

適応型テストの回答速度と得点の関係

新井一寿 舩田博之 櫻井麻野

(株式会社 リクルートマネジメントソリューションズ)

【背景と目的】

近年、コンピュータ技術の発達により、紙筆版テスト (Paper & Pencil test ; P&P) のコンピュータテスト化が急速に進みつつある。小社でも適応型のコンピュータテスト (Computer Based Test ; CBT) を提供しており、主に新卒採用選考において多くの企業にご利用いただいている。2006年度には CBT 受検者数は P&P 受検者数を上回り、新卒採用選考において CBT がスタンダードとなったといっても良いだろう。

このような受検機会の増加に関わらず、受検者は CBT の受検に慣れていないと思われる。例えば、適応型 CBT は P&P と違って「全体の問題を見渡して解きやすい問題から手をつけることができない」、「先に回答した問題に戻って見直したり解き直したりすることができない」という特徴があり、初めて受検するときにはどのように回答すればよいか戸惑ってしまい、実力を十分に発揮できない者も多数存在することが想定される。実際、2002年に小社の CBT の受検者 1150 名に実施したアンケートでは、55.8% の学生が「検査全体のボリュームがわからないため、不安だった」、55.7% の学生が「前の問題に戻って回答を修正できないため、やりにくかった」と回答している。

このように CBT に不慣れな場合、回答のペースがつかみにくいため、最初のうちは回答が遅いが最後の方になるにつれて焦ってスピードを上げるといった受検者もいるだろう。こうした場合、本来の実力よりも誤答が多くなり、得点が低くなると予測できる。また、2回目の受検時には回答の進め方を修正し、最初のうちからスピードをあげ、最後の方になるにつれてゆっくり解くといった方略を取るかもしれない。このように考えると、同じ受検者であっても、回答速度は変動しやすいものであり、受検回数を重ねて CBT に慣れることで変化することもあるのではないかと考えられる。こうした回答速度の変化によってテスト結果が変わってしまう可能性もあるだろう。

また、P&P は項目固定式テストで一定の制限時間が設定されており、正答数が多いほど得点が高くなる採点方式である。そのため、速く解くことは回答数、ひいては正答数の増加につながり、速く解き終われば見直しや解き直しもできるので、高得点につながりやすい。受検者も「正答が多いほど得点が高くなる」と考えているため、与えられた制限時間の中でできるだけ多くの問題に回答しようとしているのではないかと推測できる。一方で、適応型 CBT のような項目反応理論 (Item Response Theory ; IRT) を用いた採点方式のテストの場合、速く回答すること、つまり回答項目数が多いことは、直接高得点に結びつくわけではない。しかし、CBT に不慣れであり、P&P に慣れている受検者であれば、CBT 受検時に P&P のようにスピーディに回答し、高得点にならないケースも存在するのではないかと考えられる。

そこで本研究では、CBT における回答速度や回答速度変化を合わせて回答スタイルと呼び、回答スタイルが受検ごとに变化するかどうか、回答スタイルと得点に関係があるかどうか、さらに回答スタイルが CBT と P&P のテスト結果との関係に及ぼす影響について探索的に検討する。

回答速度モデル

ところで、適応型テストにおいて推定精度をテスト終了条件に含む場合には、出題項目数や制限時間

が受検者間で異なるため、1項目あたりの回答時間により速さを比較することになる。しかし、受検者によって出題項目が異なるため、受検者間の回答速度を比較するには、項目による所要時間の違いを考慮した速さの基準が必要となる。藤田・舩田(2005)は、回答時間分布モデルとして分布の有効性を確認した上で、受検者の項目ごとの回答時間を分布の中に位置づけて正規化することにより、個人の回答速度特性の推定を行った。この方法で、項目が異なっても同じ指標で受検者の回答速度を比較できるようにしている。本研究においても、この方法を用いて回答速度を扱うこととする。

分布モデルとは、回答所要時間データの発生モデルをエントロピー最大化原理により理論的に導出したもので、パラメータの推定が非常に容易であることが特徴である(植野,2003)。分布の密度関数は、

$$f(t) = \frac{t^{\alpha-1} \exp(-t/\beta)}{\beta^\alpha \Gamma(\alpha)}$$

で示され、回答所要時間の累積分布 $F(t)$ は、

$$F(t) = \begin{cases} 0 & t < t_0 \\ \int_{t_0}^t f(t) dt & t \geq t_0 \end{cases}$$

で示される。ここで、 t_0 は、各項目における回答に必要な最小限の時間を表す。、 は、それぞれ

$$\alpha = E^2/V \quad (E: \text{実データの平均から } t_0 \text{ をひいたもの}, V: \text{実データの分散})$$
$$\beta = E/\alpha = V/E$$

である。

【方法】

使用尺度：株式会社リクルートマネジメントソリューションズの基礎能力検査 GAT (General Ability Test) の P&P と CBT を用いた。GAT は言語検査と非言語検査で構成されており、後者を対象とした。

