

# 「情報科」の大学入試実施のための CBT システムの開発とその検討

西田 知博 (大阪学院大学), 植原 啓介 (慶應義塾大学),  
角谷 良彦 (東京大学), 中野 由章 (神戸市立科学技術高等学校)

現在、文部科学省大学入学者選抜改革推進委託事業「情報学的アプローチによる「情報科」大学入学者選抜における評価手法の研究開発」において、「思考力・判断力・表現力」を評価する「情報科」CBT システムを開発している。ここでは、2017 年度実施した試行試験用システムと大学1年生および高校生を対象に行なった試行試験の結果と分析、今後の導入を考えている出題の枠組みについて紹介する。2017 年度の試験は大学生が 176 名、高校生が 1406 名受験した。大学生と高校生では得点に差はあるものの、得点傾向は類似した結果が得られた。また、システムの操作性に関しては概ね高評価であったが、従来の紙を使った試験と比べてどちらを好むかという問いにはその評価が分かれる結果となった。

## 1 はじめに

本論文では、文部科学省大学入学者選抜改革推進委託事業「情報学的アプローチによる「情報科」大学入学者選抜における評価手法の研究開発」において、「思考力 (Thinking) ・ 判断力 (Judgement) ・ 表現力 (Expression)」 (TJE) (久野, 2017) を評価する「情報科」CBT システムに関する研究開発について、2017 年度実施した試行試験用システムとその結果を紹介する。また、2018 年度試行試験などに導入を考えている独自の出題の枠組みなど、TJE 評価のために検討している CBT システムについて論じる。

## 2 2017 年度実施 CBT システム

本事業の目的は、情報分野での TJE を評価するための問題作成方法について検討すると同時に、そのための CBT システムの検討である。この目的に向け、2017 年 7, 8 月に大学 1 年生、2018 年 2, 3 月に高校生を対象に CBT において試行試験を実施した。試行試験においては、(1)情報分野において TJE を評価できる問題作成が可能であることの検証、(2)大学入試において CBT を実施する際の注意事項の洗い出しと検証を行なった。本論文では、(2)について検証する。

### 2.1 システムの検討

システムについては、すでに行われている CBT についての調査・ヒヤリングを行った上で、出題したい問題の枠組みや解答形式が実現可能かという基準で検討した。

まず、システムの開発については、独自システムを開発するか、既存のシステムを拡張するかの選択肢がある。既存のシステムとしては、QTI (IMS, 2015) に準拠した TAO (Open Assessment Technologies, 2018) や適応型テストのプラットフォームである

Concerto (University of Cambridge, 2018) などを検討した。今回は、4.1 節で紹介している Table World のような思考力を計測するための出題の枠組みの概念実証ができるシステムを少ない工数で実装するために、独自システムの開発を選択した。

次に、問題のデータ形式や解答形式であるが、プロダクトとしてシステムを開発する場合は相互利用性を考え、広く用いられている QTI などの規格に準拠すべきであるが、概念実証を行うことが主目的である本事業のシステムとしては工数がかかり過ぎると考え、情報入試研究会が作成してきた過去の「情報入試全国模試」(中野他, 2014; 谷 他, 2016) が実施可能であることを条件として、CBT ならではのインタフェースを加えたものを開発することとした。

2017 年度の大学での試行試験の実施形態から、同時実施の人数は多くても 100 名程度と考え、ロードバランサーは使わずアプリケーションサーバとデータベースの 2 台構成とし、PHP と MySQL を使ったシステムとした。また、パソコンを用いてキーボードとマウスで解答し、タブレットの利用等は考慮していない。

### 2.2 システムの概要

図 1 が開発した 2017 年度用 CBT システムの画面例である。このシステムは Web アプリケーションとして作動し、受験者はパソコン上でブラウザを動作させて受験する。「情報入試全国模試」と同様の出題を行えるようにするため、問題に大問、中間、小問の区別を付けることができるようになっている。解答には、以下の形式を用いることができるようにした。

#### 選択型

1 つしか選択できないラジオボタンと複数選択可能なチェックボックスを用意し、それぞれ横並びか縦並びに配置可能とする。チェックボックスは、選択できる



図 1 2017 年度 CBT システム

個数を制限可能にした。

### 穴埋め型

自由記述またはプルダウンメニューの選択が可能である。また、解答欄への番号付与が可能であり、同一番号の空欄は、1つの欄に入力があれば、それに連動して他の欄も同じ内容が表示される。

