# EVALUATING THE COLLABORATIVE CRITICAL CARE NURSING PROGRAM

### Sara R. Frisch

Most Canadian hospitals have problems staffing their critical care areas, partly because of the complexity of the work. This complexity means that hospitals often provide extensive orientations for newly hired ICU nurses and expect them to participate, regardless of their critical care background. A nurse inexperienced in critical care may need six to twelve months after orientation to become comfortable and competent. The orientation and long adaptation period can leave the unit understaffed, in effect, for up to a year even if vacancies are filled quickly.

A provincially-supported demonstration project to address some of these difficulties was set up in 1984. It combined a post-secondary academic component with clinical experience at a tertiary care teaching hospital. It offered credits toward a baccalaureate degree and was designed to prepare nurses who could quickly become effective in clinical practice. Designers felt that graduates with acknowledged competence would only need a brief orientation before beginning to work for a new hospital.

The purpose of the comparative longitudinal project was to evaluate whether the experimental continuing education program was as effective as hospital-based orientations in preparing registered nurses to work in critical care settings at a beginning level. Program developers expected those completing the program to be comparable to hospital-oriented nurses in their knowledge and skills in critical care nursing and, ultimately, comparable in their clinical performance. This is a report of the results of that evaluation.

#### Literature Review

Despite the importance and widespread use of critical care orientations, few evaluations have been published. Two studies (Hansell & Foster, 1980; Houser, 1977) were evaluations of programs at single institutions. Both used measures developed in-house; reliability and validity were not reported in either study. Houser found a complex interaction between educational pre-

Sara R. Frisch, Ph.D. is Director of Nursing Research at the Montreal General Hospital, and Associate Professor in the School of Nursing, at McGill University, in Montreal.

The Canadian Journal of Nursing Research

Spring 1989, 21(1), 5-17

paration, end of orientation examination scores and clinical performance ratings at three and six months. End of course examination scores were related to clinical performance only for nurses with a diploma or baccalaureate degree who had no prior critical care experience. Hansell and Foster compared programmed instructional modules with classroom teaching. The programmed instruction group performed better than the classroom group on an end of orientation test and on head nurse performance ratings at three months. Test performance was not related to educational background nor to previous critical care experience.

These two studies illustrate one problem faced by evaluators of continuing education in nursing: finding valid and reliable instruments. Work on a measure of knowledge is in progress (Toth, 1984, 1986; Toth & Ritchey, 1984), but measures of impact on practice are also needed (Abrahamson, 1984; Dixon, 1978; Kuramoto, 1985; Lloyd & Abrahamson, 1979). Critical care performance standards (Thierer, et al., 1981; Whiteley, 1986) are necessary, but not sufficient bases, for developing such measures.

Another problem arises from the multiple influences on knowledge and practice. Cervero (Cervero & Rottet, 1984) identified three sets of variables in addition to the continuing education program: characteristics of the individual practitioner; the nature of the changes desired; and the workplace social system. Kuramoto cites a study by Cox and Baker (1981) in which supervisor support in the post-course work setting was correlated with whether or not nurses continued to use new skills in their practices.

A third problem is when to measure outcomes. Kuramoto suggests that some continuing education programs may not show immediate effects; reality shock may mask learning for some time. Houser (1977) specifies six months as the time required to adapt to the critical care setting. If masking occurs, then outcome measures should be delayed until learning can be observed. There is a risk, however, that learning may be so modified by experiences during the delay that it is no longer evident when final measures are taken.

These and other problems were encountered in the study described below.

### Methods

# Programs

The experimental program was a seven-week core course to prepare nurses for critical care work. The course was collaboratively designed by faculty at a degree-granting institution and staff at a tertiary care teaching hospital. Faculty were hired by the educational institution and classes were held there.

Clinical experience occurred at the hospital under the supervision of the faculty throughout the seven weeks. Students took examinations, gave presentations and were graded. Those passing the course received credits which could be applied toward a baccalaureate degree. Nurses hired to work at the collaborating hospital were enrolled in the core course in lieu of a critical care orientation.

