

デジタル技術を活用したタブレット入試の開発

～多面的・総合的評価に向けた技術的検討～

西郡大*・山口明德**・松高和秀***・長田聡史*・坂口幸一*・福井寿雄*・
高森裕美子*・園田泰正*・兒玉浩明*

*佐賀大学入試改革推進室 **佐賀県立唐津東高等学校 *** 佐賀県立佐賀西高等学校

佐賀大学では入試改革の1つとして、ペーパーテストでは技術的に評価することが難しい領域をタブレット等のデジタル技術を用いて評価する「佐賀大学版 CBT」の開発を進めている。これまで化学を題材に試行版テストを開発し、高校生を対象にモニター調査を実施した。問題開発では、「時間軸」の視点を取り入れるなど、デジタル技術のメリットを活かせる可能性を見出すことができた。モニター調査の結果では、多くの高校生がタブレットの解答入力に関して操作的な問題はなく、デジタル技術のメリットを活かした問題作成の効果を部分的に確認することができた。一方、実際の入試運用という観点からは、今後の検討を要する具体的な課題点が明らかになった。

1. 佐賀大学版 CBT について

「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体改革について（答申）」（中央教育審議会,2014）を発端として、高大接続システム改革会議において「最終報告書」（2016）がまとめられた。こうした背景を受け、西郡・園田・兒玉（2016）は、多面的・総合的な評価に向けた佐賀大学の入試改革構想を報告した。本稿では、その入試改革構想の1つである「佐賀大学版 CBT」の開発状況について報告する。

我が国の「CBT(Computer Based Testing)」は、国家試験として情報処理推進機構 (IPA) が実施する「IT パスポート試験」や医療系大学間共用試験などが有名であるが、近年では、英語の4技能を評価する CBT の検定試験にも注目が集まっている（日本テスト学会,2016）。

「佐賀大学版 CBT」は、こうした CBT とは異なり、ペーパーテストでは技術的に評価することが難しい領域をタブレット等のデジタル技術を用いて評価しようという試みである。デジタル技術を活用することにより、動画や音声の利用、紙媒体では表現できなかった3次元図形の提示など、PBT (Paper Based Testing) では実現できない情報量の多い問題を出題したり、記述解答からは得ることができなかった多様な解答形式を用いることで、より深い思考力や表現力の評価を目指すものである。

また、佐賀大学版 CBT では、入試で問われなければ、高校において積極的な学習活動が行われたいといわれる問題点を考慮し、大学で学ぶために重要だと思われる学習活動や学習経験を喚起するような仕組みを考えている。例えば、実験が必要な分野において、実験の学習活動や経験が有利になる問題や形式を検討し、実際の入試で問うことによって、高校の理科の授業時間などで、実験が積極的に取り入れられるようになることを期待している。これは、良いテストを検討する上で、測定信頼性や妥当性の観点からアプローチするだけでなく、社会的な文脈の視点も取り入れた検討が必要であるという主張 (Frederiksen, J. R., & Collins, A, 1989; Messick, S, 1989) と合致した考え方である。

以上の点を踏まえ、PBT で技術的に評価できない領域がどのような部分にあるのか。また、デジタル技術を利用することによって何が実現できるのか。また、それらを実現する上での課題は何かについて明らかにするために、「化学」を題材に、タブレットを利用したデジタルテスト（試行版）を開発した（以下、「試行版テスト」）。本稿では、開発の考え方と試行版テストのモニター調査の結果を報告するとともに、実際の入試運用に向けてこれから検討すべき具体的な課題点を考察する。

2. 試行版テストの開発

2.1 問題作成の基本的な考え方

開発当初、プロジェクターによるスクリーン投影によって受験者全員に動画の問題を見せ、解答は、解答用紙に記述させることを想定していた。しかし、動画の投影は、受験者全員が同じ映像を見ることができ、個別に見たい部分を巻き戻して確認することができないため、タブレットによる個別テストとして実施することにした。プラットフォームは、iPad Air（第 5 世代）、iOS（Ver.8.0 以上）を用いて、出題・解答を制御するアプリを開発した。

