

試験の解答過程分析用

コンピュータライズド・テスト・システムの開発

— 大学入試センター 藤芳衛・石塚智 —

1. はじめに

テストの研究には、単に得点だけでなく、解答所要時間等、試験の解答過程の分析が必要である。これを可能とするため、ペン・コンピュータを使用した試験の解答過程分析用コンピュータライズド・テスト・システムを開発した。

従来、解答過程の記録は困難であった。解答過程をVTR撮影する方法はデータの読み取りに膨大な手間暇がかかる。また、これまでのコンピュータライズド・テスト（以下コンピュータ・テストと略称）は問題文にメモが自由に書き込めず、操作性も悪いため、ペーパー・ペンシル・テスト（以下ペーパー・テストと略称）とは異質で利用できなかった。

本システムではペン・コンピュータを使用することにより、ペーパー・テストとほぼ同様な操作感覚で受験することができる。また、解答過程の記録もすべて自動化した。

2. 解答過程分析用コンピュータ・テスト・システムの概容

システムの開発にあたっては従来のマークシート解答方式のペーパー・テストとできるだけ同様な操作感覚で解答できるように設計した。

(1) ハードウェア

ハードウェアとしてペン・コンピュータ（アミティーSV、三菱電機製）を使用する。大きさはA4サイズで、厚さは2.54cmと薄型軽量である。主要諸元は、マイクロプロセッサ：i486SX、25MHz、主記憶容量：8MB、内蔵ハードディスク容量：80MBである。

本体上面は縦長液晶画面、縦1024ドット横

768ドットの高解像度で、白黒16階調である。画面はB5版よりやや小さい。

主入力装置は付属の電子ペンである。画面上の電子ペンのペン先の位置やスイッチの状態が画面の下のタブレットにより常に測定されている。

(2) ソフトウェア

ソフトウェアはビジュアル・ベーシック（マイクロソフト社製）で記述している。MS-Windows3.1上で動作する。

図1に数学の第4問題のコンピュータ画面を示す。画面にはメモも上書きされている。

本システムでは問題文や各設問文をペン・コンピュータの画面に表示する。画面は大学入試センター試験（以下センター試験と略称）等の問題冊子の当該ページをそのままイメージ・スキャナで読み込み、ビット・マップ・データとして作成している。国語のようにルビの付された縦書きの問題文もそのまま鮮明に表示することができる。

各画面の右端又は上部にページ欄を設けた。ペンでタッチすると問題文や設問文等、当該ページを表示する。

各設問文の画面の下部にはマークシート欄を設けた。ペンでタッチするとマーク欄が反転し、解答することができる。また、反転したマーク欄をもう一度タッチすると解答を消すことができる。

ペーパー・テストと同様、問題文にメモが自由に書き込めるようにした。問題文等が表示されている画面に電子ペンで自由に文字や記号及び図形等を、問題文を消すことなく上書きすることができる。また、上書きしたメモだけを消すこともできる。電子ペンのサイ

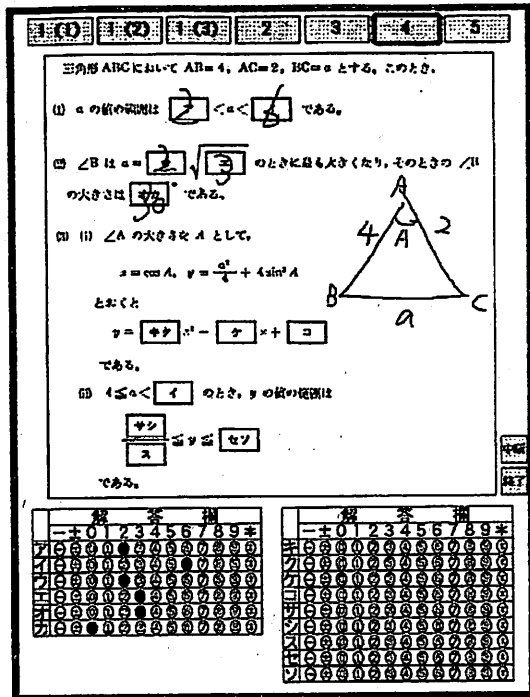


図1 コンピュータ・テストシステムの数学問題の画面

ド・スイッチを押しながら画面に斜めに線を引けば、その線を対角線とする長方形の領域のメモだけを消去できる。

受験生は、電子ペン1本で解答することができる。任意のページを自由に繰りながら問題を読み、画面に文字や記号等、メモを書き込んだり消したりしながら問題を解き、マークシート欄をタッチして解答する。