Comparison programs were located at six tertiary care teaching hospitals; three were in the same city; one in the same province, and two in a neighbouring province. Orientations were designed by hospital staff and lasted from two to six weeks. Usually several people were responsible for teaching and supervision; other staff also contributed. Classes and clinical experience were on the premises. The longer orientations appeared to be similar in their objectives, content and methods to the experimental program. However, none were collaboratively designed and sponsored, and none had a formal affiliation with a faculty of nursing.

# Subjects

All nurses in the second core program (August, 1984) through the sixth (May, 1985) were eligible as experimental subjects. Enrolments ranged from seven to 24 with a mean of 15. All (N=74) were willing to participate in the evaluation research and signed a consent form. Sixty-eight subjects (91%) completed the initial measures (seven weeks), and 50 (67%) completed the follow-up measures (six months). Eighty-five percent of those completing the follow-up were working at the collaborating hospital.

The comparison hospital group consisted of nurses hired for critical care work (full, part-time, float) who had not previously worked in that hospital's critical care area. The number per hospital ranged from five to 17. The project was discussed with approximately 70 nurses in orientation groups of one to 7 people. Sixty-five signed a consent form and entered the study. Of those entering, 48 (74%) completed all seven week measures and 27 (42%) completed all follow-up measures.

# Measures

Measures were designed or selected according to the stated outcome objectives of the experimental program.

Demographic information was collected on age, education, experience, inservice activities and language proficiency.

Knowledge of critical care nursing was measured by a 71-item multiple choice test devised by content experts prior to the start of the project.

Reliability (coefficient alpha) was .8, based on a sample of nurses who had completed earlier critical care programs in the collaborating institutions. A pilot study showed that experienced critical care nurses scored higher than novices; both these groups scored higher than nurses without critical care background.

Nursing approach referred to knowledge and use of independent and collaborative nursing interventions. Nurses' ability to identify these was measured through case studies. Respondents were asked to list the independent ("initiated by a nurse without physician input") and interdependent ("requiring collaboration with a physician") interventions they would use. A panel of experts from several institutions developed a scoring system for the data. Two raters scored the case studies; inter-rater agreement ranged between 85-93%.

Nursing approach was also measured from a description of the nurse's "most challenging patient." Approach was operationally defined by the nurse's ability to give nursing as well as medical diagnoses.

Performance. Performance was measured in several ways. The "Performance Level Self Report" was a 32-item self-rating including aspects common to all critical care areas of the co-sponsoring hospital. The seven-point rating scale ranged from 0= "cannot perform this function satisfactorily" to 6= "perform...with more than acceptable quality of work and speed."

An indirect, self-report assessment of performance was obtained from the follow-up case study. Respondents were asked to rate their *confidence* in being able to care for the patient described without help. A four-point scale was used where 1 indicated little or no help was needed.

The "Most Challenging Patient" indirectly measured performance insofar as more competent nurses were assigned more complex patients. Nurses were asked to describe this patient's physical status, psycho-social history, equipment used, medications and treatments. The description was scored for complexity by counting the challenging aspects, as suggested by the expert panel.

The "Head Nurse Rating" assessed performance through simulated patient assignments. Twelve patient vignettes were selected from 45 submitted by critical care head nurses at the collaborating hospital. Those selected showed moderate agreement among judges about the skill required to nurse the patient, covered a range of skills and represented a variety of patient problems. Head nurses rated participating nurses' competence to care for the 12 patients. Higher scores implied more competence; the maximum score was 96.

Satisfaction. Nurses reported their satisfaction with their performance of critical care nursing skills and with their preparation for work in this setting (through the experimental or orientation program). Five-point rating scales were used.

# Design

Program effectiveness was assessed using a quasi-experimental longitudinal design. Demographic data was collected at entry to the study; the knowledge test, case study and performance level self-rating were completed at the end of the core program for nurses in the experimental group and at six to eight weeks from date of hire for nurses in the comparison group. All measures were readministered at the follow-up, six months after entering the core courses (experimental) or after date-of-hire (comparison).

The experimental group was released from classes or work and data were collected on-site. Comparison hospitals could not release participants, so instructions and forms were mailed for completion on the nurse's own time. A stamped return envelope was provided. Participants were contacted by telephone and asked to return the material if it had not been received back within two weeks of mailing.

#### Results

Demographic characteristics. Demographic information was obtained at entry to the study, so changes between initial and follow-up testing reflect the effects of drop-out. Table 1 summarizes the demographic characteristics of all subjects (entering), and of those in the follow-up, by research group.

