

# 情報リテラシーを対象とした理解度テストの生成と分析

## Question Generation and Analysis of Comprehension Tests for Information Literacy

松本 陽平†  
Youhei Matsumoto

村川 猛彦†  
Takehiko Murakawa

### 1. はじめに

和歌山大学システム工学部の1年生向け情報処理教育には情報処理Ⅰ・情報処理Ⅱの2科目があり、それぞれ前期と後期に開講の必修科目である。このうち情報処理Ⅰは情報処理リテラシーを学習する科目であり、学内計算機環境やインターネットリソースなどを適切に使用するための知識や、文書作成・データ管理・プレゼンテーションを行うためのアプリケーションの使い方、ハードウェア・ソフトウェアを含む情報システムの仕組みについて学ぶ。

各授業は、教員による解説と計算機を用いた演習課題で構成されており、復習を通じた、学習事項の定着の機会がほとんどなかった。そこで、時間外学習を通じて学んだ内容の見直しと振り返りができるような、問題セットの開発および学習支援環境の提供が望まれる。

そのような学習支援環境を、問題を理解度テストのような形で学習者に提供する場合、良いテストを整備し提供することが求められる。良いテストとは、次の2つの要素を満たすこととされる[1]。一つは信頼性であり、「テストのスコアは、常に受験者の能力の大小を言い当てている」と表現できる。もう一つは妥当性と呼ばれ、「テストで問われる内容と、実際に測定されるべき能力とがマッチしている」に対応付けられる。信頼性については、問題数を多く用意することにより実現できるのに対し、妥当性のほうは「学習者がどれだけ授業の内容を理解しているか」を確かめられる証拠を考え、それを問題1つ1つに当てはめてチェックする必要がある。

本研究では、問題の自動生成によって信頼性を、項目反応理論による問題の能力値推定によって妥当性を満足するようにし、情報リテラシーに関する問題セットの開発を試みた。

### 2. 準備

#### 2.1 理解度テストの実施状況

著者らが携わってきた情報処理科目における理解度テストについて、本稿で詳述する分も含めて説明する。著者の一人(村川)は2015年度の和歌山大学システム工学部改組に合わせこの年度より情報処理Ⅰ(主に情報リテラシーに関する内容、1年前期・必修)および情報処理Ⅱ(主にプログラミング入門に関する内容、1年後期・必修)の授業を2クラスずつ担当し、他の教員と連携して教材作成に関わってきた。

情報リテラシーに関する最初の理解度テストは、2015年度の情報処理Ⅱの授業初回に実施した。5択で10問の多肢選択式問題、および3問の記述式問題を作成し、紙面により受講者に解答させた。2016年度には5択問題を20問に増やし、同じ形式で解答させた。いずれも情報処理Ⅰの内

容の復習を狙ったものであり、正解・不正解は科目の成績に入れていない。

2016年度には、時間外学習を通じて学習内容の見直しと振り返りができる学習支援環境の提供を目指して、情報処理Ⅰ関連で133問、情報処理Ⅱ関連で100問を作成するとともに、Dockerを援用したWebアプリケーションとしてサービスを構築し、2017年1月の授業時間内外で情報処理Ⅱの受講者に使用してもらった[2]。これにより1人当たりの平均解答数や思考時間などの数量情報を得たが、1問当たりの解答者数は十分ではなく、各出題の良し悪しを検討するのは困難であった。

2017年7月に、20問を精選して情報処理Ⅰの授業中に解答してもらい、101名を対象として、項目反応理論に基づく分析を試みた(5.1節)[3]。また2018年7月に、一部問題を差し替えて当該年度の情報処理Ⅰの受講者に解答してもらい(5.2節)、129名の解答を得た。前期に実施するようになったのは、理解状況をより確実に測ることが期待できるほか、BYOD(Bring Your Own Device)の利用を含む授業内容の見直しにより、情報処理Ⅱの初回に前期科目の復習をする時間が確保できなくなったためである。

なお、著者の一人(松本)は2018年度前期の情報処理Ⅰのうち1クラスのティーチングアシスタントとして勤務し、機器利用のトラブルや演習課題の内容などに関する問い合わせに応じながら、受講者の学習状況を把握してきた。教材ならびに演習課題の作成や採点には関わっていない。

