

**理解の深さの定量的評価を目指した認知診断モデルの応用と
そのプロセスの提案**
—心理学・教科教育の知見と Q 行列推定を活かした Q 行列作成—
**Application of Cognitive Diagnostic Models for
Quantitative Assessment of Depth of Understanding and
Its Proposed Process**
—Q-matrix Design Utilizing the Findings of Psychology
and Mathematical Education and Q-matrix Estimation—

佐宗 駿^{†*}, 岡 元紀[†], 植阪 友理[‡]

Shun Saso, Motonori Oka, Yuri Uesaka

[†] 東京大学大学院, [‡] 東京大学高大接続研究開発センター

The University of Tokyo, The University of Tokyo Center for Research and Development on Transition from
Secondary to Higher Education.

s-saso@g.ecc.u-tokyo.ac.jp

概要

学習者の理解の深さを定量的に診断できれば、学習内容の深い理解に向けた学習・指導改善へと繋がる。本研究では、認知診断モデルという数理モデルを活用して、理解の深さの定量的評価を試みた。認知診断モデルでは、Q 行列の作成が肝になる。そこで、学校教師と研究者が共同で行った Q 行列作成過程を示すことを目的の一つとした。さらに、認知診断モデルから得られた推定結果がどのように学校教師に受け止められるかをインタビューを通じて明らかにした。

キーワード — 深い理解, 認知診断モデル, Q 行列

1. はじめに

近年、国際的な教育動向として、教科内容の深い理解の涵養が目指されている [1, 2]。もし日々のテストの結果から学習者の教科内容の理解の深さを定量的に診断できたならば、その診断結果は理解の深化に向けた学習者の学習改善や指導者の指導法改善へと繋がるだろう。

では、どのような手段で定量的に学習者の学習状況を診断できるのだろうか。その方法の一つとして、認知診断モデル (Cognitive Diagnostic Models; [3]) という数理モデルが挙げられる。認知診断モデルでは、学習者の認知的能力をアトリビュートと呼ばれる複数の要素に分け、各アトリビュートの習得状況として学習状況を詳細に診断できる。そのためテストの合計得点だけからは、みとることができなかった学習者のつま

ずきをより詳しく把握可能となる。分析においては、学習者の各項目への「回答データ」と項目ごとに測定されているアトリビュートをまとめた表である「Q 行列 (Q-matrix; [4])」をもとに統計的推測が行われる。具体的に Q 行列は、最も基本的な 2 値型アトリビュートを用いた場合、項目の正答にどのアトリビュートが必要とされるかを、必要であれば 1、不要であれば 0 が割り振られた $J \times K$ 型行列である (J : 項目数, K : アトリビュート数)。このアトリビュートにどのような能力を対応させて Q 行列を作成するかは、認知診断モデルを適用するにあたって重要なプロセスである。従来、アトリビュートに設定されてきた能力は主として、単元固有の能力であった (e.g., [5, 6])。一方で、前述の国際的な教育目標を鑑みれば、このアトリビュートに理解の深さに関連する能力を対応させることができれば、理解の深さの定量的な評価が可能になるだろう。

Q 行列を作成すると一口にいえど、その過程は容易ではない。認知診断モデル研究の最近の発展をレビューした研究 [7] では、「Q 行列の設定は、文献調査や項目分析、複数の専門家の合議などを経て行われる時間と労力のかかるプロセスである」と述べられているほどである。しかし、これまでの認知診断モデルの応用研究を概観すると、この作成過程は必ずしも詳述されてきたわけではない。また、認知診断モデルによる診断結果が教育現場において学習・指導改善に結びつくためには、テストで測られていることに加えて、学校の授業で指導されていることを考慮して Q 行列

*本論文は第 1 著者による修士論文の一部を学会発表の形態に編集、加筆したものである。

を作成する必要がある。一方、従来は研究者のみが主体となって作成されることが多く、このような取り組みがなされていないことも課題である。そのため、これらの課題に対処しなければ、実際の学校現場で活用されるとは想像し難い。現にRPPs(Research Practice Partnerships; [8])で代表されるように、研究者と教育実践家が関わりあって作り上げていくことで実践的な意義がそこに生まれることは論を俟たない。

そこで、本研究では学校教師と共同して実践的に指導改善に生きる Q 行列の作成とその過程を提案する。さらに、作成された Q 行列やこれに基づく認知診断モデルによる推定結果が、学校教師に指導改善に生きる効果的なリソースとして受け止められるかを、インタビューを通じてその反応を検討することを目的とする。なお、Q 行列作成のプロセスにおいて本研究では Q 行列推定の手法も活用した。従来、Q 行列作成するといえば、専門家による議論という質的な作成方法が主であった。これに対し、Q 行列を定量的に検討しようとする試みが Q 行列推定であり、近年この方法論開発が盛んに行われてきている (e.g., [9])。しかし、この手法はこれまで応用実践に活用されていない。だが、これは研究者だけでなく、実践家である学校教師が Q 行列を作成する上でも、一つの拠り所となりうる手法である。また、量的側面から検討された Q 行列推定の結果も活用することでより妥当な Q 行列を作成することも可能になる。