At entry, nurses in the experimental group tended to be older, to have more education and inservice and to be less experienced in critical care than those in the comparison group. None of these differences was significant. The same relationships held among nurses remaining to the follow-up. The differences on the variables "previous inservice" and "previous critical care job" were significant (X<sup>2</sup>=4.52 and 4.579, respectively, df=1, p<.05). Comprehension of English was rated as "very good" or "excellent" by over 90% of participants (experimental and comparison) entering and remaining in the study.

Table 2 summarizes the means and standard deviations of entering and follow-up samples for each outcome measure. Data from subjects participating in the follow-up testing were analyzed by a multivariate analysis of variance to adjust for the multiple comparisons being conducted on the data set. The independent variables used were group (experimental or control) and previous critical care experience (none vs some), and the dependent variance

ables were knowledge test scores, performance level ratings, case study scores, most challenging patient, confidence in caring for this patient and head nurse ratings. Measures taken at seven weeks and six months were entered into the analysis as difference scores. Missing data were replaced with group means and F-values were adjusted for the number of such substitutions made on any given variable. All multivariate test statistics (Wilks lambda, Pillai trace and Hotelling-Lawley trace) for group were significant (p<.005) but those for previous experience and the interaction of group and experience were not significant. Table 3 shows the univariate results for group which are discussed below.

Table 1

Demographic Characteristics by Participation in Study and Group

Demographic Characteristic	All S	Subjects	Follow-up	
	Exp	Comp	Exp	Comp
N*	74	63	48	32
Age (mean years)	28.7	26.5	27.9	26.9
Educational preparation				
% hospital/community college	72.2	82.0	72.3	77.4
% degree	27.8	18.0	27.7	22.6
Previous critical care jobs				
% 1 or more	43.5	52.4	37.8	62.5**
Previous critical care courses				
% 1 or more	50.0	39.7	51.0	43.8
Previous Inservice				
% 1 or more	81.1	68.3	85.7	65.6*

<sup>\*</sup> Ns differ slightly for each characteristic due to missing data.

<sup>\*\*</sup> Differences between Experimental and Comparison groups in the Follow-up are significant; p<.05.

Table 2

Means and Standard Deviations on Outcome Measures, All Subjects and Follow-up Samples

Outcome Measure	All Subjects		Follow-up Sample			
Mean			Initial		Follow-up	
SD(N)	Exp	Comp	Exp	Comp	Exp	Comp
Knowledge test	41.73	37.77	42.34	39.11	41.48	41.43
	5.64(67	6.27(47)	5.35(46	5.64(26)	5.41(46)	5.86(26)
Nursing Approach						
Case Study - Independent	% 57.6	76.8	52.9	78.6	90.9	.856
	38(66)	17(49)	36(47)	18(31)	15(47)	.16(26)
Interdependent %	92.3	98.0	90.8	97.9	97.0	.961
	21(64)	7(48)	23(46)	6(30)	12(46)	.2(26)
Challenging Patient:						100000000000000000000000000000000000000
% Answer MedDx					100%	100%
					(30)	(24)
% Answer NDx				60% 33%*		
					(35)	(24)
Performance Self-Rating	104.96	125.79	104.64	128.39	123.02	138.35
	29.3(67)	27.2(43)	28.1(42)	28.7(23)	24.5(42)	23.1(23)
Challenging Patient:						
Difficulty					14.37	14.7
					5.8(33)	6.1(23)
Head Nurse Ratings					71.2	71.7
					14.1(43)	13.0(27)
Satisfaction: Performance					3.16	3.39
					.81(43)	.57(28)
Satisfaction: Preparation					2.72	2.7
and the second s					.79(44)	1.03(27)

Table 3

Multivariate Analysis of Variance Summary Table Experimental vs Comparison Group: Univariate test results

Dependent Variable	DF	F	p .026	
Knowledge test	1,64	5.207		
Case Study: Independent Acts	1,63	17.551	.000	
Case Study: Interdependent Acts	1,61	1.688	ns	
Performance Level	1,58	.514	ns	
Confidence, Caring for Patient	1,56	2.971	ns	
Most Challenging Patient	1,53	.199	ns	
Head Nurse Rating	1,56	.001	ns	

Knowledge. Reliability estimates based on the study samples were  $\alpha$ =.66 at seven weeks; and  $\alpha$ =.55 at six months. The F-test for group on the difference between six month and seven week knowledge test scores was significant. Scores for nurses in the experimental group were slightly lower at the follow-up testing (mean difference = -1.024) while comparison group nurses improved their performance (mean difference = 2.231). A separate analysis of variance on seven week scores, including all subjects in the study at that point, showed the collaborative program group had significantly higher scores than the comparison group (F(1,112)=12.44; p<.001).

