

# 管理会計知識の測定：項目反応理論の 会計研究への応用

尻無濱 芳 崇  
(人文社会科学部准教授)

山形大学紀要（社会科学）第48巻第1号別刷  
平成29年（2017）7月

## 研究ノート

# 管理会計知識の測定：項目反応理論の会計研究への応用

尻無濱 芳崇

(人文社会科学部准教授)

## 1. 問題意識

管理会計について人々が持っている知識の多寡を、正確に測定することは可能なのだろうか。本稿では、管理会計知識を測定する尺度を新たに開発する必要性を主張する。尺度を開発するうえで項目反応理論 (item response theory: IRT) と呼ばれる手法を利用することで、既存の尺度が持つ限界を克服できることを示す。

管理会計についての知識の多寡は会計情報利用者の判断や意思決定に影響すると考えられ、重要な変数として管理会計研究で扱われてきた (Bonner 2008, Luft and Shields 2010)。しかしこれまでの管理会計研究では、研究対象者の管理会計知識の多寡は、年齢や経歴といった代理変数に置き換えられることが多かった (e.g. Vera-Munoz 1998, Vera-Munoz et al. 2001)。管理会計の知識の多寡を問題正答率で測定する数少ない研究 (Dearman and Shields 2001, Stone et al. 2000) では、質問項目が一部しか公開されておらず、二次的な利用が難しい状況にある。さらに、これまでの研究では尺度開発のプロセスにおいて測定精度に対して十分な注意が払われてきたとは言い難い。管理会計知識の測定は経営管理者の判断・意思決定に関する研究を進めるうえで土台となるものであり、精度の高い測定尺度を開発する必要性は高い。精度の高い測定尺度が開発できれば、年齢や経歴といった必ずしも知識の多寡を正確に反映しない代理変数の利用を避け、より精緻な研究を行うことができるようになる。

本研究では、管理会計知識に関連する仮説の検証に利用することが可能な測定尺度の案を示すことで、この分野の研究の進展に貢献する。それに加えて、これまで管理会計研究で注目されてこなかった項目反応理論を紹介し、古典的テスト理論 (classical test theory: CRT) のみに基づいた現状の尺度開発を振り返るきっかけを提供する<sup>1</sup>。そして、測定精度の高い尺度の開発に向けたガイドラインを示す。

本研究では、公認会計試験の短答式試験を参考に知識測定尺度を試験的に作成した。経営学

<sup>1</sup> 本研究では、管理会計研究におけるすべての尺度開発に項目反応理論を用いるべきだと主張するつもりはない。Hambleton and Jones (1993) の指摘するように、古典的テスト理論にもサンプルサイズが小さくても適用できる、項目反応理論より単純な数理モデルであるなどといった利点もある。筆者の主張は、知識を測定するようなタイプの尺度では、項目反応理論を利用することによる恩恵が大きいというものである。

部1年生と企業内の管理会計担当者という異なる集団についてこの尺度を用いて管理会計知識を測定するという仮定のもと、項目反応理論に基づいた尺度開発がどのように行われるかを具体的に示した。

本稿の構成は以下の通りである。次節で、管理会計知識の測定を試みた先行研究をレビューし、その成果と限界、新たな尺度開発の必要性を指摘する。第3節で、項目反応理論の概要と利点を紹介し、管理会計知識を測定する尺度の開発に有用な手法であることを主張する。第4節で尺度開発の具体的なプロセスを説明する。最後に、本稿の貢献と限界、今後の課題を示す。

## 2. 管理会計知識の測定を行う研究のレビュー

管理会計研究において、管理会計についての知識の多寡が変数の1つとして利用されることがある。管理会計に関する知識は、経営者をはじめとする企業人の判断や意思決定に影響を与える重要な変数である（Bonner 2008, Luft and Shields 2010）。管理会計知識は教育歴や実務経験の長さといった代理変数で操作化されることが多い（e.g. Vera-Munoz 1998, Vera-Munoz et al. 2001）。特に、上層部理論<sup>2</sup>に基づく研究でその傾向は顕著である（Hiebl 2014）。

これらの代理変数での測定にはどのような問題があるのだろうか。一般に知識と教育歴や実務経験の長さには相関があるが、専門的な教育を受けたり実務経験が長いからと言って必ずしも知識が豊富だとは限らない（Bonner and Lewis 1990, Bonner 2008, Libby and Luft 1993）。正確な知識の測定を行うためには、知識の有無を問うような質問項目から構成された尺度を用いる必要がある。

管理会計知識を問う質問項目を設定し、具体的な測定を試みている研究は、わずかではあるが存在する。Stone et al.（2000）の研究、Dearman and Shields（2001）の研究がそれである<sup>3</sup>。

Stone et al.（2000）の研究では、米国の管理会計担当者を対象にした質問票に、管理会計の技術的知識（technical managerial accounting knowledge）を測定する尺度が含まれている。この尺度は10項目からなり、それぞれの項目は米国公認管理会計士<sup>4</sup>試験（Certified Management Accountant's exam）の問題から選ばれたものである。調査を実施した当時は米国公認管理会計士試験は4つの領域から構成されていたため、すべての領域をカバーするように各領域から少なくとも2問は出題されている。彼らの分析結果では、管理会計担当者の職階が上がるほど管理会計の技術的知識のスコアが低くなっており、経験と知識量が必ずしも正の相関にないこと

<sup>2</sup> 上層部理論（upper echelon theory）とは、CEOやCFO、トップマネジメントチームなどの組織の上層部の特性によって組織のアウトトップが変わるという理論である（Hambrick and Mason 1984; Hambrick 2007）。

<sup>3</sup> 同じ知識を測定する尺度の中でも、公認会計士の持つ財務会計・監査についての知識を測定する尺度の開発は特に進んでいる。ただし、項目反応理論は利用されていない。詳しくは、Bonner（2008）を参照。

<sup>4</sup> 米国公認管理会計士とは、管理会計や財務分析に関する専門家資格で、管理会計の専門家を育成する機関であるIMA（Institute of Management Accountants）が資格の発行を行っている。

が示されている。ただし、Stone et al. (2000) は尺度の一部しか公開しておらず、彼らの尺度を他の研究に用いることはできない。

Dearman and Shields (2001) は管理会計一般の知識を問う尺度ではなく、伝統的原価計算 (volume-based costing) と活動基準原価計算 (activity based costing: ABC) についての知識を測定する項目を開発している。伝統的原価計算に関する知識の測定尺度には5つの多肢選択型問題が含まれ、それらは1990年から94年にかけての米国公認管理会計士試験で出題されたものの中から選ばれている。この5つの質問に対する正答率で、回答者がどれだけの知識を持っているかを測定している。ABCに関する知識も同様の尺度で測定されている。しかし、この2つの尺度の質問項目も公開されておらず、こちらを利用することも不可能である。

以上のように、2つの研究で開発された尺度はどちらも公開されておらず、彼らの尺度を利用して管理会計知識を測定することはできない。また、次節で詳述するが彼らの尺度開発の過程では、質問項目の難易度は適当なものかどうか、各項目が知識のある回答者とそうでない回答者をうまく区別することができているかなどは議論されていない。知識を測定する尺度を開発するにあたって精度の高い尺度にするためには、これらの点にも注意する必要がある。次節では、項目反応理論を活用することで、これらの点に配慮した尺度開発が可能になることを説明する。

