

Psychometric Evaluation of the ProQOL Version 5 for Assessing Compassion Satisfaction, Burnout and Secondary Traumatic Stress in Nurses

Joseph L. Hagan¹

¹Department of Pediatrics, Section of Neonatology, Baylor College of Medicine, Houston, USA

Correspondence: Joseph L. Hagan, Assistant Professor, Texas Children's Hospital - Pavilion for Women, Room PFW F0720.54, 6651 Main Street, Houston, TX 77030, USA.

Received: July 27, 2019

Accepted: August 16, 2019

Online Published: August 19, 2019

doi:10.20849/ijns.v4i3.620

URL: <https://doi.org/10.20849/ijns.v4i3.620>

Abstract

Background: The Professional Quality of Life scale is commonly used to measure Burnout, Secondary Traumatic Stress, and Compassion Satisfaction in nurses despite a lack of research regarding its psychometric properties in this population. This study evaluates the reliability and validity of the ProQOL version 5 in nurses.

Methods: Chronbach's alpha was used to assess reliability and Confirmatory Factor Analysis was used to test the validity of the three factor structure.

Results: Reliability of the Burnout subscale is poor in nurses ($\alpha=.53$). Confirmatory Factor Analysis revealed the tool's three factor structure to not be valid in the nursing population ($p<.001$).

Discussion: Despite being widely used, the ProQOL is not a valid or reliable tool in the nursing population. Recommendations are provided for creating a new tool to measure Burnout, Secondary Traumatic Stress and Compassion Satisfaction in nurses.

Contribution of the Paper:

What is already known about the topic:

Burnout, Secondary Traumatic Stress and Compassion Satisfaction are important to study in the nursing population since these constructs can affect job performance and patient safety.

By assessing nurses' Burnout, Secondary Traumatic Stress and Compassion Satisfaction, targeted interventions can be created to enhance work efficiency and enjoyment, thereby improving nurse job satisfaction, nurse retention and patient outcomes.

What this paper adds:

Despite being commonly used in the nursing population, the ProQOL is not a valid or reliable tool for measuring Burnout, Secondary Traumatic Stress and Compassion Satisfaction in nurses.

A new tool for measuring Burnout, Secondary Traumatic Stress and Compassion Satisfaction in nurses is needed.

Keywords: burnout, compassion fatigue, nurses, psychometrics

1. Background

Charles Figley (1995) originally developed an instrument called the "Compassion Fatigue Self Test" which has been modified to create a number of different instruments to assess compassion fatigue. However, these instruments have been found to possess psychometric deficiencies (Figley & Stamm, 1996; Jenkins & Baird, 2002). The Professional Quality of Life (ProQOL) scale was developed in an effort to improve the psychometric properties of the Compassion Fatigue Self Test (Stamm, 2005) and to create an instrument that measures the "positive and negative effects of working with people who have experienced extremely stressful events" (Stamm 2010). The ProQOL is composed of subscales that measure three constructs: Compassion Satisfaction, Burnout, and Secondary Traumatic Stress. Compassion Satisfaction is the positive aspect of a helper's professional quality of life, while the Burnout and Secondary Traumatic Stress subscales measure the negative aspects. Exhaustion, frustration, anger and depression are characteristics of Burnout, while fear and occupational trauma are

associated with Secondary Traumatic Stress (Stamm, 2010). The terms “compassion fatigue” and Secondary Traumatic Stress are used interchangeably (Figley, 1995). The pleasure derived from helping others and/or benefiting society through work results in Compassion Satisfaction (Stamm, 2010). See Stamm (2010) for a more thorough description of the three constructs and history of the tools developed to measure them.

Use of the ProQOL for research in the nursing population is very common. In fact, the ProQOL is the tool most commonly used to assess compassion fatigue, yet a published psychometric evaluation of the ProQOL is lacking (Watts & Robertson, 2015; Bride, Radey, & Figley, 2007). A search of the term “ProQOL” in the EBSCOhost Nursing Reference Center database on June 12, 2019 yielded 489 publications. Although, according to its creators, the ProQOL is not intended to be used as a diagnostic test (Stamm, 2010, page 18) many published studies have used the ProQOL to quantify the “prevalence” of Compassion Satisfaction, Burnout and Secondary Traumatic Stress. As one example, the study by Young et al. (2011) addressed the question, “What is the prevalence of Compassion Satisfaction, Burnout, and Secondary Traumatic Stress in heart and vascular nurses?” Another recent study by Wijdenes et al. (2019) “evaluated the prevalence and severity of compassion fatigue (CF) risk among nurses employed in a large southwestern hospital system.”

