Analysis on the Correlation between Scores in Nursing Practices Given by Nursing Supervisory Instructors and Clinical Instructors of the Royal Thai Navy College of Nursing

Kwanta Phabthong a *, Haruthai Ajpru b

a, b Royal Thai Navy College of Nursing, Bangkok 10600, Thailand

Abstract

This is a descriptive research aiming at studying the scoring system of nursing supervisory instructors and clinical nursing instructors in practical application sessions and investigating the correlation between scores given by the two groups. The subjects were nursing supervisory instructors and clinical nursing instructors on wards involved in the practical application sessions of Nursing Care for Persons with Health Deviation II. The six subjects were randomly selected from three wards. The research tool was a 20-item, 5-point scale evaluation form of practical application sessions in Nursing Care for Persons with Health Deviation II. The content validity was validated by an item-objective congruence index which ranged from .67 – 1.00; data were analyzed by descriptive statistics, and inter-rater reliability was tested by an intraclass correlation coefficient. The results showed that the intraclass correlation coefficient between the 6 nursing supervisory instructors and clinical nursing instructors was at .441; there was one item with zero variance, and 6 items with a negative correlation coefficient. The intraclass correlation coefficient among the nursing supervisory instructors was higher than that among the clinical nursing instructors (ICC

© 2013 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and/or peer-review under responsibility of Academic World Education and Research Center.
Keywords: Rater Agreement Index, Clinical Practice Assessment, nursing education;

1. Introduction

Professional nursing is a practice-oriented discipline; therefore the teaching and learning management is an intertwining combination of theoretical and practical sessions. The practice of clinical nursing is therefore the core of nursing education (Chaiyasak, 2000). According to the National Education Act 1999, amended in 2002 (Office of the National Education Commission, 2002), evaluation is a critical aspect that reflects the learners' performance. The data gathered from evaluation is vital for improving students' learning through studying their learning development. At present the main challenges of evaluating the practical application sessions are related to the rater's subjectivity and evaluation methodology and tools; namely 1) inconsistency of evaluation among raters 2) low

* Corresponding name: Phabthong, A. Tel.: +662-475-2614 (22)
E-mail address: eyeangel19@hotmail.com
reliability of results 3) faults in the validity of evaluation 4) inappropriate of score weighing (Schneider, 1979 cited in Rujkorakarn, 1998). To minimize the effect of subjectivity, a multi-rater assessment is applied.

The scoring system implemented in the Royal Thai Navy College of Nursing's practical application sessions involves 2 groups of raters: nursing supervisory instructors and clinical nursing instructors. The former are instructors from the Royal Thai Navy College of Nursing who are responsible for supervising and managing the teaching and learning process conducted in clinics; the latter are nurses on wards who have received training and have been appointed to be clinical nursing instructors. The assessments are conducted separately. There has not been a study related to either the pattern of assessment or an analysis on inter-rater reliability of the two groups. Because of the significance, there should be a detailed analysis on inter-rater reliability between nursing supervisory instructors and clinical nursing instructors in order to be able to develop effective assessment tools to evaluate the nursing students' performance and to incorporate information technology for designing and developing a system of nursing practice assessment.

2. Objectives

2.1 To study the nursing practice assessment conducted by nursing supervisory instructors and clinical nursing instructors in Nursing Care for Persons with Health Deviations.
2.2 To study the intraclass correlation coefficient of the scores given by nursing supervisory instructors and clinical nursing instructors in Nursing Care for Persons with Health Deviations.

3. Research Framework

The study was conducted following Bloom's theory (1979) and a theory on inter-rater reliability (Burry-Stock and et al., 1996)

![Conceptual Framework](image)

4. Research Method

The research study is descriptive. The population were nursing supervisory instructors and clinical nursing instructors on wards involved in practical application sessions of Nursing Care for Persons with Health Deviation II. The 6 subjects were randomly selected from 3 wards.