### 3. 項目反応理論の概要<sup>5</sup>

項目反応理論とは、古典的テスト理論の問題点を克服するために1950年代以降に開発された手法である。本節では、古典的テスト理論の概要を示した後で、項目反応理論の概要とその利点を示す。

古典的テスト理論では、テスト得点は回答者の真の得点と誤差からなると考える (Hambleton and Jones 1993)。

$$y_i = t_i + e_i$$

ここで、 $t_i$  が回答者  $i$  の真の得点、 $y_i$  がテスト得点、そして  $e_i$  が誤差である。ただし、観察されるのはテスト得点だけであり、真の得点と誤差は観察されない。古典的テスト理論では真の得点と誤差は無相関であり、母集団における誤差の平均値はゼロであり、テストを繰り返した時の誤差間の相関はゼロだという前提に立つ。古典的テスト理論では回答者の知識を測定する

<sup>5</sup> 本節の記述は、Hambleton and Hones (1993), Knoll and Houts (2012), 加藤ほか (2014) に依拠している。数理的な背景は、加藤ほか (2014) の第4章から第6章を参照。

場合に、典型的には尺度を構成する項目の正答率や素点（正解した問題の配点合計）に基づいて評価を行う。例えば管理会計知識を問う尺度の場合、10個の質問があれば、そのうち8問に回答できれば正答率80%、各問の配点が1点であれば素点は8点になる。管理会計研究においては、古典的テスト理論に基づいて様々な回答者の特性が測定されている（Schäffer 2007, Van der Stede et al. 2007）

項目反応理論では、これとは全く異なったアプローチをとる。まず、項目特性曲線を用いて、個々の項目の特徴を表す。ここでは項目特性曲線を表すモデルの中でも、2パラメタ・ロジスティックモデルを用いて項目反応理論の特徴を説明する<sup>6</sup>。

図1が2パラメタ・ロジスティックモデルに基づく項目特性曲線の1例である。3つのS字曲線があるが、それぞれが尺度を構成する項目と対応している。3つの曲線は、次の式で表現される項目特性曲線の $a_j$ ,  $b_j$ に具体的な数値を代入し、図示したものである。

$$P_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}, \quad -\infty < \theta < \infty$$

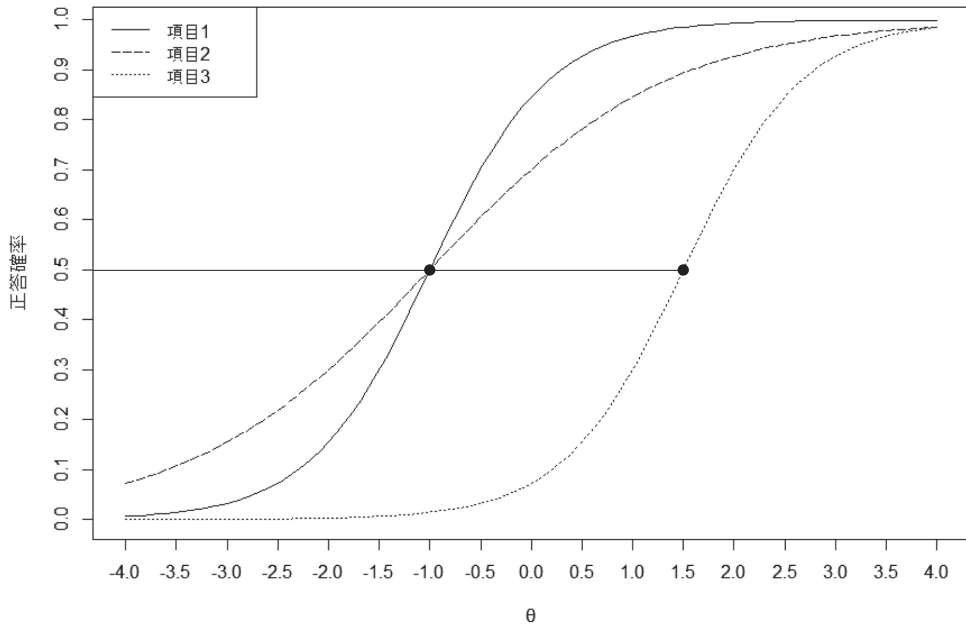
$P_j(\theta)$ は能力 $\theta$ の回答者が項目 $j$ に正答する確率を示す。ここで、 $D$ は尺度因子<sup>7</sup>で $D=1.7$ という値が設定されることが一般的である<sup>8</sup>。 $a_j$ は項目識別力、 $b_j$ は項目困難度という。2パラメタ・ロジスティックモデルでは、 $a_j$ ,  $b_j$ の2つのパラメタが推定される。項目困難度とは、項目の難しさを表すパラメタである。 $b_j$ が大きいほど難しい項目になる。 $b_j$ の値は、正答確率が50%になる回答者の能力として定まる。図1で言うと、項目1の曲線で正答確率50%となるのは $\theta = -1.0$ のときである。一方、項目3では正答確率50%になるのは $\theta = 1.5$ の時である。つまり、項目1は能力の低い回答者でも正答できる確率が高いが、項目3は能力が高い回答者でなくては正答確率が高くない。

<sup>6</sup> 後述するように、項目反応理論の2パラメタ・ロジスティックモデルにおける2つのパラメタ、項目識別力と項目困難度は、古典的テスト理論の項目識別力・正答率と対応するが、その内容は大きく異なる。この違いを示すために、本稿では2パラメタ・ロジスティックモデルに注目した。他のモデルについては、加藤ほか（2014）の第4章を参照。

<sup>7</sup> 尺度因子について、詳しくは加藤ほか（2014）の第4章を参照。

<sup>8</sup>  $D=1.7$ とするのは、項目特性曲線を表現するために、当初は正規分布の累積分布関数が用いられていた（Lord 1952）。のちにロジスティックモデルが使用されるようになるが、これを高い精度で正規累積モデルに近似できるのが、 $D=1.7$ という値であり、そのため慣習的に $D=1.7$ という値が用いられている（加藤ほか 2014）。