Some health care institutions are using ProQOL results in nurses to guide interventions. For example, a study (Branch & Klinkenburg, 2015) of nurses and allied health professionals at a large academic children’s hospital used version 5 of the ProQOL to “determine the risk for compassion fatigue among 296 direct care providers”. Branch and Klinkenburg (2015) concluded that their results “provided data necessary to support development of a compassion fatigue program for direct care providers”. But those who use the ProQOL in nurses should be aware that the validity of the ProQOL in the nursing population had not been adequately studied.

According to the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green, 2011, section 17.4.1), appropriateness of the items and domains for the population and use of the instrument to determine content-related validity, and use in a population different from the one in which the instrument was validated raises concerns about validity. The Patient Reported Outcomes Measurement Information System (PROMIS) standards (PROMIS[®], 2013, p. 45) state that “psychometric measurement equivalence of scales are basic requirements” for valid assessment of demographic subgroups, and PROMIS recommend that Confirmatory Factor Analysis be used to investigate scale validity.

The ProQOL Manual (Stamm, 2005), says the ProQOL is intended to be “used across many types of professions” (p. 6) and the previous version of the ProQOL instrument has been tested in three broad classes of workers: general health workers, child/family workers and school personnel (p.11). However the table to which the reader is referred (Stamm, 2005, Figure 5, p. 11) reveals that only 100 “general health workers” were studied and there is no mention of any further validation of version 5 of the ProQOL in any type of health care worker.

The efforts of previous researchers using the ProQOL to examine compassion fatigue in nurses are relevant and well-intentioned, however there are concerns about the validity of conclusions drawn from these studies since the psychometric properties have not been thoroughly evaluated in the nursing population. In fact, The Concise ProQOL Manual (Stamm, 2010, p. 19) specifically says, “Two very important covariates that cannot be addressed at this time through data in the data bank are work setting and types of people assisted”. Patients for whom nurses provide care are quite different compared to the clients that social workers council; and the relationship between nurses and patients is much different from the relationship between teachers and students, for example.

Given that the validity of the ProQOL has not been examined in nurses, yet this instrument is commonly used to evaluate nurses, and interventions are being created based on the results obtained, it is important to understand the instrument’s psychometric properties in the nursing population. Thus, the purpose of this study is to assess the reliability and validity of the ProQOL version 5 in nurses and to compare the subscales’ behaviors in nurses to what has been reported in the population used for validation of the tool.

2. Methods

The study was conducted at a large academic quaternary care hospital in the southern United States. The study involves secondary analysis of data used to investigate predictors of nurses considering leaving the profession due to work-related stress (Hagan et al., 2019). The survey that was administered was comprised of the Professional Quality of Life Scale (ProQOL) Version 5 (Stamm, 2010) and demographic questions. Written permission to use the ProQOL was obtained and then, after receiving approval from the Institutional Review Board, the survey was created using SurveyMonkey and distributed via institutional email with a link to the survey. In order to be included in the study, respondents had to report being a nurse who provides direct patient care.

Chronbach’s alpha coefficient was used to evaluate the reliability of the ProQOL subscales. Chronbach’s alpha

coefficients were interpreted according to Kline's (2000, p. 13) guidelines for internal consistency: $\alpha \geq 0.9$ is "Excellent", $0.7 \leq \alpha < 0.9$ is "Good", $0.6 \leq \alpha < 0.7$ is "Acceptable", $0.5 \leq \alpha < 0.6$ is "Poor" and $\alpha < 0.5$ is "Unacceptable".

Confirmatory Factor Analysis was used to examine measurement invariance in order to evaluate how well the ProQOL's three factor model fits the nursing population (Muthen, 2008; PROMIS[®], 2013, p. 45). Confirmatory Factor Analysis was used to test the hypothesis that the PropQOL version 5 three factor structure is valid in the nursing population. The Confirmatory Factor Analysis chi-square test was used to test model fit. Since the chi-square test is known to be sensitive to sample size, other methods were also used to assess model fit. Hoelter's Critical N (Hoelter, 1983) was used to determine the largest sample size for which the hypothesis of a correctly specified factor structure could be accepted ($\alpha = 0.05$). Additionally the Normed Fit Index (NFI) was used to measure model fit, with a value greater than 0.90 considered to be acceptable (Bentler & Bonett, 1980). Factor loadings ≥ 0.45 were considered adequate since this represents the threshold for a "fair" loading (Comrey & Lee, 1992).

For Exploratory Factor Analysis, Kaiser's criterion for eigenvalues (Kaiser, 1960) and the scree test (Cattell, 1966) were used to determine the appropriate number of factors as suggested by DeVellis (2011). As recommended by Ferguson and Cox (1993), principal component analysis was used for factor extraction with the Varimax procedure employed for factor rotation.