本研究で提案する Q 行列作成のプロセスは以下の通りである。まず、(1) 研究者が Q 行列推定の結果と深い理解に関する先行研究の文献調査をもとに、Q 行列の第一案を作成する。次に、(2) 学校の先生及び認知・教授学習心理学の専門家との合議を経て、Q 行列の修正を行っていく。さらに、この作成された Q 行列をもとに、(3) 認知診断モデルによって推定された学習者の各アトリビュートの習得状況を学校教師にフィードバックすることによって、診断結果の指導法改善への有用性に関する反応を検討する。このような研究者と実践家の間を往来することで、実際の教育に役に立つものとして作り上げている本研究の Q 行列作成過程をまとめたのが図 1 である。次節では、実践校の概要と Q 行列作成のプロセスを詳述していく。

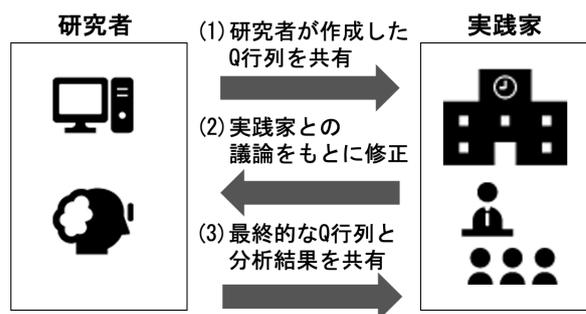


図 1 Q 行列作成のプロセス

2. 実践のフィールドと Q 行列作成のプロセス

2.1 実践のフィールド

静岡県の中立高校で行った (学力は中位層)。対象校は、以前から教員内において有志の研究グループを作り、認知・教授学習心理学の視点から、深い学びを促す授業の実践や深い学びの評価に精力的に取り組んでいた [10]。例えば数学のテスト問題の中には、単に公式に数値を代入して答えを導くだけの公式を丸暗記していても解けるような浅い理解を問う問題だけでなく、なぜその公式が成り立つのかを説明させるような深い理解を測るための問題も含まれていた。そのため、本校を対象校に選定した。分析対象とした定期テストは、数学Ⅱ「指数関数・三角関数」を範囲とした計 36 項目から構成されており、2 年生理系クラス 19 名が回答した。

2.2 Q 行列作成のプロセス

前節の定期テストに対して、Q 行列の作成を行った。作成プロセスとして、まず、(1) 研究者側が主体となって作成する。その際に、教科教育及び心理学的知見を活用するだけでなく、Q 行列推定によって得られた結果も参考にし、認知・教授学習心理学の専門家との議論のもとで作成した。次に、(2) 作成された Q 行列と推定結果をもとに、学校教師とその妥当性に関する議論を行った。最後に、この議論の中で得られた意見をもとに修正した Q 行列を最終的な Q 行列として (3) 分析結果に対する学校教師の反応も示した。

2.2.1 研究者による作成

まず、認知・教授学習心理学を専門とし、数学の中学校・高等学校教諭一種免許状を有する第 1 著者がテ

スト項目の性質、現行の学習指導要領及びこれまでの数学における深い理解に関する先行研究の知見 (e.g., [11, 12, 13, 14, 15]) を参照し、さらに認知・教授学習心理学の専門家である第3著者との協議のもとアトリビュートの特定とQ行列の作成を行った。具体的には、「A1: 定理・計算法則の理解」「A2: 概念 (本単元で習う用語) 理解」「A3: 等式・不等式が成り立つことを論理的に考察する力」「A4: 図をイメージする力」という4つのアトリビュートが特定された。それぞれのアトリビュートについて、「A1: 定理・計算法則の理解」はテスト範囲の単元で扱われた定理・計算法則を当てはめて問題解決できるかという手続きの理解、「A2: 概念 (本単元で習う用語) 理解」は、テスト範囲の単元で扱われた数学的用語の理解、「A3: 等式・不等式が成り立つことを論理的に考察する力」は、等式や不等式の計算において、各式同士のつながりを理解し、論理的に計算を遂行することができる能力、「A4: 図をイメージする力」は、図を活用しながら問題を解決することができる力とそれぞれ定義した。これらのアトリビュートの定義のもとに、教育心理学の専門家2名と教育測定学を専門とする第2著者にそれぞれ独立にQ行列を作成してもらったのち、作成されたQ行列の不一致箇所について、協議を通して統一した。

