The Admissions Process of a Bachelor of Science in Nursing Program: Initial Reliability and Validity of the Personal Interview

Barbara Carpio and Barbara Brown


Des équipes de trois membres évaluent l’éligibilité du candidat au programme en se basant sur six facteurs: la motivation, la connaissance du programme, la capacité à résoudre les difficultés, la qualité de la relation avec les autres, la capacité d’auto-évaluation et les objectifs de carrière. Chaque intervieweur accorde au candidat un pointage global sur une échelle de sept paliers. Aux fins de la présente étude, quatre équipes d’interviews ont été choisies au hasard parmi 31 équipes pour faire passer des entrevues à quatre candidats fictifs (préprogrammés). À l’aide d’une analyse des variances dans une formule de mesures repétées à deux facteurs destinée à analyser les pointages résultant des entrevues, on calcule les coefficients d’objectivité et de corrélation interne entre les équipes. La fiabilité parmi les équipes s’étendit de 0,64 à 0,97 pour les facteurs individuels, et de 0,66 à 0,89 pour les pointages d’ensemble. Pour les six facteurs, le coefficient d’objectivité et de corrélation interne s’étendit de 0,81 à 0,99 et de 0,96 à 0,99 pour les pointages d’ensemble. Les coefficients de corrélation entre les facteurs individuels et les pointages d’ensemble varient de 0,875 à 0,995. Les coefficients de corrélation de Pearson entre les postes allaient de 0,774 à 0,999. On calcule ensuite les coefficients de corrélation interne des entrevues individuelles de 108 vrais candidats au programme. Le coefficient de fiabilité, basé sur les pointages d’ensemble, était de 0,79 pour l’observation unique (à 1 évaluateur), et de 0,91 pour les observations multiples (à 3 évaluateurs). Ces conclusions appuient l’idée de poursuivre l’utilisation des entrevues, qui se révèlent un outil fiable et d’une apparente validité. Des études de validité prédictive seront également entreprises.

The undergraduate nursing degree program (B.Sc.N.) at McMaster University School of Nursing uses small groups, and is learner-centered and problem-based. A study was conducted during the 1991 admissions cycle to determine the initial reliability and validity of the semi-structured personal interview which constitutes the final component of candidate selection for this program. During the interview, three-member teams assess applicant suitability to the program based on six dimensions: applicant motivation, awareness of the program, problem-solving abilities, ability to relate to others, self-appraisal skills, and career goals. Each interviewer assigns the applicant a global rating using a seven-point scale. For the purposes of this study four interviewer teams were randomly selected from the pool of 31 teams to interview four simulated (preprogrammed) applicants. Using two-factor repeated-measures ANOVA to analyze interview ratings, inter-rater and inter-team intraclass correlation coefficients (ICC) were calculated. Inter-

Barbara Carpio, R.N., M.Sc.N. is Assistant Professor and Barbara Brown, R.N., M.Sc.N. is Associate Professor at McMaster University School of Nursing in Hamilton, Ontario
team reliability ranged from .64 to .97 for the individual dimensions, and .66 to .89 on global ratings. Inter-rater ICC for the six dimensions ranged from .81 to .99, and .96 to .99 for the global ratings. The item-to-total correlation coefficients between individual dimensions and global ratings ranged from .8 to 1.0. Pearson correlations between items ranged from .77 to 1.0. The ICC were then calculated for the interview scores of 108 actual applicants to the program. Inter-rater reliability based on global ratings was .79 for the single (1 rater) observation, and .91 for the multiple (3 rater) observation. These findings support the continued use of the interview as a reliable instrument with face validity. Studies of predictive validity will be undertaken.

The purpose of the nursing admissions process is to identify, from a pool of well qualified applicants, those who will be most successful academically and professionally. A selective admissions process is needed when the number of (eligible) applicants to a program exceeds the employment opportunities or available educational resources, or includes those who are poorly suited to the program (Haglund, 1978). Canadian nursing programs have been urged to develop selection criteria and processes that are in keeping with their educational beliefs and curricula (Registered Nurses’ Association of Ontario, 1981), and to periodically re-evaluate their selection processes and criteria.