図1 2パラメタ・ロジスティックモデルにおける項目特性曲線の例

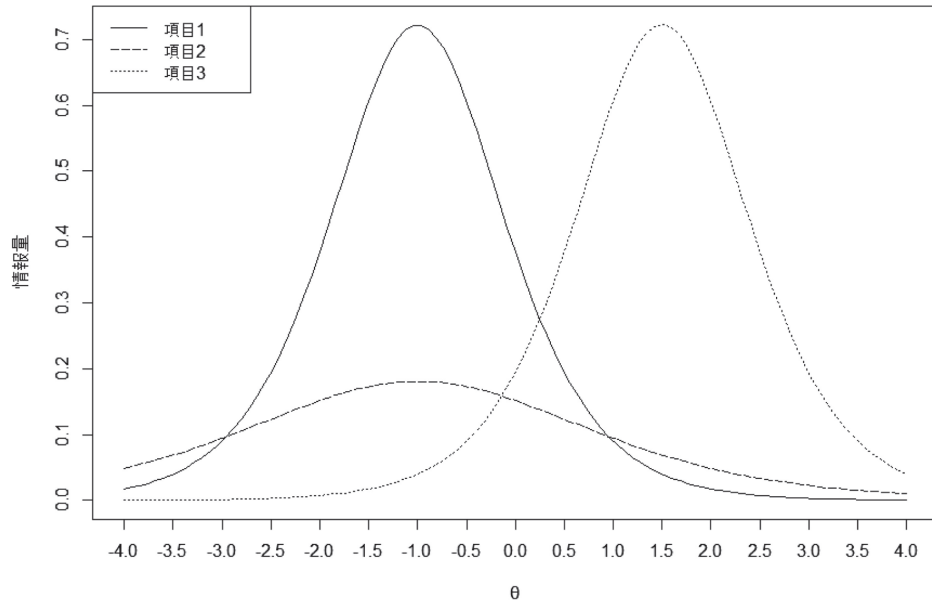


出所：加藤ほか（2014）をもとに筆者が一部修正

項目識別力 $a_j$ は、S字曲線の傾きの大きさに対応している。図1の項目1と項目2を比較すると、どちらも項目困難度は同じだが、S字曲線の傾きが異なる。項目1では $a_1=1.0$ であり急な傾きであるのに対して、項目2では $a_2=0.5$ であり、傾きが比較的緩やかである。これは何を意味するかというと、 $\theta = b_j$ （この場合は $\theta = -0.1$ ）の付近での能力の違いによってどれだけ正答確率が増加するかを示すということである。 $a_j$ の値が大きいほど識別力が高い項目となる。

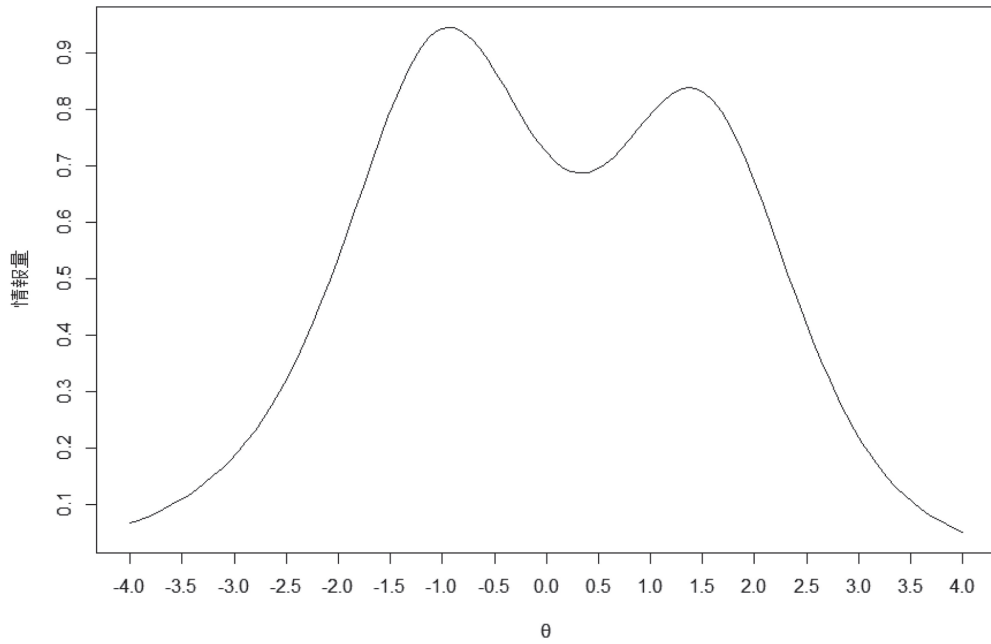
項目反応理論は、このような項目特性曲線の存在を前提として、項目パラメタ（ $a_j$ ,  $b_j$ ）と能力パラメタ（ $\theta$ ）を推定する方法である。項目反応理論は、尺度開発を行う際に威力を発揮する（Knoll and Houts 2012, 加藤ほか2014, 野口・渡辺編著 2015）。まず、項目困難度と項目識別力に基づいて尺度に含める項目の取捨選択を行うことができる。あまりにも困難な項目や容易な項目、また識別力が低すぎる項目は、尺度から除外することができるだろう。それに加えて項目の難易度と識別力が分かれば、そこから項目ごとに受験者の能力に応じた推定精度を求めることができる。この推定精度を示す項目ごとの曲線を項目情報関数といい、テスト全体の推定精度を示す曲線をテスト情報関数という。項目情報関数の和がテスト情報関数となる。図2は項目情報関数、図3はテスト情報関数をそれぞれ図示したものである。縦軸が情報量、横軸が $\theta$ を表す。この2つの図は、情報量が多くなっている $\theta$ 付近の推定精度が高くなることを示す。

図2 項目情報関数の図示



出所：加藤ほか（2014）を参考に筆者が一部修正

図3 テスト情報関数の図示



出所：加藤ほか（2014）を参考に筆者が一部修正



項目反応理論を活用することで、回答集団に応じた推定精度の高い尺度を作成することも可能になる。例えば、管理会計知識が十分でないような学部1年次の学生や、伝統的に会計情報による管理を忌避してきた業界、たとえば病院界の経営者（荒井 2009）に対しては、 $\theta$ が低い領域で精度が高くなるように質問項目を選ぶことで、精度の高い知識量の推定ができるだろう。一方で、ベテランの管理会計担当者のような管理会計のエキスパートに対しては、 $\theta$ が高い領域で測定の精度が高い尺度を用いることが有効になる。さらに、2つの尺度を構成する質問項目に違いがあっても、得点を比較可能にする等化<sup>9</sup>という手続きを行うことで、テスト結果の比較が可能になる。等化とは、2つの尺度がある場合にどちらか一方の尺度を基準として、他方の尺度の原点や目盛りの幅を調整することを言う。

項目困難度や項目識別力、測定の精度は、後述するように古典的テスト理論にも対応する指標がある（Hambleton and Jones 1993, 加藤ほか 2014）。しかし、古典的テスト理論におけるこれらの指標の値は、回答者に依存して決まるという問題を抱えている（標本依存性）。全く同じ質問項目からなる管理会計知識測定尺度を能力の異なる回答者群、例えば管理会計専門家と会計を専攻していない学部の1年生に回答してもらおうとしよう。このとき、前者は正答する人が多く各項目の難易度は簡単だと評価されるが、後者にとっては難しい問題が多く、どの問題も困難度が高いと評価されるかもしれない。このように、古典的テスト理論の立場では、項目の特性が標本に依存して変わってしまう。それに加えて、回答者の能力の推定値も、質問項目によって異なる値になる（項目依存性）。同一の回答者であっても、質問項目が変われば能力が高く評価されたり低く評価されたりするのである。このような弱点を古典的テスト理論は抱えている。

