

タブレット端末利用型 CBT における解答入力方法の検討 —オンライン手書き数式認識の利用—

国立教育政策研究所・教育課程研究センター・基礎研究部 安野 史子
Fumiko Yasuno, Department for Curriculum Research, Curriculum Research Center,
National Institute for Educational Policy Research

1 はじめに

筆者らは、Computer-Based Testing(CBT)の開発の一環として、コンピュータの特性を活かした教科・科目ベースの問題を試作し、CBTで測れる能力を明確にすることを目的として研究を行っている[3, 4, 5, 6, 7, 8, 9, 10]。開発した問題を用いて、高校生(及び大学生)を対象として調査を継続的に実施してきているが、開発初期段階は、解答用紙による解答方式を採用して、問題や解答形式の検討を行った[3, 4, 5, 6]。そして、改良を重ね、一部の問題(解答形式が選択肢式と短答式の設問)について、タブレット端末に解答を入力するシステムの開発を試み[7]、小規模パイロット調査[8, 9]を経て、モニター調査を実施した[10]。本稿では、特にモニター調査から得られたストロークデータログの分析を行うことで、数学問題におけるオンライン手書き数式認識による解答入力方法の有効性と問題点を実証的に探る。

2 数式入力のインタフェースの議論

数式入力のインタフェースは、現在、 $\text{L}^{\text{T}}\text{E}^{\text{X}}$ や MathML の類のテキスト表現、Design Science の MathType のような主にパレットなどを介した2次元表現、手書き表現の3種類に大別できる¹。文書作成ソフトウェア Microsoft Word では、これら3種類のいずれでも入力可能となっている。それぞれ長所短所があり、どれがいいかという議論をするよりも、利用環境、利用者、利用場面に応じて適したものを選択していくことであろう。

筆者らは、高校生を対象とした CBT での利用を前提にしていることから、利用者となる受験者は初めてそのシステムに触れることがほとんどである。特に、入力方法のガイダンスを短時間で行う必要があり、簡便性が求められる。そこで、オンライン手書き数式認識入力を採用した。

3 数式解答入力システムの開発

解答用紙による解答方式のみを採用していた開発初期段階において、問題冊子は、Apple Inc. が提供する電子書籍アプリケーション iBooks (現 Apple Books) による電子冊子と

¹キーボードによらない音声認識による数式入力インタフェースの研究[2]もなされているが、現時点での汎用性、実装性の観点から除外した。

した。作成においては、Apple Inc. が提供するツール iBooks Author (iBA) を用いている²。CBTにおいては、問題の順番や時間の制御が可能であるが、わが国の数学のテスト文化から鑑みて、本研究ではそのような制御は適さないと判断した。本研究ではCBTならではの数学問題として、動的オブジェクトを含む問題を開発 [4, 6, 7] してきている。そこで、開発した CBT は、動的オブジェクトが組み込まれた電子冊子の形態で、従来型の Paper-Based Testing (PBT) の問題冊子と同様に、受験者は既定の試験時間内に、任意の順番で解答することができたり、解答した問題に戻って解答を修正したりすることができる。

解答欄は、2018年に実施したパイロット調査のシステム [8, 9] を踏襲し、問題提示と一体で解答入力システムを組み込んでいる。選択肢式はHTMLのセレクトタブ <SELECT> によるプルダウンメニュー、短答式については手書き数式認識入力を採用し、どちらも iBA のウィジェット (widget) の中の「HTML5」で組み込み、手書き数式認識及び解答データの収集はオンラインで行った。

手書き数式認識の部分は、MyScript 社の MyScript CDK (Cloud Development Kit) を使って、クラウドベースの認識サービスによる手書き認識を電子書籍に統合させている。解答者は、タブレット端末画面の解答欄に指またはタッチペンで手書きで入力し、解答者自身が数式認識結果を確認して確定する。認識結果が異なっていれば再度書き直して認識結果の確認を繰り返すことができる。入力された解答データログは、最終解答のみではなく、認識イベントを実行したすべてのストロークデータ (JSON 形式) 及び認識結果 (Math ML 形式) である。

