A STUDY OF INTER-RATER
RELIABILITY USING VARIOUS
PATIENT CATEGORIZATION METHODS

by

Myrna Loy Williams Warnick

A thesis submitted to the faculty of the
University of Utah in partial fulfillment of the requirements
for the degree of

Master of Science

College of Nursing

University of Utah

August 1973
This thesis for the
Master of Science Degree
by
Myrna Loy Williams Warnick
was approved
August 1973
ACKNOWLEDGEMENTS

This writer wishes to extend grateful appreciation to the numerous people who have continually supported and encouraged her during this study: Dr. Marie Holley, Margaret Adamson, Ethel Saunders, and Gwen Luke.

Appreciation is also expressed to Mrs. Verla Collins and the Latter-day Saint Hospital Nursing Administration Staff; Mrs. Esther Sparks and the University of Utah Nursing Administration Staff; and Dr. F. R. Woolley for their suggestions and help.

A special thanks goes to those who sacrificed the most to have this study completed: my parents, Mr. and Mrs. George Williams; my children, Chris, Mark, Marla, Lelen, Curtis, and Travis.

A special thanks is extended to my husband, J. LaMar Warnick, who never waivered in his support and understanding and gave untold hours to this study.
TABLE OF CONTENTS

Acknowledgements ........................................................................ iii
List of Tables ................................................................................ v
List of Figures .............................................................................. vi
Abstract ....................................................................................... vii

Chapter

I. Introduction ............................................................................... 1
II. Method ..................................................................................... 7
III. Results and Discussion ............................................................. 12
IV. Summary, Conclusions and Recommendations ..................... 18

References ................................................................................... 21
Appendices .................................................................................... 23
Vita ............................................................................................... 31
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Raters Profile</td>
<td>8</td>
</tr>
<tr>
<td>2. Summary of Characteristics of Each Method</td>
<td>11</td>
</tr>
<tr>
<td>3. Pearson Product-Moment Correlations Among Six Raters Using the Non-Structured Method</td>
<td>12</td>
</tr>
<tr>
<td>4. Weighted Average Correlations Between Raters</td>
<td>13</td>
</tr>
<tr>
<td>5. Pearson Product-Moment Correlation Among Six Raters Using the Modified Georgette System</td>
<td>14</td>
</tr>
<tr>
<td>6. Pearson Product-Moment Correlation Among Six Raters Using the Pardee Checklist</td>
<td>14</td>
</tr>
<tr>
<td>7. Pearson Product-Moment Correlation Among Six Raters Using the Poland Point System</td>
<td>15</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Inter-Ward and Intra-Ward Comparison of Rater Reliabilities</td>
<td>17</td>
</tr>
</tbody>
</table>
ABSTRACT

Nurses have determined staffing patterns in the hospital environment for years by patient census at its historical peak. Some nursing administrators question this as being the most appropriate way to manage the workload and have suggested identifying patient's needs for care and placing the patients into categories based on those needs. Once the category or classification has been determined, the assigned category for each patient is converted to pre-determined hours of care. This system then is dependent upon the reliability of nurses using the tools for categorization and the validity of the tools themselves.

This study was designed to look at the inter-rater reliability of staff nurses using various types of categorization methods. The tools in the study ranged from a subjective, intuitive nursing assessment, to more structured format with suggested criteria (Modified Georgette), to a checklist format, (Pardee Checklist), to a point system multiple dimension tool with specific criteria, Poland Point System. These methods were used by six registered nurse raters in a general hospital assigned to two medical divisions. Each nurse rated a randomly designated patient using all four of the methods. Pearson product-moment correlations were calculated to determine the inter-rater reliability.

Of the four tools, the intuitive, subjective method of identifying patient needs was the least reliable with correlations of .18 to .85 and an overall mean correlation of .57. The reliability coefficients for the Modified Georgette ranged from .00 to 1.00 with an overall mean correlation of .73 and the Pardee Checklist had correlations between .09 and 1.00 with an overall mean correlation of .70. The Poland Point System, which is the most structured tool and has the greatest specificity of patient needs, was used most reliably as indicated by correlations of .49 to .99 and an overall mean correlation of .83.
Experience seemed to be one factor that influenced the rater's ability to use the various tools. For example, nurses with less than one year of experience were much more consistent with other raters and themselves on the more structured tool, the Poland Point System. Raters with experience of three to twenty-five years and raters with less than one year of experience were not consistent with each other. Further study on the role of experience in assessing patient needs is indicated and deemed necessary to fully understand these findings.

Nurses who assumed administrative roles in addition to their clinical assignment were not as consistent in their use of the various tools with other raters as were those nurses with primary clinical assignments. Further study is suggested on the time involved and influence of administrative tasks on clinical management of care.

Raters intra-divisionally had high reliability coefficients. This study suggested high agreement with other raters on a single division or unit but that low consistency existed between raters inter-divisionally. This certainly suggests the influence of leadership and peer group suggestion on assessment of patient needs. The scores do not suggest whether the influence was positive or not, only that raters on a unit tended to assess a patient in a similar way.
CHAPTER I
INTRODUCTION

Hospital staffing patterns and the utilization of nursing personnel has traditionally been determined by patient census alone. These patterns are usually determined without regard to other important factors which can have an effect on manpower needs. The inpatient census, or the number of patients occupying hospital beds, is highly variable which makes it difficult to predict. Nursing staff has, therefore, been allocated according to historical peak need as perceived by administrators. Many factors including the rising cost of care and cries for accountability have caused many nursing administrators to question the value of this method as an adequate guideline for staffing control.