### 短冊型

図 1 の問題に示すように短冊の内容を選択して解答欄にドラッグアンドドロップで並べる。短冊は途中挿入や並び替え、削除は自由に可能である。また、並べる個数の制限も可能である。この形式を用いて「情報入試全国模試」の短冊形プログラムの問題も出題できる。

### 記述型

単語や短文で解答を入力する。文字数制限がある問題では、現在の入力文字数がリアルタイムに表示され、解答のガイドとしている (図 2)。

なお、2017 年度のシステムでは図表はすべてイメージファイルを埋め込む形とした。CBT ならではのインタフェースとしては、短冊型解答形式の用意や穴埋め問題での空欄の連動に加え、図 1 左のナビゲーションフレームで、解答済みの問題番号に色付きのマークを示したり、各問の先頭にある見直しチェックを入れた問題が確認できる機能を提供している。



図 2 入力文字数表示

## 2.3 システムの構成

本年度システムは以下の 3 つの機能部で構成される。

### (1) 作問機能部

試験の各問題が異なる担当者によって作成されることを想定し、大問単位で試験問題を XML ファイルとして記述した大問定義書と、複数の大問定義書を統合して全体の試験問題を XML ファイルとして記述した試験問題定義書の記述内容を確認するための機能を提供する。

### (2) 試験機能部

試験会場で実施に関連する以下の機能を提供する。

- ・試験問題定義書投入
- ・受験者登録
- ・試験会場管理 (受験者管理, 試験開始・終了管理)
- ・受験

### (3) 採点・集計機能部

試験機能部で受験者が解答した結果の出力、システム外部で採点した結果の集計を行う。自動採点が可能な問題に関しては、問題定義書に書かれた正解情報を用いて採点する。また、自動採点ができない記述問題などは出力された結果 (CSV フォーマット) を採点者が採点し、得点を書き込み再度アップロードすることによって自動採点の結果と合わせて集計する。

前述の通り、本システムでは問題定義を XML で記述する。個々の問題を作成している出題者も XML によって問題を記述しなければいけないが、これは負荷が大きい。現在、問題文中の個々の記述では Markdown が使えるが、小問など問題構造の定義が Markdown のような簡易な形式で記述できるようにすることが必要であるので、現在、検討を進めている。

採点に関しては、選択や短冊の並び替えで解答する問題は順不同や組み合わせへの対応を行い自動採点する。また、単純な記述解答はパターンマッチで採点で

きる。その他の記述の解答に関しては手動採点とする。

### 3 2017 年度試行試験

2017 年度の試行試験は、大阪大学と東京大学の1年生を対象に 2017 年 7、8 月に実施した（大阪大学他、2017）。また、公募に応募してきた高等学校の生徒を対象に 2018 年 2、3 月に実施した。大学における受験者は 176 名（大阪大学 71 名、東京大学 105 名）で、文系理系の内訳は、文系 76 名、理系 99 名、不明 1 名である。なお、すべての受験者は、前期の授業として何らかの情報系の授業を受講している。高校における受験者は大学進学率が比較的高い 11 校 1406 名で、国立中等教育学校が 2 校 91 名、公立高校が 6 校 1155 名、私立高校が 3 校 159 名という内訳である。実施学年は 1 年 825 名、2 年 431 名、不明(1 年または 2 年)150 名であり、学校設定科目を含む「情報」の科目を履修した上で受験している。試験実施と同時にアンケートで尋ねた進路に関する「現在の志望を教えてください。」という問いに対しては、文系志望 640 名、理系志望 507 名、その他・無回答 259 名という内訳であった。

#### 3.1 出題問題の概要

大学において行なった試験は大問 4 問で構成され、時間は 60 分であった（大阪大学 他、2017）。その概要を以下に述べる。

第 1 問は小問集で、4 問で構成される。問 1 はネットワークに関する用語と説明の組を 4 つ作らせる問題で、「A は B である」の組を選ばせる関連的思考力を問うものである（図 1）。問 2 はルールを与えた特殊な演算を用いた計算の結果を数値で答えさせる問題で、読解的思考力を問うものである。問 3 は 2 進の数の計算の虫食い部分を答えさせる問題で、推論的思考力を問うものである。問 4 は線でつながれた頂点（グラフ）上への石の置き方が何通りあるか答えさせる問題で、読解的思考力や推論的思考力を問うものである。

