

大学における情報技術に関する能力認定試験の分析

—古典的テスト理論による分析と項目反応理論の適用可能性—

杉浦 学^{†1} 荘司 泰徳^{†1} 大岩 元^{†2}

慶應義塾大学 湘南藤沢キャンパスでは、主に新入生を対象に、情報分野の基礎的な能力を測定するための「情報技術認定試験」を実施している。本稿では1)古典的テスト理論による試験結果の分析、2)試験に対する項目反応理論の適用可能性についての検討を行った。1)の結果から、正答率が著しく低い問題であっても、その内容は概ね妥当で、試験に合格した受験者であっても理解が不十分な分野があることが分かった。また2)の検討の結果、受験者の能力推定に項目反応理論を用いることで、出題した問題の難易度に左右されない公平な合否判定が実施できる可能性が高いことが分かった。

An Analysis of Qualification Tests for IT Education at University : Through Classical Test Analysis and Applicability of Item Response Theory

MANABU SUGIURA,^{†1} YASUNORI SHOJI^{†1} and HAJIME OHIWA^{†2}

Keio University at Shonan Fujisawa Campus adopts “IT Qualification Tests” to measure basic ability of IT literacy. In this report, we describe 1) results of “IT Qualification Tests” through classical test analysis and 2) considerations for applicability of Item Response Theory (IRT) for this test. The result of classical test analysis shows that even if the questions whose correct answer rates are low, the questions themselves are appropriate. It has been revealed that many of the qualified have still not enough knowledge about some fields of the test. We have also done a preliminary investigation of IRT, and the result suggests that IRT could be effective to the reliability for the qualification criterion.

1. はじめに

慶應義塾大学 湘南藤沢キャンパスでは、2007年度に情報技術関連の科目を含む大規模なカリキュラム改定が行われた。この改定では、高校の新教科「情報」を受講した新入生の入学にともなって、入学一学期目の必修科目であった情報技術の入門科目が廃止された。新入生は入学直後に「情報技術認定試験」という能力認定試験を受験する。この試験に合格した者は入学一学期目から必修科目であるプログラミングの入門科目が履修可能になった。試験に不合格だった者は、新設された「情報基礎」という授業を履修し、試験に合格できるような基礎教育を受ける。「情報技術認定試験」はカリキュラム改定前の2004年度から実施されており、合格が卒業要件に含まれているため、2007年度以前の入学者（3年生以上の学生）も試験を受験する

必要がある。

「情報技術認定試験」はタイピングの速度と正確さ（1分間に英文150字以上、97%以上の正打鍵率が合格基準）を試験する実技試験（タイピング科目）と、情報分野の基礎的な知識を測定するオンライン試験（基礎知識科目）の2つの試験から構成されている。

本稿では「基礎知識試験」の改善を目的とし、試験の分析を行った結果について述べる。以下「認定試験」と略した場合は「基礎知識科目」のことを指す。まず、2章で認定試験の内容について述べる。次の3章で古典的テスト理論を用いた試験結果の分析について述べる。4章で認定試験に項目反応理論を適用することの可能性について検討し、5章で今後の課題について述べる。最後の6章でまとめを述べる。

