資料

テストの教示が被験者の解答行動と成績に与える影響

藤 森 進¹ 中 野 和 代²

EFFECT OF TEST DIRECTIONS UPON EXAMINEES' TEST-TAKING BEHAVIORS AND PERFORMANCE

Susumu FUJIMORI AND Kazuyo NAKANO

The purpose of this study was to investigate the effects that test directions of multiple choice tests produce on examinees' test-taking behaviors and performance. The data for this research were obtained by having students respond to 57 items of Shiba's Word Meaning Comprehension Test that constitutes the version U2. Three forms of U2 were administered under number-right scoring directions, formula-scoring (correction-for-guessing) directions, and neutral directions. There were many omitted responses by examinees under formula scoring directions. On the other hand, there were few omissions under number-right-directions. Three treatments of omitted responses were examined in item response theory : treating as wrong, ignoring, and Lord's method. Great differences existed in estimated ability parameters among these treatments under formula-scoring directions, differences under neutral directions were minor, and one under number-right directions was minimum. It was concluded that formulascoring directions should not be used, since formula-scoring directions restrict examinees' test-taking behaviors and their performance is greatly affected by treatment of omitted responses under such directions.

Key words : number-right directions, formula-scoring directions, multiple choice test, omit, item response theory.

序

テストの教示とその影響

多肢選択式テストにおいて、被験者の解答行動は、正答を知っていれば解答し、知らなければ解答しないというような単純なものではない。正しい知識による正答と、間違った知識による誤答あるいは項目に関する知識がまったくない場合の無答を両極とし、その間には部分的知識に基づくあて推量による正答や誤答、

完全にランダムなあて推量による正答や誤答など、さまざまな解答行動が生じ得る。被験者の性格や能力、問題の難しさ、テストを実施する環境、そして、テストの教示内容などにより解答行動は変化する。

本研究では、被験者の能力の正確な測定を行うテスト実施のあり方を検討するためのひとつの試みとして、被験者の解答行動（無答やあて推量など）に関与する一要因と考えられるテストの教示内容に焦点を置き、多肢選択式テストにおけるその影響について考察する。

関連する諸研究

テストの教示と、これと同時に取り上げられることの多いあて推量によるテスト得点の増分の修正に関する研究は数多い。

¹ 岡山大学教育学部 (Faculty of Education, Okayama University)

² (財団法人) 浅羽医学研究所附属岡南病院 (Kohnan Hospital Asaba Medical Research Foundation)

多肢選択テストにおけるあて推量行動と修正公式に関連する幾つかの研究 多肢選択テストで常に問題となるのが、正しい知識を持たなくとも、あて推量によって正答してしまう可能性の存在である。正答数得点には、あて推量による得点の増分が含まれているため、これを修正し、被験者の真の得点を推定するため(1)式の修正公式 (Guilford (1954), lord & Novick (1968), Lord (1980)) が知られている。

$$S = R - \frac{W}{k-1} \quad (1)$$

ここで、 k は選択肢数、 R は正答数、 W は誤答数である。この式が正しいのは、被験者が答を知っている項目には必ず正答し、知らない項目にはランダムに反応するという条件のときのみである。つまり、誤答選択肢を正答と思い込んでいるケースや、部分的知識の存在により、明らかに誤りとわかる選択肢を取り除いたりすると(1)式は成立しない。無答が存在しない場合には、(1)式の修正得点は正答数得点と相関が1となる。また(1)式は、テストの真の得点の最尤推定値を与えることが Lysterly (1951) によって示されている。

Lord (1975) は、多肢選択形式のテストでは、被験者の真の得点を知るためには、修正公式が正答数得点に勝るという研究結果を示している。この際彼は、誤答数が正答数得点から差し引かれるという採点が行われるという教示と（誤答の数とは無関係に）正答数得点によって採点されるという教示が存在すれば、同一のテストの両教示下の違いは、前者の教示では無答反応があるのに対し、後者ではランダムなあて推量が存在するという点だけであるとした。