4.1 Research Instruments

The research tool was a 20-item, 5-point scale evaluation form of practical application sessions in Nursing Care for Persons with Health Deviation II with the item-objective congruence index ranging from .67 – 1.00.

4.2 Data Collection

Data were collected through the score reports of practical application sessions of Nursing Care for Persons with Health Deviation II, submitted by nursing supervisory instructors and clinical nursing instructors on 3 wards. Each
ward consisted of 1 nursing supervisory instructor and 1 clinical nursing instructor. The instructors used the same assessment form for the 3-week nursing practices and conducted the assessment at the end of the whole practice.

4.3 Data Analysis

Data were analyzed using basic statistics which were percentage, average, and standard deviation, and the inter-rater reliability was tested by an intraclass correlation coefficient.

5. Results

All 44 nursing students were female and studying in their third year while the assessment raters were 6 nursing supervisory instructors and clinical nursing instructors. The results of inter-rater reliability were as following:

Table 1  Rater agreement between nursing supervisory instructors and clinical nursing instructors

<table>
<thead>
<tr>
<th>Items</th>
<th>Ward 1 (raters = 2)</th>
<th>Ward 2 (raters = 2)</th>
<th>Ward 3 (raters = 2)</th>
<th>clinical nursing instructors (raters = 3)</th>
<th>nursing supervisory instructors (raters = 3)</th>
<th>Intracllass correlation coefficient (raters = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Timeliness</td>
<td>-.126</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.761</td>
<td>.552</td>
</tr>
<tr>
<td>2. Responsibility</td>
<td>.516</td>
<td>.000</td>
<td>.062</td>
<td>.402</td>
<td>.353</td>
<td>.286</td>
</tr>
<tr>
<td>3. Dressing</td>
<td>Zero variance</td>
<td>Zero variance</td>
<td>.000</td>
<td>Zero variance</td>
<td>Zero variance</td>
<td>Zero variance</td>
</tr>
<tr>
<td>4. Tender</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.424</td>
<td>.139</td>
</tr>
<tr>
<td>5. Communication Skills</td>
<td>.218</td>
<td>.000</td>
<td>-.133</td>
<td>-.799</td>
<td>.330</td>
<td>.308</td>
</tr>
<tr>
<td>6. Confidence</td>
<td>.000</td>
<td>.000</td>
<td>-.878</td>
<td>.000</td>
<td>.264</td>
<td>-.042</td>
</tr>
<tr>
<td>7. Assertiveness</td>
<td>.197</td>
<td>-.365</td>
<td>-.046</td>
<td>-.398</td>
<td>.723</td>
<td>.387</td>
</tr>
<tr>
<td>8. Carefulness</td>
<td>.000</td>
<td>.000</td>
<td>.086</td>
<td>.137</td>
<td>.249</td>
<td>.232</td>
</tr>
<tr>
<td>9. Active</td>
<td>.000</td>
<td>.000</td>
<td>-.003</td>
<td>.146</td>
<td>.260</td>
<td>.287</td>
</tr>
<tr>
<td>10. Loyalty</td>
<td>Zero variance</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.145</td>
<td>.116</td>
</tr>
<tr>
<td>11. Self-Development</td>
<td>-.282</td>
<td>.000</td>
<td>-.265</td>
<td>.000</td>
<td>.448</td>
<td>.083</td>
</tr>
<tr>
<td>12. Mistake Correction</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.055</td>
<td>.033</td>
</tr>
<tr>
<td>13. Attitude on Nursing</td>
<td>.000</td>
<td>Zero variance</td>
<td>.129</td>
<td>.403</td>
<td>.000</td>
<td>.330</td>
</tr>
<tr>
<td>14. Confidentiality</td>
<td>Zero variance</td>
<td>Zero variance</td>
<td>-.292</td>
<td>.000</td>
<td>.000</td>
<td>-.175</td>
</tr>
<tr>
<td>15. Generosity</td>
<td>.000</td>
<td>.000</td>
<td>.192</td>
<td>.165</td>
<td>-.047</td>
<td>-.001</td>
</tr>
<tr>
<td>16. Respect for the patient right</td>
<td>Zero variance</td>
<td>Zero variance</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.284</td>
</tr>
<tr>
<td>17. Accuracy and Appropriate Practice</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>-.813</td>
<td>-.289</td>
<td></td>
</tr>
<tr>
<td>18. Appropriate Management</td>
<td>-.447</td>
<td>.000</td>
<td>.047</td>
<td>-.225</td>
<td>.086</td>
<td>-.597</td>
</tr>
<tr>
<td>19. Performance</td>
<td>.042</td>
<td>.000</td>
<td>-.119</td>
<td>.300</td>
<td>.273</td>
<td>.172</td>
</tr>
</tbody>
</table>
As shown in Table 1, in terms of the intraclass correlation coefficient among the 6 nursing supervisory instructors and clinical nursing instructors, the inter-rater reliability was at .441. Upon closer investigation, it showed that item 1 (timeliness) got the highest score in the inter-rater reliability with item 7 (assertiveness) the second. Only item 3 (dressing) showed zero variance while there were 6 items showing negative reliability: item 6, 14, 15, 17, 18 and 20 (confidence, confidentiality, generosity, accuracy and appropriate practice, appropriate management and writing nursing record).