The McMaster University School of Nursing offers a B.Sc.N. program that is “...learner-centered, focused on solving clinical problems through the use of inductive and deductive reasoning and requires the acquisition of appropriate knowledge, skill and personal qualities” (McMaster University School of Nursing, 1986). There are two routes to obtaining the degree: a four-year program of study (basic stream) for applicants who apply directly from secondary school or have other qualifications (referred to as mature students in some programs), and a two-year stream (post diploma) for graduates of diploma nursing programs (community college or hospital school nursing programs). Secondary school applicants are selected on the basis of academic performance alone, while those with a diploma or other qualifications undergo a selection process which assesses personal qualities in addition to academic ability. A semi-structured team interview is one component of the latter type of selection process. Since the reliability and validity of the autobiographical letter used in the admissions process was previously examined (Brown, Carpio & Roberts, 1991), the purpose of the present study, conducted during the 1991 admission cycle, was to determine the initial reliability and validity of the admission interview.

Literature Review

Admission criteria

As nursing education has evolved, academic requirements for admission have become progressively more stringent. There has been considerable research to assess the academic predictors of success in nursing (Alichnie & Bellucci, 1981; Clemence & Brink, 1978; Felts, 1986; Higgs, 1984; Jacono, Keehn &
Corrigan, 1987; Oliver, 1985; Sharp, 1984; Weinstein, Brown & Wahlstrom, 1982) and other health sciences (Balogun, Karacolof & Farina, 1986; Posthuma & Noh, 1990). It has been conjectured that personality attributes could also serve as predictors of program success, "...particularly as the [pool of] applicants becomes more homogeneous in terms of academic potential" (Beale & McCutcheon, 1980, p. 31). However, cognitive variables continue to be the most widely used selection criteria. Those that have been studied include reading ability and comprehension (Backman & Steindler, 1971; Outtz, 1979; Seither, 1980; Smith & Pervanger, 1974), verbal ability (Backman & Steindler, 1971), critical thinking (Bauwens & Gerhard, 1987; Berger, 1984), personality attributes (George & Owens, 1983), and age (Froman & Owen, 1989). Grant (1986) concluded that reading skill is the one variable that is consistently predictive of achievement in a nursing program as well as state board licensing examinations (NCLEX). The only consistent finding of studies investigating non-cognitive variables is that successful students are more mature, older, and achievement-oriented.

**Admission interview**

No studies on the use of a personal interview for admission were found in the nursing literature, although its use in selection of medical students has been widely reported (Burgess, Calkins & Richards, 1974; Calkins, Richards, McCanse, Burgess & Willoughby, 1974; McManus & Richards, 1984; Powis, Neame, Bristow & Murphy, 1988). Edwards, Johnson and Molidar (1990) identified four objectives of the personal interview in the selection process: to gather information for the applicant's profile, to assist the selection committee in making definitive decisions, to provide an opportunity to verify applicant data, and as a recruitment activity. The reported inter-rater reliability for interviews varies from a low of \( r = .32 \) (Vojir & Bronstein, 1983) to a high of \( r = .90 \) (Richards, McManus & Maitlis, 1988). Overall, the literature favors a panel or team interview to reduce the effects of individual bias. Evidence indicates that the reliability of the interview improves when a structured format is used (Edwards, Johnson & Molidar, 1990). Tarico, Smith, Altmair, Franken, and Van Velzen (1984) suggest that a formal job analysis will identify the critical factors to be included in the interview, thus improving the validity. Consistency improves with standardization of questions; reliability is increased by using sample questions and behavioral anchors to describe levels of performance such as "good," "satisfactory" or "outstanding" (Edwards, Johnson & Molidar, 1990). Moreover, because personal interviews are labor intensive, their use needs to be considered carefully (Cascio & Ramos, 1986). Overall, studies have shown that the personal interview is useful as a selection criterion to address non-academic qualities of applicants to health sciences educational programs, and the reliability and validity are improved through the use of a structured format and multiple interviewers.
Method

Applicants to the basic four-year program who are not admitted directly from high school and registered nurses applying to the degree completion stream have to meet a minimum academic standard, submit an autobiographical questionnaire and provide personal references in support of their application. Applicants who achieve the highest ratings are invited to a personal interview. After the interviews, applicant files are collated and offers of admission made.