テスト教示と解答行動 多肢選択式テストにおける教示が、被験者の解答行動に及ぼす影響についての研究で、Angoff & Schrader (1984) が2つの仮説を提唱している。ひとつは differential effects 仮説で、あて推量による正答の増分を修正するという教示（以下修正教示）のもとでは、ある種の被験者は、部分的知識があっても解答を抑制する傾向があるとするもので、これは人為的に得点を減ずる結果をもたらす、正答数得点で採点するという教示（以下正答数教示）に比べると、公平さに欠けるということを含意している。もうひとつは invariance 仮説と呼ばれるもので、正答数教示では解答するが、修正教示のもとでは解答が抑制される項目について、正答の発生は偶然レベルにすぎない、つまり、修正公式を利用する限り、成績の違いは2つの教示間でないとするものである。研究の結果、本質的に修正公式による成績に差は生じないという結論が

得られ、invariance 仮説が支持された。

この見解に対し、Albanese (1986) は、Angoff らの研究方法と結果を再検討している。彼の研究では、修正教示よりも正答数教示の場合の方が無答項目（正答が確実には分からず、しかもあて推量によって解答することを控えた項目）および、時間切れなどでまったく手をつけていない項目共に少なかった。また、修正公式を用いると、やり残した項目と無答を同質のものとして扱うことになり、実質上は能力の異なる被験者が同一得点になってしまうため、Angoff らがそれを意識していないのは問題であると指摘している。正答数教示では解答するのに、修正教示では無答とした項目の正答率を $SR (= [\text{両教示のもとでの正答数の差}] \div [\text{両教示のもとでの解答項目の差}])$ となると、能力の低い群はかなり高い SR 率を示したが、最高の能力群でも、 SR 率は偶然レベル（選択肢数分の1）より高く、この集団においても、教示の影響により解答行動が抑制されることが示された。すなわち、 SR 率の高い例があるということは、differential effects 仮説を支持することになる。また、正答数教示では被験者がより多くのテスト項目を試みるということは、内容的妥当性を考慮した場合望ましいことである。結局、彼の研究は differential effects 仮説を支持することになり、同時に、正答数教示が適当とするものである。さらに Albanese (1988) は、修正公式による採点が得点に及ぼす影響を調べ、修正教示のもとでは無答した項目に解答した場合、0.5標準偏差程度、得点が上昇するという分析結果を得ている。Albanese の指摘によると、同一被験者に対しはじめに修正教示でテストし、その後に無答項目に解答させると、偶然正答の水準よりも良い成績となる研究が多数あり、これに対して、被験者をランダムにどちらかの教示によるテスト群に割り振るという群間研究では、正答確率が偶然レベルより高くあるにもかかわらず無答してしまうケースが有意に多いという証拠はないとする研究が多い。Albanese の調査結果は以下の通りであった。修正公式のもとでは、部分的知識、誤った知識のどちらの場合でも、無答した項目に解答すると得点にかなりの影響があった。特に、部分的知識は得点上昇の可能性が大きくなるが、換言すれば、慎重な被験者にとっては、修正公式の利用は、得点減少の影響が大きくなる可能性が高いということである。さらに、解答保留のまま次の問題に進むケースが増え、スピードテストの場合には、このやり残した項目がそのまま無答になってしまうことが多い。

Cross & Frary (1978) は、正答数教示と修正教示（ひ

つとも誤答とわかる選択肢がないときはあて推量するな)の2つの状況下でのテストの信頼性と妥当性に与える選択肢の重みづけの影響を、正答数得点や修正公式による得点と比較して検討している。重みとしては、①経験的重みづけ—基準得点を上昇させるように決められた重み、②権威による正答性の重みが検討された。結果は、正答数得点や修正公式による得点に比べ、経験的重みづけによる方法は、信頼性・妥当性共に高くなった。またこの方法では、正答数教示と修正教示のどちらの妥当性にも大差はなかった。これは、選択肢の重みづけは信頼性を向上させるかもしれないが、妥当性に関しては役立たないと主張した Reilly & Jackson の研究を支持するものであるかもしれないと述べている。正答数教示も修正教示も妥当性には大差がないとすれば、正答数教示で、正答数を基準にして重みづけした経験的方法で採点した方が良いであろうと結論している。

