

## 原 著

# 臨地実習生の質の確保のための看護系大学共用試験（CBT）の開発的研究 — CBT 試験問題の作成とそのモニター試験結果の統計的分析を中心にして —

柳井 晴夫<sup>1)</sup> 亀井 智子<sup>1)</sup> 松谷美和子<sup>1)</sup> 奥 裕美<sup>2)</sup> 麻原きよみ<sup>1)</sup> 井部 俊子<sup>1)</sup>  
 及川 郁子<sup>1)</sup> 大久保暢子<sup>1)</sup> 片岡弥恵子<sup>1)</sup> 萱間 真美<sup>1)</sup> 鶴若 麻理<sup>1)</sup> 林 直子<sup>1)</sup>  
 森 明子<sup>1)</sup> 吉田 千文<sup>3)</sup> 伊藤 圭<sup>4)</sup> 小口江美子<sup>5)</sup> 菅田 勝也<sup>6)</sup> 島津 明人<sup>6)</sup>  
 佐伯圭一郎<sup>7)</sup> 西川 浩昭<sup>8)</sup>

## A Study on the Development of Computer Based Testing(CBT) for a Common Achievement Test for Nursing Colleges in Order to Maintain Students' Competency for Practical Nursing

— With Emphasis on Item Creation and Statistical Analysis of Pilot Testing —

Haruo YANAI, PhD<sup>1)</sup> Tomoko KAMEI, RN, PHN, PhD<sup>1)</sup> Miwako MATSUTANI, RN, PhD<sup>1)</sup>  
 Hiromi OKU, RN, MN<sup>2)</sup> Kiyomi ASAHARA, RN, PHN, PhD<sup>1)</sup> Toshiko IBE, RN, PhD<sup>1)</sup>  
 Ikuko OIKAWA, RN, MN<sup>3)</sup> Nobuko OKUBO, RN, PhD<sup>1)</sup> Yaeko KATAOKA, CNM, PhD<sup>1)</sup>  
 Mami KAYAMA, RN, PhD<sup>1)</sup> Mari TSURUWAKA, PhD<sup>1)</sup> Naoko HAYASHI, RN, PhD<sup>1)</sup>  
 Akiko MORI, CNM, PhD<sup>1)</sup> Chifumi YOSHIDA, RN, PhD<sup>3)</sup> Kei ITO, PhD<sup>4)</sup>  
 Emiko OGUCHI, PhD<sup>5)</sup> Katsuya KANDA, RN, PhD<sup>6)</sup> Akito SHIMAZU, MLI<sup>6)</sup>  
 Keiichiro SAIKI, PhD<sup>7)</sup> Hiroaki NISHIKAWA, PhD<sup>8)</sup>

### [Abstract]

With the rapid increase of the number of nursing universities in Japan, the importance of evaluating the level of practical nursing ability required by graduation is increasing.

Therefore, it is of urgent importance to develop a common achievement test available to nursing universities throughout Japan to evaluate the level of competency and knowledge required for a nursing student to begin clinical practice in hospitals or health facilities.

For the purpose of developing such a test, 1,120 multiple choice items were made from the 18 nursing domains, which are divided into the three areas: basic medicine, basic nursing and clinical nursing.

Then pilot testing was conducted in both paper-based and computer-based forms with the 730 students and 220 students respectively. The total of the alpha reliability of the three areas turned out to be high.

Therefore the usefulness of the computer-based examination was confirmed in comparison with the paper-based test.

**[Key words]** practical nursing, computer based testing (CBT), alpha reliability coefficient, intraclass correlation coefficient

- 
- 1) 聖路加看護大学 St. Luke's College of Nursing
  - 2) 聖路加看護大学大学院博士課程 St. Luke's College of Nursing, Graduate School, Doctoral Course
  - 3) 千葉県立保健医療大学 Chiba Prefectural University of Health Sciences
  - 4) 大学入試センター研究開発部 National Center for University Entrance Examinations
  - 5) 昭和大学保健医療学部 Showa University School of Nursing and Rehabilitation Sciences
  - 6) 東京大学大学院医学系研究科 The University of Tokyo Graduate School of Medicine
  - 7) 大分県立看護科学大学 Oita University of Nursing and Sciences
  - 8) 静岡県立大学看護学部 University of Shizuoka School of Nursing

2011年10月24日 受理

## 〔要旨〕

近年、看護系大学の急増と医療の高度化に伴い、卒業までに取得すべき看護実践能力の評価の重要性が増している。その一環として、臨地実習に入る直前の段階までに看護学生が取得すべき知識・能力を正しく評価しておくことは看護実習の適正化のための急務の課題である。このような状況に鑑み、臨地実習以前に必要とされる知識・能力の到達度を検証することを目的として、看護学領域から1,260の多肢選択形式の設問を作成し、全23大学730名の学生に紙筆形式のモニター試験、および8大学220名の学生に対するコンピュータ試験(CBT: Computer Based Testing)を実施し、そのテスト結果を分析したところ、基礎医学、看護専門科目Ⅰ、Ⅱの合計点の $\alpha$ 信頼性係数は0.9以上の高い値となった。さらにCBT試験と紙筆試験の比較によって、CBT試験の全国看護系大学共用試験としての有用性を確認した。

〔キーワード〕 臨地実習、コンピュータ試験、信頼性係数( $\alpha$ 係数)、級内相関係数

## I. はじめに

1992年の「看護師等の人材確保の促進」に関する法律の制定以降、全国の看護系大学の量的拡大は目覚しく、2011年4月の時点で194校を数え、全看護師養成課程に占める大学における看護師養成率は20%に達している。一方、少子高齢社会の進展など今日の社会変化は著しく、より安全で質の高い医療・看護が求められており、今日、各大学がとり組んでいる看護教育の改革、充実の不断の努力が全大学的規模で行われていく必要がある。

