、単位認定の手段として用いられているので、学習

(授業)の出口で用いられる手段である。インストラクショナルデザインのモデルにおいて、学習の出口に関する部分については、ガニエの学習成果の5分類がガニエほか[2]において提唱されている。よって、本稿では、溝口[1]で示した識別力を、ガニエの学習成果の5分類を参照しつつ検証することが主要な内容となる。

本稿の残りの構成は次の通りである。第2章ではガニエの学習成果の5分類について、鈴木[3]を参照しつつ、本稿での分析と関連付けながら概説する。第3章では、分析手法について説明する。第4章では、分析結果と考察について述べる。第5章ではまとめと今後の課題を記す。

2. ガニエの学習成果の5分類と本稿での分析との関連

ガニエの学習成果の5分類は、言語情報・知的技能・認知的方略・運動技能・態度で構成される³。このうち、運動技能は体の全部もしくは一部を動かして一定の課題を成し遂げられるようになることを指しているので、本稿の分析対象ではない。また、態度は物事や状況などに対する肯定的もしくは否定的な感情を指しているため、本稿の分析対象ではない。つまり、本稿の分析対象は、認知領域（あたま）の学習成果である言語情報・知的技能・認知的方略の3つとなる。

言語情報については、各種データを覚えて思い出すことができるかどうかの評価基準となる。本稿で取り上げるオンラインテストでは、空欄補充問題や多肢選択問題が言語情報に対応する問題になるだろう。

知的技能については、学んだルールなどを未知の例に適用できているかどうかの評価基準となる。本稿の分析対象であるオンラインテストでは、多肢選択問題や計算問題が対応する問題になるだろう⁴。

³ ガニエの学習成果の5分類は、多くの場面で表を提示する形で紹介されている。多くの場面で紹介されているので、その表を探索するのは容易であるが、紹介される表の出所としては、鈴木[3]を挙げるのが適切だろう。

⁴ 鈴木[3]は、いわゆる受験数学において出題される問題は、

認知的方略については、個々の学生が学習を効果的に行うための作戦を習得しているかどうかの評価基準となる。本稿の分析対象科目である「経済学概論」では、2種類のオンラインテストが提供される。1つは、単位認定の手段であり、履修学生に受験が義務付けられているオンラインテストである（以後では、期末オンラインテストとよぶ）。もう1つは、単位認定のためのオンラインテストに備える1つの手段として、ほぼ同形式で提供され、取り組むかどうかは個々の学生に任される、練習問題としてのオンラインテストである（以後では、自己学習オンラインテストとよぶ）。その点を踏まえると、練習問題に取り組んでいる学生について、自己学習オンラインテストの取り組み状況と期末オンラインテストにおける得点状況を比較することが、認知的方略に関する評価で必要になるだろう。

3. 分析手法

筆者が2018年度前期に担当した「経済学概論」において、期末オンラインテストおよび、自己学習オンラインテストの2つのオンラインテストの受験データを用いる。

まず、認知的方略について検討する。教員側が履修学生に期待している認知的方略は、自己学習オンラインテストに取り組んだ際に誤答した質問について、その質問に対応する項目を授業資料などで復習してもらうことであり、その動作をすることが期末オンラインテストで合格点を取るための最良の手段と認識してもらうことであり、実際にその行動を取ってもらうことである（以後では、この行動を誤答復習行動とよぶ）。その際、誤答した質問が空欄補充問題であれば、空欄に記入すべき言葉を、復習を通じて覚えてもらうことである。また、誤答した質問が多肢選択問題

であれば、質問された項目の正しい内容を、復習の際に覚えてもらうことである。

一方で、履修学生が実際にとる認知的方略は、自己学習オンラインテストに取り組んだ際に誤答した問題について、その問題の解答を文字通り記している箇所を授業資料から見つけ出すことであり、期末オンラインテストで同じ質問が出題されたなら、その箇所を手元で見つて解答を入力することが最良の手段と認識することであり、実際にその行動をとることかもしれない（以後では、この行動を正答探索行動とよぶ）。仮に、そのような認知的方略を実際に取られた場合には、質問が空欄補充問題であるならば、空欄に記入すべき言葉を覚えていないということになるので、その質問は言語情報を評価する質問として適切ではないといえる。また、多肢選択問題であれば、知的技能を測る質問としては適切ではないといえる。

次に、このような仮説に基づいて、本稿での分析で用いる受験データ、つまり出題した質問それぞれに対する識別力と困難度を検討する。先に記したように、教員側が履修学生に期待している認知的方略と、履修学生が実際に取っている認知的方略が一致している、つまり、履修学生が誤答復習行動を取っているならば、期末オンラインテストでの正答率は、自己学習オンラインテストにおける正答率よりも上昇するだろう。それゆえ、困難度の値は上昇するだろう⁵。さらに、この場合は、科目で取り扱った項目に関する知識を持っているのを背景として正答する受験者が多くなるはずなので、識別力の値は上昇すると考えられる。