4 モニター調査

4.1 調査目的

開発した CBT の総合的な検証及び改善に資するため。

4.2 調査実施

本調査は、数学4冊子 ($\alpha, \beta, \gamma, \delta$ 冊子)、物理3冊子、化学3冊子の計10冊子の中から、1受験者が異なる教科科目について最大2冊子の問題に解答している [10]。数学 α 冊子、 β 冊子は解答用紙へ記入する記述式で、数学 γ 冊子、 δ 冊子はタブレット端末へ直接解答するオンライン解答入力式である。

(1) 調査日 : 2019年7月下旬~11月上旬

(2) 調査対象者 : 高等学校 (全日制普通科) 第3学年 (あるいは第2学年) の生徒で、数学は、「数学I」「数学II」「数学A」「数学B」を履修している (した) 生徒。)

(3) 調査実施校 : 県立高等学校 : 4校 (3県)

²iBooks Author は今後アップデートの予定がなく、新規ユーザへの提供も中止されている (参照 : <https://support.apple.com/ja-jp/HT211136>)。)

表 1: モニター調査冊子の構成

冊子	問題番号	内 容	選択	短答	手書き数式認識による正解例
γ	第 1 問	4 次式の因数分解 (因数定理)	(2)	(1)	$(x - 11)(x - 23)(x - 29)(x - 111)$
γ	第 2 問	漸化式 (薬成分の体内残存量)	(2)	(1)	$0.6a_{n-1} + p$
γ	第 3 問	3 辺上の正三角形 (フェルマー点)	(1)~(6)		
γ	第 4 問	正十二角形の頂点からなる三角形		(1)~(4)	(1) 220 (2) 4 (3) 5 (4) 12
γ	第 5 問	色玉の取り出し (8 個)	各個数		
δ	第 1 問	絶対値を含む関数		(1)~(3)	(1) $x, x - a , x - a - a$ (2) $a < 0$ (3) $a > 0, x > \frac{a}{2}$
δ	第 2 問	三次関数の決定 (導関数)		(1)(2)	(1) $\frac{7}{2}, -\frac{9}{4}$ (2) $\frac{1}{3}x^3 - \frac{7}{2}x^2 + 10x$
δ	第 3 問	三角形の敷き詰め	(1)(2)		
δ	第 4 問	正十二角形の頂点からなる三角形		(1)~(4)	(1) 220 (2) 4 (3) 5 (4) 12
δ	第 5 問	色玉の取り出し (9 個)	各個数		

(4) 調査実施者数 : 155 人 (高 2 : 38 人, 高 3 : 117 人)

(α 冊子 : 44 人, β 冊子 : 40 人, γ 冊子 : 28 人, δ 冊子 : 43 人)

(5) 調査実施方法 : 調査実施者が所属の高等学校の教室にて, 電子問題冊子を取り込んだる 9.7 インチのタブレット端末 (iPad, OS : iOS11), タブレットペン, 解答用紙あるいはメモ用紙を配付し実施. なお, オンライン実施冊子については, 各端末を LTE の WiFi ルータ (通信事業者 : Softbank) 経由でインターネットに接続した状態で実施.

4.3 調査問題

これまでに開発してきた数学の問題の中から, 選択肢式と短答式の設問のみで構成されている問題を選び出し, オンライン解答入力冊子として γ 冊子, δ 冊子を構成した. ただし, γ 冊子と δ 冊子は, ほぼ同じ内容で, 難易度が異なるよう構成している (δ 冊子の方が難). 第 4 問は共通問題である. また, 記述式の問題で, α 冊子と β 冊子を構成し, 前述のように解答用紙へ記入する方式であるため, これらの冊子は本稿では扱わず, これ以下, γ 冊子と δ 冊子についてのみ言及する. 表 1 は各冊子の問題構成を示したものである. 手書き数式認識入力による設問は, 表 1 の「短答」の列に小問番号が付してある設問が該当し, 最右列に正解例が示してある. なお, 解答時間はいずれの冊子も 50 分とした.

