

多肢選択テストの回答時間から見た理解に関する考察

鈴木 治 郎^{†1}

正誤判定問題と多肢選択問題では回答者の回答態度の違いが想定できるために、それぞれの問題の利用目的を変えるべき理由があるとする仮説を我々は持っている。この仮説にもとづく違いを、回答者の回答時間を通じて検出できるかを本稿では試みた。その結果、2つのテスト問題の内、後で回答したテスト問題（実際には多肢選択問題）の回答所要時間のほうが十分に短く、それはテスト問題への慣れにもとづき回答時間に違いが生じたと考えてよいものであった。したがって、正誤判定問題等のテスト形式の適した問題を判断するための客観的データを得るためには、回答者の行動の測定にもっと工夫を要することがわかった。

Note on a Multiple-Choice Test associate with the Elapsed Time

SUZUKI JIRO^{†1}

We have examined the elapsed time of two type of tests, namely, a multiple-choice test and a true or false question. We have displayed two test as that the former is a true or false question and the latter is a multiple-choice test. The elapsed time of the multiple-choice test is less than other test. Our consideration has found the reason that familiar to latter test. Since we believe in that solving a multiple-choice test is much easier than solving a true or false test, we have get a result without satisfaction.

1. はじめに

オンラインテストの利用においては自動採点の容易な、穴埋め／多肢選択／正誤判定問題

は実際に多く利用されてきている。ところが、これらの問題スタイルを提供するための作問作業は、自動採点の容易さとは裏腹に、実は容易ではない。それは紛れのない正答を保証できる問題を用意することがけつして容易ではないからであり、この問題作成にまつわる課題が困難なことの例証には、正答として適切な選択肢がないために入試ミスを引き起こしたなどの社会的事件となる¹⁾ことが珍しくないことを見ても明らかであろう。

筆者はこの問題作成にまつわる困難さの問題を、逆に回答者のコンテキスト理解の問題に転換することで、正誤判定問題に積極的に活用できる可能性について論じている²⁾。そこで述べた筆者の仮説の要点は2つある。

- 正誤判定問題には学習者のコンテキスト理解を強く評価する機能がある。
- 多肢選択問題には学習者の文章解釈能力を期待して、学習課題に対する要となる情報を提供する機能がある。

ここでは²⁾、さらに回答者の心理的態度に対する考察から、

- 正誤判定問題では受験者に判断を強いることに対する心理的ストレスがある。
- 多肢選択問題では各選択肢に関するよりよい評価を判断するだけなので、判断のための心理的ストレスは小さい。

という性向が回答者に現れる可能性についても論じた。本稿では、この心理的ストレスの存在に関して、回答者の回答所要時間に関する観察から違いを見いだせる可能性について、簡単な正誤判定問題および多肢選択問題への回答時間を通じて測定することを試みたことを報告する。

なお一般には、ここで論じたようなICT利用環境における被験者の無意識な認識を把握する手段として、被験者の反応時間を利用できることがわかっており³⁾、それは実験心理学の常套手段でもある。

2. 問題の背景

本稿で論じている正誤判定問題や多肢選択問題の活用の目的は、e-Learning教材活用のためという近年増加中の需要ばかりでなく、最近の記述式レポート形式自体がある意味不毛になってきていることの克服を目指している。ここで不毛であるとする指摘は、ウィキペディアのコピー&ペーストに過ぎないレポートという論外な場合を別にしても、Google検索によって得られる適当なテキスト素材のコレクションに過ぎないレポートが、とくに記述式レポートにおいて顕著になってきたと考えるからである。そのために、検索により安易なレポートの素材を入手できることが主な原因であると考えてみる。すると、対抗策の一つが見

^{†1} 信州大学全学教育機構
Shinshu University, School of General Education

えてくる。それはコンテキストに依存した知識を探すことである。Google 検索などの一般の検索技術では、手に入れられる知識は基本的にコンテキストフリー（コンテキスト依存性のない）の知識である。

実際に検索技術の活用のための有効なノウハウは、どのような検索キーワードを追加することが、検索したい知識のもつコンテキストを確定するかにつながるかを解説しているという多くの事例⁴⁾を見ても、コンテキスト理解の重要性は明らかであろう。

2.1 正誤判定問題作成上の課題

ここでは上に述べたコンテキスト依存性の問題が、とくに正誤判定問題においてどのように現れるかを、筆者がすでに論じた事例²⁾を通じて紹介する。そこでも扱った次の例文を考える。

