

テスト受験者の能力と解答時間の関連に関する基礎的検討

堂上 哲^{†1} 菖木 禎史^{†1} 宇佐川 毅^{†1}

e-Learning システムにおいて、受験者の成績評価手法としては、素点方式や項目応答理論が用いられているが、解答時間の制限が強い試験などでは解答速度を要求される場合も少なくない。本稿では、解答時間および解答の正誤情報を分析し、受験者の能力推定の可能性について基礎的検討を行ったので、報告する。具体的には、SPIの総合能力セクションの例題を Moodle 上に構築し、各設問についての解答時間を計測するとともに、解答の正誤および総合点などとの相互関係を分析検討した。

A preliminary study on a relationship between examinee's ability and response time

SATOSHI DOUNOUE,^{†1} YOSHIFUMI CHISAKI^{†1}
and TSUYOSHI USAGAWA ^{†1}

There are various methods to evaluate examinee's ability obtained by item response theory as well as raw score. When a test is designed to allow limited time, response speed of examinee is also required. This paper reports a preliminary study on the relationship between score and response time, and discusses on a possibility to estimate examinee's ability. Response time of SPI (Synthetic Personality Inventory) -like test is measured using Moodle, and it is analyzed with the reflection of examinee's response as well as total score.

1. はじめに

大学等の教育機関では LMS(Learning Management System) が広く活用されている。特に学習者の習得状況を確認するため、多様な小テストが毎週のように利用されている。し

かし、テストの難易度の制御や学習者の能力評価については必ずしも系統的には行われておらず、教員の経験等に依存する¹⁾。テストの能力推定方式としては、素点方式及び項目応答理論に基づく手法があげられる。素点方式では重み付けされた点数の総和で受験者の能力を測定しており、問題個々の分析をすることは容易ではない。一方、項目応答理論では項目困難度数・テスト曲線などを用い受験者の能力を推定するが、大学の専門科目程度の受講者数からこれらを精度良く求めることは容易ではない。他方、項目応答理論を用いたテストにおいてはしばしば回答所要時間として時間のパラメータが利用されるが、一般には能力測定のための直接的なパラメータとしては用いられていない²⁾。さらに、CBT などの適応型テストにおいても解答時間が利用されるが、これも能力測定の直接的なパラメータとしては用いられていない³⁾。

解答時間を用いたパラメータとしては Wise と Kong により提唱された Response Time Effort がある。この Response Time Effort は、Scholastic Assessment Test(以下 SAT)において、個々の設問の解答時間を計測することでその設問に真面目に解答したのか、そうでないかを分類し、そうでないデータを取り除くために利用される⁴⁾。

日本では英語能力を評価する TOEIC や 受験者の総合能力を測定する SPI がよく利用されており、TOEIC や SPI は多岐選択式のマークシート方式が採用され、難易度の高い問題に対して短い時間で正確に解く能力が求められるとされている⁶⁾。従って、このような傾向のテストでは、設問毎の解答時間が重要なパラメータになると推測される。

本論文では、始めに LMS の 1 つである Moodle 用に作成した解答時間を測定可能なモジュールについて述べる。次に、コンテンツを実行する際に解答時間以外の動作である動作時間を測定により求め、SPI 総合能力検査の類似問題を就職試験を控えた学生を含む 7 名に受験してもらい、解答時間を測定した。その後、実験で得られた結果から Response Time Effort を算出後、閾値の検討を行った。また、結果より累積分布図を作成し、測定された解答時間とテスト受験者の能力に何らかの関係があるかについて検討を行った。

2. 従来の能力推定方式

2.1 素点方式

素点方式とは、問題に対して点数の重み付けを行い、満点(100点や50点)における得点で受験者の能力を推定する方式である。点数の重み付けは、出題者が独自に設計することになっており、その重み付けの幅は定められていない。しかし、素点方式では正解か不正解かの情報と総得点でしか能力を推定しないため、受験者が本当に理解しているかの判断がで