一方、項目反応理論を用いると、回答者から得られたデータがモデルに適合していれば、項目困難度や識別力は回答者集団に依存せずに推定できるし、回答者の能力も質問項目に依存せずに推定できる（不変性）ことが知られている（加藤ほか 2014）<sup>10</sup>。項目反応理論はこのような優れた特徴をもっているため、テスト開発において盛んに利用されている。国際的に有名な英語能力検定試験であるTOEFLや、OECD生徒の学習達成度調査（PISA）などでも項目反応理論は活用されており、わが国でもITパスポート試験、医療系大学共用試験などの活用例がある。経営学では、特に組織心理学研究において項目反応理論に基づいた尺度開発が進んでいる（野口・渡辺編著 2015）。様々な分野・研究で項目反応理論が利用されているのは、テスト問題や尺度の内容が変わっても、受験集団・研究対象が変わっても、結果の比較が可能になるからである。会計学研究では残念ながら適用例は見られないが、会計関連の資格試験では、米

<sup>9</sup> 等化について、詳しくは加藤ほか（2014）の11章を参照。この場合、等化のために2つの尺度に共通する項目を入れておく必要がある。

<sup>10</sup> 詳細は加藤ほか（2014）の第1章・第4章を参照。データがモデルに適合している場合にのみ不変性が成り立つことに注意。



国公認会計士試験の多肢選択型問題において項目反応理論を活用していることが公表されている（AICPA 2011）。このように、項目反応理論は会計資格試験での実践例もあり、会計研究での会計知識測定尺度の開発にも応用可能な手法だと考えられる。次節では、項目反応理論を用いた管理会計知識測定尺度の開発手順を示す。

#### 4. 項目反応理論に基づいた管理会計知識測定尺度開発のプロセス

本節では、加藤ほか（2014）に基づいて項目反応理論に基づいた管理会計知識測定尺度開発の手順を説明することで、尺度開発のガイドラインを示すとともにこの理論の利点を具体的に示す。

項目反応理論に基づいた尺度開発では、まず①項目作成、②予備実査と基礎分析、③項目パラメタの推定および等化を通じて、項目困難度や項目識別度が分かっている項目を増やしていくことが必要である。項目パラメタの推定値がわかっている項目の集まりを、項目プールという。項目プールが充実すると、④テストの編集、すなわち調査対象ごとに項目を選んで尺度を作ることができるようになる。その後、調査を実施し⑤テストスコアの算出とレポートを行うことになる。以下で各ステップについて説明する。なお、本研究で用いたシミュレーションデータの作成・分析には、汎用統計解析ソフトウェアであるR（R Core Team 2015）と、項目反応理論の教育・学習に有用な関数を多く含むirtosパッケージ（Partchev 2014）、項目反応理論に基づく等化を行うためのplinkパッケージ（Weeks 2010）を用いた。

##### 4. 1. 項目プールの作成：項目作成、予備実査、項目パラメタの推定および等化

まず①項目作成であるが、管理会計知識を測定する尺度の試案を作成することが本稿の目的の一つである。そこで、公認会計士試験の短答式試験をもとに質問項目を作成した。より具体的には、平成28年度第Ⅰ回・第Ⅱ回の短答式試験を利用した。管理会計知識測定尺度の開発を行った先行研究でも会計専門職試験の多肢選択式問題を質問項目として利用しているので（Dearman and Shields 2001, Stone et al. 2000）、妥当な方法だといえるだろう。

公認会計士試験の短答式試験は、計算問題と理論問題が含まれる。開発した尺度を質問票調査で利用する予定であれば、計算問題を質問項目にすると無回答が増える恐れがある。短答式試験の計算問題を解くのに時間がかかり、難易度も高いため、回答者が会計に詳しくなければ解くのをあきらめてしまう確率が高いからである。会計に詳しい回答者でも、時間がかかる調査には協力しないおそれがある。したがって、計算問題は質問項目に含めないこととした。また、短答式試験の理論問題では4つの文章の中から正しい文章の組み合わせを選ぶという問題が出題される。例えば、ア～ウに対応する4つの文章があり、「①ア・イ、②ア・ウ、③ア・

エ、④イ・ウ、⑤イ・エ、⑥ウ・エ」のうち正しい文章の組み合わせを選ぶような問題である。このような問題も解くのに時間がかかるため、回答率を低下させると考えられる。したがって、複数の選択肢の中から正しい選択肢の組み合わせを選ぶタイプの問題については、個々の文章の正誤を問う質問項目へと変更した。それに加えて、理論問題の中でも原価計算（基準）に関する正誤問題は、会計専門家以外の回答率を大幅に低下させられると思われるため、除外した。なお問題の順番は、ランダムに並び替えた。

項目作成ができれば、次は②予備実査と基礎分析である。これは、調査対象者に質問項目に回答してもらい、得られたデータを用いて項目の性能を古典的テスト理論の観点から分析するステップである。このような分析（項目分析という）は、尺度に含めることが不適当な項目をあらかじめ特定して除外する、いわゆるスクリーニングのために必要になる。極端に難しい（もしくは簡単な）項目や古典的テスト理論における識別力が非常に低い項目は、項目反応理論を適用しても項目パラメタの推定に悪影響を与える可能性がある。そのため、あらかじめ不適当な項目を除外しておくことが必要になる。

表 1 仮想データの正答率・項目識別力（学部 3・4 年生 n=500）

項目番号	質 問 内 容	正答率	項目識別力
1	管理可能利益と残余利益との違いは、事業部の業績評価を行う際に、管理可能利益は資本コストを考慮せずに算定するのに対して、残余利益は資本コストを考慮して算定することにある。	0.006	0.259
2	ある会社において、製造事業部が製造した製品を販売事業部が外部に販売する国内取引を行う場合、事業部間の振替価格の決定方法を市価基準から原価プラス基準に変更しても、その他の取引条件が変化しなければ、当該取引に関して計上される会社全体としての利益には影響を及ぼさない。	0.974	0.286
3	標準原価計算は、主として製品の製造段階における科学的・統計的な調査に基づく原価管理技法であるが、標準原価計算の限界と原価企画との関係に鑑みて、現在では標準原価計算が製品開発のスピードに対応させた原価管理技法に変化してきている。	0.850	0.552
4	事業部制では、会社内の事業を製品、顧客、地域等からグループ化し、会社が持つべき職能のうち事業遂行に必要なものを備えて、事業部が編成される。	0.992	0.171
5	本社費・共通費の各事業部への配賦は、実務上、本社や共通部門が提供するサービスの利用の程度に基づいて決定されるのが合理的であり、各事業部の利益の規模は配賦の基準とはならない。	0.084	0.479
6	最近のわが国における原価企画の概念は拡大化の傾向にあり、開発設計段階における単なる原価の引下げ活動ではなく、製品の機能向上や戦略的な製品コンセプト作りにまで及んできている。	0.132	0.506
7	原価維持とは、設定された目標原価を標準原価管理や原価改善によって管理する活動である。	0.382	0.667
8	原価改善とは、生産している製品の原価水準を維持し、さらに期待原価水準にまで計画的に引き下げる活動をいう。	0.016	0.208
9	資金管理の目的は支払不能状態の回避にあるため、資金を調達する上で、通常は自己資本による資金調達が優先される。	0.506	0.677
10	創業当初の企業のキャッシュ・フロー計算書においては、営業活動によるキャッシュ・フローと財務活動によるキャッシュ・フローがマイナスとなり、投資活動によるキャッシュ・フローによって賄われることが多い。	0.544	0.627
11	キャッシュ・サイクル（＝棚卸資産回転期間＋売上債権回転期間－買入債務回転期間）が長くなると、手元資金が不足し借入れが必要となるため、キャッシュ・サイクルは短い方が望ましい。さらに、手元資金に余裕がある場合でも、資金回収の観点からはキャッシュ・サイクルを短くすることが望ましい。	0.016	0.266
12	帳簿上は十分な利益が確保できているにもかかわらず、現金残高が不足して債務を支払うことができない状態を一般に「勘定合って銭足らず」というが、このような状態を回避するために、損益管理を徹底する必要がある。	0.796	0.553
13	売掛金回転期間を短縮し、買掛金回転期間を延ばすことによって回転差資金を増加させ、資金繰りを好転させることができる。	0.648	0.593
14	資金繰りを検討する上で、買入債務の支払について現金による場合と小切手の振出しによる場合とでは、資金繰りに与える影響が異なる場合がある。	0.096	0.389
15	手形割引を含む短期借入金を増減は、運転資金の必要額の増減に連動するのが通常であるが、運転資金の増減がない場合であっても短期借入金が増加しているケースがある。その原因として、急速な売上拡大に伴う売上債権の増加などの影響がある。	0.018	0.182

