Development and Meta-Evaluation of ‘Gerontological Nursing’
Test Items for a Nationwide Common Achievement
Test at Nursing Universities: Analysis Using Classical
Test Theory and Item-Response Theory

Tomoko Kamei, Hiroaki Nishikawa, Haruo Yanai

Key words: classical test theory, item response theory, item discrimination degree, item difficulty, gerontological nursing

Abstract

Purpose: We conducted a meta-evaluation of item characteristics such as percentage of correct answers, item discrimination degree and item difficulty using classical test theory (CTT) and item response theory (IRT) for 90 items in a gerontological nursing that created a computer based nationwide common achievement test (CBT) at nursing universities.

Subjects and Methods: We conducted a written monitoring survey by convenience sampling 730 college juniors prior to starting their on-site training at 23 public and private nursing universities. We analysed the percentage of correct answers by CTT, biserial correlations of Item Total (IT), item difficulty by IRT(2PL model), item discrimination degree and factor loading. From these results, we constructed item characteristic curves to evaluate each item’s characteristics.

Results: The average percentage of correct answers for gerontological nursing was 65.8–69.3%. The correlation coefficients were \( r = 0.30–0.41 \) for pharmacology and \( r = 0.28–0.38 \) for anatomy and pathology \( (p<0.01) \). There were large gaps for both item difficulty \( (-5.851–4.068) \) and item discrimination degree \( (0.292–2.218) \). As a result, the information content curves exhibited each item’s characteristics. The item discrimination degrees for 3 out of 90 items were low, which presents a subject for future study.

Conclusion: This study suggests that the remaining 87 items were considered to be usable in CBT however, 71.1% of the items had negative item difficulty, which indicated that they were easy items to answer. There remain some challenges in discriminating highly capable examinees in this CBT items.
要旨

【目的】看護系大学共用試験（CBT）用に作成した老年看護学の出題90項目を古典的テスト理論（CTT）、および項目反応理論（IRT）を用い正答率、および項目識別度、項目困難度などの特性によりメタ評価した。

【対象と方法】便宜的標本抽出により国公私立看護系23大学の臨地実習前の3年生730名を対象として、紙筆試験を実施した。分析は正答率、IT双列相関、項目困難度（2PL model）、項目識別度、因子負荷量を求め、項目特性曲線等を描いてすべての項目特性を評価した。

【結果】老年看護学の平均正答率は65.8～69.3%で、他科目との相関は薬理学$r=0.30\sim0.41$、解剖学・病理学$r=0.28\sim0.38$であった($p<0.01$)。項目困難度（−5.851～4.068）、および項目識別度（0.292～2.218）とも幅が広がった。情報量曲線により各項目の特性が示された。90項目中3項目の項目識別度が低かった。

【結論】90項目の問題中87項目はCBTでの利用が可能と考えられた。しかし、このうち71.1%は項目困難度が負の易しい問題であったため、能力水準の高いレベルの受検者の識別に課題が残った。

I. はじめに

平成元年に看護基礎教育課程に、老年看護学が位置づけられて以来、老年看護学教育は高齢社会のニーズに対応した在宅・入院・施設入所高齢者に対する生活機能のアセスメントと看護の観点からと拡大している（看護基礎教育の充実に関する検討会、2007）。これに伴い、科目教授内容、および実習機関は拡大しており、学生の学習評価を多様な観点から行う必要性が生じているといえる。

一方、CBT（Computer-based testing）はテスト問題群の中からランダムに問題をコンピュータ画面に提示して受検者が回答選択肢の中からマウスやキーボードにより回答し、入力、採点、分析、結果表示などすべてコンピュータを介して行うテスト方法（張、2007；西川、2007）である。1986年に米国で（Crocker & Algina, 1986）、2005年からはわが国の医療学系CBTとして導入され始め（医療学系大学共用試験実施評価機構、2005）、看護系大学のCBTについては、臨地実習前共用試験として開発途上にある（柳井ら、2009；柳井、2010；柳井ら、2012）。看護学のテスト問題（以下：項目）の精度評価に関しては、緩和ケアの知識に関する報告（Nakazawa et al., 2009）があるが、臨地実習前の学部生を対象としたCBT開発に関する報告をほとんどなく、現段階のわが国の看護系CBT開発にとって喫緊の課題であると考える。そのため看護学のCBTに用いるすべての項目特性を評価し、使用し得る項目であるかを精選することは不可欠である。