^{†1} 熊本大学
Kumamoto University

きないという問題点が存在する。

2.2 項目応答理論

項目応答理論では、テスト問題や受験者の母集団の違いに影響されないスコアを算出する方式である。この理論は項目応答曲線(テスト曲線)と呼ばれる受験者の能力ごとに、その問題に正解する確率を表した曲線を基にスコアが算出される⁷⁾。また、逆に項目応答曲線から受験者の能力を推定することも可能である。さらに、項目応答理論には各設問毎の配点は存在せず、前述の応答曲線のみを用いてスコアが算出される。その為 TOEIC のようにスコアが 10 ~ 990 点という、全問不正解でも 10 点は獲得できるシステムとなる。

しかし、項目応答理論は、項目応答曲線の精度でスコアが決まってしまうためサンプル数が多くないと(最低でも数百のサンプルが必要)利用するのが難しい、つまり大規模テストでしかを利用することが出来な。このため、大学での単独の講義の受講者を対象した規模での利用は必ずしも容易ではない⁸⁾。

従って、比較的少数の係数によって、精度良く項目応答理論を利用できれば、講義等で用いる小テストの難易度の制御との効果についてより深い分析が可能だと考えられる。

3. 解答時間に注目した先行研究

3.1 解答時間を用いた出題問題のコントロール

Computer-adaptive Testing では、受験者の解答パターンに応じて、困難度の異なるアイテムを選んで出題することが可能なテストであるが、この解答パターンに時間を加えることで、高い精度で能力の推定を可能にしている。この Computer-adaptive Testing では、項目応答理論を用いているため、この受験者ごとに異なる問題を解いても、能力推定が可能となっている⁵⁾。

3.2 Response Time Effort

解答時間に注目した先行研究として Kong らの Response Time Effort がある⁴⁾。テストによって受験者から得られた解答は、全て真面目に解いているとは限らない。そこで、テスト項目毎の解答時間閾値 T_i を求め、Response Time Effort を用いることで、真面目に受験していない可能性がある解答を取り除くことができるとしている。 i をテスト項目、 j を受験者、 k をテスト項目数とし、 RT_{ij} を受験者 j のあるテスト項目 i の解答時間とすると、式 (1) は努力していない受験者の解答時間の最大値を T_i とし、 RT_{ij} と比較することで SB_{ij} の受験者の行動値を求められることを表しており、式 (2) はこの SB_{ij} をテスト項目 k について利用し、被験者 j の Response Time Effort RTE_j を求めている。 RTE_j は

受験者 j の Response Time Effort となっている。Kong らは Response Time Effort はその値が 0.95 以上であれば、その受験者は高い確率で真面目にテストを受験しているとしている。

$$SB_{ij} = \begin{cases} 1 & \text{if } RT_{ij} \geq T_i, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$$RTE_j = \frac{\sum_{i=1}^k SB_{ij}}{k} \quad (2)$$

3.3 解答時間閾値 T_i の設定

上記の Response Time Effort における、解答時間閾値 T_i の設定について、Kong らは下記の 3 種類の方法を提案している。3 秒を閾値とする設定 (3SEC)、問題文を読む時間 (Reading Time) を考慮した設定 (READ)、専門家による見た目 (Inspect) 判断 (INSPECT) の 3 種類である。本論文では、SAT ではなく SPI を用いるが Kong らと同様にするため閾値は 3 秒とした。また、INSPECT 視察によるものであるため今回は除外した 2 種類を用いる¹⁰⁾。なお、M は Method の略である。

(M1)常に $T_i = 3s$ (3SEC)

(M2)問題文の長さに応じて、 $T_i = 3s, 5s, 10s$ とする (READ)

4. 解答時間測定システム

4.1 Moodle を用いたシステム

本論文では、解答時間を測定する為のシステムとして、Learning Management System(以下 LMS) の 1 つである Moodle を用いた。Moodle は PHP プログラム及び MySQL を用いたデータベースで構成される LMS であり、世界中で利用実績がある LMS の 1 つである。その Moodle 内に解答時間を測定する為のモジュールを組み込み、SPI 総合能力検査の類似問題コンテンツを実装し、解答時間を測定した。

本システムにおいて Moodle を用いた理由としては下記の 3 点が挙げられる。

(1) 学習管理が容易

Moodle は LMS の 1 つである為、ユーザの受験履歴、受験結果、問題内容の管理(どの問題で正解または不正解であったかの確認)が Web 上で可能である。また、コンテンツの追加なども全て GUI 上で行うことができ、PC の操作が不得意な教員でもスムーズにシステムを利用することが出来る。