項目番号	質 問 内 容	正答率	項目識別力
16	資金管理において、資本コストを上回るリターンが得られる設備投資の案件がある場合、保有する現金及び要求払預金の在高を最低限の水準まで減少させて投資を行うことがある。	0.776	0.505
17	財務会計情報の役割は利害関係者の利害調整や利益配分に重点が置かれるが、管理会計情報の役割は経営者の経営責任を明らかにすることに重点が置かれている。	0.124	0.481
18	会計情報の質について、財務会計においては正確性や適法性などの概念が適用され、管理会計においては目的適合性や適時性などの概念が適用される。	0.210	0.486
19	会計情報の測定尺度として、財務会計では財務情報すなわち貨幣価値情報が用いられ、管理会計では財務情報と、定量情報や定性情報の全ての非財務情報が用いられる。	0.156	0.486
20	財務会計情報は外部利害関係者に提供されるのに対して、管理会計情報は内部利害関係者に提供される。	0.886	0.339
21	価値連鎖（value chain）とは、製品やサービスに価値を与えるための一連の活動であり、製品やサービスへの価値は生産プロセスにおいて付加されると考えられるので、販売後に行われる製品保証等のサポート活動は価値連鎖に含まれない。	0.202	0.454
22	管理会計は、意思決定、資源配分、業績評価を行うために必要となる情報を組織内の構成員に提供する会計システムである。	0.062	0.289
23	責任センターとは組織の構成単位であり、その管理者は責任センターで行われる特定の活動について責任を持っている。	0.132	0.270
24	管理会計は外部の利害関係者との調整機能を有しているため、管理会計システムの設計を行う場合、企業内の管理者・従業員のみならず、外部の利害関係者による利用をも考慮に入れながら設計を行うべきである。	0.866	0.325
25	変動予算は、予算期間内における実際のアウトプットに基づき計算される収益・費用からなり、予算編成プロセスの一環として予算期間開始時に策定される。	0.112	0.337
26	管理会計システムを構築する際、短期的な成果と長期的な成果を評価できるような評価尺度の利用を考慮すべきである。	0.150	0.240
27	管理会計システムを選択する際、そのシステムから入手できる情報のベネフィットと情報を入手するためのコストとのバランスを考慮すべきである。	0.592	0.355
28	コスト・マネジメント・システムとは、管理者の意思決定がコストにどのような影響を与えるかを識別するための一連のツール・技法である。このため、活動基準原価計算システムはコスト・マネジメント・システムとみなされるが、全部原価計算システムはコスト・マネジメント・システムとはみなされない。	0.418	0.279

出所：筆者作成

古典的テスト理論における項目困難度は正答率、すなわち回答者全体のうち正答したものの割合で測定される。この正答率が極端に高い・低い項目は除外される。また、無回答の多さも、その項目が非常に難しい・複雑であることを示唆している可能性があるため、無回答の割合が高い項目も除外するべきである。一方、古典的テスト理論における項目識別力は、尺度の合計得点と項目得点の相関係数（この場合は項目が1か0をとる変数なので点双列相関係数）である。この値が高いほど、その項目の識別力が高いとみなされる。項目識別力が極端に低い、またはマイナスの項目は、除外することになる。

本稿では、ある4年制大学の経営学部で会計を専攻する3・4年生を対象に予備調査を行ったという仮定で、仮想データを作成した。短答式試験の問題ごとの正答率は、株式会社TACによって調査・公表されている（TAC公認会計士講座 2016）。TAC調査の正答率を参考に質問項目ごとに正答率を設定し、n=500で回答データを作成した。なお、すべての回答者がすべての質問に回答している。仮想データの各項目の正答率、項目識別力を見ると（表1）、項目番号1, 2, 4, 8, 11, 15が極端な値（正答率が5%以下もしくは95%以上、あるいは項目識別力が0.2以下）をとっている。したがって、項目反応理論を適用する前に、これらの項目を除外した。