分析対象：2008年1月～4月までに、CBTを2回、P&Pを1回受検した328名のデータを使用した。

手続き：

(1) 回答速度特性の算出

CBTの項目ごとに分布のパラメータ、を求め、回答時間分布を再現した。次に、分布の累積分布関数を用いて、受検者の項目ごとの回答時間が集団の中でどの水準にあるのかを算出し、受検者の回答時間特性は正規分布に従うという仮定のもと、その累積パーセントを標準正規分布に置き換えた。この操作によって、回答時間分布が項目ごとに異なっても、同じ指標で受検者の回答時間を評価できるようになる。それぞれの z-score を項目ごとの回答速度 z_j (項目 j の回答速度) とおき、 z_j の平均値をその受検者の回答速度とした。 z_j が小さいほど回答速度が速く、 z_j が大きいほど回答速度が遅いことを意味する。上記の作業を CBT 1回目、2回目それぞれの回答時間データに対して行った。

(2) 回答スタイルによる受検者の分類

回答速度による分類

回答速度 () の分布に従い、上位 30% を「遅い群」、下位 30% を「速い群」、残りの 40% を「中程度群」とし、3グループに分類した。

回答速度変化による分類

項目 j の回答速度 t_j とその次に回答した項目 $j+1$ の回答速度 t_{j+1} の差を $d_j = t_{j+1} - t_j$ とし、受検者ごとに d_j の平均 \bar{d}_j を求めた。この値が正の場合は回答速度が全体として遅くなる傾向があること、負の場合は全体として速くなる傾向があることを意味する。 \bar{d}_j の分布に従い、上位 30% を「遅くなる群」、下位 30% を「速くなる群」、残りの 40% を「変化なし群」とし、3 グループに分類した。

\bar{d}_j と d_j を合わせて、CBT1 回目、2 回目それぞれの受検者を $3 \times 3 = 9$ グループに分類した。

(3) 各グループの得点の平均値の算出

グループごとに、CBT1 回目の得点、CBT2 回目の得点、P&P の得点の平均値を算出した。

また、CBT1 回目と 2 回目のグループを組み合わせ、 $9 \times 9 = 81$ グループの平均値を算出した。

【結果・考察】

(1) 回答スタイルの変化

CBT2 回の受検で、回答スタイルが変化するかどうか確認するため、各スタイルの人数を表 1 に示した。回答速度が 1 回目に速かった人のうち、5割以上の人が 2 回目には「中程度」または「遅い」群になり、回答速度変化についても 1 回目に速くなる（速くなる群）群だった人のうち、約 4 割が 2 回目には変化なし（変化なし群）群、約 3 割が遅くなる（遅くなる群）群に移った。このことから、回答スタイルは変化しやすいものであることが確認された。

表 1 CBT1 回目と 2 回目における各回答スタイルの人数

		CBT2 回目									合計
		速い			中程度			遅い			
CBT1 回目	速い	407	526	342	292	459	386	121	208	189	2930
		512	872	696	359	615	567	117	201	200	4139
		322	545	554	206	372	398	74	120	164	2755
中程度		336	394	223	451	560	442	263	397	382	3448
		521	666	489	609	967	897	315	579	511	5554
		400	522	459	431	682	663	200	366	453	4176
遅い		123	132	88	263	303	233	322	471	366	2301
		201	251	152	432	611	491	467	787	721	4113
		176	228	168	324	467	495	328	615	675	3476
合計		2998	4136	3171	3367	5036	4572	2207	3744	3661	32892

注) は「速くなる群」、は「変化なし群」、は「遅くなる群」を示す。

(2) 回答スタイルと得点の関係

各回答スタイルにおける CBT 2 回受検の得点差

2 回の受検の間の回答スタイルの変化に伴い、得点が変わるかどうかが確認するため、スタイルごとに CBT2 回の受検の得点差（= 2 回目 - 1 回目）の平均値を求めた（表 2）。

CBT1 回目と 2 回目の得点差の全体での平均値は、2.5（表 2 右下）であった。この差は同じ形式のテストを 2 回受検することによる練習効果が反映されたものである。

これと比較して、回答速度が 1 回目に速く 2 回目に遅かった群の得点差は 3.5~6.4 であり、2 回目に得点上昇が大きいことがわかった。この中でも、2 回目に回答速度の遅くなる（遅くなる群）群、変化なし（変化なし群）群の受検者の得点上昇は 4.5~6.4 と顕著であり、1 回目は回答方法に慣れずに焦ってしまったが、慣れることによって 2 回目はゆっくりと確実に回答し、実力を発揮できたのではないかと推察される。一方、回答速度が 1 回目に遅く 2 回目に速かった群の得点差は -1.4~2.0 であった。中でも 2 回目に速くなる（速くなる群）群に至っては -1.4~1.2 であり、得点が下がる受検者も多く存在することがわかった。

このように焦って回答が速くなる受検者が存在する要因として、今回の CBT の仕様が挙げられる。この CBT は回答項目数が決まっていないため、回答時の目安として、回答画面右上に「テスト全体の進捗の目安」と検査が開始されてからの経過時間が比較できるように併せて表示している。さらに画面左下には項目ごとに「項目制限時間」を表示しており、受検者はそれぞれの表示を確認しながら回答を進める