- (2) 拡張・改変が可能で機能の追加が自由
Moodle にはモジュール機能と呼ばれる拡張用機能があり、他のユーザが作成した機能を利用可能である。
- (3) オープンソースである
インストールされた Moodle 内のファイルは全てオープンソースであり、全ての機能を無償で利用することができる。

4.2 Quiz + RT モジュール

解答時間は、Moodle 上に新規にモジュールを作成することで測定した。この新規に作成したモジュールは、Moodle に含まれている Quiz モジュールを基に作成し、Quiz モジュールに 解答時間 (Response Time) 情報を付加したモジュールであるため、"Quiz + RT モジュール"と呼ぶ。Quiz モジュールでは、多岐選択問題を受験する場合、MySQL データベースに受験日、問題 ID、受験者 ID、正誤情報が記録される。

そこで、Quiz + RT モジュールでは、PHP プログラムの変更、MySQL データベーステーブルの変更を行い、解答時間を測る。PHP プログラムで、タイムスタンプを取得し、MySQL データベースに追加した解答時間テーブルにテスト情報を記録した。解答時間の測定タイミングは、問題読み込み時のタイムスタンプ及び決定ボタンクリック時のタイムスタンプを利用し、これをデータベース上に保存し、結果出力時に 2 つのタイムスタンプの差を求めている。従って、Moodle のデータベースには、受験日、問題 ID、受験者 ID、正誤情報、解答時刻が記録される。

5. 実 験

5.1 実験条件

下記に、本論文で測定実験を行った際のシステム条件、コンテンツ、受験者について述べる。

5.1.1 システム条件

利用する LMS は Moodle で、Quiz + RT モジュールを利用し、解答時間を測定する。動作環境を統一するため、同一種類の PC、同一 Web ブラウザをフルスクリーン表示 (1280x800) で利用させた。Quiz + RT モジュールでは Quiz モジュールで利用できる全ての出題形式が利用できるが、今回はマウス操作のみで全ての操作が可能なる多岐選択問題形式を用いた。最大解答時間の制限はかけていない。また、Moodle ではテスト項目毎もしくは複数問毎に解答を決定することができるが、本実験では解答時間を測定するため、テスト項目毎に解答

を決定する設定とした。

5.1.2 コンテンツ

今回、Moodle に実装した解答時間測定コンテンツとしては、今回は SPI 総合能力検査に類似したテストを用いた。SPI を解答時間測定コンテンツに用いた理由は次の 4 点である。

- (a) 短時間で多数の問題を解く形式のテストである

SPI は受験者が問題を解く上で、基礎知識が求められ、比較的難易度の低い問題が多数出題される為、解答速度が求められる。このため、解答時間を測定する上で適していると判断した。

- (b) 受験者に求められる能力が明確である

SPI は下記 3 項目を総合能力として定義しており、解答時間と受験者の能力についての検討が可能であると考えた。なお、Abi は Ability の略である。

- (Abi1) 言語的理解

文の要素である語の意味を正しく理解し、文章の構成や要旨を的確に理解する力

- (Abi2) 論理的思考

すでに格闘した情報を基に、新しい情報や的確な判断を導き出す力

- (Abi3) 数量的処理

加減乗除の計算能力や、グラフ・表を正確に解釈する力

- (c) 擬似チャンスレベルが低い

SPI 試験は 8 択問題であるため、擬似チャンスレベルが 0.125 となる。

- (d) 受験者のモチベーションを維持可能

受験者のモチベーションが低いと、問題の解答を進めるにつれ、不真面目に解答する可能性が生じる。受験者として、就職活動を控えた大学学部 4 年生および大学院修士課程の学生を用いているため、就職試験に用いられる SPI の類似問題を用いることで、モチベーションが維持されることが期待される。

また、作成した問題は下記 4 項目について同難易度と思われる問題を各 5 問、計 20 問となっている。なお Ca は Category の略である。また、[] 内は求められる能力を表している。

- (Ca1) n 進法問題 [数量的処理・論理的思考]

