

現代テスト理論に基づく小テスト問題の分析

和田 武*1

Email: wada@cite.ehime-u.ac.jp

*1: 愛媛大学総合情報メディアセンター

◎Key Words 現代テスト理論, 小テスト問題, Moodle

1. はじめに

近年, Moodle (Course Management System) の多肢選択方式等を用いた小テストによる成績評価が数多く実施されているが, 作成した問題の内容によっては, 正解率に大きな差が生じて正当な評価ができない場合がある。そこで, 受験者の学習評価や指導内容と学習達成度の整合性などを分析し, 試験問題の難易度や受験者の能力値といった特性を定量化する項目反応理論やラッシュ測定理論などの現代テスト理論^(1, 2)による研究が行われている。

本稿では, 講義で利用した Moodle の小テストに対して, 現代テスト理論である項目反応理論や S-P 表理論⁽³⁾を用いた分析を試み, 試験問題の妥当性を検証し, 今後の問題作成の参考とするものである。

2. 方法と結果

2.1 項目反応理論

項目反応理論は, 教育分野で登場した理論で, 試験問題に依存しなくて受験者の能力を適切に測定でき, 留学のための語学試験や企業の入社試験などでも利用されている。また, 試験問題が適切に出題されたかなど適正レベルを把握することも可能である。

表 1 に使用データを示す。2012 年 3 月に Moodle で実施した文系学科の期末試験結果で, 受験者は 24 名, 問題数は 5 ブロック 50 問の選択問題(A~D の選択肢), 受験時間は 60 分である。1 行目は正答, 2 行目以降は個々の解答データで, 行番号, 受験者番号, そして 50 問の解答データで構成される。

解析には, TDAP⁽⁴⁾ (Test Data Analysis Program) を使用した。TDAP は, 選択肢問題で構成される小テストの結果データを分析するためのフリーソフトで, 古典テスト理論のみならず現代テスト理論の項目困難度や受験者能力値なども求めることができる。まず, 項目困難度を求める。項目困難度は-3 から+3 までの値をとり, 0 は平均的な困難度, 正の値は平均より難しい, 負の値は平均より易しい度合いを示す。ブロック(1)~(5)では, それぞれ-1.710~1.639, -1.710~1.252, -2.509~2.996, -1.710~1.849, -2.509~2.334 の項目困難度となり, 50 項目の問題は項目困難度が正の値と負の値が同程度となり, 項目困難度に関して問題は認められなかった。

表 1 使用データ

表 2 受験者能力値

Expansion factor; X = 1.292		
Person No.	Person Measure	Standard Error
10001	0.776	0.343
10002	0.660	0.339
10003	1.725	0.404
10004	0.325	0.331
10005	1.568	0.390
10006	0.660	0.339
10007	1.568	0.390
10008	0.896	0.348
10009	1.019	0.354
10010	1.280	0.369
10011	0.000	0.328
10012	0.216	0.329
10013	1.147	0.361
10014	0.216	0.329
10015	2.080	0.440
10016	0.660	0.339
10017	1.147	0.361
10018	-0.660	0.339
10019	0.435	0.333
10020	0.776	0.343
10021	2.284	0.465
10022	0.216	0.329
10023	1.147	0.361
10024	-0.108	0.328

表2に項目困難度の出力結果(受験者能力値)を示す。「0」は平均的な能力をもつ受験者、負の値は平均より低く、正の値は平均より高い能力を持つ受験者を示す。受験者10011は平均的な受験者で、受験者10018, 10024は平均より低く、受験者10021, 10015は高い能力を有する受験者である。

次に出力されるモデルの適合度分析結果では、ブロック(4)の間8(t検定結果の解釈を問う問題)はミスフィットと判断され、モデルと実際の解答とのずれが想定外であることを意味する。この場合、能力の高い受験者の誤答が多く、能力の低い受験者の正答が多い問題の可能性があり、出題時に再検討する必要がある。また、ブロック(1)の間9(区間推定に関する問題)、ブロック(5)の間47(回帰直線に関する問題)はオーバーフィットとなり、モデルと実際の解答のずれが小さすぎることを意味し、能力が高い受験者が全て正答し、低い受験者が全て誤答する場合などであり、今後の問題作成時に検討を要する。

2.2 S-P表分析

S-P表分析⁽³⁾(Student-Problem score table analysis)とは、試験問題の結果から、受験者の学習内容や教員の指導法の診断など評価情報を得るために作られた分析方法で、成績順位や得点結果からは見ることが出来ない受験者個々の育成ポイント、指導の改善ポイントなどを教員が判断することができる。