5 調査実施結果

表 2 は各冊子及び各問題の配点, 平均点, 標準偏差, 最低点, 最高点, 人数を示したものである. 冊子は, 3 校は 4 冊子 ($\alpha, \beta, \gamma, \delta$) を均等に, 1 校 (いわゆる県立トップ校 3

表 2: モニター調査実施結果

冊子	問題番号	配点	mean	sd	min	max	n
γ	第 1 問	30	18.2 (60.7%)	10.3	0	30	28
γ	第 2 問	30	20.2 (67.3%)	11.1	0	30	28
γ	第 3 問	30	20.9 (69.6%)	5.8	10	30	28
γ	第 4 問	30	5.7 (19.0%)	5.3	0	15	28
γ	第 5 問	30	22.5 (75.0%)	13.2	0	30	28
γ	全体	150	87.5 (58.3%)	23.3	42	120	28
δ	第 1 問	30	19.3 (64.5%)	8.2	5	30	43
δ	第 2 問	30	10.2 (34.0%)	13.3	0	30	43
δ	第 3 問	30	15.3 (51.2%)	12.2	0	30	43
δ	第 4 問	30	9.1 (30.5%)	7.3	0	30	43
δ	第 5 問	30	14.9 (49.6%)	15.0	0	30	43
δ	全体	150	68.9 (46.0%)	33.4	5	142	43
$\gamma \delta$	第 4 問	30	7.8 (26.0%)	6.7	0	30	71

年理系クラス)は3冊子(α, β, δ)を均等に割り当てたため、共通問題の第4問の結果からもわかるように、各冊子の母集団が若干異なる。

WiFi接続及び手書き認識入力の実施トラブルに対しては、発生に備えて記入する紙を事前に用意し、トラブルが発生した場合は、それを配付し、その用紙に解答を記入させるという対応をとった。実際には、調査実施者の瑕疵によりオフラインとなってしまった場合、手書き数式認識が上手くいかない場合が多少発生し、用紙による解答回収も行われた。ただし、問題はローカルにタブレット端末に取り込んであるため、オフラインでも問題を解くことは可能であった。

6 手書き数式認識入力に関する諸課題

以下は、手書きストロークデータについて、分析を行った結果である。手書きストロークデータの分析の有効データの人数は、55人(γ 冊子: 21人, δ 冊子: 34人)で、書き直しも含めた全ストロークデータ数は897である。なお、このデータ数には、無解答や用紙に記入した解答データは含まれない。あくまでも手書き数式認識の検証を目的とする。

6.1 ストロークデータによる最終解答

本システムでは、3で述べたように、解答者が認識結果を確認して確定するという手順を踏んでいる。ストロークデータによる最終解答は、897のうち466であった。6.2で詳述するが、一つの解答欄に対して平均1.9回解答入力になされていることになる。表3はそれらの内訳を示したものである。問題と関係ないテキストの入力が5(すべて同一受験者)、書きかけが1、ノイズのみ(無解答と同等)が1あったが、それらを欠損として除いた459の中で、手書きが正しく認識された最終の解答は437(95.2%)で、そうでない解答が22(4.8%)であった。問題となる22の解答は、解答入力欄に触れてしまった結果、

表 3: ストロークデータによる最終解答種別

種別	解答数	(%)	備考
正認識	437	(95.2%)	
ノイズ	10	(2.2%)	前方: 4 後方: 6
誤認識	11	(2.4%)	
1 → /,	2	(0.4%)	/ (スラッシュ): 1 (パイプ): 1
$x \rightarrow \chi$	3	(0.7%)	
小数点 → , ·	2	(0.4%)	, (コンマ): 1 · (中点): 1
$a \rightarrow \alpha$	1	(0.2%)	
$a < 0 \rightarrow aco$	1	(0.2%)	
$-1 \rightarrow +$	1	(0.2%)	
分母が 11 の分数 → Π	1	(0.2%)	
数式とテキスト混在	1	(0.2%)	想定外の場合分けの誤答
計	459	(100.0%)	
欠損	7		テキスト: 5 書きかけ: 1 ノイズのみ: 1

