

# 適応型テストを複数回受検した場合の形式への慣れの影響

○櫻井麻野 舛田博之

(株式会社リクルートキャリア 新卒事業本部 測定技術研究所)

## 【背景と目的】

新卒採用選考におけるコンピュータ・テスト (Computer-Based Test: CBT) の利用はテストセンターを中心として Web 上での実施も含めて一般化している。小社でもテストセンターCBT の受検者数は紙筆版テスト (Paper & Pencil test: P&P) の受検者数を大きく上回っており、受検者にとっても CBT を受検する機会の方が P&P を受検する機会よりも多くなっている。

採用における CBT はいわゆる”high-stakes”テストにあたるが、同じく CBT を採用している”high-stakes”テストである TOEFL などと比べると、企業による受検依頼を受けて初めて受検ができる (自分の意思のみで受検することができない)、公式の模擬テストなどがない、といった点が特徴である。また、小社の提供する CBT は項目反応理論 (Item Response Theory: IRT) を用いた適応型テストであるため、採用選考で使われる IRT を用いない CBT での経験を流用することは難しい。学生は複数の企業に応募して受検依頼を受け取ることにより、2 回、3 回と受検回数を増やすことができるが、初回受検の前に類似の経験をすることはできない。その一方で、短期間のうちに複数回受検するという特殊な事情があることから、複数回受検のデータを使って回答方略および受検回数を追うごとの形式への慣れを見ることができると考えられる。

新井・舛田・櫻井 (2009) は、CBT を 2 回受検した受検者の回答速度変化と得点の関係を検証し、1 回目に焦って回答してしまったが、2 回目にはゆっくりと確実に回答した受検者は、回答速度が変わらなかった受検者や速くなった受検者に比べて得点上昇の幅が相対的に大きいこと、また、P&P との比較では、2 回とも回答速度が速い受検者は P&P の方が得点が高い傾向にあるのに対し、2 回とも回答速度が遅い受検者は CBT の方が得点が高い傾向にあることを確認した。

van der Linden *et al.* (2007) は、IRT を用い、サブテストごとに能力値が推定されて次の出題が決定される適応型テストにおいて、1 回のテストの中で新たなサブテストが始まるごとに、受検者が最初の項目では余分に時間をかけ、サブテストの後半になると回答時間が速くなる“ウォームアップ効果”を確認している。

本研究では、CBT を複数回受検した場合の形式への慣れの影響を検証することを目的とし、複数回受検したデータを用い、項目ごとの回答時間と正答率に着目して、受検者の能力推定値から算出した期待正答率と実際の正答率、回答時間の変化の関係を検討する。

## 【方法】

使用尺度：株式会社リクルートキャリアの基礎能力検査 GAT (General Ability Test) の非言語検査を対象とした。同検査は計算 (割合や速度/距離など)、表読み、確率/組み合わせ/集合、推論、推論 (チェックボックス)、推論 (数直線) の 7 つの領域で構成されている。

分析対象：2013 年 1 月～5 月までに CBT を 2 回受検した 101272 名と、そのうち 3 回目も受検した 53880 名のデータを使用した。

手続き：

(1) 対象項目の絞り込み

CBTの項目ごと、1回目から3回目までの受検回数ごとに各項目が出題された受検者全体の能力推定値の平均と標準偏差を算出し、1回目と2回目、2回目と3回目で能力推定値の平均の変化が0.2以内、標準偏差が0.1以上縮小していない項目のうち、各回の受検者人数が400名以上の全110項目を対象とした。

対象項目の領域ごとの内訳は表1のとおりである。

表1: 領域ごとの対象項目数

計算	53
表読み	7
確率/組み合わせ/集合	9
推論	20
推論(チェックボックス)	10
推論(数直線)	11

(2) 回答時間の算出

CBTの項目ごとに1回目から3回目までの受検回数ごとに各項目が出題された受検者全体の回答時間の平均値を算出した。<sup>1</sup>

領域ごとの集計結果は表2のとおりである。

表2: 領域ごとの回答時間平均の推移(負の値のセルに網掛け)

	回答時間(秒)						
	1回目	2回目	3回目	4回目	2回目-1回目	3回目-2回目	4回目-3回目
計算	161.17	179.52	179.04	-	18.35	-0.48	-
表読み	279.13	267.00	273.02	276.04	-12.13	6.02	3.02
確率/組み合わせ/集合	153.84	157.48	157.51	-	3.64	0.03	-
推論	140.32	151.18	150.82	-	10.86	-0.36	-
推論(チェックボックス)	144.44	159.49	158.68	-	15.05	-0.82	-
推論(数直線)	128.95	142.40	142.05	-	13.45	-0.34	-

(3) 実際の正答率集計と期待正答率の算出

CBTの項目ごとに1回目から3回目までの受検回数ごとに各項目が出題された受検者全体の正答率を集計し、各項目が出題された受検者全体の能力推定値 $\theta$ の平均および標準偏差と各項目の特性値から、IRTの2パラメタ・モデルを使用して項目ごとの期待正答率を算出した。<sup>2</sup>

