

[Original Paper]

## Evaluation of N system of multiple-choice questions reflecting the degree of confidence of examinees

Kimihisa Nomura\*, Manabu Ashikaga\*, Tomoyo Ohno\*\*, Keiko Kawamura\*\*\*

\* Aino University

\*\* Aino Gakuin College

\*\*\* Baika Women's University

### Abstract

We developed a new method for answering multiple-choice questions which we have named the 'N system', in order to evaluate the degree of confidence in answers. Namely, when examinees are confident in their answers, they are requested to select one answer, and when they are less confident, to select two answers, one they consider the most probable and another they consider less probable. Examinees are informed in advance that the allocation of marks for a correct answer differs according to the examinee's degree of confidence.

The survey by unsigned questionnaire was performed on 85 first-year students and 77 sophomore students during the final examination at A College in February 2004. The questionnaire sheets were immediately submitted by all students.

Most students (91.3%) approved of the N system because they are able to select two answers when they are less confident in their answer. 46.9% remarked that they had difficulty in deciding whether to select a single answer or two. 43.5% reported that the consumption of time to select two answers was short. The opinion that N system provides a useful self-assessment of confidence on examinee's knowledge was supported by 64.6% of students, while 37.0% of students thought that N system reflected the knowledge of each student and were supportive of its use as a method of evaluation. The allocation of marks was generally approved by examinees.

**Key words** : multiple-choice question, new method of answering (N system), evaluation, examinee, confidence

(原 著)

## N 式実力反映型多肢選択試験に対する受験者側の評価

野村公寿\*, 足利 学\*, 大野知代\*\*, 河村圭子\*\*\*

**【要旨】** 我々は解答に対する確信度を加えた新しい解答方法 (N 式) を考案し, 実施してきた。N 式では受験者が解答に自信があるときは答えを一つ, 自信がなく迷うときは正しいと思う方から順位をつけて二つの答えを記入させた。その際, 正答の場合, 配点に差をつけることを明記した。2004 年 2 月の定期試験期間中に A 短大看護学科の 1 年生 85 名, 2 年生 77 名に無記名で, N 式についてのアンケート調査を行なった。回収率は 100% であった。

解答に迷ったとき二つの答えを選べることには 91.3% が賛成であった。解答を一つにするか二つにするか決断しにくいのが 46.9%, 二つの答えを選ぶときの所要時間は短くて済むのが 43.5% であった。N 式は知識の確実性を自己評価する機会になると答えたのが 64.6% であったのに対し, 実力が反映される方式であると認めているのは, 予測に反して 37.0% であった。そのほか配点などについては基本的には現行の方法が支持された。

キーワード: 多肢選択試験, 新解答方式 (N 式), 評価, 受験者, 確信度

### I. はじめに

多肢選択試験のうち四肢択一法は看護師国家試験に長年用いられており, 現在は試験問題のプール制も導入されている。しかし, 正解を四つの選択肢から選ぶ以上, まぐれ当たりを避けることができない。

我々は看護教育機関において, 定期試験に多肢選択試験を取り入れているが, 解答方式に工夫を加えて, その設問に対する解答者の確信度ができるだけ反映される方法 (N 式多肢選択試験, 以下 N 式), 言い換えれば実力反映型多肢選択試験ともいべき解答方法を考案し, 1997 年から採用してきた<sup>1)</sup>。つまり, 解答者がその解答に自信があるときは一つ, どちらか迷うときは, 正答の可能性が高いと思う方から A, B の順位

をつけて二つの解答を選ばせ, それが正答であったときの得点に差をつけることを明示した。したがって解答欄は「自信あり」「迷う A」「迷う B」の三つになる。通常は 40 問を出題し, 「自信あり」で一つだけ選んだ答えが正答であったときは 3 点, 「迷う A」「迷う B」の二つを選んで「迷う A」が正答であったときは 2 点, 本来は間違いである「迷う B」が正答であったときにも 1 点を与えた。今回受験者である学生にこの方式に対する評価を求めて, N 式をさらに改良することを研究の目的とした。

