The job performance of 153 graduates of an associate degree nursing program was evaluated by doctors and nurses working in close supervision of the graduates. The evaluation instrument was constructed to represent both the curriculum objectives of the training institution as well as performance criteria of supervisors. Graduates also rated their preparation on a similar instrument. A number of curriculum implications of the results are discussed. Results also indicate no significant relationship between measures of scholastic success (GPA and State Board scores) and rated job performance. The study provides an example of successful implementation of a basic evaluation model. (Author)
Evaluation and Prediction of Rated Job Performance of Nursing Graduates

American Educational Research Association
1975 Annual Meeting

Session #12.01

Presented by
Dr. Barry J. Wilson, University of Northern Iowa
Mr. Gene Packwood, Delta College

This study was prepared under contract with Delta College, University Center, Michigan, with funds awarded through CAPITATION Grant Number (E04.N000601-01), by the Division of Nursing, Bureau of Health Professions Development, Health Resources Administration, Public Health Service, DHHS.
The Study was designed to assess potential strengths and weaknesses of nursing preparation and training as reflected in job performance of nursing school graduates and also to investigate the predictive relationship of measures of scholastic success such as GPA and Nursing Board Scores with graduate job performance. A 62 item rating scale was constructed and designed to tap the following dimensions of nursing performance: planning for nursing care, implementing nursing care, interpersonal relationships and communication, leadership and group procedures, evaluating and reporting nursing care, and professional involvement. Input for the rating scale reflected curriculum objectives as well as a field survey of performance criteria. Graduates were rated by a nurse and a doctor who worked in closest supervision of their job. Graduates completed a similar rating scale in which they were asked to rate the adequacy of their preparation for the various performance dimensions. Ratings were obtained from a sample of 153 graduates of the associate degree nursing program at Delta College, University Center, Michigan. Results indicated a need for additional clinical experience requiring total involvement of nursing students, advanced courses in pharmacology, anatomy, physiology, and nutrition, and leadership experience. Results also suggested no relationship between various indices of GPA and Board Scores with rated performance. It is probable that rated performance is influenced by a number of personality variables. Also
doctors appear to perceive the performance of nurses from different perspectives than do supervising nurses.
INTRODUCTION

The current study was undertaken to investigate the post-graduate job status and performance of associate degree nursing graduates at Delta College, University Center, Michigan. The study was designed to provide data relevant to the following questions: (1) What are some possible strengths and weaknesses of the Delta Nursing Program as reflected in the job performance of nursing graduates? (2) What are the predictive relationships among measures of scholastic achievement such as grade point average (GPA) and State Board of Nursing Examinations and the job performance of nursing graduates? Some assumptions implicit in these questions include: (a) That valid and reliable indices of job performance can be secured; (b) That rated strengths and weaknesses of job performance may reflect strengths and weaknesses of the nursing curriculum; (c) That grades and Board scores are sufficiently reliable and valid to allow prediction to job performance; and (d) That factors contributing to good grades and high Board scores are also significant factors in later performance in the nursing profession.
THE PROBLEM

The task of gathering valid information regarding nursing performance is not as easy as Magnusson (1967) notes. Of central importance is the problem of constructing a concise definition of nursing success. It is an evident fact that the nursing profession is not a single occupation but a whole range of occupations and different areas of nursing place different demands upon individual capabilities. What constitutes a good nurse in one field does not necessarily hold in another. The situation is further complicated in that evaluations of job performance depend upon the observation of more than one judge. Judges who rate the same individuals can have seen these individuals in different situations and different individuals can have been rated by different judges. These considerations indicate the problem of establishing an absolute criterion of nursing success.

Given the task of assessing the job performance of nurses, the author conducted a search for a standardized instrument to evaluate nursing performance. Inspection of the standard reference by Buros (1971) revealed no such instrument in the area of nursing evaluation which was designed for on-the-job performance. Construction of such an instrument was therefore indicated.