- (Ca2) 命題 [言語的理解・論理的思考]

- (Ca3) 順列・組み合わせ [数量的処理・論理的思考]

- (Ca4) 鶴亀算 [言語的理解・数量的処理]



図 1 Moodle 上に表示された SPI 総合能力問題の一例で、受験者は問題文を読んだ後、a～h のラジオボタンをクリックし、決定ボタンを押すことで解答時間を測定することができる



図 2 Moodle 上に表示された解答時間バイアス測定問題で、受験者は ○ の選択肢 (この問題では e) を選び、決定ボタンを押すことで、解答時間バイアスを測定することができる

上記いずれも 8 択問題であり、出題順序、選択肢順序はランダムとした。

図 1 は、Moodle 上に SPI 総合能力検査の類似問題コンテンツを実装し、実際のテスト画面を表示したものである。

また、Moodle 上に実装したコンテンツにおいて、測定できる解答時間は、次の (Act1)～(Act4) の 4 つの解答に関わる動作の合計時間となる。なお Act は Action の略である。

(Act1)問題が表示され、その問題を読む

(Act2)解答を考える

(Act3)選択肢のラジオボタンをクリックする

(Act4)決定ボタンをクリックする

この上記 4 動作のうち (Act3) 及び (Act4) の動作に対応する時間は、解答を得た後の動作となり、厳密な意味での問題を解答するために必要な時間ではないため、本論文では以下”解答時間バイアス”と呼ぶこととする。また、(AC1)の問題表示時間は、イントラネット環境での受験であるため、十分に小さいと考えられ、ここでは近似的に 0 と見なした。

5.1.3 受験者

受験者には就職活動を控えた大学学部 4 年生および大学院修士課程 7 名を用いた。また、受験者には受験前に「このテストは解答時間を測定しており、可能な限り速く問題を解くように」との教示を与えた。

5.2 解答時間バイアスの測定

前述の解答時間の定義で述べた (Act1) 問題を読む時間、(Act2) 解答を考える時間をほぼ 0 秒にすることで解答測定を試みた。具体的には、図 2 に例示するような「8 択の中に 1 つだけある ○ を選択してもらおう」というという教示をしておけば個々のテスト項目の問題を読む必要はなくなり、解答を考える時間も必要ではないため、(Act3) 選択肢のラジオボタンを選択・クリックし (Act4) 決定ボタンをクリックするだけである。この問題を受験者 7 名に対して 20 問提示して「解答時間バイアス」を測定した。

その結果、全受験者全問題の平均解答時間は 2.7 秒であった。本論文では、この値を解答時間バイアスの推定値とする。

5.3 解答時間測定

(Ca1)～(Ca4) の 4 項目を各 5 問、計 20 問について、7 名の受験者に対して解答時間を測定した。図 3 は全受験者の解答時間と正誤情報 (0 が不正解、1 が正解) の分布図である。分布図の横の数値はそれぞれ合計点、テスト項目 (Ca1) ～ (Ca4) における点数を表している。また、縦に引かれた実線は正解の解答時間の中での最小値、最大値を示しており、破線は正解の解答時間の平均値を示している。

表 1 各受験者の全問題に対する Response Time Effort(ExaA ~ ExaG は各受験者を表している)

	ExaA	ExaB	ExaC	ExaD	ExaE	ExaF	ExaG
3SEC	1.00	1.00	1.00	1.00	1.00	1.00	1.00
READ	1.00	0.95	1.00	1.00	0.89	1.00	1.00

5.4 Response Time Effort の算出

5.4.1 3SEC による T_i の設定

Kong らが用いた $T_i = 3s$ を本実験でも用いる。この $T_i = 3s$ は問題文を読むことなく、不真面目に解答した時の閾値であると考えられる。

5.4.2 READ による T_i の設定

READ による T_i の設定として、本論文では 8 秒を用いる。その理由として、Kong らの手法では問題の平均文字数を用いて T_i を設定しており、英語と日本語では 1 文字の意味合いが異なる可能性があるが、今回はこの平均値をそのまま利用する。作成したコンテンツの問題文の平均文字数は 60 文字となっており、この値を日本人の平均読書速度である 700 文字/分を用いることで⁹⁾、問題を読む時間を平均 5 秒とし、これに解答時間バイアスの推定値 2.7 秒を考慮した 8 秒を中間の閾値とした。また、4 秒, 8 秒, 12 秒を T_i の値と設定する。それぞれ問題カテゴリの平均文字数が 30 文字以内, 30 文字より多く 90 文字より少ない, 90 文字以上の時に利用される。(Ca1) で $T_i = 4s$, (Ca2) で $T_i = 12s$, (Ca3), (Ca4) で $T_i = 8s$ とした..