「大学における看護系人材養成の在り方に関する検討会」（文部科学省、2011）は学士課程のコアとなる看護実践能力と卒業時到達目標を示し、看護系大学では、これをもとに看護実践能力の評価内容を具体化していく必要がある。看護教育は講義と実習で構成（演習・関連実習を講義に含む）されており（厚生労働省、1989）、とりわけ老年看護学実習では保健医療福祉の多様な場で看護を展開する機会となっている。そのため実習開始前に看護系学生自身が自分の看護実践能力を確認し、学習ニーズに気づき、自己理解して実習に臨むことが重要となる。このことは、看護系学生の全体の実習の質を向上することにつながり、それを担保する方法として、看護系大学のカリキュラムに位置づけてCBTを導入していくことは意義があると考える。

本研究プロジェクトでは、看護系大学が共同利用できるCBT開発を進めることを目的として、全国看護系大学教員を対象としてCBT導入の意向を調査し、教員の58%以上が導入に肯定的であることを示した（柳井ら、2009）。次に、基礎医学6科目（生理学、生化学、解剖学、微生物学、病理学、薬理学）、専門科目11科目（公衆衛生学、基礎・小児・母性・成人・老年・精神・在宅・地域看護学、看護教育学、看護管理学）の項目プール1,120題を作成した。これを用いて臨地実習前の学部3年次生に紙筆試験、およびコンピュータを用いたモニタ調査を行い、項目特性とCBT実施上の課題を示し（柳井、2010；柳井ら、2012）、実用化の準備を進めている。

CBTでは項目プールの中から受検者の能力に適した出題が可能であるが（仁田ら、2005）、各受検者が回答するすべての項目の難易度と測定精度を適切に評価しておくことが重要となる（豊田、2011）。

そこで、本研究では看護系大学のための CBT 用に開発した項目群組から、全科目得点の主成分分析による主成分分析率が高値で、科目全体の関連性が高かった老年看護学を取り上げ、両理論を用いて出題したすべての項目特性を評価した。

II. 研究目的

看護系大学 CBT システムの開発用に作成した、四肢択一回答による老年看護学の出題 90 項目について、臨地実習前の学部 3 年次生を対象とした紙筆試験を行い、CTT、および IRT を用いて各項目特性をメタ評価し、CBT に使用するうえでの課題を示す。

III. 老年看護学の項目作成過程

1. 出題枠組み、および項目修正プロセス

看護系大学生が老年看護学臨地実習に臨むために必要な知識、および態度を A:「老年期の心身の加齢性変化、および生涯発達支援（齢別・心理社会的加齢性変化、老年看護の基本等）」、B: 「老年期に多い急性・慢性疾患と治療、リハビリテーションの基本（認知症、骨折、廃用症状群、外科の疾患、慢性疾患、せん妄、摂食障害、薬物療法等）」、C: 「高齢者への看護過程（包括的アセスメント、看護計画、実施、評価等）」、D: 「高齢者看護の態度（コミュニケーション方法、倫理観等）」、E: 「高齢者ケアに関連する保建医療福祉制度の適用」に大別し、これを出題領域として、科目的学年配置は所属大学のカリキュラムを参照した。所属大学は出題時に看護師保健師教育を統合したカリキュラム構成で、大学基準協会相互評価で高く評価され、カリキュラムの質は担保されていると判断したため、科目、教育内容、学年配置について、臨地実習開始前までに履修する内容を参考にした。出題形式は単一問題、連続問題、組合せ回答問題、状況設定連続問題、状況設定組合せ回答問題とし、回答はすべて四肢択一とした。作成は、筆者ら大学看護教育経験 9 ～ 20 年の 2 大学の老年看護学専任教員 3 名が行った。

老年看護学の項目と回答択肢の表面・内容妥当性の検討は、本研究プロジェクトとして参加 12 大学の教員計 39 名が行い、問題文の表現の修正、正答する教員数の確認を行った。なお、生化学は日本医学6学科、公衆衛生学、基礎看護学は看護専門 11 科目の出題も、この教員 39 名が中心的に行い、表面・内容妥当性的確認を行った。

2. モニター調査による項目特性のメタ評価の方法

1) 対象と方法

モニター調査の対象は、全国看護系大学協議会に加盟している大学中、2009年9月までに本研究参加の同意を得た23大学（国立5, 公立7, 私立11大学）の臨地実習前の学部3年次在籍者730名（国立121, 公立182, 私立427名）である。