2. 認定試験の内容

認定試験の目的は「基礎的な情報関連の知識^{*1}を有し、プログラミングの入門科目を受講するに相応しいかどうかを判定する」ことにある。現行の試験で出題

^{†1} 慶應義塾大学 政策・メディア研究科

Graduate School of Media and Governance, Keio University

^{†2} 慶應義塾大学 環境情報学部

Faculty of Environmental Information, Keio University

*1 基本的に、教科「情報」で扱われている内容。

表 1 認定試験の出題範囲

分野 1: 情報科学基礎 (計 130 問)	分野 2: コンピュータシステム (計 62 問)
1-1 : 情報量の単位と接頭辞 (13 問)	2-1 : ファイルの概念 (4 問)
1-2 : ビット数と表現できる情報量 (8 問)	2-2 : ディレクトリの役割とツリー構造 (4 問)
1-3 : ビットマップ画像の持つ情報量の計算 (6 問)	2-3 : 相対パスと絶対パスの概念 (4 問)
1-4 : 10 進数 2 進数の基数変換 (20 問)	2-4 : 相対パスの指定方法 (9 問)
1-5 : 2 進数同士の加算 (18 問)	2-5 : 絶対パスの指定方法 (8 問)
1-6 : アナログとデジタルの概念と特徴 (6 問)	2-6 : CPU の機能とクロック周波数の理解 (5 問)
1-7 : アナログ/デジタル変換のプロセス (11 問)	2-7 : 出力・入力装置の具体例 (6 問)
1-8 : 文字コード (6 問)	2-8 : 主記憶装置と補助記憶装置の性質と役割 (8 問)
1-9 : ビットマップ画像とベクトル画像の特徴 (6 問)	2-9 : 基本ソフトウェアと応用ソフトウェア (7 問)
1-10 : 可逆圧縮と非可逆圧縮 (7 問)	2-10 : OS の役割 (7 問)
1-11 : 推論と論理思考 (6 問)	
1-12 : 真理関数 (7 問)	
1-13 : 検索式 (6 問)	
分野 3: コンピュータネットワーク (計 49 問)	分野 4: 情報セキュリティと法 (計 35 問)
3-1 : サーバ・クライアントモデル (7 問)	4-1 : 不正アクセスとその対策 (8 問)
3-2 : P2P の用途と違法性の誤認識 (5 問)	4-2 : スパムメールとフィッシング詐欺 (6 問)
3-3 : パケットの概念 (4 問)	4-3 : サイト証明書と認証機関 (4 問)
3-4 : 回線交換方式とパケット交換方式 (6 問)	4-4 : ウイルスとスパイウェアの概念と対策 (6 問)
3-5 : IP アドレス (3 問)	4-5 : 共通鍵暗号と公開鍵暗号 (6 問)
3-6 : DNS とドメイン名 (4 問)	4-6 : 著作権の概念と事例 (5 問)
3-7 : 経路制御 (3 問)	
3-8 : 電子メールの仕組みとプロトコル (4 問)	
3-9 : メーリングリストの概念 (3 問)	
3-10 : 電子メールを利用するときの注意点 (4 問)	
3-11 : HTTP と URL (6 問)	

している試験範囲を表 1 に示す。

出題範囲は「情報科学基礎」、「コンピュータシステム」、「コンピュータネットワーク」、「情報セキュリティと法」の 4 つの分野から構成した。各分野は更に出題内容によってカテゴリに分割して、合計 40 のカテゴリを作成した。問題は選択肢が 3 個から 9 個の多義選択式とした。

これらの 40 カテゴリに属する問題を約 270 問作成し、それを問題プールとした。表 1 の括弧内に各分野、各カテゴリに属する問題数を示した。作問にあたっては、問題のレビュー支援システムを作成し¹⁾、問題の質の向上をはかった。

試験の際は、1 つのカテゴリごとに 1 問をランダムに選び、合計 40 問の問題を出題する。制限時間は 20 分で、合格基準は 7 割 (28 問) 以上の正答率としている。受験回数に制限はなく、合格するまで何度でも受験することができる。

3. 古典的テスト理論による受験結果の分析

本章では認定試験の受験結果を古典的テスト理論を用いて分析²⁾し、問題の妥当性と受験者集団の特性について検討を行う。

3.1 分析対象

2007 年 3 月 27 日から 7 月 31 日にかけて 164 回の

試験が実施され、合計 1123 名が受験した。この結果 (複数回受験している受験者は初回受験のデータのみを抽出) を用いて、古典的テスト理論を用いた分析を実施した。受験者の学年は 1 年生から 4 年生までが混在している^{*1}。