Job performance may be evaluated in essentially three ways: by means of a rating scale, a checklist, or product evaluation. The latter was eliminated as unfeasible in this case; the second eliminated because it does not lend itself to quantitative analysis. Construction of a rating scale was therefore indicated.
A primary consideration in the construction of the rating scale was content validity. In this case, given the basic objectives of the study, content validity implies that the content or items of the instrument accurately sample both the curricular objectives of the nursing program at Delta and on-the-job performance criteria. Therefore, construction of the rating scale included not only curricular objectives provided by the Division of Nursing at Delta College but also performance criteria collected from the field.

A good rating scale demands qualified observers who perform the rating. In the present study it was decided that one primary source of qualified observers should be the nurse who exercises closest supervision over the graduate and who is presumably in the best position to evaluate all aspects of the job performed by the graduate. This person is designated "nurse supervisor" in this study although this appellation should not be interpreted as the job title of the rater and should not be confused with the individual so designated in a typical hospital setting. The "nurse supervisor" in the present case is that person designated by the Delta graduates as "the nurse in closest supervision of their work." It was decided that an additional rating be obtained from a doctor who worked closely with the graduate. A third rater in the study is the Delta graduate. In this case, however, rather than rating their own performance, graduates were asked to rate how they perceived the adequacy of their preparation at Delta in relation to the performance criteria sampled by the rating scale.

In summary, the basic problems posed by the study were: (1) Design of a rating scale which accurately sampled the domain of "nursing success;" and (2) Soliciting ratings from the best possible sources.
REVIEW OF RELATED LITERATURE

Measures of scholastic success such as GPA and standardized test scores have traditionally played an important role in the selection of individuals for jobs and advanced study. This role has come under attack in recent years. Kirschenbaum, Simon, and Napier (1971) summarize a number of studies investigating grades as predictors of academic and occupational performance. Among these, a study by Prince, Taylor, Richards, and Jacobsen (1969) found no relation of grades in medical school to twenty-four performance characteristics of physicians although there was a slight relationship observed between grades and doctors who contributed to the professional literature. Kappell (1962) reported a slight positive correlation between college grades and final salaries attained by employees of the American Telephone and Telegraph Company but no relationship with other performance criteria. Barr (1961) summarized thirty-three studies in the area of prediction of teacher effectiveness, and found a median correlation of .09 (negligible) between supervisor ratings and college GPA of teachers. On the other hand, there are studies which do report higher relationships between grades and later performance (e.g. Breckenridge, 1932). When grades are used as predictors, data can be obtained as to the accuracy with which they do predict. This is something which must be determined in each setting and for each criterion that is being predicted.

The construction of rating scales has a long history which will not be summarized here. As Thorndike and Hagen (1961) note, there are two basic problems in obtaining sound ratings: (1) The willingness of the rater to rate honestly and conscientiously in accordance with the instructions given to him;
and (2) his ability to rate consistently and correctly even with the best of intentions. The implications of these problems are, first of all, raters must be "sold" on the importance of the ratings or else judgments may be hurried and superficial; secondly, attempts must be made to prevent the rater from being overly influenced by a general bias with respect to the person being rated. Specific attempts to solve these problems were undertaken in the present study. Correspondence with graduates and raters included cover letters from the Division of Nursing as well as the author. These letters fully described the nature of and need for the study and its relevance to nursing education. Raters were cautioned in their instructions regarding rating according to general impressions. All items contained provisions for making "Not Applicable".

THE DESIGN OF THE STUDY

The variables involved in the study included four indices of GPA (Nursing GPA, Science GPA, Overall Delta GPA, and High School GPA), scores from the State Board Examinations in five areas (Medical, Surgical, Obstetrical, Pediatric, Psychiatric), age, experience, seven subtest scores from the job performance rating scales.

The question posed concerning potential strengths and weaknesses of the nursing program in this case involved an essentially descriptive analysis of the job performance ratings. It must be emphasized, however, that due to lack of a control group - i.e. ratings of a group of nurses with similar training and experience from other institutions - interpretation of these results must be cautious and tentative. It is quite possible that some strengths or weaknesses
as rated reflect perceptions of the rater which have little to do with any particular training program, but may be based upon perceptions of certain "types" of nurses, or nurses of a certain age, or with other possibly non-relevant characteristics which the Delta nursing sample were in common with the general population of recent nursing graduates. In the present study, the only control available was utilized - i.e. ratings were obtained from several sources who presumably do not share the same set of biases. Ratings were obtained from the graduate, from a doctor, and from a nurse who exercised some form of supervision over the graduate.