Poland et al (1970) reported that they were using a new method for measuring patient care based on physical needs, replacing the bed as the basis for planning patient care. Economically, hospital administrators are no longer willing to accept the inpatient census as the gauge for allocation of nursing personnel. They are seeking an accurate, objective measure that will reflect patient’s needs in order to make adjustments in the workload. Georgette (1970) suggested that responsible nursing administrators could benefit by turning to industrial management methodology for sound and scientific control of the workload. Cost control and efficiency techniques of industry have greatly improved manpower allocation problems. These techniques consider not only numbers, but also necessary skills to meet production needs. The need for assessing along multiple dimensions in determining nursing staff needs has been acclaimed by Young (1968, p. 85) in his proposal that:

Traditionally, most nursing unit staffing in short-term hospitals has been guided by rules of thumb that provide fixed amounts of nursing hours per patient day, based on historical measures of peak need as implied by average bed occupancy. Such
procedures usually have not attempted to respond directly to the highly variable demand for care; they were frequently based on relatively long-term estimates of the number of patients to be cared for rather than an actual and immediate aggregate nursing care required by individual patients. A more effective procedure, and one that can be shown to require fewer total nursing hours when confronted by stochastic demand, is to detect and respond cybernetically to increases and decreases in the demand for patient care when and where it occurs within the hospital system.

If staffing needs are to be efficiently determined, then there must be some logical system developed which will render quality care to the patient. A system of allocating nursing personnel can only be as effective as the information received by nursing administration in determining workload management. The number of staff and needed skills must be accurately predicted based on the type and number of tasks which need to be performed for each patient during a given time. If the information provided by nursing is not reliable and does not accurately predict the scope of patient's needs, administrators will look for other ways to analyze the workload. A major problem in determining staff needs is finding instruments, or methods, for nurses to use which are reliable and valid.

A review of the literature reveals that several types of categorization systems have been reported. Many of these systems attempt to give a numerical value to patient's needs. Methods range from short, subjective, intuitive need assessments made by nurses to elaborate checklists of the tasks involved in providing care to the patient.

Wolfe and Young (1965) developed a categorization system which classifies patients into self-care, intermediate care, and intensive or total care categories. The criteria for placement into these three categories is primarily based on the patient's ability to care for himself. With this method, nursing staff identified the primary factors which would best indicate self-sufficiency; namely ambulation, feeding, bathing, and major therapy.
A patient who is mobile who can feed and bathe himself and who can generally take care of his personal needs without assistance is also not likely to be very ill (although in some cases "illness" may not be apparent). He will tend to require somewhat less direct care nursing time than a patient who is bedfast, and who must be assisted in feeding, roughly on the basis of self-sufficiency and avoid the many difficulties associated with attempts to determine just how ill is ill. (Wolfe and Young, 1965, p. 5.)

Historically, the work of Wolfe and Young are among the earliest attempts to define staff needs based on criteria other than occupied beds. Many of the categorization methods discussed hereafter are based on their efforts.

The categorization system developed by Georgette (1970) elaborated on the self-sufficiency concept. Observable nursing behaviors necessary to provide adequate care were the parameters assessed in assigning patient categories. Five parameters, activities of daily living (ADL), grooming, eating, excretion and comfort; general health; treatments; medications; and teaching and emotional support. These are included in each evaluation to determine placement of a patient into a particular category. Criteria have been established for each parameter to help nurses evaluate the patient with less subjectivity. When all the parameters have been assessed, the patient is given a number that is consistent with the category in which he is placed most often, i.e., the nearly self-sufficient patient is placed in Category I with more intensive type of nursing needs indicated in progressive categories, II, III, and IV. With categorization complete the total amount of nursing time can be approximated based on time standards for each category of patient.

The Salt Lake City Latter-day Saints (LDS) Hospital Nursing Administrative Staff (1971) modified the categorization system of Georgette (See Appendix A). Some of the parameters were changed to more closely identify the needs of patients within their hospital. The areas which they included were: activities of daily living, e.g., personal
hygiene, diet, turning and/or assisted activity and excretion; diagnostic evaluation; medications; treatments and emotional support; and teaching. As with the Georgette System, detailed criteria were established for inclusion of patients into certain categories. The nurses assess the patient's needs and assign the patient to categories I - IV, category I indicating self-sufficient patients to category IV for patients needing complete nursing care. Again, the number of hours of care allotted to each patient is based on the category in which the patient was placed. As applied by LDS Hospital, care needs are reassessed prior to each oncoming shift. Personnel allocation is completed after the information to classify patients needs is converted by nursing administration into hours of nursing care needed by each nursing unit.

White, Quade and White (1967) completed a study using self-care, intermediate care, and intensive care (strict and moderate) categories. Physician estimates of patient care needs were included in this study as an added measure. White noted that there was agreement between physicians and nurses on patient classification in the intensive care category, but there was a lack of agreement in nurse-physician classifications of intermediate care and self-care classes. These differences were explained as follows:

Nurses may be placing a large proportion of patients in intermediate care because of present checklist criteria, which are based on objective nursing tasks. In contrast, the classification by physicians is based largely on subjective definitions which perhaps relate more closely to a general impression of the patient's degree of mobility. (White, 1967, p. 3.)