細かい点が付き、それも一緒に認識され、ノイズが付与されたものが10(2.2%)、誤認識のまま目視での見逃しが11(2.4%)、数式とテキストの混在が1であった。誤認識結果の目視での見逃しは、形が似ている記号やギリシャ文字が目立つ。また、本研究で用いている手書き認識入力システムは、数式とテキストの混在を想定しておらず、解答欄はテキストか数式のいずれかに設定している。 δ 冊子の第1問(1)の想定していない誤答は「 $x \geq 5$ のとき $x - 10$, $x < 5$ のとき $-x$ 」であったが、テキストを含むため、そもそもこのシステムでは認識できない。参考までに言うと、物理、化学では、日本語での手書き文字認識入力を試みたが、数式よりも認識率が高く受験者からも高評価であった [10]。

6.2 解答入力の頻度

解答者が、解答欄に手書きデータを入力した頻度をまとめたものが表4である。三つのブロックの左側が欠損を含む全897のデータによる度数分布で、中央が前出の表3と

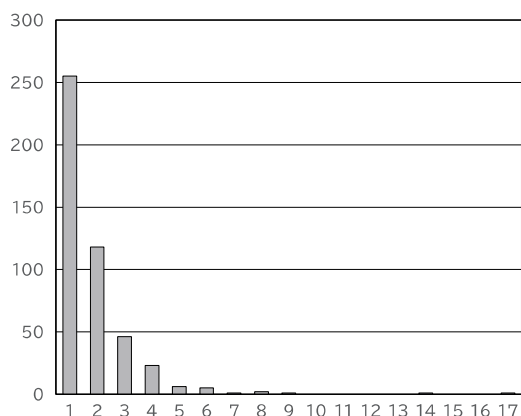


図 1: 解答入力の頻度

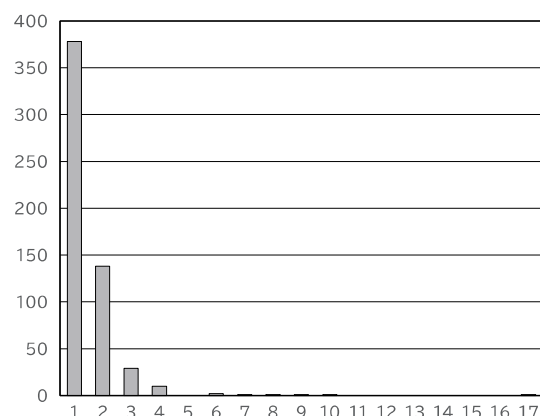


図 2: 解答入力の頻度 (同一解答別)

表 4: 解答入力の種類

回数	解答入力の頻度 (欠損含む)			解答入力の頻度			解答入力の頻度 (同一解答別)		
	度数	割合 (%)	回数×度数	度数	割合 (%)	回数×度数	度数	割合 (%)	回数×度数
1	253	54.3%	253	255	55.6%	255	378	67.3%	378
2	115	24.7%	230	118	25.7%	236	138	24.6%	276
3	54	11.6%	162	46	10.0%	138	29	5.2%	87
4	21	4.5%	84	23	5.0%	92	10	1.8%	40
5	8	1.7%	40	6	1.3%	30	0	0.0%	0
6	7	1.5%	42	5	1.1%	30	2	0.4%	12
7	1	0.2%	7	1	0.2%	7	1	0.2%	7
8	2	0.4%	16	2	0.4%	16	1	0.2%	8
9	1	0.2%	9	1	0.2%	9	1	0.2%	9
10	1	0.2%	10	0	0.0%	0	1	0.2%	10
11	0	0.0%	0	0	0.0%	0	0	0.0%	0
12	1	0.2%	12	0	0.0%	0	0	0%	0
13	0	0.0%	0	0	0.0%	0	0	0.0%	0
14	1	0.2%	14	1	0.2%	14	0	0.0%	0
15	0	0.0%	0	0	0.0%	0	0	0.0%	0
16	0	0.0%	0	0	0.0%	0	0	0.0%	0
17	0	0.0%	0	1	0.2%	17	1	0.2%	17
18	1	0.2%	18	0	0.0%	0	0	0.0%	0
計	466	100%	897	459	100%	844	562	100%	844
	欠損 (テキスト)			5		12	↑		
	欠損 (書きかけ)			1		18	459 に解答修正の 103 を加算		
	欠損 (ノイズのみ)			1		23			

