

# プレイスメントテストの項目分析の利用について

—入学者選抜と入学後教育の追跡調査を目指して—

大竹 洋平 (山梨英和大学)

入試選抜制度の評価のためには、入学後の学習状況を追跡調査することが考えられる。しかし、現行の入試制度には様々な方式が存在するので、異なる入試制度を経てきた学生の学力を相互に比較するのは困難である。また、選抜効果もあることにより、入試と入学後の状況の連関をみるのも難しい。よって本論文では、新入生のプレイスメントテストを軸として、入試区分に遡って比較・追跡することを提案する。また、その足がかりとして、プレイスメントテストの問題項目の項目分析を行った結果も示す。

キーワード：追跡調査、プレイスメントテスト、項目分析、項目特性図

## 1 はじめに：研究の背景

入試選抜の評価としては、入学後の学習状況を追跡することが第一に考えられる。しかし、学生の学力を比較・追跡するのに、入試自体を軸とすると以下のような困難性がともなう。

まず、現行の大学入試制度には、国公私の別を問わず、AO 入試・推薦入試・一般入試（今後は、総合型選抜・学校推薦型選抜・一般選抜と名称変更される）や、その他、様々な選抜区分が存在する。したがって、入試区分の異なる学生の学力を比較することは、困難である。

また、大学入学後の学習状況（成績など）と、入学時の試験における得点との関連（相関）が分析されるが、そこには「選抜効果」が存在する（芝・渡部, 1988）。すなわち、入学試験の受験者の中には、未入学者（不合格者および合格しても入学手続きをしない者など）が存在するため、データの打ち切りがあることになるため、受験時の学力と入学後の成績との相関が低く見積もられてしまうことである。

したがって、本研究では、受験生集団の入学試験の成績を元に追跡調査を設計するのではなく、入学者集団を考え、プレイスメントテストを軸として、追跡調査を設計することを試みる。

そのための最初の段階として、プレイスメントテストの問題項目の分析をおこなったので、その結果を報告する。

### 1.1 本論文の構成

本稿の構成は以下である。まず第 2 章で、先行研究を紹介する。そこでは、先行研究を、入試選抜区分ごとの比較（2.1 節）、追跡調査（2.2 節）およびテ

ストの項目分析（2.3 節）の区分において整理する。

次に第 3 章で、本稿のメインであるプレイスメントテストの項目分析の方法を紹介する。基本的な項目分析の方法として、まず、2 値反応データ行列（3.1 節）について紹介する。次にその行列を分析する、困難度（3.2 節）、識別力（3.3 節）、項目特性図（3.4 節）についてまとめる。

第 4 章では、Z 大学 Y 学部の新入生全員に実施したプレイスメントテストの結果とその項目分析結果を示す。

第 5 章では、結果を精査することで、今後、追跡調査へと拡張していける可能性を論じるとともに、テストの縮刷版が作成できることにより、非常時への備えもできることを論じる。

## 2 先行研究

### 2.1 入試選抜区分ごとの比較調査研究

ここでは、入試選抜区分ごとの比較を整理する。AO 入試、推薦入試、一般入試（前期・後期）などの入試区分のうち、どの形態で入ってきた学生が、入学後に高いパフォーマンスをあげているかを調べた調査研究である。AO 入試を中心として、新規で設定した特定の入試制度の評価を目的としている場合には、入学後の成績と関連させることが多いので、次の 2.2 節で議論する。個別大学の AO 入試の分析については、多くの報告がある。ほとんどの大学で AO 入試を実施している現状では、個別大学の内部資料として存在すると思われるが、その中でも、公表されているのは、選抜が成功しており、かつ公表するだけのマンパワーのある大学だと考えられる。国立大学の最初期の AO 入試として、筑波大学（白川ほか, 2004）、東北大学

(倉元・大津, 2011), 九州大学(林ほか, 2012)がある。その他にも, 山口大学(林, 2011), 京都工芸繊維大学(山本・内村, 2011), 福井大学(大久保ほか, 2012)など多く存在する。

次に, 幾つかの入試区分の併願について論じる。私立大学では, 同じ入試区分を数回に渡って実施することが一般的であるから当然として, 分離分割方式をとってきた国公立大学においても, 入試方式が多様化しているため, 一人の受験生がAO入試・公募制推薦入試・一般入試前期日程・一般入試後期日程のように, 複数回の受験が可能である場合が増えている。それら同一年度内での複数の受験方式の受験生の出願(併願)・入学状況を公開している大学は少ないが, 東北大学(倉元・大津, 2011)や名古屋大学(石井, 2014)の例がある。