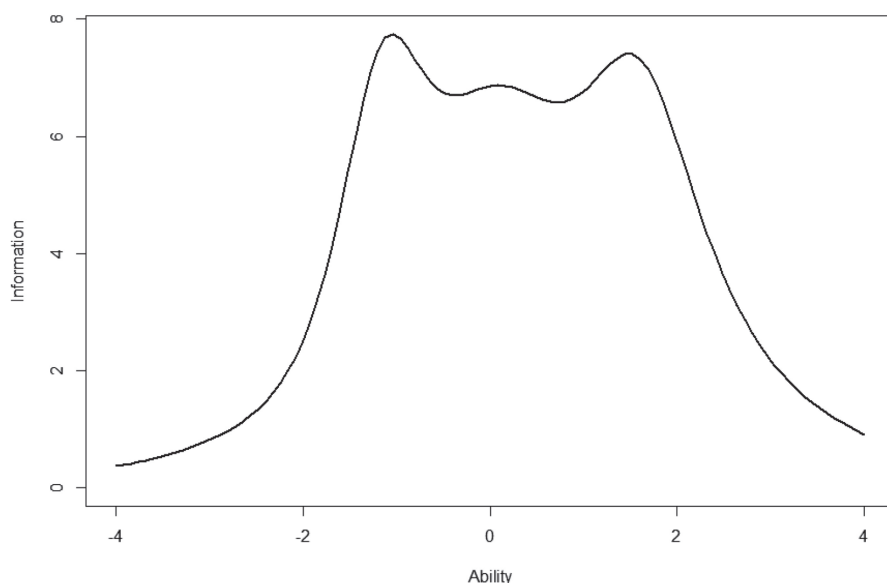
表2 仮想データの項目識別力・項目困難度（学部3・4年生 n=500）

項目番号	質 問 内 容	項 目 識別力 (a <sub>i</sub> )	項 目 困難度 (b <sub>j</sub> )
3	標準原価計算は、主として製品の製造段階における科学的・統計的な調査に基づく原価管理技法であるが、標準原価計算の限界と原価企画との関係に鑑みて、現在では標準原価計算が製品開発のスピードに対応させた原価管理技法に変化してきている。	3.75	-1.16
5	本社費・共通費の各事業部への配賦は、実務上、本社や共通部門が提供するサービスの利用の程度に基づいて決定されるのが合理的であり、各事業部の利益の規模は配賦の基準とはならない。	2.94	1.61
6	最近のわが国における原価企画の概念は拡大化の傾向にあり、開発設計段階における単なる原価の引下げ活動ではなく、製品の機能向上や戦略的な製品コンセプト作りにまで及んできている。	2.20	1.43
7	原価維持とは、設定された目標原価を標準原価管理や原価改善によって管理する活動である。	2.42	0.38
9	資金管理の目的は支払不能状態の回避にあるため、資金を調達する上で、通常は自己資本による資金調達が優先される。	2.50	-0.02
10	創業当初の企業のキャッシュ・フロー計算書においては、営業活動によるキャッシュ・フローと財務活動によるキャッシュ・フローがマイナスとなり、投資活動によるキャッシュ・フローによって賄われることが多い。	2.00	-0.14
12	帳簿上は十分な利益が確保できているにもかかわらず、現金残高が不足して債務を支払うことができない状態を一般に「勘定合って銭足らず」というが、このような状態を回避するために、損益管理を徹底する必要がある。	2.21	-1.06
13	売掛金回転期間を短縮し、買掛金回転期間を延ばすことによって回転差資金を増加させ、資金繰りを好転させることができる。	1.76	-0.53
14	資金繰りを検討する上で、買入債務の支払について現金による場合と小切手の振出しによる場合とでは、資金繰りに与える影響が異なる場合がある。	1.42	2.08
16	資金管理において、資本コストを上回るリターンが得られる設備投資の案件がある場合、保有する現金及び要求払預金の在高位を最低限の水準まで減少させて投資を行うことがある。	1.59	-1.12
17	財務会計情報の役割は利害関係者の利害調整や利益配分に重点が置かれるが、管理会計情報の役割は経営者の経営責任を明らかにすることに重点が置かれている。	1.77	1.63
18	会計情報の質について、財務会計においては正確性や適法性などの概念が適用され、管理会計においては目的適合性や適時性などの概念が適用される。	1.39	1.28
19	会計情報の測定尺度として、財務会計では財務情報すなわち貨幣価値情報が用いられ、管理会計では財務情報と、定量情報や定性情報の全ての非財務情報が用いられる。	1.67	1.46
20	財務会計情報は外部利害関係者に提供されるのに対して、管理会計情報は内部利害関係者に提供される。	1.16	-2.17
21	価値連鎖（value chain）とは、製品やサービスに価値を与えるための一連の活動であり、製品やサービスへの価値は生産プロセスにおいて付加されると考えられるので、販売後に行われる製品保証等のサポート活動は価値連鎖に含まれない。	1.15	1.49
22	管理会計は、意思決定、資源配分、業績評価を行うために必要となる情報を組織内の構成員に提供する会計システムである。	1.10	2.91
23	責任センターとは組織の構成単位であり、その管理者は責任センターで行われる特定の活動について責任を持っている。	0.69	2.97
24	管理会計は外部の利害関係者との調整機能を有しているため、管理会計システムの設計を行う場合、企業内の管理者・従業員のみならず、外部の利害関係者による利用をも考慮に入れながら設計を行うべきである。	0.88	-2.43
25	変動予算は、予算期間内における実際のアウトプットに基づき計算される収益・費用からなり、予算編成プロセスの一環として予算期間開始時に策定される。	1.01	2.41
26	管理会計システムを構築する際、短期的な成果と長期的な成果を評価できるような評価尺度の利用を考慮すべきである。	0.48	3.75
27	管理会計システムを選択する際、そのシステムから入手できる情報のベネフィットと情報を入手するためのコストとのバランスを考慮すべきである。	0.58	-0.70
28	コスト・マネジメント・システムとは、管理者の意思決定がコストにどのような影響を与えるかを識別するための一連のツール・技法である。このため、活動基準原価計算システムはコスト・マネジメント・システムとみなされるが、全部原価計算システムはコスト・マネジメント・システムとはみなされない。	0.36	0.95

出所：筆者作成

③項目パラメタの推定および等化では、予備調査で得た回答データに対して項目反応理論を適用する。そうすることで、項目困難度や項目識別力を推定することができる<sup>11</sup>。先に発生させた3・4年次学生データに対して2パラメタ・ロジスティックモデルで項目困難度と項目識別力を推定した結果が表2である。項目識別力は0.36～3.75、項目困難度は-2.43～3.75の値をとっている。テスト情報関数（図4）を見ると、中程度の能力をよく識別できることが分かる。学生の能力推定値の平均は-0.02、標準偏差1.14であった。全体として、簡単な項目も難しい項目もそろっており、識別力も低くはないので、この項目プールから項目を選んで尺度を作成することに大きな問題はないだろう。

図4 仮想データのテスト情報関数（学部3・4年生 n=500）



出所：筆者作成

なお、項目パラメタの推定を行う際には、各項目が項目反応理論の前提（一次元性・局所独立性）を満たしているかをチェックし、データが項目反応理論のモデルに適合しているかもチェックする必要がある（適合度の評価）<sup>12</sup>。

<sup>11</sup> 項目パラメタおよび能力パラメタの推定方法の詳細は、加藤ほか（2014）の8章を参照。

<sup>12</sup> 詳しくは、加藤ほか（2014）の第7章、第9章、第13章を参照。



#### 4. 2. 項目プールの利用：テストの編集，テストスコアの算出とレポート

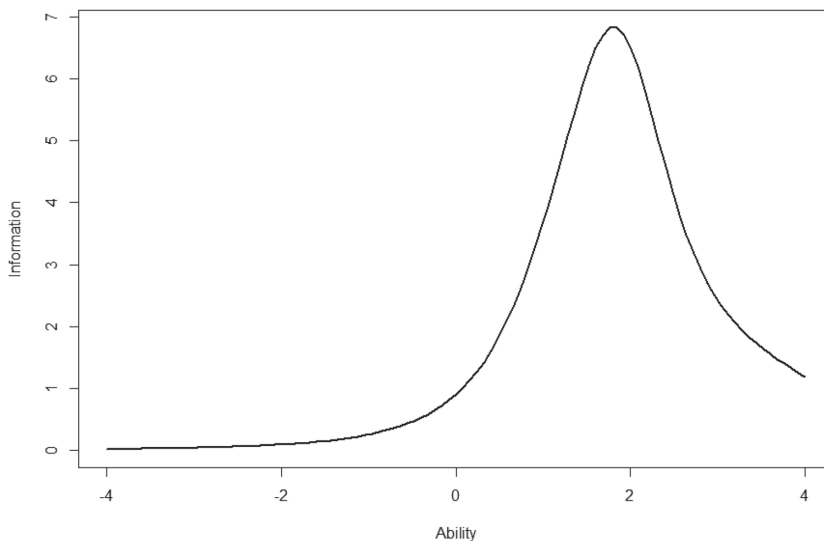
項目プールが作成できれば、調査対象ごとに④テストの編集を行うことになる。テスト情報関数を参照しつつ、調査対象の能力をよく識別できるように尺度を構成する。本稿では、企業の管理会計担当者と、経営学部1年生を対象に管理会計知識の多寡を推定すると仮定する。企業の管理会計担当者（ $n=200$ ）に対しては、項目プールから困難度の高い項目を選ぶことで、精度の高い能力値の推定ができるだろう。すべての質問項目を入れると質問票が長くなりすぎて回答率が低くなることが予想されるため、ここでは質問項目を10問に絞った尺度を作成した。管理会計担当者の回答データは、能力値が高くなるような設定で作成した。この仮想データに対して、項目反応理論を適用した。その後、項目パラメタと能力パラメタの比較が可能なように、学部3・4年生対象データを分析して得た結果を用いて等化の手続きを行った。得られた項目困難度・項目識別力の値は表3の通りであり、表2の結果と同じような結果であった。テスト情報関数を見ると、ねらい通り高い能力値の推定が正確な尺度になっていることがわかる（図5）。なお、管理会計担当者の能力推定値の平均は、2.68、標準偏差0.89であった。