解答過程の記録はすべて自動化した。ページが繰られるごとにそのページ番号と時刻が、また、マークシート欄がタッチされるごとに解答番号と選択肢番号及び時刻が内蔵のハードディスクに時系列データとして自動的に記録される。

ページ欄をタッチしてから1画面が変わる速さは0.75秒である。参考のため、センター試験の問題冊子を手で1ページ繰る速さを測定した。9名の成人の中央値は、国語が1.06秒、英語が1.18秒である。

3. ペーパー・テストとコンピュータ・テストのテスト方式の違いが試験結果に及ぼす影響に関する実験

3.1 目的

本コンピュータ・テスト・システムの性能を評価するため、従来のペーパー・テストと本コンピュータ・テストとを比較する実験を行う。両テスト方式が試験の解答所要時間、ページを繰った回数及び得点等に及ぼす影響を吟味する。

3.2 方法

ペーパー・テストの解答過程は従来のVTR撮影法により記録する。このため、VTRで記録できるよう、問題冊子を工夫する。各問題は問題文と設問文ごとにページを変える。各ページの上部にはページ番号を印刷する。また、設問文の上部には解答のためのマークシート欄を設ける。

受験生の前方上部からビデオ・カメラで問題冊子を撮影し、その画像信号と時刻表示装置の画像信号を合わせてVTRで記録する。実験終了後、検査者がVTRを再生しながら目視でページが繰られるごとにページ番号と時刻を読みとり記録する。

コンピュータ・テストの解答過程は本システムで記録する。コンピュータ・テストの各画面はペーパー・テストの問題冊子の各ページと対応している。

実験課題は、国語・数学・英語の3教科についてそれぞれ2課題ずつである。各課題は昭和60年度及び61年度の共通1次試験の問題で構成する。国語の各課題は現代文と古文の2問題ずつ。数学の課題は、数学Iの問題を2問題ずつ。英語は、課題1が6問題、課題2が7問題である。

表1に教科・課題別のページ数と配点を示す。尚、ページ数は、英語だけが2課題ともペーパー・テストの方が1ページずつ少ない。これは、両課題とも左のページに長文を表示し、右のページにその設問文を表示した見開き形式の問題が1問題ずつ含まれているためである。

表1 教科・課題別のページ数および配点

課題	ページ数	得点 (配点)
国語1	18	81
国語2	20	87
数学1	2	76
数学2	2	104
英語1	10*	87
英語2	9*	94

*が付いているものはコンピュータ・テストの場合更に1ページ多い。

被験者は、共通1次試験又はセンター試験で国語・数学・英語の3教科を受験した国公立大学の1年生である。ペーパー・テストの被験者は平成1年度16名、平成2年度16名の計32名。コンピュータ・テストの被験者は平成6年度の55名。両被験者群の在籍学部系統は文系及び理系ほぼ同数である。

手続きは、教示語、試験時間を制限しない作業制限法の個人検査である。教科の実施順序は午前中に数学、午後に英語と国語を順次実施する。国語及び英語における二つの課題の実施順序はランダムとする。また、各課題の中の問題の解答順序は任意とする。一方、数学は2課題の計4問題の解答順序は任意とする。

尚、数学の計算用紙については、ペーパー・テストでは問題冊子の左側のページが白紙なのでそれを使用させる。また、コンピュータ・テストでは画面への書き込み以外に、あらかじめ計算用紙を5枚配布する。

実験期間及び場所は、ペーパー・テストが平成1年及び2年の10月から12月、コンピュータ・テストが平成6年10月から12月。実験場所は大学入試センター視聴覚室である。