#### 2.2 項目反応理論

項目反応理論[1][4]は項目応答理論とも書かれ、英語のItem Response Theoryの頭文字からIRTと表記されることもある。正答率や素点に基づくテスト(古典的テスト理論と呼ばれる)では、正解率が高かった問題について、問題が易しかったからか、それとも解答者(「受験者」または「受検者」と書かれることもあるが、本稿では「解答者」で統一する)の能力が高かったからなのかを判断できない。それに対し、適切な計画を立ててテストを行い、項目反応理論を用いて分析を行うことで、解答者の能力に関する情報と、各出題(項目)の難易度などに関する情報とを、切り離して推定することなどが可能となる。

分析にあたり本研究では、行を解答者、列を出題として、正答には1、誤答には0を格納した「項目反応パターン」を作成し、後述する統計解析ソフトウェアへの入力とする。また項目の推定にあたり、識別力と困難度という2つの実数値に基づく2パラメタ・ロジスティックモデル(2PLM)を採用している。各解答者の能力値は1個の実数値で表される。識別力と困難度の値が得られれば、2次元平面上に項目特性曲線と呼ばれる曲線を描き、他の項目との比較ができる。項目特性曲線の例を図2-1に示す。横軸は能力値、縦軸は正答確率(その能力値の解答者集団のうちどれだけの割合が正答するか)を表す。番号1, 2, 3の曲線につい

て、識別力と困難度の値はそれぞれ(1, 0), (1, 2), (0.5, 0)であり、識別力が同一の場合(番号1と2)には曲線の形状は同一であるが、困難度が高い番号2は右に曲線が描かれる。識別力が異なる場合(番号1と3)には、曲線の傾きが異なる。能力値に応じて識別を行うという観点から、識別力の値は高く、項目特性曲線は中心に大きな傾きを持つ右上がりのS字を描くものがよい。それに対し困難度は高ければよいとは限らない。基本的知識を測る理解度テストにおいては、低い困難度は「きちんと聞いていたか」を把握するための設問として有用であると言える。

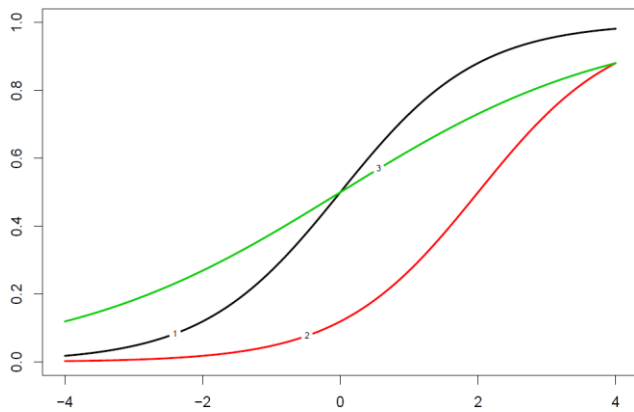


図 2-1 項目反応曲線の例

項目反応理論を用いることで得られる他の利点として、異なるテストの結果を同一の尺度に乗せ、比較可能にできる点が挙げられる。その手続きは等化と呼ばれる。異なるテストの比較のためには、異なるテストを共通の解答者が解答する「共通受験者計画」、テスト間に共通する問題が含まれている「共通項目計画」、係留テストと呼ばれるテストを用意して解答させる「係留テスト計画」などが考えられるが、本研究では共通項目計画を採用している。線形等化関数を用いた場合、一次変換の傾きと切片という2つの実数値を得ることができ、合わせて等化係数と呼ばれる。

項目反応理論に基づく分析に関して統計解析ソフトウェアはRのバージョン3.4.4を用いた。パラメタ推定を行うためにirtoolsパッケージを、また等化に関してはplinkパッケージを導入した。OSはUbuntu 18.04, CPUはIntel Core i7-7700K, 主記憶は32GBの計算機を使用し、対話型環境でコマンドを実行した。各ステップは瞬時(時間がかかるものでも1秒以内)に処理が完了した。