表3 仮想データの項目識別力・項目困難度（管理会計担当者  $n=200$ ）

項目番号	質 問 内 容	項 目 識別力 (ai)	項 目 困難度 (bi)
5	本社費・共通費の各事業部への配賦は、実務上、本社や共通部門が提供するサービスの利用の程度に基づいて決定されるのが合理的であり、各事業部の利益の規模は配賦の基準とはならない。	3.07	1.83
6	最近のわが国における原価企画の概念は拡大化の傾向にあり、開発設計段階における単なる原価の引下げ活動ではなく、製品の機能向上や戦略的な製品コンセプト作りにまで及んできている。	2.40	1.28
14	資金繰りを検討する上で、買入債務の支払について現金による場合と小切手の振出しによる場合とでは、資金繰りに与える影響が異なる場合がある。	2.69	1.92
17	財務会計情報の役割は利害関係者の利害調整や利益配分に重点が置かれるが、管理会計情報の役割は経営者の経営責任を明らかにすることに重点が置かれている。	1.25	1.33
19	会計情報の測定尺度として、財務会計では財務情報すなわち貨幣価値情報が用いられ、管理会計では財務情報と、定量情報や定性情報の全ての非財務情報が用いられる。	0.61	1.38
21	価値連鎖（value chain）とは、製品やサービスに価値を与えるための一連の活動であり、製品やサービスへの価値は生産プロセスにおいて付加されると考えられるので、販売後に行われる製品保証等のサポート活動は価値連鎖に含まれない。	1.38	1.79
22	管理会計は、意思決定、資源配分、業績評価を行うために必要となる情報を組織内の構成員に提供する会計システムである。	1.10	3.40
23	責任センターとは組織の構成単位であり、その管理者は責任センターで行われる特定の活動について責任を持っている。	1.09	2.65
25	変動予算は、予算期間内における実際のアウトプットに基づき計算される収益・費用からなり、予算編成プロセスの一環として予算期間開始時に策定される。	1.36	3.05
26	管理会計システムを構築する際、短期的な成果と長期的な成果を評価できるような評価尺度の利用を考慮すべきである。	0.94	3.12

出所：筆者作成

図5 仮想データのテスト情報関数（管理会計担当者 n=200）



出所：筆者作成

表4 仮想データの項目識別力・項目困難度（学部1年生 n=200）

項目番号	質 問 内 容	項 目 識別力 (a <sub>j</sub> )	項 目 困難度 (b <sub>j</sub> )
3	標準原価計算は、主として製品の製造段階における科学的・統計的な調査に基づく原価管理技法であるが、標準原価計算の限界と原価企画との関係に鑑みて、現在では標準原価計算が製品開発のスピードに対応させた原価管理技法に変化してきている。	2.57	-1.32
7	原価維持とは、設定された目標原価を標準原価管理や原価改善によって管理する活動である。	2.38	0.17
9	資金管理の目的は支払不能状態の回避にあるため、資金を調達する上で、通常は自己資本による資金調達が優先される。	2.37	0.17
10	創業当初の企業のキャッシュ・フロー計算書においては、営業活動によるキャッシュ・フローと財務活動によるキャッシュ・フローがマイナスとなり、投資活動によるキャッシュ・フローによって賄われることが多い。	2.06	-0.12
12	帳簿上は十分な利益が確保できているにもかかわらず、現金残高が不足して債務を支払うことができない状態を一般に「勘定合って銭足らず」というが、このような状態を回避するために、損益管理を徹底する必要がある。	1.85	-0.87
13	売掛金回転期間を短縮し、買掛金回転期間を延ばすことによって回転差資金を増加させ、資金繰りを好転させることができる。	1.94	-0.26
16	資金管理において、資本コストを上回るリターンが得られる設備投資の案件がある場合、保有する現金及び要求払預金の在高位を最低限の水準まで減少させて投資を行うことがある。	1.96	-1.24
20	財務会計情報は外部利害関係者に提供されるのに対して、管理会計情報は内部利害関係者に提供される。	1.07	-2.05
24	管理会計は外部の利害関係者との調整機能を有しているため、管理会計システムの設計を行う場合、企業内の管理者・従業員のみならず、外部の利害関係者による利用をも考慮に入れながら設計を行うべきである。	1.00	-2.43
27	管理会計システムを選択する際、そのシステムから入手できる情報のベネフィットと情報を入手するためのコストとのバランスを考慮すべきである。	0.47	-1.00

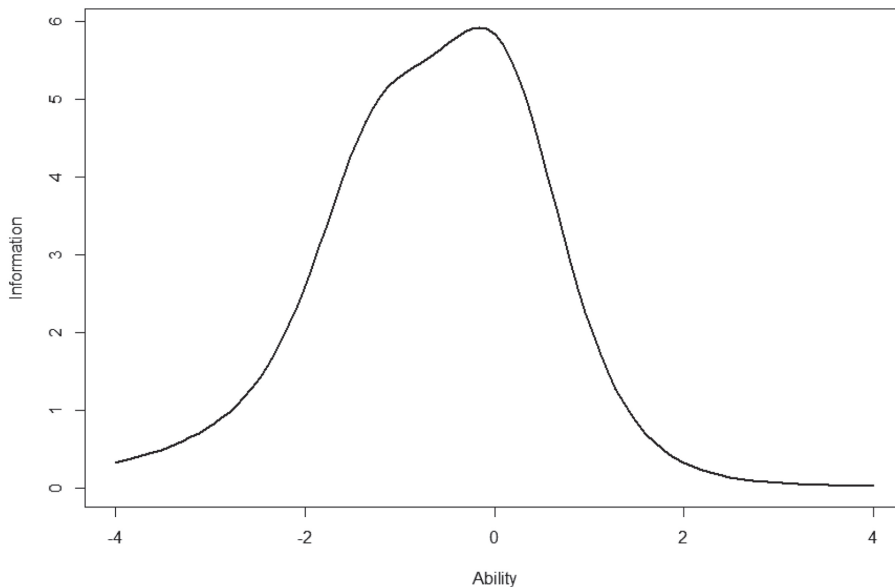
出所：筆者作成

一方、経営学部の1年生（n=200）を対象にした尺度では、項目困難度の低い10項目を選んで尺度を編成することになる。管理会計担当者についての分析と同じように、3・4年次学生