同様の欠損を除外したデータによる度数分布 (そのヒストグラムは図 1) をまとめたものである。解答入力の回数が、1 回で済んでいるのが約 56%、2 回 (1 回の書き直し) までが約 81% で、4 回 (3 回の書き直し) までで約 96% に達し、ほぼ 4 回以内で入力を終えていることがわかる。しかし、のべ 459 の解答入力欄に対して、全体で 103 回解答そのものが修正されている。手書きによる数式認識の精度を確認するために、同一の解答別に解答入力の頻度をまとめたものが、表 4 の右側のブロックである (そのヒストグラムは図 2)。1 解答につき平均 1.5 回解答入力している計算になる。こちらの場合では、1 回で済んでいるのが約 67%、2 回 (1 回の書き直し) までが約 92% で、3 回 (2 回の書き直し) までで約 97% に達し、ほぼ 3 回以内で入力を終えていることがわかる。なお、解答入力回数が 5 回以上の場合の多くは、解答欄を全画面表示にして手書き数式認識入力を行いなさいという指示に従っていない場合である。全画面表示にせず問題表示モードの状態解答欄に手書き入力を行うと、画面が左右に動いて非常に書きにくい状態となり、手書き入力ができても、判読不能に近い字になることがほとんどである。

表 5: 解答の書き直し

種別	度数	割合 (%)
誤認識	160	57.1%
ノイズ	87	31.1%
誤認識+ノイズ	5	1.8%
続け字	2	0.7%
続け字+ノイズ	1	0.4%
重ね文字	1	0.4%
重ね書き	2	0.7%
重ね書き+継ぎ足し+ノイズ	1	0.4%
認識不能 (${}_nC_r$)	2	0.7%
数式とテキスト混在	1	0.4%
その他 (判読不能)	18	6.4%
計	280	100%

表 6: 誤認識例

手書き	誤認識	頻度
小数点	,	15
小数点	1	6
1	(5
1	/	14
7	Π	14
7	π	14
a	A	5
x	χ	8
< 0	co	6
\leq	\leq	5
-	=	6
a_{n-1}	$an - 1$	11

6.3 解答の書き直し

手書き数式認識で書き直しがなされたデータは、表 4 の右下の 844 データのうち 280 である³。ここでの書き直しとは、解答の修正は含まず、同一の数式の書き直しを意味する。表 5 は、それらを分類したものである。誤認識による書き直しは、書き直し全体の 60% 近くに上る。また、ノイズ (解答入力欄に触れてしまった結果、細かい点が付く、それも一緒に認識されたもの) による書き直しは、全体の 33.5% であった。ノイズは、ストロークデータとしての点であるような目視で確認しにくい程度のものもかなりあり、数式認識させてはじめて気が付くものも少なくない。逆に、小数点を書き込んだ (ストロークデータが生成された) が、目視で確認できない大きさの点であったため、再度小数点を付与し、認識結果として点がダブル (..) になっている事例も見られる。誤認識とノイズ以外の多くは、続け字や重ね書きなど手書き数式認識に不向きな記述であった。

誤認識は多岐に渡るが、頻度が 5 以上であったものを表 6 に示す。小数点の誤認識は、(コンマ) や 1 が多いが、 \cdot (中点), 1 (添字の 1) も観察されている。数字の誤認識は、表 6 に示す 1 と 7 が目立つが、形が近い「0 が o, σ, θ 」, 「6 が $0, b$ 」, 「9 が q 」等も散見される。また、1, | (パイプ), / (スラッシュ), [の間の誤認識も見られる。アルファベットは、表 6 に示す a と x が目立つが、それ以外は僅かであった。数学記号は、表 1 の正解例からもわかるように、不等号と絶対値記号のみであったため、参考となるような結果は得られなかった。これ以外に、接近あるいは重なってしまったことで、) (が $x, |-, -1, 1-$ が + と認識されてしまった事例が若干あった。最後に、上付き (指数) や下付き (添え字) がうまく認識されずに、書き直している事例も一定数観察されている。

[1] に示されているように、オンライン手書き数式の認識は、数学記号の認識と構造認識の組み合わせで行われる。隠れマルコフモデル (HMM) が数学記号を認識するために

³表 4 の右側のブロックの 844 と 562 の差の 282 と一致していないのは、正しく認識されたにもかかわらず、再度書き直した事例が 2 例あったからである。