### 3. 関連研究およびテスト事例

文献[4]の「はじめに」では、項目反応理論に基づいて設計・運営されているテストが8つ並べられている。その3番目以降は、語彙・読解力検定、GTEC、日本語能力試験など、語学系テストで占められている。国際的な英語能力測定試験であるTOEFLも、同じページに例示されている。

語学系テスト以外にも、項目反応理論を用いたテストが普及しつつある。例えば「医療系大学間共用試験」[5]は、日本の医学部(6年制課程)の学生が受験し、その合格が、臨床実習を行う条件となっているテストの一つである。医療系大学間共用試験実施評価機構が運営しており、CBT(Computer-Based Testing)においては320の設問を6時間で解答させる。問題プールから、受験者ごとに問題が異なる問題をランダムに出題し、平均難易度に差がないよう調整が

なされている。2013年10月現在でプール問題数は2万を超えている。合否判定は各大学が実施するが、判定基準について「IRT標準スコア」の数値が設けられている。

また情報処理技術者試験の試験区分のうち、ITパスポート試験においては、2011年11月より解答方法をCBT方式に変更するのに合わせて、採点は項目応答理論に基づいて解答結果から評価点を算出している[6]。

項目反応理論を活用したテスト分析の研究についていくつか示す。まず、斉田[7]は、共通の情報が含まれていないテスト間に事後的に共通尺度を設定するという、事後的等化の手法を提案し、1995年度から2008年度までで約20万人の高校生が解答した「茨城県高等学校英語学力テスト」に適用して尺度値を求め、高等学校入学時の英語学力が年々低下していることなどを明らかにしている。

月原ら[8]では、九州大学情報工学部に入学したばかりの学生432名を対象に、高校で習得すべき基本的な数学の知識を問う35問を解かせたテストについて、IRT評価と素点評価とを比較検討した結果、学生の能力を正確に、また効率的に(少ない問題数で)推定するには、IRT評価がより有効であることを報告している。

尻無濱[9]は、管理会計知識を測定する尺度の試案を作成することを目的の一つとして、公認会計士試験の短答式試験をもとに、正誤問題による質問項目を作成し、予備調査・基礎分析により項目パラメタの推定に悪影響を与える可能性のある項目を除外したのち、大学の経営学部で会計を専攻する3・4年生の解答データをもとに項目プールを作成し、経営学部1年生について管理会計知識の能力推定を試みている。なお分析に使用した解答は仮想データであることが明記されている。

## 4. 問題の自動生成

### 4.1 概要

本研究で行われる問題生成の手順を図4-1に示す。まず、授業で指定している参考書[10]の索引ページをもとに、索引語の単語リストを作成しておく。こうして作成した単語リストの中から問題作成者に問題として利用する索引語を選択してもらい。その後、その単語が正解となる問題文と単語に関連する誤選択肢について、DBpedia[11]を利用して単語の要約と関連語を取得し、これらを編成して1つの問題とする。その後は出題にあたって差し支えないかを人手により確認し、必要に応じて修正する。

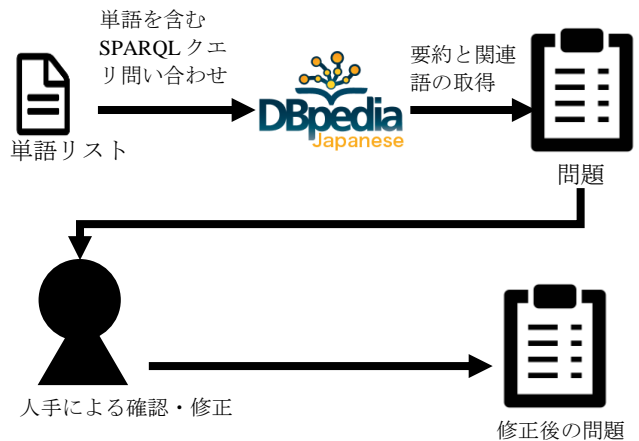


図 4-1 問題自動生成の流れ

## 4.2 索引語の取得

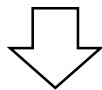
単語の取得については、参考書[10]の索引ページに記載されている索引語を対象に取得を行う。抽出したテキスト部分にはページ番号や点線からなるリーダー線が含まれているため、それらを除去し、単語ごとに改行することで単語リストを作成している。単語リストの作成について行うのは最初の一回だけである。