既に述べたように、Angoff & Scharader は、修正公式を用いる限り正答数教示でも修正教示でも成績の違いは認められないとしているが、Cross & Frary (1977) は、被験者が無答にした項目をあて推量で置き換えさせると、正答が偶然より有意に増すという結果を報告している。

Frary (1989) の研究結果は、修正公式とその利用という教示下では、中程度の困難度のテスト項目の場合、誤った情報を持つ傾向が低く、無答傾向が少なくとも中程度の被験者の場合には、無答行動による得点の増減が無視し得ない比率で生じるとしている。Albanese (1988) も不適切な無答が得点減になると主張している。

項目反応理論による被験者能力の推定

本研究では、被験者の能力を推定する方法として項目反応理論を利用する。項目反応理論によれば個々の被験者の能力の推定値の精度が分かることや異なったテスト結果も比較できるなどの利点がある。

項目反応理論に従うモデルにはいくつかのものが知られているが、ここでは、(2)式によるあて推量をモデル化した3母数ロジスティックモデル (Birnbau, 1968) を利用することにする。

$$P(x=1|\theta) = c + \frac{1-c}{1 + \exp(-Da(\theta-b))} \quad (2)$$

ここで、 x は正答のとき1、誤答のとき0をとるダミー変数であり、 $D=1.7$ の定数である。また、 θ は被験者の能力、 a は項目の識別力、 b は項目の困難度、 c は偶然的正答が生じる確率をあらわす母数であり、

$P(x=1|\theta)$ は能力 θ の被験者が項目に正答する確率を表わす。このモデルを用いて被験者の能力を推定する場合、最尤推定法とベイズ推定による方法があるが、本研究ではベイズ推定 (Bayes modal estimate) を用いることにする。すなわち、次式を最大とする θ を能力の推定値とする。

$$\Pi_j P_j(\theta)^{x_j} Q_j(\theta)^{1-x_j} f(\theta) \quad (3)$$

ここで、 $f(\theta)$ は θ に関する事前の知識を表現する事前分布である。

項目反応理論における無答処理の問題 教示の影響と密接に関係するのが無答の処理方法についての問題である。項目反応理論における無答の処理についての研究結果を見てみよう。無答項目はどのように採点すべきであろうか。正答が分からないとき解答することは、まったくの偶然から正答するという事態が起こらない限りは、恥をかくだけだと被験者は思うかもしれない。無答を採点からはずすと、確実に正答が分かっているものだけを解答し、望むがままの高い能力推定値を得ることになるだろう (Lord, 1980, p.228)。無解答にもペナルティが必要ということになる。この対極の姿勢が無答を誤答として扱う方法であるが、必然的に、被験者にあて推量などの無理な解答を強いることになる。Lord (1974) は、被験者の能力の推定に利用する尤度関数において、無答をランダムなあて推量による反応に置き換えることによって、最尤推定値よりも良い推定値が得られるとしている。この Lord の無答の処理は、(3)式において無答項目は $x=1$ /選択肢数として処理することを意味する。

本研究の目的

本研究では、多肢選択形式テストの教示として、解答方法のみを記述した中立的教示と採点方法についても言及した正答数教示(正答数得点による採点)、修正教示(正答数と誤答数の差をとる採点)の合わせて3種類の教示を考え、被験者の解答行動への影響について検討する。教示と関連する採点方法として、正答数得点、修正公式による得点、本研究の修正教示に示した正答数と誤答数の差による得点、項目反応理論による能力母数の推定値をとりあげ、採点方法の違いが持つ影響について検討する。さらに、項目反応理論による能力母数の推定に関連した、無答を誤答として処理する方法と、無答を被験者の能力の推定に利用しない方法、そして Lord による方法の比較検討も行う。