使用され、確率文脈自由文法 (SCFG) がこれらの記号間の関係をモデル化するために使用されている。数式には、アルファベット、ギリシャ文字、特殊文字 (記号)、アクセント記号、演算子等、特殊で似通った形状の記号が多いこともあり、今回のデータセットの認識結果から、数学記号の認識、構造認識の両方において、ある程度の誤認識が発生することが確認された。

6.4 数学特有の数式ストロークの順序

最後に、数学特有の数式ストロークの順序について触れる。簡単な例を挙げると、 $x^2 - x - 12$ を因数分解する際、 $(x - 3)(x - 4)$ を先に書き、後から片方の $-$ に縦棒を付与して $+$ にし、 $(x + 3)(x - 4)$ とすることは当然ある。数式を必ず左から右へ順に書いていくとは限らないし、左から右へ順に書かないことのほうが合理的であることさえあり得る。本データにおいても、図 3 に例示した書き順で入力した解答のストロークデータが存在する。この問題は δ 冊子第 2 問 (2) の誤答である。この問題は導関数の問題であることから、 x^2 の項の係数、定数項、 x^3 の項の係数の順に決定していった、その順に書いていった可能性は否定できない。因みに、このストロークの認識結果は、

$$\frac{2,6}{77}x^3 \prod x^2 + \frac{780}{77}$$

であった。

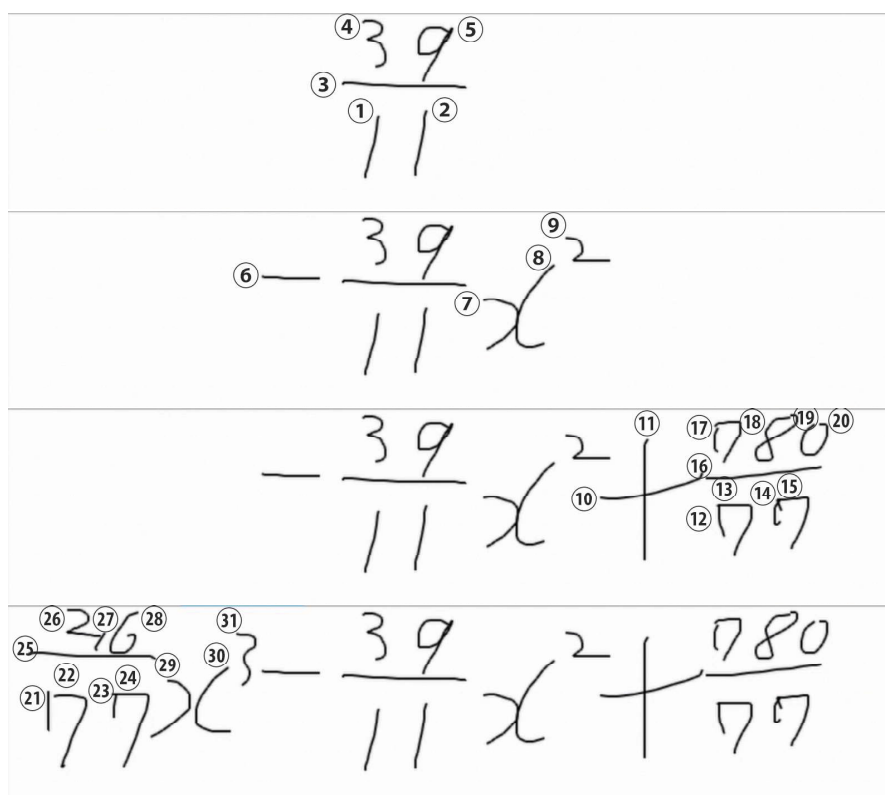


図 3: 数式の書き順例

7 まとめと今後の課題

本稿では、CBT の特性を生かした数学の問題で、解答形式が選択肢式と短答式の設問であるものに対して解答入力システムを開発し、タブレット端末に解答を入力する方式を試みた。モニター調査で得られたストロークデータセットにより、手書き数式認識入力は、95%程度は正しく認識され、解答の書き直し回数も、それほど多くないことが実証された。この結果から、キーボードで入力が困難な数式記号などを受験者が容易に入力できるという点において、CBT の数式解答入力方法として、手書き数式認識入力は、非常に有効な方法であると考えられる。