Considering only the raters who were nursing supervisory instructors, the results showed that inter-rater reliability was at .761. There were 4 items showing zero variance which were items 3, 13, 14 and 16 (dressing, attitude on nursing, confidentiality and respect for the patient right) while there were 3 items reflecting negative reliability which were items 15, 17 and 19 (generosity, accuracy and appropriate practice). On the other hand, focusing on the raters who were the 3 clinical nursing instructors on the ward, the results reflected that inter-rater reliability was at .183, and there were 2 items showing the reliability as higher than .30. The rest showed zero variance or negative correlation.

Focusing on each particular ward, the results reflected that the inter-rater reliability of the nursing supervisory instructor and the clinical nursing instructor on ward 1 was at -.18. Considering each item, there was one item showing reliability higher than .30 which was item no.2 (responsibility). The rest showed zero variance or negative reliability. The reliability of wards 2 and 3 was at -.061 and .03 respectively. There was no item which received a reliability score of more than .30 in these 2 wards.

6. Discussion

As the inter-rater reliability among the 6 nursing supervisory instructors and clinical nursing instructors was at .441, it showed that the reliability of raters' assessment was at 44.10 per cent which can be considered as only moderate. There was one item that showed zero variance which means every rater gave the same score. The 6 items showing negative reliability reflects the wide range of scores given to those items. Some raters gave a high score while the others gave low ones on the same items. Therefore, an understanding of each item needs to be established and agreement on each criterion also needs to be sought. The different contexts of assessment procedures in each ward had affected the performance of nursing students practicing in each ward.

The inter-rater reliability among the 3 raters who were nursing supervisory instructors was at .761 and at .183 among the other 3 raters who were clinical nursing instructors in the ward. This reflected the higher reliability gained from the nursing supervisory instructors. Therefore, in order to judge the learning results, the scores given by nursing supervisory instructors should be weighed higher and put more focus on than those from clinical nursing instructors.

However, the inter-rater reliability in each ward is quite low, and the majority of items received a reliability level at .00. Considering the raw scores, it showed that the raters who were clinical nursing instructors gave similar scores to every student; therefore, there was no deviation of scores in particular items. From the interview with the clinical nursing instructors, they have mentioned that “The work load was too large which prevented them from rating students individually. Therefore, the score was given to groups of students” “There were some difficulties in giving scores for some items” and “They didn't understand some items.” Hence, they rated each student similarly.

Recommendations

1. The Royal Thai Navy College of Nursing should consider amending the evaluation form of practical application sessions, especially the items receiving zero variance and negative reliability in order to use it in each ward in the same way.
2. A clear understanding of each item and criterion also needs to be established among the nursing supervisory instructors and clinical nursing instructors.

References