## 2.2 入試と入学後とをつなぐ追跡調査

追跡調査というと, 学生の受験状況から, 入学後の成績, さらに, 卒後の進路のように順を追って追跡する調査が主である。公表されているかどうかは別として, 個別大学によって, 様々な調査がなされてきていると想定される。

『大学入試データの解析[理論と応用]』に所収の宮原ほか(1999), 千野(1999)をはじめ, 古くから調査研究がある。特に, 入試成績と入学後の成績との相関を調べようとする試みは, 多くなされようとしてきているが, 入試を挟むと不合格者のデータが得られないことによる打ち切りの効果(選抜効果)により, 相関が低く出てしまうため, 注意が必要である(芝・渡部, 1988)。

前の2.1節で, 個別大学のAO入試を評価する研究群を紹介したが, それ以外でも, 入試と入学後の成績(特に, GPA)との関連を調べた調査研究が多い。最近では多くの大学でGPA(Grade Point Average)が算出されるようになってきているため, それを追跡調査に用いることが多くなってきている。(例えば, 吉村・木村(2011)や山田・西本(2014)などを参照。)しかし, そもそもGPA制度自体に批判もある。GPA自体を精緻化する研究(半田, 2012)もある。また, 履修単位数が少ない方が, 平均であるGPAの値が上がりやすいという特性を補正するために, Grade Pointの平均だけでなく, Grade Pointの合計を利用することも考えられる。実際に, Z大学Y学部のように, 併用している例もある。

一方で, 学生自身の意見や卒業後のデータなどを利用した調査研究もある。そもそもGPAは, 学生の学修行動の教員による間接評価であると考えられるため,

直接評価を利用して検討する研究もある(林, 2015)。さらに, 西郡(2011)のように, 卒後情報も含め, どのような情報で追跡調査をすればよいかを検討した研究もある。追跡調査には, 丁寧なデータベース設計(北川, 2004)が不可欠であり, 入試に関する注意事項をまとめたものに, 大竹(2020)がある。

また, 入学前ないし受験前まで追跡しようと思うと, 入試広報活動の分析も必要であり, それを俯瞰するものに, 倉元・泉(2014)や大竹(2019)などがある。

## 2.3 テストの項目分析

本論文では項目の困難度・識別力を求めていくが, これらの量はIRTの困難度(いわゆるbパラメタ)・識別力(いわゆるaパラメタ)に対応する。

IRTなど現代的なテスト理論(池田, 1994; 熊谷・荘島, 2015; 光永, 2017; 村木, 2011; 芝, 1991; 豊田, 2012)を用いずとも, 古典的なテスト理論の範囲でテストの項目分析は可能である(肥田野, 1972; 池田, 1972; 池田, 1992)ので, 本論文でも, 古典的なテスト理論によるテストの項目分析を行なっていく。ちなみに, 現代テスト理論の文献には, 古典的テスト理論の記載を含むものが多い。

## 3 方法

本研究では, 第1章「はじめに: 研究の背景」でも述べたように, 入試問題自体ではなく, プレイスメントテスト問題の項目分析を行う。

### 3.1 反応データ行列

一般に, 受検者 $N$ 人が問題項目数 $M$ のテスト冊子に回答した場合を考える。受検者 $i$ ( $i=1, 2, \dots, N$ )が, 問題項目 $j$ ( $j=1, 2, \dots, M$ )に正答した場合を1, 不正解(空白無回答を含む)の場合には0を与えると, 0と1のみを要素とする $N \times M$ の要素をもつ行列 $U$ ができあがる(2値反応データ行列)。それら进行操作することで, 様々な分析が可能になる。

### 3.2 各項目の困難度(正答率・通過率)

第 $i$ 行の要素を合計すると, 受検者 $i$ の正答項目数が, 第 $j$ 列の要素を合計すると, 問題項目 $j$ の正答者数が得られる。したがって, 行列 $U$ の要素を $u_{ij}$ ( $i=1, 2, \dots, N; j=1, 2, \dots, M$ )とすると,  $j$ 列の平均 $p_j$ は以下のように表される。

$$p_j = \frac{1}{N} \sum_{i=1}^N u_{ij}$$

この $p_j$ によって、問題項目  $j$  の正答率（通過率）が得られる。これが、その問題の難易度を表していると言える。

ちなみに、項目  $j$  の分散は、 $p_j(1 - p_j)$ であるが、それ自体がテストの項目分析で用いられることは稀である<sup>2)</sup>。

### 3.3 各項目の識別力（正答数との相関係数）

次に、各項目の識別力を算出する。上記行列  $U$ （2 値反応データ行列）を、さらに操作する。受検者  $i$  の正答項目数 $x_i$ は、第  $i$  行の要素の合計、すなわち、