スティーブ・ジョブズはアップル・コンピュータの社長である

この主張を扱う授業において、そこでの知識の精度要求に応じて、たとえば次のような正しいとする理由、あるいは誤りとする理由を考えることができることを我々は論じた²⁾。

正しいとすべき理由 たとえば以下のような状況（授業コンテキスト）の設定によって正しいと判断すべき文になる。

- 現在の社名はアップル社が正しいが、多くのコンテキストではアップル社とアップル・コンピュータ社とは同一性をもつ企業名として解釈することに問題はない。よって多くのコンテキストでは2つの会社名を区別する必要はなく、社名の間違いとすべき理由はない。
- 社長と呼ばれる地位は、日本では多くの場合、最高経営責任者（CEO）と同じものとして解釈としてよい。よって CEO と記述してないから誤りとするようなこだわりは無視しても問題ない。

誤りとすべき理由 たとえば以下のような状況（授業コンテキスト）の設定によって誤り、あるいは出題自体が不当と判断すべき文になる。

- アップル社がアップル・コンピュータ社から社名変更したという事実は、現在のアップル社の儲け頭が iTunes にまつわるビジネスであることに起因すると考えられる。したがって、たとえばインターネットをメディアとして捉える授業においては、「アップル社」という社名の記述に「コンピュータ」の文字が入るかどうかが、無視できない大きな違いとなる。

- アメリカの企業の経営形態を扱う授業では、CEO/CTO/CIO などの役割を明確に区別できなくてはならない。そのようなコンテキストにおいては、社長のような担当責任のあやふやな呼称は明らかに不十分である。

以上の例で論じたような正誤判定問題が成立するためには、出題者と回答者とがコンテキストを共有できなければならない。コミュニケーション一般の問題であれば、コンテキストの共有をどうすれば可能になるかは大きな問題である。ところが適当な授業に対して、その受講者が授業を理解しているのかどうかを確認するという問題として正誤判定問題が活用されるのであれば、テスト問題自体には記述されていないコンテキストを理解した上で問題の回答に取り組むことが期待されて当然とはいえないだろうか。このように考えることで、正誤判定問題は問題自身の知識確認ばかりでなく、その問題でテストされる授業に対するコンテキスト理解の判定問題の役割を与えることができるのである。

2.2 正誤判定問題を題材とした記述問題化

ここまで本節で論じてきた正誤判定問題に関する悩みを、次のような出題としてみれば、それは記述式問題として活用できるものになる。

「○○○という問題の正答は△△△である」が正しい問題／正答の組となるように、○○○という問題の状況設定を補いなさい。

我々はこの例にあげた**問題に対して状況設定を補う行為**、いいかえれば問題のためのコンテキスト設定が、実は暗黙の了解の中で行われることが多くのテストにおける実情であり、そのことを欠点と考えるのではなく、適当な授業等でテストが行われるためには、**テスト受験者は授業等に対するコンテキストを理解しているかどうか**がむしろ問われるべきであると考えるのである。

3. 仮説の測定環境

ここで検討するデータは、以下のようにして測定した。オンラインテストの実施環境は信州大学が全学生に対して提供している学修支援環境 eALPS を利用した。eALPS はオープンソースの CMS である Moodle を利用して構築されている。以下で測定環境の機能について述べるときは Moodle の名前を用いることにする。学生に対する eALPS の利用サポートは全学生への利用支援を担う e-Learning センターが行なっているため、利用自体の手助けを教員が担う必要はない。

実際に提供した問題は、Moodle 上の小テスト機能を利用して正誤判定問題および多肢選択問題（4 択問題）とした。Moodle の小テスト機能では次の 6 つの情報を取得できる。

- 被験者 ID
- 回答開始時刻
- 回答終了時刻
- 回答所要時間
- テスト得点
- 各設問ごとの回答

以上のレポート情報をもとに、被験者 ID による各データ間の相互参照を行うことにより我々は以下の比較を試みる事ができた。

- 2つのテスト（正誤判定問題および多肢選択問題）に関して、その回答開始時間にもとづきテストを嗜好する順序を調べる
- 回答所要時間とテスト得点との相関を調べる
- テスト得点間の相関を調べる