\* 藍野大学

\*\* 藍野学院短期大学

\*\*\* 梅花女子大学

## II. 対象と方法

研究の対象としたのは2004年度に在籍し、N式の試験を3回以上受けた経験のあるA短期大学看護学科の学生である。1年生85名、2年生77名を対象として2004年2月に無記名でアンケート調査を実施した。いずれも定期試験終了後その場で記入の上、直ちに提出してもらった。回収率はいずれも100%であった。1年生はN式による定期試験の直後で、N式は3回目の経験であった。2年生はN式による定期試験は2003年10月が最後であったので4か月後ということになるが、N式の試験は過去に5回経験していた。

## III. 結果

まず、どれが正答か迷ったときに二つの解答を選ぶことについての質問では、「非常に良い」という学生は1年生と2年生の平均（以下1年生と2年生で明らかな差があるもの以外は、回答者数も近似しているので両学年の平均を用いた）が58.2%と過半数を占め、「良い」を合わせると91.3%に達した（表1）。この場合、解答を一つにするか二つ選ぶかの決断については、「決断しにくい」学生が38.3%で、「非常に

決断しにくい」学生を合わせると46.9%と半数近くを占めた（表2）。解答を二つ選ぶときの所要時間は、一つを選ぶときに比べて「やや短い」学生が33.6%、「やや時間がかかる」と答えた人が31.7%で拮抗していた。時間的に「変わらない」という学生は21.7%で、やや少なかった（表3）。

100点満点で40問出題すれば、正答1問についての配点は2.5点になるが、N式では「迷う」の場合、二つ選んだということで、「迷うA」での正答は1問2点にしてある。このことについて2点のときと2.5点のときとで解答を選ぶ数がどのように変わるかを訊ねた。まず、いくつ答えを選ぶか決断できないときは二つ選べる方、つまり「迷う」を選んだ学生が54.7%に対して、一つしか選べない「自信あり」を選んだ学生は27.3%と少数であった（表4）。もし「迷うA」で正答したときの配点が2.5点であったらばどうするかを訊ねたところ、二つ選べる方を選ぶという学生が62.5%になり、7.8ポイント増加した。逆に、配点は最も高いが一つしか選べない方を選ぶ学生は11.9%で、15.4ポイント減少した。

次に、N式について評価を求めたところ、以下のような結果が得られた。この方式が受験者の実力を反映する方式かどうかについては、どちらともいえない

表1 解答に迷ったとき複数の解答を選ぶことは

	1年生	%	2年生	%	1・2年生	%
非常に良い	43	51.2	50	65.8	93	58.2
良い	32	38.1	21	27.6	53	33.1
どちらともいえない	7	8.3	2	2.6	9	5.6
悪い	2	2.4	3	4.0	5	3.1
非常に悪い	0	0	0	0	0	0
合計	84		76		160	

表2 解答を一つにするか二つにするかの決断は

	1年生	%	2年生	%	1・2年生	%
素早く決断できた	4	4.7	2	2.6	6	3.7
比較的簡単に決断できた	16	18.8	16	20.8	32	19.8
普通に決断できた	24	28.2	24	31.2	48	29.6
決断しにくかった	32	37.7	30	38.9	62	38.3
非常に決断しにくかった	9	10.6	5	6.5	14	8.6
合計	85		77		162	

表3 解答を一つ選ぶのに比べて二つ選ぶときの所要時間は

	1年生	%	2年生	%	1・2年生	%
かなり短い	6	7.2	10	13.0	16	9.9
やや短い	29	34.5	25	32.4	54	33.6
変わらない	19	22.6	16	20.8	35	21.7
やや時間がかかる	28	33.3	23	29.9	51	31.7
大変時間がかかる	2	2.4	3	3.9	5	3.1
合計	84		77		161	