方 法

語彙理解力検査

本研究では、テストとして芝の語彙理解力検査U 2版(全57項目、各問とも5肢選択)を利用した(芝, 1978; 芝・野口, 1982)。これは、項目の困難度母数の推定値が既に得られており、被験者の能力と項目の困難度との組合わせによる教示の影響を見るのに都合がよいからである。この研究では、3母数ロジスティックモデルをテスト得点を説明するモデルとして用い、項目母数の推定値としては、芝(1978)の標準化データと、その後収集されたデータを用いた藤森(1993)による推定値を利用した。なお藤森(1993)では、(3)式の θ の事前分布として平均0、標準偏差6の正規分布を採用しているため、本研究においても同一の事前分布を適用している。

教示

教示にはTABLE 1に示した3種類を用いる。すなわち、①中立的教示：解答方法のみを記述した教示であり、芝の語彙理解力検査で一般的に利用されている教示に従っている。②正答数教示：解答方法の説明に加え、正答数得点で採点するという教示、③修正教示：解答方法の説明に加え、無答を除外した(正答数-誤答数)で採点するという教示の3種類である。修正教示による得点の修正内容は、(1)式の修正公式に対応していないが、被験者に誤答を差し引いた形で得点の修正が行われているということを明確にするためには、実例を示す必要があると判断したため、短時間で理解可能な「(正答数-誤答数)で採点する」という教示にした。

TABLE 1 検査に用いた教示3種

教示群	教 示 内 容
中立的教示	「これは、ことばの意味をきくテストです。各問のはじめにあげたことばと、もっとも意味の近いことばを、5つの中から、ひとつだけ選んで○をつけてください。」
正答数教示	「(中立的教示に加えて) このテストは、1問1点で57点満点です。57問のうち、正答できた数があなたの得点になります。たとえば、27問正答し、残りの30問のうち、10問が誤答で20問が無答であれば、得点は、正答の数だけの27点となります。」
修正教示	「(中立的教示に加えて) 採点は、正答から誤答をひく方法で行います。ただし、無答の場合は正答からひきません。このテストは、1問1点で57点満点です。このとき、27問正答し、残りの30問のうち、10問が誤答で、20問が無答であれば、得点は、正答数-誤答数=27-10=17によって、17点となります。」

TABLE 2 各教示群の無答発生状況

教示群	被験者数	全問解答者数(%)	最多無答数	平均無答数
中立的	377	257(68.2)	49	2.8
正答数	136	107(78.7)	22	0.9
修正	272	38(14.0)	55	24.3
全 体	785	402(51.2)	55	9.9

被験者

調査はある地方国立大学の学生785名を対象とし、1991年10月～1993年4月に実施したものである。785名の内訳は、中立的教示群377名、正答数教示群136名、修正教示群272名に講義時間中に集団方式で調査した。入学年度、学年等は統制していないが、「教育心理学概説」などの講義の受講者に調査を行ったものである。各群の人数が異なるのは、中立的教示は芝の語彙理解力検査のオリジナルのものである関係上、調査したケースが多いことによるためである。正答数教示群が最も少ないのは、本研究のために新たに調査したデータであるためである。

結果と考察

正答数得点及び修正公式による分析

全問解答者は中立的教示群68.2%、正答数教示群78.7%、修正教示群14.0%である(TABLE 2参照)。全問解答者の各群の比率の差の検定を行ったところ、 χ^2 値=235.64(自由度=2)となり1%水準で有意な結果となった。正答数教示群では約90%の被験者が無答数0～1で、無答数の平均は0.9、標準偏差3.0、最高22である(FIGURE 1参照)。スピードテストでなく、正答数で採点するという教示にもかかわらず、少数にしても無答が生じているのは、教示内容の理解不足があったか、成績に関係しないテストであったため、動機づけが不十分となったことの現れであろう。これとは対照的に、修正教示群では無答数の平均24.3、標準偏差14.7、最高55であり、ほとんど解答していないケースもある(FIGURE 2参照)。中立的教示群は正答数教示群と同様、全問解答者が多い。同群の無答数の平均は2.8、標準偏差7.8、最高49で、散らばりが大きい点では修正教示群と似ている(FIGURE 3参照)。3群間の無答数のダンカンの多重比較検定の結果では、修正教示群と他の2群との間に5%水準で有意差が見られた。修正教示群は、教示により解答行動が明らかに抑制され、あて推量を控えて完全な知識がある場合に限り解答しているのに対し、正答数教示群はあて推量行動をかなりとって