第 2 問は表現力や推論的思考力を問うアルゴリズムに関する問題で、小問 2 つで構成される。問 1 は  $7 \times 7$  の格子点に印を描く問題で、(1) は条件から描かれるパターンを選ぶもの、(2) は与えられたパターンを描くために必要な条件式の数値を選ばせるものである。問 2 は 2 次元空間での二分探索に関する問題で、(1) は探索途中での状況を問うもの、(2) はアルゴリズムを短冊型で答えさせるものである。

第 3 問は情報社会系の長文問題で、判断力や読解的思考力を問うものである。SNS をテーマにした問題

で、小問 3 つで構成される。問 1 はプライバシー侵害の行為を選択する問題である。問 2 は問題行為と理由の妥当な組み合わせを個数は指定せずに選択させる問題と、文に書かれたトラブルに関する理由と取るべき行動について短文（20 字以内）で記述させる問題である。問 3 は問題行為の例と理由をそれぞれ 20 字以内の短文で 2 組記述させる問題である。

第 4 問はプログラミングに関する問題で、表現力や読解的思考力を問うものである。整数の集合に関する問題で、小問 3 つで構成される。問 1 は問題を読み取らせ、条件を満たす整数をすべて選択するものである。問 2、3 はステップに分け、短冊を並び替えてプログラムを作成させる問題である。

高校においては、授業 1 校時（45～50 分）で試験とアンケートが行えるように、大学で出題した問題のうち、第 1 問の問 1、2、4、第 2 問の問 1、第 3 問の問 1、2 を選択し、配点を調整して 50 点満点、35 分の試験とした。

#### 3.2 試験結果

大学における試行試験の各大問の平均点と合計の平均・標準偏差(SD)を表 1 に示す。全体（文理不明を含む）の平均点は 55.9 であり、理系と文系には約 15 点の差がある。各問題を見ると、文系では情報社会系の第 3 問の得点が他の問題に比べて高くなっている。一方で、プログラミングを行わせる第 4 問は理系と文系で大きな差がつき、図 3 に示すようにばらつきが大きいこともわかる。また、小問を集めた第 1 問はばらつきは少ないが、文理で 4 点の差がついている。アルゴリズムを問うた第 2 問に関しては、下位 25% の得点と平均点が一致するという結果になっている。これは、格子点に印を描く問 1 が全問正解で、2 次元空間での二分探索を扱う問 2 が 0 点であった受験者が約 4 割居たためである。

高校における試行試験の各大問の平均点と合計の平均・標準偏差を表 2 に示す。全体（文理不明を含む）の平均点は 23.0（50 点満点）であり、理系志望と文系志望には大きな差はない。各問題を見ると、小問を集めた第 1 問は 25 点満点で平均点が 6.1 であり、得点率が最も低い。表 3 に各小問の平均点（高校も大学と同じ配点で求めたもの）を示すが、第 1 問の問 1（1-1）、問 4（1-4）の得点率が低く、特に高校生の得点率が低いことがわかる。1-1 はネットワークに関する知識面の差が要因になっている可能性がある。また、1-4 は石の置き方を考える問題で、すべての場合をみれなく見つけることは高校生には難易度が高かった



ことが伺える。第 2 問は中央値が満点となる (図 4) という結果になっており、簡単すぎる設問であったことが伺える。情報社会系の第 3 問は、短文の作文を含む問 2 (3-2) が大学生と比べて低い結果となっている。

表 1 大学における試験の平均と標準偏差

	1	2	3	4	計	SD
文系	10.4	13.7	16.0	7.5	47.7	14.3
理系	14.6	16.0	15.2	16.4	62.3	17.0
全体	12.8	15.0	15.6	12.5	55.9	17.4

(各問 25 点の 100 点満点)