表4 「迷うA」での正答が2点の場合、すぐ決断できないときは

	1年生	%	2年生	%	1・2年生	%
「自信あり」を選ぶことが多かった	23	27.1	21	27.6	44	27.3
「迷う」を選ぶことが多かった	46	54.1	42	55.3	88	54.7
どちらともいえない(半々に近い)	16	18.8	13	17.1	29	18.0
合計	85		76		161	

表5 この解答方式は実力が反映される方式である

	1年生	%	2年生	%	1・2年生	%
はい	37	43.5	23	29.9	60	37.0
どちらでもない	39	45.9	36	46.7	75	46.3
いいえ	9	10.6	18	23.4	27	16.7
合計	85		77		162	

表6 この解答方式は知識の確実性を自己評価する機会になる

	1年生	%	2年生	%	1・2年生	%
はい	54	64.3	50	64.9	104	64.6
どちらでもない	21	25.0	19	24.7	40	24.8
いいえ	9	10.7	8	10.4	17	10.6
合計	84		77		161	

表7 「迷うA」での正答は2.5点にすべきである

	1年生	%	2年生	%	1・2年生	%
はい	30	35.3	26	33.8	56	34.6
どちらでもない	21	24.7	26	33.8	47	29.0
いいえ	34	40.0	25	32.4	59	36.4
合計	85		77		162	

表8 「迷うA」で40問全問を正答しても80点なのは不合理である

	1年生	%	2年生	%	1・2年生	%
はい	12	14.1	10	13.0	22	13.6
どちらでもない	29	34.1	23	29.9	52	32.1
いいえ	44	51.8	44	57.1	88	54.3
合計	85		77		162	

表9 「自信あり」で34問以上正答しても99点または100点なのは不合理である

	1年生	%	2年生	%	1・2年生	%
はい	17	20.0	5	6.5	22	13.6
どちらでもない	30	35.3	29	37.7	59	36.4
いいえ	38	44.7	43	55.8	81	50.0
合計	85		77		162	

という解答が最も多く46.3%で、実力が反映されるという回答37.0%を上回っていた(表5)。次に、この方式が知識の確実性を自己評価する機会になるか否かを訊ねたところ、64.6%が肯定し、否定したのは10.6%に過ぎなかった(表6)。

N式の問題点の一つに配点がある。つまり「自信あり」で全問正解すると120点になること、「迷うA」での正解は本来2.5点になることが解決できていない。これらの点について以下のような場合の意見を求めた。まず「迷うA」での正答には2.5点を与える

べきであるという質問については意見が分かれた。賛成、中立、反対が2年生ではほぼ三等分され、1年生では反対が40.0%とやや多く、中立がやや少なかった(表7)。次に、「迷うA」で全問正答しても80点にしかならないのは不合理であるという質問に対しては、54.3%が不合理ではないと答えていた(表8)。逆に、34問以上正答しても99点あるいは100点止まりなのは不合理であるという質問では、50.0%が不合理ではないと回答していた(表9)。

#### IV. 考 察

N式の特徴である迷ったときに複数の解答を選ぶことについては、多くの支持が得られた。設問に答えるのに、解答を一つにするか二つにするかの判断およびそれに要する時間については意見が三つに分かれた。このことは、この方式が煩雑とはいえず、各自手間のかかり方は異なることを示していた。田中ら<sup>2)</sup>は客観試験において解答に確信度を加味し、確信度を5段階の中から選ばせた試験で、解答に要する時間は従来の方法の約1.5倍であったと述べている。しかし、N式では解答とは別に確信度を記入させるのではなく、両者が一体化しているため、特にこの点での問題は無いと思われる。