ると考えられる。

FIGURE 1 正答数教示群の無答数

FIGURE 2 修正教示群の無答数

FIGURE 3 中立教示群の無答数

続いて項目の正答率、つまり、項目の難しさと無答の発生数について検討する。本研究の全データに基づき項目の正答率を求め、これと修正教示群における各

項目の無答発生数との相関を求めたところ、 -0.605 となった。また、中立的教示群では -0.591 の相関となり、いずれも負の相関となった。すなわち、難しい項目でより多くの無答が発生していることが明らかとなった。なお、正答数教示群は無答の発生が少ないため、この分析からは除外した。

次に、平均正答数は、中立的教示、正答数教示、修正教示群の順に23.0, 18.3, 15.1で、修正教示群が最も低い(TABLE 3 参照)。3群間の平均正答数のダンカンの多重比較検定の結果では、3群間に、いずれも5%水準の有意差があった。これは、修正教示群では解答行動が慎重になるため無答数が増え、解答数が少ない分、正答数も減少したと捉えられる。逆に、正答数教示群では解答行動が抑制されない分、あて推量により正答する部分が増えていると思われる。TABLE 3より、平均誤答数は中立的教示群31.1、正答数教示群37.8、修正教示群17.7である。正答数教示群は知識の有無にかかわらず解答する傾向を持つため、誤答数は当然最も多い。知識がない場合はもちろん、部分的知識があっても解答を抑える傾向にある修正教示群と比較すると2倍以上である。

TABLE 3 各教示群の解答結果

教示群	正答数の 平均(SD)	誤答数の 平均(SD)	CF1*の 平均(SD)	CF2**の 平均(SD)
中立的 正答数	23.0(7.5)	31.1(8.7)	- 8.1(14.2)	15.2(8.8)
修正	18.3(5.3)	37.8(6.0)	-19.5(10.9)	8.9(6.7)
修正 全体	15.1(7.6)	17.7(11.2)	- 2.6(12.3)	10.7(7.6)
	19.5(8.0)	27.6(12.0)	- 8.2(14.2)	12.5(8.5)

* CF1=正答数-誤答数 ** CF2=正答数- $\frac{\text{誤答数}}{\text{選択肢数}-1}$

なお、修正教示群と正答数教示群の正答数得点による比較では、当然、正答数教示群の平均成績が上位となるが、修正教示群の教示による方法(正答数-誤答数)で採点すると、修正教示群の平均値が上位になる(修正=-2.6, 正答数=-19.5)。そして、多肢選択テストにおけるあて推量の修正公式による採点でも修正教示群が上位になる(修正=10.7, 正答数=8.9)。このように、教示と採点方法が異なることによって成績の順位も異なる結果を得る。

以上のことから、3種の教示に関し以下のような示唆をすることができる。正答数教示は被験者のあて推量が増加するという問題を含んでいる。しかし、テストの内容的妥当性の観点からは、無答項目が少なくなるため望ましいとも言える。修正教示は、あて推量を

抑制できるが、安全策を選ぶ被験者は無答するため不利になる。部分的知識がある者も、まったく知識がない者も同様に無答し、誤答と測定されるのは、真の知識レベルを評価する観点からは望ましくない。その上、正答が明確でない時に、解答するかどうかの付加的決定に要する解答時間も必要となり、テスト実施面で負担となる可能性がある。中立的教示は、正答数教示と修正教示の中間的性格を持つが、無答行動の発生に関しては、正答数教示と似通ったものとなっている。