The second question concerning the relationship among indices of GPA and Nursing Board Scores and rated job performance was investigated by a correlational analysis. If moderate correlations of GPA and Board Scores were observed, the results would be analyzed by step-wise regression procedures to determine the best combination of predictors to the criterion variable of job performance.

**SELECTION OF SAMPLE**

There were two considerations in the selection of the nursing graduate sample. First of all, the graduate should have been on the job sufficiently long to allow valid assessment of performance, and yet also be a recent graduate in order to avoid contamination of ratings by lengthy experience and post-graduate training. Second, the size of the sample should be large enough to allow valid inferences. The decision was made to include all graduates of 1970, 1971, and 1972. This yielded a population of 364 graduates. Initial letters to 75 graduates were returned as unforwardable, address unknown.
Initial contact was made with approximately 180 graduates. Of these, 111 completed a rating scale while 93 were rated by a nurse in a supervisory capacity and 57 were rated by a doctor. All graduates were included in the final analysis if one of the three ratings was returned. This yielded a total of 153 graduates although one or two of the ratings called for may have been missing.

Graduates who responded were found to work in a variety of nursing settings but primarily in a hospital setting. It should be noted that graduates who responded were found to have slightly higher but statistically significant GPA and Board Scores than non-respondents.

THE RATING SCALES

The 62 item rating scale completed by nurses in a supervisory capacity and doctors were identical. The rating scale completed by graduates was also identical except that the last 4 items and 1 item in the Leadership scale were deleted as not appropriate. Also the graduates were asked to rate the adequacy of their preparation at Delta rather than rating their perceived performance.

Doctors and nurses who rated graduates were nominated by the graduates themselves as the nurse or doctor who worked in closest supervision with them. These nominations were secured by the initial questionnaire sent to the 364 graduates. Those nominated were contacted by mail, informed of the nature of the study, and requested to complete an open-ended questionnaire. This questionnaire asked them to define criteria of nursing performance which were most important as they perceived their situation. These collected criteria together
with a statement of performance objectives provided by the Division of Nursing at Delta, served as the content base for the rating scale items. The initial set of scale items written by the author was revised and edited in accordance with a review by faculty from the Division of Nursing which included a trial administration of all scales by a number of individuals not in the designated graduate population.

ANALYSIS OF THE DATA

The statistical analysis was performed with the assistance of the Statistical Package for the Social Sciences (SPSS), a system of computer programs available at the Computer Center at the University of Northern Iowa. In the present study, a complete set of data for each of the 153 graduates in the sample was not available. In order to make maximum use of available data, a case was deleted from the analysis only when the missing data was required for that specific analysis. In the case of correlations, missing data from either pair caused pair-wise deletion for that specific correlation. Analysis of all rating scales was conducted in a similar fashion. The mean response for each item was computed and represents the weighted sum of responses to the item. The categories were weighted from 1 to 5 with 1 being poor and 5 being excellent. The N/A category as well as items not responded to were in all cases ignored by the analysis. Subtest means were computed by first summing within cases the weighted item responses in the subtest and dividing by the number of items in the subtest. These were summed across cases and divided by the number of cases to get the subtest mean.

12
RESULTS AND DISCUSSION

The reporting and discussion of results has been separated in accordance with the two basic questions posed by the study. Results and discussion of possible strengths and weaknesses are taken up first and followed by results and discussion of the relationship of GPA and Board Scores with rated job performance. A final section discusses the merits of the rating scale with possible suggestions or revision for future use.

POTENTIAL STRENGTHS AND WEAKNESSES

It should be noted again that due to lack of a control group of graduates from other nursing programs, results should in no way be interpreted as placing graduates or the nursing program on a good-bad or strong-weak continuum relative to other graduates or other programs. It is quite possible that an area designated "weak" in the present context could be representative of all recently graduated nurses and that if data collected on a representative sample of all recent graduates might demonstrate that an area designated "weak" in the present study is a "strong" area relative to other institutions. Of course, the opposite could also be true. The fact of the matter is that the present data cannot be used to substantiate any such external comparisons. Assessment of strengths and weaknesses in the present study refers only to relative rated performance of Delta graduates.
RESULTS AND DISCUSSION

Averaging of subtest rank according to the three ratings indicate basic agreement among graduates, nurse supervisors, and doctors that areas of strength include those of "implementing nursing care" and "interpersonal relations" while areas of weakness include "leadership and unit procedures" and "professional involvement." While the category of professional involvement was rated relatively high by doctors, fifty percent of the doctors checked the Not/Applicable category.