必要がある。特に初めて受検した者にとっては、これらの表示に慣れていないため、回答を急かされている感覚に陥り、結果として焦って回答している可能性がある。

各回答スタイルにおけるCBTとP&Pの得点差

回答スタイルによって、CBTの得点がP&Pの得点と相対的にどのくらい異なるか確認するため、各回答スタイルにおいてCBTとP&Pの得点差(=CBT-P&P)の平均を求めた(表3)。なお、CBTの得点は1回目と2回目の結果の平均値を用いている。

P&PとCBTの得点差の全体での平均値は、-0.3(表3右下)であった。これと比較して、2回とも回答速度が速い群の得点差は-2.0~-2.8と大きく、P&Pの得点の方が相対的に高かった。このことから回答速度の速い回答スタイルの受検者はP&Pの方が得意であると考えられる。先に述べたように、速く解くことで、P&Pでは回答数が増えて見直しや解き直しが可能になる一方、CBTでは速く解いても前に解いた問題には戻ることができず、高得点につながりにくいからであると考えられる。

一方、回答速度が2回とも遅い群の得点差は0.4~2.5であり、CBTの得点の方が相対的に高い傾向にあった。CBTは回答項目数が決まっておらず、回答項目数が少なくても正答が続けば高得点になる一方、P&Pでは回答項目数の少なさはそのまま正答数の少なさにつながり、得点が高くなりにくい。このことから、回答速度の遅い回答スタイルの受検者は、CBTの方が得意であると考えられる。

表2 各回答スタイルにおけるCBT1回目と2回目の得点差(CBT2回目-1回目)

		CBT2回目									合計
		速い			中程度			遅い			
CBT 1回 目	速い	2.1	3.1	3.4	2.3	3.9	4.7	4.7	5.5	6.4	3.7
		1.9	2.3	2.9	3.1	3.8	4.0	3.6	4.5	4.6	3.1
		1.3	2.5	2.0	3.2	3.1	3.7	3.5	5.2	4.7	2.8
	中程度	1.4	2.4	3.2	2.0	3.0	4.1	2.7	3.2	4.4	2.9
		1.0	2.0	2.1	1.9	2.6	3.4	1.9	2.7	4.1	2.5
		0.4	0.7	1.1	0.7	2.0	2.1	1.4	2.0	2.9	1.5
	遅い	1.2	0.5	2.0	2.1	3.1	3.7	2.2	3.5	4.0	2.9
		0.1	0.3	1.2	1.6	1.1	3.1	2.1	2.4	3.3	2.0
		-1.4	0.1	1.3	-0.1	1.0	2.0	0.8	1.5	2.2	1.2
	合計	1.1	1.9	2.2	1.8	2.6	3.3	2.2	2.9	3.6	2.5

注) は 速くなる群, は 変化なし群, は 遅くなる群, を示す。

表3 各回答スタイルにおけるCBTとP&Pの得点差(CBT-P&P)

		CBT2回目									合計
		速い			中程度			遅い			
CBT 1回 目	速い	-2.5	-2.5	-2.2	-2.7	-1.2	-0.9	-1.0	-1.2	0.4	-1.7
		-2.8	-2.4	-2.2	-1.6	-1.2	-0.8	-1.5	0.1	-0.5	-1.7
		-2.8	-2.3	-2.0	-1.5	-1.1	-0.7	-0.1	-0.2	-0.5	-1.6
	中程度	-1.3	-1.4	-0.9	-0.7	-0.3	-0.3	0.0	-0.2	1.0	-0.4
		-1.5	-1.2	-1.2	-0.8	-0.4	0.1	0.0	0.1	1.6	-0.4
		-0.9	-1.5	-0.6	0.8	0.3	0.3	1.9	0.8	1.7	0.2
	遅い	0.9	-1.5	0.2	0.2	0.7	0.6	0.4	0.7	1.6	0.6
		-1.0	0.4	1.0	0.4	0.9	0.7	1.1	1.5	2.4	1.1
		-0.5	0.0	0.5	0.5	1.3	1.1	2.5	2.1	2.2	1.4
	合計	-1.7	-1.7	-1.3	-0.5	-0.1	0.0	0.7	0.8	1.6	-0.3

注) は 速くなる群, は 変化なし群, は 遅くなる群, を示す。

本研究の結果、今回のCBTでは速く解くことが必ずしも高得点には結びつかず、P&Pと比較すると得点が高い傾向にあることが示された。一度受検することで二度目に回答スタイルを変えて実力を発揮している受検者がいることを考えると、CBTの形式に慣れておくことが重要であると考えられる。

今後は、CBTに慣れていない受検者が焦って失敗しないようしくみを構築することが必要となるだろう。また、本研究の結果が、今回のCBTにおける独自のものなのか、それとも適応型テスト一般にもあてはまるものなのか、仕様との関係について検討していく必要がある。

【引用文献】

藤田彩子・舩田博之 2005 CBTにおける受検者の回答速度特性推定の試み 日本テスト学会第3回大会発表論文抄録集, 106.

植野真臣 2003 ガンマ分布におけるWB T所要時間データのオンライン解析システム 教育情報システム学会第28回全国大会論文集, 295-296.