3.2 分析結果

1123 名中、合格者は 204 名 (合格率 18.2%) であった。得点 (1 問 1 点で 40 点満点) の分布状況を図 1 に示す。全体の分布 (図 1 中 $f(x)$) は正規分布からはずれているが、受験者を平均得点が 18.5 点の集団 (図 1 中 $f_1(x)$ 85%) と 32.5 点の集団 (図 1 中 $f_2(x)$ 15%) に分割すると、それぞれがほぼ正規分布している。即ち、受験者には内容をよく理解している集団と、理解できていない集団に分れているものと推測される。

全問題についての正答率について調べてみると、最も正答率の高かった問題は「4-1: 不正アクセスとその対策」内の問題で約 96.8%、最も正答率の低かった問題は「3-11: HTTP と URL」内の問題で約 7.2%だった。

カテゴリに含まれる複数の問題の平均正答率 (カテゴリの正答率) が特に低いものは「1-3: ビットマップ画像の持つ情報量の計算」、「2-3: 相対パスと絶対

*1 大学側から、受験者の学年や学部に関するデータの提供が無かったため、内訳は不明である。

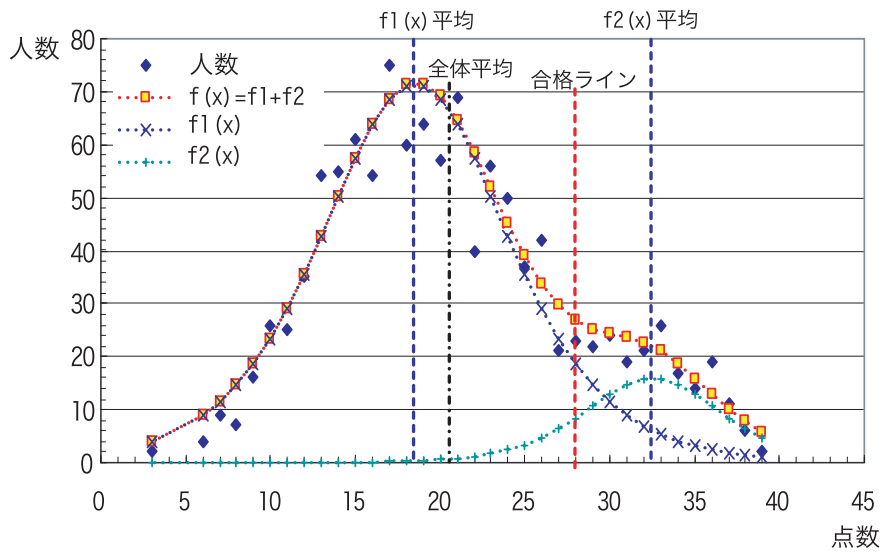


図 1 得点の分布状況

0000000 から 1111111 までの 7 桁の 2 進数は何ビットの情報量を持つか。

- A) 7 ビット (8.84%)
- B) 14 ビット (6.12%)
- C) 128 ビット(50.34%)
- D) 256 ビット(28.57%)
- E) 512 ビット(6.12%)

図 2 正答率の低い問題例 その 1

10 × 10 ドットのビットマップ画像は何ビットの情報量を持つか。1 ドットは 8 色を表現できるものとする。

- A) 300 ビット(13.04%)
- B) 400 ビット(21.20%)
- C) 550 ビット(5.43%)
- D) 650 ビット(11.96%)
- E) 800 ビット(48.37%)

図 3 正答率の低い問題例 その 2

パスの概念」,「2-4: 相対パスの指定方法」であった。これらのカテゴリに属する問題は、情報量やパスの概念が正しく理解できていないと正解できない、つまり用語を暗記しただけでは正解できない問題が属するカテゴリである。

また、正答率が著しく低い(正答率が 1/選択肢数を下回っている)ものは 13 問であった。そうした問題のうち、代表的なものを図 2, 図 3, 図 4 に示す。正解の選択肢については「」を併記し、括弧内に各選択肢の選択率を示した。この 13 問が属する 11 のカテゴリについて、表 1 中に 印を付記した。