5.4.3 閾値の検討

表 1 は各受験者の全問題に対する手法別の RTE_j である。表 1 の結果より 3SEC では不真面目に解答しているテスト項目は存在しないが、READ では不真面目に解答している可能性のあるテスト項目があり、解答時間閾値以下の解答が 4 問 確認された。この 4 問について真面目にテストを受けていないのであれば、解答時間閾値 T_i 以下の解答は擬似チャンスレベルに近くなるはずであるが、解答時間閾値以下の解答は全て正解であった。つまり、本実験においての閾値として 3SEC を用いる事が妥当であると考えられる。

6. 検 討

6.1 特異点に関する検討

図 3 で示された、各受験者の解答時間及び正誤の分布図では、他の解答時間から大きく外れている点が存在する。本論文ではこの大きく外れた点を特異点と呼ぶことにし、その特

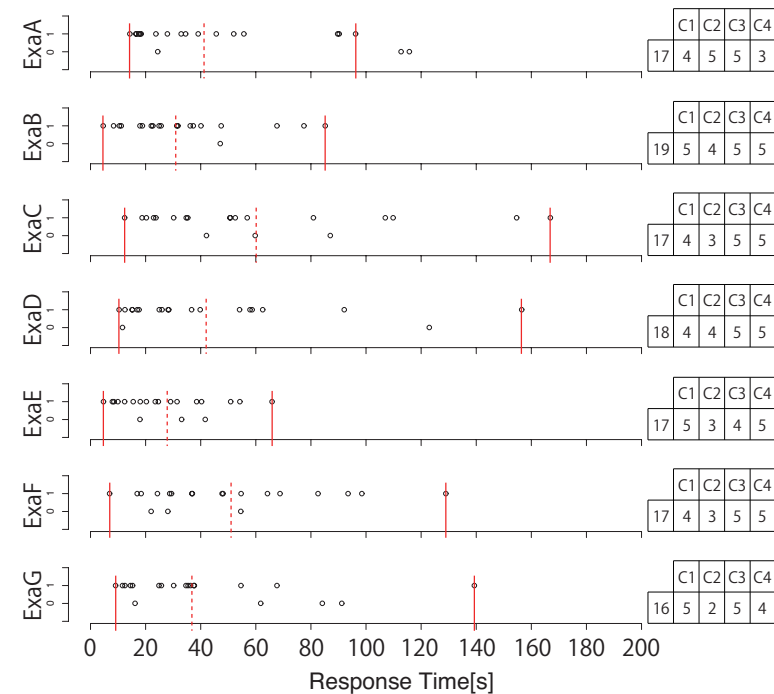


図 3 各受験者 ExaA ~ ExaG の解答時間及び正誤並びにテスト項目毎の正解数 (正答を 1, 誤答を 0 として plot。縦線は最小最大解答時間、破線は平均解答時間を表す。また、総得点 (20 問中)、各テスト項目別正解数 (5 問中) を分布図の横に示す。

異点について検討する。

例えば、受験者 D, E, G の特異点であるが、受験後、受験者 G に聴取したところ、「n 進法の問題で計算間違いに気づき、再計算をしたため非常に時間がかかってしまった。」と口述された。このように、(Ca1), (Ca4) の問題では再計算の可能性があり、この再計算によって受験者 D, E, G の特異点が発生したと思われる。この特異点の検出に関しては、Moodle のモジュール内でラジオボタンのクリック動作を検出することにより、より正確な検討を行えると考え、今後の課題とする。