$$x_i = \frac{1}{M} \sum_{j=1}^M u_{ij}$$

で表せる。

項目  $j$  の識別力 $r_j$ は、反応行列の  $j$  列と正答項目数との相関係数（点双列相関係数である項目テスト間相関：IT 相関）で定義される（熊谷・荘島, 2015; 豊田, 2012）。したがって、

$$r_j = \frac{\sum_{i=1}^N (u_{ij} - p_j)(x_i - \bar{x})}{\sqrt{\sum_{i=1}^N (u_{ij} - p_j)^2} \sqrt{\sum_{i=1}^N (x_i - \bar{x})^2}}$$

のように、項目識別力 $r_j$ が定義できたことになる。ただし、 $p_j = \frac{1}{N} \sum_{i=1}^N u_{ij}$  と  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$  である。

### 3.4 各項目への応答（項目特性図）

識別力以外にも、問題項目  $j$  についての弁別性を表すものに、項目応答関（項目特性図）がある（豊田, 2012; 熊谷・荘島, 2015）。これは、横軸に学力（図ろうとしている特性）を、縦軸に正答確率を取ってグラフ化したものである。

テスト得点（各受検者の正答数）を元に、グループ化（3 階層ないし 5 階層程度）に区分した集団ごとの、各項目の正答数を、プロットした図である。

## 4 問題項目分析の結果

Z 大学 Y 学部では入学時オリエンテーションの際に、国語・数学・英語などの教科ごとに、実施している。各教科のテストは、基礎学力の診断だけでなく、習熟度別の授業のクラス分けにも利用されることとなる。

本研究では、分析の手始めとして、ある年度の新生 180 人を対象とした。各科目の出題形式についてだが、数学は記述式ではあるものの、計算の途中過程などではなく、解答のみを採点の対象としており、0-1 の 2 値正誤データが得られる。一方で、国語・英

語は、論述式の項目も含まれるため、部分点等が存在する。その意味で、各教科のなかでも、正解・不正解を明確に分離した 0-1 データを得ることができる「数学」を対象とした。問題項目数は 20 であった。

プレースメントテスト数学の各項目の内容についてみてみると、全ての新生が学習した範囲から出題するために、中学校 1 年生から高等学校 1 年生までの学習内容から出題されている。

### 4.1 各項目の困難度および識別力

20 項目の困難度として、項目正答率（通過率）を、識別力として、先に述べた相関係数（点双列相関係数）を、以下の表 1 に示す。また、参考までに、各項目の学習時期も示す。

表 1 プレースメントテストの項目分析表

項目	学習時期	正答率	相関係数
1	中 1	40.1	0.55
2	中 1	50.8	0.70
3	中 1	58.8	0.65
4	中 1	24.3	0.51
5	中 2	51.4	0.55
6	中 2	33.3	0.69
7	中 2	84.7	0.41
8	中 2	74.0	0.50
9	中 2	67.8	0.35
10	中 2	31.6	0.47
11	中 3	51.4	0.62
12	中 3	26.6	0.64
13	中 3	57.1	0.71
14	中 3	4.0	0.26
15	中 3	23.2	0.63
16	高 1	43.5	0.65
17	高 1	11.9	0.52
18	高 1	20.3	0.63
19	高 1	56.5	0.53
20	高 1	41.8	0.47
平均		42.7	

各項目の識別力をみると、相関係数の絶対値が極端に小さかったり、マイナスだったりする項目がないことから、概ね良好であると言える。

また、出題範囲（学習学年）ごとに、困難度も適度に散らばっている。

ただし、一部正答率が低い（困難度が高い）項目については、簡単に理由を説明する。高校1年の学習範囲である項目17は、放物線の問題だが、変数 $x$ と定数 $a$ とを使い分ける応用問題であるため、正答率が低かったと考えられる。また、中学3年の範囲である項目14も、放物線の問題で、定義域に対応した変域を求める応用問題で、正答率が低かったと考えられる。中学1年の範囲である項目4は、扇型の面積を求める問題であるが、中学・高校のその後の学習ではほとんど出てこない図形の問題であるため、学習年次の割には、正答率が低かったものと考えられる。

数学という科目の性質上、（語学のように）多くの範囲が段階的に習得されるものではなく、一部、学習年次が下であっても、その分野だけに固有の能力が存在することが考えられる。すなわち、数学という能力の次元性を完全には仮定できないことが考えられる。