The interview format used at McMaster University is a 30-minute semi-structured, panel interview. Interview teams consist of a faculty member, a student or alumnus of the program, and a community representative. The community representative may be “nursing service personnel, other faculty and consumers of health care [who have] an awareness of those personal characteristics perceived as necessary in health care workers” (McMaster University School of Nursing, 1983). The only information interviewers have about the applicant is a one-page autobiographical sketch provided by the applicant at the time of the interview.

The measures taken to establish face validity of the interview were similar to those reported in the study of the autobiographical letter (Brown, Carpio & Roberts, 1991). A formal job analysis to identify attributes of the successful nursing student was not conducted at the time that the interview guidelines were developed in 1981. However, six themes or dimensions were identified by faculty as being essential prerequisites to success in this small-group, student-centered, problem-based nursing program: motivation, awareness of the McMaster University nursing program, problem-solving abilities, ability to relate to others, self-appraisal skills and career goals. Interviewers independently assess the applicant in each of the six dimensions using a five-point Likert scale. Since interviewers are not provided with formal behavioral descriptions of each point on the scale, they are asked to provide written descriptions of the applicant’s responses to support their assessment. Interview guidelines and sample questions are provided in the interviewer packages (Figure 1).

Interviewers, particularly those new to the interview process, attend an annual orientation session immediately prior to the interview dates. These briefing sessions provide an opportunity to ask questions, meet the other team members, practice interviewing, and receive instruction packages.

During the interview, each rater independently assesses an applicant and records his/her comments and ratings on the personal interview assessment form. Each interviewer assigns the applicant a 1-5 rating for each of the six
Figure 1

Sample interview guidelines for the dimension “self-appraisal ability”

<table>
<thead>
<tr>
<th>Self-appraisal ability</th>
<th>Suggested Approaches and Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students in the B.Sc.N. program are expected to increasingly assume responsibility for self-evaluation.</td>
<td>What is your greatest asset in working with people?</td>
</tr>
<tr>
<td>Does the candidate demonstrate insight into:</td>
<td>What do you do if peers become angry with you?</td>
</tr>
<tr>
<td>1. strengths/weaknesses in working with people;</td>
<td>What do you do if authority figures tell you that your work performance is falling below what is expected?</td>
</tr>
<tr>
<td>2. strengths/weaknesses in educational pursuits;</td>
<td>How do you handle this yourself?</td>
</tr>
<tr>
<td>3. strengths/weaknesses in being independent;</td>
<td>What is your greatest asset in learning?</td>
</tr>
<tr>
<td>4. feeling self-confident;</td>
<td>Have other people told you this?</td>
</tr>
<tr>
<td>5. seeking out other people to validate self-performance.</td>
<td>Is this what you think?</td>
</tr>
<tr>
<td></td>
<td>How independent do you think you are?</td>
</tr>
<tr>
<td></td>
<td>Give an example of when you exercised independence.</td>
</tr>
<tr>
<td></td>
<td>Have you ever had an experience where you had to ask others to tell you what they think of your work?</td>
</tr>
<tr>
<td></td>
<td>How do you react when the feedback is negative?</td>
</tr>
</tbody>
</table>

dimensions, and a global rating using a 7-point Likert scale categorizing them as “unacceptable” (1-2), “acceptable” (3-5), “good” (6), or “outstanding” (7) for the program. The 7-point scale for overall assessment is distinct from the sum or average of the numerical rating assigned to each dimension.

The three interviewer ratings are summed, providing an interview score out of a maximum possible 21 (three ratings of 7). This number is then combined with the academic rating (out of a maximum possible 6), the personal references (maximum 5), and the autobiographical questionnaire (maximum 21) to calculate the applicant’s overall admission score (out of a maximum possible 53). During the collation process faculty-student teams independently review the complete files for consistency of comments and ratings within and between teams. Although some variation between interviewers in terms of their comments and ratings of an applicant is desired, wide discrepancies warrant careful review of the file by the collation group. Interviewers’ comments are reviewed, and unusually lenient or harsh ratings by a team are thoroughly discussed.