Since nurses and physicians found that they were unable to agree on specific patient problems when they were evaluated separately, attention was turned to overall implications of the nursing criteria being used. In doing so, specific nursing criteria were identified as being of greatest significance in classification and these were chosen for analysis. For example, the bath was singled out as the best single criteria for distinguishing self-care patients from others. To produce a workable list of measures, five elements were extracted from the original list. These were: temperature, pulse, respiration and/or blood pressure, oxygen therapy, suction, cleanliness, and dietary.
Poland (1970) further developed these five criteria, resulting in two additional dimensions: turning and/or assisted activity and toileting-output (See Appendix B).

In order to obtain greater precision in determining staffing needs, nurses and assistants were timed while they performed the various tasks listed under the seven categories. Using the observations of many nursing personnel over several days, average time periods were obtained for each task. The seven levels of patient care needs were broken down into subcategories with point values. The number of points a patient receives may then be converted to nursing care hours.

Pardee (1970) has also established criteria to assess patient’s needs utilizing a checklist format. The criteria used are virtually the same as White, namely pulse and respiration and/or blood pressure, personal hygiene, activity, diet, oxygen, therapy and medical treatments. Nursing divisions adjust the criteria to meet their own particular needs. The criteria used on a surgical unit are different from the neurology unit which are, in turn, different from the medical unit (See Appendix C).

The ward clerks are responsible for categorizing patients every eight hours, basing their data on information obtained from the nursing care plan. Because of limited clinical background, criteria used with clinical personnel were stated as simply as possible. Categorization is begun with the clerk marking on a scale of one to three for each of twenty-one need areas. A category is then assigned the patient according to the category with the most number of checks. An unusual feature of the Pardee Checklist was the predetermination of care of certain patient conditions. For example, on a surgery day a patient is automatically given a level three check by the clerk, if he is being suctioned a level two or three check is required.

Nearly every system or method reported in the literature uses some method of quantifying patient needs on which to predict nursing workload. This figure must be based on a realistic evaluation of patient care needs rather than patient census. Identifying a system and a way to utilize that system that is both economical and valid which
can be used reliably by registered nurses is of prime importance. Hospital nursing ad-
ministrations cannot count on information from any system until it has been shown to
be reliable as well as valid in its use. It is assumed that all of the above mentioned
methods, if properly applied, have some validity and represent substantial improvements
over determining staffing needs using traditional patient census methods. A question
of prime importance is which of the methods can be applied most reliably and efficiently
with the existing variability among the nursing staff who collect the data.

Thus, the purpose of this study is to determine the inter-rater reliability among
registered nurses using four categorization methods; e.g., the Modified Georgette, the
Poland, the Pardee, and an intuitive nursing assessment which approximates current
practice.
CHAPTER II

METHOD

Study Setting. This study was conducted in a 284-bed university hospital located in Salt Lake City, Utah. The hospital is comprised of four intensive care units, two medical units, two surgical units and six additional specialty areas.

The hospital, associated with the University of Utah College of Medicine, has an atypical patient population for several reasons. The factors which influence the number and type of patients encountered include: teaching needs of the University, serving as a regional hospital to a large portion of the Intermountain West, providing specialty services not available elsewhere, and serving as a major provider of medical services in the Salt Lake City area. These factors often create special problems in anticipating and providing adequate nursing coverage.

Administrative personnel in the nursing service at University Hospital also hold faculty appointments with the University of Utah College of Nursing. This affiliation brings an unusual added dimension not found in many hospitals. These nursing administrators are well aware of current thinking and trends within academic circles as well as the practical application of new ideas to patient care. The result is a dynamic staff who are aware of the necessity for research and evaluation and are willing to accept responsibilities in these areas. Excellent cooperation in performing this study was obtained from Nursing Administration staff members and patients.

The study was conducted during the month of July, 1972, on two divisions, caring for medical patients, with a total of 60 beds. A heavy patient load was encountered during the period of the study with census running at over 95%. The six members of nursing staff used as raters in this study were permanently assigned to one of these divisions. The head nurse on each division was responsible for assigning personnel and completing work schedules. If the division was short of personnel, either because of illness or needs
of patients exceeding what the assigned staff could provide, a manpower resource pool from the nursing office supplied additional help. The head nurse or in some cases a staff nurse could request help from the resource pool. These requests were based on experience and intuitive assessment of the staff needed to provide adequate care in that situation. Although the nursing staff on the medical divisions had been oriented to a categorization similar to Georgette's method for determining staffing needs, it was not being used at the time of the study.

Raters. Nine registered nurses were selected to participate in the study. The initial group of raters comprised all the nurses on the two divisions. However, only six nurses could meet the criteria below. Minimum criteria for the raters were established as follows:

1) Graduate of either a baccalaureate, associate or diploma nursing program.
2) Licensure as a registered nurse, (R.N.)
3) Commitment to complete data collection using all four methods.
4) Time available to attend orientation classes.
5) Scheduled on either day or afternoon shifts during the period of the study.

Table 1 presents a summary of the education, experience, position and shift assignment for each of the six raters completing the study.