3.3 結果

(1) 両テスト群の学習到達度

図2(1)に両テスト群の共通1次試験又はセンター試験における国語・数学・英語の3教科

の得点の偏差値の箱型図を示す。平成1年度の被験者は共通1次試験を、一方、平成2年度及び6年度の受験生はセンター試験を受験している。試験制度及び受験年度の異なる被験者の得点を合併して分析するため、各年度の被験者の得点をその年度の全受験生の得点で標準化し偏差値を求めた。図2中の両テスト群の箱型図の横幅の比は両テスト群の人数の比を表す。

両テスト群の国語・数学・英語の3教科ともその得点分布はほぼ同様である。両テスト群の教科別の偏差値の下隣接値、下ヒンジ、中央値、上ヒンジ、上隣接値は各々ほぼ同様である。また、中央値等、両テスト群の得点分布に差異があるかどうかを見るため、Wilcoxonの順位和検定を行った。両テスト群間にすべて有意差は認められない。

(2) 両テスト方式が解答所要時間、ページを繰った回数及び得点に及ぼす影響

表2に解答所要時間、ページを繰った回数及び得点に関する両テスト群の分布に差異があるかどうかを見るため、Wilcoxonの順位和検定を行い、その結果を示す。

図2(2)に教科・課題別の解答所要時間の箱型図を示す。被験者が各ページを開いている時間をその問題文を読んでいる時間あるいはその設問を解いている時間とみなし、教科・課題別に集計した。

解答所要時間の分布に関しては両テスト群は3教科ともほぼ同様である。解答所要時間の中央値はペーパー・テスト群の方がコンピュータ・テスト群よりも100ないし200秒程度長い。また、順位和検定の結果も国語の第2課題に有意差が認められる。しかし、この1課題を除き、他の課題には有意差は認められない。また、解答所要時間の中央値の差異もあまり大きくない。このため、両テスト群の分布はほぼ同様と見ることができる。

