CMS 小テスト問題分析による授業改善の試み

The attempt for the improvement of the teaching approaches based on Moodle quizzes analysis

和田　武 †
Takeshi Wada †

wada@cite.ehime-u.ac.jp

† 愛媛大学総合情報メディアセンター

† Center Information Technology, Ehime University

概要

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，作成した設問の内容によっては，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．Moodleでは，小テストの結果に対する分析としてアイテム分析機能が備わっており，テストの信頼性の指標を表す識別指数や判別指数を求めることができる．さらにより詳細な分析を行うためにTDAP [1]を用いた項目反応理論に基づく分析[2]など数多くの研究が行われている．

本稿では，今後の設問作成の参考とするために，講義で利用したMoodleの小テストに対して，これらの手法を用い，平均点や標準偏差などの基本統計量を求めて分析する古典的テスト理論および現代テスト理論である項目反応理論やS-P表理論[3]を用いた分析を行い，試験問題の妥当性を検証したので報告する．

キーワード

テスト理論，小テスト，Moodle

1. はじめに

情報系センターは, ネットワーク管理, 学術研究支援, 地域貢献などの情報基盤の統括に加え，情報教育の中核としての役割がある．最近では，センターシステムの教育用システムに Moodle（CMS: Course Management System）を組み込み，情報系センターの教員等が情報基礎教育を担当している大学が少なくない. CMS を用いた授業では, 教材の提示以外に各回ごとの確認問題や期末試験を Moodle の多肢選択方式等を用いた小テストによる成績評価が実施されているが，作成した設問の内容によっては，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従来からの古典的テスト理論を用いた研究や，項目反応理論やラッシュ測定理論などを用いた現代テスト理論による研究が行われている．

本稿では，今後の設問作成の参考とするために，講義で利用した Moodle の小テストに対して，Moodle のアイテム分析や TDAP による分析および 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のでここに報

告する.

2. 方法と結果

Moodle の小テスト問題を検証するために，Moodle のアイテム分析を用いる方法，TDAP を用いる方法，および S-P 表による分析を行った.

2.1 Moodle のアイテム分析

Moodle の小テストにはアイテム分析機能が備わっており，(1)ファシリティ指標，(2)識別指数，(3)判別係数などの統計データが古典的テスト理論によって算出される.

(1)ファシリティ指標は，項目容易度とも言い，小テストの受験者にとって設問がどれだけ難しいか簡単かを示す指標であり，今回使用したデータに対しては，全て多肢選択方式の設問で構成されていたので，正しく答えた受験者の割合を示す．(2)識別指数は，弁別力ともいい，受験者間の特性の違いをどの程度正確に捉えられるかを示す指標で，優秀な受験者ほど正答しやすく，そうでない受験者ほど誤答しやすい傾向が顕著な場合にこの項目の指数が高くなる．(3)判別係数は，小テスト全体と設問の点数との相関係数で，この係数が負の設問は，優秀な受験者が間違って回答したことを意味し，優秀な受験生に対してペナルティになってしまうので，このような設問は避けるべきである．判別係数は受験者全体の情報を利用するので，上位中位下位グループ分けの情報を利用する識別指数に比べてより詳細に分析できる.

開始日時	受験完了	所要時間	評点/10	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
2012年02月2日13:05	2012年02月2日13:19	13分44秒	8	1/1	1/1	1/1	1/1	0/1	1/1	1/1	1/1	0/1	1/1
2012年02月2日12:57	2012年02月2日13:02	5分25秒	8	1/1	1/1	1/1	1/1	0/1	1/1	1/1	1/1	0/1	1/1
2012年02月2日12:59	2012年02月2日13:04	5分10秒	9	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	0/1	1/1
2012年02月2日13:00	2012年02月2日13:42	41分54秒	7	0/1	1/1	1/1	0/1	1/1	0/1	1/1	1/1	1/1	1/1
2012年02月2日12:55	2012年02月2日13:01	6分4秒	7	1/1	1/1	1/1	0/1	0/1	1/1	1/1	1/1	0/1	1/1
2012年02月2日13:02	2012年02月2日13:12	10分	6	0/1	1/1	1/1	1/1	0/1	0/1	1/1	0/1	1/1	0/1
2012年02月2日12:56	2012年02月2日13:02	5分48秒	8	1/1	1/1	1/1	0/1	1/1	1/1	1/1	0/1	1/1	1/1

図１．Moodel のアイテム分析

2.1.1 使用データ

今回，Moodle のアイテム分析で使用したデータは，2012 年 2 月に実施した文系学科の担当科目 24 名の期末試験結果の一部である．（図１参照）

表１に今回使用した小テストの出題項目を示す．50 個の設問がブロック(1)から(5)に分かれて出題されている.