<table>
<thead>
<tr>
<th>Rater No.</th>
<th>Degree</th>
<th>Years Exper.</th>
<th>Position</th>
<th>Normal Shift</th>
<th>Completed Observations By Raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B. S.</td>
<td>10</td>
<td>Head Nurse</td>
<td>Days</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>3-yr. Diploma</td>
<td>25</td>
<td>Team Leader</td>
<td>Days</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>2-yr. Associate</td>
<td>3</td>
<td>Team Leader</td>
<td>Rotates</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>2-yr. Associate</td>
<td>¾</td>
<td>Team Leader</td>
<td>Afternoons</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>B. S.</td>
<td>5</td>
<td>Head Nurse</td>
<td>Days</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>B. S.</td>
<td>1</td>
<td>Team Leader</td>
<td>Days</td>
<td>7</td>
</tr>
</tbody>
</table>
It can be seen that half of the group held bachelors degrees, and that the raters' experience ranged from about six weeks to over twenty-five years. The average for the group was approximately 14.5 years. Most of the data were collected during day shift.

Methods for Categorization. Four different categorization methods were used by each of the raters in the study. A complete copy of each method or protocol is shown in Appendices A, B, and C. Method Two, Three, and Four correspond to the Modified Georgette (developed at Salt Lake LDS Hospital), Pardee Checklist, and Poland Point System respectively. Method one was virtually unstructured, using nurse's overall intuitive judgement to rate patient care needs on a one-to-three scale. A fourth level of need was available to describe long-term or chronic patients. This level was used on a limited basis, and does not represent a true end point of the continuum represented by levels one, two, and three. Method one was an attempt to quantify current practice then in effect at University Hospital. In practice, the patients were rated on an alphabetic scale (A, B, or C) which was later converted to numerics for analysis.

Table 2 summarizes the major features of each of the four methods used in the study.

Procedure. Each day, one patient was selected according to a random table of numbers, to be categorized on each of the four tools. The ratings were completed at the change of shift when the greatest number of registered nurses could evaluate the patient within a reasonable period of time. The time lapse between the first observation to the last was held to less than two hours. Each rater evaluated the patient independently utilizing all four tools. The number of raters varied each day according to the number of nurses scheduled between the two divisions.

After the first series of observations were made, any patient who had a sudden change in condition such as, vital signs becoming unstable or the nursing or medical plan of care altered, were eliminated from the study. If, however, the patient's condition was unstable at the outset and remained so, his data was retained. Four patients were eliminated because of inadequate data collecting or unstable patient conditions. A total of sixteen patients were assessed and categorized. Categorization
of the patients was completed during a one-month period excluding weekends when there was neither a representative staffing pattern nor patient load.
<table>
<thead>
<tr>
<th></th>
<th>TYPE</th>
<th>NUMBER OF CARE AREAS EVALUATED</th>
<th># OF CATEG. LEVELS</th>
<th>METHOD FOR DETERMINING STAFF NEEDS</th>
<th>MAJOR AREAS REVIEWED</th>
<th>CRITERIA FOR CATEG.</th>
<th>METHOD OF REPORT. DATA</th>
<th>PROTOCOL FORM USED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Unstruct.</td>
<td>Unspecified, essentially 1 overall</td>
<td>3 Plus 1</td>
<td>Estimated from exper.</td>
<td>Unspecified</td>
<td>None</td>
<td>Overall impression (A,B,C,D)</td>
<td>None</td>
</tr>
<tr>
<td>2.</td>
<td>Modified</td>
<td>7</td>
<td>4</td>
<td>Calculated from overall category level</td>
<td>Hygiene, diet turning and/or assisted activity excretion, diagnostic evaluations, medications, teaching and emotional support.</td>
<td>Extensive</td>
<td>Average level of care (I, II, III, IV)</td>
<td>None (referred to list)</td>
</tr>
<tr>
<td>3.</td>
<td>Pardee</td>
<td>Up to 21</td>
<td>3</td>
<td>Calculated from overall category level</td>
<td>Diet, activity, vital signs, IPPB, urine analysis, monitor, dressings and drains, suction, isolation, incontinence, admit, dialysis, surgery or major exam</td>
<td>Minimal</td>
<td>Average level of care (I, II, III)</td>
<td>Used special medical form (new form for each patient)</td>
</tr>
<tr>
<td>4.</td>
<td>Poland</td>
<td>7</td>
<td>Varies from 1-12 points</td>
<td>Minutes of care needed calculated from points received</td>
<td>Respiratory aides, suction, cleanliness, turning and/or assist activity, diet, toilet-output, vital signs and measurements</td>
<td>Moderate</td>
<td>Total number of points</td>
<td>Circed points on checklist (new form for each patient)</td>
</tr>
</tbody>
</table>
CHAPTER III
RESULTS AND DISCUSSION

Analyses of Data. The data were analyzed at the University of Utah Computer Center (UU/CC) using a Univac 1108 computer and the UU/CC library program CORREL, written by Edward Sharp. Pearson Product-moment correlations were calculated between each of six raters on a given method. The results are four six-by-six square inter-correlation matrices, representing the four methods. Since all matrices obtained are symmetrical only the lower half is displayed.

Pearson Product-moment correlations between the six raters on tool number one, the unstructured intuitive nursing assessment, are displayed in Table 3.

TABLE 3
Pearson Product-moment Correlations among six Raters using the Non-Structured Method.

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.65(9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.51(11)</td>
<td>.85(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.63(10)</td>
<td>.85(4)</td>
<td>.76(8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.18(9)</td>
<td>.65(5)</td>
<td>.73(7)</td>
<td>.39(6)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.26(7)</td>
<td>.58(6)</td>
<td>.50(3)</td>
<td>.00(4)</td>
<td>.85(5)</td>
</tr>
</tbody>
</table>

( ) = Number of patient assessments used to calculate correlation

The correlations ranged from .00 between raters four and six to .85 between raters two and three, three and four, and five and six. These correlations indicated a high degree of variability among raters.