一方で、教員側が履修学生に期待している認知的方略と、履修学生が実際に取っている認知的方略が一致していない、つまり、履修学生が正答探索行動を取っているならば、期末オンラインテストでの正答率は、自己学習オンラインテストにおける正答率よりも上昇するだろう。Blackboard

暗記された問題のパターンを思い出すという言語情報に属する部分と、思い出したパターンの中で計算処理をするという知的技能の部分が混じっており、本来の学習課題の性質が試験条件によって歪められていると主張している。この主張を踏まえると、筆者が実施しているオンラインテストでの計算問題は、パラメーターを複数設定しているとはいえ、知的技能を純粋に測る問題になっていないかもしれない。

⁵ ここでの困難度とは、出題された質問に対する正答率で定義される。よって、困難度の値が上昇するとは、正答率が上昇することを意味する。「困難度」という単語の持つイメージから生じる誤解を防ぐために、本稿では適宜「正答率」という単語を用いる。

Learn の項目分析機能で導出される難易度の値は、質問に正答した受験者の比率で表されるので、困難度の値は上昇するだろう。しかし、科目で取り扱った項目に関する知識がないにも関わらず正答している受験者の比率が上昇するので、識別力の値は減少するだろう。

以上を踏まえると、オンラインテストで出題している質問が、学習成果を測る上で適切な質問になっているかどうかを検討する際には、本稿での分析手法に従うと、識別力の変化に注目することになる。そして、識別力の値が減少している場合に、学習成果を測る上での適切な質問であるかどうか疑問視するという形になる。

4. 分析結果と考察

本稿で分析対象とする 2 回のオンラインテストで共通して出題されている質問は 35 問であった。そのうち、自己学習オンラインテストから、期末オンラインテストにかけて、識別力が減少した質問は 17 問であった。内訳は、複数穴埋め問題が 9 問、正誤問題が 3 問、多肢選択問題が 3 問、数式計算問題が 2 問であった。識別力の減少幅は 0.02 から 0.33 にわたっている⁶。また、この 17 問すべてについて、困難度は上昇している。つまり、自己学習オンラインテストから、期末オンラインテストにかけて、受験者の正答率は上昇している。上昇幅は、2.5%から 49.32%にわたっている。これらの結果は、識別力の減少した質問については、第 3 章で記した仮説のうち、教員側が期待していない認知的方略、つまり、正答探索行動を履修学生は取っていることを示唆している。上記した結果をもとにしながら、オンラインテストで出題する質問の形式が、ガニエの学習成果の 5 分類、特に認知的方略に照らしたうえで、どの程度適切であるのかを検証したい。

まず、数式計算問題は学習成果を測る上での良問とは限らない。識別力が減少した数式計算問題 2 問について、それぞれの正答率は、

8.16%、11.13%上昇し、識別力は、0.14、0.25 減少している。出題の際には、受験者が練習問題に取り組む過程で、答えの数値を覚えるという行動を防ぐために、パラメーターを複数設置したうえでランダムに出題する方式を採用している。しかし、受験者が練習問題に取り組む過程で行っていることは、出題されるごとに計算をするのではなく、出題されるごとに答えの数値を記録すること、つまり、正答探索行動のように推察される。鈴木[3] は、受験数学において出題される問題は、本来の学習課題が試験条件によって歪められているとの主張をしている。本稿での予備的な分析は、その主張を裏付ける結果の 1 つと位置付けられるかもしれない。

正誤問題は、質問の作り方によって対照的な結果がでる問題形式といえる。出題した正誤問題のなかには、正答率は 2.5%上昇し、識別力は 0.25 減少している質問があった。この質問は、教員側が期待していない認知的方略、つまり、正答探索行動を、履修学生が採用していることを表す典型的な問題と位置付けられる。一方で、正誤問題は一律に学習成果を測る上で不適切な問題形式とも言い切れない。正答率が 25.9%上昇しているものの識別力も 0.22 上昇している質問もある。第 3 章で示した仮説に従うと、教員側が期待している認知的方略、つまり誤答復習行動を履修学生が採用しており、その認知的方略が履修学生にとってのメリットになっているのであれば、正答率と識別力はともに上昇すると考えられる。よって、この正誤問題は学習成果を測る上で適切な質問と位置付けることもできる。

正誤問題と比較すると、多少は適切な問題形式であるものの、多肢選択問題も質問の作り方によって対照的な結果がでる問題形式だろう。正答率が 87.5%から 92.86%へと上昇しながらも識別力は 0.77 から 0.75 への減少にとどまっている質問がある。また、正答率が 63.6%から 97.3%に上昇しながらも識別力が 0.33 から 0.72 に上昇している質問もある。このような質問は、正答率の高さから Blackboard Learn の項目分析機能では簡単な質