表１．　小テストの出題項目

2011年度 法文 試験項目

	ブロック(1)	ブロック(2)	ブロック(3)	ブロック(4)	ブロック(5)
問1	区間推定(2)	ばらつき	クロス集計(1)	クロス集計(4)	ヒストグラム(3)
問2	基本統計量(1)	回帰直線(1)	区間推定(1)	クロス集計(2)	レーダーチャート
問3	基本統計量(2)	回帰直線(2)	仮説の検定(1)	差がある	回帰直線(3)
問4	外れ値	データの種類	仮説の検定(2)	分散分析(3)	基本統計量(3)
問5	分散分析(1)	正規分布(1)	分散分析(2)	危険率	時系列分析(2)
問6	度数分布とヒストグラム(1)	相関係数(1)	指数平滑化法	中心極限定理	正規分布(3)
問7	度数分布とヒストグラム(2)	相関係数(2)	正規分布(2)	帰無仮説	相関係数(3)
問8	時系列データ(1)	確率分布(1)	分布(3)	母集団と標本	相関係数の有意性検定(4)
問9	棒グラフ	確率分布(2)	t-検定(1)	t-検定(3)	相対参照と絶対参照
問10	Paird t検定	移動平均	クロス集計(3)	スマホのセキュリティ	乱数

(136308) 正答	確率分布(2): 確率分布の記述の中で、誤っているものを選んでください。	確率変数がどのような値になるかを示す分布である。	(0.00)	6/24	(25%)	58%	0.504	-0.20	-0.00
		正規分布、カイ2乗分布、t分布、F分布などが該当する。	(0.00)	2/24	(8%)				
		標準正規分布は、平均が1、標準偏差が0の理論的に設定された正規分布である。	(1.00)	14/24	(58%)				
		正規分布は、連続分布の中で最もよく現れる代表的な分布である。	(0.00)	2/24	(8%)				

図２．　Moodle の設問とアイテム分析結果

2.1.2 アイテム分析結果

図２は，表１のブロック(2)-9（確率分布(2)）の設問とアイテム分析結果を示す. 右欄の７個の係数はそれぞれ，部分点（ここでは選択肢の３番目が部分点１となっているので正解を表す），解答数（24 名中 14 名が正解），解答%，ファシリティ指標（58%），標準偏差（0.504），識別指数（-0.20），そして判別係数（-0.00）を示す．識別指数が負値となった設問は 50 問中図２以外にブロック(2)-問１（バラつきに関する問題）があった．残りの 48 問は識別係数が正の値となり問題は見受けられなかった.

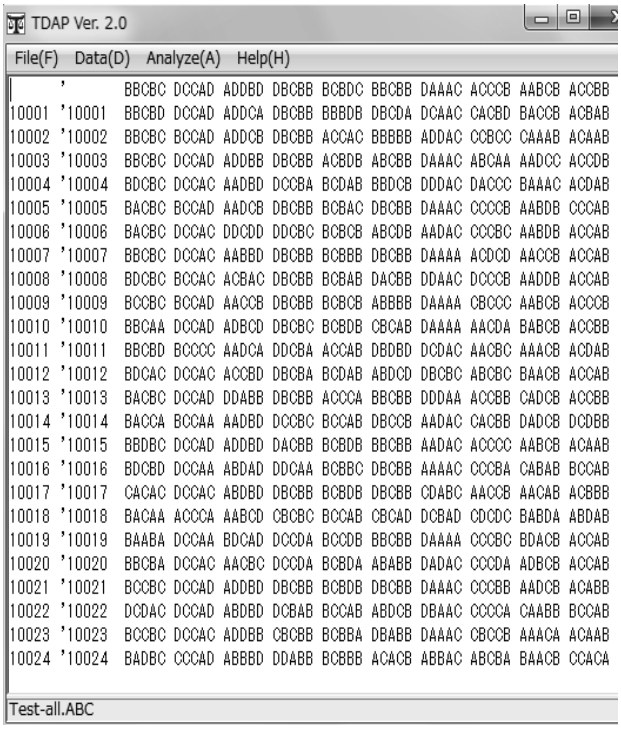
判別係数が負の値をとるのは，成績が良くない学生が良い学生よりも正解数が多いことを示し，このような設問は小テスト問題全体の精度を下げるといわれているので，今後除外すべきである．判別係数が負の値となるのは，図２以外にはブロック(3)-問４（仮説の検定に関する設問②　-0.09），ブロック(5)-問１（ヒストグラムに関する設問（-0.07））の２つの設問であったが，わずかにゼロを下回っていた. 残り 47 問の設問については問題は見受けられなかった.

