THE DEVELOPMENT OF A TOOL TO ASSESS LEVELS OF STRESS AND BURNOUT

Virginia Skinner, RN, RM, BHsc, MNH,, Research Project Officer, Northern Sydney Central Coast Health, New South Wales, Australia. PhD candidate, University of New England, Armidale, Australia.

vskinner@nsccahs.health.nsw.gov.au

Kingsley Agho PhD, Msc, Associate Lecturer, School of Public Health, The University of Sydney, New South Wales, Australia.

Trish Lee-White, RN, RM, BSoc Sci, Obstetric Clinical Nurse Consultant, Northern Sydney Central Coast Health, New South Wales, Australia.

Dr Judy Harris PhD, Lecturer, University of New England, Armidale, New South Wales, Australia.

Accepted for publication October 2006

Key words: face validity, internal consistency, test re-test.

ABSTRACT

Objective:

To pilot test the reliability and validity of a newly developed tool measuring nursing and midwifery staff stress and burnout.

Design:

Descriptive survey.

Setting:

Public hospital, aged care facility and university.

Subjects:

For the pilot study a total of forty-nine (n=49) nurses and midwives, selected by convenience sampling, were sent an initial pilot questionnaire. The return rate was seventy per cent initially and the return rate on the second mail out was forty-nine per cent.

Main outcome measure:

To determine reliability and validity of a new tool that explores nurses' and midwives' perceptions of stress, burnout and control over their working environment.

Results:

Face validity, test-retest reliability, internal consistency and principal component analysis were established. Overall Cronbach's alpha was 0.87 indicating good internal consistency for the stress/ burnout element of the questionnaire. The test-retest reliability intraclass correlation coefficient reported 0.30 - 0.90 for all six sub scales which were developed for both parts of the questionnaire.

Conclusion:

The pilot study indicates that it is possible to construct a valid and reliable instrument to assess nurses' and midwives' perception of stress and burnout.

INTRODUCTION

There have been various tools and instruments used previously in the literature to assess stress (Maslach and Jackson 1981; Jewell and Siegall 1990; Stordeur et al 2001; Goldberg 1978; Rahe and Tolles 2002). These tools were reviewed but seemed dated and no longer pertinent to current issues and concerns faced by midwives and nurses in their challenging contemporary clinical work environments. Tools must possess basic attributes (validity and reliability) that assure dependable measurement of the variables under investigation (Waltz et al 1991).

Norbeck (1985) suggests that there are four minimum standards necessary for the adequate evaluation of an instrument for use in research. These standards should include at least one type of content validity, one type of construct (or criterion-related) validity and two types of reliability testing (internal consistency and test-retest). This paper will explore the issues concerning validity and reliability as they relate to the development of a new, original questionnaire.

RELEVANT LITERATURE

Validity and reliability

Validity is the extent to which a study using a particular tool measures what it sets out to measure (Polit and Hungler 1997). The testing of validity is not exactly proven but rather supported by an accumulation of evidence. A researcher does not totally validate a tool per se but more an application of the tool. Unlike reliability, there are no simple statistical calculations to assess validity (Polit and Hungler 1997).

Polit and Hungler (1997) define content validity as the adequacy of the content area being considered. A subtype of content validity is face validity and this investigates whether an instrument is calculating the appropriate construct (Dempsey and Dempsey 1992). Dempsey and Dempsey (1992) define face validity as whether the items within an instrument measure the variables in a specific content area. Construct validity is another standard to be achieved in developing a new tool. It measures a specific construct or hypothetical trait, such as: grief, intelligence or prejudice pertaining to an instrument (Dempsey and Dempsey 1992). Factor analysis is one way of establishing construct validity. Factor analysis is calculated to statistically define subgroups for the indexes created by the researcher (Field 2005).

The reliability of a tool is a criterion for assessing quality (Polit and Hungler 1997). A tool is reliable when a repeat use of the tool consistently measures what it is measuring in exactly the same way (Dempsey and Dempsey 1992). This is also an assessment of the stability of a tool. This approach has certain disadvantages. Mood, physical condition, knowledge and attitudes do change between measurements despite the stability of a tool (Polit and Hungler 1997). The time period for test-retest reliability was chosen (two weeks) so that it was long enough for individuals not to remember specific responses and not too long so that maturation and learning would most likely not occur in this time frame, affecting the answers. Reliability is expressed as a number (a coefficient). The higher the number the more reliable. Rarely is a tool perfectly reliable and is often reported as 0.80, 0.70 or 0.60, as opposed to 1.0 (Dempsey and Dempsey 1992).

