広島工業大学紀要教育編
第17巻（2018）71–77
論　　文

テスト問題の配点と得点調整に関する一考察：

項目反応理論との比較

廣瀬　英雄*

（平成29年９月15日受付）

A Consideration on Point Allotments to Items and Score Adjustment in Testing:

Comparison to the Item Response Theory

Hideo HIROSE

(Received Sep. 15, 2017)

Abstract

In many situations, regardless of whether we are aware of it or not, point allotment to items in testing and score adjustment are often performed without deep considerations; for example, in final examination in universities, teachers try to find the appropriate point allotments for scores, or to set the appropriate threshold to discriminate the successful group and the failed group by using their own methods. However, such the methods of point allotment and score adjustment may affect the examinees' accurate evaluation for abilities. The item response theory, IRT, is one of the famous methods to evaluate examinees' abilities and items (problems) difficulties simultaneously accurately and efficiently. In this paper, we consider typical situations in testing to compare the results from point allotment methods and score adjustment methods with those from the item response theory. We have found that score adjustment methods have possibilities to disturb the orders of scores arranged by using the raw scores. However, the IRT may consistently leave the orders as they were. In addition, the IRT arranges appropriate scores. To evaluate the examinees' abilities accurately, it is recommended to use the IRT method rather than to use point allotment or score adjustment methods.

Key Words: point allotment, score adjustment, item response theory, difference reduction method between two empirical distribution functions, score transform via median value shifting

1　はじめに

期末試験，あるいはそれに類するテストでは，普通，指定された分野から複数の問題が適切に選び出され，各問題に割り当てられた配点の下で採点が行なわれた後，各問題で得られた得点の合計点を受験者の習熟度や能力を総合的に測る指標としている。例えば，100点満点を総合点とするとき，各問題への配点は問題の難易度によってテストを実施する側で，４問ならすべて25点というように，事前に設定されている。この方法を使うとき，配点によってはスコアが上下するだけでなく，総合順位も変わってくる可能性がある。

そこで，筆者はできるだけ公平で公正なテストを目指す意味で，これまで，問題の難易度も自動的に計算してそれを評価に取り込むことができる「項目反応理論」（item response theory, IRT[1-8]）を用いた評価法を期末試験など

* 広島工業大学データサイエンス研究センター＆環境学部建築デザイン学科

に取り込んできた。こうすることで，事前に配点を意図的にあるいは恣意的に決めておくということからは解放され，ある程度，公平・公正性は保たれていると考えていた。

しかし，そもそも配点の配分法について，従来の事前配置を用いた方法，あるいは項目反応理論などの現代テスト理論を用いた方法などについて，それらの方法が公正性や公平性を持っているかどうかについてはあまり議論されていないように思われる。

ここでは，従来から用いられてきた配点の配分法を用いた総合得点と項目反応理論を用いた総合得点の比較，あるいは，それらを用いた場合の受験者の順位の変化に与える影響に注目してみたい。従来からの配点法については，（平均点が上がるように，あるいは下がるように）配点を意図的に変えることで総合得点や順位にどのような影響が出るかも考察してみたい。

特に，素点（ここでは各問題の得点の合計）の経験分布をシフトさせることによって得られる得点調整の結果，調整前後の順位に与える影響についても観ていく。得点調整には，センター試験で用いられている「**分位点差縮小法**[9]」や私立大でよく使われている「**中央値補正法**[10]」などがある。前者は複数のテストの結果に開きがある場合に使われ，後者は単一のテストでの得点分布変更を行なう場合に使われる。ここでは比較的簡単に計算できる後者についてのみ述べる。得点調整の影響についても，従来法による結果と項目反応理論を用いた場合と比較してみたい。

2　問題の難易度と配点の重み

どの問題も難易度が同程度の場合には，すべての問題に等しい配点を与えるのは自然である。配点について困難が生じるのは問題の難易度に差が出る場合である。ここでは次の３つの方法について考えている。