Inspection of individual stems in the scale indicate agreement of strength in the following areas: identification of patient needs, providing nursing care until medical orders are obtained, preparing patients for diagnostic procedures, and kindness and consideration shown toward patients. Agreement of weak areas include use of community resources, application of facts and principles of nutrition, providing for a group of patients, conducting nursing care conferences, assuming leadership and management responsibilities, and assisting in the instruction of auxiliary personnel. It should be noted that with the exception of the item concerning leadership and management these "weak" items were checked as Not/Applicable by a relatively high percentage of raters.

Certain items show substantial disagreement among raters which may reflect differences of professional perceptions. Graduates felt relatively well prepared to develop a written plan of care for the patient, however, their supervisors feel they don't perform that task well. Nurse supervisors were relatively well satisfied with graduates' application of anatomy and physiology, while doctors and graduates were quite dissatisfied. Two other items are of particular interest in that graduates rated their preparation very low in the areas of taking
appropriate actions in emergencies and carrying out commonly occurring techniques for patient care. Quite the contrary say supervising nurses and doctors who gave these items very high ratings. Finally, consultation and cooperation with physicians and others ranked very high with doctors but very low with nurse supervisors. Graduates perceive this as an area needing further development in their preparation.

RESULTS AND DISCUSSION OF OPEN-ENDED QUESTIONS

GRADUATE RESPONSES

Fifty-five per cent of the graduates mentioned the Auto-tutorial Laboratory (ATL) as a major strength of the program. Twenty-five per cent of the graduates felt that the excellence of some of their instructors was a major strength. Other strengths which were mentioned by at least ten per cent of the graduates were: good instruction in principles and theories of nursing, weekly quizzes, the stress on Nursing Care Plans, and the use of area hospitals.

Nearly all graduates indicated that they felt that more clinical experience was needed, especially experience which would allow total involvement. Likewise nearly all graduates felt the need for additional course work, especially in the areas of pharmacology, physiology, anatomy, and nutrition. Opportunity for leadership and instruction in team leading was also a major area of concern.

Some suggestions were also made concerning changes in the learning environment. These included better selection of instructors, smaller student-instructor ratio, system of grading, and method of resolving student-instructor conflicts. Finally, a number of miscellaneous suggestions were made including approximately ten...
per cent of the graduates who felt that the open-door admissions policy should be ended and the course of studies extended to a three-year program.

NURSE SUPERVISOR AND DOCTOR RESPONSES

Interpersonal relations were seen by nurse supervisors equally as a strength or weakness of the graduate. This information at first glance would not appear to be particularly informative. Close inspection of the comments, however, suggests that interpersonal relations designated as a strength referred to relations with patients, families, or hospital personnel at the occupational level of the graduate or below. Interpersonal relations mentioned as a weakness typically referred to the relationship between the graduates and their superiors. On the other hand, doctors see the relationships of graduates with patients and subordinates as a major strength with no mention of weakness in dealing with superiors. These comments substantiate the contrasting viewpoints of nurse supervisors and doctors in one of the items in the rating scale.

Nurse supervisors frequently mentioned continued efforts at improvement of self and nursing knowledge as a strength but both nurse supervisors and doctors perceive a need for continued education to improve technical skills and keep knowledge current.

Leadership was frequently mentioned by nurse supervisors as a strength for some graduates but a weakness of others. Approximately fifty per cent of the nurse supervisors mentioned the leadership area as a strength or weakness which seems to indicate leadership as an important performance variable for many of the graduates occupational categories.
Self confidence, organizational efficiency, decision making ability, and ability to set priorities were seen as weak areas by nurse supervisors and doctors. It would seem that these performance variables relate most directly to clinical experience.