2002年にまとめられた「看護学教育のあり方に関する検討会報告(文部科学省)」においては、「看護実践能力」を向上させる最も有効な手段として「臨地実習」はきわめて重要で、各大学は、「臨地実習」に臨む学生について、当該実習の到達目標に沿った実習開始前の習得レベルの確認、および実習終了後の到達レベルを評価するなど、大学としての評価システムを構築すべきであるということが述べられている。しかし、現状では臨地実習前の学生のレディネスが必ずしも十分でないことも指摘されており、加えて、病院などの施設における学生の受け入れ体制の不十分さ、実習指導者の不足など、臨地実習における環境条件の整備を推進することが求められている。

ところで、2000年3月に文部科学省高等教育局の諮問機関として設置された「医学・歯学のあり方に関する調査研究者協力会議」は、2001年3月に「21世紀における医学・歯学教育の改善方策について—学部教育の再構築のために—」をまとめ、医学・歯学教育のモデル・コア・カリキュラムを提示した。そしてそれに基づいて、臨床実習以前の学生の適切な評価システムの構築のための大学間「共用試験」システムへの参加の意向確認と試験内容、実施方法について検討され、その結果、臨床実習に必要な知識・能力を測るCBT(Computer Based Testing)と客観的構造化臨床試験(OSCE)が2005年12月から行われることとなった。なお、このCBTとOSCEを含めた臨床実習のためのレディネスを測定する

試験が医・歯学部における「共用試験」と呼ばれるものである。そして修業年限が6年間となった薬学部も、2009年から薬学部における「共用試験」を開始した。

## II. 研究方法

### 1. 研究目的

全国の看護系大学で臨地実習生の質の確保のために利用可能な共用試験(CBT)の開発を行い、将来における本格的な導入を見据えた試験の実施のための諸課題を明らかにする。

### 2. 問題作成の手順

現行の医・歯学部の「共用試験(CBT)」を参考に、主に知識・能力(認知能力を含む)等を問う「共用試験(CBT)」のための問題作成にあたり、看護学、および看護学関連領域を大きく、①基礎医学領域6科目(生理学、生化学、解剖学、病理学、微生物学、薬理学)、②看護専門科目Ⅰ7科目(公衆衛生学、基礎看護学、地域看護学、在宅看護学、看護教育学、看護管理学、生命倫理学)、③看護専門科目Ⅱ5科目(成人看護学、老年看護学、小児看護学、母性看護学、精神看護学)の3領域18科目に分類し、本研究の分担者、研究協力者にテスト項目は4肢、または5肢の多肢選択方式で、臨地実習を受ける3年生のうちほぼ70%の正答を得るような問題の作題を依頼した。そして、当該領域の専門家によって作成された問題は他領域の専門家によって問題の適切性、正答が一通りに定まるかについて入念にチェックされた。なお、上記18科目のうち、老年看護学、看護管理学、生命倫理学、小児看護学、母性看護学、および基礎看護学に関する出題経緯とその結果については、文献1)~5)に述べてある。

### 3. データ収集の方法

作成された問題はなるべく多数の項目についての情報が必要であるが、同一被験者に解答させることは時間的

に無理が生じるため、看護専門科目Ⅰ、看護専門科目Ⅱのそれぞれについては、160問からなる3つのセット（計  $3 \times 160 \times 2 = 960$  問）を精選し、研究分担者が所属する大学のうち、実施に関して承諾の得られた23大学730名の大学生を対象として、モニター試験を実施した。さらにこの23大学を設置形態（国立、公立、私立）、規模、地域性を考慮して3グループに分類した。分類結果は、グループ1（9大学、266名）、グループ2（7大学、220名）、グループ3（7大学、244名）であり、それぞれのグループ別に一つずつのセットの問題を解答させた。ただし、これらのグループの問題の難易度、および測定内容はほぼ等しくなるように配慮し、これらのグループ間の学力差の有無を検証するために、基礎医学問題は3グループ共通とした。

この結果に基づき、作成された設問の識別力・困難度を推定し、コンピュータによる「看護系大学共用試験（CBT）」を開発し、前述の23大学のうち、実施するための機器、施設が整っている大学で実施に関する承諾が得られた8校の看護系大学の220名の学生に対してパソコンを用いた「共用試験（CBT）」を実施した。試験問題については、聖路加看護大学に設置したサーバーを通して、各大学に送信された。なお、今回の研究における「共用試験（CBT）」はコンピュータによる多肢選択形式のテストを中心とした知識試験（CBT）の部分開発に限定し、実技試験（OSCE）については今後の課題とする。

倫理的配慮として、モニター試験、CBT試験への参加は23大学の学生の自由意思に基づき行った。なお、本研究は研究倫理審査委員会において承認後（承認番号：10-008）に実施した。

#### 4. 分析方法

(1) 紙筆試験の調査結果は、グループ1、2、3別に、基礎医学、看護専門科目Ⅰ、看護専門科目Ⅱの記述統計（平均値、標準偏差、相関係数）及び $\alpha$ 信頼性係数をもとめる。  
 (2) 設問1つあたりの信頼性係数である級内相関係数（intraclass correlation coefficient）<sup>6,7)</sup>をもとめる。さらに看護専門科目Ⅰ、Ⅱに含まれる科目に、平行とみられる100の設問が含まれているとみなし、そのテストの合計点の信頼性係数を求める。級内相関係数（ $r_I$ ）は1項目、 $\alpha$ 係数はテスト全体の信頼性係数ということになり、テスト全体の項目数が $p$ のとき、 $\alpha$ と $r_I$ の間には次の(1)式および(2)式が成立する。

$$r_I = \alpha / (p + (1-p)\alpha) \cdots (1), \Rightarrow \alpha = pr_I / (1 + (p-1)r_I) \cdots (2)$$

(2)式は $p$ 個の平行となる項目の項目間相関係数が一定値 $r = r_I$ に等しいときの合計点の信頼性係数を示すものでスピアマンブラウン（Spearman Brown）の公式と呼ばれる。(2)式において、 $r_I = 0.1$ のとき、 $\alpha = p/(p+q)$ となり、 $\alpha$ は $p$ の単調増加関数となる。