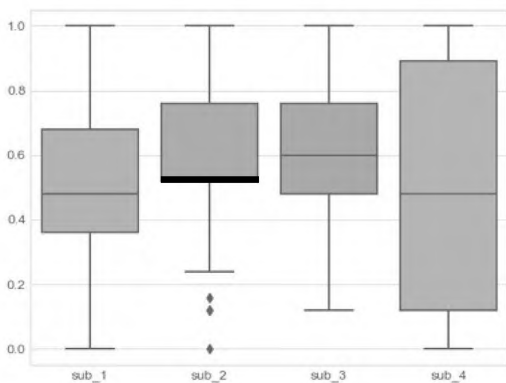
表 2 高校における試験の平均と標準偏差

	1	2	3	計	SD
文系	5.9	6.8	9.6	22.3	7.8
理系	6.6	8.0	10.2	24.9	7.2
全体	6.1	7.1	9.8	23.0	7.9

(第 1 問 25 点, 第 2 問 10 点, 第 3 問 15 点, 50 点満点)

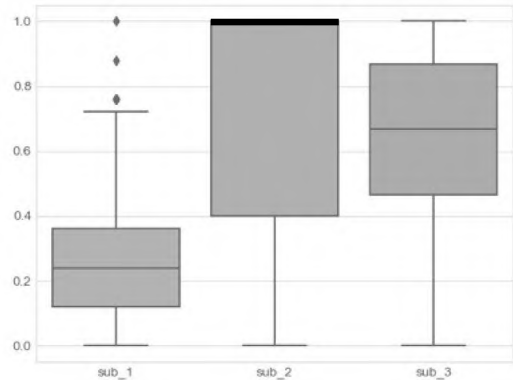
表 3 小問の平均点

	1-1	1-2	1-4	2-1	3-1	3-2
大学	2.83	4.40	2.97	9.42	2.95	7.72
高校	1.04	3.58	1.13	7.10	2.69	5.67
配点	8	6	8	10	3	10



(sub\_?は大問第?問を表す)

図 3 得点のばらつき (大学)



(sub\_?は大問第?問を表す)

図 4 得点のばらつき (高校)

大問間と大問と総得点の相関を調べると、大学においては、大問間の相関は第 3 問のみ他との相関が 0.03~0.12 とほとんど相関がなく、他の各問は 0.40~0.50 と互いにやや相関を持つという結果になった。総得点との相関においても、第 3 問は 0.34 と弱い相関で、第 1, 4 問はそれぞれ 0.75, 0.84 と強い相関を持つという結果になった。第 3 問は記述問題を含む情報社会系の長文問題であり、文理の平均点に差がなく、ちらばりも小さいという結果と合わせて考えると、他の問題とは異質であることが伺える。一方で、第 1, 4 問、特にプログラミングを扱う第 4 問は受験者の差が大きく、それが総得点の差に影響を与えていることが伺える。高校においては、大問間の相関は 0.20~0.31 と低く、各大問と総得点との間には 0.69~0.73 と同程度の相関があるという結果となった。

図 5 は大学生の得点のヒストグラムである。2 つのピークがある分布となったが、文系と理系それぞれのピークが現れていると考えられる。ただし、文系、理系それぞれもきれいな山状の分布にはなっていない。図 6 は高校で行なった試験と同じ問題の大学生の試験結果を配点も同一にして得点分布を比較したものである。全体に分布は大学生の方が得点が高い結果となっているが、双方ともきれいな山状の分布になっている。

### 3.3 アンケート結果

受験直後に主に CBT システムについて、アンケートを行った。図 7 に集計の抜粋を示す。

「操作はスムーズにできましたか?」に関しては、大学生、高校生ともに 9 割以上が「非常に操作しやすかった」「操作しやすかった」と答え、大学生では「非常に操作しにくかった」という回答はなかった。操作性に関しては、概ね問題なかったと考えられる。

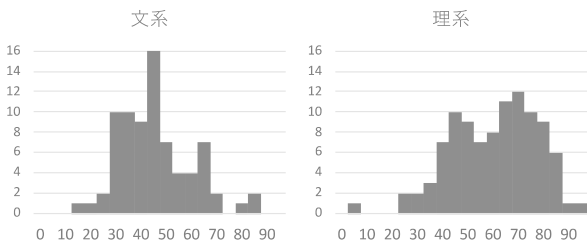
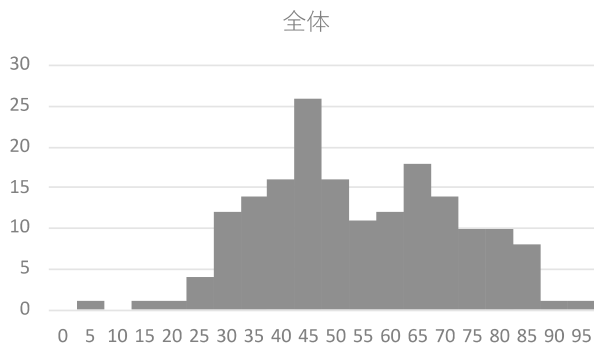