"The Concise ProQOL Manual" (Stamm, 2010, pp. 15-18) describes a scoring methodology when the survey is administered in a group setting that standardizes subscale raw scores by the sample mean and standard deviation, and provides cutoffs based on percentiles (<25th percentile, 25th through 75th percentile and >75th percentile) of the standardized t scores in the population used for validation. A different interpretation of subscale scores is provided for situations when an individual takes the ProQOL alone (i.e., not in a group setting), since there is no sample mean and standard deviation available, with raw scores ≥ 42 putting the individual in the "high" category for each of the three subscales (Stamm, 2010, p. 29-30). The proportion of nurses categorized as "high" for each subscale using the interpretations for group versus individual administration was compared using McNemar's test. And the relationship of raw scores with t scores in the validation population provided in the manual (Stamm, 2010, p. 31-34) were compared to the relationships observed in the study's sample of nurses by plotting the two linear regression lines on the same graph, separately for each of the three subscales. The null hypotheses that the median Burnout, Secondary Traumatic Stress and Compassion Satisfaction raw scores in nurses are the same as the corresponding population medians reported by Stamm (2010) were tested using the Wilcoxon signed-rank test.

"The Concise ProQOL Manual" (Stamm, 2010, pp. 19-21) reports associations of gender, age (18-35 years vs. ≥ 36 years), race (white vs. non-white), years at current employer (<5 years, 5-15 years and >15 years) and years in the field (<5 years, 5-15 years and >15 years), with ProQOL subscale t scores based on analysis of "a data bank of 1,289 cases created from multiple studies". These analyses were replicated to examine the associations of demographic categories with ProQOL subscale scores in our sample of nurses using Student's t-test for the dichotomous variables gender, age (18-35 years vs. ≥ 36 years), and race (white vs. non-white) and using ANOVA for years at current employer (<5 years, 5-15 years and >15 years) and years in the field (<5 years, 5-15 years and >15 years).

As mentioned previously, this psychometric evaluation of the ProQOL version 5 is conducted as a secondary analysis of data reported by study Hagan et al. (2019). Therefore, no *a priori* power analysis was performed specifically for this psychometric evaluation. For Exploratory Factor Analysis, the suggested sample size is a minimum ratio of 5 subjects per item up to 300 subjects, with 300 generally considered adequate irrespective of the number of items (Tinsley & Tinsley, 1987). For Confirmatory Factor Analysis, 10 or more respondents per item is preferred for structural equation models (Bentler & Chou, 1986), of which Confirmatory Factor Analysis is a special case. Thus, the sample size of 496 nurses who completed all 30 ProQOL items is more than adequate for examining the factor structure of the ProQOL in nurses. SAS version 9.4 (SAS Institute, Inc., Cary, North Carolina) was used for all data analysis.

3. Results

A total of 864 participants started the survey however only 496 responded to all ProQOL items. These 496 participants' data were analyzed for the current study. The human resources department estimated that there were 3,300 nurses employed at the institution when the survey was administered, yielding an estimated response rate of $496/3300 = 15\%$. The majority of subjects were female (98%) and Caucasian (58%) with an average age of 40.3 years (Table 1).

Table 1. Respondent characteristics

Characteristic	Frequency ¹ (%)
Gender (n=488)	
Female	478 (98)
Male	10 (2)
Ethnicity (n=493)	
African American	50 (10)
Asian	86 (17)
Caucasian	285 (58)
Hispanic / Latino	58 (12)
Other	14 (3)
Education level (n=495)	
Certificate	4 (1)
Diploma	11 (2)
Associates degree	72 (15)
Bachelor's degree	335 (68)
Master's degree	65 (13)
Post-master's degree	4 (1)
Doctorate	4 (1)
Patient care setting (n=494)	
Inpatient	379 (77)
Outpatient	115 (23)
Typical work shift (n=490)	
Days	326 (67)
Nights	150 (31)
Combination	14 (3)
Certified in specialty area (n=491)	
Yes	232 (47)
No	259 (53)

Chronbach's alpha coefficient for the three ProQOL subscales was $\alpha = 0.72$ for Compassion Satisfaction, $\alpha = 0.53$ for Burnout, and $\alpha = 0.77$ for Secondary Traumatic Stress (Table 2). The Confirmatory Factor Analysis chi-square test indicated a highly significant lack of model fit for the three factor structure of the ProQOL ($X^2 = 1068$, $df = 399$, $p < 0.001$). Hoelter's critical $N = 207$ is much lower than the actual sample size of $n = 496$, which again leads us to reject the hypothesis of acceptable model fit ($p < 0.05$). Similarly, the NFI = 0.678 which is substantially lower than the 0.90 threshold considered to be the minimum for an acceptable model.