問題の正答率が低くなる原因として、問題の内容や記述に不備がある場合、受験者が問題を解く能力がない場合の 2 つが考えられる。13 問の問題を検討した結果、その内容は概ね妥当であり、問題の記述にも不備はないと考えられるため、正答率が極端に低いのは受験者の理解不足であると推測できる。

これらの 13 問に関しては、合格者(正答率が 7 割以上)の受験者集団のみを抽出した場合であっても、

デジタル化と音質の関係について述べた次の文章の空欄(a)を埋めるのに、最も適した語句を選べ。

電話より CD の音質が良い要因の一つは、電話と比較して、CD のデジタル化の際の (a) が短いからである。

- A) 標本化間隔(18.48%)
- B) ビット間隔(23.91%)
- C) 周波数間隔(39.13%)

図 4 正答率の低い問題例 その 3

正答率は 50%以下となった。つまり試験に合格できた受験者であっても、これら 13 問の問題に正答するための知識は不足していることが分かった。現行の正答率による合格判定では、一部の知識(例えば情報量の基礎概念)が欠けている場合であっても合格できる可能性があるという意味で、十分な能力保証はできていない。

合格者の知識に欠けた分野がないことを保証するた

めには、現行の合格判定方法を改善する必要がある。また、合格判定の方法の改善に加え、出題した問題の難易度に関わらず、受験者の能力を正確に測定することも必要になる。これには項目反応理論を用いることが望ましい。そこで次章では、認定試験に対して項目反応理論を適用することの可能性を検討する。

4. 項目反応理論の適用可能性の検討

項目反応理論を用いた TOEFL や TOEIC では、出題された問題に左右されずに受験者の能力を判定することが可能である。現行の認定試験では、出題した問題の難易度を考慮せず、正答率（7割以上で合格）による合格判定を行っている。このため、出題された問題が難しければ合格が難しくなり、出題された問題が簡単であれば合格は容易になる。本章では、認定試験に項目反応理論を適用することにより、この問題が解決できるかどうかを検討する。

4.1 項目反応理論

項目反応（応答）理論^{3),4)}は、1950年代以降に米国より発達した現代テスト理論である。

項目反応理論をテストに用いることによって、以下のような効果を得ることができる。

- (1) 複数のテスト間の結果の比較が容易になる
- (2) テストの測定精度をきめ細かく確認できる
- (3) テストの平均点をテスト前に制御できる
- (4) 受験者ごとに最適な問題をその場で出題できる

項目反応理論に関する研究は言語テストや心理学の分野などで盛んであるが、情報技術に関するテストに対する適用事例も報告⁵⁾されている。

次に、本稿で利用した項目反応理論における2パラメータロジスティックモデルについて述べる。受験者の能力を θ とし、能力 θ を持つ被験者の j 番目の問題の正答率を p_j とした場合、2パラメータロジスティックモデルは以下の式で表現される。

$$p_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

θ は受験者の能力値（尺度値）である。受験者の能力について定量的に推定した値であり、テストにおいて測定すべき受験者の特性（学力など）を示す。

b は困難度である。問題の難易度を表し、受験者の能力値に応じた正答率の指針となる。困難度が能力値と同値の時、すなわち $\theta = b$ の時に正答率は0.5である。 b は $-\infty < b < \infty$ の範囲で定義される。能力値分布が平均値0、分散1に標準化されている時、困

難度はおよそ-3.0から3.0の範囲で推定され、負の値で小さくなるほどその問題は易しく、正の値で大きくなるほど問題は難しいと解釈する。

a は識別力（項目識別力、または項目弁別力）を表す。識別力は能力に対する正答確率の変化率を表現する。識別力が大きい問題は、能力が低い受験者と能力が高い受験者の正答確率の差が大きくなる。

D は正規分布を近似するために導入された尺度要素であり、一般的に1.7を用いる。

4.2 項目（問題）の特性値の推定

現行の認定試験は、各受験者がそれぞれのカテゴリごとに異なる40問の問題を受験しているため、実際の試験結果から問題の特性値（困難度 b や識別力 a ）を推定することが難しい。そこで「情報基礎」の授業内に実施した、認定試験の模擬試験の結果を用いて、問題プールの一部の問題に関する特性値の推定を行った。