項目反応理論による分析

項目反応モデルによる分析には、既に述べたように、3母数ロジスティックモデルを用いた。項目反応理論を適用したときの無答の処理はLordの方法の他に、誤答にカウントする方法と、能力 θ の計算に含めないという方法が存在する。前者の処理は、無答は結局被験者が正答が分からなかったということを意味すると考え、誤答とするわけである。これに対して後者の処理は、時間切れなどによる未解答の無答は、必ずしも被験者が当該項目を受けたときに誤答することを意味しないと考えるものであろう。テスト冊子の最後の方に集中的に発生する未解答はこの種の代表的なものであるが、途中部分に存在する未解答項目も、被験者が解答するのを後回しにしたため未解答項目となった可能性があり、必ずしも誤答として考えられるべきものではない。テスト時間を十分にとった場合は、どちらかというとき無答を誤答として処理すべきようにも思えるが、長時間にわたるテストは、被験者の解答意欲の低下を招くため、必ずしも誤答処理を正当化できるものではない。ここでは、無答の処理方法として項目反応理論で一般的に利用されている3種類のものを扱う。すなわち、当該項目がテストに存在しなかったかのように扱う方法を無答処理1とし、誤答として扱う方法を無答処理2、Lordの方法を無答処理3として比較検討することにする。

無答処理1の場合の能力母数の推定値の平均は3.57、標準偏差0.87となり、無答処理2の平均は3.24、標準偏差1.07、無答処理3の平均は3.45、標準偏差0.82となった(TABLE 4 参照)。無答処理1の場合の方が最も高く、無答処理3が続き、最低が無答処理2となっているが、これは、無答処理2では誤答数が確実に増えるため、当然の結果といえる。興味深いのはTABLE 4に示した教示ごとに見た場合の能力母数の平均値である。まず第1にどの無答処理でも中立的教示が最も高い値となっていることが特徴的である。無答処理1では、修正教示の方が正答数教示より高い値となっているが、

無答処理2では、正答数教示が逆に高い値となっている。無答処理3では、無答処理1と2の中間となっている。

TABLE 4 各教示群の能力母数の推定値の平均

教示群	無答処理1 θ の平均(SD)	無答処理2 θ の平均(SD)	無答処理3 θ の平均(SD)
中立的	3.73(0.79)	3.64(0.86)	3.69(0.76)
正答数	3.12(0.84)	3.10(0.84)	3.12(0.83)
修正	3.57(0.92)	2.75(1.21)	3.27(0.79)
全体	3.57(0.87)	3.24(1.07)	3.45(0.82)

正答数教示の場合は、どの処理方法でもほぼ同一の値を得ているのに対し、修正教示の場合は、無答処理の違いにより、かなり能力母数の推定値の平均が異なる。通常項目反応理論では、無答が存在しない場合、母数の推定に利用する尤度において誤答を0、正答を1として計算する。無答が存在するとき、Lordの方法では無答を1/選択肢数とし、無答処理2では無答を誤答0とすることになる。本研究結果では、Lordの無答処理3が無答処理1と2の中間的特徴を示したわけであるから、無答をノーカウントとする無答処理1は、結果的に被験者の能力推定を他の処理に比較してやや高めに推定するはずであり、実際、結果もその通りとなった。實際上、処理の違いによって θ の平均が大きく異なっているのは修正教示の場合であるが、この原因として、修正教示の場合の無答数がかなり多いため、処理の違いの影響も大きくなったことがあげられるが、無答の処理の効果そのものは、どの教示の場合でも同様に機能していることは当然である。

Lord(1980)の指摘にあるように、無答がテスト項目として存在しなかったかのように処理されるようになると、完全に知っているものだけに解答することになるという問題点が生じると考えられる。本研究の修正教示にはこの問題が生じているようである。実際に修正教示群では、無答を除いた場合の正答率の平均は、TABLE 3から読み取れるように、他の2群に比較してはるかに高い値となっている。

TABLE 5, 6は、各種の能力推定値と無答数、誤答数などの変数間相関行列を全体のデータ及び教示別に示してある。ここにはいくつか特徴的なことがある。無答処理1の θ と無答数はいずれのデータにおいてもほとんど無相関となっているが、無答処理2の θ は、修正教示群では-0.5589とマイナスに大きな相関と