Since dimensions of patient care assessed by tool one were not well-defined with a particular format, the categorizations were probably on estimates from previous experiences which also varied considerably. In looking at the correlations between the raters...
using tool one, raters two and three with more than three years of experience used this method most reliably or consistently. The two raters with less than one year of experience used this intuitive method least reliably with a correlation of .00. The correlations suggest a marked increase in reliability with increased experience.

Weighted mean correlations were calculated on each of the four tools to provide an estimate of the overall agreement between the six raters using each method. These means were calculated by multiplying the number of observations made in determining each correlation coefficient with the correlation coefficient itself, then adding the products and dividing by the total number of observations. The resultant numbers are shown in Table 4 which provides a simultaneous, reliability score among raters on all of the methods.

TABLE 4
Individual and Overall Weighted Mean Correlations on Four Categorization Tools.

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Overall Weighted Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.46</td>
<td>.66</td>
<td>.69</td>
<td>.57</td>
<td>.52</td>
<td>.44</td>
<td>.57</td>
</tr>
<tr>
<td>2</td>
<td>.68</td>
<td>.80</td>
<td>.87</td>
<td>.72</td>
<td>.77</td>
<td>.49</td>
<td>.73</td>
</tr>
<tr>
<td>3</td>
<td>.72</td>
<td>.67</td>
<td>.88</td>
<td>.66</td>
<td>.83</td>
<td>.37</td>
<td>.70</td>
</tr>
<tr>
<td>4</td>
<td>.82</td>
<td>.90</td>
<td>.89</td>
<td>.90</td>
<td>.69</td>
<td>.82</td>
<td>.83</td>
</tr>
</tbody>
</table>

As indicated in Table 4, there is an increase in the overall weighted mean from .57 on tool one to .73 on tool two. The increased structure of tool two apparently facilitated the consistency among raters. It is of importance to note that the two raters with the least experience did no better with this tool than on tool one. Only those nurses with experience attained greater consistency with this tool.
Table 5 displays the intercorrelations between raters on tool two, the Modified Georgette. The range of correlations from .00, raters four and six, to 1.00, raters three and six indicates great variability in the overall use of the tool by all six raters. The two raters with the least experience, four and six, had the correlation of .00.

### TABLE 5

Pearson Product-moment Correlations among six Raters using the Modified Georgette Method.

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.83(9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.84(11)</td>
<td>.89(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.81(10)</td>
<td>.94(4)</td>
<td>.81(8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.50(9)</td>
<td>.97(5)</td>
<td>.90(7)</td>
<td>.78(6)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.32(7)</td>
<td>.45(6)</td>
<td>1.00(3)</td>
<td>.00(4)</td>
<td>.84(5)</td>
</tr>
</tbody>
</table>

( ) = Number of patients assessed to calculate correlation

Correlations between the six raters on tool three, the Pardee Checklist extend from -.09, raters one and six, to 1.00, raters three and six. The weighted mean correlation for this tool was .70 which is about the same as for tool two (.73). However, there was greater variability between raters as indicated by the wider range in correlations as demonstrated on Table 6.

### TABLE 6

Pearson Product-moment Correlations among six Raters using the Pardee Checklist.

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.89(9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.90(11)</td>
<td></td>
<td>1.00(4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.76(10)</td>
<td>.33(4)</td>
<td>.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.92(9)</td>
<td>.87(5)</td>
<td>.92(7)</td>
<td>.71(6)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-.09(7)</td>
<td>.20(6)</td>
<td>1.00(3)</td>
<td>.58(4)</td>
<td>.67(5)</td>
</tr>
</tbody>
</table>

( ) = Number of patient assessments used to calculate correlation
The two nurses with minimal experience improved their correlations from .00 on tools one and two to .58 on tool three. According to these statistics nurses with less experience do better on a highly structured tool. Those nurses with three to 25 years of experience also used the tools consistently with correlations between .87 and 1.00. The low reliability coefficients occurred between those nurses with little experience and the nurses with over three years experience. They seemed to view patient's needs differently.

Table 7 shows the inter rater correlations obtained with tool four, the Poland Point System. These correlations ranged from .49 to .99. The overall mean for this particular tool showed an increase from .53 on tool one, .73 on tool two, and .70 on tool three to a mean of .83 (See Table 4).

### TABLE 7

Pearson Product-moment Correlations among six Raters using the Poland Point System.

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.81(9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.93(11)</td>
<td>.99(4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.87(10)</td>
<td>.98(4)</td>
<td>.94(8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.62(9)</td>
<td>.85(5)</td>
<td>.67(7)</td>
<td>.83(6)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.84(7)</td>
<td>.95(6)</td>
<td>.97(3)</td>
<td>.91(4)</td>
<td>.49(5)</td>
</tr>
</tbody>
</table>

( ) = Number of patient assessments used to calculate correlation

There is greater specificity, or structure, written into this particular tool for identifying patient needs. The other methods had restricted ranges of patient assessment criteria which must be considered as a possibility for allowing more variability to occur than with tool four.

The two nurses with one year or less experience were able to use this method with a correlation of .91 suggesting that the more structure and more specificity a tool has, the more reliability it can be used by those with minimal experience. The nurses with
experience used this tool consistently with correlations ranging from .62 to .99. This particular tool also seemed to lend itself to better utilization between the more experienced nurses and those with minimal experience as indicated by the decreased range of the correlations on tool four.