方法は、各大学の担当教員を通じて、研究協力学生の募集と各大学で試験日を定めて形式別による紙筆試験を行った。学生に過去の時間拘束による負担をかけないこと、すべての科目の出題1,120項目の特性を検討できるようにするため、協力大学を設置形態別に層化し、ラウンドに3グループ（以後、グループ1, グループ2, グループ3）に割り付けた。基礎医学はアンカー科目（すべての受験者が回答する共通問題）とし、専門科目は重複テスト分冊法（日本テスト学会, 2011）により各科目の出題をブロックに分け各グループに480項目をランダムに割り当てた。出題数は学生一人あたり基礎医学160項目（生理学30, 生化学25, 解剖学25, 病理学30, 微生物学25, 薬理学25）、看護専門科目320項目（公衆衛生学20, 基礎看護35, 地域看護30, 在宅看護35, 看護教育10, 看護管理20, 看護倫理20, 成人看護35, 老年看護30, 母性看護30, 小児看護30, 精神保健看護35）
とし、試験時間は休憩を含め4時間とした。

2) 分析方法

CTTによる正答率、アンカー科目との相関、IT 双
列相関（正答、誤答の2値データ、および合計点が2
変量正規分布におよぶ）およびIRTにより項目
識別度（a），項目困難度（b）を分析し、項目特性曲
線（item characteristic curve；ICC）(Hambleton
& Swaminathan, 2010)を描いた。IRTはあら
2つのパラメータによる2次元モデルで表し、2PLモ
デル（two-parameter logistic model）により次式に
によって求めた（P＝正答率，θ＝能力値（−∞＜θ＜∞），
D＝1.7，a＝識別力，b＝困難度，Ij(θ)＝テスト項目
jの情報量，Qj＝誤差率(1－Pj)。

\[
I_j(\theta) = D^2a_j^2 P(\theta) Q(\theta)
\]

\[
P_j(\theta) = [1 + e^{-(D_a(D_b-b)-b)\theta}]^{-1} \quad (\text{村木, 2011})
\]

の

解析ソフトはBLOG-MG3（Scientific Software
International社）を用いた。

3) 倫理的配慮

モニター調査に際しては、協力学生の任意性を保証
した、担当教員自身は協力した学生名を把握しない方
法をとった。結果は成績に影響しないことを説明し、
文書により同意を得た。本調査は所属大学、および各
協力大学の研究倫理審査委員会の承認を得て実施した。

IV. 結果

1. 出題領域別回答形式

老年看護学の出題領域別項目数は、A：老年期の心
身の加齢性変化、および生涯発達支援16 (17.8%),
B: 老年期に多い急性・慢性疾患と治療、リハビリ
テーションの基本33 (36.7%), C: 高齢者者への看護
過程33 (36.7%), D: 高齢者者者の態度6 (6.7%),
E: 高齢者ケアに関連する保健医療福祉制度の適用2
(2.2%)であった。出題形式は単一問題61 (67.8%),
連続問題3 (3.3%), 組合せ回答問題4 (4.4%), 状況
設定連続問題6 (22.2%)、状況設定組合せ回答問題
2 (2.2%)であった (表1).

2. 表面妥当性の検討結果

正答が複数ありうるなど5項目の問題文、または回
答に修正の必要性があげられ、これらを修正した。

3. 項目のモニター調査結果

1) 老年看護学の項目の平均正答率、他科目との関
連

平均正答率はグループ1. 2. 3の順に69.3%,
65.8%, 67.1%, 信頼性係数は0.547, 0.589, 0.537と
差異はほとんどなかった。総得点とアンカー科目得点
との相関係数は、解剖学、病理解、薬理学とは0.28
以上（p < 0.01）、生理学とは0.15～0.33、生化学と
は0.16～0.21であった（p < 0.05：p < 0.01）(表2).

大学設置形態別の正答率は公立68.7% > 国立
67.9%、私立66.9%の順に高かったが、有意差はな
かった。

2) CTT、およびIRTによる老年看護学の項目特性
の評価

全項目の出題領域、出題形式、CTTによる正答率、

| 出題領域 | 項目 | 項目数 | 出題形式 | 項目識別度低値
または項目困難度高値
(因子負荷量<0.3)の項目数 | 修正課題があった
項目番号 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A.老年期の心身の加齢性変化、および生涯発達支援</td>
<td>16</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>1 (6.3%)</td>
</tr>
<tr>
<td>B.老年期に多い急性・慢性疾患と治療、リハビリテーションの基本</td>
<td>33</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>C.高齢者者への看護過程</td>
<td>33</td>
<td>22</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>D.高齢者者者の態度</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>E.高齢者ケアに関連する保健医療福祉制度の適用</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>計</td>
<td>90</td>
<td>61</td>
<td>3</td>
<td>4</td>
<td>20</td>
</tr>
</tbody>
</table>