なっているのに対し、正答数教示群では -0.0911 とほとんど無相関となっている。無答処理2による θ が、無答処理2により修正教示群で誤答数の増加をもたらしたことの影響を大きく受けたためであろう。これに対して、無答は無視するという処理1の θ は、この影響を余り受けていないようである。無答処理3の場合は、無答処理1と2の中間の性格をここでも示している。また、修正公式による採点と能力母数の相関は、修正教示群で無答処理3の場合にやや高くなっている。正答数得点との相関は、修正教示群で無答処理2、3が高くなっている。

TABLE 5 各種能力推定値及び無答数などの変数間相関（全データ・修正教示群）〔上三角行列は全データ、下三角行列は修正教示群〕

	無答数	正答数	誤答数	正答－ 誤答	修正公式	無答処理 1の θ	無答処理 2の θ	無答処理 3の θ
無 答 数	1.0000	-.5760	-.8363	.3796	-.2504	-.0294	-.5281	-.2423
正 答 数	-.6650	1.0000	.0335	.5376	.9356	.7405	.9015	.8689
誤 答 数	-.8605	.1917	1.0000	-.8247	-.3215	-.4606	.0410	-.2866
正答－誤答	.3709	.4470	-.7923	1.0000	.8006	.8076	.4756	.7335
修正公式	-.3507	.9326	-.1754	.7397	1.0000	.8642	.8397	.9245
処理 1 θ	-.0762	.6954	-.3743	.7735	.8352	1.0000	.8208	.9490
処理 2 θ	-.5589	.8792	.1347	.4238	.8324	.7887	1.0000	.9195
処理 3 θ	-.3211	.8645	-.1678	.6904	.9288	.9242	.9203	1.0000

1%有意水準（片側）上三角部分 $|r| > 0.08$
下三角部分 $|r| > 0.14$

TABLE 6 各種能力推定値及び無答数などの変数間相関（中立教示群・正答数教示群）〔上三角行列は中立教示群、下三角行列は正答数教示群〕

	無答数	正答数	誤答数	正答－ 誤答	修正公式	無答処理 1の θ	無答処理 2の θ	無答処理 3の θ
無 答 数	1.0000	-.3521	-.5950	.1791	-.1512	-.0637	-.3801	-.1467
正 答 数	-.0742	1.0000	-.5427	.8577	.9784	.8738	.9174	.9188
誤 答 数	-.4457	-.8596	1.0000	-.8973	-.7045	-.6932	-.4466	-.6574
正答－誤答	.2071	.9602	-.9681	1.0000	.9454	.8836	.7555	.8853
修正公式	.0400	.9935	-.9123	.9858	1.0000	.9089	.8850	.9380
処理 1 θ	.0215	.8674	-.7896	.8569	.8715	1.0000	.9239	.9753
処理 2 θ	-.0911	.8737	-.7376	.8316	.8650	.9916	1.0000	.9475
処理 3 θ	-.0135	.8726	-.7764	.8523	.8728	.9982	.9966	1.0000

1%有意水準（片側）上三角部分 $|r| > 0.12$
下三角部分 $|r| > 0.20$

項目反応理論の適用において、能力の推定値が無答数の影響を受けにくいという意味では、無答処理1は魅力的であると言える。しか、被験者は正答を知っ

ているものだけに解答することになるというLordの指摘は問題として残る。特に、被験者が採点から無答が除外されることを知れば、この傾向は著しく助長されるであろう。結局、解答行動の著しい抑制、無答の処理方法の違いによる能力の推定値の違いに大きな差が生まれる修正教示は適当ではないと考えられる。

展望と関連する考察

TABLE 5, 6 から、正答数と誤答数が修正教示群でのみ正の相関となっている。テストでは、正答数が増加すれば誤答数は必然的に減少するのであるから、両者の関係は負の相関となることが一般的であるが、正の相関となったということは、解答数が多い積極的解答被験者と少ない消極的被験者が修正教示で存在していることを意味している。つまりこのことは、Angoff らの言う differential effects 仮説が成立していることを示唆する。