図2(3)に教科・課題別のページを繰った回数の箱型図を示す。

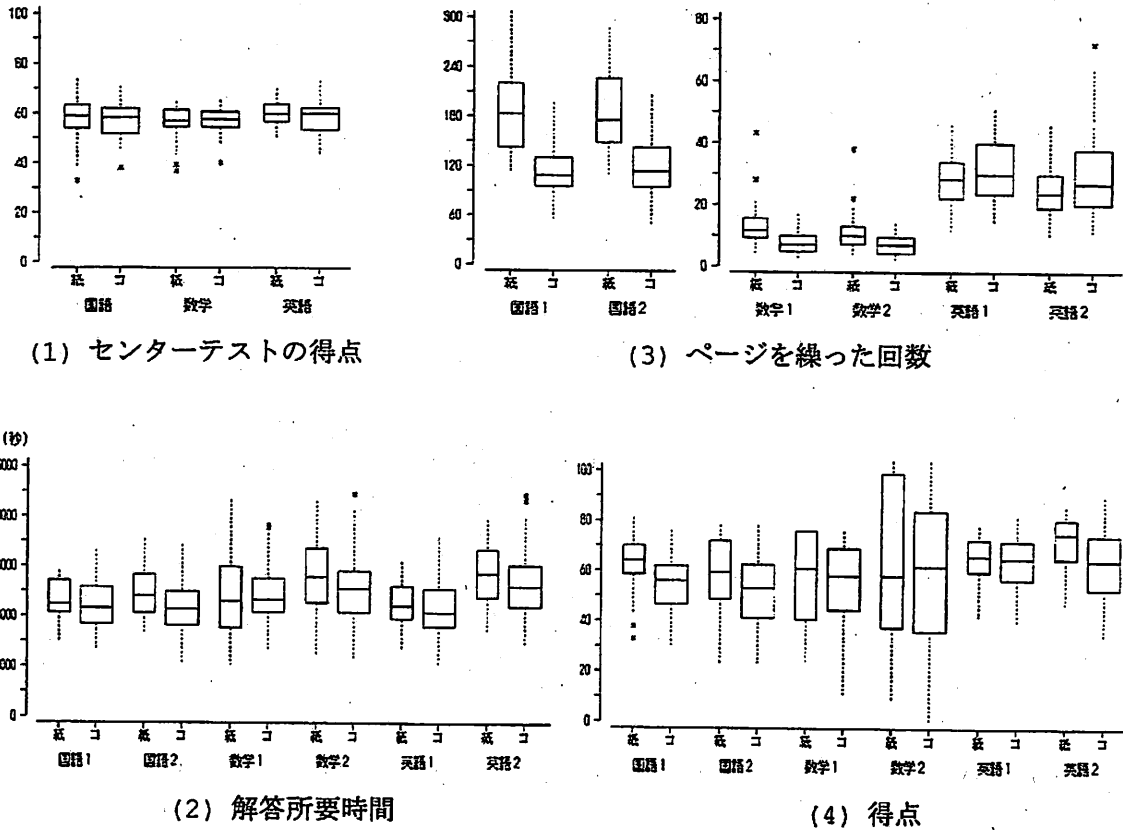


図2 センターテストの得点、ペーパー・テスト群とコンピュータ・テスト群の解答所要時間、ページを繰った回数および得点分布の箱形図
 図中、「紙」はペーパー・テスト群、「コ」はコンピュータ・テスト群を示す。

表2 解答所要時間・ページを繰った回数・得点の順位和検定

課題	解答所要時間		ページを繰った回数		得点	
	順位和	有意水準	順位和	有意水準	順位和	有意水準
国語1	1542	n. s.	2104	***	1765.5	**
国語2	1683	*	2032.5	***	1685	*
数学1	1373	n. s.	1883.5	***	1467.5	n. s.
数学2	1566.5	n. s.	1772	**	1450	n. s.
英語1	1559.5	n. s.	1262.5	n. s.	1472.5	n. s.
英語2	1567	n. s.	1207.5	n. s.	1749	**

表中、「*」は $p < 0.05$ 、「**」は $p < 0.01$ 、「***」は $p < 0.001$ 、「n. s.」は non-significant を表す。

問題のページを繰った回数の分布に関しては両テスト群の差異は教科によって異なる。国語と数学はペーパー・テストの方が有意に多い。国語と数学の4課題ともページを繰った回数の中央値は、ペーパー・テストの方

がコンピュータ・テストよりも3割以上も多い。検定の結果もすべて有意である。一方、英語の分布はほぼ同様である。両テスト群の分布の間の差異もさほど大きくなり、有意差も認められない。確かに英語では2課

題ともペーパー・テストの方が少ない。しかし、英語は2課題ともペーパー・テストの方がコンピュータ・テストよりもページ数が1ページずつ少ない(表1)ため、ページを繰った回数が少なくなったものと推測する。もしページ数が等しければ、ページを繰った回数分布も同様となろう。

図2(4)に教科・課題別の得点の箱型図を示す。

得点分布に関しては両テスト群の差異は教科によって異なる。

国語の得点は、2課題ともペーパー・テストの方がコンピュータ・テストよりも1割程度高い。その差異も有意である。

一方、数学の得点は2課題ともほぼ同様である。有意差も認められない。

英語は課題によって異なる。第1課題はほぼ同様である。しかし、第2課題はペーパー・テストの方が1割程度高く、有意である。

(3) 集団応答曲線

両テスト群の解答所要時間の概容を比較するために、図3に集団応答曲線を示す。集団応答曲線は横軸に解答所要時間を、縦軸に解答を終了した被験者数の相対累積度数をとってプロットしたものである。両テスト群の解答所要時間を詳細に比較するため、教科別に第1課題の両テスト群の二つの集団応答曲線を重ねて図示した。図中、個々の点を実測値を表す。また、右上上がりの二つの曲線が当てはめられたワイブル分布関数である。

国語はペーパー・テストの方がコンピュータ・テストより200秒程度中央付近では右にずれており、解答に時間をかけていることが知られる。しかし、形状はほぼ同様である。

数学は、二つの曲線の位置はほぼ同様である。しかし、形状が若干異なる。

英語は、ペーパー・テストの方が下から中央付近ではやや右にずれており、ペーパー・テストの方が若干時間をかけていることが知られる。しかし、形状はほぼ同様である。

本実験の集団応答曲線にはワイブル分布関数がよく適合する。今回の実験の6課題ともワイブル・パラメータの推定が可能であった。実測値とワイブル曲線とがよく一致していることが図3から読みとられる。