Table 2. Chronbach's alpha coefficients(α), interpretation of internal consistency and comparison with Chronbach's alpha coefficients reported by Stamm (2010)

ProQOL Subscale	α	Interpretation ²	α reported by Stamm, 2010
Compassion Satisfaction	0.72	Good	0.88
Burnout	0.53	Poor	0.75
Secondary Traumatic Stress	0.77	Good	0.81

For Exploratory Factor Analysis, only two components had eigenvalues greater than 1 and there were two components above the “elbow” of the scree plot (Figure 1), so both Kaiser’s criterion for eigenvalues and the scree test indicated that a two factor structure for the ProQOL is more appropriate in the nursing population than the hypothesized three factor structure. Two (20%) of the Compassion Satisfaction items, 4 (40%) of the Secondary Traumatic Stress items and 9 (90%) of the Burnout items did not adequately load (i.e., factor loading <0.45) onto their respective factors (Table 3).

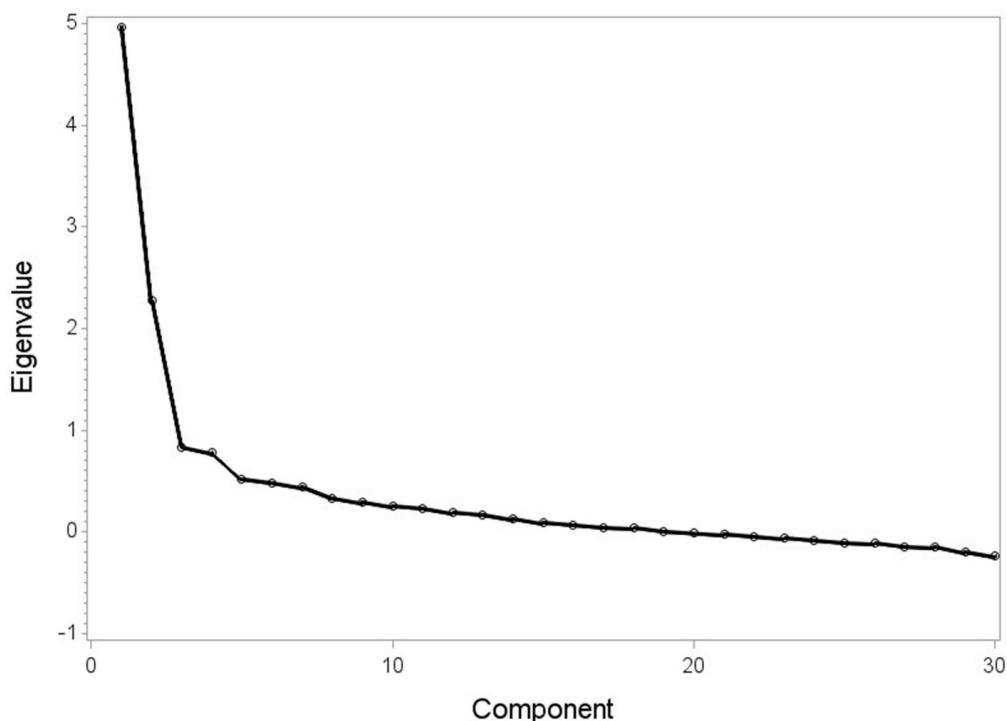


Figure 1. Scree plot of the ProQOL components

Table 3. Factor loadings of the ProQOL items. (*Indicates item does not load onto factor sufficiently, i.e., factor loading <0.45)

ProQOL Subscale	ProQOL Item	Factor Loading
Compassion Satisfaction	I get satisfaction from being able to care for people	0.512
	I feel invigorated after working with those I care for	0.012*
	I like my work as a nurse	0.640
	I am pleased with how I am able to keep up with nursing techniques and protocols	0.271*
	My work makes me feel satisfied	0.429*
	I have happy thoughts and feelings about those I care for and how I could help them	0.493
	I believe I can make a difference through my work	0.477
	I am proud of what I can do to help	0.630
	I have thoughts that I am a "success" as a nurse	0.352
	I am happy that I chose to do this work	0.623

Burnout	I am happy	0.263*
	I feel connected to others	0.302*
	I have beliefs that sustain me	0.119*
	I am the person I always wanted to be	0.167*
	I am a very caring person	0.383*
	I am not as productive at work because I am losing sleep over traumatic experiences of a person I care for	0.582
	I feel trapped by my job as a nurse	0.444*
	I feel worn out because of my work as a nurse	0.101*
	I feel overwhelmed because my work load seems endless	0.117*
	I feel "bogged down" by the system	0.290*
Secondary Traumatic Stress	I am preoccupied with more than one person I care for	0.062*
	I jump or am startled by unexpected sounds	0.190*
	I find it difficult to separate my personal life from my life as a nurse	0.386*
	I think that I might have been affected by the traumatic stress of those I care for	0.643
	Because of my caregiving, I have felt "on edge" about various things	0.523
	I feel depressed because of the traumatic experiences of the people I care for	0.713
	I feel as though I am experiencing the trauma of someone I have cared for	0.735
	I avoid certain activities or situations because they remind me of frightening experiences of the people I care for	0.649
	As a result of my caring, I have intrusive, frightening thoughts	0.290*
I can't recall important parts of my work with trauma victims	0.488	

The relationships between subscale raw scores and t scores in the validation population provided in the manual (Stamm, 2010, p. 31-34) were substantially different from the relationships observed in the sample of 496 nurses, with nurses' regression lines shifted below the population regression lines for Burnout and Secondary Traumatic Stress (Figure 2A and 2B), and the nurses' regression line for Compassion Satisfaction shifted above the validation population regression line (Figure 2C), meaning that the same raw score for nurses resulted in a lower Burnout and Secondary Traumatic Stress t score and a higher Compassion Satisfaction t score compared to the validation population, although the difference was not as large for Compassion Satisfaction.