注）単一：単一問題、連続：連続問題、組合せ：組合せ回答問題、状況設定：状況設定問題、状況設定かつ組合せ回答問題

表1 出題領域出題数、出題形式、および修正課題があった出題

6 日本看護科学会誌 33 巻 2 号 (2013)
表2 グループ別老年看護学の出題の概要分析結果

<table>
<thead>
<tr>
<th>項目</th>
<th>グループ1</th>
<th>グループ2</th>
<th>グループ3</th>
</tr>
</thead>
<tbody>
<tr>
<td>項目数</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>正答率</td>
<td>69.3</td>
<td>65.8</td>
<td>67.1</td>
</tr>
<tr>
<td>正答率最大値</td>
<td>34</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>正答率最小値</td>
<td>10</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>信頼性（クロネバックα）係数</td>
<td>0.547</td>
<td>0.589</td>
<td>0.537</td>
</tr>
</tbody>
</table>

アンカーコンサートに基づいた出題の特徴とその相関 (r) 4)

| 生理学 | 0.23** | 0.15 * | 0.15 * |
| 生化学 | 0.19** | 0.21** | 0.28** |
| 解剖学 | 0.39** | 0.28** | 0.30** |
| 病理学 | 0.36** | 0.28** | 0.30** |
| 微生物学 | 0.24** | 0.25** | 0.28** |
| 薬理学 | 0.30** | 0.30** | 0.41** |

4) r = ピアソン乗積相関係数
*: p<0.05, **: p<0.01

表3 老年看護学項目の古典的テスト理論 (CTT)、および項目反応理論 (IRT, 2-PL model) による出題特性の分析結果

<table>
<thead>
<tr>
<th>項目番号</th>
<th>出題領域</th>
<th>正答率</th>
<th>CTT</th>
<th>IRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001 A</td>
<td>項目1</td>
<td>0.86</td>
<td>0.181</td>
<td>0.800</td>
</tr>
<tr>
<td>1002 A</td>
<td>項目2</td>
<td>0.51</td>
<td>-0.011</td>
<td>0.343</td>
</tr>
<tr>
<td>1003 A</td>
<td>項目3</td>
<td>0.24</td>
<td>0.202</td>
<td>0.549</td>
</tr>
<tr>
<td>1004 B</td>
<td>項目4</td>
<td>0.36</td>
<td>0.122</td>
<td>0.300</td>
</tr>
<tr>
<td>1005 C</td>
<td>項目5</td>
<td>0.79</td>
<td>0.216</td>
<td>0.966</td>
</tr>
<tr>
<td>1006 A</td>
<td>項目6</td>
<td>0.73</td>
<td>0.248</td>
<td>0.588</td>
</tr>
<tr>
<td>1007 B</td>
<td>項目7</td>
<td>0.93</td>
<td>0.269</td>
<td>0.703</td>
</tr>
<tr>
<td>1008 D</td>
<td>項目8</td>
<td>0.87</td>
<td>0.238</td>
<td>0.636</td>
</tr>
<tr>
<td>1009 B</td>
<td>項目9</td>
<td>0.73</td>
<td>0.162</td>
<td>0.598</td>
</tr>
<tr>
<td>1010 A</td>
<td>項目10</td>
<td>0.90</td>
<td>0.109</td>
<td>0.423</td>
</tr>
<tr>
<td>1011 B</td>
<td>項目11</td>
<td>0.68</td>
<td>0.108</td>
<td>0.433</td>
</tr>
<tr>
<td>1012 C</td>
<td>項目12</td>
<td>0.95</td>
<td>0.111</td>
<td>0.526</td>
</tr>
<tr>
<td>1013 B</td>
<td>項目13</td>
<td>0.85</td>
<td>0.285</td>
<td>0.767</td>
</tr>
<tr>
<td>1014 B</td>
<td>項目14</td>
<td>0.29</td>
<td>0.157</td>
<td>0.487</td>
</tr>
<tr>
<td>1015 A</td>
<td>項目15</td>
<td>0.49</td>
<td>0.109</td>
<td>0.243</td>
</tr>
<tr>
<td>1016 B</td>
<td>項目16</td>
<td>0.48</td>
<td>0.104</td>
<td>0.416</td>
</tr>
<tr>
<td>1017 A</td>
<td>項目17</td>
<td>0.51</td>
<td>0.104</td>
<td>0.531</td>
</tr>
<tr>
<td>1018 A</td>
<td>項目18</td>
<td>0.87</td>
<td>0.106</td>
<td>0.512</td>
</tr>
<tr>
<td>1019 A</td>
<td>項目19</td>
<td>0.95</td>
<td>0.124</td>
<td>0.718</td>
</tr>
<tr>
<td>1020 A</td>
<td>項目20</td>
<td>0.95</td>
<td>0.624</td>
<td>2.218</td>
</tr>
<tr>
<td>1021 B</td>
<td>項目21</td>
<td>0.74</td>
<td>0.162</td>
<td>0.546</td>
</tr>
<tr>
<td>1022 B</td>
<td>項目22</td>
<td>0.37</td>
<td>0.143</td>
<td>0.518</td>
</tr>
<tr>
<td>1023 B</td>
<td>項目23</td>
<td>0.95</td>
<td>0.127</td>
<td>0.548</td>
</tr>
<tr>
<td>1024 B</td>
<td>項目24</td>
<td>0.87</td>
<td>0.845</td>
<td>1.141</td>
</tr>
<tr>
<td>1025 B</td>
<td>項目25</td>
<td>0.95</td>
<td>0.401</td>
<td>1.072</td>
</tr>
<tr>
<td>1026 D</td>
<td>項目26</td>
<td>0.96</td>
<td>0.111</td>
<td>0.958</td>
</tr>
<tr>
<td>1027 B</td>
<td>項目27</td>
<td>0.88</td>
<td>0.259</td>
<td>0.707</td>
</tr>
<tr>
<td>1028 B</td>
<td>項目28</td>
<td>0.64</td>
<td>0.301</td>
<td>0.525</td>
</tr>
<tr>
<td>1029 D</td>
<td>項目29</td>
<td>0.47</td>
<td>0.133</td>
<td>0.443</td>
</tr>
<tr>
<td>1030 C</td>
<td>項目30</td>
<td>0.49</td>
<td>0.074</td>
<td>0.372</td>
</tr>
</tbody>
</table>