今回は、汎用的なクラウドベースの認識サービスを利用したが、現時点で、数式とテキストが混在する解答についての自動認識、「 $_5C_2$ 」のように添え字から始まる数式の認識、ができないという問題点があることを付け加えておく。また、特記すべき点は、一定数発生しているノイズは、対応策を検討する必要がある。

CBT のプロトタイプとして電子書籍フォーマットをベースに開発を進めてきたが、様々な OS での利用を前提にすると、HTML と JavaScript の組み合わせでの Web 版が最適と考えられる。これまでに開発した問題及びシステムを学習管理システム (Learning Management System) の小テストや CBT プラットフォームシステムに移行することを現在検討している。特に、本研究でその有効性が確認できた手書き数式認識入力システムの組み込みが課題となっている。

最後に、本稿では解答の修正回数など、これまで PBT では知り得ることがなかった試験における解答過程の情報が得られている。このことは、CBT でならであり、新たな分析手法や評価手法へと繋がる可能性を秘めている。しかし、試験における解答過程の情報を評価等に活かしていけるのかどうかについては、これからの課題である。

註

iPad, iPad mini, Apple Books, iBooks, iBooks Author は、米国および他の国々で登録された Apple Inc. の商標である。

附 記

本研究は JSPS 科研費 JP17H00822, JP18K18684 の助成を受けたものです。

参考文献

- [1] Francisco Álvaro, Joan-Andreu Sánchez, José-Miguel Benedí : Recognition of on-line handwritten mathematical expressions using 2D stochastic context-free grammars and hidden Markov models, Pattern Recognition Letters, 35, 58–67, 2014.
- [2] 前田秋吐, 鈴木昌和 : 「音声認識を用いた日本語による数式インタフェース」, 京都大学数理解析研究所 講究録 No.1286, 「数学解析の計算機上での理論的展開とその遂行可能性」, pp.170–177, 2002.

- [3] 安野史子：「高大接続を視野に入れたタブレットを用いる評価問題の試作 —映像や動的オブジェクトを含む問題—」, 『大学入試研究ジャーナル』 Vol. 27, pp.71–79, 2017.
- [4] 安野史子, 西村圭一, 根上生也, 浪川幸彦, 三宅正武：「高大接続に資する動的オブジェクトを含む数学の評価問題の開発」, 京都大学数理解析研究所 講究録 No.2022, 「数学ソフトウェアとその効果的教育利用に関する研究」, pp.177–184, 2017.
- [5] 安野史子：「高大接続を視野に入れたタブレットを用いる評価問題の試作 (2) —映像や動的オブジェクトを含む問題—」, 『大学入試研究ジャーナル』 Vol. 28, pp.155–162, 2018.
- [6] 安野史子, 西村圭一, 根上生也, 祖慶良謙, 高橋広明, 浪川幸彦, 伊藤仁一, 三宅正武：「動的オブジェクトを有する CBT 数学問題の開発」, 日本数学教育学会誌, 第 100 巻 第 5 号, 『数学教育』 72-3, pp.2–14, 2018.
- [7] Yasuno,F., Nishimura,K., Negami,S. and Namikawa,Y. : Development of Mathematics Items with Dynamic Objects for Computer-Based Testing Using Tablet PC, International Journal for Technology in Mathematics Education, 26(3), 132–137, 2019.
- [8] 安野史子：「電子書籍を利用したタブレット端末による CBT の数式入力システムの検討」, 京都大学数理解析研究所 講究録 No.2142, 「数学ソフトウェアとその効果的教育利用に関する研究」, pp.71–78, 2019.
- [9] 安野史子：「電子書籍を利用したタブレット端末による CBT の数式入力システムの検討—手書き認識入力を中心に—」, 『大学入試研究ジャーナル』 Vol. 30, pp.221–226, 2020.
- [10] 安野史子：「タブレット端末利用型 CBT の開発とモニター調査による評価」, 『大学入試研究ジャーナル』 Vol. 31, pp.146–153, 2021.