索引語は 585 語あり、そのうち 281 語（約 48%）について、2018 年度の情報処理 I の授業で使用（解説用スライドに記載）されていた。

## 4.3 問題文の取得

問題文の取得には Wikipedia の情報を LOD (Linked Open Data) として公開している DBpedia[11] から単語の要約を取得することで実現している。LOD の実現に用いられている RDF (Resource Description Framework) のデータモデルは、リソース情報を主語 (Subject)、述語 (Predicate)、目的語 (Object) の 3 つの要素 (トリプル) で表す。主語にはリソースを同定するための識別子 (URI) を使用する。述語には主語と目的語の関係を同定するための URI を使用する。目的語には、数値や文字列、日付などのリテラルか、リソースを表す URI を使用する。RDF (LOD) 用の汎用クエリ言語として SPARQL を利用し、DBpedia のエンドポイントにクエリを用いて検索することで、単語の要約を取得する。そのためクエリは、主語に検索したい索引語を入れた URI を指定し、述語に主語と目的語の関係が要約を表す URI を指定することで、目的語に索引語の要約のリテラルが現れるので、それを取得する。利用するクエリとして「ドメイン名を与えた場合の問い合わせ文およびその結果を図 4-2 に示す。

```
select distinct * where
{
  <http://ja.dbpedia.org/resource/ドメイン名>
  <http://dbpedia.org/ontology/abstract> ?o .
}
```



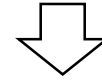
ドメイン名 (ドメインめい、domain name) は、IP ネットワークにおいて個々のコンピュータを識別する名称の一部。

図 4-2 問題文の問い合わせクエリ例

## 4.4 誤選択肢の獲得

誤選択肢の獲得の獲得には DBpedia を用いて、索引語の関連語を取得することで獲得している。その際 SPARQL クエリは主語に検索したい索引語を入れた URI を指定し、述語に主語と目的語の関係が関連語を表す URI を指定することで、目的語に索引語の関連語が含まれた URI が現れるので、そこから関連語となる部分だけを抽出することで、関連語の取得としている。利用するクエリの例を図 4-3 に示す。

```
select distinct * where
{
  <http://ja.dbpedia.org/resource/ドメイン名>
  <http://dbpedia.org/ontology/wikiPageWikiLink> ?o .
}
```



IP アドレス、レジストラ、インターネット、ホスト名... など

図 4-3 誤選択肢の問い合わせクエリ例

## 4.5 人手による点検および修正

自動生成した問題をそのまま解答者に出题するのは、問題が正しく作成されているかどうか不明瞭であるため、適切ではない。そこで、最後に人手による点検を行い、それで授業の復習として出题するには不適切だと考える問題を修正する。修正に関しては、問題文を一部わかりやすくするために修正したり、授業で出現しない語があれば、その選択肢を別のものに変更したりする。また、問題自体が不適切だと考える場合は、その問題を基に新たな問題が作成できるようにする。

## 4.6 生成例

提案手法によって作成した問題は (i) そのまま出題に利用できるもの、(ii) 授業実施者の判断により問題文や選択肢を変更したもの、(iii) 自動生成した問題をもとに授業実施者が問題文を新たに作成したものに分けられる。このうち (i) に該当するものを図 4-4 に示す。また (ii)、(iii) については修正前、作成前と修正後、作成後についてそれぞれ図 4-5、4-6 に示す。