さらに我々の仮説からは、回答者である学生は正誤判定問題よりも回答においてストレスの少ない択一問題を好む傾向が予想されるため、この嗜好を測定できる可能性があるよう、Moodle 上のテスト問題選択表示においては

R の平均・分散計算に関する問題（正誤判定）

R の平均・分散計算に関する問題（4 択）

と学生に表示するようにした。この表示により、多肢選択問題をより好む回答者は、2つの問題を表示順序によらずに多肢選択問題を先に回答することが予想され、その回答情報を、我々は回答開始時間として取得できる。

3.1 テスト問題の難易度と出題

被験者である学生たちに対しては、オンラインテストの回答はインターネット検索や自前のノートを参照しながらの「いわゆるカンニング」自由のテストだと想定した知識確認問題であると事前に伝えてある。また、事前に自前のノートを用意してある学生にはきわめて容易に回答することのできる問題であるが、問題を見てから初めてオンライン検索を行なっても、時間内の回答には間に合わないよう回答時間を制限してあるとも伝えてある。なお、この目的のために実際に競っていた回答時間の制限は3分（= 180秒）である。測定の用いた設問は、大学初年次の学生を対象とする統計学を主題とした授業で扱う問題を対象とした。この授業では、統計学を実際の応用問題に適用するための分析ツールとして統計解

析ソフトウェア R の利用を設定している。今回、測定に利用した問題は以下の通りである。これらの問題は、事前に指示された学習を済ましていれば全問正解であることが期待されるものである。

正誤判定問題 正誤判定問題は以下の5つの知識確認問題とした。

- (1) R で合計を求める関数名は sum である。
- (2) R で標本の大きさを求める関数名は size である。
- (3) R で平均を求める関数名は mean である。
- (4) R で分散を求める関数名は variance である。
- (5) R で標準偏差を求める関数名は var である。

多肢選択問題 多肢選択問題では選択肢の数を4つに設定した。その際に、「選択肢の数だけ4つの正誤判定問題を並べる」ことにならないよう配慮し、以下の通り作成した。

- (1) R で平均を求める関数名を以下から選びなさい。
{ a. length , b. average , c. sum , d. mean }
- (2) R で合計を求める関数名を以下から選びなさい。
{ a. length , b. mean . c. average, d. sum }
- (3) R で標本の大きさを求める関数名を以下から選びなさい。
{ a. size , b. mean , c. sum , d. length }
- (4) R で分散を求める関数名を以下から選びなさい。
{ a. se , b. variance , c. sd , d. var }
- (5) R で標準偏差を求める関数名を以下から選びなさい。
{ a. variance , b. var , c. se , d. sd }

統計解析ソフトウェア R に関する知識をもたない人でも想像がつくように、多肢選択問題は正誤判定問題と比べて、設問自体に正答を考えるためのヒントが現れる。この一般的事実を踏まえれば、多肢選択問題を先に回答し、そこでの自身の知識の不備を補った後に、正誤判定問題を回答する回答者は珍しくないものと考えられる。このような行動をすると推測できる回答者の情報を部分的にも得るために、表示順序そのものによって多肢選択問題を先に回答することへ回答者を誘導することのないように、前述の正誤問題と択一問題の表示順序に関する配慮を行なったのである。

4. 測定結果

前節の問題に対して34名の回答を得た。以下で主な統計を示す。

4.1 テスト問題の選択順序

34名中多肢選択問題を先に回答した者は1名、CMSでの表示順と同じ正誤判定問題から回答した者が33名と、我々の予想の反映はなかった。また、先に回答した問題における経験を踏まえて、後で回答する問題のための学習準備をする、という学習行動を示すものと想定すれば、2つのテストへの回答時刻には学習時間のための空きが生じるはずだが、回答者34名中1分以上の時間を空けて後で回答した問題に取り組んだ学生は1名のみであった。このため、以下では多肢選択問題への選好については議論しないとともに、先に受験した問題にもとづき、もう一つの問題受験のために学習準備をするなどの学習行動についても論じない。

4.2 得点分布

表1, 図2に得点分布情報を示しておく。図中、Formerは先に回答した得点、Latterは後で回答した得点を示す。表中、「差」は正誤判定問題の得点から多肢選択問題の得点を引いたものを表す。表1にあげた分布から分かるように正規分布で近似される分布ではない。

表1 得点分布
Table 1 Distribution of Test Score

問題種別	平均点	最低点	第1四分位	中央値	第3四分位	最高点
正誤判定	3.6	2	3	4	4	5
多肢選択	3.5	1	3	4	4	5
合計	7.1	4	6	7	8	10
差	0.1	-2	0	0	1	3

