

## 試験問題の分析支援システムの提案

### Proposal on an Analysis Support System for Exam Questions

林 貴史<sup>\*1</sup>, 高木 正則<sup>\*1</sup>

Takafumi HAYASHI<sup>\*1</sup>, Masanori TAKAGI<sup>\*1</sup>

<sup>\*1</sup> 岩手県立大学ソフトウェア情報学部

<sup>\*1</sup> Faculty of Software and Information Science, Iwate Prefectural University

Email: g031j126@s.iwate-pu.ac.jp

あらまし：近年、多種多様な資格試験や検定が行われているが、試験の解答データに基づいて出題した問題を評価することは次回の試験問題を作成する際に有効である。しかし、試験問題の分析にはテスト理論や統計などの専門的知識が必要となるため、十分な試験分析をしきれていない。本研究では、試験問題の分析負担軽減を目的とし、試験問題の分析結果や次回の作問アドバイスを提示する試験問題の分析支援システムを提案する。

キーワード：テスト理論、作問支援システム、多肢選択形式問題、試験作成、ご当地検定

#### 1. はじめに

多くの教育現場では、学習者の理解度や能力を測定するために様々な試験が実施されている。それらの試験では、試験の得点から学習者を評価することはするが、出題した試験問題の妥当性や信頼性を評価することは少ない。その要因として、試験問題の分析に必要なテスト理論や統計などの専門的知識が不足しているため、十分な試験分析をしきれていないことが考えられる。そこで、我々は専門知識を持たない試験作成者の試験問題の分析負担軽減を目的とし、試験問題の分析結果や次回作問時のアドバイスを生成する試験問題の分析支援システムを提案する。本稿では、岩手県盛岡市で開催されているご当地検定試験<sup>(1)</sup>「盛岡もの識り検定試験」(以下、もりけん)の検定試験の被験者データを用いて、問題分析の例を示す。

#### 2. 提案内容

提案システムの概要図を図1に示す。図1①の分析では、古典的テスト理論や項目反応理論などを駆使して項目難易度や項目弁別力、さらには信頼性に関する分析などを行う。また、項目分析から明らかになった項目特性を考慮し、次回作問の際に参考になる作問アドバイスを生成する。

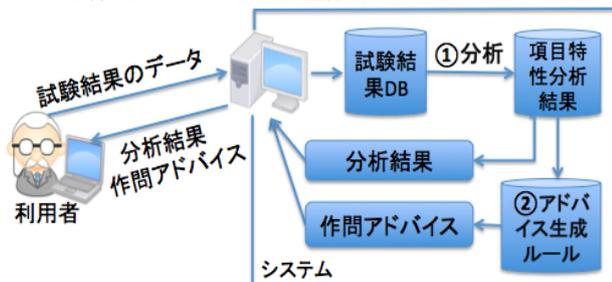


図1 システム概要図

#### 3. 手動による試験問題の分析

本稿では、試験問題を手動で分析した結果から、項目分析の具体的な手法やアドバイスの生成ルールを検討する。分析したのはもりけんの2012年3級で出題された問題100問である。受験者数は、全体で120名であった。分析手法としては、試験の信頼性の指標としてクロンバックの $\alpha$ 係数、折半法、再テスト法、平行テスト法、など数多くの手法があるが、その中でも、一般的に広く用いられているクロンバックの $\alpha$ 係数を用いた。また、項目難易度と項目弁別力を古典的項目分析手法で算出した。表1に各項目の評価基準を示す。

##### 3.1 クロンバックの $\alpha$ 係数

クロンバックの $\alpha$ 係数はテスト得点の信頼性係数の値を内的整合性の観点から推定する公式である<sup>(2)</sup>。計算式は次の式で表される。

$$\alpha = \left\{ \frac{\text{項目数}}{(\text{項目数}) - 1} \right\} \times \left\{ 1 - \frac{(\text{各項目の分散の合計})}{\text{合計点の分散}} \right\}$$