SUMMARY

Taken as a composite, responses to the rating scale and to questions concerning major strengths and weaknesses may indicate the following generalizations concerning the present sample of Delta nursing graduates of 1970, 1971, and 1972:

1. The ATL lab was perceived by graduates to be an effective instructional aid.

2. Graduates perceived themselves as providing compassionate and understanding care to patients and able to relate well with patients and able to relate well with patients and subordinates. This perception was generally substantiated by their raters.

3. Graduates may have difficulty in relating well with superiors, especially nurse superiors although information from doctors does not substantiate this generalization. (Given the fact that doctor raters were nominated by the graduates and return rate from doctors was low it is possible that the ratings obtained from doctors may not be representative.)
4. Graduates perceived a need for additional clinical experience which allows for the total involvement of the graduate. This lack of clinical experience may be substantiated by ratings from nurse supervisors and doctors who perceived the lack of clinical experience in terms of lack of self-confidence, lack of organization and setting of priorities, and lack of leadership capabilities.

5. Graduates are often placed in positions of leadership. Some graduates were rated well in this area, others poorly. Graduates perceived a need for leadership experience and possibly leadership classes in their nursing preparation.

6. Graduates perceived themselves as lacking some basic technical skills but this did not appear to be entirely substantiated by their supervisors.

7. Graduates perceived the need for additional coursework especially in the areas of pharmacology, physiology, anatomy, and nutrition. Doctors tended to agree with this assessment but it was less evident in the returns from nurse supervisors. It is probable that the particular course perceived as "needed" is one which relates to the particular job situation in which the graduate works. There is agreement, however, that continuing education is needed for graduates.
GPA AND BOARD SCORES AS PREDICTORS OF PERFORMANCE

RESULTS

Tables of correlations between GPA and age with the three ratings are presented in Table 1. Graduate ratings of the program show low positive correlations with nursing GPA and age. A similar relationship between age and nursing GPA was noted in an earlier study by the Office of Research and Development at Delta College. The correlation of .24 found between nursing GPA and graduate ratings suggests a slight tendency for graduates with high GPA to rate the program more favorably.

Table 2 contains correlations between the three ratings, experience and the five areas of State Nursing Board scores. Experience was operationally defined as months of work experience as a nurse. Experience as a licensed practical nurse was included. Negligible correlations among these variables ranged from -.15 to +.14.

Correlations among the three ratings were quite low and not statistically significant. A negative correlation between doctor's rating and experience is worthy of note (r = -.28) and is discussed below.
DISCUSSION

The data suggest no significant linear relationships between measures of scholastic success and the job performance of nursing graduates as rated by nurses and doctors who work with the graduates. As noted in the discussion of strengths and weaknesses, there are indications that doctors who work with graduates perceive job performance from a different perspective than do the nurses who rated the graduates. This may account for some portion of the low correlation between doctor and nurse ratings ($r = .14, N = 31$). If the predictor variables of GPA and Board scores had demonstrated a positive correlation with doctor ratings as well as nurse supervisor ratings, some significant relationship might have been produced by pooling the ratings of doctors and nurse supervisors into a composite rating. Such a procedure has been shown to enhance the reliability of ratings (Thorndike and Hagen, 1961), but does not appear to be warranted in the present case.

The significant negative correlation (-.28) found between doctor's ratings and length of experiences is somewhat surprising. The factors which contribute to this negative relationship are a matter of conjecture but may reemphasize the point that personal attributes and personality dimensions influence perceptions of job performance.

Given the fact that graduates work in many different settings with different demands placed upon their capabilities, it is possible that scholastic success may show a relationship with some types of nursing occupations, but not others. As Goldman (1961) observes, not all members of a particular group are equally predictable and by sub-dividing a group in certain ways we can sometimes raise the efficiency of prediction for at least some of its members.
### TABLE 1