Web ブラウザと Web サーバの間で HTML などのコンテンツの送受信に用いられる通信プロトコルは【      】である

HTTP  
SMTP  
TCP  
URL  
WWW

図 4-4 問題例

IP ネットワークにおいて個々のコンピュータを識別する名称の一部を【 】という。

ドメイン名  
ホスト名  
IPアドレス  
レジストラ  
ICANN



「wakayama-u.ac.jp」など、インターネットにおいて個々のコンピュータを識別する文字列を【 】という。

ドメイン名  
MACアドレス  
IPアドレス  
レジストラ  
コモンネーム

図 4-5 修正前と修正後の問題

OSI 参照モデルにおける 7 階層の内の第 3 層の事である【 】はエンド・ツー・エンド（起点から終点まで）のバケット配送に対する責任が有る。

ネットワーク層  
データリンク層  
トランスポート層  
アプリケーション層  
セッション層



TCP/IP モデルは、【 】の階層で構成される。

4つ  
3つ  
5つ  
6つ  
7つ

図 4-6 変更前と変更後の問題

## 5. 問題セットの作成および実施

### 5.1 2017 年度実施

出題する問題は 20 問とし、受検者（情報処理 I 受講者）は全て問題に解答することとした。前年度[2]に作成した 133 問の中から、15 問をそのまま採用し、別に 3 問については修正を加えて使用した。新たに 2 問を作成した。いずれも、4 節で述べた自動生成によるものではない。また出題内容がそれぞれ独立であるよう注意した。

解答にあたっては、[2]と同じ Docker に基づく独自の Web サービスに受講者の BYOD PC およびブラウザを用い

てアクセスする形で実施した。出題画面の例を図 5-1 に示す。サーバは筆者らの研究室にある 2 台の Linux PC で別々に稼働させた。学内環境ということもあり、数十人の同時アクセスにおいて通信上の不具合は特に見られなかった。授業終了後にそれぞれの PC からアクセスログを取得し、送信元 IP アドレスとアクセス時刻をもとに解答を獲得した。

### 問題1

pptや【 】は、PowerPointで編集や表示するためのファイルの拡張子である。

1 docx 2 xls 3xlsx 4 pptx 5 doc

図 5-1 2017 年度の出題画面

### 5.2 2018 年度実施

今年度実施した理解度テストでは、共通項目計画に基づく等化を行うことを念頭に置き、昨年度実施した理解度テストからいくつか問題を変更することとした。ただし出題は 20 問で、全てを解答してもらう方針は、変わらない。

6.1 節で述べる通り 2017 年実施において、識別力が低い項目が 2 問発生したためそれらを除外して、残り 18 問を 2:1 に分け、12 問をどちらの年度も解答し、残る 6 問は 2017 年度と 2018 年度とで異なる問題とした。その結果、変更するのは 2+6 で 8 問となった。この 8 問に関しては、4 節で述べた手法を用いて自動生成を試みた。そのうち 1 問はそのまま問題に使用した。5 問は人手による修正により問題文や選択肢を変更した。2 問は自動生成した問題をもとに、人手で問題文を作成した。2018 年度に解答しない 6 問は、2017 年度の 18 問の中からランダムに選んだ。

解答にあたっては、和歌山大学内で稼働している学習管理システム Moodle の「小テスト」機能を用いることとした。出題画面の例を図 5-2 に示す。この実施においても、受講者の BYOD PC およびブラウザを用いてアクセスしてもらい、通信上の不具合は見られなかった。授業終了後に「評価」の Excel ファイルを取得し、人物情報を取り除いて、項目反応パターンを得た。

### 問題 12

未解答

最大評点 1.00

▼ 問題にフラグを付ける

不特定多数からのアクセスを受け付けるWebアプリケーションにおいて、サービス提供例は、ロードバランサ、アプリケーションサーバ、【 】サーバといった機器で構成される。