2.1　すべての問題に同じ配点を与える均等配点法

問題の難易度にかかわらず，すべての問題の配点を均等にするという方法である。これをここでは均等配点法とよぶ。

2.2　難しい問題に高い配点を行なう加速配点法

易しい問題には多くの受験者が正解を出して，難しい問題に正解を出す受験者は少ないと考えられる。そこで，習熟度や能力を精度良く測るには，難しい問題に高い配点を与え，易しい問題には配点を低くするという方法である。易しい問題では能力に応じた差は出ないが，難しい問題では差が出る，という意味である。言い方を変えると，受験者の順位付けに重きを置いているとも考えられる。ただし，どの程度難しくなるとどの程度の配点の比率を考えればよいかということについてはあまり深く考えられておらず，例えば，４問の問題の難易度が１，２，３，４というような場合（この数値は例えば正答率が４：３：２：１というように考えるとわかりやすい），配点比率を１：２：３：４というように与えるという程度である。これをここでは加速配点法とよぶ。

容易に想像できるが，この場合受験者全員の平均点は均等配点法の場合の平均点よりも低い。

2.3　難しい問題に低い配点を行なう減速配点法

易しい問題に正解を出すことができるのは基本的なことであるため，基本を押さえておくという意味では易しい問題には高い配点を与え，難しい問題には配点を低くするという方法である。これは習熟度や能力を測るという意味からは精度が悪くなるように考えられる。しかし，見方を変えると，一定程度の習熟度を満たしているかどうかを測るために，あるしきい値以上なら合格でそれ以下なら不合格というように，受験者のグループを２つに分けるという２値分類を行なっているとも考えられる。４問の問題の難易度が１，２，３，４というような場合，配点比率を４：３：２：１というように与えるということに相当する。これをここでは減速配点法とよぶ。

容易に想像できるが，この場合受験者全員の平均点は均等配点法の場合の平均点よりも高い。

3　得点調整法

3.1　単一科目だけでの調整

１つの科目でも，受験者の平均点が非常に高かったとき，あるいは思わぬ低さを示したときで，平均点によってある意思決定を行なうことに意味がある場合，平均点の数値そのものが意味を持つことがある。例えば，単位取得条件は素点が60点以上となっているにもかかわらず，期末試験の平均点が40点だとしたら，（対称分布なら）受験者の半数が40点以下になり，半数以上が不合格になることにある。これは成績をつける上で好ましいことではない。そこで，何らかの得点調整を行って平均点を上げようとする場合がある。

平均を上げるには減速配点法を使うことができる。実際の採点の現場で，あらかじめ配点が受験生に示されていない場合にはそのことは可能で，易しい問題に高い配点を与えて全体的に素点をジャックアップさせるやり方である。これは採点時に行なうことができる。

もう１つの方法は，中央値を操作し（上下させ），それにともない，素点が中央値以下なら中央値操作に比例する操作を，以上ならその対称の操作を行うという方法である。これを**中央値補正法**とよぶ。数式で示すと，中央値を m，

素点を x，補正後の点数を y とすると，

$$y = a \cdot x \qquad (x < m)$$
$$y = \frac{100 - a \cdot m}{100 - m}(x - m \cdot x) + a \cdot m \qquad (x \geq m)$$

となる。a の値は，平均が低いときには $a > 1$ で，高いときには $a < 1$ である。これは採点が終わって素点が出揃った時に行なうことができる。

3.2　複数科目間の調整

複数科目の場合，例えば理科の科目で，生物，化学，物理の３科目から１科目選択して理科の点数とする場合で，例えば，化学が全体的に非常に高い点数だった場合，化学の素点はそのままにしておき，生物，物理の点数に得点調整を行なう。化学と物理の経験分布関数を並べた後，物理の経験分布と化学の経験分布の間に，補正した経験分布関数を作るというものである。

これはセンター試験にも使われている方法で，同じ分野でも選択科目が異なることによって不公平が起こらないように配慮する方法である。**分位点差縮小法**と呼ばれる。具体的な変換法についてはここでは述べない。