加えて、Q行列を定量的に検討するために、Q行列のベイズ推定を行った。小サンプル下での推定であるため、探索的なQ行列推定ではなく、協議の上で作成した本Q行列を事前情報として事前分布に反映させて、Q行列の事後分布の推定を行う確認的なQ行列推定を行った。Q行列推定には、先行研究 [9] により提案されたギブスサンプラーによるDINAモデル (deterministic inputs, noisy and gate model; [16] 詳細は3.2.1節) のもとでのQ行列推定法を採用した。なお、Q行列推定の際の事前分布にはベータ分布を用いた。具体的な事前分布の設定としては、協議の上で作成したQ行列において1に当たる成分に対応するQ行列の成分が、同じく1をとる確率が約0.66になるように設定した。また同様に、0にあたる成分に対応するQ行列の成分が同じく0をとる確率が約0.66になるように設定した。推定の際には、先行研究 [9] にて示されたマルコフ連鎖モンカルロ (Markov Chain Monte Carlo; MCMC) 法の設定に従い、イタレーション数は100000、バーンインは50000、チェーン数は3とした。推定結果は、Q行列の各成分の事後期待値及びMAP (maximum a posteriori) 推定値を参照し、本Q行列との不一致箇所に関するQ行列の成分の設定を再検討した。

2.3 現場教師との協議

再検討後のQ行列とこれを用いて認知診断モデルによる分析を行った推定結果をもとに、本テストの作成者を含む高校教師2名と第1,3著者でQ行列と設定したアトリビュートの内容について議論を行った。議論はWeb会議システムツールの一つであるZoomを用いて1時間程度で行われ、はじめの20分では、第1著者がパワーポイントを用いて、認知診断モデルの概要及びQ行列と推定結果の読み取り方について説明した。その後、Q行列や推定結果についての質疑と協議を行った。その結果、本定期テストを作成し、本テストを実施したクラスの数学の授業を担当する教師から「A2: 概念 (本単元で習う用語) 理解」と「A4: 図をイメージする力」のアトリビュートの関連性に関しての発話が得られた。また同様に、「A1: 定理・計算法則の理解」と「A3: 等式・不等式が成り立つことを論理的に考察する力」についても関連性が指摘された。つまり、「A2: 概念 (本単元で習う用語) 理解」というアトリビュートのより高い習得水準に「A4: 図をイメージする力」が位置付き、「A1: 定理・計算法則の理解」のより高い習得水準として「A3: 等式・不等式が成り立つことを論理的に考察する力」が位置づくものであると表現する方法の提案である。これらの議論は、先行研究での知見とも整合的である。用語の理解においては「そもそも...とはどのようなものか」という用語や概念の意味について、用語の意味を断片的な知識として有しているだけでなく、グラフや図などの表象との関連づけが重要とされている (e.g., [11, 12, 13])。よって、「A2: 概念 (本単元で習う用語) 理解」と「A4: 図をイメージする力」を「数学的用語の理解」を診断するアトリビュートとして一括りにして定義し直すこととした。

また、「どのようにその手続きを行うのか」という「手続きの理解」だけでなく、どのような場面でその計算を用いれば良いかや公式がなぜ成り立つのかを考えることが深い理解には必要とされてきた (e.g., [15, 16])。よって、「A1: 定理・計算法則の理解」と「A3: 等式・不等式が成り立つことを論理的に考察する力」を「公式の運用」という名前のもと手続きの理解を診断するアトリビュートとして再定義した。

ここで提案された「数学的用語の理解」と「公式の運用」は授業担当者かつテスト作成者が普段の授業から意識して指導を行っていた理解の要素であった。指導と評価の一体化の観点からも、授業内で意識的に指導したことを評価することが学習・指導改善には効果

的である [17]。そのため、この2つのアトリビュートについて学習者がどの程度深く理解できているかを診断することが、学習改善・指導改善に有用であると考えた。そこで、「数学的用語の理解」と「公式の運用」についてそれぞれ習得か未習得かの2値型で表現するのではなく、理解の深さという習得水準を多値型アトリビュート (e.g., [18]) を用いて表現することとした (i.e., 0:未習得, 1:浅い理解, 2:深い理解)。

加えて、議論の結果、授業内では明示的に扱っていないが、授業内で扱った内容を応用して問題を解く力(「学習内容の転移」アトリビュートと名付けた)、指数・累乗根・三角関数を含む式の計算を遂行する能力(「計算力」アトリビュートと名付けた)を新たに追加した。