(3) 問題を解くのに必要とされる能力（ $\theta$ 、平均0、分散1を仮定）を想定し、 $\theta$ の関数として、それぞれの設問 $x_j$ に正答する条件付確率を、 $P_j(x_j/\theta)$ と表現する。この関数を規定する項目の母数と被験者の能力 $\theta$ を推定する。 $P_j(x_j/\theta)$ としては、 $j$ 番目の項目母数を、 $a_j$ （識別力）、 $b_j$ （困難度）によって定義すれば、被験者 $i$ が項目 $x_j$ に正答する確率（ $P_{ij} = 1$ ）は、

$$P_{ij}(a_j, b_j, \theta_i) = 1 / (1 + \exp(-1.7a_j(\theta_i - b_j))) \cdots (3)$$

によって表される。上記関数によって、各設問の識別力 $a_j$ と困難度 $b_j$ を推定する方法が項目反応理論<sup>9)</sup>（Item Response Theory, IRTと略記）と呼ばれるもので、1960年代にLord & Novick<sup>8)</sup>によって提唱され、1980年代以降に実用化されるようになったものである。

項目反応理論の標準的ソフトウェアBILOGの適用により、紙筆試験によるモニター調査データの結果を3グループ別に解析し、上記未知母数である $a_j$ （識別力）、 $b_j$ （困難度）を推定する（ただし、BILOGの場合は、 $a_j$ は非負と仮定される）。

(4) 全部で3領域240項目からなるパソコンによるCBT試験を8大学220名の学生に実施し、その結果を分析し、紙筆試験結果と比較した。さらに、紙筆試験とCBT試験の結果を科目別に分析した。

### Ⅲ. 結果

#### 1. 基本属性

3グループ間の受験者数を設置形態別に表1に示した。

グループ3に公立大学、グループ1に私立大学が多い傾向がみられた（自由度4、カイ二乗値 = 69.05,  $p = 0.001$ ）。

表1 モニター試験（紙筆検査）における受験者構成

	大学設置形態			合計
	国立大学	公立大学	私立大学	
グループ1	2校(43名)	2校(26名)	5校(197名)	9校(266名)
グループ2	2校(45名)	2校(58名)	3校(117名)	7校(220名)
グループ3	1校(33名)	3校(98名)	3校(113名)	7校(244名)
合計	5校(121名)	7校(182名)	11校(427名)	23校(730名)

#### 2. 領域別試験成績

表2に、3領域（基礎医学、看護専門科目Ⅰ、看護専門科目Ⅱ）別の試験成績の平均値を示した。基礎医学は、3グループ別に74.6、72.4、72.7（全体平均は73.29）、平均正答率は46.6%、45.3%、45.5%、看護専門科目Ⅰは、3グループ別に104.7、97.3、101.2、平均正答率は65.4%、60.8%、63.2%であった。一方、看護専門科目Ⅱの平均値は90.5、88.9、91.6、正答率は56.5%、56.2%、57.2%であった。したがって、基礎医学の困難度

表2 3領域試験科目成績のグループ別基本統計量

大 学 グループ	科目名	人数	平均値 (点)	標準偏差	最高点	最低点
グループ1	基礎医学	266	74.56	11.60	111	31
	看護専門科目 I	266	104.65	14.07	130	28
	看護専門科目 II	266	90.54	12.67	120	28
	(I + II) 合計点	266	195.19	24.32	245	56
グループ2	基礎医学	220	72.40	12.41	114	40
	看護専門科目 I	220	97.32(103.20)	12.16(14.35)	122	41
	看護専門科目 II	220	88.92(89.04)	13.41(13.01)	116	41
	(I + II) 合計点	220	186.25(192.24)	23.78	238	82
グループ3	基礎医学	244	72.70	10.85	107	44
	看護専門科目 I	244	101.16 (103.42)	12.80(13.78)	125	48
	看護専門科目 II	244	91.62 (89.22)	11.57(12.31)	117	39
	(I + II) 合計点	244	192.77(192.64)	22.08	235	97

表3 3領域試験科目間相関係数と偏相関係数

グループ1	看護専門科目 I	看護専門科目 II	(I + II) 合計点	全体合計点
基礎医学	0.522(0.160)	0.659(0.472)	0.645	0.827
看護専門科目 I		0.654(0.483)	0.919	0.860
看護専門科目 II			0.899	0.894
総合点 (I + II)				0.963
グループ2	看護専門科目 I	看護専門科目 II	(I + II) 合計点	全体合計点
基礎医学	0.606(0.261)	0.644(0.372)	0.673	0.851
看護専門科目 I		0.729(0.556)	0.922	0.881
看護専門科目 II			0.937	0.906
総合点 (I + II)				0.962
グループ3	看護専門科目 I	看護専門科目 II	(I + II) 合計点	全体合計点
基礎医学	0.539(0.242)	0.613(0.414)	0.643	0.824
看護専門科目 I		0.640(0.466)	0.915	0.864
看護専門科目 II			0.895	0.876
総合点 (I + II)				0.960

表4 グループ別3領域18科目の主成分分析の結果

グループ	第1主成分固有値	第1主成分寄与率	第2主成分固有値	第2主成分寄与率	第1主成分負荷量最大値	第1主成分負荷量最小値
グループ1	5.981	33.27%	1.863	44.62%	0.723	0.291
グループ2	6.314	35.08%	1.499	43.40%	0.734	0.428
グループ3	5.565	30.92%	1.691	40.31%	0.646	0.348

が最も高く、看護専門科目 I が最も低く、看護専門科目 II の困難度は、その中間にあることが示された。

看護専門科目 I、および II について、グループ 2、3 の値をグループ 1 に等化 (equating) して得られた平均値 (修正平均値) を表 2 の括弧 ( ) 内に示した (等化の方法は Tucker の回帰法<sup>10)</sup> による)。等化によって、基礎医学の共通問題では、グループ 1 の平均値がグループ 2、3 に比べ高かったため、看護専門科目 I、II の修正平均値は、グループ 1 の平均値に近い値となった。