領域ごとの算出結果は表3-表5のとおりである。

表3: 領域ごとの回答時間平均の推移(1回目と2回目: 負の値のセルに網掛け)

[1回目と2回目]	期待正答率			実際の正答率			実際の正答率		実際の正答率上昇分 -期待正答率上昇分
	1回目	2回目	2回目-1回目	1回目	2回目	2回目-1回目	1回目	2回目	
計算	0.509	0.540	0.032	0.526	0.580	0.054	0.017	0.040	0.022
表読み	0.722	0.745	0.023	0.559	0.527	-0.032	-0.162	-0.217	-0.055
確率/組み合わせ/集合	0.380	0.410	0.030	0.516	0.544	0.028	0.137	0.134	-0.003
推論	0.499	0.531	0.032	0.454	0.504	0.049	-0.045	-0.027	0.017
推論(チェックボックス)	0.455	0.467	0.013	0.338	0.375	0.037	-0.116	-0.092	0.024
推論(数直線)	0.252	0.270	0.018	0.392	0.432	0.040	0.141	0.162	0.022

表4: 領域ごとの回答時間平均の推移(2回目と3回目: 負の値のセルに網掛け)

[2回目と3回目]	期待正答率			実際の正答率			実際の正答率		実際の正答率上昇分 -期待正答率上昇分
	2回目	3回目	3回目-2回目	2回目	3回目	3回目-2回目	2回目	3回目	
計算	0.540	0.547	0.006	0.580	0.586	0.006	0.040	0.039	0.000
表読み	0.745	0.748	0.003	0.527	0.527	-0.001	-0.217	-0.221	-0.004
確率/組み合わせ/集合	0.410	0.446	0.035	0.544	0.565	0.021	0.134	0.120	-0.014
推論	0.531	0.530	-0.001	0.504	0.501	-0.003	-0.027	-0.029	-0.002
推論(チェックボックス)	0.467	0.483	0.015	0.375	0.400	0.024	-0.092	-0.083	0.009
推論(数直線)	0.270	0.280	0.011	0.432	0.437	0.005	0.162	0.157	-0.006

表5: 表読みの回答時間平均の推移 (3回目と4回目: 負の値のセルに網掛け)

[3回目と4回目]	期待正答率			実際の正答率			実際の正答率		実際の正答率上昇分 -期待正答率上昇分
	3回目	4回目	4回目- 3回目	3回目	4回目	4回目- 3回目	3回目	4回目	
表読み	0.741	0.815	0.074	0.531	0.524	-0.007	-0.210	-0.291	-0.081

【結果・考察】

(1) 回答時間と正答率

回答時間は1回目と2回目を比較すると長くなり、2回目と3回目ではほとんど変わらなかった。正答率は1回目と2回目を比較すると上昇し、2回目と3回目ではほとんど変わらなかった。ただし、表読みという領域の問題のみ、回答時間が2回目で短くなったあと3回目で再び長くなっており、正答率も2回目で下がり、3回目以降はほとんど変化がなかったことから、3回目と4回目の変化についても検証した。4回目では受検人数が少なくなっているため、200人以上に出題されている6項目を対象とした。この結果、3回目と4回目の変化がほかの領域の2回目と3回目の変化に類似していることがわかった。これは、ほかの領域の項目は1回の受検で2題-3題出題されるのに対し、同領域の項目は1回の受検で1題しか出題されないことから、2回受検した段階で初めてほかの領域の1回受検分に相当する項目に回答したことになるためと考えられる。

この結果から、CBTでの受検に際し、受検者が各項目にかける回答時間は、それぞれの領域の問題に2, 3回程度回答した段階で最適な値に落ち着くと推察される。

(2) 期待正答率と実際の正答率

受検回数ごとに期待正答率と実際の正答率を比較すると、1回目から2回目では実際の正答率上昇分が期待正答率上昇分を上回り、2回目から3回目では実際の正答率上昇分と期待正答率上昇分にはほとんど差がない。このことから、1回目と2回目では、期待正答率の上昇分を上回る実際の正答率上昇分に表れていると考えられる。推論(チェックボックス)、推論(数直線)の2領域においては、実際の正答率上昇分と期待正答率の上昇分の差が大きかった。この2領域はその他の領域が択一式であるのに対し、該当する選択肢をすべて選ぶ回答方式で、2013年より新たに搭載された形式である。そのため、既存の領域と比べて慣れによる影響がより大きかったものと推察される。

(3) 回答時間変化の割合と慣れの影響

実際の正答率上昇分と期待正答率上昇分の差を慣れの影響と置き、その数値が大きかった計算と推論(チェックボックスおよび数直線)について、回答時間の差を先に受検した回の回答時間で割った回答時間変化の割合と慣れの影響の関係を見るため、項目ごとの回答時間変化の割合と慣れの影響を図1-図4に示した。1回目から2回目では正の相関があり、2回目から3回目ではほとんど変化がなかった。