試行版テストの流れは、受験者情報入力画面から始まり、試験に関する指示画面、操作説明・練習画面を経て各問題が出題される。受験者情報及び試験に関する指示は、PBT でも行われる一般的なプロセスだが、操作説明・練習画面は、センター試験のリスニング試験と同様に、動画、画像、選択回答、記述回答など、操作に関する説明を目的としている（図 1）。

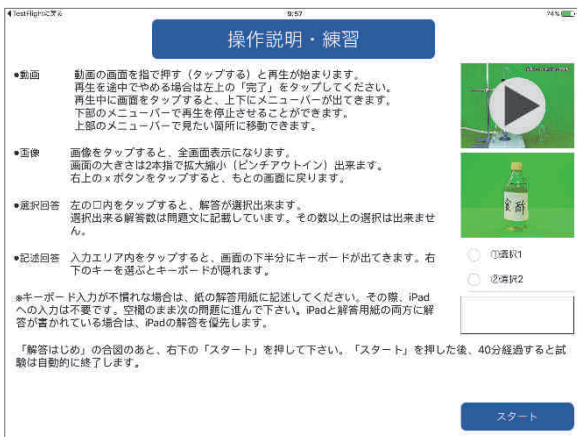


図 1. 操作説明・練習画面

出題および解答の形式は、実験動画から考えられる現象を選択肢より解答させる問題。様々な実験器具を並べた写真から必要な器具を選ばせる問題。数値データをもとに計算させる問題。ある現象が生じる理由を記述によって解答させる問題で構成した。そのため、ラジオボタンから回答を選ぶ多肢選択方式、計算問題や理由を問う記述式による解答形式で構成される。

出題に関しては、以下の 4 点をデジタル技術のメリットとして活かした。1 点目は、一連の

実験動作やプロセスを動画によって示すことにより、PBT では実現できなかった「時間軸」を入れることである。例えば、PBT において、ある実験過程に注目して受験者に考えさせたい場合、関係する実験の過程を問題文中で説明しなければならない。そうすると、出題者の意図的な誘導が回避的に介入してしまう。しかし、一連の実験過程を動画により示すことができ、受験者自身が操作して確認できるようになれば、そのポイントとなる実験過程がどこであるかを考えさせることから問うことができる。これにより、想定外の実験結果等を提示して、実験手順や操作のどこに間違いがあったのかなどを動画の流れの中で確認させる問題を出題することができる。

2 点目は、タブレットのピンチアウト機能である。PBT では、複雑な構造をもつ実験器具や物質などを示す場合、精巧な印刷を要するが、タブレットであれば、問題文中に小さな画像で示しておいても、画像の解像度を高くしておくことが可能である。これにより、複雑な構造を持つ写真や画像を提示しても、指で拡大（ピンチアウト）して確認できる。このメリットを利用して、様々な種類の器具や同じ器具でも目盛りの異なる器具から、適切なものを選択させる問題などを出題することが可能となる。

3 点目は、解答の順次性を制御できることである。複数の設問で構成する問題を PBT で出題する場合、先行設問の答えや知識を用いて、次の設問を解く形式が用いられることがある。この形式では、先行する設問が解けなければ、次の設問の解き方が分かっていたとしても正答は難しい。これを避けるために、各設問の解答確定後には、前の設問に戻れないように制御し、先行設問の解説や正答を提示する。これにより、提示された解説や正答の内容を踏まえて、次の設問に解答することが可能となる（図 2）。

4 点目は、時間管理である。PBT では試験監督が試験時間を管理するのが一般的だが、デジタル技術を用いれば、終了時刻到来時に自動的にアプリを終了させることで、厳密な試験時間の管理ができる。さらに、各設問にどの程度の時間を要しているかなど、PBT では取得できなかった情報を得ることができ、新たな観点から受験者の解答を分析することが可能である。



図 2. 前の設問に戻れないことを注意する画面

一方、タブレットを利用したテストであるゆえに生じる課題点もある。例えば、PBTであれば、問題文の重要な部分に線を引いたり、メモを残したりすることができる。これらの機能を今後どのように位置づけるのかは1つの課題といえる。