2.2 TDAP による分析

TDAP(Test Data Analysis Program）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[1]で，基本統計量などの古典テスト理論のみならず現代テスト理論の項目分析や項目応答理論の指標なども求めることができる．

表２に TDAP で使用したデータを示す．これは「2.1 Moodle のアイテム分析」で利用したデータを TDAP で扱えるようにしたものである．A～D は各設問の選択肢を表す．1 行目は正答，2 行目以降は個々の解答データで，左から行番号，受験者番号，そして 50 問の解答データで構成される．

表２． 使用データ



2.2.1 古典的テスト理論

古典的テスト理論によるテストデータの分析には，(1)基本統計量から受験者グループを分析する方法，(2)項目分析により分析する方法などがある．

(1) 基本統計量

従来の基本統計量（グループの平均値，分散，偏差値など）から受験者グループの特性を分析する手法で，受験者グループの大まかな特性を把握するのに便利な分析手法である．これは一般的にも説得力がある方法と言われている．図３は，TDAP より基本統計量を求めた結果の一部である．24 名の受験者の平均点は 50 点満点の 33 点で，最高点が 43 点，最低点が 20 点であり，尖度(Kurtosis)が負の値を示しているので正規分布よりも平坦で分布の裾野が広く，歪度(Skewness)も負の値となっているので左に裾野が広く（右に偏った分布）であることがわかり，比較的今回の試験問題はよくできたといえる．

図３の後半には，標準得点(Standard Scores)が示されている．Z-s(Z Score)は，受験者各自の得点と平均点の差を標準偏差で割った値で，偏差値とも言われており，各受験者の得点が平均点からどのくらい離れているかを示すもので，負の値は平均点よりも低く，正の値は平均点よりも高い度合いを示しているので受験者ごとに成績がわかる．ここで 5s は 5 段階，9s は 9 段階の評定を表している．

```
*** BASIC STATISTICS ***

Name of *.CUT file = Test-all.CUT

Number of examinees ------------    24
Sum of the raw scores ----------   793
Minimum score ------------------    20
Maximum score ------------------    43
Median -------------------------    33
Range --------------------------    23
Mean ---------------------------    33.042
Variance -----------------------    30.623
Standard deviation -------------     5.534
Skewness -----------------------    -0.235
Kurtosis -----------------------    -0.357


*** STANDARD SCORES ***

   * NOTES *
     ID = Identification Number of examinee
     R-s = Raw score,  Z-s = Z-score,  T-s = T-score
     5s = Five point scale,  9s = Stanine scale

     ID    R-s     Z-s      T-s     5s   9s
==========================================
    10001    33   -0.008    49.925    3    5
    10002    32   -0.188    48.118    3    5
    10003    40    1.257    62.574    4    7
    10004    29   -0.730    42.696    2    4
    10005    39    1.077    60.767    4    7
    10006    32   -0.188    48.118    3    5
    10007    39    1.077    60.767    4    7
    10008    34    0.173    51.732    3    5
    10009    35    0.354    53.539    3    6
    10010    37    0.715    57.153    4    6
    10011    26   -1.272    37.275    2    3
    10012    28   -0.911    40.889    2    3
    10013    36    0.535    55.346    4    6
    10014    28   -0.911    40.889    2    3
    10015    42    1.619    66.188    5    8
```