注1) 大学の3年は課題が出題する出題を含む。
注2) 出題領域（重複する場合は、主な領域とした）: A: 老年期の心身の加齢性変化、および生活発達支援、B: 老年期に多い激変・慢性病の治療、リハビリテーションの基本、C: 高齢者への看護過程、D: 視野看護の態度、E: 高齢者ケアに関する保健医療福祉制度の適用
注3) 主題形式: 単一・単一問題、組合せ問題、回答、連続問題、状態、状態設定かつ連続問題、状態、状態設定かつ組合せ回答問題
注4) IT双列相関係数、IRTによる項目識別度、項目困難度、因子負荷量を表3に示した。
注5) 正答率: 正答率は最小5.3% (No. 1006: 出題領域A; 出題形式単一問題)、正答率90%以上は21項目 (No. 1004, 1012, 1015, 1020, 1026, 1036, 1037, 1040, 1042, 1043, 1044, 1058, 1059, 1060, 1063, 1071, 1073, 1074, 1076)
1078，1085，1087：出題領域A = 3，B = 5，C = 10，D = 3，E = 0；出題形式単一問題 = 15，連続問題 = 2，状況設定連続問題 = 4）が該当した。このほか、正答率30〜80％に39項目（43.3％）が該当していた。
② IT 双列相関：各項目と合計得点との IT 双列相関係数は最少 -0.082（No. 1064；出題領域 B；出題形式単一問題）〜最大 1.034（No. 1040；出題領域 C；出題形式単一問題）に分布し、これも項目による差異を認めた。
IT 双列相関係数 0.3 以上で総得点との関連が強かったのは 34 項目（37.8％；出題領域 A = 4，B = 10，C = 17，D = 2，E = 1；出題形式単一問題 = 27，連続問題 = 1，状況設定連続問題 = 5，状況設定組合せ回答問題 = 1）であった。
一方、IT 双列相関係数 0.1 未満で総得点との関連がなかったのは計 16 項目（17.8％；出題領域 A = 5，B = 7，C = 4，D = 0，E = 0；出題形式単一問題 = 9，連続問題 = 1，組合せ回答問題 = 3，状況設定連続問題 = 3）であった。
③ 項目識別度：No. 1006 は IT 双列相関係数が低値で、項目識別度が算出できなかった。その他の項目では項目識別度は最少 0.292（No. 1064；出題領域 B；出題形式単一問題）〜最大 2.218（No. 1020；出題領域 C；出題形式単一問題）に分布し、項目識別度が 2 を超えたものはこのうち No. 1020 の 1 項目のみであった。項目識別度が 1 以上 2 未満は 16 項目（No. 1004，1012，1015，1019，1024，1025，1035，1036，1037，1040，1050，1059，1063，1073，1078，1087；出題領域 A = 2，B = 3，C = 8，D = 2，E = 1；出題形式単一問題 = 12，連続問題 = 1，状況設定連続問題 = 2，状況設定組合せ回答問題 = 1）で、これらの正答率はすべて 80％以上，IT 双列相関係数はすべて 0.4 以上と高値を示した。
④ 項目困難度：項目困難度の分布は最少 -5.851（No. 1026；出題領域 D；出題形式状況設定連続問題）〜最大 4.068（No. 1064；出題領域 B；出題形式単一問題）であった。項目困難度が正値（難しい）を示したのは 25 項目（27.8％；出題領域 A = 4，B = 15，C = 6，D = 0，E = 0；出題形式単一問題 = 14，連続問題 = 0，組合せ回答問題 = 2，状況設定連続問題 = 8，状況設定組合せ回答問題 = 1）で，負値（容易）を示したのは 64 項目（71.1％；出題領域 A = 11，B = 18，C = 27，D = 6，E = 2；出題形式単一問題 = 46，連続問題 = 3，組合せ回答問題 = 2，状況設定連続問題 = 12，状況設定組合せ回答問題 = 1）で，約 7 割は項目困難度が負である是平易な項目となっていた。
項目困難度 4.0 以上の難しすぎるものは 1 項目（No. 1064；出題領域 B；出題形式単一問題），-4.0 未満の易しすぎるものは 5 項目（No. 1004，1013，1026，1060，1085）（出題領域 A = 1，B = 2，C = 1，D = 1，E = 0；出題形式単一問題 = 2，連続問題 = 1，状況設定連続問題 = 2）であった。
⑤ 因子負荷量，項目識別度，正答率による特性：因子負荷量はテストの質を評価する指標に用い，0.3 以下は質が低いと判断できる。因子負荷量（a）は a = r/√1−r により項目識別度 r = 0.314 に相当する（Hambleton & Swaminathan, 2010）。正答率 30〜80％，かつ項目識別度 0.314 以上に該当したものは 38 項目（42.2％；出題領域 A = 7，B = 13，C = 15，D = 2，E = 1；出題形式単一問題 = 23，連続問題 = 0，組合せ回答問題 = 2，状況設定連続問題 = 12，状況設定組合せ回答問題 = 1）であった。
3）課題を認めめた出題
正答率，項目識別度，因子負荷量が低値か算定不能を示したものの出題領域は表 1 に示した。No. 1006（正答率 5.3％，因子負荷量・項目識別度とも算定不能），No. 1031（正答率 35.5％，項目識別度 0.308，項目困難度 1.994，因子負荷量 0.294），No. 1064（正答率 23.8％，項目識別度 0.292，項目困難度 4.068，因子負荷量 0.281）の 3 項目が課題のある項目として挙げられ，出題領域は A と B である。
No. 1006 は出題領域 A の単一問題である，一般的な加齢性変性を問うものである。「加齢に伴う変化で，適切なのはどれか」に対し 60％の回答者は「筋肉プリント」とし，「心筋の代謝性変化」の回答者は 5.3％であった。これはいずれも正答であり，結果的に回答選択肢に 2 つの正答が含まれたものとなっていた。
No. 1031 は高齢者の聴覚について問うもので，4 種の記述の中から 2 種類の正しい組み合わせを選ぶ出題領域 B の組合せ回答問題であった。回答文に否定形の文章が含まれていたため，受検者を惑わせ，正答率と識別度が低くなっていたようであった。
No. 1064 は無記の感染症，感染経路，発症リスク，症状について問うもので，正しい知識がなければ回答が困難な出題領域 B の単一問題であった。正答率は 23.8％と低く，55.7％の回答者は感染症を誤っていた。項目識別度は 0.292 と低く，項目困難度は 4 を超えて
表4 項目No.1089, No.1064の問題文と回答選択肢、項目情報量曲線および特性曲線