解答法の説明の欄に「迷うA」で正答のときは2点を与えることを明示しておいたが、「自信あり」を選ぶか「迷う」を選ぶかすぐに決断できなかったときは、二つ答えられる「迷う」を選ぶ学生が54.7%と半数を超えていた。これに関連して、もし「迷うA」で正答したとき2.5点、つまり通常40問の四肢択一方式で与えられる点数が与えられるとしたらどうするかを訊ねたところ、62.5%が「迷う」を選ぶであろうと答えた。つまり、点数が高くなった方を選ぶという回答が7.8ポイント増加していた。その分だけ「自信あり」が減少するわけで、解答に本当に自信のある人が「自信あり」を選ぶことになる可能性が高くなる。現状では「自信あり」を選択しても、今までN式を実施した私立看護系短大（2校）、学年に関係なく正答率が63.2～69.4%と必ずしも高くないので<sup>1)</sup>、正答になったときの配点が多いために「自信あり」の方を選んだ可能性が否定できないからである。しかし、「迷うA」の正答を2.5点にすることは、期待値が「自信あり」を上回ることであり、適切ではない。

N式における期待値を検討した場合、すべて「自信あり」ででたらめな解答をしたときは40問中10問が正答となり30点、「迷う」で同じくでたらめな解答をしたときは、「迷うA」で10問正答の20点に、「迷うB」で残りの30問中「迷うA」で取り上げた選択肢を除いた3肢の中から1肢を選ぶので、10問正答で10点、合計30点となる。自信があってもなくても期待値に差がみられない点が問題となる。しかし、実際にはすべての解答をでたらめに記入する学生はいないと思われ、自信があるときは配点の高い「自信あり」の3点を、自信がないときは正答で少なくとも2点あるいは1点は得られる「迷う」の方を選ぶことは

アンケートの結果からも推測できる。そのため、数字上は期待値に差がなくても、N式は確信度とは無関係の方法ではないと考えられる。

ここで視点を変えて、すべての問題について四肢のうち正答である可能性が高いと思う方から順位をつけて、1問につき二つずつ、つまり必ず「No. 1」「No. 2」の方式で（N式の「迷うA」「迷うB」のように）解答を記入してもらおう方法を考えてみたい。N式の「自信あり」での正答はこの方法でも「No. 1」で正答になるし、「迷うA」での正答も「No. 1」に含まれ、確信度の区別はつかなくなる。一方、「迷うB」ここでは「No. 2」はたとえ正答（当然「No. 1」は誤答）であっても点数は与えない。しかし、正答にあと一步という可能性はある。一方、「No. 1」「No. 2」のどちらも誤答であったときは、4肢中2肢を選んでも誤答ということで、ほとんど理解していないか、まれには質問を取り違えて逆に答えているかであろう。a～dまでの組合せから二つの答えを選ぶ方式では、通常は正しいものの組合せを選ばせるが、誤っている組合せを選ぶという設問にすると、学生により慎重さを求めるということになる。採点の対象は「No. 1」だけなので、得点は通常行われている四肢択一式と同じになる。N式に比べての得点の増減はN式の「自信あり」「迷うA」「迷うB」の正答数に左右される。また、得点分布も高低どちらへも移動する。この方式の欠点は、自信をもって正答は一つと考えている学生に、間違いと分かっている選択肢も一つ選ばせるという不合理さがあり、N式を超える方式ではないと考えられる。

N式が受験者の実力を反映する方式かどうかについては、37.0%の賛同が得られたのみで、我々の目論見が必ずしもその通りには受け取られていなかった。受験者である学生に十分結果が還元されていなかったためとも考えられ、今後の改善点の一つである。これに関連して、N式が知識の確実性を自己評価する機会になるかどうかについては64.6%が賛成していた。解答に自信があったのか、それとも迷ったのか、その結果が正答であったのか、誤答であったのか、端的に言えば一つの答えを選んで正答できたのか、二つの答えを選んだにもかかわらずどちらも誤答であったのかによって、自分が選んだ答えがどの程度信頼性があったかについて考える機会になりうるものと思われる。