3. 作成された Q 行列に基づく分析と現場教師からの反応

3.1 最終的な Q 行列

最終的なアトリビュートの定義を表 1, Q 行列を表 2 に示す。なお、「A1:公式の運用」と「A2:数学用語の理解」は多値型アトリビュート (i.e., A1b(A2b)が必要:2, A1a(A2a)のみ必要:1, いずれも不要:0), 「A3:学習内容の転移」「A4:計算力」には2値型アトリビュート (i.e., 必要:1, 不要:0)。「A1:公式の運用」において、「A1a:単一の公式の手続き的運用」のみが必要とされる場合、1が割り振られ、これよりも習得水準の高い「A1b:複数の公式の論理的運用」が必要とされる項目に対しては2が割り振られている。同様に、「A2:数学用語の理解」において、「A2a:数学用語の一面的理解」のみが必要とされる場合、1が割り振られ、これよりも習得水準の高い「A2b:数学用語の図を介した多面的理解」が必要とされる項目に対しては2が割り振られている。この Q 行列を用いて、認知診断モデルによる分析を行った。

3.2 分析方法

認知診断モデルによる分析では、2値型の DINA モデルを多値型へと拡張した RPa-DINA モデル (reparameterized polytomous attributes DINA model; [19]) を用いた。分析に先立ち、DINA モデル及び RPa-DINA モデルを簡単に導入しておく。

3.2.1 認知診断モデルの数理的導入

DINA モデルは、項目に関連する複数のアトリビュートのうち、一つでも習得できていない場合には、正答確率が大きく低下するという特徴を有する [7]。DINA モデルでは、誤差がない場合に、個人 i が項目 j に正答できるか否かを表す理想反応 η_{ij} が次のように定義される。

$$\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \quad (1)$$

ここで、 $q_{jk} \in \{0, 1\}$ は Q 行列の j 行 k 列成分であり、項目 j がアトリビュート k を必要とするかどうかを表しており、 $\alpha_{ik} \in \{0, 1\}$ は個人 i がアトリビュート k を習得しているかを表す指示変数である。この η_{ij} で条件付けた項目反応確率を

$$P(x_{ij} = 0 | \eta_{ij} = 1) = s_j \quad (2)$$

$$P(x_{ij} = 1 | \eta_{ij} = 0) = g_j \quad (3)$$

とする。(2)式は、項目 j の正答に必要とされるアトリビュートを全て習得しているにもかかわらず、誤答してしまう確率(不注意パラメタ)である。(3)式は、項目 j の正答に必要なアトリビュートが全て揃っていないにもかかわらず正答してしまう確率(当て推量パラメタ)を表している。これら2つの項目パラメタを用いて項目反応関数は、

$$P(x_{ij} = 1 | s_j, g_j, \alpha_i, \mathbf{q}_j) = \left((1 - s_j)^{x_{ij}} s_j^{1-x_{ij}} \right)^{\eta_{ij}} \left(g_j^{x_{ij}} (1 - g_j)^{1-x_{ij}} \right)^{1-\eta_{ij}} \quad (4)$$

と表される。このような特徴を持つのが DINA モデルである。そして、この DINA モデルにおける理想反応 (i.e., (1)式)を

$$\eta_{ij} = \prod_{k=1}^K I(\alpha_{ik} \geq q_{jk})^{I(q_{jk} > 0)} \quad (5)$$

多値型アトリビュートに拡張したのが RPa-DINA モデルである。理想反応以外の要素は全て DINA モデルと同じである。なお、(5)式で $I(\cdot)$ は指示関数であり、括弧内の条件式が正しい場合に1、正しくない場合に0を返す。理想反応以外の要素は全て DINA モデルと同じである。

3.2.2 RPa-DINA モデルによる分析

分析において、統計ソフトウェアとして R ver 4.0.2 と JAGS(just another Gibbs sampler) を用いた。ま

表 1 アトリビュートの定義

アトリビュート名		アトリビュートの定義
A1：公式の運用	A1a：単一の公式の手続き的運用	単一の公式に数値を代入して適用する力
	A1b：複数の公式の論理的運用	いくつかの公式をもとにそれらを組み合わせ式同士の繋がりを理解して問題を解決する力
A2：数学用語の理解	A2a：数学用語の一面的理解	文章中の数学用語が示すことばを把握・処理する力 (ただし、図をもとに把握・処理する必要はない)
	A2b：数学用語の図を介した多面的理解	数学的用語や三角関数 (i.e., sin, cos, tan) が示すことばを介して理解する力
A3：学習内容の転移		授業内では明示的に取り扱われていない問題だが、授業で習った内容を応用して正解を導くことができる力
A4：計算力		指数・累乗根・三角関数同士の計算を遂行して答えを導く力

表 2 最終的な Q 行列

項目	アトリビュート				項目	アトリビュート			
	公式の運用	数学的用語の理解	学習内容の転移	計算力		公式の運用	数学的用語の理解	学習内容の転移	計算力
1	1	1	0	1	19	0	2	0	0
2	1	1	0	1	20	0	2	0	0
3	1	0	0	1	21	0	2	0	1
4	1	0	0	1	22	0	2	0	0
5	1	0	0	1	23	0	2	0	0
6	1	0	0	1	24	1	0	0	1
7	1	0	0	1	25	2	1	1	1
8	1	0	0	1	26	2	2	0	1
9	2	0	0	1	27	1	1	0	1
10	0	2	0	0	28	2	2	0	1
11	0	2	0	0	29	2	2	1	1
12	1	2	0	1	30	2	2	0	1
13	0	2	0	0	31	1	1	0	1
14	0	2	0	0	32	1	1	0	0
15	0	2	0	0	33	2	0	1	1
16	0	2	0	0	34	1	0	1	1
17	0	2	0	0	35	2	2	1	1
18	0	2	0	0	36	2	2	1	1