参考までに2つのテスト得点の平均値の差に関するt検定をしてみるとp値0.443であり、違いがあるとはいえない。

4.3 回答所要時間分布

表2に所要時間分布、図??には所要時間の差の分布情報を示しておく。時間は秒単位で与えてある。回答に課した時間制限3分を超えると回答は自動終了かつ採点される。このときの所要時間の記録は181秒(3分1秒)となる。「差」は正誤判定問題の回答所要時間から多肢選択問題の回答所要時間を引いたものである。比1は正誤判定問題の回答所要時間を多肢選択問題の回答所要時間で割ったものであり、比2は比1の逆数である。比2の分布を与えた理由は、比1において1より小さな比はすべて一つの階級にまとめられるためである。表2の分布から分かるように正規分布で近似される分布ではない。参考までに2

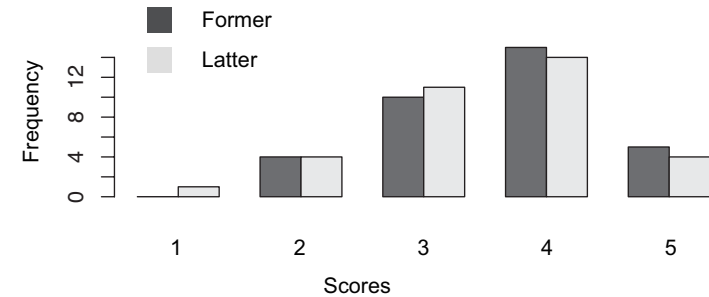


図1 得点分布
Fig. 1 Distribution of Test Score

表2 回答所要時間の分布
Table 2 Distribution of Elapsed Time

問題種別	平均値	最小値	第1四分位	中央値	第3四分位	最大値
正誤判定	109.3	19.0	54.8	108.5	162.5	181.0
多肢選択	88.3	22.0	42.3	74.0	136.2	181.0
合計	197.6	51.0	129.0	180.0	267.5	362.0
差	21.0	-112.0	-4.5	9.0	39.0	130.0
比1	1.68	0.32	0.91	1.09	1.85	6.20
比2	0.99	0.16	0.54	0.92	1.02	3.20

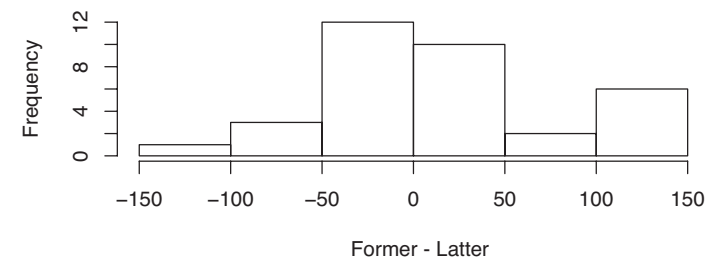


図2 回答所要時間の差の分布
Fig. 2 Distribution of Test Score

つのテストの所要時間の差について平均値に関する t 検定をしてみると p 値 0.043 になり、回答所要時間は多肢選択問題のほうが短い可能性が示唆される。

5. 2種類のテストの回答時間差の考察

統計的に有意ではないが、ここで得られたデータからは、多肢選択問題のほうが回答所要時間が全体的に短くなる可能性が示唆されている。この要因として次の2つを我々は考えた。

- (1) 多肢選択問題のほうが正誤判定問題よりも回答時間が一般に短くなる理由がある（我々の仮説）
- (2) 類似の問題を2つ続けて回答した結果、後で回答した問題のほうが回答にかかる時間が短くなる（慣れ）

以下では後者の慣れの要因に関する検証を試みる。

5.1 慣れに関する考察

慣れによる影響を調べるのに、我々は先に回答した正誤判定問題と、後で回答した多肢選択問題との得点差に注目した。元々の出題自体が知識確認のやさしい問題なので、慣れによる影響が十分にあれば、後で回答したほうが回答者の学生がよくなる可能性が高いと考えられる。そこで、

正誤判定問題の得点 \leq 多肢選択問題の得点
である回答者について、回答所要時間の比較を試みた。この条件に該当する回答者は34名中22名である。この表3に対する t 検定の結果は p 値が0.012であり、上記仮説の後者