**INTER-CORRELATIONS OF AGE AND GPA WITH RATINGS**

<table>
<thead>
<tr>
<th></th>
<th>Nursing GPA</th>
<th>Science GPA</th>
<th>Overall GPA</th>
<th>Age</th>
<th>High School GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doctors’ Ratings</td>
<td>.07</td>
<td>.14</td>
<td>-.03</td>
<td>-.04</td>
<td>.02</td>
</tr>
<tr>
<td>Nurse Supervisor</td>
<td>.04</td>
<td>.01</td>
<td>.02</td>
<td>.06</td>
<td>-.02</td>
</tr>
<tr>
<td>Graduate Ratings</td>
<td>.24*</td>
<td>-.12</td>
<td>.10</td>
<td>.20**</td>
<td>-.02</td>
</tr>
</tbody>
</table>

**sig .01**

### TABLE 2

**INTER-CORRELATIONS OF STATE BOARD SCORES WITH EXPERIENCE AND RATINGS**

<table>
<thead>
<tr>
<th></th>
<th>Medical</th>
<th>Surgical</th>
<th>Obstetrical</th>
<th>Pediatric</th>
<th>Psychiatric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse Supervisor</td>
<td>-.10</td>
<td>.01</td>
<td>-.07</td>
<td>-.15</td>
<td>-.01</td>
</tr>
<tr>
<td>Rating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate Rating</td>
<td>-.09</td>
<td>-.08</td>
<td>-.01</td>
<td>.00</td>
<td>-.15</td>
</tr>
<tr>
<td>Experience</td>
<td>.06</td>
<td>.00</td>
<td>.04</td>
<td>.09</td>
<td>-.14*</td>
</tr>
</tbody>
</table>

*sig .05
Inspection of rating scales as well as subtest intercorrelations suggest that a halo effect governed some of the ratings especially those completed by the doctors. The halo effect is the tendency to rate in terms of over-all general impression without differentiating specific aspects. In the present case, this was suggested when some ratings were returned with virtually all items checked in the same category.

**SUMMARY**

Results of the present study suggest little or no relationship between GPA, Board scores, and rated job performance. These results are not inconsistent with a number of studies cited above concerning grades and occupational performance. There are a number of possible explanations for such results. For instance, one might question the reliability of not only the rating scale, but also the indices of GPA and the Board scores as well. An important consideration in the evaluation of the rating scale is that of inter-rater reliability. An unbiased estimate of inter-rater reliability requires two equally qualified raters who bear essentially the same relationship to the ratee. In the present case, this was not possible and no unbiased estimate can be made of the inter-rater reliability of the scale.

Inspection of the responses to the open-ended questions regarding strengths and weaknesses suggests that a number of factors weigh heavily in assessment of job performance which probably have little to do with ability to get grades or perform well on Board scores. These include variables such as interpersonal relations, dependability, enthusiasm, reactions under stress, professional appearance, punctuality, adaptability, and decisiveness.
The following generalizations appear to be warranted by the results:

1. There was no demonstrated relationship between measures of scholastic success such as GPA and State Board scores with rated job performance of Delta graduates.

2. The inter-rater reliability of the scale may need improvement for future use although an unbiased estimate of inter-rater reliability was not obtained in the study.

3. Rated job performance appears to be greatly influenced by a number of personality traits. Prediction of job performance might be enhanced by attempting to estimate these attributes during training.

4. Ratings, especially those ratings completed by doctors, appear to have been influenced by a halo effect.
RECOMMENDATIONS

The following recommendations were provided to the Nursing division at Delta College with acknowledgment that (1) appropriate curricular changes may already have been undertaken (2) suggestions for implementation are not (and probably can not) be specific.

1.) It was recommended that the Division of Nursing investigate the feasibility of incorporating additional clinical experience into the nursing program.

2.) That the Division of Nursing investigate the feasibility of incorporating leadership training and/or experience into the nursing program.

3.) That advanced courses in the area of pharmacology, physiology, anatomy, and nutrition may be needed as optional courses or post-graduate courses.

4.) That the present data to be used as a baseline for evaluation of the effectiveness of curricular changes.

5.) That the rating scale be revised for future use.

6.) That further efforts to predict job performance include personality variables in the prediction.
7.) That recommendations of graduates for jobs and post-graduate study be based on information in addition to GPA and Nursing Board Scores.

8.) That attempts be made to make the scale available to other nursing institutions that comparative data may be gathered.
SELECTED BIBLIOGRAPHY