invariance 仮説が、被験者のあて推量行動の発生に伴うランダム誤差の大きさを検討しているのに対し、differential effects 仮説が問題としているのは、修正教示が、被験者の性格などの違いにより、解答の抑制傾向、つまり、同一の教示の下でも、人により無答の発生が異なるという問題提起をするものである。この種の誤差は、あて推量によるランダム誤差よりも重大なのか、それとも状況によって異なるのかという問題が生じる。本研究は、differential effects 仮説を相対的に支持する結果を得たが、この問題の検討のためには、2種類の誤差の大きさの評価と、影響する要因に関する検討が欠かせない。また、冒頭に記した differential effects 仮説と invariance 仮説をめぐる研究結果の間には若干のくい違いがある。これらの研究では、どのような教示のもとで、どのような被験者がどのような項目に解答し、あるいは解答しないかということが定量的に示されていない。つまり、テストの教示内容によって、被験者が解答するテスト項目が変化し、しかもその変化には、テスト項目の困難度と被験者の能力が関係しているわけであるから、諸研究の結果が必ずしも一致しないのは、この点に関する論議が正確さを欠いているのも一因といえる。本研究ではそのような試みの第一歩として項目反応理論を利用して被験者の能力について推定した。項目反応理論に基礎を置いたコンピュータを利用した適応型テストなどを利用して、被験者の能力と提示される項目の困難度を正確に統制することにより、被験者の解答行動がどのように変化するかなどを検討することが今後意義のある研究になるであろう。

引用文献

- Albanese, M.A. 1986 The correction for guessing : A further analysis of Angoff and Schrader. *Journal of Educational Measurement*, **23**, 3, 225—235.
- Albanese, M.A. 1988 The projected impact of the correction for guessing on individual scores. *Journal of Educational Measurement*, **25**, 2, 149—157.
- Angoff, W.H. & Schrader, W.B. 1984 A study of hypotheses basic to the use of rights and formula scores. *Journal of Educational Measurement*, **21**, 1, 1—17.
- Birnbaum, A. 1968 Some latent trait models and their use in inferring an examinee's ability. In F.M.Lord & M.R.Novick, *Statistical theories of mental test scores*. Reading, Mass : Addison-Wesley, 1968. Chapters 17—20.
- Cross, L.H. & Frary R.B. 1977 An empirical test of Lord's theoretical results regarding formula scoring of multiple-choice tests. *Journal of Educational Measurement*, **14**, 4, 313—321.
- Cross, L.H. & Frary, R.B. 1978 Empirical choice weighting under "Guess" and "Do not guess" directions. *Educational and Psychological Measurement*, **38**, 613—620.
- Frary, R.B. 1989 The effects of inappropriate omissions on formula scores : A simulation study. *Journal of Educational Measurement*, **26**, 1, 41—53.
- 藤森 進 1993 芝の語彙理解力検査の項目パラメタの推定結果についての報告 東京工業大学工学部教育群主催 IRT セミナー資料
- Guilford, J.P. 1954 *Psychometric Methods*, 2nd ed. McGraw-Hill, NY. 秋重義治 (監訳) 1959 精神測定法 培風館
- Lord, F.M. 1974 Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, **39**, 2, 247—264.
- Lord, F.M. 1975 Formula scoring and number-right scoring. *Journal of Educational Measurement*, **12**, 7—11.
- Lord, F.M. 1980 *Applications of item response theory to practical testing problems*. Hillsdale, NJ : Lawrence Erlbaum Associates.
- Lord, F.M. & Novick, M.R. 1968 *Statistical theories of mental test scores*. Reading, Mass. : Addison-Wesley.
- Lyerly, S.B. 1951 A note on correcting for chance success in objective tests. *Psychometrika*, **16**, 21—30.
- 芝 祐順 1978 語彙理解尺度作成の試み 東京大学教育学部紀要, **17**, 47—58.
- 芝 祐順・野口裕之 1982 語彙理解尺度の研究 I 東京大学教育学部紀要, **22**, 31—42.

謝 辞

語彙理解力検査の利用を許可下さいました東京大学名誉教授芝祐順先生には心から感謝します。被験者としてご協力頂いた岡山大学学生の皆さんにも感謝いたします。また本研究のために光岡美貴さんよりデータを提供して頂きました。この場を借りて御礼申し上げます。

(1993.12.23受稿, '94.9.2受理)