4　素点と得点変更後の点数の比較

4.1　テスト１の場合（平均点が50点に近い場合）

現象をわかりやすく説明するため，４人の受験生 A，B，C，D が４問（Q1，Q2，Q3，Q4）を受験した結果が表１に示すとおりになったと仮定する。表の数値は１問の点数が100点に換算したときの得点の割合を示している。この例は，平均点が50点に近い場合の例になっている（これをテスト１とよぶ）。

容易にわかるように，習熟度は A が最も低く D が最も高い。また，問題の難易度は Q1が最も易しく Q4が最も難しい。

表１　テスト１を受験したときの得点率（パーセント）

		問題				得点
		Q1	Q2	Q3	Q4	平均
受験者	A	50	40	30	20	35
	B	60	50	40	30	45
	C	70	60	50	40	55
	D	80	70	60	50	65
得点	平均	65	55	45	35	50

次に，表１のテストの結果に対して，加速配点，均等配点，減速配点，IRT の評価を行なってみた結果を表２に示す。加速配点法では Q1，Q2，Q3，Q4にそれぞれ10%，20%，30%，40%の配点の重みを置き，減速配点法では Q1，Q2，Q3，Q4にそれぞれ40%，30%，20%，10%の配点の重みを置いている。表の中で ability は IRT による ability 値を表し，IRT 点は ability を50倍して50を加えるという変換を行っている。IRT の計算には応答マトリクスの要素の値が0/1の２値をとることが求められるが，ここでは EM タイプ IRT を使って応答に [0, 1] の有理数値（実際的には実数値でもよい）をとることを許している。

表２を見ると，加速配点により平均は下がり，減速配点により平均が上がっていることが確認できる。IRT を用いた IRT 点はこの場合には均等配点の結果と全く同じになっている。

表２　加速配点，均等配点，減速配点，IRT の結果の比較（テスト１）

		評価法				
		加速	均等	減速	ability	IRT 点
受験者	A	30	35	40	−0.295	35
	B	40	45	50	−0.097	45
	C	50	55	60	0.097	55
	D	60	65	70	0.295	65
得点	平均	45	50	55	0	50

次に，加速配点，均等配点，減速配点の結果に対して中央値補正法によって得点調整を行なった例を見てみよう。ここでは目標とする変換後の中央値を50点とした。表３にこの結果を示す。

すべての平均点が50点に統一され，得点調整の効果が見える。また，加速配点や減速配点の影響はそれほど受けていないこともわかる。

表３　得点調整を行なった結果（テスト１）

		評価法		
		加速	均等	減速
受験者	A	33	35	36
	B	44	45	45
	C	56	55	55
	D	67	65	64
得点	平均	50	50	50

ただ，このテスト１の例は得点調整を行なわなくてもよいような比較的おだやかな例であった。しかし，もし，得点調整を行わないままであると平均点が50点（あるいは60点）から大きく外れる場合にはどうであろうか。得点調整を行う場面はこういった局面であるため，次に２つの例（１つは平均点が50点から離れ小さくなった場合（テスト２とよぶ），もう１つは大きくなった場合（テスト３とよぶ））について，得点配分の変更による影響と得点調整の結果について見てみたい。

4.2　テスト２の場合（平均点が50点よりも小さい場合）

表４に，テスト２を受験したときの得点率を示す。素点

での平均点が30点になっており，平均点を上げたい動機がある場合には得点操作が有効に働くと思われる。

表 4　テスト 2 を受験したときの得点率（パーセント）

		問題				得点
		Q11	Q12	Q13	Q44	平均
受験者	A	30	20	10	0	15
	B	40	30	20	10	25
	C	50	40	30	20	35
	D	60	50	40	30	45
得点	平均	45	35	25	15	30

表 4 のテストの結果に対して，加速配点，均等配点，減速配点，IRT の評価を行なってみた結果を表 5 に示す。配点の重みづけはテスト 1 と同じである。

表 5 を見ると，均等配点と比較して，加速配点では平均は下がり，減速配点では平均が上がっていることが確認できるが，減速配点でも50点に及ばない。しかしながら，IRT 点は平均点が52点になっており，均等配点より全体的に20点程度加算されていることがわかる。