Other tools used in the literature

The Maslach (Maslach and Jackson 1981) Burnout Inventory primarily studied three dimensions of burnout in nurses. These included: providing nursing care in an atmosphere of depersonalisation; depicting nurses attending to tasks and patients without any emotional feeling; emotional exhaustion and perceptions of reduced personal accomplishment; all indicative of feelings of low morale. It did not study specific stressors such as high patient acuity or low staffing levels. Jewell and Siegall's (1990) Nurse Stress Index correlated stress scores with job satisfaction, not behavioural aspects. It aimed to identify issues of occupational stress.

Stordeur, D'Hoore and Vandenberghe (2001) used a nursing stress scale which identified three sources of stress: physical, psychological and social environments. They conducted their study on leadership, organisation stress, and emotional exhaustion among nursing staff. These researchers did not look at the behavioural aspects of individual nurses. Goldberg (1978) designed the General Health Questionnaire-12 which detected psychological indicators of ill health. It was used in occupational and community settings as opposed to nursing workplaces. The questionnaire's main purpose evaluated psychiatric morbidity.

AIMS

A pilot study undertaken before embarking on a main study is of the utmost importance (Hundley and van Teijlingen 2002). A pilot study was undertaken to primarily establish a feedback mechanism, ensure the survey was user-friendly, ensure the items in the survey covered the content area of interest and establish a degree of reliability. This paper explores issues associated with the development of a new, original questionnaire and reports on reliability and validity determined by a pilot study. The paper discusses the imperative issues of a comprehensive pilot process that assesses not only the questionnaire but ensures that it is possible to acquire meaningful data and analysis.

METHOD

Development of survey

Generation of items for the draft questionnaire

The process of developing a comprehensive questionnaire commenced with the accumulation of literature and other questionnaires from the area of interest. The questionnaire was designed and developed specifically for the study. Items were generated from a literature review. Search terms used were: stress, burnout, personality, and behavioural characteristics. The questionnaire consisted of three sections. The first section obtained demographic information. The second section comprised thirty-eight items related to stress and burnout. The third section comprised fifteen questions related to personality and behavioural aspects exhibited in particular scenarios, known as vignettes (Polit and Hungler 1985). Additionally a comprehensive accumulation of information from the area of interest came from the first hand knowledge and experience of the researcher in the health care workforce.

Expert advice from academic and clinical experts in the field included a university presentation feedback session. The questionnaire was presented for comment to a post-graduate research residential school in October 2003 which included students and lecturers at the university. This university research residential school is an opportunity for students to present their research and receive feedback, which provided an excellent venue to present the questionnaire in its first draft form. The original questionnaire was presented and many changes were implemented from this presentation.

Changes to questions relating to stress and behavioural aspects

The residential school feedback provided a comprehensive list of changes and additions from the original draft questionnaire. All changes suggested were incorporated into the new questionnaire. The first addition included: 'Tick one box' to ensure participants answered single responses only. The word 'inefficient' replaced 'poor'. The words 'In your work' were added to the question: 'Are you constantly looking for a challenge?' The word 'suffered' as in suffering pain was changed to 'experienced'; 'stress that keeps you moving' was changed to 'stress that keeps you motivated'.

The questions: 'How long was your holiday at one time over the last year?' and 'How often did you holiday for more than one day over the last year?' were condensed into one question: 'How long was your longest holiday at one time over the last year?' One of the options for this question was changed to 'seven to ten weeks' from 'seven or eight weeks'. The phrase 'not having enough time to attend to the quality of care of clients' was changed to 'is your workload too excessive at times to provide quality of care for clients'. The question: 'Do you think you are well-suited to this particular type of work?' was changed to 'Do you think you are suited to the particular type of work you are doing?'

Changes to questions relating to demographics

Changes were made as indicated. The area of 'nursing management' was included in the 'areas of work' which had not been included in the original draft questionnaire. The sentence: 'If you work in more than one area please write in the space provided starting with the most predominant area first' was added to elicit the predominant area. It was recognised that nurses often work in many different specialties and it was considered important to know the primary or most usual specialty work place. Added to the demographics were: the hours worked and whether the nurse or midwife worked permanent, fulltime, part time or casual.

Sampling population

Forty-nine respondents were included in this pilot study. The sample size was chosen to provide adequate information on reliability and a certain degree of face validity. Respondents included eighteen registered nurses from aged care facilities, thirty midwives from the central coast of New South Wales and one doctoral student (who was a registered nurse) from the university. The subjects were selected by convenience sampling. The sample was selected because of geographical accessibility. Ethics approval was received from the appropriate authorities.