図３．古典的テスト理論（基本統計量）

(2) 項目分析

図４は，項目分析により求めた，①項目困難度(DIFF)，②項目弁別力指数(DISC)，③実質選択肢数(AENO)等の結果の一部である．

① 項目困難度（DIFF）

項目困難度は，テスト項目がどのくらい難しかったかを検討する指標で，正答率とも呼ばれている．0.000 から +1.000 の値をとり，1.000 に近いほど易しい設問で，0.000 に近いほど難しい設問だったと解釈できる．図４では，設問 7,8 が全問正解，0.5 以下が設問 2,10,12 をはじめ 11 問で全体的に易しかった設問といえる．

```
〈NOTES〉
 DIFF  = Item difficulty index (p)
 DISC  = Discrimination power index (r pbi)
 AENO  = Actual equivalent number of options (k)
 ADIF  = Appropriateness of difficulty
 ADIS  = Appropriateness of discrimination power index
 AAEN  = Appropriateness of actual equivalent number of options
 SADIF = Standard appropriateness of difficulty
 SADIS = Standard appropriateness of discrimination power index
 SAAEN = Standard appropriateness of actual equivalent number of options
 SATOT = Standard appropriateness total
 RANK  = Rank in SATOT order

*** ITEM ANALYSIS SUMMARY TABLE ***
```

NO.	DIFF	DISC	AENO	ADIF	ADIS	AAEN	SADIF	SADIS	SAAEN	SATOT
1	0.917	0.057	1.411	0.417	0.003	0.895	0.430	0.416	0.557	1.403
2	0.333	0.282	3.780	0.417	0.087	0.948	0.430	0.467	0.584	1.481
3	0.833	0.145	1.723	0.583	0.021	0.818	0.495	0.427	0.519	1.441
4	0.750	0.370	1.964	0.750	0.158	0.736	0.561	0.511	0.478	1.550
5	0.667	0.405	2.356	0.917	0.196	0.786	0.627	0.535	0.503	1.664
6	0.667	0.405	2.415	0.917	0.196	0.820	0.627	0.535	0.520	1.681
7	1.000	0.000	1.000	0.250	0.000	0.000	0.364	0.414	0.110	0.888
8	1.000	0.000	1.000	0.250	0.000	0.000	0.364	0.414	0.110	0.888
9	0.917	0.547	1.332	0.417	0.427	0.722	0.430	0.677	0.471	1.578
10	0.458	0.401	2.784	0.667	0.192	0.683	0.528	0.532	0.451	1.511
11	0.875	0.026	1.578	0.500	0.001	0.860	0.463	0.414	0.540	1.417
12	0.417	0.467	3.444	0.583	0.279	0.891	0.495	0.586	0.555	1.636
13	0.542	0.218	3.059	0.833	0.050	0.896	0.594	0.445	0.558	1.597
14	0.542	0.203	2.956	0.833	0.043	0.851	0.594	0.440	0.535	1.570
15	0.583	-0.207	2.930	0.917	0.045	0.912	0.627	0.442	0.566	1.634
16	0.917	0.275	1.332	0.417	0.082	0.722	0.430	0.464	0.471	1.365
17	0.583	0.388	2.922	0.917	0.177	0.908	0.627	0.523	0.564	1.714
18	0.917	0.356	1.411	0.417	0.146	0.895	0.430	0.504	0.557	1.490
19	0.833	0.185	1.761	0.583	0.036	0.861	0.495	0.436	0.540	1.471
20	0.583	0.541	2.611	0.917	0.414	0.761	0.627	0.669	0.490	1.786

図４．項目困難度，項目弁別力指数，実質選択肢数

② 項目弁別力指数（DISC）

優秀な受験者とそうでない受験者を弁別（識別）できたかどうか検討する指数で, -1.000 から+1.000 の値をとり, +1.000 に近くなるほど弁別力が高いと判断され, 能力の高い受験者よりも低い受験者のほうが多く正答した場合には負の値が算出される．図４では設問 15 が負の値を示しており，その他，設問 21, 設問 39 が負の値を示した.

③ 実質選択肢数（AENO）

多肢選択形式の設問で，用意した選択肢が偏りなく選択されているかどうかを検討する指標で, 0.000 から選択肢の数までの値（今回は４）の値をとる．図４では，実質選択肢数は，設問１から 50 までで，22 問が 2.0 以下となっており，４つの選択肢を用意したにもかかわらず，２つ分の働きしかしていなかった．今後，問題作成の際にはこの点に特に注意が必要である.