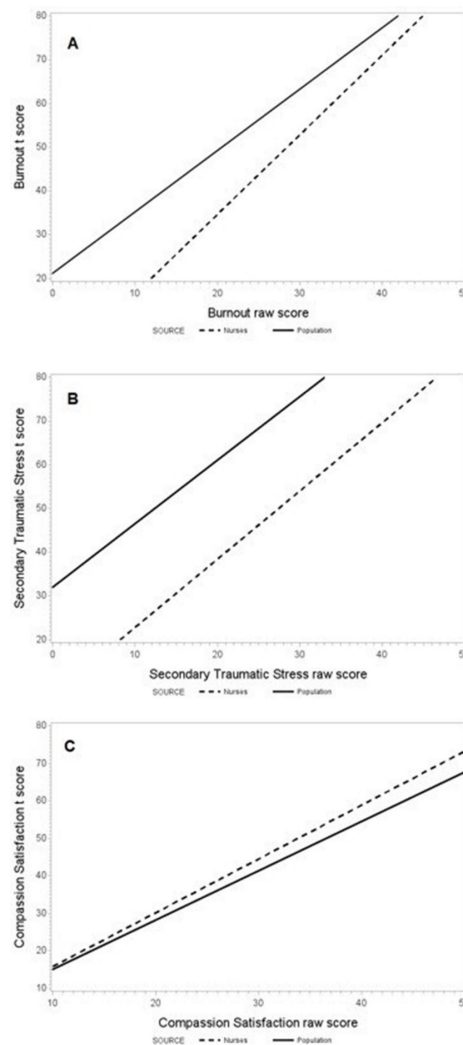


Figure 2. Linear regression lines displaying, separately for the validation population and for nurses, the relationship between t scores and raw scores for A) Burnout B) Secondary Traumatic Stress and C) Compassion Satisfaction

The population median Burnout raw score reported by Stamm (2010) was 20 compared to 28.4 in the sample of nurses ($p < 0.001$). The population median Compassion Satisfaction raw score reported by Stamm (2010) was 37 compared to 33.8 in the sample of nurses ($p < 0.001$). The population median Secondary Traumatic Stress raw score reported by Stamm (2010) was 11 compared to 27.4 in the sample of nurses ($p < 0.001$).

When the criteria for scoring in a group setting based on standardized t scores were applied, 162 (33%) had high levels of burnout compared to none with high burnout based on the individual scoring interpretations ($p < 0.001$), 122 (25%) had high Compassion Satisfaction according to group scoring criteria compared to 76 (15%) using individual scoring cutoffs ($p < 0.001$), and 147 (30%) had high Secondary Traumatic Stress based on group scoring criteria compared to 2 (0.4%) when applying individual scoring cutoffs ($p < 0.001$).

For our sample of nurses, gender was not significantly associated with Compassion Satisfaction ($t = 0.9$, $df = 486$, $p = 0.389$), Burnout ($t = 0.1$, $df = 486$, $p = 0.908$) or Secondary Traumatic Stress ($t = 0.7$, $df = 486$, $p = 0.556$) t scores. White respondents had significantly higher Burnout t scores ($t = 2.5$, $df = 494$, $p = 0.013$), but no significant differences in Compassion Satisfaction ($t = 1.3$, $df = 494$, $p = 0.207$) or Secondary Traumatic Stress ($t = 0.7$, $df = 494$, $p = 0.463$) t scores. Nurses 36 years of age or older had significantly higher Compassion Satisfaction t scores ($t = 2.1$, $df = 485$, $p = 0.036$) compared to nurses less than 36 years of age but Secondary Traumatic Stress ($t = 1.7$, $df = 485$, $p = 0.089$) and Burnout ($t = 0.1$, $df = 485$, $p = 0.902$) t scores were not significantly different between the two age groups. Compassion Satisfaction t scores were significantly higher for nurses with more years of experience

($F=5.2$, $df_1=1$, $df_2=475$, $p=0.024$) but Secondary Traumatic Stress ($F=1.4$, $df_1=1$, $df_2=475$, $p=0.246$) and Burnout ($F=0.8$, $df_1=1$, $df_2=475$, $p=0.376$) t scores did not vary significantly with duration of employment as a nurse. Similarly, Compassion Satisfaction t scores increased significantly with duration of employment at the institution ($F=4.2$, $df_1=1$, $df_2=455$, $p=0.043$) but Secondary Traumatic Stress ($F=1.1$, $df_1=1$, $df_2=455$, $p=0.293$) and Burnout ($F=1.0$, $df_1=1$, $df_2=455$, $p=0.320$) t scores did not.