模擬試験の受験者数は155名で、問題は認定試験で出題されているもののうち、情報基礎の授業の前半で学習した内容に関するものを選択し、制限時間や資料の持ち込みなどは本番の認定試験と同じとした。155名の受験者集団は近似的な等質グループとみなし、42問の問題の特性値を推定した。推定には4.1節で述べた2パラメータロジスティックモデルを採用した分析ソフト「EasyEstimation」⁶⁾を利用した。一般に、2パラメータロジスティックモデルを用いた特性値の推定には、200~300名分の受験データが必要と言われているため、今回の推定に用いたデータ量は十分ではない。

各問題の推定結果に関しては困難度、識別力とも、正常値の範囲内であった。

4.3 適用実験

4.2節で述べた推定結果を使って、模擬試験の受験者の能力推定が公平に行えるかの実験を行った。特性値の推定を行った42問は全部で26のカテゴリに属している。うち12のカテゴリに関しては、同一のカテゴリに属する複数の問題の特性値が推定できた。そこで、12のそれぞれのカテゴリから、困難度の一番高いものと低いものを1問ずつ選択し、12問の問題セットを2つ作成した。以下、困難度の高い問題を集めたセットを「難セット」、困難度の低い問題を集めたセットを「易セット」と略記する。

現行の試験では、各カテゴリから1問がランダムに選択されて出題されているため、「難セット」は難しい問題のみが選ばれてしまった受験者にとってのワーストケースであり、「易セット」は簡単な問題のみが選ば

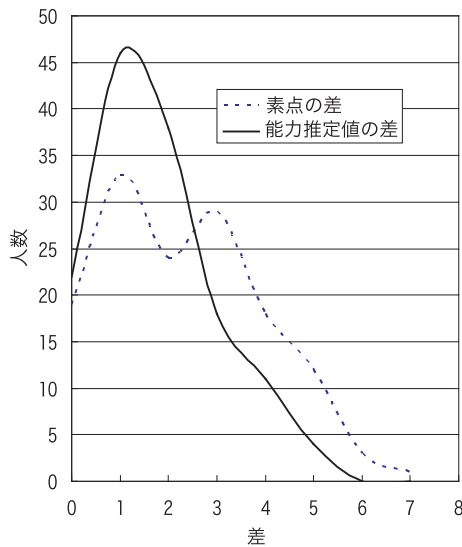


図5 「素点の差」と「能力推定値の差」の分布

れたベストケース^{*1}である。現行の認定試験では「難セット」が出題された場合であっても「易セット」が出題された場合であっても合格基準は固定であり、7割以上の正答率が要求されている。

特性値の推定に利用したデータから「難セット」のみに関する結果と「易セット」のみに関する結果を抽出した。図5中の点線に同一の受験者における「難セット」と「易セット」の合計点(12点満点)の差の分布を示す。1点から3点程度の差がついた受験者が多いことが分かる。

「難セット」と「易セット」はそれぞれ12問であるから、正答率が受験者の能力であるとする古典的テスト理論に基づけば、0~12までの13段階で能力を判定していることになる。つまり「難セット」を受験した場合は低く能力が推定され「易セット」を受験した場合は高く能力が推定されている。ここで、古典的テスト理論による合計点と項目反応理論を使った能力推定値の差を比較するため、次のような方法で実験を行った。

まず、「難セット」と「易セット」のそれぞれの受験結果から、模擬試験の受験者の能力を「Easy Est-Theta」⁶⁾を用いて推定した。「Easy Est-Theta」は最尤推定法による推定を行うため、全問正解や全問不正解の受験者の能力値を判定することが不可能であった。そのため、どちらかのセットで0点や12点であった受験者16名を除いた139名についてを実験の対象とした。