最後に、上に述べたようなN式の矛盾点に関して質問をした。まず、「迷うA」は通常的方式ではこれだけを選ぶことになるので、正答のときは2.5点を与

えるという設問に対しては意見が分かれ、一つを選ぶのに比べて、ある程度配点上の差をつけることを可とする意見が多かった。次に、本来ならば「迷う A」で全問正答すれば 100 点なのに N 式では 80 点ということの不合理性については、54.3% の学生が不合理ではないと答えて、N 式の意義を認めているように思われた。これに関連して、「自信あり」で 34 問以上正答しても 100 点止まりなのは不合理であるという回答は 13.6% と少なく、50.0% の学生が不合理ではないと答えていた。100 点を超える学生は少ないので、このような回答になった可能性もある。実際、出題された 40 問のうち半数の 20 問を「自信あり」で正答すれば 60 点（仮に合格ラインとする）になるという学生に有利な、最初からいわゆる「ゲタ」を履かせた方法になっている。これが N 式の矛盾点でもあり、まだ明確に解決できない問題である。

「自信あり」と「迷う」で期待値が同じになる点を改善するためには、「自信あり」で正答のときは加点、誤答のときは減点する方法も考えられる。その場合、「自信あり」で正答するほど加点も大きくなり、100 点を超えるという解決しにくい問題点がさらに広がるので、100 点以上をどう扱うか、合格ラインをどこに設定するかということを決める必要がある。101 点以上は算入されないで、「迷う」を選ぶというバイアスも考えられるが、この件について不合理と答えた学生は 13.6% に過ぎなかった。もっとも、成績の良い学生がそう答えているのかどうかは、無記名アンケートのため推測できない。

N 式は「自信あり」で正答できる学生には、40 問中 33 問の正答で 99 点取れることで有利であり、できない学生には、自信がなく「迷う B」で正答になった分も加算されるため有利であるということから、合格ラインを 60 点に置く限り、成績の良し悪しによって起こる不公平さは特にないものと思われる。

以上のように、N 式は基本的には学生に支持されていると考えられる。今後、受験者の実力をさらに良く反映させるためには選択肢の数を増やすことが一つの解決策である。そのためには、近年英米の医学部の試験に取り入れられ、米国の医師国家試験の一部にも採用されている Extended Matching Items<sup>3, 4)</sup> が参

考になる。この方式は 6～25 肢（あるいはそれ以上）択一方式<sup>3)</sup>というべき方法である。つまり、いくつかの問題に共通した多数の選択肢からなるリストを一つ設け、そこから各問題の解答を選ばせる方法である。Case et al.<sup>3)</sup> は、例えばさまざまな胸痛について診断を求める問題で、医学部の学生には 10 肢、卒後 1 年目のレジデントには 20 肢、循環器科の研修医には更に多くの選択肢を設けて、レベルに応じた調整ができるので、自由に解答する試験と五肢択一試験の良い折衷案であると述べている。N 式に Extended Matching Items の考えを導入したとき、選択肢の数と、迷ったときにいくつまで解答を求めるか、その配点をいくらにするかを検討する必要がある。現在実施している四肢択一方式で a, b, c, d のうち二つの組み合わせから正答を選ばせる形式でも、a～d のすべての組み合わせを用いれば解答は 6 種類になり、四肢択一方式よりも正確な知識レベルを知ることができると思われる。この場合、1 肢だけを選択する問題については選択肢を増やすことで対応できるが、偶然正答になる確率が変わってくる点で、従来の四肢択一方式と異なってくる。この点の解決も図る必要があると思われる。

本論文の要旨は第 36 回日本医学教育学会（高知市）において報告した。

#### 引用文献

- 1) 野村公寿, 足利 学, 大野知代. 学力を反映させた多肢選択試験の検討. 藍野学院紀要 2002; 16: 39-45.
- 2) 田中 勲, 尾島昭次, 原田研介, 畑尾正彦, 石田 清, 石田尚志, 石井哲夫, 川田志明, 斎藤泰一, 梅里良正, 牛場大蔵. 確信度を加味した客観試験解答方法の検討. 医学教育 1990; 21: 329.
- 3) Case SM, Swanson DB. Extended-Matching Items: a practical alternative to free-response questions. Teach Learn Med 1993; 5: 107-15.
- 4) Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, Rubin E. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. Hum Pathol 1997; 28 (5): 526-32.