Although the sample was a mixture of midwives and registered nurses working in competely different areas, the sample seemed indicative of what the main study sample would resemble. The surveys were distributed through the nursing unit managers to the registered nurses and midwives and directly to the university student. The main study could have included some of the pilot study respondents, but the researcher was unable to determine if there was such overlap.

Classifications and coding legends for questionnaire

Classifications or categories were developed for the first two parts of the questionnaire with numerical values attached to extreme, moderate, fair and nil/negligible levels of stress, control and self-imposed pressure. Questions from each part of the questionnaire were categorised into the following sub-scales: work environment, burnout, control, job satisfaction, stressors and prefers working alone. Ordinal data for the stress/burnout and personality surveys was obtained using a Likert ranking scale which designated level of stress or behaviour expected to be found in that situation. For example, the answers were: 'never', 'occasionally', 'frequently', 'most of the time' and 'always'. The answers were ranked from nil stress to extremely stressed on opposite ends of the scale.

There were eight questions which included 'never/occasionally' in the same box as the answer or option. For example, the question: 'How often do you feel emotionally drained at work?' would rate a high stress rating if the respondent answered 'always' and a nil/negligible stress rating if the answer was 'never/occasionally'.

All parts of the survey used different indexes of coding to accommodate the varying, required levels. For both the stress/burnout and personality/behaviour components of the survey, specific indexes were used for the subscales (see table 1).

Table 1:

Stress/burnout and personality/behaviour subscales/indexes

Subscales	Questions / Indexes
Work environment	Frequency of stress / 0-4 Excessive workload / 0-4 Rush to complete tasks / 0-4 Finishing late / 0-4 Treated with respect by clients / 0-4 Organisational support / 0-4 Work colleagues unsupportive / 0-4 Expectation of 'stress' free environment / 0-2
Burnout	Apathy / 0-4 Low morale / 0-4 Feeling undervalued / 0-4 Feeling overwhelmed / 0-4 Feelings of incompetence / 0-4 Increasing anxiety / 1-6 Fatigue / 1-5 Emotionally drained / 1-5 Loss of empathy for colleagues / 0-2 Loss of empathy for clients / 0-2 Burnout unavoidable / 0-2
Control	Powerlessness / 0-4 Decision-making / 0-4 Motivated by maintaining control / 0-4
Job satisfaction	Suited to work / 0-4 Enjoying type of work / 0-4 Change area of practice / 0-3 Leave professional discipline / 0-3 Frequency of job dissatisfaction / 0-7
Psychosocial stressors and symptoms	Sleeplessness / 1-6 Depression / 1-6 Frequency of sleeplessness / 1-5 Headaches / 1-5 Stress considered healthy / 0-2 Stress requiring treatment / 0-7 Mental health leave / 0-6 Helplessness / 1-7 Frequency of depression / 0-6 Length of holiday / 0-8
Personality / behaviour	Working independently / 0-4 Achieving more than time allows / 0-4 Expect more than reasonably possible / 0-4 Irritability / 0-4 Pushed for time / 0-4 Difficulty slowing down for procedures / 0-4 Working at high performance / 0-4 Arriving early for appointments / 0-4 Reporting sick if unwell / 0-4 Continuing work if unwell / 0-4 Keyed up on most days / 0-4 Tendency to perform many tasks / 0-4 Constantly looking for challenge / 0-4 Strong sense of commitment / 0-4 Reactions when irritable / 0-4

PROCESSES FOR DETERMINING TEST-RETEST RELIABILITY

For test-retest reliability, scores on the two sets of responses are correlated statistically to yield a coefficient referred to as the correlation coefficient. If the results are the same or similar, the coefficient will be high – say 0.90 and the instrument is said to have high test-retest reliability. The first survey was distributed by the nursing unit manager and the second survey was mailed by the researcher. There was no way of controlling the location where the respondent completed the survey. The directions to complete the survey were exactly the same.

RESULTS

Response rate

The average age of the pilot respondents was 47.9 years and the average number of years in the nursing profession was 24.2 years. Thirty-five questionnaires were returned out of forty-nine distributed resulting in a seventy-one per cent return rate initially. A test–retest procedure was followed. Eight respondents remained anonymous and necessarily these respondents were not sent another questionnaire. Respondents were able to be identified by placing their contact details on the questionnaire. The remaining twenty-seven respondents were sent another questionnaire two weeks later. Twenty-four respondents returned the second questionnaire, giving a test-retest return rate of forty-nine per cent of the postal questionnaires distributed. Twenty-four respondents were therefore used in the pilot data analysis.