表 5　加速配点，均等配点，減速配点，IRT の結果の比較（テスト 2）

		評価法				
		加速	均等	減速	ability	IRT 点
受験者	A	10	15	20	−0.313	34
	B	20	25	30	−0.069	47
	C	30	35	40	0.154	58
	D	40	45	50	0.363	68
得点	平均	25	30	35	0	52

次に，加速配点，均等配点，減速配点の結果に対して中央値補正法によって得点調整を行なった例を見てみる。目標とする変換後の中央値はテスト 1 と同様50点とした。表 6 にこの結果を示す。

すべての平均点が50点に統一され，得点調整の効果が見える。また，加速配点や減速配点の影響はそれほど受けていないこともわかる。

表 6　得点調整を行なった結果（テスト 2）

		評価法		
		加速	均等	減速
受験者	A	20	25	29
	B	40	42	43
	C	60	58	57
	D	80	75	71
得点	平均	50	50	50

4.3　テスト 3 の場合（平均点が50点よりも大きい場合）

表 7 に，テスト 3 を受験したときの得点率を示す。素点での平均点が70点になっており，平均点を下げたい動機がある場合には得点操作が有効に働くと思われる。

表 7　テスト 3 を受験したときの得点率（パーセント）

		問題				得点
		Q21	Q22	Q23	Q24	平均
受験者	A	70	60	50	40	55
	B	80	70	60	50	65
	C	90	80	70	60	75
	D	100	90	80	70	85
得点	平均	85	75	65	55	70

表 7 のテストの結果に対して，加速配点，均等配点，減速配点，IRT の評価を行なってみた結果を表 8 に示す。配点の重みづけはテスト 1，2 と同じである。

表 8 を見ると，均等配点と比較して，加速配点では平均は下がり，減速配点では平均が上がっていることが確認できるが，加速配点でも50点に及ばない。しかしながら，IRT 点は平均点が48点になっており，均等配点より全体的に20点程度減算されていることがわかる。

表 8　加速配点，均等配点，減速配点，IRT の結果の比較（テスト 3）

		評価法				
		加速	均等	減速	ability	IRT 点
受験者	A	50	55	60	−0.363	32
	B	60	65	70	−0.154	42
	C	70	75	80	0.069	53
	D	80	85	90	0.313	66
得点	平均	65	70	75	0	48

次に，加速配点，均等配点，減速配点の結果に対して中央値補正法によって得点調整を行なった例を見てみる。目標とする変換後の中央値はテスト 1，2 と同様50点とした。表 9 にこの結果を示す。

すべての平均点が50点に統一され，得点調整の効果が見える。また，加速配点や減速配点の影響はそれほど受けていないこともテスト 2 の場合と同様である。

表 9　得点調整を行なった結果（テスト 3）

		評価法		
		加速	均等	減速
受験者	A	38	39	40
	B	46	46	47
	C	54	54	53
	D	62	61	60
得点	平均	50	50	50

4.4　複数科目の合計点を評価点とする場合

これまでの例では単一科目の中での調整の結果を調べたものであった。ここでは，総合的な学力を測るというように，複数科目（例えば，数学，理科，英語の３科目の合計点，あるいはそれらの平均点）で評価点を表そうとする場合について調べる。 テスト１，テスト２，テスト３では，テスト１には得点調整は不要で，テスト２，テスト３は平均がテスト１と大きく離れているので得点調整の必要性を感じさせるが，ここでは，テスト１，２，３すべてに得点の変更を試みた結果について述べる。

今度は，受験者A，B，C，Dがテスト１，２，３を受け，科目それぞれに加速配点，均等配点，減速配点，IRTの結果の総合点（テスト１，２，３で得られた得点の平均値である）について調べてみる。表10に，テスト１，２，３を受験したときの加速配点，均等配点，減速配点，IRTの結果の総合点の比較を示す。