To look at an individual's overall performance with each of the four methods, an estimate of the raters overall reliability was calculated into a weighted mean score. This was done by multiplying the correlation coefficient with the number of observations. This is demonstrated on Table 4.

Rater six, a baccalaureate graduate nurse with one year of experience lacked consistency with all other raters as demonstrated by her overall mean correlations on the tools of .44, .49, .37, .82. This study did not attempt to identify individual motivation, attitude toward the study, or clinical assessment skills which could have been the basis for the low reliability coefficients of this rater.

There were high correlations, .73, .85, .75 and .87, among all nurses within each division, as indicated on Figure 1. This division consistency seemed to indicate nurses of one division assess patient needs and categorize similarly. Whether they all categorize the patient's needs adequately was not determined in this study. They have simply assessed the needs consistently with each other. Inter-division scores however, indicated a great lack of consistency by the overall mean correlations of .41, .62, .64 and .85.

Nursing administrations, using patient assessment tools as a means of providing adequate nursing staff to units must necessarily have information about patient care needs that is consistently obtained from each unit so that it is comparable throughout the institution. Otherwise, a nursing administrator's ability to establish realistic staffing priorities will be less than desirable.
CATEGORIZATION TOOLS

Figure 1—Inter-ward and intra-ward comparison of overall rater reliabilities.

- - - - Mean intra-ward reliability
- - - - Mean inter-ward reliability
- - - - Mean average reliability
CHAPTER IV
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Nurses have determined staffing patterns in the hospital environment for years by patient census at its historical peak. Some nursing administrators question this as being the most appropriate way to manage the workload and have suggested identifying patient's needs for care and placing the patients into categories based on those needs. Once the category or classification has been determined, the assigned category for each patient is converted to pre-determined hours of care. This system then is dependent upon the reliability of nurses using the tools for categorization and the validity of the tools themselves.

This study was designed to look at the inter-rater reliability of staff nurses using various types of categorization methods. The tools in the study ranged from a subjective, intuitive nursing assessment, to more structured format with suggested criteria (Modified Georgette), to a checklist format, (Pardee Checklist), to a point system multiple dimension tool with specific criteria, Poland Point System. These methods were used by six registered nurse raters in a general hospital assigned to two medical divisions. Each nurse rated a randomly designated patient using all four of the methods. Pearson Product-moment correlations were calculated to determine the inter-rater reliability.

Of the four tools, the intuitive, subjective method of identifying patient needs was the least reliable with correlations of .18 to .85 and an overall mean correlation of .57. The reliability coefficients for the Modified Georgette ranged from .00 to 1.00 with an overall mean correlation of .73 and the Pardee Checklist had correlations between .90 and 1.00 with an overall mean correlation of .70. The Poland Point System, which is the most structured tool and has the greatest specificity of patient needs, was used most reliably as indicated by correlations of .49 to .99 and an overall mean correlation of .83.
A possible source of error that may have been introduced into this study is the effect of repeated use of the various tools when assessing needs of patients. The overall weighted mean increased from .57 to .73, to .70, to .83 on tools one, two, three, and four respectively.

In every case the four tools were administered in a one, two, three, four order subsequently more data was available for assessment by the time a rater reached tool four. Repeated use of these tools also increased the individual raters consistency with other nurses. Those nurses with the most numbers of observations had the highest reliability coefficients. A recommendation is made that the study be repeated using one tool per patient by raters to establish whether the observed increase in reliability is due to internal features of the tool or the learned behavior of the raters.

Experience seemed to be another factor that influenced the rater's ability to use the various tools. For example, nurses with less than one year of experience were much more consistent on the more structured tool, the Poland Point System. Raters with experience of three to twenty-five years and raters with less than one year of experience were not consistent with each other. Further study on the role of experience in assessing patient needs is indicated and deemed necessary to fully understand these findings.

Nurses who assumed administrative roles in addition to their clinical assignment were not as consistent in their use of the various tools as were those nurses with primary clinical assignments. Further study is suggested on the time involved and influence of administrative tasks on clinical responsibilities.

Raters intra-divisionally had high reliability coefficients. This study suggested high agreement between raters on a single division or unit but that low consistency existed between raters inter-divisionally. This suggests the influence of leadership and peer group suggestion on assessment of patient needs. The scores do not suggest whether the influence was positive, only that raters on a unit tended to assess a patient in a similar way.
With the limitations of sample size and the lack of consistency among raters on the various tools, it is recommended that further investigation be made on the reliability of these tools. Once that has been completed subsequent validity studies need to be made to assure that nurses are measuring what needs to be measured. To quote Sanford (1965, p. 131):

Once we succeed in making a reliable observation or in setting down a reliable test score, we still face the question of what it means, if anything.
REFERENCES


Wolfe, H. and Young J. P. Staffing the nursing unit: controlled variable staffing. *Nursing research* 1965, 14(3) (a).

Wolfe, H. and Young, J. P. Staffing the nursing unit: the multiple assignment technique. *Nursing research* 1965, 14(4) (b).