No.1089 「70歳男性、50年後のヘビースモーカーである。最近、散歩時の息切れが激しくなり、咳、痰の症状が強くなった。禁煙をしようと思っていた矢先に多量の血液の嘔出が見られたため総合病院を受診し、胸部X線撮影の結果、肺がんの疑いで、詳細な検査を勧められ入院となった」
No.1089-1 「入院後の検査の結果、肺がんの診断で右上葉切除術が予定された。術前指導で最優先されるのはどれか」（出題領域 C：状況設定連続問題の1題）
回答選択肢 a. ベッド上の体動の指導 回答率（3.3%）
b. ベッド上の体動の仕方の指導 （15.2%）
c. 臥床による食事摂取の指導 （4.6%）
d. 禁煙、横隔膜（腹式）呼吸の指導 ○（79.1%）
（正答率 79.1%、IT双列相関係数 0.372、項目識別度 0.883、項目困難度 1.745、因子負荷量 0.662）

No.1089 項目情報量曲線
No.1089 項目特性曲線

No.1064 「即興について適切なのはどれか」（出題領域 B：単一問題）
回答選択肢 a. 即興の原因は白黒面である 回答率（55.7%）
b. 即興は試と試が接触した場合のみ感染する （4.1%）
c. 副腎皮質ホルモン外用薬の長期使用者では感染しやすい ○（23.9%）
d. 角化型の扁平をおまかな症状は疼痛と丘疹である （18.9%）
非回答 （0.6%）
（正答率 23.8%、IT双列相関係数 0.082、項目識別度 0.232、項目困難度 0.498、因子負荷量 0.281）