1つ選択してください

- Web
- セキュリティ
- ファイアウォール
- クライアント
- データベース

図 5-2 2018 年度の出題画面

## 6. 理解度テストの実施結果

### 6.1 2017 年度実施の結果

2017 年 7 月の情報処理 I 最終回の授業で、2 教室の受講者に解答を依頼した。20 問全てに解答し、適切に記録が保

存された 101 名の解答をもとに、項目反応パターンを作成し、分析を行った。

各項目（問題）の正答率、および統計処理言語 R および `irt` ライブラリを用いて算出した困難度と識別力を、表 6-1 に示す。

表 6-1 2017 年度実施の問題ごとの状況

問題番号	正答率	項目除去前		項目除去後	
		識別力	困難度	識別力	困難度
1	0.455	0.569	0.331	0.545	0.344
2	0.267	0.626	1.742	0.637	1.716
3	0.297	0.411	2.174	0.441	2.038
4	0.515	0.663	-0.107	0.680	-0.105
5	0.257	0.966	1.291	0.967	1.289
6	0.228	0.577	2.263	0.585	2.234
7	0.426	1.249	0.294	1.287	0.287
8	0.683	0.618	-1.352	0.617	-1.355
9	0.267	0.391	2.668	0.409	2.557
10	0.525	0.144	-0.694		
11	0.475	0.369	0.274	0.364	0.278
12	0.267	0.933	1.261	0.884	1.313
13	0.535	0.621	-0.250	0.638	-0.246
14	0.317	0.699	1.209	0.717	1.183
15	0.871	0.918	-2.387	0.916	-2.390
16	0.693	1.174	-0.888	1.182	-0.884
17	0.950	-0.324	9.272		
18	0.248	1.263	1.126	1.207	1.158
19	0.624	0.482	-1.109	0.476	-1.122
20	0.723	0.432	-2.311	0.428	-2.334

「項目除去前」の識別力・困難度をもとに得られた項目反応曲線を図 6-1 に示す。横軸は能力値、縦軸は正答率である。

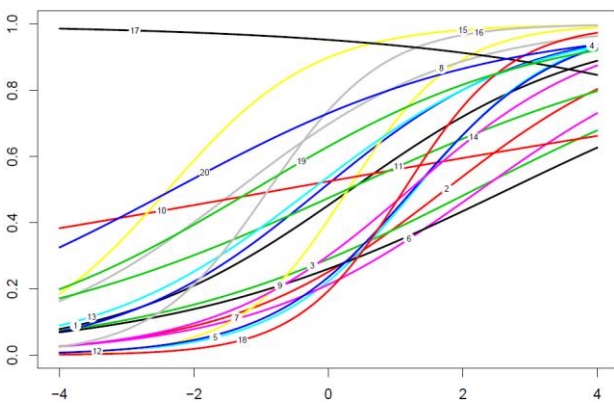


図 6-1 2017 年度の項目特性曲線

最も識別力の高かったのは問題番号 7 で、その問題文（正解をカッコ書きにし、誤選択肢を省略している。次節も同様）は「検索エンジンやデータベースにおいて、すぐに検索結果を返せるよう構成された情報を、（インデックス）」であった。

それに対し、右下がりの曲線（問題番号 17）、および緩やかな右上がりの曲線（問題番号 10）に関しては、項目反応モデルに基づく能力値の推定に適切とは言えない。そこ

でこれらを除外し、18 問のみを対象として識別力と困難度を算出し直したものが、表 6-1 の「項目除去後」の列となる。値に若干の変動はあるが、項目反応曲線は図 6-1 と大きく変わらなかった。

## 6.2 2018 年度実施の結果

2018 年 7 月の情報処理 I の第 14 回の授業で、2 教室の受講者に解答を依頼した。20 問全てに解答し、適切に記録が保存された 129 名の解答をもとに、項目反応パターンを作成し、6.1 節と同様に分析を行った。2018 年度の解答状況を表 6-2 に示す。また「項目除去前」の識別力・困難度をもとにした項目反応曲線を図 6-2 に示す。

表 6-2 2018 年度実施の問題ごとの状況

問題番号	正答率	項目除去前		項目除去後	
		識別力	困難度	識別力	困難度
1	0.605	0.119	-3.592		
2	0.589	0.464	-0.817	0.466	-0.815
3	0.302	0.892	1.089	0.825	1.155
4	0.519	1.553	-0.080	1.647	-0.080
5	0.481	1.131	0.081	1.137	0.079
6	0.248	0.529	2.221	0.529	2.223
7	0.395	-0.020	-20.748		
8	0.806	0.339	-4.302		
9	0.519	0.660	-0.131	0.697	-0.127
10	0.512	1.101	-0.058	1.082	-0.060
11	0.519	0.220	-0.357		
12	0.171	1.939	1.276	1.982	1.264
13	0.612	0.781	-0.666	0.704	-0.725
14	0.434	-0.124	-2.153		
15	0.829	0.339	-4.777		
16	0.736	0.897	-1.333	0.907	-1.323
17	0.744	0.480	-2.337	0.479	-2.344
18	0.729	0.456	-2.267	0.410	-2.503
19	0.643	-0.245	2.441		
20	0.698	0.679	-1.356	0.673	-1.367