### 3. 領域別試験成績の相関係数、偏相関係数

表 3 に 3 グループ別に、基礎医学、看護専門科目 I、看護専門科目 II、看護専門科目 (I + II) 合計点、全体

合計点間の相関係数を示した。全体合計点との相関係数が最も高かったのは、すべてのグループにおいて看護専門科目 II であり、続いて看護専門科目 I、基礎医学の順であった。また、看護専門科目 I と看護専門科目 II の合計点 (I + II) と基礎医学の合計点の間にはグループ別にみると、0.5 ~ 0.6 の相関がみられ、看護専門科目 I と看護専門科目 II のうちでは、看護専門科目 II のほうが看護専門科目 I より、基礎医学と強い相関がみられた。グループ別に、基礎医学、看護専門科目 I、看護専門科目 II の任意の 2 つの科目から他の 1 つの科目の影響を取り除いた偏相関係数 (partial correlation coefficient) を求め、その結果をグループごとに表 3 の括弧 ( ) 内に示した。グループ 1 においては、看護専門科目 I と看護専門科

目Ⅱから基礎医学の影響を除去した偏相関係数は0.160、グループ2、3においては0.261、0.242とやや小さめの値が得られた。このことは、基礎医学の成績は、看護専門科目Ⅰおよび看護専門科目Ⅱの相関に与える影響が強いことを示すものである。

4. 18の教科科目間の主成分分析

18の科目成績についてグループごとに主成分分析を行ったところ、表4の結果が得られた。3グループとも第1主成分の固有値および寄与率もともに大きくすべての科目の主成分負荷量がプラス0.3を超えていた。なお、グループ1の場合、主成分負荷量の最も大きかった科目は、小児看護、つづいて基礎看護学、看護教育学、老年看護学で、第2成分で負荷量の高かった科目は生命倫理学、看護管理学であった。

5. 信頼性係数と級内相関係数による分析

3つのグループ別に計算された看護専門科目Ⅰ（計160問）の信頼性係数を表5に示した。それぞれ、0.871、0.822、0.843であった。同様に、表5から看護専門科目Ⅱ（160問）の信頼性係数は、それぞれ、0.820、0.834、0.792であった。一方、基礎医学（計160問）は3グループ共通の問題であるが、3グループ別に信頼性係数を求めると、0.769、0.799、0.736と看護専門科目Ⅰ、および看護

表5 試験成績の信頼性係数（α係数）

	グループ1	グループ2	グループ3
基礎医学	0.769	0.799	0.736
看護専門科目Ⅰ	0.871	0.822	0.843
看護専門科目Ⅱ	0.820	0.834	0.792
(Ⅰ+Ⅱ)合計点	0.908	0.900	0.890
全体合計点	0.921	0.922	0.907

専門科目Ⅱに比べやや低めとなった。看護専門科目ⅠとⅡの合計点の信頼性係数は、グループ別に、0.908、0.900、0.890となった。さらに、3教科18目全体（480問）の合計点の信頼性係数は0.921、0.922、0.907ときわめて高い値になった。次に各科目別に信頼性係数の値を求めその結果を表6に示した。

表6 基礎医学科目別 信頼性係数（α係数）、級内相関係数、項目数を100にしたときの信頼性係数

	信頼性係数 1*	級内相関係数	信頼性係数 2**
生理学	0.324	0.016	0.619
生化学	0.347	0.021	0.682
解剖学	0.449	0.032	0.767
病理学	0.382	0.020	0.671
微生物学	0.361	0.022	0.692
薬理学	0.577	0.052	0.845

信頼性係数 1\*：α信頼性係数

信頼性係数 2\*\*：項目数100にしたときのα信頼性係数

最も高い信頼性係数を示した科目は、基礎医学6科目の中では薬理学0.577、続いて解剖学が0.449であった。看護専門科目Ⅰ・Ⅱの中で、生命倫理学の信頼性係数は突出して高く、その値はグループ別に、0.944、0.812、0.900であった。看護専門科目Ⅰの中で、生命倫理学に次いで信頼性の高かった科目は看護管理学でその値は0.784、0.753、0.641、続いて、看護専門科目Ⅱのうちで信頼性の高かった科目は母性看護学（0.645、0.660、0.565）、精神看護学（0.637、0.545、0.589）、信頼性の低い科目は成人看護学（0.414、0.416、0.407）であった。表7に示したように、級内相関係数は、生命倫理学、看護管理学を除く10科目のすべてにおいて、0.1未満と小さめの値となった。特に、看護専門科目Ⅰの公衆衛生学、基礎看護学、地域看護学の級内相関係数は低かった。一方、看護専門科目Ⅱにおいては、成人看護学の級内相関係数が最も低

表7 科目別信頼性係数（α係数）級内相関係数、項目数を100にしたときの信頼性係数

科目名	設問数	グループ1			グループ2			グループ3		
		信頼性係数 1*	級内相関係数	信頼性係数 2**	信頼性係数 1*	級内相関係数	信頼性係数 2**	信頼性係数 1*	級内相関係数	信頼性係数 2**
公衆衛生学	20	0.204	0.013	0.569	0.350	0.026	0.727	0.202	0.013	0.568
基礎看護学	35	0.513	0.029	0.749	0.475	0.025	0.719	0.396	0.018	0.647
地域看護学	30	0.510	0.034	0.778	0.298	0.014	0.587	0.367	0.019	0.659
在宅看護学	25	0.567	0.050	0.840	0.479	0.035	0.784	0.581	0.053	0.848
看護教育学	10	0.480	0.085	0.902	0.449	0.075	0.890	0.432	0.071	0.884
看護管理学	20	0.784	0.154	0.948	0.753	0.132	0.938	0.641	0.082	0.893
生命倫理学	20	0.944	0.450	0.988	0.812	0.178	0.956	0.900	0.311	0.978
成人看護学	35	0.414	0.020	0.677	0.416	0.020	0.677	0.407	0.019	0.659
老年看護学	30	0.547	0.039	0.802	0.589	0.046	0.826	0.537	0.037	0.793
小児看護学	30	0.544	0.038	0.797	0.578	0.044	0.821	0.458	0.028	0.742
母性看護学	30	0.645	0.057	0.821	0.660	0.061	0.866	0.565	0.041	0.810
精神看護学	35	0.637	0.040	0.834	0.545	0.033	0.733	0.589	0.099	0.916