Young, John P. Information nexus guides decision system. *Modern hospital* 1966, 106(2) 101-105.
## APPENDIX A

### LATTER DAY SAINTS HOSPITAL

**GUIDELINES FOR USE OF PATIENT CATEGORIZATION**

<table>
<thead>
<tr>
<th>AREA OF CARE</th>
<th>CATEGORY I</th>
<th>CATEGORY II</th>
<th>CATEGORY III</th>
<th>CATEGORY IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Activities of Daily Living</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Eating</td>
<td>1. Feeds self or needs little help.</td>
<td>1. Needs some help in preparing food for eating. May need encouragement.</td>
<td>1. Cannot feed self but is able to chew and swallow all right.</td>
<td>1. Cannot feed self at all and may have difficulty chewing and swallowing food.</td>
</tr>
<tr>
<td>B. Grooming</td>
<td></td>
<td>1. Needs some help with bathing, oral hygiene, hair combing, etc.</td>
<td>1. Unable to do much for self.</td>
<td>1. Completely dependent.</td>
</tr>
<tr>
<td>C. Excretion</td>
<td>1. Up and to BR alone or almost alone.</td>
<td>1. Needs some help in getting up to BR or using urinal.</td>
<td>1. In bed and needs bedpan or urinal to be placed and removed after use. May be able to partially turn or lift self.</td>
<td></td>
</tr>
<tr>
<td>II. General Health</td>
<td>1. Good--in for a diagnostic procedure or a simple treatment or surgery procedure (biopsy, D &amp; C simple and minor.)</td>
<td>1. Mild symptoms, more than one mild illness, mild debility, mild emotional reaction, mild incontinence (not more than once/shift)</td>
<td>1. Acute symptoms, severe emotional reaction to illness or surgery, more than one acute med/surg. problem, severe or frequent incontinence.</td>
<td>1. Critically ill, may have a very severe emotional reaction.</td>
</tr>
</tbody>
</table>
III. Treatments

1. Simple—supervised ambulation, pedangle, simple dressing, test procedure preparation not requiring medication, reinforcement of surgical dressing, X-pad, vital signs once/shift.

IV. Medications

1. Simple, routine, not needing pre or post evaluation, P.R.N. medications no more than once/shift.

V. Teaching and Emotional Support

1. Routine follow-up teaching, patients with no unusual or adverse emotional reactions.

1. Diabetic, cardiac, hypohyptensive, diuretic, anticoagulant medications, P.R.N. medications more than once/shift, medication needing pre or post evaluation.

1. Initial teaching of care of ostomies, new diabetics, tubes that will be in place for periods of time, etc. Patients with conditions that require a major change in eating, living, or excretory practices. Patients with mild adverse reactions to the illness—depressions, overly demanding, etc.

1. Unusual amount of Category II medications, control of refractory diabetics (need to be monitored more than every 4 hours).

1. More intensive Category II items, teaching of apprehensive or mildly resistive patients, care of moderately upset or apprehensive patients, confused or disoriented patients.

1. More intensive Category III medications, I.V.'s with frequent, close observation and regulation.

1. Teaching of resistive patients, care and support of patients with severe emotional reactions.

1. Any Category I treatment more than once/shift. Foley catheter care, I & O, bladder irrigations, sitzbaths, compresses, test procedures requiring medication or follow-ups, simple (nonmedicated) I.V.'s, Clinitest-Acetest, simple enema for evacuation, vital signs every 4 hours.

1. Any treatment more than twice/shift, medicated i.v.'s, complicated dressings, sterile procedure, care of tracheotomy, Harris flush, suctioning, tube feeding, vital signs more than every 4 hours.

1. Any elaborate or delicate procedure, one requiring 2 nurses, vital signs, more often than every 2 hours.
<table>
<thead>
<tr>
<th>AREA OF CARE</th>
<th>CATEGORY I</th>
<th>CATEGORY II</th>
<th>CATEGORY III</th>
<th>CATEGORY IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>II. Diagnostic Evaluation</td>
<td>1. Preparation for simple diagnostic tests &amp; procedures (i.e., phisohex bath, routine enema, G.B. pills)</td>
<td>1. Preparing for multiple tests or procedures, or gathering multiple specimens. (i.e., douche &amp; enema, time scrub for skin prep., enemas till clear with no problems.)</td>
<td>1. Assistance of one nurse with diagnostic procedure requiring 15-30 minutes.</td>
<td>1. Diagnostic procedure requiring nurse assistance of more than 30 minutes.</td>
</tr>
</tbody>
</table>
APPENDIX B
LATTER-DAY SAINTS HOSPITAL
GENERAL GUIDELINES FOR USE IN PATIENT CLASSIFICATION

CLASSIFICATION I
A patient requiring minimal nursing care whose condition is characterized by:
1. Self-sufficient in activities of daily living.
2. Few diagnostic tests.
3. Simple, uncomplicated treatments.
4. Few medications.
5. Acceptable behavior patterns
6. Requirements for simple orientation and teaching to meet patient’s needs.

CLASSIFICATION II
A patient requiring a moderate amount of nursing care whose condition is
characterized by:
1. Need for assistance in activities of daily living.
2. Preparation for multiple tests or procedures or gathering of multiple specimens.
3. Periodic treatment and/or observation.
4. Periodic administration of medications requiring evaluation.
5. Occasional deviations from acceptable behavior patterns.
6. Requirements for more detailed teaching.

CLASSIFICATION III
A patient requiring a considerable amount of nursing care whose condition is
characterized by:
1. Almost complete or total care required as to activities of daily living.
2. Frequent, time consuming diagnostic tests and procedures.
3. Frequent treatments and/or observation.
4. Numerous medications.
5. Significant deviation from acceptable behavior patterns.
6. Requiring specific teaching.