4. Discussion

Confirmatory Factor Analysis revealed that the three factor structure of the ProQOL has poor validity in nurses. Also the reliability of all three subscales was substantially lower in our sample of nurses compared to what Stamm (2010) reported in the validation population. The Burnout subscale has poor internal consistency in nurses with only one (10%) item adequately loading onto this factor. In concordance with the results of the current study, a study (Sacco, Ciurzynski, Harvey, & Ingersoll, 2015) using the ProQOL in a sample of 221 critical care nurses found much lower reliability of the Burnout ($\alpha=0.45$) and Secondary Traumatic Stress ($\alpha=0.73$) subscales than reported by Stamm (2010). The poor internal consistency of the Burnout subscale is particularly problematic.

With regard to associations of demographic characteristics with ProQOL subscale t scores, in contrast to Stamm (2010), the current study did not find non-white respondents to have significantly higher Secondary Traumatic Stress t scores but did find that nurses 36 years of age or older had significantly higher Compassion Satisfaction and Burnout t scores, and Compassion Satisfaction t scores varied significantly with duration of employment as a nurse and duration of employment at the institution, unlike Stamm (2010).

The nurses in our sample had significantly higher Burnout and Secondary Traumatic Stress median raw scores and significantly lower median Compassion Satisfaction raw scores compared to the validation population, but the absolute magnitude of the subscale scores is lost after conversion to t scores since the t scores are standardized to render interpretations relative to the subscale sample means. The same raw score in our sample of nurses correspond to a much lower Secondary Traumatic Stress t score, a lower Burnout t score and a higher Compassion Satisfaction t score compared to the population used for ProQOL validation. In reference to the reported subscale percentiles, Stamm (2010, p. 31), states that “[c]alculations using the SPSS scoring scheme will have some variance in comparison to the table. The variance is trivial, and only applies if a person is on the on the border of a cut score.” However the observed variances in our sample of nurses were very large.

Additionally, it does not make sense that an individual’s measured degree of Burnout, Secondary Traumatic Stress and Compassion Satisfaction depends so heavily on the group to whom the survey was administered. The 99th percentiles of the Burnout and Secondary Traumatic Stress raw scores reported by Stamm (2010, p 34) are 37 and 31, respectively, both of which are well below the raw score cutoffs of 42 used as the cutoff for “high” Burnout and Secondary Traumatic Stress when assessing ProQOL results from individual administration (Stamm 2010, p 29-10), so there appears to be an error in the cutoffs prescribed for individual scoring.

These issues have led researchers to miscategorize nurses’ Compassion Satisfaction, Burnout and Secondary Traumatic Stress. For example, Young et al. (2011) conducted a study to address the question, “What is the prevalence of Compassion Satisfaction, Burnout, and Secondary Traumatic Stress in heart and vascular nurses?” Based on the results of their study, they report the following:

When examining individual HVIMC nurses scores, 40% of the group had average Compassion Satisfaction scores and 60% scored in the high range. Eighty-four percent of the group had low levels of Burnout and 16% had average Burnout levels. Seventy-six percent had low levels of Secondary Traumatic Stress with 24% experiencing average levels. Review of HVICU nurses’ individual scores showed that approximately 82% had average levels of Compassion Satisfaction and about 18% had high levels. The level of low Burnout was 36% and 64% had average levels. The level of Secondary Traumatic Stress for the low category was 56% and 44% fell in the average category.

Although not specified in the Methods section, since none of the nurse had high Burnout or high Secondary Traumatic Stress, it is obvious that Young et al. (2011) used the interpretations of subscale scores for situations when an individual takes the ProQOL alone, with raw scores ≥ 42 putting the individual in the “high” category for each of the three subscales (Stamm, 2010, p. 29-30). If the t score cutoffs were used (Stamm, 2010, pp. 15-18), approximately 25% of the nurses should have high Burnout and Secondary Traumatic Stress, since these percentiles are standardized relative to the sample mean.

In another study, Branch and Klinkenberg (2015 p 163) administered the ProQOL version 5 to 274 healthcare providers including 179 RNs to categorize participants. They converted raw scores to t scores and categorized

participants with t scores above the population 75th percentile provided by The Concise ProQOL Manual” (2010, page 16) as high Burnout and Secondary Traumatic Stress and participants with t scores below the 25th percentiles as low Compassion Satisfaction. Branch and Klinkenberg (2015) report 25%, 31% and 27% of participants in the high category for Compassion Satisfaction, Burnout and Secondary Traumatic Stress, respectively. As discussed previously, use of the t score cutoffs (Stamm, 2010, pp. 15-18), will put roughly 25% of the sample in the high category for each of the three subscales.