*1 実際の試験では40問が1セットで出題される。

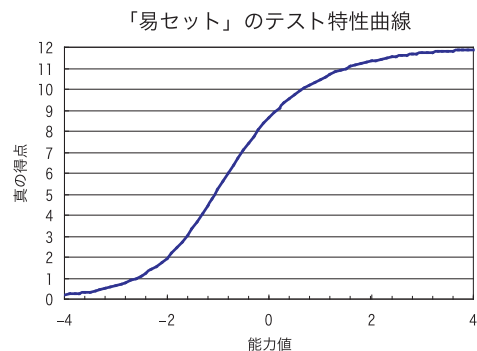
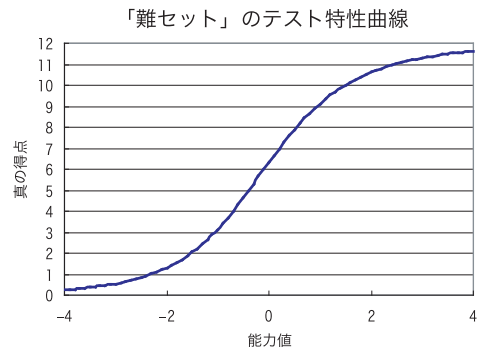


図6 各セットのテスト特性曲線

次に「難セット」と「易セット」で測定できる能力の推定値の範囲を確認した。図6に「難セット」と「易セット」のテスト特性曲線を示す。縦軸は1問を1点とした場合の真の得点、横軸は受験者の能力値を示す。例えば、能力値が0の受験者が「難セット」を受験した場合は真の得点は6になり、「易セット」を受験した場合は9になることを示している。

図6から「難セット」で真の得点が1である受験者の能力推定値は-2.3であり、11の受験者の能力推定値は2.4であることが分かる。つまり「難セット」を受験し、1点から11点だった受験者は概ね-2.3~2.4の範囲で能力値が推定されることになる。同じように「易セット」の場合は概ね-2.5~1.5の間で能力値が推定される。そこでそれぞれのセットで推定された能力値を11段階に分類すれば、1点から11点という古典的テスト理論における合計点と比較できることになる^{*2}。図5中の実線が「難セット」と「易セット」のそれぞれから推定された11段階の能力値の差の分布である。

*2 真の得点はある能力を持つ受験者集団の平均点であるため、正解する問題によってはこの範囲外の能力値が推定されることもある。実験では7%の受験者がこの範囲外であった。能力値が想定範囲より低かった者は1、高かった者は11として分類した。

同一の受験者の能力を推定しているのに、推定の差は0になることが期待されるが、分布のピークは1である。原因として、問題数が12問と少ないこと、問題の項目特性の推定精度が十分でないことが考えられる。しかし、項目反応理論を用いた推定値の差の分布は、素点合計の差の分布に比較して0に近く、問題の難易度の影響を受けにくい能力推定が行えることが確認できた。

今回は意図的に「難セット」と「易セット」を作った実験を行ったため、「難セット」と「易セット」の両方を受験している同一の受験者集団が存在する。このような場合は得点を線形変換することにより、得点の調整を行うことができる。しかし、実際の認定試験では、難易度の異なる問題セットを同一の受験者集団に実施することは難しいため、得点の調整は不可能である。今回推定を行った42問に関しては、特性値の推定精度は十分とはいえない。しかし、これらの42問に関しては、どのような組み合わせで出題をしたとしても同一の尺度で受験者の能力が推定が可能であり、問題の難易度を考慮しない現行の正答率による合格判定に比較し、判定精度の向上が期待できる。

5. 今後の課題

本章では、項目反応理論を認定試験に適用し、改善を行う上での課題について整理する。

5.1 測定対象の能力による問題のカテゴリ化

3.2節でも述べたように、試験全体の正答率によって合格判定を実施すれば、ある特定の分野に関する理解が不十分な受験者であっても試験に合格してしまう可能性がある。これは項目反応理論による能力推定値で合格判定を行ったとしても解決できない。試験全体で一つの能力を測定していると仮定すれば、その能力値が基準以上であれば合格できてしまうからである。