表10　加速配点，均等配点，減速配点，IRTの結果の比較
（テスト１，２，３の総合点）

		評価法				
		加速	均等	減速	ability	IRT点
受験者	A	30	35	40	−0.405	30
	B	40	45	50	−0.133	43
	C	50	55	60	0.133	57
	D	60	65	70	0.405	70
得点	平均	45	50	55	0	50

また，表11に，テスト１，２，３を受験したときの加速配点，均等配点，減速配点の結果に対して中央値補正法によって得点調整を行なったときの総合点の比較を示す。これは，テスト１，２，３で得られた得点の平均値である。

表11　得点調整を行なった結果
（テスト１，２，３の総合点）

		評価法		
		加速	均等	減速
受験者	A	31	33	35
	B	44	44	45
	C	56	56	55
	D	69	67	65
得点	平均	50	50	50

これまで見てきたところ，テスト１，２，３のいずれの場合でも，また総合的に見ても，たとえ得点の大きさが変更されても，A，B，C，Dの得点評価の順位は変わっていない。しかし，場合によってはこの順位が変わる場合が起こる。期末試験の成績のように評価値そのものが意味を持つ場合には得点の加速や減速あるいは得点調整は有効に働くように思えるが，受験者の能力評価の順位だけが問題になる場合にはそれほどの意味はない。しかし，変換した評価値の順位が素点の順位と異なる場合（逆転現象）にはこのことが問題になってくる。次の例でこのようなケースを確認する。

5　得点調整によって順位が逆転する場合

A，B，C，Dが３つの科目（例えば，ここでは数学（テスト４），理科（テスト５），英語（テスト６）と仮定する）を受験したときの得点率が表12のとおりであったとする。このときの，それぞれの科目についての加速配点，均等配点，減速配点，IRTの結果を表13に，得点調整を行なった結果を表14に示す。

表12　テスト４，５，６を受験したときの得点率

科目	受験者	問題				得点
数学（テスト４）		Q31	Q32	Q33	Q34	
	A	72	62	52	42	57
	B	71	61	51	41	56
	C	69	59	49	39	54
	D	68	58	48	38	53
	平均	70	60	50	40	55
理科（テスト５）		Q41	Q42	Q43	Q44	
	A	42	32	22	12	27
	B	45	35	25	15	30
	C	48	38	28	18	33
	D	51	41	31	21	36
	平均	47	37	27	17	32
英語（テスト６）		Q51	Q52	Q53	Q54	
	A	100	90	80	70	85
	B	97	87	77	67	82
	C	94	84	74	64	79
	D	91	81	71	61	76
	平均	96	86	76	66	81

表13　加速配点，均等配点，減速配点，IRTの結果の比較
（テスト４，５，６それぞれ）

科目	受験者	評価法				
数学（テスト４）		加速	均等	減速	ability	IRT点
	A	52	57	62	0.031	52
	B	51	56	61	0.011	51
	C	49	54	59	−0.028	49
	D	48	53	58	−0.048	48
	平均	50	55	60	−0.008	50
理科（テスト５）		加速	均等	減速	ability	IRT点
	A	22	27	32	−0.068	47
	B	25	30	35	−0.001	50
	C	28	33	38	0.064	53
	D	31	36	41	0.128	56
	平均	27	32	37	0.031	52
英語（テスト６）		加速	均等	減速	ability	IRT点
	A	80	85	90	0.082	54
	B	77	82	87	−0.004	50
	C	74	79	84	−0.086	46
	D	71	76	81	−0.166	42
	平均	76	81	86	−0.044	48

表14 得点調整を行なった結果
（テスト４，５，６それぞれ）

科目	受験者	評価法		
数学（テスト４）		加速	均等	減速
	A	52	57	62
	B	51	56	61
	C	49	54	59
	D	48	53	58
	平均	50	55	60
理科（テスト５）		加速	均等	減速
	A	44	45	46
	B	50	50	50
	C	56	55	54
	D	62	60	59
	平均	53	53	53
英語（テスト６）		加速	均等	減速
	A	62	61	60
	B	59	59	58
	C	57	56	56
	D	55	54	54
	平均	58	58	57