Unsuccessful applicants can discuss the outcome of the selection process and their individual performance with the Chairperson of the Undergraduate
Nursing Admissions Committee. The Faculty of Health Sciences does not store application information; unsuccessful applicants can reapply in subsequent years without being recognized as repeat applicants. Thus, there are no data to indicate whether any of the applicants in the 1991 admission cycle had previously applied unsuccessfully, or sought assistance to prepare themselves for the interview.

Two study designs were used: four pre-programmed simulated applicants were each interviewed by four selected interviewer teams, and 140 actual applicants were interviewed by 31 teams in the regular admission cycle.

**Simulated applicant design**

To assess the reliability and validity of the personal interview, four applicant profiles, representative of “unsatisfactory,” “acceptable,” “good,” and “outstanding” suitability to the program were developed by the research team based on the six dimensions addressed in the interview guide. These profiles were then used as the basis for training four simulated applicants. All four simulations were of applicants to the basic stream, as it would have been very difficult to develop a credible simulation of a registered nurse, and local nurses might have been recognized by the interviewers and/or been future applicants to the program.

The Standardized Patient Program of the Faculty of Health Sciences helped the research team to recruit and train three women and one man. Their own personal experiences were integrated into the simulations. To avoid their being recognized as actors by the interviewers, none of the simulated applicants had acted as simulated patients for the B.Sc.N. program during the previous few years. After the training sessions, each applicant participated in a mock interview with the research team in order to fine-tune the simulation.

Four interviewer teams were randomly selected from the pool of interviewers who had provided written consent to participate in the study. All interviewers were blind to the final selection of study teams. The study teams were identical in composition to the other interview teams, participated in the regular orientation session, and received the same instructions during the actual interview process. All four simulated applicants were interviewed by each of the study interviewer teams, but in different order to control for order effect.

**Actual applicants**

Thirty-one teams interviewed a total of 140 applicants; 86 were applying to the post diploma stream and 54 to the basic stream. Inter-rater and inter-
team reliability of interview scores of all applicants were assessed. Due to
logistic problems, including last-minute cancellations by interviewers, only
108 applicants were interviewed by complete three-member teams and in-
cluded in the study. For the purposes of data analysis, applicants were identi-
fified only by an assigned number, and interviewers by team and category (e.g.
the faculty member of the third team was coded as 3F, the student as 3S, the
community representative as 3C, and so forth).

**Statistical analysis**

The Biomedical Data Package Statistical Software (University of California,
1990) (BMDP) routines were used to analyze the data. A two-factor repeated-
measures ANOVA was used where the explicit factors were team (level=4)
and rater (level=3). The mean square error terms for subjects and the factors
of interest (rater and then team) were used to calculate intra-class correlation
coefficients (ICC) (also known as weighted Kappa or generalizability coeffi-
cients). Generalizability theory on which calculations were based allows iden-
tification and measurement of the sources of variation which contribute error
to an estimate (Streiner & Norman, 1989). The interview ratings were ana-
yzed to measure ICC among interviewers and teams. Reliability was
measured through inter-rater and inter-team correlations of global ratings.

**Results**

**Simulated applicant design**

The ICC of global ratings and of the ratings on the six dimensions were
calculated to determine both inter-rater and inter-team reliability (Table 1).
The ICC ranged from .66 for the single observation inter-team correlation, to
.99 for the multiple observation inter-rater correlation. In order to determine
if the global ratings were a true reflection of the ratings on the six component
dimensions, ICC were calculated. The ICC ranged from .64 on the inter-team
single observation for the dimension of “self-appraisal” to .99 for the inter-
rater multiple observation for the dimension of “motivation.”