##### 3.2 項目難易度

項目難易度は、正答した受験者の全体受験者に対する割合で、正答率と言い換えることもできる。値の範囲は0から1までで、0に近いほどその項目が難しいことを意味し、1に近いほど項目が易いことを意味する。当て推量による正解を考慮すると、適切な項目難易度は0.625である<sup>(3)</sup>。古典的テスト理論や項目反応理論を用いて算出することができる。

##### 3.3 項目弁別力

項目弁別力は各項目がどれだけ成績上位者と下位者を分別できるかを表す指標である。得点分布の形状や信頼性に影響を与える。項目弁別力の値はプラスの値が大きいほど弁別力が高く、0から1までの値をとる。古典的テスト理論や項目反応理論を用いて算出することができる。以下は古典的テスト理論を用いた算出式である。

$$\left( \frac{\text{成績上位群} / \text{全体}}{\text{成績下位群} / \text{全体}} \right) - \left( \frac{\text{成績上位群} \cdot \text{下位群の人数}}{\text{合計点でそれぞれ}} \right)$$

上位・下位 3 分の 1 あるいは 27% で求められる。  
表 1 項目の分析

| 指標  | 手法                        | 基準   |
|-----|---------------------------|--|
| 信頼性 | 再テスト法                     | 0.80 以上であれば、信頼性が高い <sup>(4)</sup>                          |
|     | 平行テスト法                    |  |
|     | 折半法                       |  |
| 難易度 | 内部一貫法 (クロンバックの $\alpha$ ) | 0.72 以上であれば、信頼性が高い <sup>(3)</sup>                          |
|     | 古典的テスト理論<br>項目反応理論        | .39 以下かつ .81 以上の項目は難易度が不適切、.29 以下かつ .91 以上の項目は難易度が極端に高い・低い |
| 弁別力 | 古典的テスト理論                  | .29 以下の項目は弁別力が低く不十分、.19 以下の項目は弁別力が極端に低い                    |
|     | 項目反応理論                    |  |

#### 4. 分析結果

クロンバックの  $\alpha$  係数を求めた結果、0.911 であり、信頼性が高いテストであったことが分かった。古典的項目分析の項目難易度、項目弁別力で分析した結果の一部を表 2 に示す。なお、表 2 中の項目番号は実際のテストで出題された項目番号と異なる番号を付与している。表 2 の難易度、弁別力で下線が引かれている項目は表 1 の評価基準で「不適切」または「不十分」と判定された項目である。また、二重下線が引かれている項目は「極端に高い・低い」と判定された項目である。下線や二重下線が引かれている項目に関しては、項目の削除や修正が必要であることが示唆される。

表 2 項目分析結果

| 項目 | 難易度         | 弁別力         | 項目 | 難易度         | 弁別力         |
|----|-------------|-------------|----|-------------|-------------|
| 1  | <u>0.92</u> | <u>0.19</u> | 26 | <u>0.98</u> | <u>0.09</u> |
| 2  | <u>0.95</u> | <u>0.03</u> | 27 | <u>0.96</u> | <u>0.13</u> |
| 3  | <u>0.94</u> | <u>0.09</u> | 28 | <u>0.98</u> | <u>0.03</u> |
| 4  | <u>0.96</u> | <u>0.16</u> | 29 | <u>0.91</u> | <u>0.19</u> |
| 5  | <u>0.96</u> | <u>0.13</u> | 30 | 0.88        | 0.28        |
| 6  | <u>0.97</u> | <u>0.13</u> | 31 | 0.83        | <u>0.09</u> |
| 7  | <u>0.93</u> | <u>0.19</u> | 32 | <u>0.81</u> | <u>0.28</u> |
| 8  | <u>0.95</u> | <u>0.16</u> | 33 | 0.93        | <u>0.22</u> |
| 9  | 0.77        | 0.41        | 34 | <u>0.87</u> | 0.31        |
| 10 | 0.77        | 0.28        | 35 | <u>0.91</u> | 0.28        |
| 11 | 0.71        | <u>0.19</u> | 36 | 0.80        | <u>0.25</u> |
| 12 | <u>0.87</u> | <u>0.09</u> | 37 | 0.79        | 0.50        |
| 13 | 0.83        | 0.31        | 38 | 0.63        | 0.31        |
| 14 | 0.62        | 0.56        | 39 | <u>0.93</u> | 0.25        |
| 15 | <u>0.88</u> | <u>0.28</u> | 40 | <u>0.89</u> | <u>0.16</u> |
| 16 | 0.69        | 0.44        | 41 | 0.70        | <u>0.22</u> |
| 17 | 0.64        | 0.34        | 42 | <u>0.36</u> | 0.31        |
| 18 | 0.70        | 0.66        | 43 | 0.54        | 0.63        |
| 19 | 0.42        | 0.56        | 44 | 0.53        | 0.41        |
| 20 | <u>0.81</u> | 0.38        | 45 | 0.53        | 0.63        |
| 21 | 0.58        | 0.34        | 46 | 0.73        | 0.31        |
| 22 | 0.57        | 0.38        | 47 | <u>0.35</u> | <u>0.44</u> |
| 23 | 0.80        | 0.38        | 48 | 0.65        | <u>0.16</u> |
| 24 | 0.66        | 0.44        | 49 | 0.78        | 0.34        |