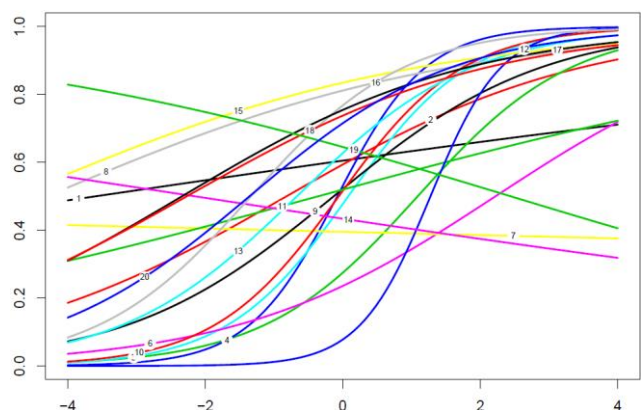


図 6-2 2018 年度実施の項目反応曲線

最も識別力の高かったのは問題番号 12 で、その問題文は「不特定多数からのアクセスを受け付ける Web アプリケーションにおいて、サービス提供側は、ロードバランサ、アプリケーションサーバ、（データベース）サーバといった機器で構成される。」であった。この問題は 2017 年度にも

同じ番号で出題しており、そこでの識別力の高さは20問中5位であった。

図 6-1 よりも図 6-2 のほうが、不適切な曲線が多いように見える。そこで識別力と困難度の値に数値基準を設けて除外することとした。具体的には、識別力が 0.3 未満、または困難度の絶対値が 4 を超えるものを除外する。この基準は、2017 年度実施分で除外したものについても当てはまる。最終的に 2018 年度実施分においては、7 問を除外することとなり、残りの 13 問を対象とし識別力と困難度を再計算した結果が、表 6-2 の「項目除去後」の列となる。

### 6.3 等化による年度間比較

共通項目計画に基づく等化において、対象とする問題番号は、2017 年度・2018 年度共通で問題番号 4, 12, 13, 16, 20, および 2017 年度の問題番号 3 と 2018 年度の問題番号 6 の合計 6 問となった。基準集団を 2017 年度の理解度テスト解答者とし、等化法には Mean & Sigma を指定して、R の plink パッケージを用いて計算した結果、等化係数として、傾き  $A=1.066956$  および切片  $K=-0.035072$  を得た。

この値を用いて 2018 年度実施分について等化後の項目パラメタ（識別力および困難度）ならびに各解答者の能力パラメタを算出した。2017 年度実施および 2018 年度実施について、正答率および能力値の散布図を図 6-3 に示す。この図において横軸は正答率（2017 年度は正答数÷18, 2018 年度は正答数÷13）、縦軸は等化後の能力値を表す。2017 年度の全問正解者は 1 名、2018 年度は 2 名であった。

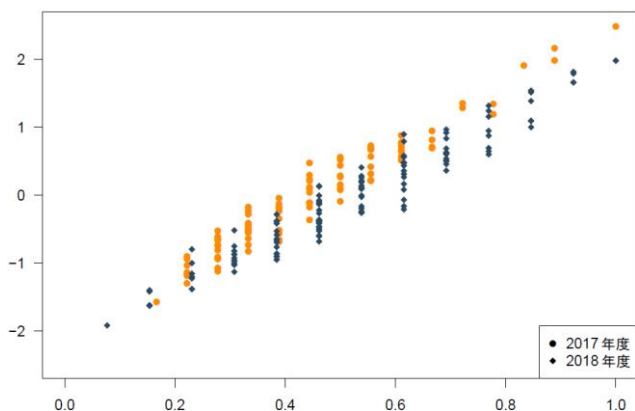


図 6-3 正答率と等化済み能力値の散布図

## 7. 考察

自動生成によって作成した問題の 8 問のうち分析に除去された問題は 1 問であり、学生の能力を測る問題として採用できること推定でき、自動生成の結果としてはおおむねうまくいった。

2017 年度の正答率が低い問題番号 6 と 18 を識別力・困難度で見ると問題番号 6 の識別力は 0.577 で困難度は 2.263 であり、問題番号 18 の識別力は 1.263 で困難度は 1.126 であった。これら 2 つの問題では識別力が大きく異なっており、項目反応理論を用いた分析では識別力が高い問題番号 18 は良問、問題番号 6 は悪問であると判断することができた。このような比較検討は、正答率のみでは行うことが困難であり、項目反応理論の有用性を示唆するものとなる。