To examine the contribution of each of the six dimensions to the overall
rating, item-to-total correlations were calculated for each of the dimensions
as compared with the global rating. The Pearson correlation coefficients be-
tween any pairing of dimensions, or of individual dimensions with the global
rating were all very high: the correlations for community interviewers ranged
from .88 to 1.00; for faculty raters, from .88 and 1.00; and for student/alum-
nus interviewers, .77 to 1.00. For each of the interviewer categories, it was the
dimension “career goals” that yielded the lowest correlation coefficient when
paired with another dimension.
Table 1

Inter-team and Inter-rater Intraclass Correlation Coefficients (r) of Interviews of Four Simulated Applicants (n=4 applicants x 4 teams x 3 raters/team)

<table>
<thead>
<tr>
<th></th>
<th>Inter-Team</th>
<th></th>
<th>Inter-Rater</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Team</td>
<td>4 Teams</td>
<td>1 Rater</td>
<td>3 Raters</td>
</tr>
<tr>
<td>Motivation</td>
<td>.75</td>
<td>.93</td>
<td>.92</td>
<td>.99</td>
</tr>
<tr>
<td>Awareness</td>
<td>.88</td>
<td>.97</td>
<td>.93</td>
<td>.98</td>
</tr>
<tr>
<td>Problem-solving</td>
<td>.72</td>
<td>.93</td>
<td>.84</td>
<td>.97</td>
</tr>
<tr>
<td>Self-appraisal</td>
<td>.64</td>
<td>.90</td>
<td>.88</td>
<td>.98</td>
</tr>
<tr>
<td>Ability to Relate</td>
<td>.69</td>
<td>.93</td>
<td>.85</td>
<td>.98</td>
</tr>
<tr>
<td>Career Goals</td>
<td>.67</td>
<td>.93</td>
<td>.81</td>
<td>.97</td>
</tr>
<tr>
<td>Global Rating</td>
<td>.66</td>
<td>.89</td>
<td>.96</td>
<td>.99</td>
</tr>
</tbody>
</table>

To ascertain that interviewers could discriminate between applicants highly suited to the program and those less suited, the four randomly selected interviewer teams (each consisting of three members) interviewed all four simulated applicants. This resulted in a total of 12 interview scores for each applicant.

The differences between mean global ratings assigned by the four teams to each of the four applicants (between-subject difference, wherein the implicit between-subject factor is assigned by default) was statistically significant ANOVA F(1,3)=19.91, p=.02). The differences were statistically significant at p=.01 for five of the six dimensions, and at p=.02 for the dimension of “awareness of the McMaster program” and for global ratings (Table 2).

The mean global ratings reflected the simulations, in that the applicant programmed to be unsuitable (Applicant D) was identified as such by all four

Table 2

Repeated Measures ANOVA to Detect a Rater or Team Effect in the Interview Subscale Scores and Global Rating (n=4 applicants x 3 teams x 3 raters/team)

<table>
<thead>
<tr>
<th></th>
<th>F (1, 3)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>24.15</td>
<td>.01*</td>
</tr>
<tr>
<td>Awareness</td>
<td>20.32</td>
<td>.02*</td>
</tr>
<tr>
<td>Problem-solving</td>
<td>31.14</td>
<td>.01*</td>
</tr>
<tr>
<td>Self-appraisal</td>
<td>33.54</td>
<td>.01*</td>
</tr>
<tr>
<td>Ability to Relate</td>
<td>34.37</td>
<td>.01*</td>
</tr>
<tr>
<td>Career Goals</td>
<td>28.81</td>
<td>.01*</td>
</tr>
<tr>
<td>Global Rating</td>
<td>19.91</td>
<td>.02*</td>
</tr>
</tbody>
</table>

*p<0.05
teams, the outstanding applicant (Applicant B) was rated outstanding, and so forth. This established a measure of concurrent validity. Not only did the teams all consistently distinguish between suitable and unsuitable applicants, but there was also consistency between teams in terms of overall ratings. When many different interviewers are used, teams may differ in terms of their assessment of the same applicant (Burns & Grove, 1987). To examine this possibility, the mean scores of each team were compared and it was found that there was no significant difference between the means of the global ratings across teams ANOVA $F(3,3) = .50, p = ns$).