ず、観測データに対する RPa-DINA モデルの適切性を検討するために、ピアソン残差に基づく、事後予測 p 値を計算した [20]。分析の結果、事後予測 p 値は 0.45 であり、この値は 0.5 に近いほど指定したモデルからの事後予測値と観測値に乖離がないことを示すため、RPa-DINA モデルの適切さが支持された。よって、以下 RPa-DINA モデルを用いて分析を行った。

本推定では、MCMC 法によるベイズ推定を行った。各パラメタに対する事前分布は先行研究 [9] を参考に

$$s_j \sim \text{Beta}(1, 1) \quad (6)$$

$$g_j \sim \text{Beta}(1, 1)_{T(0, 1-s_j)} \quad (7)$$

$$c_i \sim \text{Categorical}(\boldsymbol{\pi}) \quad (8)$$

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\mathbf{1}) \quad (9)$$

のように定めた。なお、項目パラメタには、 $0 \leq g_j \leq 1 - s_j \leq 1$ という単調性制約 (monotonicity restriction) を課している [21]。

MCMC 推定のイタレーションは 30000 回、バーンインは 10000 回、チェーン数を 4 とした。MCMC 推定における収束判定指標 \hat{R} は 1.01 未満 [22] であった。またトレースプロットも視覚的に確認したところ、各チェーンにおける乖離は見られなかった。分析に際し、アトリビュートに関しては、各個人のアトリビュート習得程度の事後期待値の算出とアトリビュート習得程度の事後期待値に基づく階層的クラスタリング (Ward 法) を行った。階層的クラスタリングのクラスタ数は、得られたデンドログラムから視覚的に 3 が適切だと判断した。以下では、得られた結果を示しながら、解釈を行っていく。

3.3 分析結果とその解釈

3.3.1 各生徒のアトリビュート習得状況

まず、各生徒のアトリビュート習得状況を捉えるために、19 名分の各生徒のアトリビュート習得程度の事後期待値を推定した。その結果から 4 名分を抜粋したのが表 3 である。

表 3 から、生徒 ID10 と 14 の生徒はそれぞれ総合得点が 52 点、54 点であった。定期テストのフィードバックにおいてよく与えられる総合得点の情報のみからはこれら二人の今回のテスト範囲の理解度は同程度であると考えられる。しかし、アトリビュート習得状況からは、各学習者の学習上のつまずきについて、より詳細な診断情報を得ることができる。以下、アトリビュート習得程度の推定結果の解釈の基準として、

多値型アトリビュートである A1, A2 は 0.5 未満が習得できていない、0.5 以上 1.5 未満は浅い理解とみなされる習得水準 (i.e., A1a, A2a) に達している、1.5 以上は深い理解とみなされる習得水準 (i.e., A1b, A2b) に達している状態とする。また、2 値型アトリビュートを設定した A3, A4 については [23] に倣い、0.5 以上が習得、0.5 未満が未習得も状態であると解釈する。

生徒 ID10 の生徒は「A2: 数学的用語の理解」については、アトリビュート事後期待値が 1.93 のため「A2b: 数学的用語の多面的理解」という深い理解まで習得していると考えられる。しかし、その他のアトリビュートについてはいずれも習得しているとされる基準を下回っているため、習得できていないと考えられる。一方、個人 ID14 の生徒のアトリビュート習得程度を見てみると、「A1: 式の理解」はアトリビュート事後期待値が 1.59 であり、「A1b: 複数の公式の論理的運用」という深い理解まで習得できていると考えられ、A3, A4 についても習得していると考えられる。しかし、「A2: 数学的用語の理解について」は、アトリビュート事後期待値が 1.09 であり、「A2a: 数学的用語の片面的理解」にとどまっていることがわかる。このように、同程度の総合得点だったとしても、習得しているアトリビュートは異なり、特に習得の水準も異なるという診断情報を得ることもできる。

この結果は、総合得点のような一次元的な尺度よりも、より詳細な診断情報を与えてくれる。個人 ID10 の生徒には特に「公式の運用」について、例えば、どのような公式が存在し、その公式はいつ使うことができるのかを意識した学習や指導の必要があるだろうと学習・指導方針を立てることもできる。一方、個人 ID14 の生徒には、「数学的用語の理解」について、数学的用語を図を介して理解するために、例えば、三角関数が表している図的な意味は何かを意識した学習や指導の必要があると学習・指導方針への示唆を得ることができる。このように、各個人のアトリビュート習得状況の推定結果から学習・指導改善に有益な情報を得ることができた。