2.2.2 項目反応理論

項目反応理論は，項目応答理論とも言われ，テスト問題の特性（例えば，難しさや易しさなど）と受験者の能力との間にある確率的な関係を仮定すれば学力の絶対値が測れるという理論である．ここで項目反応理論について説明する．項目反応理論のモデルには、計算に使用するパラメータ数によって、主に３つのモデルがある。(1) 項目困難度パラメータ（item difficulty parameter=b）のみを想定する１パラメータモデル、(2) 項目困難度パラメータ(b)と項目弁別力パラメータ（item discrimination parameter=a）を含む２パラメータモデル、そして、(3) 当て推量パラネータ（guessing parameter）を含む３パラメータモデルである。TDAP では、以下のモデル式で表される(1)の １パラメータラッシュモデル（ロジスティックモデル）を使用している。このモデルは、パラメータの解釈が比較的簡単で、少ないサンプル（100～200）でも安定した推定が可能である。

$$P_j(\theta) = \frac{1}{1+e^{-(\theta - b_j)}}$$

$P_j(\theta)$ ：θ で表される能力を持つ受験者が、項目j に正答する確率

θ ：受験者能力パラメータ

b_j ：項目 j の項目困難度パラメータ

ここで$-(\theta - b_j)$は、問題の難しさが学力をどれだけ上回っているかを表し、一般的にb_j、θは、それぞれ-3～+3 の値をとり、$b_j = 3$であれば大変難しい設問で、例えば, $\theta = -1.4$であればあまり成績が良いとは言えない学生である。上のモデル式から、学力の値と難しさの値が一致している場合、$-(\theta - b_j)=0$ となり、$P_j = 0.5$ となるので、解ける確率は 50%となる。また、設問の難しさが学力を上回れば上回るほど、ギャップ x が大きくなると分母が大きくなるので、正解率P_jは小さくなるから、正解の確率は低くなる。

この項目反応理論は，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.

図５は，１パラメータラッシュモデル（PROX 法)を用いた項目反応理論によるテストデータ分析結果の最初の画面である．全設問の初期項目困難度を示している．実際のデータ解析には, 図６の最終項目困難度パラメータ, 図７のモデルとの適合度の検討に焦点をあてる.

図６-１は, 項目反応理論により求められた最終項目困難度（最終能力パラメータ）を表したものである．Item No（設問）7,8 は「All Correct Responses」となっており，この設問は全受験者が正解だったので，この設問では受験者の能力を推定することができないので除外されている．設問が易しすぎたようである．ここで，負の値は平均的な困難度より易しく，正の値は平均的な困難度よりも難しい度合いを示すもので，-3.000 から+3.000 の値を示す.