(4) 両テスト得点の内的整合性

表3に教科別のテスト得点の内的整合性の指標としてクロンバックの α 係数を示す。

α 係数を概観すると両テスト群はほぼ同様の傾向を示している。全課題とも0.65から0.70でほぼ同様と考えられる。

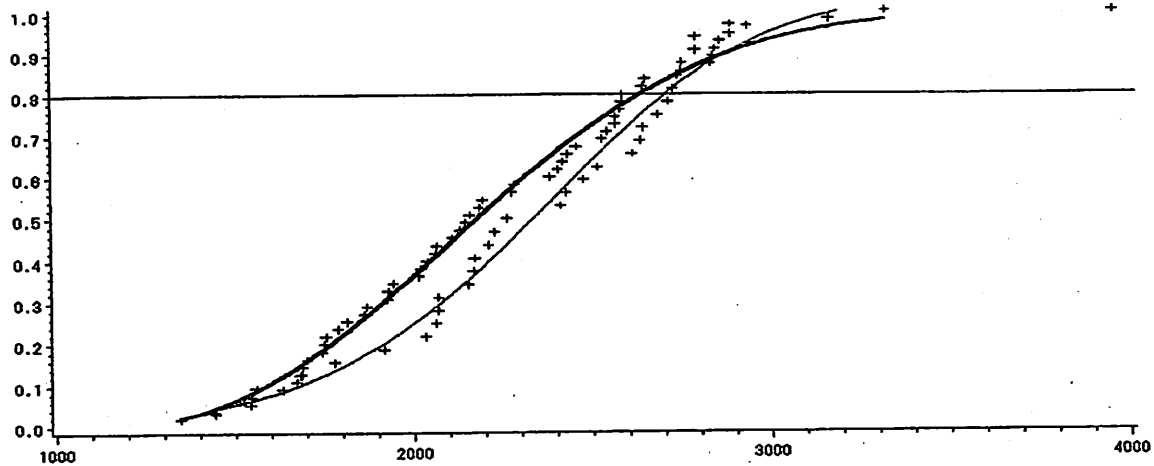
3.4 考察

ペーパー・テストと本コンピュータ・テストの両テスト方式の違いが試験結果に及ぼす影響を比較分析した結果、解答所要時間等、3教科ともほとんど影響しない要因と、ページを繰った回数や得点等、教科によっては影響する要因とが認められるけれども、この傾向を前提に分析すれば、本コンピュータ・テストの結果からペーパー・テストに関する有益な知見を得ることは可能である。

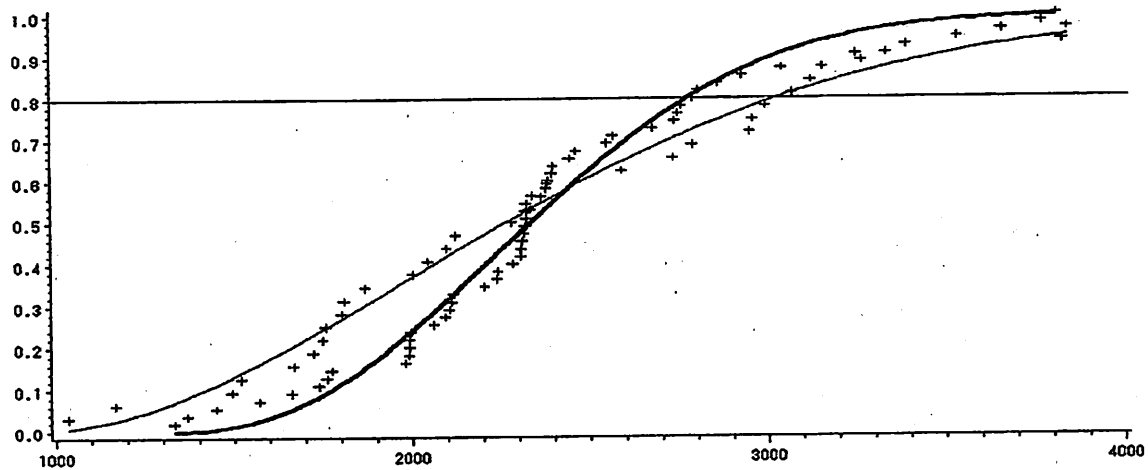
まず、両テスト群は比較可能である。すなわち、両テスト群はセンター試験等、時間制限法で測定した得点分布に関して同様であり(図2(1))、両テスト群は学習到達度に関して等質と仮定できる。