（テスト４については平均点が50点に近いので得点調整は行なっていない。テスト５，６のみに対して行なっている）

科目毎に見ると，変換した評価値の順位が素点の順位と同じになっている。この３科目を合計して平均をとってみる。その結果を表15に示す。また，テスト４，５，６を受験したときの加速配点，均等配点，減速配点の結果に対して中央値補正法によって得点調整を行なったときの総合点の比較を表16に示す。更に，IRT の評価法については，３科目を同時に評価することができるので，その結果を表17に示した。

表15 加速配点，均等配点，減速配点，IRT の結果の比較
（テスト４，５，６の総合評価）

		評価法				
		加速	均等	減速	ability	IRT 点
受験者	A	51.3	56.3	61.3	0.015	50.7
	B	51.0	56.0	61.0	0.002	50.1
	C	50.3	55.3	60.3	−0.017	49.2
	D	50.0	55.0	60.0	−0.028	48.6
得点	平均	50.7	55.7	60.7	−0.007	49.6

表16 得点調整を行なった結果
（テスト４，５，６の総合評価）

		評価法		
		加速	均等	減速
受験者	A	48.8	50.9	52.8
	B	49.7	51.5	53.3
	C	50.3	51.8	53.4
	D	51.2	52.4	53.8
得点	平均	50.0	51.7	53.3

表17 IRT による同時評価
（テスト４，５，６を同時に用いて計算）

		IRT による同時評価	
		ability	IRT 点
受験者	A	0.039	51.9
	B	0.029	51.4
	C	0.009	50.5
	D	−0.001	50.0
得点	平均	0.019	51.0

A，B，C，D の能力値を θ_A，θ_B，θ_C，θ_D とするとき，

得点調整なしでは，　$\theta_A > \theta_B > \theta_C > \theta_D$

得点調整を行なうと，　$\theta_A < \theta_B < \theta_C < \theta_D$

のように，評価の順位が逆転していることがわかる。しかし，IRT 評価の場合には，科目毎に評価した IRT 点の平均値，あるいは３科目を同時に用いて評価した IRT 点のいずれも，素点，あるいは得点調整なしの評価順位と同じになっている。

中央値補正による得点調整法では，単独科目では得点調整前後での評価の順位は素点を使ったときの順位と同じに保たれるが（これは中央値を境にして線形変換していることから明らか），複数科目の得点から総合評価値を求める場合，科目の得点平均値から大きく離れている場合に得点調整法を用いて総合評価値を求めた結果では，その評価値の順位が素点での順位と逆転することがある。特に，総合評価値が似通った受験生の僅差を問題にする場合には，得点調整法を行なった場合と行なわずに素点のまま評価した場合とで結果が異なってくることがある。

しかしながら，IRT を用いた方法では，

１）評価値は常に50点を中心としたばらつきを示し，
２）受験者の能力値を適切に反映した結果が得られ，
３）素点による複数科目の総合点から順位を求めた結果は IRT の順位の結果と整合性がある

ことがわかった。IRT は，公正で公平な評価法であることが特徴であるが，単一科目での期末試験だけでなく，複数科目の総合評価を求めるような場合にも有用な評価法であることが示された。

6　考　察

どのような状況でも，能力の非常に高い人やその逆の人の習熟度はテストに特段の工夫がなくても測ることが可能であるが，能力の似通った人の中から合格と不合格の２つのグループに分けるしきい値を設定するのはとても困難である。なぜなら，習熟度のわずかな違いは確率的な変動の幅の中に埋もれてしまい，チャンスのいたずらによって悲劇が生まれることもあるからである。しきい値以上は合格でそれ未満は不合格であるという絶対的な指標をそれほど

の疑問もなく多くの人が受け入れている理由はよくわからないが，このような分類法をいつまで続けるのであろうか。２人の習熟度が等しい，あるいはある人の習熟度はこの値である，というような仮説を立てて検定すると容易に棄却できないということが言えるはずなのに，いまだに一度のチャンスの結果だけによって分類されている。