また、PBTであれば、全体の問題内容を把握してから、自分が解きたい問題から解くことが可能であるが、試行版テストでは、順次的な出題について検証したいという思惑もあり、全体の問題を俯瞰的に確認することができないようにしている。そのため、試験問題に入る前の全体の説明画面において、問題数と各問題の解答目安時間をあらかじめ示すことにした。

なお、試行版テストでは、前の問題に戻ることができないように制御しているため、PBTのように前の問題に戻って解答を見直したり修

正したりすることができない。受験者の解答の利便性という観点からはネガティブな機能ともいえる。しかし、見直しができるということは、作題者が1つ1つの問題に想定する解答時間で、受験者が必ずしも解答していない可能性が指摘できる。例えば、ある問題が解けていなかった場合、「その問題を解ける学力があったにも関わらず、他の問題の見直し等に時間を掛けたために、解答が出来なかった」のか、「その問題に時間を割り当てたのにもかかわらず分からなくて解けなかった」のかについて弁別することができない。こうした観点についても、デジタル技術を活用した出題管理方法の在り方を検討していくことが必要である。

2.2 問題内容

問題内容は、高校2年生で学習する範囲の中から、「水の状態変化と沸騰実験」(5問)、「食酢の酢酸濃度を中和滴定で求める実験」(9問)を出題した。解答時間の目安は、前者が15分、後者が25分を想定し、全体で40分の試験とした。問題内容については、化学を専門とする大学教員だけでなく、高校で化学を担当する教員も問題作成に関わり、実際の高校現場の視点からも問題の適正性を検討した。

各問題の出題意図と難易度及びタブレットテストとして工夫した点を表1に示す。PBTでは評価できない領域を全ての問題で測定しようとするのではなく、PBTで問う内容とも相関が考えられる問題を含めることで、普段の化学の学力との関係を分析できるように構成した。

表 1. 問題の出題意図、難易度およびタブレットテストとして工夫した点

問題	出題意図	難易度	タブレットテストとして工夫した点
問 1-1	水の状態変化を理解し、動画で示される実験の様子が説明できる。	易	
問 1-2	水の状態変化を理解し、動画で示される実験の様子が説明できる。	易	
問 1-3	水の状態変化を理解し、動画で示される実験の様子が説明できる。	易	実験の動画を用いることで、PBTでは出題出来ない水の状態変化の様子を出題し、実験で起こっている現象を説明させる。
問 1-4	水の状態変化を理解し、容器内の圧力減少により沸点が低下することを理解している。	難	
問 1-5	水の状態変化を理解し、容器内の圧力変化を理解している。	やや易	
問 2-1	メスフラスコとメスシリンダーやホールビペットとメスピペットなど体積計の性質を理解し、画像に示された器具の中から、指示された希釈操作に必要な器具の組合せを正しく選ぶことができる。	普通	タブレット上で拡大が可能な画像を用いることで、器具の種類や容量、器具の目盛りなどの情報を元に解答することを意図している。
問 2-2	実験経験から、実験の手順と器具の役割を理解しており、必要な操作を正しく理解しているか。	普通	

問 2-3	中和滴定の基本的事項を知っている。滴定に用いる酸と塩基の性質と、指示薬の染色域の知識から正しい指示薬を選択できる。	易	
問 2-4	実験経験から実験操作で誤りやすい事項の候補を見出し、それをもとに動画を観察し、誤った実験操作を見つけ出すことができる。	難	実験の動画を用いることで、PBT では出題出来ない実験操作に関する出題をした。また、実験から学ぶ操作上の注意点を出题し、実験経験の観点を加えた。
問 2-5	実験の動画を見て、その中に示された滴定値から基本的な濃度の計算ができる。	普通	
問 2-6	実験の動画を見て、その中から濃度計算に必要な情報を見つけ出し、それをもとに基本的な濃度の計算ができる。	普通	計算は、PBT でも類似の出題が可能。試薬の分子量を問題文で与えず、試薬ビンから読み取らせることで、実験経験の観点を加えた。
問 2-7	与えられた滴定値を用いて、基本的な濃度の計算ができる。	普通	
問 2-8	水酸化ナトリウム水溶液の性質を理解している。	やや難	
問 2-9	題意の実験過程を理解して、正しい酢酸の濃度を求めることができる。	やや難	