実験結果から両テスト群の解答所要時間の分布はほぼ同様であり(図2(2)及び図3)、両テスト方式の違いは解答所要時間の分布にはあまり影響しないことが知られる。このため、コンピュータ・テストの解答所要時間の分布からペーパー・テストの解答所要時間の分布を推定することが可能である。

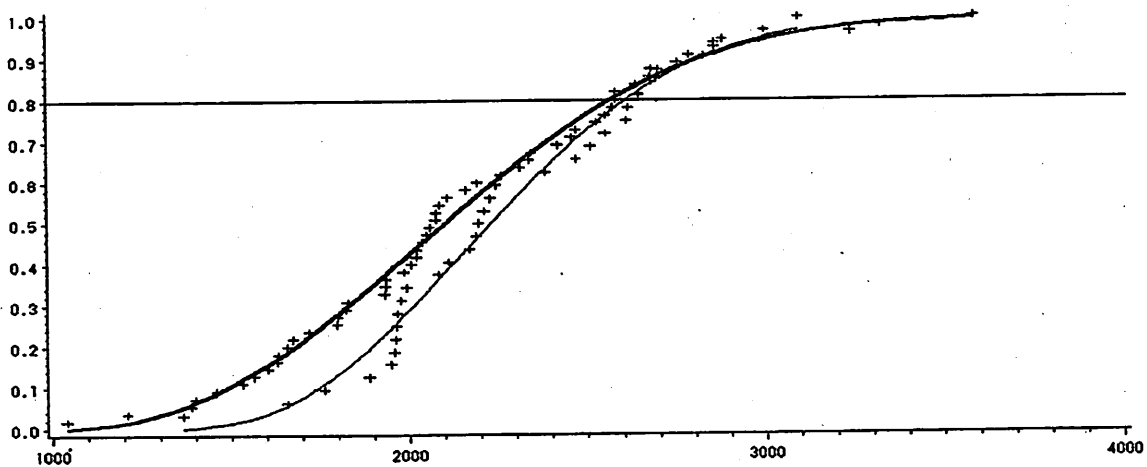
一方、国語の得点分布のように両テスト方式の違いが影響する要因についてはその差異を前提にすれば、コンピュータ・テストの分布からペーパー・テストの分布を推定することも可能である。得点分布に関して数学のように両テスト方式の違いが影響しない教科と



(1) 国語 1



(2) 数学 1



(3) 英語 1

図 3 ペーパー・テスト群とコンピュータ・テスト群の国語 1, 数学 1, 英語 1 の集団応答曲線
 図中, 細線はペーパー・テスト群, 太線はコンピュータ・テスト群を示す. また, 縦軸は相対累積度数, 横軸は解答所要時間 (秒) である.

表3 クロンバックのAlpha係数

科目	問題数	ペーパー・テスト	コンピュータ・テスト
国語	4	0.6975	0.6810
数学	5	0.6532	0.6799
英語	13	0.4509	0.6784

国語のように影響する教科とがある(図2(4))。しかし、この国語の得点の差異も1割程度である。また、テスト得点の内的整合性の指標としての α 係数もほぼ同様である(表3)。このため、コンピュータ・テストの得点分布からペーパー・テストの得点分布を推測することは可能である。

ところで、この国語のコンピュータ・テストの得点が低下した原因としていくつかの要因が考えられる。

第一にペーパー・テストのページの繰りやすさの要因が考えられる。国語は、各問題とも2、3ページにもわたる長文の問題を読みながら各設問を解かねばならない。問題文の見返しが多くなり、見返しが容易なペーパー・テストの方が得点を取りやすいのではないかと推測できる。事実、国語のペーパー・テストのページを繰った回数はコンピュータ・テストよりも3割以上も多くなっている(図2(3))。一方、数学のように見返しをさほど必要としない教科は得点への影響は少ないものと推測できる。