Data analysis

As the sample was small, an average was calculated for any numerical data that were missing (mean imputation method). Overall reliability of the scale was calculated by Cronbach's alpha indicating internal consistency. The Spearman's rank order correlation was employed to analyse inter-item, item-total correlation and correlations between subscales. Spearman's rank order correlation was also used with the intraclass correlation coefficient to estimate the degree of resemblance or reliability of the subscales for the preliminary and final versions of the pilot questionnaire (that is, the test and retest scores). The intraclass correlation coefficient was used with the continuous data. Principle component analysis using factor analysis was employed to produce the variables that are highly loaded or pertinent to midwives and nurses.

Internal consistency

The overrall Cronbach reliability level for internal consistency for the total and subscales between the preliminary (first test) and final (second test) versions of the questionnaire were calculated. The result for the first test was 0.87 and 0.82 for the second test for the stress and burnout component of the questionnaire. The second correlation coefficient is only marginally lower than the first, to be expected when questions pertain to aspects of behaviour and stress. This level should be at least 0.70.

One question was not included because one hundred per cent of the nurse and midwife participants responded 'no' to the question: 'Would you expect your environment to be 'stress' free?' Nurse and midwife respondents perhaps felt there is an expectancy that there will be certain levels of workplace stress with which to contend. For the personality traits component of the questionnaire, only the first five questions were used as they showed good reliability for internal consistency (> 0.70) for the preliminary (first test) and final (second test) versions of the questionnaire (see table 2).

Table 2:

Internal consistency reliability values (Cronbach's alpha) for preliminary and final versions of personality/ behaviour component

Question	Preliminary	Final
Achieving more	0.8354	0.7957
Unreasonable expectations	0.8146	0.7671
Irritability	0.8454	0.736
Pushed for time	0.8387	0.7938
Difficulty slowing down	0.8781	0.826
Total	0.8708	0.8208

Test-retest reliability

Table 3 shows the test-retest reliability estimates (see column 3). From the table, all subscales in retest reliability were moderately correlated (r = 0.47 - 0.69) (see table 3, column 2) and the correlation coefficients are between 0.30 - 0.62 (see table 3, column 1) for the preliminary version of the test however stress and burnout subscales show lower correlation coefficients. This fact is supported by Stevens (1992) who reports that the strengths of the relationship or association depends on context and in some cases where the correlation is low does not imply that the outcome has no useful significance. Kline (1999) reaffirms this by reporting that when looking at psychological constructs, realistically lower correlation coefficients are more acceptable because of diversity of constructs being measured. This may be an indication that the data aggregated for these particular questions were multi-dimensional, not uni-dimensional. Stevens (1992) reports that most things have multiple causes and in these cases it is difficult to account for a big variance with just one single cause. Even though a small correlation is identified, this could make a substantial contribution for determining and evaluating strategies for reduction of stress. Losing this descriptive information might be detrimental for understanding symptoms of stress. For example, if high correlation coefficients are identified, these areas may be easier to address in possibly alleviating stress areas. The intraclass correlation coefficient ranged from 0.30 to 0.92 with stress and burnout showing a high resemblance between pre and post test (see table 3, column 4) which indicates that the instrument has high test-retest reliability for these two subscales. Average paired responses (for test-retest scores) were 27.5 out of 38 (73%) the same for the stress/burnout component of the questionnaire and 9.3 out of 15 (61%) the same for the behavioural aspect.

Table 3:

Spearman's rank order correlation- rho, intraclass correlation coefficient (ICC) and factor analysis method for the six sub-scales

Subscales	Consistency (Spearman's rho)	Retest	ICC	Factor1
Work environment (8-item scale)	0.62	0.70	0.44	0.33
Burnout (11-item scale)	0.35	0.47	0.92	0.90
Control (3-item scale)	0.62	0.74	0.30	-0.30
Job satisfaction (5-item scale)	0.58	0.68	0.73	0.54
Stress (10-item scale)	0.30	0.50	0.81	0.91
Personality (15-item scale)	0.52	0.64	0.52	0.53

Factor analysis

In social sciences, issues or items are often measured that cannot be directly measured (latent variables). Stress and burnout cannot be measured directly: they have numerous facets. However different aspects of stress and burnout can be measured. Stress levels, ideas of motivation, and enthusiasm can be assessed. Factor analysis shows whether these measures reflect a single variable. Specifically to test whether these numerous variables driven by one underlying variable. Principal component analysis and factor analysis are techniques for identifying clusters or groups of variables (Field 2005). The factor analysis for this pilot exercise indicated that stress and burnout had high loading factors (see table 3, column 4). The principal component analysis identified stress and burnout to be two major factors that nurses' and midwives' experience in their work environment.