表3 回答所要時間の分布（正誤判定問題の得点 \leq 多肢選択問題の得点）

Table 3 Distribution of Elapsed Time with Some Score Condition

問題種別	平均値	最小値	第1四分位	中央値	第3四分位	最大値
正誤判定	114.7	19.0	58.0	135.0	162.0	181.0
多肢選択	82.6	22.0	38.5	74.0	111.0	181.0
合計	197.6	51.0	129.0	180.0	267.5	362.0
差	32.1	-61.0	-2.0	21.0	68.0	130.0
比1	1.89	0.36	0.99	1.29	2.15	6.20

(2) を支持している可能性が高い。つまりテスト問題で扱う知識への慣れが回答所要時間の短縮につながっている可能性が高い。

5.2 高得点群と慣れの影響

この回答所要時間が後で回答したテストのほうで短縮されるという傾向を、先に回答した

問題、つまり正誤判定問題において高得点（5問中4問正答）だった者に限って考察してみると、もちろん得点向上の余地はないので該当者は少なくなるが、該当者は34名中13名あった。このとき t 検定の p 値を求めると $p = 0.183$ になる。したがって先に回答した際の高得点群では時間短縮効果はあまりないと考えられる。

5.3 低得点群と慣れの影響

低得点群（5問中正答は3問以下；前述の問題で低得点ということは事前学習が不十分であるとの裏返しにもなるが）について、時間短縮傾向を調べると、 t 検定の p 値を求めると $p = 0.038$ であり、回答所要時間の平均の差は大きい（ p 値が高得点群/低得点群に分ける前と比べて大きくなったのは、標本の大きさに依存する自由度が小さくなったためである）。したがって低得点群の回答者が時間短縮に対して大きく寄与していると考えられる。

6. まとめ

- (1) 正誤判定問題に対する回答時間よりも多肢選択問題に対する回答時間のほうが短い可能性が示唆される。
- (2) これらの2種類の問題に対する回答所要時間の差は、我々の仮説「正誤判定問題への回答は選択問題への回答よりも時間がかかる」に原因を求めるよりも、後で受験する問題においては、慣れから回答所要時間が短縮されると解釈するほうが自然である。
- (3) 回答所要時間の2つ目のテスト受験における時間短縮は、とくに1つ目のテスト受験における低得点群において顕著である。

という特徴が、今回のテスト問題を通じた回答所要時間の比較からわかった。これらの観察結果は、テストの繰り返しをうまく編成するだけでも、学習への効果があがることを示唆している。

一方で、我々の仮説を確かめることで問題形式の特徴を有効に活かせるテストを構成するための情報を得るためには、少なくとも当面以下の性質をもった問題による測定を行う必要があると考えられる。

- (1) 正誤判定問題よりも多肢選択問題のほうが回答しやすいと、回答者の判断しやすい問題による測定を行う。
- (2) 正誤判定問題の回答に心理的ストレスを生じやすい、客観性の弱い設問の記述となるような問題による測定を行う。

なお、ここで論じたような記述式回答の抱える欠陥を指摘するものには、学生などの記述式回答における「わかった」、「よかった」などの感性語彙は学習効果の指標として有効でな

いとす研究⁵⁾ などがある。

(平成?年?月?日受付)

(平成?年?月?日採録)

参 考 文 献

- 1) たとえば「高卒認定試験で出題ミス, 正しい選択肢なし」, 産経ニュース 2010年11月14日 (2010).
- 2) 鈴木治郎: 学習支援のためのオンラインテスト利用に関する考察, CIEC 春季研究会 2010 報告集, (2010).
- 3) バイロン・リーブス他: 人はなぜコンピューターを人間として扱うか, 翔泳社 (2001).
- 4) 安藤進: Google に聞け! 英語の疑問を瞬時に解決, 丸善 (2004).
- 5) 藤田悠, 鈴木治郎: e ラーニングシステム上での教材レビューで得られた評価語彙の研究, コンピュータ&エデュケーション, Vol.26, pp.61-65 (2009).

鈴木 治郎 (正会員)

1958年生. 1990年上智大学大学院工学研究科数学専攻博士後期課程単位取得退学. 1990年信州大学医療技術短期大学部講師, 2002年同学教授. 2006年より信州大学全学教育機構教授(現職). 日本数学会, 情報処理学会, CIEC, JPCATS 各会員.