|    |             |             |    |      |      |
|----|-------------|-------------|----|------|------|
| 25 | <u>0.81</u> | <u>0.25</u> | 50 | 0.79 | 0.50 |
|----|-------------|-------------|----|------|------|

#### 5. 作問アドバイスの検討

4 章の分析結果から、作問アドバイスを検討した。項目難易度と項目弁別力の値により、表 3 のような作問アドバイスを生成することができる。と考える。

表 3 作問アドバイス生成ルール

| 条件   | 作問アドバイス例   |
|--|--|
| $\alpha < 0.72$  | テスト全体の信頼性が低く、良いテストではない可能性があります。また、項目 n が信頼性を下げている可能性があります。             |
| 項目難易度 $< 0.40$<br>または<br>$0.80 < \text{項目難易度}$                   | 項目 n は難しいまたは簡単な問題の為、悪い問題の可能性があり、見直す必要があります。                            |
| $0.30 < \text{項目弁別力}$  | 項目 n は成績上位と下位に差が出ており、成績上位の人が比較的解ける良い問題だといえます。                          |
| $0.39 < \text{項目難易度}$<br>かつ<br>$\text{項目弁別力} < 0.30$             | 項目 n は標準的な問題であり、成績上位と下位に差がない問題の為、差をつけるのであれば、少し問題を難しくすると良いでしょう。         |
| 項目難易度 $< 0.40$<br>かつ<br>$0.30 < \text{項目弁別力}$                    | 項目 n は難しい問題であり、成績上位と下位に差がでている為、上位者が比較的解ける問題として下位者と差を付けたい時に出題すると良いでしょう。 |
| $0.39 < \text{項目難易度}$<br>$< 0.81$<br>かつ<br>$\text{項目弁別力} < 0.30$ | 項目 n はかなり易しい問題であり、成績に関係がなく解ける問題の為、テスト序盤に使う問題としては適しています。                |

#### 6. まとめと今後の検討課題

本稿では、試験問題の分析支援システムを提案した。また、項目分析の方法や作問アドバイスの生成機能を検討するために、もりけんの試験結果を手動で分析した。分析の結果から、項目難易度と項目弁別力の結果から次回作問時のアドバイスを生成できる可能性が示唆された。今後は、被験者集団に依存しない値を算出できる項目応答理論を用いて計算した場合についても検討する。また、試験分析支援システムを開発し、作問負担軽減につながる作問アドバイスを生成できるシステムを開発する。

#### 参考文献

- (1) 盛岡商工会議所, 盛岡もの識り検定,  
<http://www.ccimorioka.or.jp/jinzai/moriken.html>
- (2) 大友賢二:『言語テスト・データの新しい分析法 項目応答理論入門』, 大修館書店, 1996
- (3) Nunnally, JumC. Psychometric Theory 2ndEdition. New York:McGraw-Hill, 1978.
- (4) 山森光陽, 前田啓朗(編)『英語教師のための教育データ分析入門: 授業が変わるテスト・評価・研究』(pp.4-12), 東京:大修館, 2004.
- (5) 豊田秀樹:『項目反応理論[入門編]-テストと測定の科学-』, 朝倉書店, 2002.