社会的に認知されている非常に重要な試験には，IRT などのより公正で公平な能力測定法を評価に加えることや，いくつかの面から多面的な評価を行なうことや，あるいはいったん受け入れてその後習熟度の向上を確認したりするとかの柔軟な方法がそろそろ始まっても良い頃だと考える。

ここで取り上げた議論は，数学的にいつでも成り立つというようなことではない。ある局面においてはその取り扱いに対して理解を深めておく必要があるということだと考える。１つの例ですべてを説得することはできないが，少なくとも反例のような事例にはなっている。

IRT はここで取り上げたような局面でも合理的な結果をもたらしてくれたが，IRT にも確率的な変動がともなっていることに注意したい。

7　まとめ

期末試験，あるいはそれに類するテストでは，問題に事前に配点をあたえておくのが普通である。しかし，採点時に平均点が期待する値から離れていた場合，難しい問題に高得点を与える加速配点法や易しい問題に高得点を与える減速配点法を用いて全体の得点分布を調整することがある。もっと積極的には，試験の後で平均点が満点の50％から遠く離れた場合とか，複数の科目間での得点分布が離れないような得点調整が行なわれる。前者では中央値補正法が，後者には分位点縮小法がその例である。

加速配点法，減速配点法，あるいは得点調整によって全体の得点分布を変えることで，名目上の平均を移動させても受験者の得点順位に変更がなければ特段の問題はない。しかし，単一科目では発生しなかった得点順位の不変性も，科目が複数になり総合的な得点で受験者の習熟度を評価しようとする場合，受験者の得点順位が得点調整を行なう前のもとの得点順位と整合しなくなる可能性が出てくる。本論文ではそのような実際の典型例を示した。

しかし，IRT を用いた方法では，１）評価値は常に50点を中心としたばらつきを示し，２）受験者の能力値を適切に反映した結果が得られ，３）素点による複数科目の総合点から順位を求めた結果は IRT の順位の結果と整合性があることがわかった。IRT は，公正で公平な評価法であることが特徴であるが，単一科目の期末試験だけでなく，総合的な習熟度を問う総合試験のような場合にも有用な評価法であることが示された。

文　献

1）R. K. Hambleton and H. Swaminathan, Item Response Theory: Principles and Applications. Springer, 1984.

2）R. Hambleton, H. Swaminathan, and H. J. Rogers, Fundamentals of Item Response Theory. Sage Publications, 1991.

3）W. J. D. Linden and R. K. Hambleton, Handbook of Modern Item Response Theory. Springer, 1996.

4）月原，鈴木，廣瀬：項目反応理論による評価を加味した数学テストと e-learning システムへの実装の試み，コンピュータ＆エデュケーション（CIEC），Vol. 24, pp. 70–76, 2008.

5）作村，徳永，廣瀬：EM タイプ IRT による不完全マトリクスの完全化とその応用，情報処理学会論文誌，数理モデル化と応用 Vol. 7, No. 2, pp. 17–26, 2014.

6）H. Hirose, T. Sakumura, Item Response Prediction for Incomplete Response Matrix Using the EM-type Item Response Theory with Application to Adaptive Online Ability Evaluation System, IEEE International Conference on Teaching, Assessment, and Learning for Engineering 2012, pp. 8–12, August 20–23, 2012.

7）H. Hirose and T. Sakumura, An Accurate Ability Evaluation Method for Every Student with Small Problem Items Using The Item Response Theory, Proceedings of the International Conference on Computer and Advanced TEchnology in Education（CATE 2010）, pp. 152–158, August 23–25 2010.

8）H. Hirose, T. Sakumura, T. Kuwahata, Score allotment optimization method with application to comparison of ability evaluation in testing between classical test theory and item response theory, Information, Vol. 17, No. 2, pp. 391–410, 2014.

9）前川：大学入試センター試験における選択科目間の得点調整について，計測と制御，40（8）, pp. 568–571, 2001.

10）伊藤：入学試験における得点調整の理論と実態，久留米大学商学研究，8（1）, pp. 196–218, 2002.