3. モニター調査

3.1 調査概要

試行版テストを検証するために、S 県内の 2 つの普通科高校理系クラスに在籍する 2 年生を対象にモニター調査を実施した。A 高校では、3 クラス 104 名、B 高校では 3 クラス 119 名の協力を得た（男女比は同程度）。調査時期は、試行版テストで出題する範囲の学習後に設定し、全クラスのモニター調査を 3 日間で終えた。

モニター調査では、50 台の iPad を用いて実施した。出題と解答は、タブレットで行うことを原則としたが、タブレットに解答を入力できない者や、トラブルが生じた場合などを想定して解答用紙を用意した。また、メモや計算を行うための下書き用紙も配布している。

まず、最初に調査責任者が、調査目的と試行版テストの操作方法について約 5 分間説明し、40 分の試験を行った。試験終了後に、本テストに関するアンケート調査を実施し、帰宅までに記入・提出を指示した。解答データは、受験者別にリストでタブレット内に保存することができ、調査終了後に端末ごとにデータを一括収集できる仕組みになっている。

3.2 調査結果

初日の調査において、アプリの不具合により 8 件の解答データが取得できなかったため、急遽、アプリを改修して以降の調査を実施した。改修後の調査では、試験時間の途中でアプリが停止するトラブルが 12 件あったが、再起動によって対応した。最終的に、215 件の解答データを得ることができた。

当初、タブレットへの入力解答が技術的に困難な生徒が一定数生じると想定していたが、ほぼ全員がタブレットでの入力解答を行えたのは 1 つの発見であった。ただし、計算問題において累乗の表現方法が分からないなど、部分的に入力ができなかったものについては、当該部分のみ解答用紙に記入したケースがみられる。

各問題の採点は、完全正答に 1 点、それ以外は 0 点とし、計算問題や記述問題における部分点や設問ごとの重み付けは行わなかった。したがって、14 点満点での採点の結果、平均 5.16、標準偏差 1.94、最高点 10、最低点 1 であった。問 1 は 5 点満点で平均 2.01、標準偏差 0.79、最高点 5、最低点 0。問 2 は 9 点満点で平均 3.14、標準偏差 1.69、最高点 8、最低点 0 であった。各設問の正答率（通過率）は表 2 に示す。

表 2. 各設問の正答率（通過率）

問 1-1	問 1-2	問 1-3	問 1-4	問 1-5	問 1 全体				
0.80	0.24	0.83	0.06	0.08	0.40				
問 2-1	問 2-2	問 2-3	問 2-4	問 2-5	問 2-6	問 2-7	問 2-8	問 2-9	問 2 全体
0.59	0.54	0.83	0.30	0.26	0.38	0.13	0.09	0.03	0.35

受験者における日頃の化学の学力について、A～Fの6区分にレベル別に分けた成績を高校から提供してもらい、試行版テストとの関係性について分析した。6区分の学力レベルについては、日頃の模試等で偏差値70以上を成績上位者としてA段階から分けてもらっている。各学力レベルに該当する生徒数を表3に示す。

表3. 受験者の化学の学力レベル

レベル	該当者数	%
A	31	14.7
B	34	16.1
C	43	20.4
D	47	22.3
E	38	18.0
F	18	8.5
合計	211	100

※4名のみ成績データなし

表4は、設問ごとに学力レベルの分布をまとめたものである。数値は、学力レベル別の該当者に占める正解者数の割合である。問1は、学力レベルによる正答率の違いはみられないことから、通常のPBT成績と試行版テストとの相関関係は高くない。したがって、日常の模試等で評価する化学の学力面とは異なる側面を評価できている可能性が考えられる。