Nursing approach. Case study data were scored for accuracy (correctly identifying interventions as independent or collaborative). Accuracy scores in identifying independent interventions ranged from 53%-91%. The univariate F on the difference between seven week and six month scores was significant. The experimental group performed more poorly than the comparison group at the seven-week testing [all subjects: F(1,113)=11.4985, p<.001; subjects in follow-up: F(1,76)=14.175, p<.001]. Unlike the knowledge test, participants in the experimental program improved their performance between testings more than those in the comparison programs (mean difference: .399 vs .069).

Accuracy was over 90% for identification of interdependent interventions at seven weeks and six months (Table 2). The univariate F for groups on the difference score (six months - seven weeks) was not significant.

Diagnoses listed for "most challenging patient" were scored for success in giving an accepted diagnosis. These data were not included in the multivariate ANOVA. All nurses in experimental and comparison groups responded to the medical diagnosis question. More collaborative program nurses gave an accepted medical diagnosis than nurses in comparison settings, but the difference was not significant. However, for nursing diagnoses, significantly more nurses in the experimental group answered the question  $(X^2 = 5.50, df=1, p<.05)$ , and more of their answers (52%) were actual diagnoses.

Performance. The self-report measure's reliability (coefficient  $\alpha$ ) was .95 at seven weeks and .96 at six months. The univariate F-test on the difference between seven week and six month scores was not significant, indicating the two groups had similar changes. There was, however, a significant difference at each testing in how the groups rated themselves. Nurses in the comparison group gave consistently higher ratings than nurses in the experimental group. [see Table 3; initial rating, all subjects, F(1,108)=13.976, p<.001; initial rating, subjects in follow-up: F(1,63)=10.43,p<.01)]. The pattern held for all four face valid subscales on this measure (nursing process, professionally-related, skills/equipment and cognitive/assessment) as well as the total score.

A similar pattern was seen in nurses' ratings of confidence in their abilities to care for the patient described in the follow-up case study. Nurses in the comparison group rated themselves as more able to care for the patient without help (mean = 1.52) than those in the experimental group (mean = 1.87). However, the univariate F-test on this six month measure was not significant.

As shown in Table 3, there were no significant differences between the groups on performance measures that were not self-ratings. The mean difficulty scores for the nurse's "most challenging patient" were 14.3 for experimental and 14.7 for comparison nurses. This suggests both groups were being assigned patients of similar difficulty within settings, and thus were perceived as similarly competent.

The means on the head nurse scale were 71.2 for the experimental group and 71.7 for comparison subjects. The difference was not statistically significant. The reliability (coefficient  $\alpha$ ) of this measure was .933.

Satisfaction. Table 2 shows nurses' mean ratings on two global satisfaction items - preparation for critical care nursing and performance as a critical care nurse. These were not significantly different. There were differences on specific items. Nurses in the experimental group perceived "performance as a professional nurse (self-directed, ethically and legally responsible)" to be more important (mean = 4.79) than nurses in comparison settings (mean = 4.44; F=8.16, p<.01). Nurses in the comparison group gave higher ratings on items such as "perform tasks required", "operate specialized equipment", "perform technical skills", "give total patient care" and "patients' positive perception of me". Nurses in the experimental group gave higher ratings to items such as "support patients' families", "cope with legal responsibilities" and "take a stand on moral/ethical issues".

#### Discussion

The results reported above must be interpreted with caution. They do not demonstrate any overall advantage or disadvantage for the collaborative program compared to the other orientations. In terms of knowledge, collaborative program nurses performed better at the end of their critical care preparation than the comparison group, but the difference disappeared by six months. Generally, the data showed no significant differences between the two groups at six months on measures that were not self-report: the knowledge test, the difficulty of "challenging patients", and the head nurse ratings. On measures that were self-report - the performance self-rating and confidence in nursing a patient like the one in the case study - collaborative program nurses rated themselves lower than those in the comparison group. The nursing approach measures and satisfaction items suggest the two groups differed somewhat in their nursing focus.