等化を行った結果に関しては、図 6-3 を見ると 8 問正解で、能力値が最高である解答者と 11 問正解で能力値が最低の解答者ほぼ同等の能力値となっている。このことから、正解数に基づき得点を算出する素点方式とは別の方法で、より詳細に解答者の能力値を推定できると考えられる。また、図 6-3 では 2017 年度と 2018 年度を同じ尺度で測ることができ、目視では、2017 年度の解答者のほうが能力値が全体的に上位であると見ることができる。

今後の課題として、より正確に測るためにより多くの人のシステムの利用が必要であると考えられる。また、2017 年度は 2 問、2018 年度は 7 問を分析対象として除去する必要があった。これに関しては授業担当教員が問題作成者を兼ねていたことが 1 つの原因であると考えられる。

## 8. おわりに

本研究では学習者に良い理解度テストを提供することを目的として、信頼性の実現のために問題の自動生成を、また妥当性の確保のために項目反応理論による識別力・困難度・能力値推定に関する分析を行ってきた。その結果、自動生成によって問題を生成することができ、能力値などの推定によって「どの問題が授業を理解しているか」ということを判断する基準になりえる値を得ることができた。

今後の課題として、自動生成手法の見直しが挙げられる。現段階で今回自動生成によって作成された問題のうち、そのまま量できた問題はたった 1 問であり、それ以外は教員による修正・新規作成が行われた。これでは、信頼性が確保されるだけの問題を作るために大きな負担が教員にかかることになる。情報リテラシーとして学ぶべき内容に則した問題を作成できるように手法を見直す必要がある。

そして、より正確な分析を行うために多くの解答が必要であるため、継続的な解答と年度間比較を行っていきたいと考えている。

**謝辞** 理解度テストの解答に協力してくださった受講生の皆様に感謝申し上げます。2017 年度の理解度テストの実施にあたっては登田純太氏の協力を得ました。合わせて御礼申し上げます。

## 参考文献

- [1] 光永悠彦: テストは何を測るのか, ナカニシヤ出版 (2017).
- [2] 松本陽平, 藤原敬介, 村川猛彦: 情報処理教育を対象とした e ラーニングシステムの構築, 情報知識学会誌, Vol.27, No.2, pp.155-160 (2017).
- [3] 村川猛彦, 登田純太, 松本陽平: 項目反応理論に基づく大学初年度向け情報リテラシー科目の理解度テスト分析, 2018 年電子情報通信学会総合大会, 情報・システム講演論文集 1, p.155, D-15-23 (2018).
- [4] 加藤健太郎, 山田剛史, 川端一光: R による項目反応理論, オーム社 (2014).
- [5] 高木康: 共用試験実施機構と共用試験, 第 1 回医師国家試験改善検討部会資料 3. <https://www.mhlw.go.jp/file/05-Shingikai-10803000-Iseikyoku-Ijika/0000048625.pdf> (2018-07-27 参照).
- [6] 情報処理技術者試験・情報処理安全確保支援士試験 [https://www.jitec.ipa.go.jp/1\\_04hanni\\_sukiru/\\_index\\_hanni\\_skill.html](https://www.jitec.ipa.go.jp/1_04hanni_sukiru/_index_hanni_skill.html) (2018-07-27 参照)

- [7] 齊田智里: 英語学力の経年変化に関する研究—項目応答理論を用いた事後的等化法による共通尺度化—, 風間書房 (2014).
- [8] 月原由紀, 鈴木敬一, 廣瀬英雄: 項目反応理論による評価を加味した数学テストと e-learning システムへの実装の試み, コンピュータ & エデュケーション, Vol.24, pp.70-76 (2008).
- [9] 尻無濱芳崇: 管理会計知識の測定: 項目反応理論の会計研究への応用, 山形大学紀要 社会科学, Vol.48, No.1, pp.49-66 (2017)
- [10] 岡本敏雄(監修), 安齊公士他(著): よくわかる情報リテラシー, 技術評論社 (2012).
- [11] DBpedia Japanese. <http://ja.dbpedia.org/> (2018-07-27 参照).