一方、問2（問2-9を除く）については、学力レベルによる正答率に明らかな相関関係が確認され、学力レベルが高いほど正答率が高くなっている。つまり、PBTで評価している側面と類似した面を評価していると解釈できる。この理由は、PBTでも類似した出題が可能である計算問題や理由説明といった内容が、問2に含まれているためだと考えられる。

表4. 学力レベル別にみる設問ごとの正答率 (%)

レベル	問1-1	問1-2	問1-3	問1-4	問1-5	問2-1	問2-2	問2-3	問2-4	問2-5	問2-6	問2-7	問2-8	問2-9
A	83.9	16.1	80.6	6.5	3.2	77.4	77.4	96.8	51.6	32.3	61.3	32.3	12.9	6.5
B	73.5	29.4	88.2	2.9	8.8	70.6	67.6	94.1	32.4	44.1	55.9	14.7	11.8	0.0
C	79.1	23.3	81.4	7.0	16.3	53.5	62.8	79.1	37.2	27.9	39.5	14.0	9.3	2.3
D	85.1	29.8	85.1	6.4	4.3	61.7	46.8	89.4	25.5	29.8	36.2	6.4	10.6	6.4
E	76.3	13.2	81.6	7.9	10.5	50.0	34.2	71.1	15.8	5.3	18.4	7.9	0.0	0.0
F	88.9	27.8	77.8	5.6	5.6	38.9	38.9	55.6	11.1	11.1	11.1	5.6	5.6	0.0

本研究では、実験の学習経験が有利に作用する問題内容や形式とは何かということも課題の1つとしている。したがって、アンケート調査で尋ねている問1と問2の類似実験の経験有無（自己認識）によって、試行テストの得点に差が生じるかを分析した。

その結果、問1（問1の合計点）では、「経験あり群」(N=51)の平均が2.18 (SD=0.89)。「経験なし群」(N=164)の平均が1.96 (SD=0.76)となり、t検定の結果、 $t(213)=1.68, p<.10$ と僅かながら「経験あり群」において得点が高かった。問2（問2の合計点）では、「経験あり群」(N=139)の平均が3.19 (SD=1.69)。「経験なし群」(N=76)の平均が3.05 (SD=1.69)と得点差はみられなかった。

さらに、上記2群に学力レベルを加えたのが表5である。二元配置による分散分析を行った結果、問1については主効果、交互作用ともみられなかったが、問2では、学力レベルに主効果が確認された ($F(6,201)=25.4, p<.001$)。ただし、設問ごとにみた場合、問2-1と問2-6につ

いては、「経験あり群」の方が「経験なし群」とよりも僅かながら平均点が高かった。

以上のことから、問1は、類似した実験の経験が得点に影響を及ぼしているものの、問2については、実験経験よりも通常の化学の基礎学力と相関する問題であることが示された。

表5. 学力レベルと実験経験有無別の平均点

レベル	問1		問2	
	有り	無し	有り	無し
A	1.75	1.93	4.33	4.69
B	2.00	2.03	4.09	3.58
C	2.38	1.93	3.31	3.14
D	2.08	2.12	3.33	2.64
E	2.17	1.77	2.04	2.00
F	2.33	1.92	1.50	2.13

最後に、受験者の心理的側面から試行版テストがどのように捉えられたのかを整理する。アンケートの結果からは、「通常のペーパーテスト（筆記試験）の出題される内容とどのような違いを感じましたか」という質問に対して、図

3 のような回答が得られた。

代表的な意見として、動画による実験映像という一連の流れの中から問題を解くために必要な情報を自分で読み取らなければならないことを指摘するものがみられた。これは、デジタル技術のメリットとして想定していた部分が機能していることを示す意見として捉えることができる。また、PBT では、絵や図表によって実験の一部分しか提示されないものの、動画であれば具体的場面の提示が可能であり、自分が経験した実際の実験場面を思い出しながら解答できるといった意見や、映像によって説明されるため、問題を早く理解できるという意見がみられた。これは、評価の妥当性を高められる可能性を示唆する。PBT であれば、実験の様子を文章や図・絵等でしか表現できないため