図 5 得点分布 (大学)

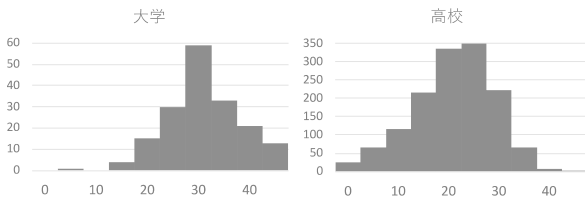


図 6 得点分布 (高校生試験との比較)

今回のシステムで提供したサポートインタフェースである「解答見直しのためのチェック」を活用したかという問いに関しては、大学生の半数以上が「活用した」と回答したのに対し、高校生は3割に留まり、約36%が「気づかなかった」と回答している。

紙での出題(PBT)とコンピュータでの出題(CBT)のどちらが良いかを問う質問に対しては、その答えが割れる結果となったが、大学生の方が CBT の方が良いという答えが若干多いという結果であった。大学生の回答を文理で分析すると、CBT を選択した割合が文系 58%、理系 46%となり、文系が CBT を好むという結果になった。自由記述をあわせて見ると、CBT に対してネガティブな回答としては、

- ・メモは紙、入力は計算機というのが煩わしい
- ・紙だと問題に線を引ける
- ・目が痛い、疲れる
- ・タイピング能力差別だ

などがあつた。一方で、ポジティブな回答としては、

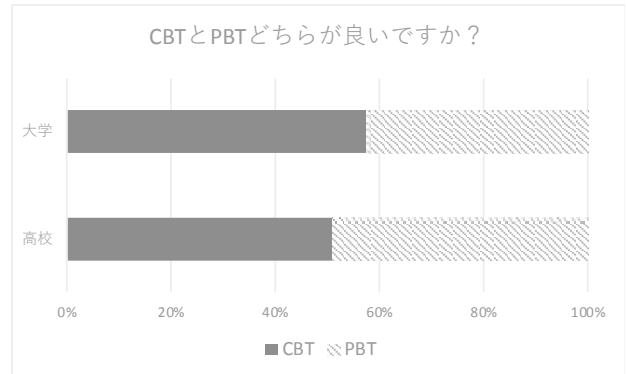
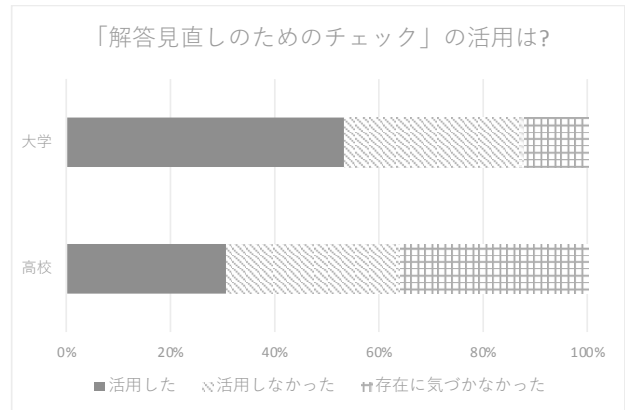
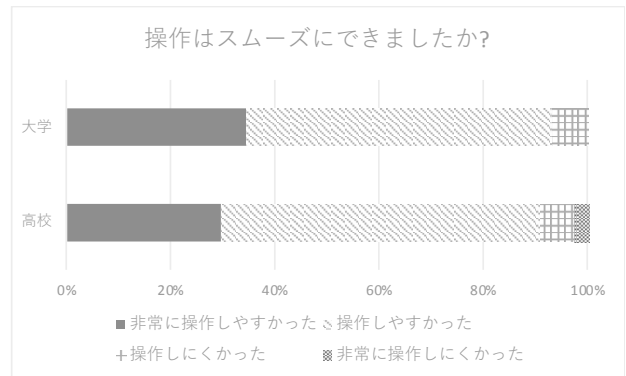