CLASSIFICATION IV
A patient requiring complete nursing care whose condition is characterized by:
1. Total dependency on the nurse for activities of daily living.
2. Excessively time consuming diagnostic tests and procedures.
3. Comprehensive treatments and/or close observation.
5. Severe deviation from acceptable behavior requiring intensive emotional support.
### PARDEE PATIENT CLASSIFICATION CHECKLIST

**DIET**
- Regular
- Special
- N.P.O.

**ACTIVITY**
- Ambulate
- B.R.P.
- Commode
- Positioning/Turn or/ROM

**VITAL SIGNS**
- q4H - qid
- q6H

**I.P.P.B.**
- q2H - q4H

**URINE**
- Spec. Brav./Frac./Vol.

**MONITOR**

**DRESSINGS, DRAINAGE**

**SUCTION**
- N.G./N.T.

**ISOLATION**
- Dressing

**INCONTINENCE**

**DAY OF ADMISSION**

**ISOLATION**

**DAY OF SURGERY/OR MAJ. MED. EXAM.**
### APPENDIX C

**POLAND POINT SYSTEM**

**ELEMENTS OF PHYSICAL CARE AND THEIR LEVELS OF INTENSITY**

**AS DENOTED BY POINT ASSESSMENT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Criterion</th>
<th>Points Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diet</strong></td>
<td>Feeds self without supervision, or parent feeds patient.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feeds self with supervision by staff.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feeds self but needs constant presence of staff, or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>gastrostomy feeding q4h.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Total feeding by personnel, instructing the parent, continuous I.V., or</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>blood transfusion.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tube feedings more frequently than q4h.</td>
<td>12</td>
</tr>
<tr>
<td><strong>Toileting—Output</strong></td>
<td>Toilets without supervision.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Toilets with supervision, specimen to be collected, or uses bedpan.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Up to toilet with stand-by supervision, or output</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>measurement every hour, or daily colostomy irrigation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incontinent, average output.</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Incontinent with diarrhea, or immediate postoperative colostomy or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>urethrostomy, or drainage with frequent dressing change.</td>
<td>12</td>
</tr>
<tr>
<td><strong>Vital Signs and</strong></td>
<td>Routine—daily temperature, pulse, and respiration.</td>
<td>1</td>
</tr>
<tr>
<td>Measures</td>
<td>Vital signs q4h, or night observation q1h.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Vital signs monitored plus hypothermia, or vital signs q2h.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Vital signs and observation every hour, or vital signs monitored plus</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>hypothermia and neurologic evaluation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BP, pulse, respirations, and neurologic evaluation q3/4h.</td>
<td>12</td>
</tr>
<tr>
<td><strong>Respiratory Aids</strong></td>
<td>Beside humidifier, or “blow bottle,”</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mist or Croupette when sleeping, or cough and deep breathe q2h, or IPPB</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>without supervision q4h.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Continuous oxygen, or cough and deep breathe q1h, or continuous assisted</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>ventilation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mechanical respiratory aid, or IPPB with supervision q4h.</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>PPB continuously with intermittent Ambu “bagging.”</td>
<td>12</td>
</tr>
<tr>
<td><strong>Suction</strong></td>
<td>Routine postoperative standby.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Nasopharyngeal or oral suction prn.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Tracheostomy suction every hour, or nasogastric tube irrigation q2h.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Tracheostomy suction q3/4h, patient responsive</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Tracheostomy suction q3/4h, patient not responsive</td>
<td>12</td>
</tr>
</tbody>
</table>
Cleanliness
- Bathes self, bed straightened.  
- Bathes self with help or supervision, daily change of bed.  
- Bathed and dressed by personnel, or partial bath given, daily change of linen.  
- Bathed and dressed by personnel, special skin care, occupied bed.

Turning and/or Assisted Activity
- Up in chair with assistance once in 8 hours.  
- Up in chair with assistance twice in 8 hours, or walking with assistance.  
- Bedfast with assistance in turning q2h, or up walking with assistance of two people twice in 8 hours.  
- Bedfast with assistance in turning q1h.  
- Turning on Foster frame or CircOlectric bed q1h.

<table>
<thead>
<tr>
<th>Points</th>
<th>PCU's</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>4–11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12–19</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>20–27</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>28–35</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>36–43</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>44–51</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>52–59</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>60–67</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>68–75</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>76–80</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

TO ARRIVE AT an intensity of care for a patient, add the appropriate points from each category and use this total to find the PCU's (patient care units in hours) from the conversion table. Point assessment was worked out from a time study, and flexibility was built in by rounding out minutes to the hour to allow for intangibles and unplanned incidents.
VITA

Name: Myrna Loy Williams Warnick
Birthplace: Emery, Utah
Birthdate: June 22, 1936
High School: Carbon High School, Price, Utah
College: Carbon Junior College, Price, Utah
University: Brigham Young University, Provo, Utah
University: University of Utah, Salt Lake City, Utah
Degree: B. S. University of Utah, Salt Lake City, Utah
Professional Organizations: Utah Nurses Association, Inservice Conference Group, American Nurses Association
Professional Positions:
Assistant Director of Nursing
Latter-day Saints Hospital
Salt Lake City, Utah
1967-1969

Director Inservice Education
Latter-day Saints Hospital
Salt Lake City, Utah
1971-1972

Instructor
Department of Community and Family Medicine
University of Utah
Salt Lake City, Utah
1972-