に、問題文が複雑で難解になることがある。この場合、化学に関する原理や現象の理解(学力)ではなく文章読解力によって正答率が変化することになりかねない。問いたいことをわかりやすく正確に表現できる技術は、妥当性の面からテストの精度を高めることになると考える。

一方、「その設問の問題文しかなく、リード文をもどって見ることができなく、覚えておるか、メモしておかなければならない」「先の問題が分からないので、時間配分が難しい」といったように、当初予想していた解答における利便性の低下に対して不安を感じる意見がみられた。これらのことから、どのように受験者に問題文を提示し、回答させるかというタブレットテストの機能や仕組みに関して、さらに改善を要する必要があることが示された。

- ・映像がある点。すべてが文字でなく、情報処理能力が必要と思う。
- ・画像や動画を見て行うものが多く、動画では見る時間がかかる上、その中から必要な情報を取り出す過程が手間だった
- ・教科書に載っていないことが出題される。
- ・計算に必要な情報を自分で読み取らなければならないことに戸惑った。
- ・実験方法の誤りを答えさせる問題は動画でしかできないと思う。
- ・実際に映像があったから文字でいっぱい書かれるよりは何かは何をしているのか理解できた。
- ・ペーパーテストの問題よりももっと実際に行ったときの対応力が問われているかなと思った。
- ・操作が言葉の説明ではなく実際の映像で説明されるので、問題を早く理解できた。
- ・動画が見られることで、実験した時の状況が思い出しやすく、解きやすかった。
- ・動画で必要な情報を読み取るのはいままでにない新たな感覚だった。
- ・動画をみて答えるので、誤操作の問題など。実験に関する細かいところまで問題にできる点。
- ・動画を見て、実験の誤りを示す問題は、ペーパーテストでは出題できないものだと感じた。
- ・動画を見ることができたり、カラーの写真を見ることができたので考えやすかった。
- ・動画を見るという作業があってペーパーテストより、さらに実験の過程を理解しなくてはいけないと感じた。
- ・普段のテストではあまり聞かれないような実験を用いたものが多かった。

図 3. 「通常のペーパーテスト(筆記試験)の出題される内容とどのような違いを感じましたか」という質問に対する自由記述回答(代表的なものを抜粋)

4. 今後の課題

「佐賀大学版 CBT」の理念を実現するためには、さらにデジタル技術のメリットを見出し、問題開発や解答方式を検討する必要がある。そのためには、PBT を題材に、どのような問題や問い方が、PBT の範疇で思考力や判断力等を評価しているかを明らかにすることが不可欠である。これらの分析を踏まえて、PBT では評価することが難しい領域を、より深く掘り下げて検討しなければならない。

また、アンケートには、「仮に、大学入試と

して本テストを実施する場合、どのような点が気になりますか」という質問もあり、「予測変換で答えがわかる」「システム上のトラブル」

「前の問題に戻れない」「早く終わってからの確認ができない」「キーボード入力に時間がかかる」「時間配分が気になる」「入力しにくい文字や数字がある」「問題文にメモができないのが困る」といった意見が多くみられた。これ以外にも、システムの安定性、問題作成の継続性、入試問題としての秘匿性の確保、動画撮影の環境や技術など、実際の入試で運用するために解

決しなければならない課題がある。今後、試行版テストをもとに、より実践的な観点から開発を進めていきたい。

参考資料

- 中央教育審議会(2014)「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、入学者選抜者の一体改革について～すべての若者が夢や目標を芽吹かせ、未来に花咲かせるために～」(答申)。
- Frederiksen, J. R., & Collins, A.(1989), A systems approach to educational testing. *Educational Researcher*, 18, 27-32.
- 高大接続システム改革会議(2016)「高大接続システム改革会議最終報告」。
- Messick, S.(1989), Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: American Council on Education/Macmillan. 13-103.
- 日本テスト学会(2016),「公開シンポジウム1:大規模 e テスティングの運用モデルの確立を目指して」,日本テスト学会第 14 大会発表論文抄録集,22-31.
- 西郡大・園田泰正・兒玉浩明(2016).「『多面的・総合的評価』に向けた佐賀大学の入試改革」,大学入試研究ジャーナル, No26,pp.23-28.