これらの分析より、このテスト冊子は、全体として、幅の広い学力を、概ね適切に計測していると言える<sup>3)</sup>。

#### 4.2 各項目の応答図（項目特性図）

問題項目 $j$ への応答（項目特性図）を示す。ここでは、5群（ここでは、低得点者群、中の低得点者群、中の中得点者群、中の高得点者群、高得点者群と名付ける。）に分けた結果を示す。

前節のように、各項目の識別力が概ね良好であり、ほとんどの項目が、豊田（2012）の分類で言うことの、識別力の高い項目（G, L, H型）であることになる。よって本稿では、典型的なL型, G型, H型の3項目の結果のみ示す。L型（下位識別項目）としては、正答率が高い項目（項目8：通過率74.0, 識別力0.50）、G型（高識別力項目）としては、正答率が中程度の項目（項目20：通過率41.8, 識別力0.47）、H型（上位識別項目）としては、正答率の低い項目（項目17：通過率11.9, 識別力0.52）の項目応答図を、それぞれ、以下の図1に示す。

L型は、低得点者群は不正解だが、それ以外の層では、正答の割合が高くなり、ほぼ横ばいである。G型は、低得点者群は不正解の割合が多く、そこから単調増加を示し、高得点者群での正答率が高くなっていく。H型は、中の中得点者群（もしくは、中の高得点者）まで、不正解の割合が多いが、高得点者群になると、正解の割合が高くなる。

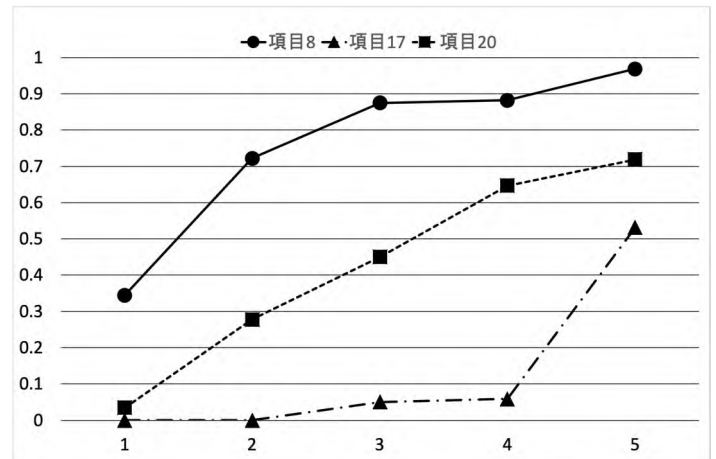


図1 典型的な項目の応答図（項目特性図）

#### 5 考察とまとめ

Z大学Y学部のプレイスメントテスト・数学のテスト冊子の項目分析を行ったところ、全20項目について、識別力の問題はなく、難易度が適度に散らばっていることがわかった。よって、同じ問題を使って、ある程度新入生の学力を計測できると考えられる。

今後、同じ問題項目を使った経年変化の分析や、入学後の成績との関連の分析、さらに、元々の目的であった入試区分ごとの分析・入試制度の評価へもつなげていけると考えられる。

また、今回の項目分析の応用として、緊急事態への対処が考えられる。以下で、追加の議論・考察として、テスト冊子の再構成を議論して、本稿を閉じる。

#### 5.1 テストの項目分析による利点：テスト冊子再構成への活用（追加の議論・考察）

2020年初頭より、新型コロナ・ウイルスが蔓延したことにより、2020年度は授業開始を遅らせる大学が多く、また、大規模集会の自粛により、新入生オリエンテーションも規模を縮小しての実施を余儀なくされている。プレイスメントテストも、規模を縮小して実施したり、受検時間も短縮したりすることが想定される。

受検時間を短縮したうえで、プレイスメントテストを実施するという事は、テスト冊子全体の縮刷版を構成するという事でもある。その場合、テストの項目分析を事前にしておいたことにより、縮刷版のテスト構成が比較的容易に行うことができるようになった。

具体的には、40分で構成していたテスト冊子を、30分で実施する場合、テスト冊子を選択ないし分割して実施することが可能になる。

注

- 1) 新入生全員を対象とした基礎学力テスト。習熟度別授業などのクラス分けなどにも用いられる。配置テストとも呼ばれる (Linn, 1989 池田訳 1981)。
- 2) 豊田 (2012) の p.11 には、「分散には、通過率 (平均値) 以上の情報がないので、項目分析に使用することは少ない」とある。しかし、本稿で相関係数 (点双列相関係数) を算出しているように、得点と項目正答との相関を考慮することはよく行われている。したがって、分散自体の情報を直接的に利用しているわけではないが、分散の情報を間接的に利用しているとも考えられる。
- 3) ちなみに、このテストについてのクロンバックの  $\alpha$  係数 (熊谷・荘島 (2015) など、多くの文献に記載がある) を求めたところ、0.89 となった。十分な信頼性があると言える。

参考文献

千野直仁 (1999). 「入試形態、就職先と入学後の成績との関連について」柳井晴夫・前川眞一編『大学入試データの解析：理論と応用』現代数学社, 188-201.