These results beg the question, why does interpretation of a nurse’s score depend so much upon whether the ProQOL was taken alone or as part of a larger group? The guidelines for interpreting individual scores appear to be flawed since it is not possible to have a high level of Burnout or Secondary Traumatic Stress when applying these criteria. And when using the group administration criteria for interpreting t scores, is it desirable to always have appropriately 25% of the sample categorized as having a high level of Burnout, Secondary Traumatic Stress and Compassion Satisfaction? The ProQOL Version 5 Concise Manual (Stamm, 2010) states that “Regardless of the system effects of people who are experiencing high levels of both Burnout and Secondary Traumatic Stress, it is important that interventions be pursued for them as individuals. Supervisors should consider job reassignment during the intervention time.” Although this study of the psychometrics of the ProQOL might seem to be overly critical and even derogatory, it is important that such a widely used tool be valid and reliable for assessing nurses since interventions and job reassignments are being recommended for employees with high levels of Burnout and Secondary Traumatic Stress. But the current study shows that the ProQOL has poor reliability and validity in nurses despite being frequently misused to guide interventions in this population. These results indicate a great need for development of a valid tool for nurses, similar to the Compassion Fatigue Self-Test (CFST) used to specifically measure Secondary Traumatic Stress in psychotherapists – a tool with psychometric properties that have been studied (Jenkins and Baird, 2002). The low response rate is a limitation of the present study, which does create the potential for sample bias, but it seems unlikely that a higher response rate would change the study’s conclusions.

5. Conclusions

Numerous published nursing studies have stated that interventions were being implemented based on ProQOL results. However, the current study has shown that the burnout scale has poor reliability and the instrument as a whole lacks validity in the nursing population. Interpretations of subscale results depend very much on whether the criteria for group versus individual administration are used. ProQOL results should be interpreted with skepticism. A new instrument is needed to measure Burnout, Secondary Traumatic Stress and Compassion Satisfaction in nurses.

Creating a New Tool

Although development of a new tool is beyond the scope of this study, some general guidelines can be given. If Burnout, Secondary Traumatic Stress and Compassion Satisfaction are of interest to measure in nurses, when creating a new tool for this purpose, the theoretical constructs of Burnout, Secondary Traumatic Stress and Compassion Satisfaction should be defined and a list of potential items measuring aspects of these constructs in nurses should be created and subsequently scrutinized in order to winnow it down to a more concise item list. The items should be more specific to the nursing profession than those contained in the ProQOL, so experienced nurses should be the ones to undertake item generation. At this stage of the process the item scales (e.g., Likert scale responses) should be specified. Then item evaluation and refinement should be undertaken. Once the item list has been generated, Exploratory Factor Analysis is used to investigate the new tool’s factor structure. Numerous references can be found (e.g., DeVellis, 2011) that describe these steps in detail.

Confirmatory Factor Analysis to independently validate the factor structure derived from Exploratory Factor Analysis could be accomplished using either of the following approaches:

Use a split sample design (i.e., randomly allocate a proportion of the respondents to a training dataset used to extract the factors and create the subscales with Exploratory Factor Analysis, and allocate the other respondents to a validation dataset for subsequent Confirmatory Factor Analysis).

Obtain a sample of responses from nurses to be used for tool subscale development via Exploratory Factor Analysis, followed by a distinct sample of nurses obtained for Confirmatory Factor Analysis.

There are a multitude of references available that explain Confirmatory Factor Analysis in detail (e.g., Brown, 2015).

An important consideration for the researcher attempting to undertake this task is the appropriate sample size for Exploratory Factor Analysis and Confirmatory Factor Analysis. As previously mentioned, at least 5

respondents per item up to a total of 300 respondents are recommended for Exploratory Factor Analysis (Tinsley & Tinsley, 1987) and at least 10 respondents per item is recommended for Confirmatory Factor Analysis (Bentler & Chou, 1987). Obtaining nurse-level data for some real-world outcomes (e.g., nurse termination/resignation) in order to assess aspects of construct validity would also be informative. Practical uses of the new tool such as describing the population of nurses in terms of these constructs, investigating associations of nurse characteristics with these constructs and assessing individual nurses or groups of nurses for potential intervention should be considered from the beginning when the tool is being developed. The intended application of the new tool will guide its development and evaluation.

Acknowledgements: The author is immensely grateful to Dr. Marlene Walden, Dr. Barbara McFadden, Pamela Krinock, Aaron Loeb, Lindsay Stephens and Darlene Acorda for their invaluable participation in the original study.