3.3.2 アトリビュート事後期待値に基づく階層的クラスタリング

次に、各生徒のアトリビュート事後期待値に基づいて階層的クラスタリングを行い、回答者全体を 3 つの理解度別のクラスタに分類した。この結果を表 4 に示す。

この結果は、特に教師がクラス全体の理解度の傾向

表3 各生徒のアトリビュート習得程度の事後期待値 (一部抜粋)

生徒 ID	総合得点	公式の運用	数学用語の理解	学習内容の転移	計算力
3	38	0.39	1.51	0.50	0.35
4	37	0.86	0.30	0.41	0.55
10	52	0.34	1.93	0.49	0.38
14	54	1.59	1.09	0.53	0.99

表4 クラスごとのアトリビュート習得程度の事後期待値と平均総合得点

クラス ID	平均総合得点	公式の運用	数学用語の理解	学習内容の転移	計算力
1	67.8	1.59	1.75	0.62	0.90
2	40.5	0.42	1.59	0.49	0.39
3	34.0	0.96	0.26	0.38	0.64

をマクロに捉え、今後の指導改善に活かす際に有用な情報である。クラス番号1では、今回のテスト範囲においては、すべてのアトリビュートを習得しており、深い理解まで到達できていることがわかる。一方、クラス番号2では、「A1:数学的用語の理解」については、深い理解まで到達できているもののそれ以外のアトリビュートについては習得できていないことが読み取れる。そして、クラス番号3に属する生徒は「A4:計算力」については習得できているものの、「A1:公式の運用」は浅い理解にとどまっており、A2、A3は習得できていない傾向があることがわかる。

以上のような解釈に基づけば、クラス番号1に属する生徒は、本単元において習得してほしいアトリビュートは全て習得できているため、次の学習範囲に円滑に進むことができるだろう。しかし、クラス番号2に属する生徒には、既出の公式の再確認とどのように使うのか、そして、それらを計算する力を指導する必要がある。そして、クラス番号3に属する生徒には、式の適用場面を意識させ、いくつかの式を論理的に組み合わせて答えを導く力を養い、数学的用語については、基本的な用語を確認した後にそれらがどのように図やグラフなどの表象と結びついてくるのかについて教授していくべきであろう。このように、各個人のアトリビュートだけでなく、クラス全体をアトリビュートの習得状況によって分類することで、どのようなつまづきのグループが存在するのかを把握することができる。これらは、今後の指導方針を決定していく際にも、どの説明に重点を置いていくべきかを考える上で、有用な診断情報となるだろう。

3.4 推定結果の教師との共有

このような推定結果やQ行列は、学校教師にどのように受け止められたのだろうか。本節では、高校教師2名と第1・3著者で行ったグループインタビューの結果をもとにこれを検討する。はじめに、得られた推定結果である「各生徒のアトリビュート習得程度の事後期待値」「クラス分析の結果」及び、最終的なQ行列を学校教師と共有した。なお、結果の共有にあたっては、Q行列作成時の学校教師及び認知心理学の専門家との協議時と同様に、まず第1著者がパワーポイントの資料を用いて、改めて認知診断モデルの概要と今回得られた各推定結果が表す意味と解釈の方法について説明した。その後、学校教師から推定結果に対してインタビューを行った。

以下、本節では、認知診断モデルによる推定結果を高校教師2名と共有した際にどのように捉えられたかを紹介する。なお発話において、T1は本定期テスト作成者かつ本定期テストを実施したクラスの数学を担当している教師、T2は本定期テストを実施したクラスの物理を担当している教師の発言を表している。まず、認知診断モデルによる分析結果を初めて見た時の教師の反応を紹介する。

何か分からない箇所や不明点はあるかということを探ったところ、認知診断モデルの枠組みは学校現場の教師において、理解しやすい枠組みであり、多次元的能力をアトリビュートとして測ることができるため、観点別評価への応用の可能性も示唆された。さらに、Q行列という枠組みでテストで測ろうとしていた能力を再確認することができるという点においても有用性を感じていたことが以下の発話から読み取れる(特に、下線部)。

T1:

「なんだこれっていうのは特になく、ずっと入ってきて非常に分析していただいてありがたいなっていう感じなんですけど、表(Q行列)でこの1の数とかみた時に自分自身もテスト作っていて、まあバランスがちょっと偏っていたりだとか、そういう点では非常にいい指標で問題むしろ作る上でこういう意識して作りたいなって今後思ったところとあとこのアトリビュートっていうところで、教科の観点別評価にかかわるなって思って」