Assuming the actual competence of the two groups of nurses was comparable, the self-report data may reflect a lack of confidence on the part of experimental group nurses. Perceptions of competence have been shown to be related to confidence but not necessarily to more objective measures of performance (Bucher & Stelling, 1977). Lower confidence could result if supervision during the core program was closer and the opportunity for independent decision making was rarer than in the hospital orientations.

The work setting could also have affected follow-up knowledge test performance and confidence. Data from head nurse ratings suggest the critical care units in the collaborating hospital were more specialized than those in comparison hospitals. The knowledge test was a general test. Nurses at the collaborating hospital may have been tested on general knowledge after experiencing work emphasizing specialized knowledge, while nurses in comparison hospitals may have received a broader experience more congruent with the test content. Confidence might also have been affected by setting effects. Again, head nurse data indicated that patients treated at the collaborating hospital had more difficult and complex problems than those at the comparison hospitals. Nurses might take longer to learn the ropes in this setting.

Overall, the two groups of nurses seemed satisfied with their respective programs. The greater reported satisfaction of comparison nurses on items dealing with tasks and skills may also be related to differences in confidence.

This study suffered many of the problems inherent in program evaluations in field settings. It compared two groups of unknown equivalence because assignment was determined by the hiring decisions of participating hospitals; randomization was not feasible. Measures were administered under different circumstances - on-site for the experimental group, by mail for the comparison group. Mailing measures was undesirable for many reasons but was unavoidable. Most problematic was the drop-out rate for nurses in the comparison group. This seriously affects the internal validity of the study.

The absence of accepted, widely used ways to assess the research variables meant that most measures were developed within the context of the research project. All were paper and pencil; observation and qualitative data gathering methods could not be used for reasons of resource constraints.

The impact of setting on performance could not be examined because setting was confounded with research group. Variation in work conditions was greater for nurses in the comparison groups than those at the collaborating hospital, potentially making it more difficult to find "true" differences between the groups.

On the other hand, the research had its strengths. The initial sample size was larger than that used in other studies. Detailed demographic information was collected in order to assess similarity between the groups because random assignment was not possible. The project was multi-site rather than confined to a single hospital. The knowledge test was developed independently of the core program and the orientations and was not used to assess performance in these learning situations. Several nurses in the core program commented that they felt the test was a better measure of critical care knowledge than the course examinations. Finally, efforts were made to get at the issues of knowledge and performance in a variety of ways because of the measurement problems. Belief in the validity of the findings is stronger insofar as results converge.

Some of the difficulties encountered in this study will be faced by any nursing administrator who is trying to decide if a new or revised critical care program is better than its predecessor. Nursing staff will not be randomly selected or assigned to programs, there will not be equivalent comparison groups, learning will be affected by experiences in the setting. One problem, though, can be addressed: that of developing valid and reliable measures of performance. A paper and pencil measure, such as the knowledge test used here, may be useful, but it will not necessarily reflect performance in the setting. Other approaches should be tried, including performance ratings and analyses of the nursing care required - and provided - on the unit.

The general problem of evaluating critical care preparation cannot be addressed without greater agreement about what constitutes acceptable performance. It is also important to examine the work setting as well as the educational program to understand what nurses learn and how they perform. The setting can affect what is retained, what more is learned and what knowledge and skills are possessed and expressed. These phenomena should be studied directly, not in the context of a program evaluation.

The formal evaluation results were not clear, but there is other evidence of the program's success. The collaborating hospital has continued to cooperate and head nurses' have expressed satisfaction with the results. This suggests the core program was adequately preparing nurses. Furthermore, the program has become attractive to other hospitals. A model has evolved whereby the educational institution provides the classroom content on its premises to nurses from many hospitals. Clinical experience is provided and supervised by the hiring hospital. The impact of this modified program on critical care staffing problems remains to be investigated.