\* 信頼性係数 1：α信頼性係数、\*\* 信頼性係数 2：項目数100にしたときのα信頼性係数

く、続いて老年看護学、小児看護学の値も低めであった。

次に、各科目の誤差成分のみが異なる複数個の平行テストの項目数を100とした場合に到達可能な信頼性係数の値を(2)式で示したスピアマンブラウンの公式を用いて計算し、表7に信頼性係数2として示した。

級内相関係数の高い生命倫理学は100個の平行テストを加算することによって $\alpha$ 信頼性係数は0.90台に到達した。一方、看護専門科目Ⅱのなかで、級内相関係数で0.5~0.6台の $\alpha$ 信頼性係数であった老年看護学、小児看護学、母性看護学、精神看護学は100の平行テストを加算することにより、 $\alpha$ 係数は0.75台から0.8台に増加した。

以上、本研究において作成され、モニター調査として実施された、基礎医学160問、看護専門科目Ⅰ(3グループ別)160問、看護専門科目Ⅱ160問(3グループ別)160問、計480の設問の合計点の信頼性係数は0.9を超えた。しかし、科目別の合計点の信頼性係数は、概して低くなった。

### 6. 項目反応理論による看護専門科目Ⅰ、Ⅱの分析結果

項目反応理論<sup>9)</sup>における標準的プログラムであるBILOGによって、上記未知母数である $a_i$ (識別力)、 $b_j$ (困難度)を推定した(ただし、BILOGの場合は、 $a_i$ は非負と仮定される)。

図1-1に、看護専門科目ⅠとⅡに関するそれぞれ160項目についてのグループ別正答率の箱ひげ図を示した。看護専門科目Ⅰの正答率の中央値は70%強、看護専門科目Ⅱの正答率の中央値は60%弱となっている。つづいて図1-2に項目反応理論によって計算された困難度(難易度)の箱ひげ図を示した。中央値はいずれのグループにおいても困難度の低い方向に偏っているが、正答率に比べ、看護専門科目ⅠとⅡの中央値の差が小さくなっている。図1-3には、識別力の値を箱ひげ図で示した。看護専門科目Ⅰの識別力の平均値は0.792、0.582、0.662、看護専門科目Ⅱの識別力の平均値は0.505、0.523、0.507と看護専門科目Ⅰの平均値がⅡに比べ高くなっているが、看護専門科目Ⅰにおいては、特にグループ1と3において、識別力の値が2を超える項目が多く、そのため、中央値は看護専門科目ⅠとⅡではほぼ同一となった。

このようにして計算された識別力と困難度を用い、式(3)により項目特性曲線を描いた(ページ数の制約により、グループ1の結果のみ図示した)。図2-1、図2-2は看護専門科目Ⅰの160問の項目特性曲線、図3-1、図3-2はグループ1の看護専門科目Ⅱの160問の項目特性曲線を描いたものである。

いずれのグループにおいても、看護専門科目の140番~160番の項目特性曲線は識別力が高く、困難度が低いという傾向がみられた。

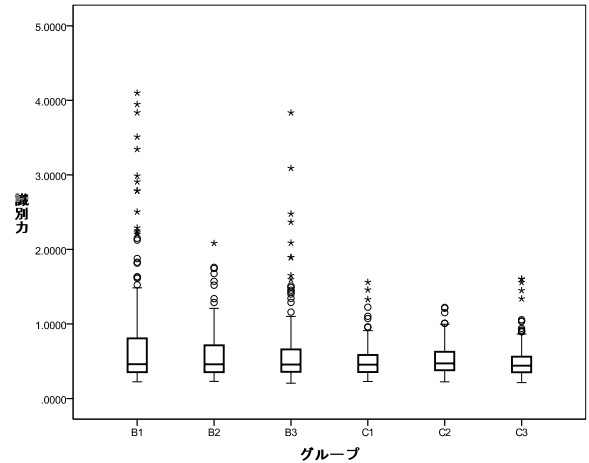


図1-1 看護専門科目Ⅰ(B1,B2,B3)とⅡ(C1,C2,C3)の識別度(グループ別)

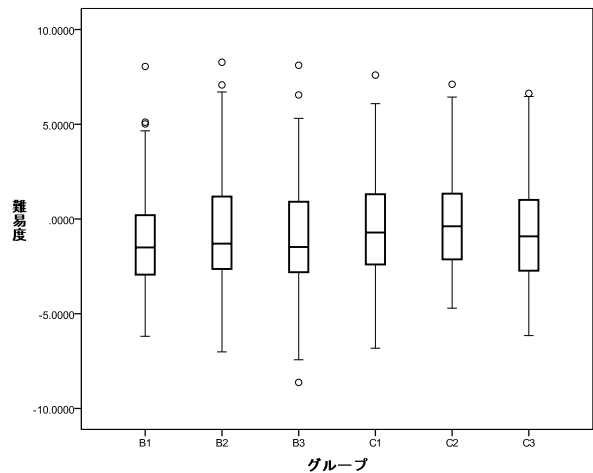


図1-2 看護専門科目Ⅰ(B1,B2,B3)とⅡ(C1,C2,C3)の困難度(グループ別)