表3 S-P表分析

macro	D _t = 0.446	1	2	5	3	4	6	%	Blank	9	C.S
3	10003	1	1	1	1	1	5	100	0	0	0
5	10005	1	1	1	1	1	5	100	0	0	0
7	10007	1	1	1	1	1	5	100	0	0	0
15	10015	1	1	1	1	1	5	100	0	0	0
21	10021	1	1	1	1	1	5	100	0	0	0
10	10010	1	1	1	1	0	4	80	0	0	0
13	10013	1	1	1	0	0	3	60	0	0	0
17	10017	0	1	1	1	0	3	60	0	0.67	0
20	10020	1	0	1	1	0	3	60	0	0.67	0
23	10023	1	1	0	0	1	3	60	0	0	1
1	10001	1	1	0	0	0	2	40	0	0	0
2	10002	1	1	0	0	0	2	40	0	0	0
4	10004	1	1	0	0	0	2	40	0	0	0
6	10006	1	0	1	0	0	2	40	0	0	0
8	10008	0	0	1	0	1	2	40	0	1.5	0
9	10009	1	0	1	0	0	2	40	0	0	0
12	10012	0	1	1	0	0	2	40	0	0	0
11	10011	0	0	1	0	0	1	20	0	0	0
14	10014	0	1	0	0	0	1	20	0	0	0
16	10016	0	0	0	1	0	1	20	0	0	2
19	10019	0	0	0	1	0	1	20	0	0	2
22	10022	0	0	0	0	1	1	20	0	0	3
18	10018	0	0	0	0	0	0	0	0	0	0
24	10024	0	0	0	0	0	0	0	0	0	0
		14	14	14	10	8	60				
%		41.2	41.2	41.2	29.4	23.5	35.3				
cp		0.04	0.08	0.08	0.17	0.19					

S-P表作成には、高知県庁ホームページで公開されているS-P表作成ワークシートを用いた。表3は、表1の使用データを5ブロックに分け、10点満点のデータを基に平均点以上を「1」、平均点以下を「0」としてS-P表を作成した。

S曲線(実線・青色)は、得点の度数分布を表し、曲線の左側に「1」、右側に「0」が並ぶのが理想型で、受験者の達成状況や学習状況が把握される。このS曲線左上の「0」(誤答)は、成績の良い受験者が簡単な問題を間違えているので単なる間違いであり、右下の「0」は内容が理解されていないことに原因があると思われる。また、受験者10008は左側に正答がなく、右側に正答があるので、易しい問題ができず、難しい問題しかできていないので、欠席など学習上での問題の可能性が考えられる。

P曲線(点線・赤色)は、正答数の度数分布を表し、曲線の上側に「1」、下側に「0」となるのが理想型で、個々の問題の適切さ、指導の効果などを読み取ることができる。ブロック(1)~(5)の問題は上側の正答数が下側の正答数を上まわっているため、得点が高い受験者は正答し、低い受験者は誤答しているため、問題としては適切と考えられる。

S曲線とP曲線は一般に接近しているが、若干離れているのは、受験者の学習が不十分で学力のばらつきが大きいか、出題問題としての不適切さが考えられる。さらに、曲線とS-P曲線の乖離は0.446と小さくなく、標準的な試験問題といえよう。

なお、表中のCPは問題注意係数を示し、0.5以上の受験者が7名、CSは受験者注意係数を示し、この値が0.2以下なので試験問題として適切だと判断される。

3. おわりに

講義で利用したMoodleの小テストに対して、現代テスト理論である項目反応理論やS-P表理論を用いた分析を試み、試験問題の妥当性を検証した。

項目反応理論(項目困難度)では、小テスト問題50問はバランスよく出題されていたと判定されたが、受験者能力値については、ブロック(4)の間8がミスフィット、ブロック(1)の間9、ブロック(5)の間47がオーバーフィットと判定され、今後見直す必要がある。

S-P表分析による方法では、受験者によっては難しい問題しかできず易しい問題ができないなどの結果が出たが、これは欠席などの学習上の問題が考えられる。

今回のS-P表分析に使用したデータは、5つのブロック単位で分析したので、今後は50問の問題ごとに分析する必要もある。また、受験者数を増やし、古典テスト理論による評価を含めて、さらに試験問題の妥当性を検証していきたい。

参考文献

- (1) 中村洋一著、テストで言語能力は測れるか、桐原書店、2004
- (2) 竹内俊彦、項目反応理論入門、青山学院大学総合研究所、2006.
- (3) S-P表の入門、佐藤隆博、明治書店。