6.2 累積分布図を用いた検討

図 4 は各受験者の解答時間及び正誤情報の累積分布図である。縦に引かれた直線は、最

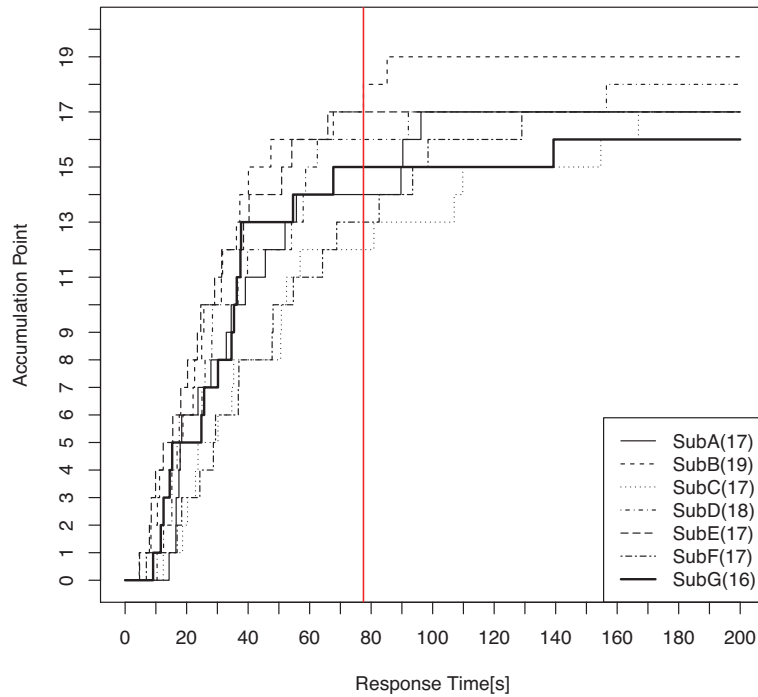


図 4 各受験者の解答時間及び正誤の累積分布図 (凡例の ExaA ~ ExaG は各受験者を表し、() 内の数字は各受験者の合計点となる。縦に引かれた直線は、最も点数の良かった受験者が点数の 90%(18 点)を取った解答時間を示している)

も点数の良かった受験者が点数の 90%(18 点)を取った解答時間を示している。この結果を見ると、各受験者において獲得点数に到達するための立ち上がりの時間にある程度の差が見られる。つまり、この分布で示された立ち上がりが受験者の能力に関わっているのではないかと推測される。

さらに、この累積分布図では、最終的な成績順と受験者 B が点数の 90%(18 点)を取った解答時間における成績順は

B, D, A, C, E, F, G

B, E, D, G, A, F, C

となり、成績順序が大きく変わる。

表 2 各受験者の非線形回帰分析による近似曲線の各値 (ExaA ~ ExaG は各受験者を表しており、 α は傾き、 β は合計点、 γ は 1 点目を獲得した時間となる)

	ExaA	ExaB	ExaC	ExaD	ExaE	ExaF	ExaG
α	0.046	0.034	0.023	0.038	0.044	0.020	0.041
β	17	19	17	18	17	17	16
γ	14.3	4.6	12.4	10.4	4.7	6.9	9.1