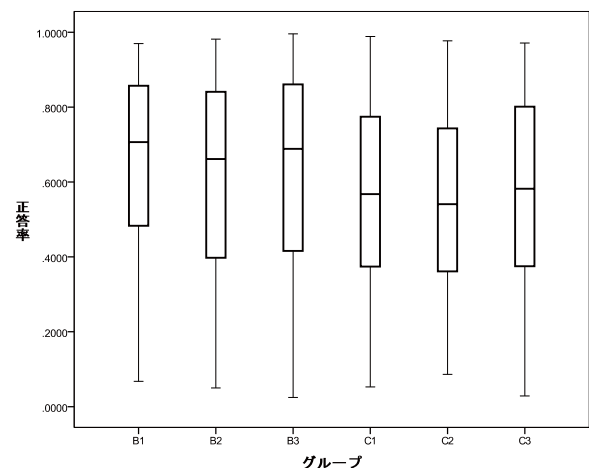


図1-3 看護専門科目Ⅰ(B1,B2,B3)とⅡ(C1,C2,C3)の正答率(グループ別)

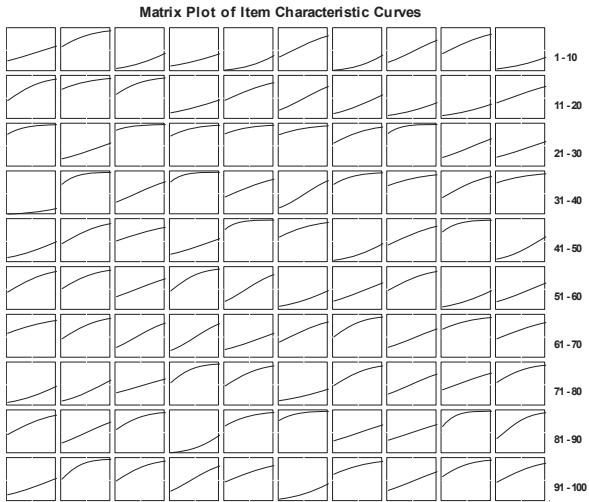


図2-1 項目特性曲線（看護専門科目Ⅰ，問1～問100）

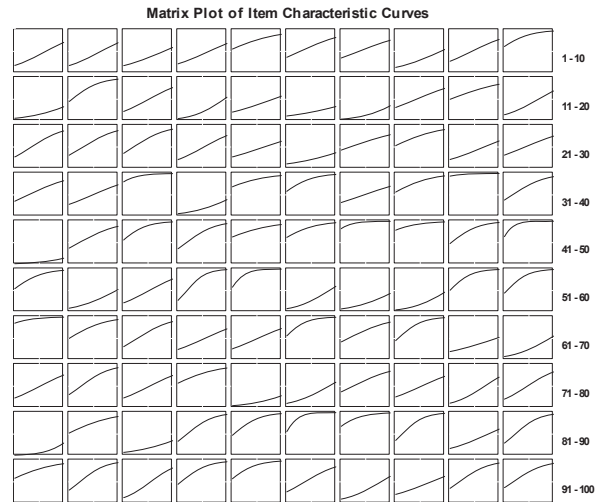


図3-1 項目特性曲線（看護専門科目Ⅱ，問1～問100）

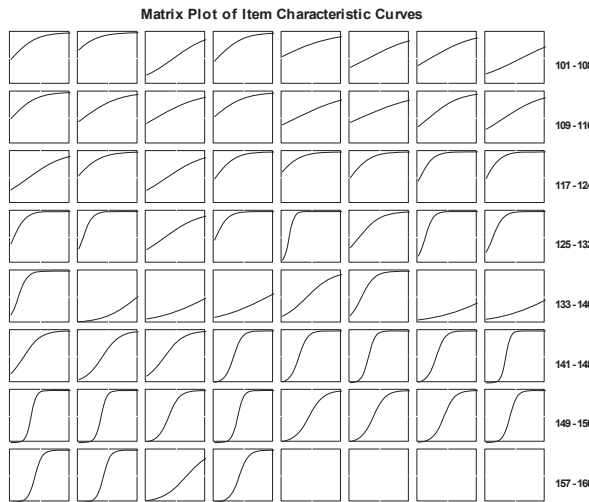


図2-2 項目特性曲線（看護専門科目Ⅰ，問101～問160）

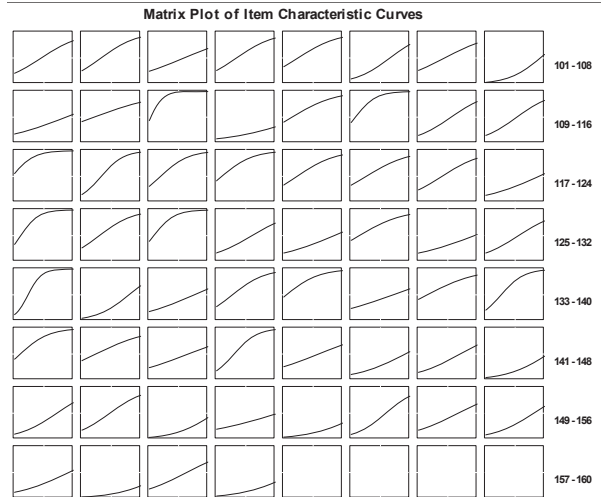


図3-2 項目特性曲線（看護専門科目Ⅱ，問101～問160）

## 7. パソコンを使用した CBT 試験の分析結果

紙筆形式モニター試験に比べ、コンピュータによる CBT 試験問題の困難度は低くなっている傾向がみられた。実際に、試験問題数を 80 とした CBT 試験の基礎医学、看護専門科目Ⅰ、Ⅱの平均値はそれぞれ 43.03、58.90、50.91 となり、この値を 2 倍した 86.0、117.8、102.8 は、表 2 で示したモニター試験（グループ 1）の平均値（74.6、104.6、90.5）を上回っていた。