The researchers questioned whether the high correlations based on the global ratings might have resulted from the interview teams adjusting their individual ratings to a common or team score, despite being specifically instructed not to do so at the orientation session. In an attempt to explore this possibility, the data were reanalyzed using a total score, a number obtained by summing the ratings on the six dimensions. However, as shown in Table 3, this maneuver resulted in even higher ICCs.

| Table 3 |
|---|---|---|---|---|
| **Inter-team and Inter-rater Intraclass Correlation Coefficients (r) of Global Rating and Team Scores (n=4 applicants x 4 teams x 3 raters/team)** | **Inter-Team** | **Inter-Rater** |
| | 1 Team | 4 Teams | 1 Rater | 3 Raters |
| Global Rating | .66 | .89 | .96 | .99 |
| Total Score | .81 | .95 | .97 | .99 |

**Actual applicants**

Team interview scores (the sum of the three assigned global ratings) for individual applicants ranged from 3/21 to 21/21, with faculty, student/alumnus, and community interviewers using the full 7-point range in assigning ratings. Inter-team and inter-rater reliability of global ratings for the total applicant group ranged from .79 for the single observation, to .91 for the multiple (3 rater) observation.

Concurrent validity was assessed by comparing interview scores with scores obtained by the applicants on other admission criteria (i.e. academic, personal references, and autobiographical questionnaire). Since only those applicants who achieve high ratings on each criterion proceed to the subsequent step in the selection process, this can reduce the range of ratings. Only those applicants who achieved high ratings (5/7 to 7/7) on their autobiographical questionnaires were invited to be interviewed. The effect of this reduction in variance can be illustrated by comparing the correlation between
interview scores and overall collation ranking in the total interview pool with the correlation among these measures in the subgroup accepted into the program. The correlation was .82 (p=.01) in the total interview pool, and .77 (p=.02) among the accepted applicants.

Discussion

The personal interview is the final step in candidate selection and contributes proportionately more weight to the overall process than any other step. It is therefore important to assess its reliability and validity. The high correlations between applicant ratings (inter-rater and inter-team ICC), demonstrate that multiple interviewers were able to clearly differentiate between applicants who had various degrees of suitability to the program using the existing interview criteria and format. The results support the continued use of a semi-structured team interview, and are consistent with other recent research findings (Edwards, Johnson & Molidar, 1990; Posthuma & Noh, 1990; Richards, McManus & Maitlis, 1988).

The fact that “career goals” was the dimension that yielded the lowest correlations with the global rating may be due to the fact that it assesses the clarity of the applicant’s goals rather than their appropriateness. Thus, an applicant with a high rating in this dimension might not have a high global rating, whereas the other five dimensions may be more closely linked to applicant suitability. The guidelines for the dimension “career goals” will be revised to bring them more in line with the other dimensions.

The personal interview assessment form was reported by the interviewers to have a useful format. The interviewers stated that the orientation sessions held immediately prior to the interviews increased the consistency of expectations and ratings, a statement that was supported by the research findings. Since the ICCs for the total score were even higher than those for the team global rating, the global rating will continue to be used as an accurate measure of applicant assessment.

However, as Haglund (1978) cautioned: “When a faculty decides to develop specialized programs to prepare individuals to meet the requirement for admissions it needs to be aware of the decreased efficiency of the admission process” (p. 241, italics added). Through repeated applications for admission to programs, applicants may merely “learn the process” rather than truly improve their suitability. The Admissions Committees of the Faculty of Health Sciences undergraduate programs are currently reviewing the policy and procedures for providing feedback to unsuccessful applicants to be equitable for all applicants, yet not provide coaching for future applications.
Both in-course progress and licensure examination results of students in the B.Sc.N. program will be monitored to ascertain the levels of predictive validity of both the autobiographical letter (Brown, Carpio & Roberts, 1991) and the personal interview.

References


This research was supported by the Neimeier Fund of the School of Nursing, and the statistical analysis provided by the Program for Educational Development of the Faculty of Health Sciences, McMaster University.