半田智久 (2012). 『GPA制度の研究：functional GPAに向けて：the functional grade point average system』大学教育出版.

肥田野直 (編) (1972). 『テスト I』東京大学出版会. シリーズ「心理学研究法」No.7.

林篤裕・副島雄児・田尾周一郎 (2012). 「21世紀プログラムの10年」『大学入試研究ジャーナル』, 22, 155-161.

林寛子 (2011). 「新たな入学者追跡調査における選抜方法評価」『大学入試研究ジャーナル』, 21, 159-164.

林寛子 (2015). 「入学後の成功と資質・能力自己評価にみる入試の評価——山口大学入学者追跡調査データ分析より」『大学入試研究ジャーナル』, 25, 151-156.

池田央 (編) (1972). 『テスト II』東京大学出版会. シリーズ「心理学研究法」No.8.

池田央 (編) (1992). 『テストの科学：試験にかかわるすべての人に』日本文化科学社. (2007年: 教育測定研究所.)

池田央 (編) (1994). 『現代テスト理論』朝倉書店. 行動計量学シリーズ.

石井秀宗 (2014). 「推薦及び一般入試の受験者層の推移に関する検討」『大学入試研究ジャーナル』, 24, 35-40.

北川博之 (2014). 『データベースシステム』オーム社.

熊谷龍一・荘島宏二郎 (2015). 『教育心理学のための統計学——テストでココロをはかる』誠信書房. 心理学のための統計学シリーズ No.4.

倉元直樹・泉毅 (2014). 「東北大学工学部AO入試受験者にみる大学入試広報——その意義と発信型、対面型広報の効果」『日本テスト学会誌』, 10(1), 125-146.

倉元直樹・大津起夫 (2011). 「追跡調査に基づく東北大学AO入試の評価」『大学入試研究ジャーナル』, 21, 39-48.

Linn, Robert L. (ed.) (1989). *Educational Measurement Third Edition*, Macmillan Publishing Company (池田央ら訳 (1992). 『教育測定学 原著第3版 (上・下)』C.S.L.学習評価研究所, みくに出版 (発売).

光永悠彦 (2017). 『テストは何を測るのか：項目反応理論の考え方』ナカニシヤ出版.

宮原英夫・浜崎道子・清水和彦・白鷹増男 (1999). 「入試時の成績と在学中の成績との関連」柳井晴夫・前川眞一編『大学入試データの解析：理論と応用』現代数学社, 178-187.

村木英治 (2011). 『項目反応理論』朝倉書店. シリーズ「行動計量の科学」No.8.

西郡大 (2011). 「個別大学の追跡調査に関するレビュー研究」『大学入試研究ジャーナル』, 21, 31-38.

大久保貢・金澤悠介・倉元直樹 (2012). 「AO入試入学生の追跡調査——福井大学工学部の事例」『大学入試研究ジャーナル』, 22, 145-153.

大竹洋平 (2019). 「入試広報活動の戦略策定に関する論点整理——入試データベースの利活用および大学組織における連携を題材として」『大学入試研究ジャーナル』, 29, 287-294.

大竹洋平 (2020). 「入学志願者動向を知るための追跡データベース整備についての基礎的な考察——入試広報活動から入学後の学習活動までを接続する追跡データベースを目指して」『山梨英和大学紀要』, 18, 29-39.

芝祐順 (編) (1991). 『項目反応理論：基礎と応用』東京大学出版会.

芝祐順・渡部洋 (1988). 『入試データの解析』新曜社.

白川友紀・島田康行・渡邊公夫ほか (2004). 「筑波大学AC入学者の追跡調査——平成12年度入学者の3年目と14年度入学者」『大学入試研究ジャーナル』, 14, 65-71.

豊田秀樹 (2012). 『項目反応理論[入門編] (第2版)』朝倉書店. 統計ライブラリー.

山田美都雄・西本裕輝 (2014). 「追跡データを用いた大学生の成績推移の分析」『大学入試研究ジャーナル』, 24, 29-34.

山本以和子・内村浩 (2011). 「AO入試入学者の学習活動追跡による傾向分析」『大学入試研究ジャーナル』, 21, 119-123.

吉村宰・木村拓也 (2011). 「志願・入試・学務データに見られる入学者選抜方法の特徴」『大学入試研究ジャーナル』, 21, 165-170.