上記の結果から、個人ごとに選択される問題が異なるパソコンによる CBT モニター試験（2010 年 10 月実施）の結果と、紙筆試験によるモニター試験（2009 年 9～10 月実施）の結果（項目別解答率、項目別年度間相関係数）の間に著しい差はみられなかった。しかし、科目別にみると、紙筆試験に比べ、CBT 試験の方が正答率がやや高くなる傾向がみられた。例えば、CBT になると正答率が高まる科目には、生命倫理学、公衆衛生学、看護管理学、薬理学、成人看護学、老年看護学、小児看護学があり、正答率に差がみられない科目は生化学、解剖学、生理学、母性看護学、基礎看護学であった。

## IV. 考察

### 1. 合計点の信頼性係数

信頼性係数とは、合計点の全分散に対する真の得点の分散の割合を示すものである。一方、 $(1 - \text{信頼性})$ によって定義される非信頼性係数 (unreliability)<sup>11)</sup> は全分散に対する誤差分散の割合に相当する。したがって、信頼性係数が 0.9 であれば、非信頼性係数は 0.1 となり、誤差分散の全分散に対する割合は 10% と低めの値となる。しかし、非信頼性係数が 0.4、0.5、0.6 と増加するにつれ、誤差分散の全分散に対する割合が 40%、50%、60% と増加するので、得られたテスト得点の推定値の信頼区間は広がる。

看護専門科目ⅠとⅡの合計点 (Ⅰ + Ⅱ) の信頼性係数は、グループ別に、0.908、0.900、0.890 となった。さらに、3 教科 18 科目全体 (480 問) の合計点の信頼性係数は 0.921、0.922、0.907 ときわめて高く、十分に実用に耐えうるものであることが判明した。

## 2. テストの信頼性と内容妥当性をめぐって

基礎医学 160 問の合計点の信頼性係数は 0.70 台、看護専門科目 I, II 160 問合計点の信頼性係数は 0.80 台であったので、これらの 3 領域のうち、1 領域のみのテストを用いることは信頼性の観点から推奨されない。実際、各領域の科目別の信頼性係数は、看護管理学、生命倫理学を除くと 0.4 ~ 0.6 というやや低めの値が得られた。各科目の信頼性を高めるには、各項目の主成分分析により、一次元性を確認する必要がある。本研究において、主成分分析を行うと、多くの科目の主成分寄与率は 10% 以下となった。また、主成分分析によって得られた第 1 主成分負荷量はいくつかの項目でマイナスの負荷量が得られ、信頼性係数を低める働きをしていることになる。

しかし、テストの測定理論の立場からは、信頼性をあげることにより、テスト内容の幅を狭めいわゆる内容妥当性を低めるという通信理論の言葉で帯域幅と忠実度のジレンマ<sup>12)</sup>に陥る危険性がある。モニター調査のグループ 1 のデータに含まれる成人看護の 35 項目について主成分分析すると、第 1, 2, 3 の固有値は 2.153, 1.743, 1.654 となり、3 因子が推定された。この結果をプロマックス回転したところ、第 1 因子：肝炎関連、第 2 因子：SLE、メニエール病患者および糖尿病入院患者の生活指導など、第 3 因子：透析患者の合併症、血糖値・インシュリン、検査全般に関する記述、と 3 つの異なる因子が抽出された。成人看護学は、7 ~ 8 冊のテキストからなる広範囲な分野であるため、多因子化したものと推測される。このように、内容的に 3 つの異なる因子に包含される成人看護の設問を 1 つの科目にまとめることの適切性については、今後検討の余地があろう。同様の傾向は、看護専門科目 II において信頼性の低い小児看護学や老年看護学についてもあてはまる。しかし、看護専門科目 I に含まれる生命倫理学、看護管理学については、 $\alpha$  信頼性係数、および主成分分析によって得られる第 1 主成分の寄与率が高く、1 次元性が強い科目といえる。

## 3. 項目反応理論による項目選択の有効性

II. 研究方法の 4. 分析方法で示した (2) 式によって描かれる項目特性曲線 (図 2, 3) が右よりの設問ほど困難度が高く、また、曲線の勾配が大きな項目ほど、識別力が高い項目といわれる。さらに、設問全体の情報量曲線を求めることにより、与えられた項目がどの程度の能力群の被験者をより識別するかについての情報を得ることができる。頁数の制約により本稿で情報量曲線の記述は省いたが、看護専門科目 I, II の各項目は、比較的低い能力群の被験者を識別し、基礎医学は、能力の低い群から高い群にわたって広い能力層を識別する問題が含まれていた。

こういう結果を用いることによって、米国の語学試

験 TOEFL で実施されているようなパソコンによる適応型 (Adaptive) 試験が開発可能となる。2005 年にわが国で開発された医学系共用試験には項目反応理論が適用されている<sup>13)</sup>が、わが国の看護学の分野においては、これまで項目反応理論を用いた分析結果は報告されてない。今後、看護系共用試験の開発に際し、必要となる項目バンクの設計には項目反応理論の利用が必要不可欠となることを強調しておく。

## V. 結語

本研究によって 3 領域 [基礎医学 (160 問)、看護専門科目 I (480 問、160 問 × 3 セット)、看護専門科目 II (480 問、160 問 × 3 セット)] 計 1,120 の設問を作成し、それらを 3 つのグループに分け 23 大学 730 名の学生に紙質検査によるモニター調査を実施した。さらにそのうち 240 問を選んで 8 大学 220 名にパソコンによる CBT 試験を実施した。その主要結果は次の通りであった。