No.1064 項目情報量曲線
No.1064 項目特性曲線

項目識別度 0.314 未満であった項目例としてNo.1064（出題領域 B：単一問題）を取り上げ、問題文、回答、項目情報量曲線、および項目特性曲線を表4に示した。項目特性曲線は、横軸を潜在特性能力（Ability）（θ：－3<θ<3）、縦軸を正答率として得られるもので、曲線が右よりの項目ほど項目困難度が高く、項目情報量曲線は勾配が大きいほど項目識別度が高い（Hambleton & Swaminathan, 2010）。
No.1089 は正答率 79.1%、IT双列相関係数 0.372、
項目識別度 0.883，項目困難度 1.745 で，全科目総得点との相関が高く，平易だが，識別度が高い項目で，項目情報量曲線は 0.1 をピークに右下がりを示し，項目特性曲線は，最下位群から最上位群に向けて正答率が急激に右上がりとなり，能力値が 2 － 1 のより低いレベルの能力値の推定に適している項目であった。
一方，No. 1064 は正答率 23.8%，IT 双列相関係数 0.082，項目識別度 0.292，項目困難度 4.066，難しい項目であった，項目情報量曲線は，情報値 0.34 をピークに急激に下降し，項目特性曲線は 0.65 から始まり，能力値の低いレベルの受検者の識別は困難な項目であった。

Ⅴ. 考察
1. 老年看護学の出題領域について
欧米やわが国の医師試 CBT の評価に IRT を用いることが増えてきている（仁田ら，2005），本研究では，看護系大学 CBT 用の老年看護学の出題 90 項目について，CCT，および IRT を用いて各項目の特性をメタ評価したものである。
出題領域は老年看護の臨床実習前の看護系大学生にとって広範囲を網羅していたと考えられ，また，単一問題のみではなく，連続問題，組合せ回答問題，状況設定成績問題，状況設定組み合わせ回答問題などバリエーションのある多様な出題となっていた。
大学設置形態別の老年看護学全体の正答率は公立＞国立＞私立の順であったが有意差はなく，3 グループ別でも老年看護学全体の正答率 65.8～69.3% ではなくまたわずかに，いずれも高値とからえられた，しかし，グループ別の信頼性係数 0.537～0.589 とやや低く，受検者の等質性は臨床実習開始前という点で保持されていたと考えられるため，これ出題の難易度が多様であり，易し問題から難しい問題まで多岐にわたる出題であったため，信頼性係数がやや低くなっていたと解釈できた。

2. 各出題の項目特性について
項目識別度は，能力値と正答率にプラスの強い関連があるかを示すパラメータであり，今回の項目では 0.290 から 2.218 までと幅があったが，90 項目中 87 項目は 0.3 以上を示しており，これらの項目識別度に問題なかった項目は，項目得点がテスト全体で測っている特性を適切に反映し，受検者を区別している項目であり，出題領域や主題形式も大きく偏っていないと考えられた。
一方，No. 1006，No. 1031，No. 1064 の 3 項目識別度が低かったため，回答者の能力の推定を行うことができず，項目情報値が小さいので，テストの信頼性が小さくなると考えられる，そのため，この A と B 領域の 3 題については出題内容や回答選択肢を改善する必要があると考えられた。
項目の難易度を示すパラメータである項目困難度は，およそ 2.0 から 2.0 の間で推定され，2.0 に近い項目は易しく，2.0 に近い項目は難しいと解釈される（豊田，2011；高橋，2002），本分析では項目困難度が負の項目が 71.1% を占めていた，CBT では，難度に高低がある多様な項目がブレーキされている必要があるため，一概にこれらの難度を上げることが必要とは言えず，正答率が 90% を超える単一問題についてはより難易度の高い問題に修正してよいと考える。
項目識別度が低い例として示した No. 1064 は高齢者の伝染性疾患の感染経路や症状などに関して，正確な推論を求める問題であった。一方，項目識別度が高い例として示した No. 1089 は，看護の優先度を問う問題で，対象患者の状況を理解できれば，正答を導くことに困難はない問いであると考えられ，No. 1089 の項目特性曲線は，豊田（2011）が示す A 型に該当し，受検者の特性レベルの全体にわたって識別が可能であることを示していると考えられた。また情報量曲線は，全体の得点に及ぼす個々の項目の影響度を示し，No. 1089 で能力値 2 付近をピークとして，その近辺の能力を有する受検者の能力を把握することができる項目であった。
これらのように，項目難易度の低い項目によるテストでは能力水準の高いレベルにある受検者を十分に識別することができなくなり，逆に項目困難度の高い項目によるテストでは，能力水準の低いレベルの受検者を識別することが困難になる。そのため，すべての受検者に適応するテストにするためには，困難度を多様にする必要がある。また，項目識別度の高い項目をいかに開発するかがテストの信頼性を高めることにつながり，受検者の能力水準に適したテスト問題を個別に出題する適応型テストの開発にもつながると考えられる。
本研究により，出題したすべての項目の特性をメタ評価したことで，各項目の識別度と難易度を明示することができた。これによって制約ある受検時間の中で
各受検者にとって簡単すぎる問題や難しいすぎる問題をあらかじめ排除し、個々のレベルにあたる問題を出題可能になるなど、測定精度を保つため問題数を減らす CBT の開発を行うことが可能となり、作成した問題の利用法を提示することができたと考えられた。