第二に過去の問題の解答経験の有無が考えられる。共通1次試験の本課題の問題はペーパー・テストの被験者にとっては5年ほど前に実施された過去の問題である。一方、コンピュータ・テストの被験者にとっては10年ほど前に実施された過去の問題である。これらの問題を一度見た経験はペーパー・テストの被験者の方が多いためと推測する。一方、数学の得点は過去の解答経験のような知識よりも問題を解く能力により依存しており、得点に差異が認められなかったものと推測する。

第三は実験条件の問題である。ペーパー・

テストは二人ずつ実験を実施した。このため、あまり周りを気にせずに最後まで集中して解答することができたと思われる。一方、コンピュータ・テストは10名ずつ実験を実施したため、半数以上が解答を終了すると、集団心理が働き、この被験者は途中で解答を急ごうとする傾向が出て来る恐れがある。事実、有意ではないけれども、国語のコンピュータ・テストの解答所要時間は両課題とも中央値付近では200秒程度ペーパー・テストよりも短い。

このように、本実験結果には両テスト方式の違い以外の要因も関係している恐れがあるため、吟味実験を必要とする。同一の被験者にペーパー・テストとコンピュータ・テストとを実施して分析する必要がある。

4. 質問紙調査

(1) 目的

今回開発したコンピュータ・テスト・システムを評価するため、質問紙調査を実施する。

(2) 方法

質問紙項目は3種類、20項目である。コンピュータ・テストの画面の見やすさに関して8項目、操作のしやすさに関して5項目である。また、ペーパー・テストとコンピュータ・テストとの比較に関して7項目である。

解答は、5件法で求める。コンピュータ・テストの画面の見やすさ及び操作のしやすさに関しては、「非常に悪い」は-2点、「悪い」は-1点、「普通」は0点、「よい」は1点、「非常によい」は2点とする。また、ペーパー・テストとコンピュータ・テストとの比較に関しては、コンピュータ・テストの方が「非常

表4 質問紙調査の解答結果

	項目	相対度数 (%)				
		-2	-1	0	1	2
コンピュータ・テストの画面の見やすさ	1) 画面の大きさ	4	18	33	26	20
	2) 問題文の量	2	24	35	22	18
	3) 文字の大きさ	7	31	33	15	15
	4) 文字の鮮明度	2	13	44	31	11
	5) ルビの見やすさ	7	16	29	27	20
	6) ページ欄の見やすさ	2	4	16	38	40
	7) ページ欄の状態の見やすさ	2	13	26	24	36
	8) マークシート欄の見やすさ			16	27	56
コンピュータ・テストの操作のしやすさ	9) ペン・タッチのスムーズさ	2	13	35	33	18
	10) ページの繰りやすさ	2	7	35	27	29
	11) ページが変わる速さ		6	24	26	46
	12) マークのしやすさ	4	16	33	24	24
	13) マークの反転の速さ			22	40	38
paper・test とコンピュータ・テストとの比較	14) 問題文の読みやすさ	30	28	28	13	2
	15) 問題文の量	11	35	36	9	9
	16) ページの繰りやすさ	13	35	16	18	18
	17) マーク解答のしやすさ		4	4	6	87
	18) 疲れやすくない	22	9	26	26	18
	19) 単調ではない	6	7	50	19	19
	20) 全体的な受けやすさ	9	31	26	24	11

に悪い」は-2点、「悪い」は-1点、ペーパー・テストとコンピュータ・テストでは「変わらない」は0点、コンピュータ・テストの方が「よい」は1点、「非常によい」は2点とする。

被験者は上述の評価実験でコンピュータ・テストを受験した55名である。

(3) 結果

表4に項目ごとに評価点の相対度数分布(パーセント)を示す。尚、調査項目の15と20にそれぞれ1箇所ずつ無回答があった。

コンピュータ・テストの画面の見やすさについてはほとんどの項目で見やすいとの評価が得られた。1)表示「画面の大きさ」及び、2)1画面に表示される「問題文の量」は問題を解くのに十分である。3)「文字の大きさ」