また、項目反応理論を適用する場合の前提となる「一次元性」についても検討が必要となる。一次元性とは、試験で測定しようとする能力が単一でなければならないということを目指す。今回の分析では詳細な統計分析を実施していないが、現行の認定試験に一次元性が確保されているかは疑わしい。

こうした問題を解決するために、現行の40カテゴリのうち、同一の能力を判定していると思われるカテゴリをまとめ、測定対象の能力を基準にした「能力カテゴリ」を作成する必要がある。この「能力カテゴリ」ごとに項目反応理論を使った能力推定を実施し、合格者には全ての「能力カテゴリ」に関して一定以上の能力を保証することが必要である。

5.2 問題の見直しと指導設計

「能力カテゴリ」の作成にあたっては、測定すべき能力の設計と問題内容の再検討が必要になる。

指導設計 (ID:Instructional Design) 技術⁷⁾では、まず教育の目的を明らかにし、次に具体的に評価が可能な教育目標を設定する。これには、学習者がどのような前提能力を持つのか、またそれを出発点にしてどのような能力に到達することを想定するかを明らかにしなければならない。次に、目標が達成されたかどうかを評価する問題を作成してから、学習システムの設計に入る。こうした手順に従い「測定すべき能力とは何か」と「その能力を測定する問題はどうか」を同時に検討すべきである。

5.3 特性値の推定に必要なデータ収集

「能力カテゴリ」の設計、問題の見直しに加えて、問題の特性値の推定を効率よく進める必要がある。これには実際の試験のデータを推定に利用するのが有効である。複数回の試験で同じ問題セットの試験を実施できれば、推定に必要なデータの蓄積が簡単に行える。しかし同一の受験者が複数回の試験を受験することを考慮すると、同じ問題セットばかりを出題すれば、正答を暗記する受験者が発生する可能性が高い。そこで、同一の「能力カテゴリ」から出題する問題セットを数種類用意しておき、それらのセットから1つをランダムに選んで出題するといった方式が有効である。試験全体が複数の「能力カテゴリ」から構成されていれば、複数回の受験にも対応できる。特性値の推定は単一の「能力カテゴリ」内で実施すればよいから、用意した問題セットごとにデータを集めて推定が行える。

今後、試験システムを改良し、特性値の推定が実施できるような受験データを蓄積する予定である。

6. まとめ

本稿では、慶應義塾大学 湘南藤沢キャンパスで実施されている認定試験を対象とした分析について述べた。古典的テスト理論による受験データの分析により、正答率が著しく低い問題であっても、その内容は概ね妥当で、試験に合格した受験者であっても理解が不十分な分野があることが分かった。また、認定試験に項目反応理論を適用するための予備実験を試みた。その結果、出題された問題の難易度に左右されない公平な合格判定が実施できる可能性が高いことが分かった。認定試験に項目反応理論を導入するための課題が明らかになったので、今後は導入のための準備作業を順次進めていく予定である。

参 考 文 献

- 1) 杉浦学, 大岩元. テスト問題を改善するための協
調作業を支援する環境構築. 日本教育工学会論文
誌, Vol.30, No.3, pp. 171-181, 2006.
 - 2) 荘司泰徳. 情報教育における現代テスト理論の
適用 - 項目反応理論を用いた学習到達度評価作成
の試み - . 慶應義塾大学 政策・メディア研究科 修
士論文, 2007.
 - 3) 大友賢二. 項目応答理論入門. 大修館書店, 1996.
 - 4) 豊田秀樹. 項目反応理論 [入門編]. 朝倉書店,
2002.
 - 5) 山川修, 田中武之, 菊沢正裕. IRT を利用した新
入生のコンピュータ利用能力の測定. 第 3 回 日
本 WebCT 研究会, pp. 49-52, 2005.
 - 6) 熊谷龍一. 項目パラメタ推定プログラム Easy
Estimation. <http://irtanalysis.main.jp/>,
2005.
 - 7) Walter Dick 他著, 角行之監訳. はじめてのイン
ストラクショナルデザイン. ピアソンエデュケー
ション, 2004.
-