図 7 CBT システムの評価

- ・ CBT は作文がやりやすい
- ・ 記述で文字数をカウントしてくれるのが良い
- ・ 字が下手なのでありがたい
- ・ かさばらず、ページをすぐに見つけられる
- ・ 問題の下に解答欄があるのが良い

などがあつた。理系に PBT の方が良いという答えが若干多いことの1つの要因として、計算は計算用紙でなければいけないことに対する不満があると考えられる。一方で、作文問題の解答に関しては、図2のように入力文字数を表示する機能が評価されており、文系に CBT の方が良いという答えが若干多くなった要因の1つとして考えられる。

この他「選択肢問題や計算問題は紙、プログラミン

「CBT だ」や「どうせならタブレットで」という回答もあった。また、「CBT でやるならプログラム等を動かせるようにすべき」という回答もあったが、これは次期システムで実現している。

### 3.4 2017 年度 CBT システムの評価

2017 年度夏に実施した大学 1 年生を対象とした試行試験では、システムのトラブル等は発生しなかった。その点では、システムは概ね想定通りに動作し、またアンケート結果でも操作しやすかったとの回答が大半を占め、システム設計に大きな瑕疵はなかったことが確認された。一方で、CBT については賛否両論の結果となっている。これは、今回のシステムが PBT を前提として作題されたものを、ほぼそのままの形で CBT に搭載したためと考えられる。このことから、CBT を実施する際には、CBT を前提とした作問が必要であると考えられる。また、高等学校における試行試験はブラウザが推奨のものでない場合もあったが、解答中にシステムが操作できなくなったトラブルが 3 件発生したのみで留まった。

## 4 今後のシステム

前節で紹介したシステムは、2017 年度に行った試行試験のためのものであるが、現在は TJE を評価する独自機能を加えた CBT システムを開発し、2018 年度の試行試験を行なっている。ここでは未実装のものも含めて、システムの開発で検討しているものを紹介する。

### 4.1 出題の枠組みの検討

2017 年度の CBT での試行試験は、これまでの紙ベースの試験に準ずるものであったが、2018 年度の試行試験では、CBT ならではの機能を使って次のような作題をし、TJE を測定することを検討した。

#### 手順やプログラムを動かして検証できる問題

プログラミングやアルゴリズムに関する問題は、今回は短冊を並べて解答する形式で出題しているが、解答したプログラムや手順を実際に動作させ、その結果を確認できるような出題形式が考えられる。このような出題の枠組みでは、単なる最終解答だけではなく、解答の過程も見ることにより、TJE の評価ができると考える。解答形式も、図 8 のようなブロックプログラミングエディタを使い 2018 年度試験を実施している。



図 8 ブロックプログラミングエディタ

### Table World

Table World は二次元テーブルのデータの変化から思考力を問う出題の枠組みである。この枠組みでは、プログラム実行前後のデータ例がいくつか与えられる(図 9)。また、入力例を受験者が与えて、対応する出力を確認することができる。

このテーブルを元にした設問としては、入力から出力を答えるものや、入力と出力の関係からプログラムがどのようなものであったかを考えさせ、ブロックなどを使ってプログラムを作成させるものなどが考えられる。このような問題では、作成したプログラムを動作させることにより、出力結果を模範解答の出力と比較することによってプログラムを修正可能とし、その過程を評価することも可能である。

入力	出力																		
<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>11</td><td></td></tr></table>	1	2	11		<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>13</td><td></td></tr></table>	1	2	13											
1	2																		
11																			
1	2																		
13																			
入力	出力																		
<table border="1"><tr><td>1</td><td>2</td><td>3</td></tr><tr><td>11</td><td></td><td></td></tr><tr><td>298</td><td></td><td></td></tr></table>	1	2	3	11			298			<table border="1"><tr><td>1</td><td>2</td><td>3</td></tr><tr><td>13</td><td></td><td></td></tr><tr><td>2100</td><td></td><td></td></tr></table>	1	2	3	13			2100		
1	2	3																	
11																			
298																			
1	2	3																	
13																			
2100																			
入力	出力																		
<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>12</td><td></td></tr><tr><td>23</td><td></td></tr></table>	1	2	12		23		<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>14</td><td></td></tr><tr><td>25</td><td></td></tr></table>	1	2	14		25							
1	2																		
12																			
23																			
1	2																		
14																			
25																			