も普通である。4)「文字の鮮明度」及び、5)「ルビの見やすさ」は見やすい。また、6)「ページ欄の見やすさ」、7)ページ欄で今どのページを読んでいるかを表示している「ページ欄の状態の見やすさ」及び、8)「マークシート欄の見やすさ」についてもすべて見やすいと評価されている。

コンピュータ・テストの操作性についても、すべての項目でよいとの評価が得られた。9)電子ペン入力における「ペン・タッチのスムーズさ」、10)「ページの繰りやすさ」、11)「ページが変わる速さ」、12)マークシート欄の「マークのしやすさ」、13)「マークの反転の速さ」に関して操作性が高いと評価されている。

一方、ペーパー・テストとコンピュータ・テストとの比較に関しては、二つのテスト方

式の違いを反映する項目が認められる。14)「問題文の読みやすさ」はペーパー・テストの方が非常に読みやすい。一方、17)「マーク解答のしやすさ」はコンピュータ・テストの方が非常にしやすいと評価されている。

その他の項目はほぼ同様である。15)1ページ又は1画面に表示される「問題文の量」の十分さは変わらない。16)「ページの繰りやすさ」は若干ペーパー・テストの方がよいようであるけれどもほとんど変わらない。また、18)「疲れやすくない」もほとんど変わらない。19)「単調ではない」は、コンピュータ・テストの方がペーパー・テストよりも若干単調ではないという結果が得られた。

最後に、20)「全体的な受けやすさ」に関してはペーパー・テストとコンピュータ・テストとを比較してもほとんど変わらないという評価結果が得られた。

(4) 考察

本コンピュータ・テスト・システムは、従来のマークシート解答方式のペーパー・テストと比較しても本質問紙調査の解答者に好意的に受け入れられている。

画面については見やすいと評価されている。確かに問題文の読みやすさはペーパー・テストと比較すればペーパー・テストの方が優れている。しかし、小・中学生の頃から毎日長時間コンピュータ・ゲームに慣れ親しんできたこの世代の解答者には液晶画面の見にくさもさほど苦にはならないようである。

また、操作についても非常にしやすいと評価されている。従来のマウスを使用して画面の指示に応答するコンピュータ・テスト・システムには抵抗があることが指摘されている¹⁾。しかし、ペン・タッチ入力による解答は鉛筆よりかえって解答しやすいという評価が得られた。

5 結論

今回ペン・コンピュータを使用して開発し

たコンピュータ・テスト・システムはセンター試験等、マークシート解答方式のペーパー・テストの解答過程の分析用として有用と考えられる。大学1年生に対する比較実験結果によれば、ペーパー・テストとコンピュータ・テスト方式の違いは、解答所要時間等、国語・数学・英語の3教科ともほとんど影響しない要因と、国語の得点等、教科によっては影響する要因とが認められる。しかし、それを前提に分析すれば、本コンピュータ・テストの結果からペーパー・テストに関する知見を得ることは可能である。

質問紙調査の結果からも調査対象の大学1年生の世代には本コンピュータ・テスト・システムは好意的に受け入れられている。画面も見やすく、操作性も優れており、従来のマークシート解答方式のペーパー・テストと比較しても全体的に受けやすさはほとんど変わらないと評価されていた。

本システムの開発は、試験の解答過程の定量的分析を可能にする。従来、センター試験等、入学試験に関する研究は主として得点や正答・誤答に関する分析に限定されてきた。本システムの利用により解答過程をすべて自動的に記録し、分析できるため、試験問題の新しい評価方法の開発を可能にする。

謝辞

データ解析にご援助賜りました静岡大学の山田文康先生及び面倒なペーパー・テストのデータの読みとりを担当していただきました大学入試センター研究補佐員の高瀬道、神田幸子、田上佐江子の皆様方に厚く御礼申し上げます。

参考文献

- 1)野口浩之(1994) パーソナル・コンピュータをベースにした識別性検査に対する受験者の反応。名古屋大学教育学部研究紀要, 41, 25-38.