また、教師がテスト作成時に測ろうとしていた能力をアトリビュートという形で明確に表すことが、無意識的にテスト作成時に重要視していることがらの意識化にも有用であったことが示唆された。さらに、本研究で作成したQ行列や推定結果が指導改善に活かせることも、以下の発話から読み取れる(特に、下線部)。

T1:

「確かに、式の理解、概念理解、応用的なものも入れようかなと確かに思って作成している感じがあるので、多分テスト作成している時はここまで意識はしていないと思うんですけど、こういうふうに分けられたことで確かに自分でそういうふうで作っている感じはありますし、なんかこう今後の指導にはすごい生きるような感じがするので、例えば生徒によって、概念理解はしっかりしているけど結局計算力がないなって場合には計算力のところをつけなきゃいけないなって思いますし、ただテストやって終わりじゃなくてこの分析結果を知ることで今後の指導には非常に生きると感じるので、分け方については自分は数学でいうとしっくりくる感じがします」

T2:

「我々問題作るときは、ここまで深く考えていないですよ、正直。でやっぱ(Q行列を)作った後に振り返ってみて、こういう骨格が見えてくる。自分にはこういう意図があったんだなあって、概念理解の問題が入っていたりとか、計算の問題が入っていたりとか、後から。」

このように、認知診断モデルによる推定結果を指導法改善に活かせる可能性だけでなく、Q行列という枠組みでテスト問題を捉え直すことで、教師が無意識のうちに重要だと思っていたことが明確化されることが示唆された。さらに、これらによってテストで測って

いる能力が適切かどうかを再検討でき、教師の評価リテラシーの向上にも寄与する可能性が示唆された。

4. まとめと考察

本研究では、認知診断モデルを用いて、学習者の理解の深さを定量的に診断する方法の提案を行った。さらに、認知診断モデルの実践上、課題であったQ行列作成の具体的なプロセスの提案も行った。具体的には、(1)研究者が心理学・教科教育の知見を参考にQ行列を作成し、学校教師へ共有。(2)学校教師との協議を通じ、普段の指導で重視している能力をアトリビュートの定義に取り入れ、Q行列を修正。(3)最終的なQ行列と認知診断モデルによる分析結果を学校教師へフィードバックするというプロセスで構成された。

結果として、理解の深さを反映したアトリビュートとQ行列を特定したことで、各学習者の理解度におけるつまづきを各生徒のアトリビュートとして詳細に診断可能となった。また、各生徒のつまづきをミクロに捉えられるだけでなく、アトリビュート習得程度の事後期待値に基づいたクラスタ分類を行ったことで、マクロな視点から生徒全体をつまづきのタイプ別に分けて捉えることも可能となった。これらの結果は、習熟度別のクラス編成にも一役買えるであろう。学校現場ではしばしば、普段のクラスとは別に、主に数学や英語の授業において、習熟度別クラスを設けることがある。この習熟度別の基準として、定期テストの総合得点の順位が活用され、次学期のクラスが構成される。しかし、ここで総合得点の高低に基づくクラス編成ではなく、アトリビュート習得状況のクラスタを参考にしたクラス分けをすることで理解度につまづきに応じた数学の指導に役立つ可能性が十分に考えられる。

また、本研究では、認知診断モデルの実践的活用に向けた提案を行ってきた。しかし、学校教師が統計ソフトウェアを用い認知診断モデルによる分析を行えるかといえば難しい。そのため、将来的には、学校教師が認知診断モデルの結果をより容易かつ適切に算出できるような、より利用しやすいインターフェースの開発などが必要であろう。

本研究で提案したアトリビュートは、数学の「指数関数・三角関数」を題材としたため、アトリビュートの定義は出題内容に依存する部分も含まれる。そのため、本研究で提案したアトリビュートの定義自体をそのまま他のテストへと適用することは難しい。しかし、本アトリビュートを設定する上で重視された視点や作成過程自体は今後、他のテストに認知診断モデルが応用される際に重要な知見となるだろう。ここでい

う視点は具体的に、多値型アトリビュートを設定して理解度の深さを表してきた「公式の運用」や「数学的用語の理解」という本研究で学校教師からも重視された視点のことである。実際、本研究で提案されたアトリビュートの視点は、大規模テストへの分析にも活用されており、その結果として、従来、テスト開発会社が設定してきた Q 行列よりもモデル適合度指標の観点から当てはまりが良いことが示されている [24]。

最後に、当然ではあるが、本研究で提案されたアトリビュートが唯一解というわけではない。アトリビュートは、診断の目的に応じて設定されるべきである。本研究では、深い理解を涵養するために、学習者の理解の深さを診断したいという目的があった。だからこそ、理解の深さに基づいたアトリビュートを提案してきた。そのため本研究の知見は、従来、教育実践現場において活用されることが少なかった認知診断モデルが、学習・指導改善へ実際に活用されるための一つの足がかりとなれば幸いである。