データと比較可能のように等化の手続きを行った。作成した1年次学生の仮想データをもとに項目困難度・項目識別力を推定した（表4）。テスト情報関数から、能力値の低い部分での推定精度が高くなっていることがわかる（図6）。なお、学部1年生の能力推定値の平均は-2.12、標準偏差は3.33であった。

図6 仮想データのテスト情報関数（学部1年生 n=200）



出所：筆者作成

テストの編集ができれば、⑤テストスコアの算出とレポートである。回答者から得たデータに基づいて、各回答者の能力値を推定し、それを解釈のしやすい適当なスコアに変換する。

項目反応理論に基づいた一連の尺度開発プロセスは、古典的テスト理論に基づく尺度開発と比較して明確なメリットを持っている。まず、本節で行ったように項目を組み替えて尺度を構成することで、回答者群の能力の高さに合わせた推定精度の高い尺度を作成することができる。さらに、2つの尺度の内容が違っていても、等化ができるように工夫をしておくことで、回答者の能力を比較することが可能になる。管理会計研究においては、管理会計知識の多寡が異なる集団を対象に研究が行われている（管理会計担当者、企業経営者、医療職、学生など）。項目反応理論を用いることで、各研究の研究対象に適した推定精度の高い尺度を利用することができ、さらに研究結果の比較も可能になる。これは、古典的テスト理論に基づいた尺度では得ら

れない利点である<sup>13</sup>。このように、項目反応理論が持つ特徴は、管理会計研究に対して大きなメリットをもたらすと考えられる。

## 5. 結論と今後の課題

本稿では、これまで管理会計知識の測定を試みてきた研究を振り返り、新たな尺度を開発する必要性を指摘した。そのうえで、知識を測定するタイプの尺度開発に有用な項目反応理論の利点を紹介し、それを管理会計研究に応用するためのガイドラインを、仮想データを用いて示した。本研究は、管理会計研究における知識測定方法を改善する方向性を示したという点で貢献がある。

ただし、本研究はあくまで仮想データに基づいて尺度開発のプロセスを示しただけであり、実際の大学生や管理会計担当者などから得たデータを基に尺度開発を行ったわけではない。管理会計知識を測定する尺度の開発は、管理会計における判断・意思決定研究を促進する基礎になると考えられ、その重要性は高い（Bonner 2008, Luft and Shields 2010）。今後は、様々な回答集団から得た実際のデータを項目反応理論に基づいて分析し、管理会計知識測定に利用可能な質問項目を増やしていくことが必要だろう。

## 謝 辞

本研究はJSPS科研費JP15K17155助成を受けたものである。

## 参考文献

- AICPA. 2011. How is the CPA exam scored?  
[http://www.aicpa.org/BecomeACPA/CPAExam/PsychometricsandScoring/ScoringInformation/DownloadableDocuments/How\\_the\\_CPA\\_Exam\\_is\\_Scored.pdf](http://www.aicpa.org/BecomeACPA/CPAExam/PsychometricsandScoring/ScoringInformation/DownloadableDocuments/How_the_CPA_Exam_is_Scored.pdf)
- Bonner, S. E. and B. L. Lewis. 1990. Determinants of auditor expertise. *Journal of Accounting Research*. 28, Supplement: 1-20.
- Bonner, S. E. 2008. *Judgement and decision making in accounting*. Pearson Education. 田口聡志監訳. 2012. 『心理会計学：会計における判断と意思決定』中央経済社.

---

<sup>13</sup> 古典的テスト理論では、研究間で研究対象に応じた尺度の変更を行うと研究結果の比較ができなくなる。その一方で研究間で同じ尺度を用いると、研究対象によっては能力の推定精度の低い結果しか得られないことになる。

- Dearman, D. T. and M. D. Shields. 2001. Cost knowledge and cost-based judgement performance. *Journal of Management Accounting Research*. 13(1): 1-18.
- Hambleton, R. K. and R. W. Jones. 1993. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*. 12(3): 38-47.
- Hambrick, D. 2007. Upper echelons theory: An update. *Academy of Management Review*. 32(2): 334-343.
- Hambrick, D., and P. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*. 9(2): 193-206.
- Hiebl, M. R. W. 2014. Upper echelons theory in management accounting and control research. *Journal of Management Control*. 24(3): 223-240
- Knoll, M. A. Z. and C. R. Houts. 2012. The financial knowledge scale: An application of item response theory to the assessment of financial literacy. *The Journal of Consumer affairs* 46(3): 381-410.
- Libby, R. and J. Luft. 1993 Determinants of judgement performance in accounting settings: ability, knowledge, motivation and environment. *Accounting, Organizations and Society*. 18(5): 425-450.
- Lord, F. 1952. A Theory of Test Scores (Psychometric Monograph No.7). Richmond, VA: Psychometric Corporation.
- URL <https://www.psychometricsociety.org/sites/default/files/pdf/MN07.pdf>
- Luft, J. and M. D. Shields 2010. *Psychology models of management accounting*. Now Publishers.
- Partchev, I. 2014. irtoys: Simple interface to the estimation and plotting of IRT models. R package version 0.1.7. <http://CRAN.R-project.org/package=irtoys>
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Schäffer, U. 2007. *Management Accounting and Control Scales Handbook*. Wiesbaden: Gabler.
- Stone, D. N., J. E. Hunton and B. Wier. 2000. Succeeding in managerial accounting. Part1: knowledge, ability and rank. *Accounting, Organizations and Society* 25: 697-715.
- Van der Stede, Wim A., S. M. Young and C. X. Chen. 2007. Doing management accounting survey research In: Chapman, C. S., A. G. Hopwood and M. D. Shields. (eds.) *Handbook of Management Accounting Research*. Elsevier Science, 445-478.
- Vera-Munoz, S. C. 1998. The effects of accounting knowledge and context on the omission of opportunity costs in resource allocation decisions. *The Accounting Review*. 73(1): 47-72.

- Vera-Munoz, S. C., W. R. Kinney Jr. and S. E. Bonner. 2001. The effects of domain experience and task presentation format on accountants' information relevance assurance. *The Accounting Review*. 76(3): 405-429.
- Weeks, P. J. 2010. plink: An R Package for Linking Mixed-Format Tests Using IRT-Based Methods. *Journal of Statistical Software*. 35(12): 1-33.  
URL <http://www.jstatsoft.org/v35/i12/>.
- 荒井耕. 2009. 『病院原価計算：医療制度適応への経営改革』 中央経済社.
- 加藤健太郎・山田剛史・川端一光. 2014. 『Rによる項目反応理論』 オーム社.
- TAC公認会計士講座. 2016. 『2017年度版公認会計士試験短答式試験過去問題集』 TAC出版.
- 野口裕之・渡辺直登編著. 2015. 『組織・心理テストの科学：項目反応理論による組織行動の探求』 白桃書房.