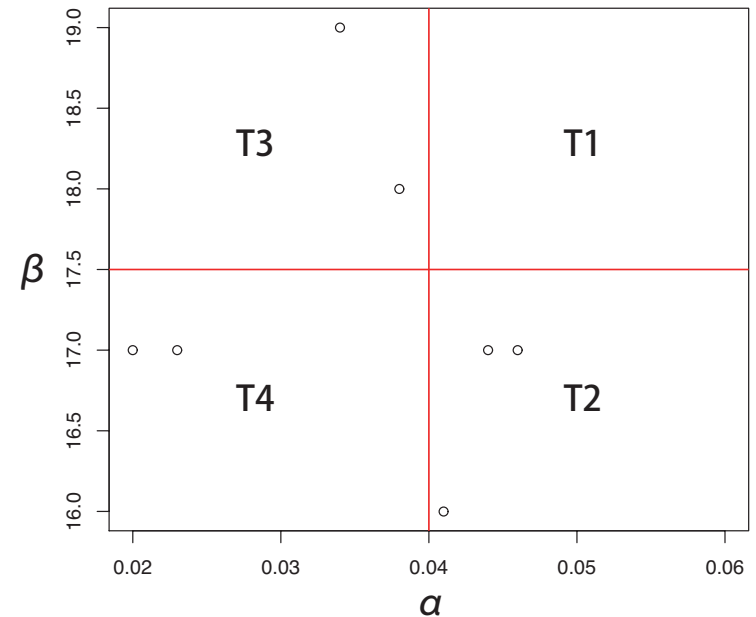


図 5 解答速度に対応する α を横軸に、総得点に対応する β を縦軸にとった 2 次元配列

従って、もし解答時間に制限があった場合、受験者 D は制限時間内に解答することが出来ず、結果的に点数が低くなることが予想される。この累積分布図を基に非線形重回帰分析を行い、近似式の算出を行った。この近似式は式 (3) のようになると考えられる。

$$\theta = \beta(1 - \exp(-\alpha(t - \gamma))) \quad (3)$$

(α は累積点数の上昇の大きさであり解答速度に対応する、 θ は累積点数、 β は総得点、 γ はテスト項目のうち正答した最短の解答時間)

表 2 は、各受験者の推定パラメータである。解答速度に対応する α を横軸に、総得点に対応する β を縦軸にとり 2 次元配列にしたものが図 5 である。

図 5 より受験者の特性として、次の 4 タイプが考えられる。() 内の T は Type の略である。

(T1) 解答時間が速く、正確に解くことが出来る (α 大, β 大)

(T2) 解答時間は速いが、正確さに欠ける (α 大, β 小)

(T3) 解答時間は遅いが、正確に解くことが出来る (α 小, β 大)

(T4) 解答時間は遅く、正確さにもかける (α 小, β 小)

上記 4 タイプがあると仮定すると、解答時間の平均値より A,E,G は (T2)、B,D は (T3)、他は (T4) に分類できる。

7. おわりに

本論文では、SPI 総合能力検査の類似問題で用いられている分野に対して、テスト項目毎の解答時間を計測し、解答時間と受験者の能力との関係について検討した。まず、Kong らの Response Time Effort に基づく分析を行い、彼らの 3SEC 法による閾値が本実験でも適応できることを確かめた。また 3SEC で利用された閾値 3 秒は実験で求めた解答時間バイアス 2.7 s と対応していると考えられる。さらに、累積回答数の非線形重回帰分析の近似により、総得点と解答速度の関係から受験者の能力を分析できる可能性の端緒がみられた。

参 考 文 献

- 1) 緒方, 宇佐川, 学習管理システムを活用した信号処理教育, 信学技報 IEICE Technical Report SP2010-24, 2010
- 2) ソンムアン, 植野, e テスティングにおける得点・時間予測システムの開発, 電子情報通信学会論文誌 D, Vol.J91-D, No.9, 2225-2235, 2008
- 3) 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, 106-109, 2005
- 4) Steven L. Wise, Christine E. DeMars, An Application of Item Response Time: The Effort-Moderated IRT Model, Journal of Educational Measurement Volume 43, Issue 1, pages 19-38, March, 2006
- 5) Wim J. van der Linden, David J. Scrams, Deborah L. Schnipke, Using Response-Time Constraints to Control for Differential Speededness in Computerized Adaptive Testing, Applied Psychological Measurement, Vol.23 No.3 195-210, 1999
- 6) SPI ノートの会, 直前でも OK! [パソコン版 SPI2] これが本当のテストセンターだ!, 洋

泉社, 2008

- 7) 高橋 正規, 項目反応理論入門—新しい絶対評価, イデア出版局, 2002
- 8) Cynthia G. Kromrey, Jeffrey D. Chason, Walter M. Yi, Qing, Evaluation of Parameter Estimation under Modified IRT Models and Small Samples, Proc. of the Annual Meeting of the Psychometric Society Parshall, 1997)
- 9) 佐々木 豊文, 絶妙な「速読」の技術, 明日香出版社, 2005
- 10) Xiaojing J. Kong, Steven L. Wise, Dennison S. Bholra, Setting the Response Time Threshold Parameter to Differentiate Solution Behavior From Rapid-Guessing Behavior, Educational and Psychological Measurement August 2007 vol. 67 no. 4 606-619