#### REFERENCES

- Abrahamson, S. (1984) Evaluation of continuing education in the health professions. Evaluation and the health professions, 7(1), 3-23.
- Bucher, R. & Stelling, J.G. (1977) Becoming professional. Beverly Hills, CA: Sage.
- Cervero, R.M. & Rottet, S. (1984) Analyzing the effectiveness of continuing professional education: An exploratory study. Adult Education Quarterly, 34, 135-146.
- Cox, C.L. & Baker, M.G. (1981) Evaluation: The key to accountability in continuing education. Journal of Continuing Education in Nursing, 12(1), 11-19.
- Dixon, J. (1978) Evaluation criteria in studies of continuing education in the health professions: A critical review and suggested strategy. Evaluation and the health professions, 1(2), 47-65
- Hansell, H.N. & Foster, S.B. (1980) Critical care nursing orientation: A comparison of teaching methods. Heart and Lung, 9, 1066-1072.
- Houser, D.M. (1977) A study of nurses new to special care units. Supervisor Nurse, 8(7), 15-22.
- Kuramoto, A.M. (1985), Research on continuing education in nursing. Annual Review of Nursing Research. 149-170.
- Lloyd, J.S. and Abrahamson, S. (1979) Effectiveness of continuing medical education: A review of the evidence. Evaluation and the health professions, 2, 251-280.
- Thierer, J., Perhus, S., McCracken, M.L., Reynolds, M.A., Holmes, A.M., and Berkowitz, D.S.(Eds) (1981) Standards for nursing care of the critically ill. American Association of Critical-Care Nurses. Reston Publishing Company: Reston, VA.
- Toth, J.C. (1984) Evaluating the use of the Basic Knowledge Assessment Tool (BKAT) in critical care nursing with baccalaureate nursing students. *Image: The Journal of Nursing Scholarship*, 16, 67-71.
- Toth, J.C. (1986) The Basic Knowledge Assessment Tool (BKAT) validity and reliability: A national study of critical care nursing knowledge. Western Journal of Nursing Research, 8, 181-195.
- Toth, J.C. and Ritchey, K.A. (1984) New from nursing research: The Basic Knowledge Assessment Tool (BKAT) for critical care nursing. *Heart and Lung*, 13, 272-279.
- Whiteley, M. (1986) Developing national standards for critical care nurses. Canadian Critical Care Nursing Journal, 3(4), 7.

This research was conducted with the collaboration and cooperation of many people. The contribution of Margaret Malone, B.A.A.N., M.A. as research assistant is particularly acknowledged.

The paper is based on one presented at the CAUSN National Nursing Research Conference, Hamilton, Ontario, June 6-7, 1987.

The project was supported by a grant from the Ontario Ministry of Health: I.J. Bajnock, Ryerson Polytechnical Institute and D. Wylie, Toronto General Hospital - Principal Investigators.

# RÉSUMÉ

# Évaluation d'un programme collaboratif de soins intensifs

Un projet de démonstration élaboré sur un mode collaboratif afin de préparer les infirmières à exercer dans les unités de soins intensifs a été évalué par rapport aux cours d'initiation en milieu hospitalier, à l'aide d'un schéma longitudinal quasi expérimental. L'échantillon comprend 74 infirmières inscrites au programme d'études entre août 1984 et mai 1985 et 65 infirmières exerçant dans l'une des six unités de soins tertiaires utilisées à des fins de comparaison. L'évaluation des connaissances, du rendement et de la satisfaction a été effectuée sept semaines et six mois après l'inscription au programme. Pour le premier test, les infirmières du programme collaboratif affichent un rendement meilleur au chapitre des connaissances mais évaluent moins bien leur rendement que les autres infirmières. Cinquante infirmières du programme collaboratif (67 %) et 27 infirmières des autres unités (42 %) se sont prêtées au test après six mois. Le résultat des tests de connaissances ne permet pas d'établir de distinction entre les deux groupes; les infirmières du projet de démonstration continuent d'évaluer leur rendement moins bien que les autres infirmières mais les évaluations de l'infirmière en chef et autres indices de rendement ne permettent pas de déceler de différences entre les deux groupes. Le degré de satisfaction suscité par la préparation au travail dans une unité de soins intensifs (programme ou initiation) est identique dans les deux groupes. Les difficultés propres à l'évaluation dans le cadre des programmes d'éducation permanente sont ensuite débattues.