- ①基礎医学、看護専門科目 I, II の計 480 問を合計得点の  $\alpha$  信頼性係数は、3 つのグループのそれぞれにおいて 0.921, 0.922, 0.907 と高い値が得られた。なお、これらの合計点には、看護専門科目 I の方が看護専門科目 II に比べ強い影響を与えていた。
- ②基礎医学、看護専門科目 I, II の各科目の合計点の信頼性係数は 0.7 ~ 0.8 と全体の合計点に比べるとやや低めになる。さらに、科目別に信頼性係数を求めると、生命倫理学が 0.8 ~ 0.9 で高めの信頼性係数が示されたが、他の科目の信頼性係数は 0.2 ~ 0.7 と低い値を示した。
- ③  $\alpha$  信頼性係数は科目に含まれる設問数の増加に伴い増加する傾向がみられるので、1 問あたりの信頼性とみなされる級内相関係数をもとめ、テスト全体が、十分に大きな信頼性係数を得るために必要となる項目数をスピアマンブラウンの式を用いて検討した。
- ④基礎医学、看護専門科目 I, II の各項目について、項目反応理論による分析を BILOG を用いて行った。その結果、看護専門科目 I および II に含まれる各項目の識別力の中央値は 0.5 程度であったが、看護専門科目 I には II に比べ識別力が高い項目が含まれることがわかった。さらに、項目特性曲線の図から、看護専門科目 I の問題番号 141 から 160 (主に、生命倫理学、看護管理学の設問) の項目は、識別力が高く困難度が低いことが読み取れる。
- ⑤各科目から困難度がほぼ等しくなるようなセットを作成しパソコンによる CBT 試験を実施したところ、一見不公平にみえる CBT 試験と紙筆試験の成績に関しては、著しい差がみられないことが確認されたが、その詳細についての検討は今後の課題としたい。



⑥本研究で開発した CBT 試験が全国の看護系大学で利用されるようになるには、④ CBT 試験問題の作成法、⑤ CBT 試験の妥当性および信頼性に関するさらなる情報の蓄積、および、③ CBT 試験の各大学のカリキュラムにおける導入法、についての検討が急務の課題である。

本稿は平成 20 年 4 月～平成 22 年度科学研究費補助金の援助をうけた基盤研究 A「臨地実習生の質の確保のための看護系大学共用試験（CBT）開発研究」の研究成果の一部である。ここに、研究分担者と連携研究者を加えて 32 名（平成 20 年度）、57 名（平成 21 年度）、56 名（平成 22 年度）の共同研究者の先生方に、限らない感謝の意を表す。

なお、本稿の執筆者の役割は以下の通りである。

(1) 執筆（柳井晴夫）、(2) モニター試験問題の作成、老年看護学（亀井智子）、在宅看護学（亀井智子）、看護教育学（松谷美和子）、看護管理学（奥裕美、井部俊子、菅田勝也）、地域看護学（麻原きよみ）、小児看護学（及川郁子）、基礎看護学（大久保暢子）、母性看護学（片岡弥恵子、森明子）、精神保健看護学（萱間真美、島津明人）、生命倫理学（鶴若麻里、吉田千文）、成人看護学（林直子）、公衆衛生学（西川浩昭、佐伯圭一郎）、薬理学（小口江美子）、(3) データ分析（伊藤圭、佐伯圭一郎、西川浩昭、柳井晴夫）

#### 引用文献

- 1) 亀井智子. (2011). 老年看護に関する出題の経緯とその結果, 柳井晴夫編:平成 20～22 年度「臨地実習生の質の確保のための看護系大学共用試験 (CBT) の開発的研究」科学研究費基盤研究 A 総合報告書. 91-94.
- 2) 奥裕美, 井部俊子, 鶴若麻里, 吉田知文. (2011). 臨床実習場面の行動に関する問題—(看護管理学, 倫理学問題) の検討. 柳井晴夫編:平成 20～22 年度「臨地実習生の質の確保のための看護系大学共用試験 (CBT) の開発的研究」科学研究費基盤研究 A 総合報告書. 103-106.
- 3) 小泉麗, 及川郁子. (2011). 臨地実習前看護共用試験問題の開発, 柳井晴夫編:平成 20～22 年度「臨地実習生の質の確保のための看護系大学共用試験 (CBT) の開発的研究」科学研究費基盤研究 A 総合報告書. 107-109.
- 4) 片岡弥恵子, 森明子. (2011). 臨地実習前看護共用試験問題—母性看護—に関する出題の経緯とその結果, 110-111. 柳井晴夫編:平成 20～22 年度「臨地実習生の質の確保のための看護系大学共用試験 (CBT) の開発的研究」科学研究費基盤研究 A 総合報告書. 91-94.
- 5) 安ヶ平伸江, 大久保暢子, 小山真理子, 志自岐康子. (2011). 臨地実習前看護系共用試験問題—基礎看護領域—の開発, 柳井晴夫編:平成 20～22 年度「臨地実習生の質の確保のための看護系大学共用試験 (CBT) の開発的研究」科学研究費基盤研究 A 総合報告書. 91-94.
- 6) Shrout P.E., Fleiss, J.L. (1979). Intraclass Correlations. Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.
- 7) 繫榊算男, 柳井晴夫, 森敏昭. (2008). Q110: 信頼性係数と級内相関係数. Q & A で知る統計データ解析 (第 2 版). 東京:サイエンス社, 2008, 222-224.
- 8) Lord, M. & Novick, M. (1968). Statistical Theories of Mental Test, Reading MA: Addison & Wesley: 1-576.
- 9) 村木英治. (2011). シリーズ行動計量の科学 8 項目反応理論. 日本行動計量学会. 東京:朝倉書店, 1-142.
- 10) 前川眞一. (1999). 第 II 部 3. 得点調整の方法. 柳井晴夫, 前川眞一. 大学入試データの解析—理論と応用 (初版). 京都:現代数学社, 1999, 88-109.
- 11) 水野欽司, 野嶋栄一郎. (1979). 人間科学の統計学 7 テストの信頼性と妥当性. E.G. カーマイン, R. A. ツエラー著. 東京:朝倉書店, 1-102.
- 12) Cronbach, L.J. (1984). Essentials of Psychological Testing, 4th ED. New York: Harper and Row. 1-726.
- 13) 仁田善雄, 前川眞一他. (2008). 項目反応理論を用いた第 1 回共用試験医学系 CBT の統計解析. 医学教育. 36 (1). 3-16.