3．研究の限界
今回の調査対象は看護系大学協会加盟校の中から研究参加大学の学生を便宜的な母集団としたため、今後より多くの看護系大学学生を対象として項目のメタ評価を行うことが必要である。

4．今後の課題
試験終了と同時に結果を確認できる CBT の利点を生かし、達成度を直ちにフィードバックし、実習前の学習へと導入する教育への活用方法を検討することが重要である。また、臨地実習で用いる看護実践の確認のために、動画などによる出題を検討する必要がある。今後更なる看護実践能力と出題を対応させるため、臨地実習前後の実践能力の評価が可能となると考えられる。

VI．結論

看護系大学共用試験（CBT）は、日本看護学会の出題計 90 項目を作成し、23 大学の臨地実習前の 3 年次生 730 名を対象としてモニター調査を実施し、CTT およびIRT によりすべての項目特性をメタ評価した。

90 項目中 87 項目は正答率、項目識別度、因子負荷量から評価して問題のない出題であったが、3 項目は課題が認められ、修正が必要であった。今回作成した項目の 71.1% は項目困難度が低い難易度の項目であったため、能力水準の高いレベルの受検者の識別に本 CBT を利用するうえでは課題が残った。

謝辞：本研究にご協力いただいた看護系大学教員の皆様、モニター調査にご協力いただいた学生の皆様に深謝します。聖路加看護大学 祐井文子准教授、千葉県立保健医療大学杉本知子准教授に、問題作成に協力をいただきました。本研究は平成 20 年度～22 年度文部科学省科学研究費補助金基盤研究 (A)（課題番号 20249084）「臨地実習生の質の確保のための看護系大学共用試験（CBT）開発的研究」（研究代表者：柳井晴夫）による分担研究の一部である。

文献
張一平等（2007）：確信度テスト法と項目反応理論、1-4。東京大学出版会、東京。
文部科学省（2011）：大学における看護師系人材養成の在り方に関する検討会最終報告、http://www.mext.go.jp/b_menu/shingi/cho usa/koutou/40/toushin/icsFiles/afiedf10\112003\11/1302921\1_1.pdf（検索日：2011年3月26日）
村本英治（2011）：項目反応理論、50。朝倉書店、東京。
Nakazawa Y., Mitishita M., Morita T., et al. (2009): The palliative care knowledge test: reliability and validity of an instrument to measure palliative care knowledge among health professionals, Palliative Medicine, 23(8), 754-766。
日本テスト学会（2011）：見直そう、テストを支える基本的技術と教育、1-83。金子書房、東京。
西川浩昭（2007）：看護師等国家試験の CBT（Computer-Based Testing）化と CBT の解説、日本赤十字豊田看護大学紀要、3(1), 13-19。
仁井田洲雄、前川真一、柳本武美、他（2005）：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), 3-9。
高橋正誠（2002）：項目反応理論入門、97-150。イデア出版局、東京。
豊田秀樹（2011）：項目反応理論入門編テストと帰納の科学、朝倉書店、東京。
柳井晴夫、鈴木和幸、他（2009）：臨地実習生の質の確保のための看護系大学共用試験（CBT）開発研究のためのアンケート調査結果について、聖路加看護学会誌、13(3), 61。
柳井晴夫（2010）：臨地実習生の質の確保のための看護系大学共用試験（CBT）開発の研究—モニター調査の試験項目に関する統計的分析—、平成 20～22 年度科学研究費補助金基盤研究（A）成果報告書、No.2。
柳井晴夫、亀井智子、松谷美和子、他（2012）：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 試験問題の作成とモニター試験結果の統計的分析を中心にして—、聖路加看護大学紀要、第 38 号、1-9。