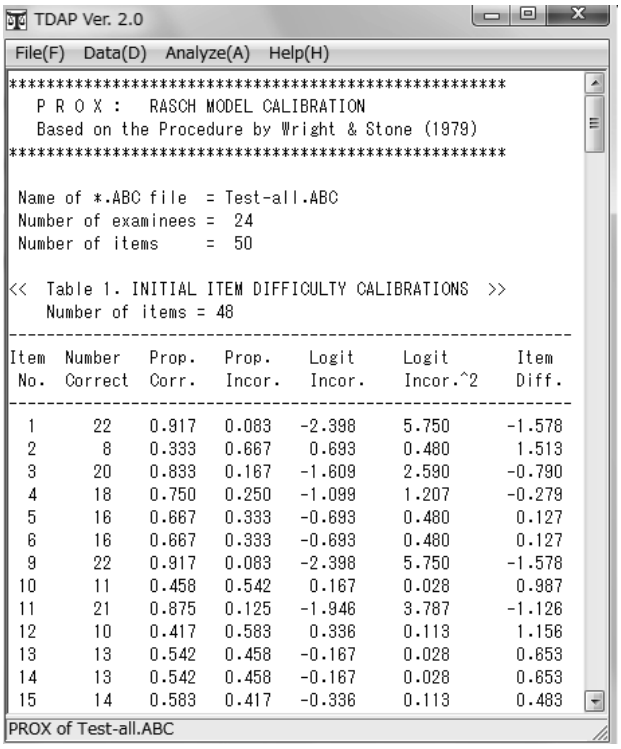


図５．項目反応理論

Expansion factor; Y = 1.083

Item No.	Final Calib.	Standard Error
7	All Correct Responses	
8	All Correct Responses	
1	-1.710	0.769
2	1.639	0.451
3	-0.856	0.570
4	-0.302	0.491
5	0.137	0.451
6	0.137	0.451
9	-1.710	0.769
10	1.069	0.426
11	-1.220	0.642
12	1.252	0.431
13	0.707	0.426
14	0.707	0.426
15	0.523	0.431
16	-1.710	0.769
17	0.523	0.431
18	-1.710	0.769
19	-0.856	0.570
20	0.523	0.431
21	-0.856	0.570
22	-2.509	1.063
23	0.335	0.439
24	1.639	0.451

図６-１．最終項目困難度（１）

図６-１の設問 1, 3, 4, 9, 11, 16, 18, 19, 21, 22 をはじめ計 21 問が負の値を示し，難しい設問だったことを示すが，残りの 29 問は正の値を示したので，設問としてはほぼ妥当だと判断できよう．

図６-２は，最終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 は平均より高い能力を有する受験者である．

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

図６-２．最終項目困難度（２）

<< Table 4. ANALYSIS OF FIT >>

Item No.	Sigma(Z^2)	t
1	26.429	0.488
2	24.618	0.235
3	31.932	1.215
4	20.906	-0.316
5	20.423	-0.391
6	21.382	-0.243
9	10.228	-2.316
10	22.111	-0.132
11	25.888	0.413
12	20.560	-0.370
13	24.684	0.244
14	24.892	0.274
15	32.327	1.265
16	22.222	-0.116
17	23.480	0.070
18	13.850	-1.535
19	20.768	-0.338
20	18.999	-0.619
21	32.243	1.254
22	27.743	0.668
23	17.269	-0.908
24	19.911	-0.472
25	27.938	0.694
26	29.321	0.878
27	20.600	-0.364
28	21.117	-0.284
29	18.125	-0.763
30	15.397	-1.241
31	29.112	0.850
32	19.018	-0.616
33	21.330	-0.251
34	26.429	0.488
35	29.081	0.846
36	27.604	0.649

図７-１．　モデルとの適合度の検討（１）

図７は，PROX 処理を行った全項目と全受験者について，モデルとの適合度の度合いを求めたものである．ｔはモデルとの適合度の度合いを示す値である。ｔ>2 の場合,その項目または受験者はミスフィットと判断される．ミスフィットは，例えば，能力の高い受験者の誤答が多く，逆に低い受験者の正答が多い場合などがミスフィット項目となり，受験者のミスフィットの場合には，例えば，当て推量だけで解答したために，能力パラメータから判断すれば，当然正答してもよいような易しい設問に誤答したり，反対にとても正答できないような難しい設問に正答するなどの場合にミスフィット受験者となる．従って，より正確な受験者能力パラメータなどを算出するためには，このようなミスフィットは除外することが望ましい．図７-１の場合，設問 38 がミスフィット項目に該当した．図７-２では，10017 の受験者がミスフィット受験者に該当した．

一方，ｔ< -2 の場合，オーバフィットと判断されて，例えば，能力の高い受験者が全て正答し，能力の低い受験者が全て誤答する場合など，オーバフィット項目となる．受験者の場合には，ある受験者の解答が，易しい設問には全て正答し,難しい設問には全て誤答するように，モデルが想定しているパターンとのずれが極端に小さい場合などがオーバフィット受験者となる．ただし，オーバフィットの場合には，除外して再計算することはしないが，原因を明らかにして改善していくことが必要となる．図７-１では設問 9, 47 がオーバフィット項目に，図７-２では，10021 の受験者がオーバフィット受験者となった．