⁶ 識別力は Pearson の相関係数をもとに算出されているので、値の範囲は-1 から+1 である。

問と判定され、質問の修正を勧められる。しかし、教員側が期待している認知的方略と、履修学生にとっての認知的方略が一致する、つまり、履修学生が誤答復習行動を取っている可能性が高い質問と位置付けるのも可能と考えられる。つまり、修正せず、むしろ残しておく質問ともいえるだろう。一方で、正答率の 27.1%上昇が識別力の 0.33 減少をもたらしている質問も存在する。この質問は、教員側が期待していない認知的方略、つまり、正答探索行動を履修学生が採用するのが適切になっている質問であり、学習成果を測る上では不適切な質問だろう。

複数穴埋め問題は、本稿で取り上げている学習成果を測る上で、適切な質問を作成できる可能性が相対的に高い問題形式といえる。正答率が 48.96%上昇しているにもかかわらず、識別力の減少が 0.05 にとどまっている質問がある。ほかにも、正答率が 33.06%上昇しているが識別力の減少が 0.06 にとどまっている質問や、正答率の 24.6%上昇が識別力の減少 0.07 にとどまっている質問がある。さらには、正答率が 35.9%から 74.75%に上昇していると共に識別力が 0.71 から 0.94 に上昇している質問がある。これらのデータを踏まえると、複数穴埋め問題については、教員側が期待している認知的方略、つまり、誤答復習行動を履修学生が採用しており、その認知的方略が履修学生にとってのメリットになっていると考えられる。つまり、複数穴埋め問題は学習成果を測る上で適切な問題形式といえるだろう。

5. まとめと今後の課題

本稿では、筆者の担当科目で実施しているオンラインテストで出題している質問が、教員側が期待している学習行動を履修者に促すような質問になっているかどうかを、Blackboard Learn の項目分析機能で導出される数値を用いつつ、ガニエの学習成果の 5 分類に照らしあわせる形で、予備的な検証を試みた。

本稿での分析で主に示唆された事柄は 2 点ある。第 1 点は、数式計算問題は学習成果を測る

上で適切な問題形式とは必ずしもいえず、本来の学習課題が試験条件によって歪められている可能性がある点である。第 2 点は、複数穴埋め問題は学習成果を測る上で適切な問題形式である可能性が相対的に高い点である。

今後の課題は多岐にわたると考えられる。差し当たり 4 点あげておきたい。第 1 点は、数式計算問題をはじめとして、学習課題が試験条件によって歪められているかどうかを判別するには、どのような点に注目する必要があるのか明らかにすることである⁷。

第 2 点は、第 1 点とも関連するが、学習課題が試験条件によって歪められないような質問にするには、どのような質問にする必要があるのか、質問作成の際の留意点について明らかにすることである。

第 3 点は、本稿での分析結果の普遍性に関わる点である。本稿の分析結果はあくまでも 2018 年度に担当した「経済学概論」で実施したオンラインテストでの受験データのみを用いている。それゆえ、別の年度で実施したオンラインテストでも、本稿での分析と同様の結果が導出されるのかどうか、違った結果が導出されるのであれば、何か違った結果の導出に影響をもたらしているのかを確かめる必要がある。この点は、本稿で紹介した数値の変化 (差) を有意な変化 (差) とみなせるのかどうかという点にも関わると思われる。

第 4 点は、第 2 点とも関連するが、学習成果の分類と出題する質問の形式との対応に関する点である。本稿の分析では、ガニエの学習成果の 5 分類を意識していた。この分類を今後も意識しながら、学習の出口に関する分析を進める際や、担当科目で試験問題を作成する際には、この分類と出題する質問の形式は、いわゆる 1 対 1 に対応する必要があるのかどうか、といった点を明らかにする必要はあるだろう。例えば、言語情報を測るうえでの問題形式は空欄補充問題に限定されるの

⁷ これは単に、筆者が先行研究を知らないだけの可能性が非常に高い。よって、この点は筆者自身の今後の課題という側面が非常に強い。今後の課題として挙げている第 2 点と第 4 点についても、同様のことがいえる。

かどうか、空欄補充問題に限定される必要があるのかどうか、といった点である。

謝 辞 本稿は Japan Blackboard User Group 第 11 回会合や第 148 回 LT セミナーでの発表に対して寄せられたコメントに着想を得ている。これらの発表に対してコメントを寄せてくれた先生方や参加いただいた先生方に感謝する。特に、宮崎誠先生（帝京大学）からのコメントが本稿に深くかかわっており、記して感謝する。また、筆者の教育工学分野での研究活動や本稿の体裁に対しては、帝京大学 LT 開発室の教職員の方々から多大なご支援をいただいている。記して感謝する。もちろん、本稿における誤りは、すべて筆者の責任である。コメントを歓迎する。

参考文献

- [1] 溝口佳宏, “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”, 帝京大学ラーニングテクノロジー開発室年報, 第 16 巻, pp87-93, 2019
- [2] ロバート・M. ガニエ, キャサリン・C. ゴラス, ジョン・M. ケラー, (鈴木克明, 岩崎信 監訳), “インストラクショナルデザインの原理”, 北大路書房, 2007
- [3] 鈴木克明, “放送利用からの授業デザイナー入門～若い先生へのメッセージ～”, 日本放送教育協会, 1995