DISCUSSION

Based on feedback from respondents in the research residential school, the format was changed to enhance user friendliness. The initial questionnaire contained the demographic information at the beginning, which was then placed at the end of the document. The residential school respondents suggested that immediately requiring personal details to be divulged might inhibit or restrict subsequent responses.

The pilot study and accompanying analysis for this newly devised questionnaire showed good overall reliability. For the personality/behaviour component of the questionnaire, only the first five questions were found to be reliable based on test-retest processes. Despite this finding, the remaining eleven questions were not deleted from the final version of the questionnaire as it was believed that these questions could also provide a basis for descriptive statistics. If reliability is found to be low in the main study, this would be a limitation of this tool. It was concluded that there was a strong statistical correlation (X = 0.86)

between stress and burnout ($p<0.05$) which suggests that nurses and midwives experiences of stress may increase burnout levels.

Limitations

Meticulous attention to appropriate piloting strategies identified weaknesses in the original questionnaire discussed in this paper. If not addressed, questions regarding coding, classifying, analysing and discussing subsequent findings could be expected. One of the limitations of this tool was the inclusion of questions with low correlation co-efficients. The sample size of the pilot survey would probably not be indicative of the general population, but increased numbers in the main study could address this issue. The effects of or reasons for non-responders cannot be analysed as who was given the questionnaires was not known.

A forty-nine per cent return rate was considered adequate for this pilot study but non-response bias needs to be considered (Polit and Hungler 1997). Those who returned anonymous questionnaires but did not want to participate in the second test were of varying characteristics to those who participated. They were mostly in the aged care, neonatal or midwifery professions and had worked in the profession anywhere from two to forty-one years.

The final questionnaire was distributed to 1366 nurses and midwives in Australia and ultimately achieved a forty-one per cent response rate. The time, cost and energy to reach the distribution and collection stage of this research project warranted careful piloting as described here. It is expected that the findings will include information useful to the nursing and midwifery professions, as well as employers in the Australian health care industry.

CONCLUSION

The pilot study associated with the development of a new questionnaire demonstrated that it is possible to construct a reliable instrument to assess nurses' and midwives' perceptions of stress and burnout in their working environment. Comprehensive attention to careful survey development, adequate feedback from appropriately selected pilot respondents and detailed adherance to high quality piloting principles and strategies yielded significant information that informed the final survey. Future studies with this instrument on bigger populations and in different cultural and socio-economic settings are needed to develop a generalisable conclusion about this questionnaire as well as on nurses' and midwives' perceptions of stress and burnout.

REFERENCES

Dempsey, P.A. and Dempsey, A.D. 1992. *Nursing research with basic statistical applications*. Boston: Jones and Bartlett Publishers.

Field, A. 2005. *Discovering Statistics using SPPS*. London: SAGE Publications.

Goldberg, D. 1978. *Manual of the General Health Questionnaire*. Windsor: NFER-Wilson.

Hundley, V. and van Teijlingen, E. 2002. *The role of pilot studies in midwifery research*. Midwives, 5(11):372-374.

Jewell, L. and Siegall, M. 1990. *Contemporary Industrial Organisational Psychology*. St Paul, New York: West Publishing Company.

Kline, P. 1999. *The handbook of psychological testing*. London: Routledge.

Maslach, C. and Jackson, S.E. 1981. *The measurement of experienced burnout in C.L. Cooper The international library of clinical writings in psychology 1: industrial and organizational psychology*. C.L. Cooper. New York: New York University.

Norbeck, J.S. 1985. *What constitutes a publishable report of instrument development?* Nursing Research, 34(6):380-381.

Polit, D.F. and Hungler, B.P. 1997. *Essentials of Nursing Research. Philadelphia*: Lippincott.

Rahe, R. and Tolles, R. 2002. *The brief stress and coping inventory: a useful stress management instrument*. International Journal of Stress Management, 9(2):61-70.

Stevens, J. 1992. *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Stordeur, S., D'Hoore, W. and Vandenberghe, C. 2001. *Leadership, organizational stress, and emotional exhaustion among hospital nursing staff*. Journal of Advanced Nursing, 35(4):533-542.

Waltz, C.F., Strickland, O.L. and Lenz, E.R. 1991. *Measurement in Nursing Research*. Philadelphia: F.A. Davis Company.