Person No.	Sigma(Z^2)	t
10001	60.591	1.317
10002	62.321	1.474
10003	46.083	-0.095
10004	35.081	-1.324
10005	40.285	-0.720
10006	35.376	-1.288
10007	51.535	0.457
10008	40.688	-0.675
10009	29.979	-1.968
10010	41.998	-0.531
10011	41.963	-0.535
10012	37.070	-1.087
10013	54.225	0.719
10014	41.034	-0.637
10015	50.989	0.403
10016	45.538	-0.152
10017	70.128	2.163
10018	48.472	0.151
10019	57.572	1.037
10020	47.543	0.056
10021	21.465	-3.216
10022	52.595	0.561
10023	48.980	0.202
10024	55.731	0.863

図７-２．モデルとの適合度の検討（２）

2.3 S-P 表分析

S-P 表分析（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* = 0.446										
		1	2	5	3	4	6	%	Blank	9	c.s
3	10003	1	1	1	1	1	5	100	0		0
5	10005	1	1	1	1	1	5	100	0		0
7	10007	1	1	1	1	1	5	100	0		0
15	10015	1	1	1	1	1	5	100	0		0
21	10021	1	1	1	1	1	5	100	0		0
10	10010	1	1	1	1	0	4	80	0		0
13	10013	1	1	1	0	0	3	60	0		0
17	10017	0	1	1	1	0	3	60	0		0.67
20	10020	1	0	1	1	0	3	60	0		0.67
23	10023	1	1	0	0	1	3	60	0		1
1	10001	1	1	0	0	0	2	40	0		0
2	10002	1	1	0	0	0	2	40	0		0
4	10004	1	1	0	0	0	2	40	0		0
6	10006	1	0	1	0	0	2	40	0		0
8	10008	0	0	1	0	1	2	40	0		1.5
9	10009	1	0	1	0	0	2	40	0		0
12	10012	0	1	1	0	0	2	40	0		0
11	10011	0	0	1	0	0	1	20	0		0
14	10014	0	1	0	0	0	1	20	0		0
16	10016	0	0	0	1	0	1	20	0		2
19	10019	0	0	0	1	0	1	20	0		2
22	10022	0	0	0	0	1	1	20	0		3
18	10018	0	0	0	0	0	0	0	0		0
24	10024	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，50 の設問を 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設問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，易しい設問ができず，難しい設問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

の設問の適切さ，指導の効果などを読み取ることができる．ブロック(1)～(5)の設問は上側の正答数が下側の正答数を上まわっているので，得点が高い受験者は正答し，低い受験者は誤答しているので，設問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 と大きくなく，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，全て 0.2 以下となり，試験問題として適切だと判断された．CS は受験者注意係数を示し，0.5 以上の受験者 7 名が要注意の受験生と判断された．

３．　まとめ

今回，受験者の正当な評価を行うために，Moodle の小テスト問題に対して，古典的テスト理論および現在テスト理論によるテスト問題の分析を行った．その結果，いくつかの課題がみつかった．

まず，Moodle のアイテム分析による分析では，判別係数が負の値となった３つの設問が，また，識別指数が負の値となった２つの設問が見直しの対象となったが，他のほとんどの設問においては，Moodle のアイテム分析では問題は認められなかった．

TDAP による項目分析では，設問 7, 8 が全員正答となり，また，項目困難度が 0.5 以下となった設問 11 個以外の 39 問は易しい設問と判定され，全体的にはほぼ妥当な結果が得られた．項目弁別力指数では，設問 15, 21, 39 が負の値を示し，能力の低い受験者が高い受験者よりも多く正答したと判断され，今後検討が必要である．今回の項目分析で最も課題が明らかになったのは，実質選択肢数が 2 以下の設問が 50 問中 22 問もあり，４つの選択肢を用意したにも関わらず，２つ分の働きしかしていなかったことがわかった．今後，選択肢の問題作成の際には大いに注意を払わなければならない．

現代テスト理論の項目反応理論による分析では，最終項目困難度は 50 問中難しい設問と判断されたのは 21 問だったので，全体としてはバランスがとれた 50 問だったといえよう．一方，モデルの適合度の検討では，ミスフィット項目として設問 38 が該当したので，今後検討を要する．

S-P 表分析では，得点が高い受験者は正答し，低い受験者は誤答しているので，設問としては適切と判断される．

今後は，表１の小テストの出題項目から明らかなように，出題項目とブロックが対応していないので，分野別にまとめる必要がある．また，より多い受験生の小テスト結果について，試験問題の妥当性の検証を行いたい。

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.