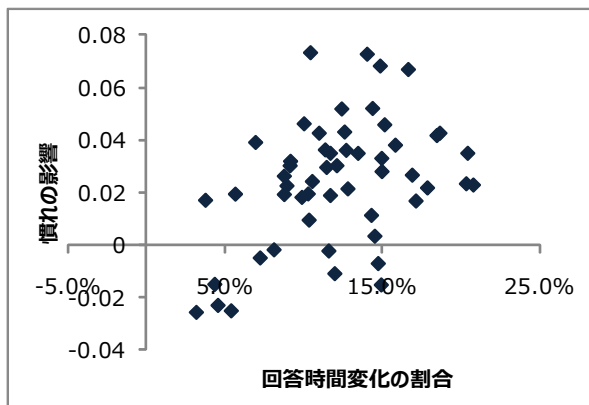


図1 計算: 回答時間変化の割合と慣れの影響 (2回目-1回目)

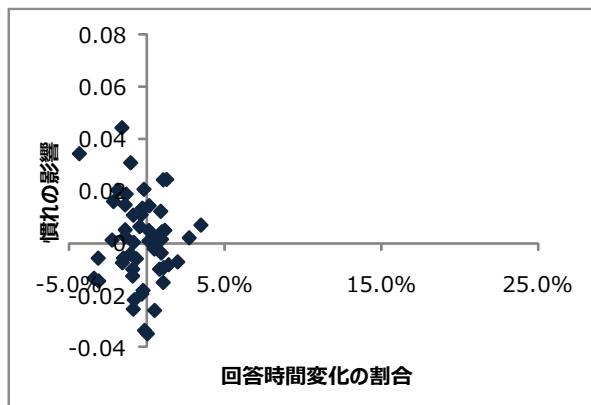


図2 計算: 回答時間変化の割合と慣れの影響 (3回目-2回目)

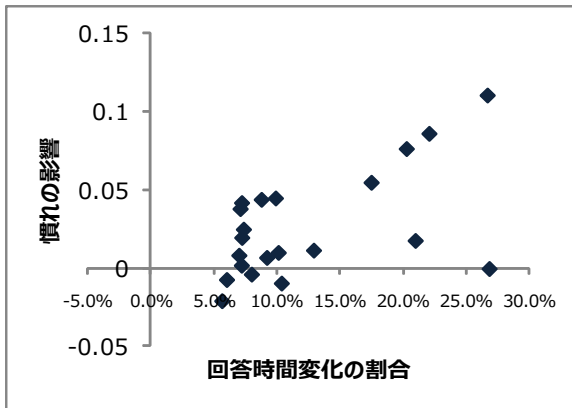


図3: 推論 (チェックボックスおよび数直線) 回答時間変化の割合と慣れの影響 (1回目-2回目)

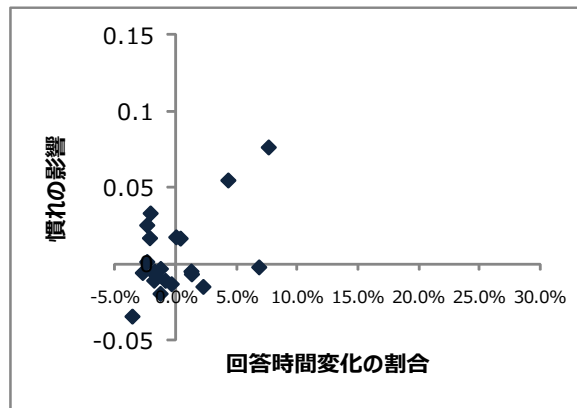


図4: 推論 (チェックボックスおよび数直線) 回答時間変化の割合と慣れの影響 (2回目-3回目)

本研究の結果、CBTを2回以上受検した場合、1回目の受検では十分な時間をかけずに回答した結果、実力を存分に発揮できなかった受検者が、2回目以降に少し落ち着いて回答する傾向があることが示された。また、落ち着いて回答した結果は、期待正答率の上昇分を上回る正答率の上昇にも表れている。この現象は、初回受検の経験を経て、2回目には落ち着いて問題に取り組むという“カームダウン効果”とも呼ぶべきものである。適応型テストでは正答していくと項目の難度が上がるため、P&Pとは手ごたえが異なること、また、初回受検の前に類似の経験をする機会がないことから、受検者はこの“カームダウン効果”を経て初めて本来の実力を発揮できるとも考えられる。今後は受検者が初回受検時から実力を発揮できるような仕組みを整えることが必要となるだろう。

#### 【引用文献】

新井一寿, 舛田博之, 櫻井麻野 (2009) 適応型テストの回答速度と得点の関係 日本テスト学会 第7回大会発表論文抄録集

van der Linden, W.J., Breithaupt, K., Chuah, S.C., & Zhang, Y. (2007). *Detecting Differential Speededness in Multistage Testing*. *Journal of Educational Measurement*, 44(2), 117-130

<sup>1</sup>表読みの領域のみ4回目まで算出(受検人数は31008名)。

<sup>2</sup>表読みの領域のみ4回目まで算出(受検人数は31008名)。