図 9 Table World のデータ例

### ゲームブック形式の問題

医学系 CBT では「順次解答連問形式」と呼ばれる同一の症例について、関連する 4 つの設問が連続して問われ設問形式がある。この形式は、医療面接から始まり、診察、診断に至る思考過程とその背景にある病態の理解を問うような作問が行われている(吉岡, 2014)。この形式をさらに進め、RPG あるいはゲームブックのような形式で対話的な問いかけをし、トラブルシューティングやプログラムのテストなどを行わせ、そのプロセスを評価する出題の枠組みが考えられ、2018 年度試験に一部導入している。

## データベース操作や設計の問題

CBT の特性を用い、データベース操作を通じて大量のデータを処理させる問題が考えられる。また、操作だけではなく提供したい検索インタフェースに応じたクエリを考えさせるような出題も可能である。

また、既存の CBT は IRT を用い、小問によって評価を行うものが多い(西田他, 2017)。しかし、それらは知識や判断を問うものが多く、思考力や表現力を十分に評価できるかは不明である。本事業では、上に挙げた CBT ならではの設問と組み合わせることにより、そのような小問形式の試験でも TJE を評価することが可能かを検証する。

また、IRT を用いるためには多くの小問を作成する必要がある。このために、数値などを置き換えるだけで多数の問題が生成できるようにする作題法も検討している。

## 4.2 採点

大規模な試験実施を考慮した場合、自動採点や採点補助の提供が必須となる。問題となるのは、記述問題の採点である。数値や語句レベルはパターンマッチで採点できるが、自由記述の場合、自動採点は難しい。キーワードを抽出してアンダーラインを引いたり、出現数をカウント、模範解答との類似度の算出など採点補助機能の提供を検討している。また、機械学習を用いての前処理を行い、採点補助をすることも考えられる。また、プログラムを答えさせる問題ではテストケースをどのように提供するかを検討する必要がある。

## 5 まとめ

本稿では、文部科学省委託事業「情報学的アプローチによる「情報科」大学入学者選抜における評価手法の研究開発」の中で、「思考力・判断力・表現力」を評価する CBT システムの研究開発について、2017 年度に実施した試験用システムと、それを用いて行った試行試験の実施の概要について紹介した。また、現在開発・検討している新しいシステムにおいての出題形式の紹介を行った。今後は、2018 年度に実施した試験およびそれに用いた CBT システムの評価を行なっていきたい。また、QTI などの規格に準拠したシステムの開発は今後の課題である。

## 謝辞

本稿は文科省大学入学者選抜改革推進委託事業(情報分野)の支援による研究成果の一部である。

## 参考文献

- IMS Global Learning Consortium(2015), 「QTI」.  
<<https://www.imsglobal.org/question/>>(2018 年 11 月 29 日)
- 大阪大学 他(2017), 「2017 年度「情報科」大学入学者選抜試行試験問題」.  
<[http://www.uarp.ist.osaka-u.ac.jp/pdf/CBT-V1\\_Examination.pdf](http://www.uarp.ist.osaka-u.ac.jp/pdf/CBT-V1_Examination.pdf)>(2018 年 11 月 29 日)
- Open Assessment Technologies(2018), 「TAO」.  
<<https://www.taotesting.com/>>(2018 年 11 月 29 日)
- 久野靖(2017). 「思考力・判断力・表現力を測るには?」, 情報処理, Vol.58, No. 8.
- 谷聖一 他(2016). 「「第 3 回・第 4 回大学情報入試全国模擬試験」の実施と評価」, SSS2016 情報処理学会情報教育シンポジウム論文集, 7-14.
- 中野由章 他(2014). 「「大学情報入試全国模擬試験」の実施と評価」, SSS2014 情報処理学会情報教育シンポジウム論文集, 11-17.
- 西田知博 他(2017). 「「情報科」大学入学者選抜における CBT システムの研究開発」, SSS2017 情報処理学会情報教育シンポジウム論文集, 182-187.
- University of Cambridge  
The Psychometrics Centre (2018). 「Concerto」,  
<<https://concertoplatform.com/>>(2018 年 11 月 29 日)
- 吉岡俊正(2004). 「医学系共用試験 CBT の医学部における取り組み」, 日本テスト学会 第 1 回研究会.  
<[http://www.jartest.jp/pdf/jirei1\\_1.pdf](http://www.jartest.jp/pdf/jirei1_1.pdf)> (2017 年 10 月 15 日)