参考文献

- [1] 文部科学省 (2017) “新しい学習指導要領の考え方—中央教育審議会における議論から改訂そして実施へ—”, Retrived from https://www.mext.go.jp/a_menu/shotou/new-cs/_icsFiles/afielddfile/2017/09/28/1396716_1.pdf (2022年4月21日閲覧).
- [2] Bellanca, J. A. (2014). “Deeper learning: Beyond 21st century skills.” Bloomington, IN: Solution Tree Press.
- [3] Rupp, A. A., Templin, J., & Henson, R. A. (2010). “Diagnostic measurement. Theory, Methods, and Applications.” New York: Guilford.
- [4] Tatsuoka, K. K. (1983). “Rule space: An approach for dealing with misconceptions based on item response theory.” *Journal of educational measurement*, 345–354.
- [5] Sun, Y., Suzuki, M., & Toyota, T. (2013). “Designing effective feedback for cognitive diagnostic assessment in web-based learning environment.” In *Proceedings of the 21st International Conference on Computers in Education.*, 115–120.
- [6] Madison, M. J., & Bradshaw, L. (2018). “Evaluating intervention effects in a diagnostic classification model framework: Intervention in a diagnostic classification model.” *Journal of Educational Measurement*, 55, No. 1, 32–51.
- [7] 山口一大・岡田謙介 (2017). “近年の認知診断モデルの展開”, 行動計量学, 44, No.2, 181–198.
- [8] Welsh, R. O. (2021). “Assessing the quality of education research through its relevance to practice: An integrative review of research-practice partnerships.” *Review of Research in Education*, 45(1), 170–194.
- [9] Chung, M. (2019). “A Gibbs sampling algorithm that estimates the Q-matrix for the DINA model.” *Journal of Mathematical Psychology*, 93, 102275.
- [10] 植阪友理・廣澤一徳・坂口卓也・水野木綿・太田絵梨子・柴里実・富田真永・眺野翠・梶山佳明 (2020). “深い学びを評価する高校における定期考査の提案とその効果—高大連携による公立学校での実践的検討—”, 日本教育心理学会総会発表論文集第 62 回総会発表論文集, 102, 日本教育心理学会.
- [11] Manalo, E., Uesaka, Y., Chen, O., & Ayabe, H. (2019). “Showing what it looks like: Teaching students to use diagrams in problem solving, communication, and thinking. In *Deeper Learning, Dialogic Learning, and Critical Thinking.*” Routledge. pp. 230–246.
- [12] Rittle-Johnson, B., & Alibali, M. W. (1999). “Conceptual and procedural knowledge of mathematics: Does one lead to the other?” *Journal of educational psychology*, 91, No. 1, 175.
- [13] 市川伸一 (2000). “勉強法が変わる本—心理学からのアドバイス”, 岩波書店.
- [14] Lachner, A., & Nückles, M. (2016). “Tell me why! Content knowledge predicts process-orientation of math researchers’ and math teachers’ explanations.” *Instructional Science*, 44, No.3, 221–242.
- [15] 藤村宣之 (2012). “数学的・科学的リテラシーの心理学: 子どもの学力はどう高まるか”, 有斐閣.
- [16] Junker, B. W., & Sijtsma, K. (2001). “Cognitive assessment models with few assumptions, and connections with nonparametric item response theory.” *Applied Psychological Measurement*, 25, No. 3, 258–272.
- [17] 国立教育政策研究所 (2020). “「指導と評価の一体化」のための学習評価に関する参考資料 (高等学校数学)”, 東洋館出版社
- [18] Karelitz, T. M. (2004). “Ordered category attribute coding framework for cognitive assessments” (unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.
- [19] Zhan, P., Jiao, H., Man, K., & Wang, L. (2019). “Using JAGS for bayesian cognitive diagnosis modeling: A tutorial.” *Journal of Educational and Behavioral Statistics*, 44, No. 4, 473–503.
- [20] Levy, R., & Mislevy, R. J. (2016). “Bayesian psychometric modeling.” CRC Press.
- [21] Ma, W., & Jiang, Z. (2021). “Estimating cognitive diagnosis models in small samples: Bayes modal estimation and monotonic constraints.” *Applied Psychological Measurement*, 45, No. 2, 95–111.
- [22] Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2020). “Rank-Normalization, Folding, and Localization: An Improved \hat{R} for Assessing Convergence of MCMC (with Discussion).” *Bayesian Analysis*, 16, No.2, 667–718.
- [23] de la Torre, J., Hong, Y., & Deng, W. (2010). “Factors affecting the item parameter estimation and classification accuracy of the DINA model.” *Journal of Educational Measurement*, 47, No. 2, 227–249.
- [24] 佐宗駿・岡元紀・植阪友理 (2021) “認知診断モデルを通じた深い理解の実証的解析—大規模学力調査を用いた分析と従来の観点との比較—”, 日本テスト学会第 19 回大会発表論文妙録集, 118–121.