Funding sources: No funding was received for this research.

Conflict of interest: The author has no financial interest or benefit from this research and has no conflict of interest to disclose.

References

- Bentler, P. M., & Chou, C. H. (1987). Practical issues in structural equation modeling. *Sociological Methods & Research*, 16, 78-117. <https://doi.org/10.1177/0049124187016001004>
- Bentler, P., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-600. <https://doi.org/10.1037/0033-2909.88.3.588>
- Branch, C., & Klinkenburg, D. (2015). Compassion fatigue among pediatric healthcare providers. *MCN, The American Journal of Maternal / Child Nursing*, 40(3), 160-6. <https://doi.org/10.1097/NMC.0000000000000133>
- Bride, B. E., Radey, M., & Figley, C. R. (2007). Measuring compassion fatigue. *Clinical Social Work Journal*, 35, 155-163.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 629-637. <https://doi.org/10.1007/s10615-007-0091-7>
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.
- DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: SAGE.
- Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A user's guide. *International Journal of Selection and Assessment*, 1(2), 84-94. <https://doi.org/10.1111/j.1468-2389.1993.tb00092.x>
- Figley, C. (1995). *Compassion Fatigue: Coping with Secondary Traumatic Stress Disorders in Those Who Treat the Traumatized*. London: Brunner-Routledge.
- Figley, C. R., & Stamm, B. H. (1996). Psychometric Review of Compassion Fatigue Self Test. In B. H. Stamm (Ed.), *Measurement of Stress, Trauma and Adaptation*. Baltimore: Sidran Press.
- Hagan, J., Tyer-Viola, L., & Graves, K. (2019). Predictors of Nurses Considering Leaving the Profession Due to Work-Related Stress in a Large Pediatric and Women's Hospital in the United States. *Journal of Hospital Administration*. 8(1), 27-33. <https://doi.org/10.5430/jha.v8n1p27>
- Higgins, J., & Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions, version 5.1.0, The Cochrane Collaboration®. Retrieved June 19, 2016, from <http://handbook.cochrane.org/>
- Hoelter, J. W. (1983). The Analysis of Covariance Structures: Goodness-of-Fit Indices. *Sociological Methods and Research*, 11, 325-344. <https://doi.org/10.1177/0049124183011003003>
- Jenkins, S. R., & Baird, S. A. (2002). Secondary traumatic stress and vicarious trauma: A validation study. *Journal of Traumatic Stress*, 15(5), 423-432. <https://doi.org/10.1023/A:1020193526843>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151. <https://doi.org/10.1177/001316446002000116>
- Kline, P. (2000). *The Handbook of Psychological Testing* (2nd ed.). London: Routledge.
- Muthen, B. O. (2008). Factor Analysis. UCLA: Statistical Consulting Group. Retrieved June 21, 2016, from http://www.ats.ucla.edu/stat/seminars/muthen_08/part2.pdf
- PROMIS®. (2013). PROMIS® Instrument Development and Validation Scientific Standards Version 2.0.

- Retrieved June 19, 2016, from http://www.nihpromis.org/Documents/PROMISStandards_Vers2.0_Final.pdf
- Sacco, T. L., Ciurzynski, S. M., Harvey, M. E., & Ingersoll, G. L. (2015). Compassion satisfaction and compassion fatigue among critical care nurses. *Critical Care Nursing*, 35(4), 32-42. <https://doi.org/10.4037/ccn2015392>
- Stamm, B. H. (2005). The ProQOL manual: The Professional Quality of Life Scale. Retrieved May 14, 2017, from <http://www.compassionfatigue.org/pages/ProQOLManualOct05.pdf>
- Stamm, B. H. (2010). The Concise ProQOL Manual, 2nd ed. Pocatello, ID. Retrieved May 16, 2017, from http://www.proqol.org/uploads/ProQOL_Concise_2ndEd_12-2010.pdf
- Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, 34, 414-424. <https://doi.org/10.1037/0022-0167.34.4.414>
- Watts, J., & Robertson, N. (2015). Selecting a measure for assessing secondary trauma in nurses. *Nurse Researcher*, 23(2), 30-35.
- Wijdenes, K. L., Badger, T., & Sheppard, K. (2019). Assessing compassion fatigue risk among nurses in a large urban trauma center. *The Journal of Nursing Administration*, 49(1), 19-23. <https://doi.org/10.1097/NNA.0000000000000702>
- Young, J. L., Derr, D. M., Cicchillo, V. J., & Bressler, S. (2011). Compassion satisfaction, burnout, and secondary traumatic stress in heart and vascular nurses. *Critical Care Nursing Quarterly*, 34(3), 227-234. <https://doi.org/10.1097/CNQ.0b013e31821c67d5>

Notes

Note 1. Sample sizes are not the same for each characteristic because a different number of responses were missing.

Note 2. From Kline (2000).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).