Construct Validity of the Professional Quality of Life (ProQoL) Scale in a Sample of Child Protection Workers

Steve Geoffrion1,2, Josianne Lamothe1,2,3, Julien Morizot,1 and Charles-Édouard Giguère2

1School of Psychoeducation, Université de Montréal, Montreal, Quebec, Canada
2Trauma Studies Center, Institut Universitaire en Santé Mentale de Montréal, Montreal, Quebec, Canada
3School of Criminology, Université de Montréal, Montréal, Quebec, Canada

The Professional Quality of Life (ProQOL) scale is one of the most widely used measures of compassion satisfaction and fatigue despite there being little publicly available evidence to support its validity. This study, conducted among a sample of 310 child protection workers, assessed the construct validity of this measure using confirmatory factor analysis (CFA) and bifactor modeling. The CFA failed to confirm the adequacy of the three-factor structure proposed by Stamm (2010). In response, a bifactor model postulating a factor structure with a general factor in addition to independent factors (compassion satisfaction, job burnout, and secondary traumatic stress) was proposed, highlighting the unidimensionality of the ProQOL while allowing for each subscale to be used separately. Moreover, this bifactor model of the ProQOL was moderately correlated with the Posttraumatic Disorder Checklist, r = −.427, p < .001, and strongly correlated with scales of well-being at work, r = .694, p < .001, and psychological distress at work, r = −.666, p < .001, thus supporting the ProQOL’s convergent validity. No associations were found between the ProQOL and the Life Event Checklist, which supports the ProQOL’s discriminant validity. Overall, the results indicated that compassion satisfaction and compassion fatigue represent higher and lower levels of the same construct rather than two different constructs. Researchers and clinicians could therefore compute a single score to rate professionals’ individual levels of professional quality of life.

The concept of compassion fatigue (CF) is currently one of the dominant theoretical frameworks in work-related stress studies, which examine the consequences of caring for others. Compassion fatigue refers to the cumulative psychological toll associated with working with survivors of trauma or perpetrators of violence and crime as part of everyday work (Osofsky, Putnam, & Lederman, 2008); it is sometimes thought of as the inevitable cost of caregiving (Figley, 1999). Manifesting itself through symptoms similar to those found in posttraumatic stress disorder (PTSD) and burnout (BO; Bride, Radey, & Figley, 2007; Stamm, 2005), CF has been shown to affect workers’ cognition, emotions, behaviors, and even physiology (Berzoff & Kita, 2010). In addition, clinicians experiencing CF have reported symptoms akin to those experienced by their traumatized clients (Bride et al., 2007). Given these known consequences, it is important to develop tools capable of accurately assessing individuals in need of support.

Stamm’s (2010) Professional Quality of Life (ProQOL) instrument is one of the most popular tools available to measure CF. According to Stamm, over 200 published studies have used the ProQOL to assess CF since the measure’s inception. Literature reviews on the subject confirm the popularity of the ProQOL as a research instrument (Cieslak et al., 2014; Watts & Robertson, 2015). Yet despite the scale’s popularity, the ProQOL’s psychometric properties remain nebulous (Bride et al., 2007; Hemsworth, Baregheh, Aoun, & Kazanjian, 2018), which is problematic for any researcher who wishes to study the concept further (Watts & Robertson, 2015). More specifically, there is a need to clearly demonstrate that the ProQOL has adequate construct validity (Hemsworth et al., 2018; Watts & Robertson, 2015). Construct validity has been defined as a unifying form of validity requiring the integration of different complementary sources of information (Simms & Watson, 2007). This article

This research was supported by scholarships awarded to the first author by the Fonds de la Recherche du Québec - Société et culture (FRQSC), the Institut de recherche Robert-Sauvé en santé et sécurité au travail (IRSST), the International Center for Comparative Criminology (ICCC) and the foundation of Institut Universitaire en santé mentale de Montréal as well as by a research grant awarded to the VISAGE research team by the Canadian Institutes of Health Research (CIHR: TVG251591).

Correspondence concerning this article should be addressed to Steve Geoffrion, School of Psychoeducation, Université de Montréal, C.P. 6128 succursale Centre-Ville, Montréal, Québec, Canada, H3C 3J7. E-mail: s.geoffrion@umontreal.ca.

© 2019 The Authors. Journal of Traumatic Stress published by Wiley Periodicals, Inc. on behalf of International Society for Traumatic Stress Studies. View this article online at wileyonlinelibrary.com
DOI: 10.1002/jts.22410

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
details the results of an evaluation of the ProQOL’s factor validity, internal consistency, and reliability, as well as convergent validity and discriminant validity.

The ProQOL framework as proposed by Stamm (2010) aims to explain the consequences of working as a “helper” to traumatized individuals. The term “helper” refers to psychologists, trauma therapists, social workers, nurses, disaster relief workers, and other professionals who are mandated to intervene with traumatized victims. Two broad overarching concepts in this framework are used to explain the work realities of helpers: CF and compassion satisfaction (CS). According to this framework, CF is composed of two subsets of symptoms: job BO and secondary traumatic stress (STS). Stamm (2010) defines BO as lingering feelings of hopelessness and fatigue that interfere with the ability to perform effectively at work. Symptoms of BO may include feelings of being trapped, overwhelmed, “bogged down,” and unsatisfied by one’s job. Workers who experience STS are described as being “preoccupied with thoughts of people one has helped. Caregivers report feeling trapped, on edge, exhausted, overwhelmed and ‘infected’ by others’ trauma” (Stamm, 2010; p. 21). Symptoms of STS may include fear, sleep difficulties, intrusive images, and avoiding traumatic narratives. Stamm insists that STS and BO are two distinct but related concepts: “[they] both measure negative affect but are clearly different; the BO scale does not address fear while the STS does” (pp. 13–14).

The ProQOL framework, however, is not solely concerned with the negative consequences of helping others—it also seeks to capture the positive outcomes of care. Indeed, research has shown that not all caregivers inevitably experience CF (Søndena, Lauvrudd, Sandvik, Nonstad, & Whittington, 2013). For instance, using a representative sample of 532 self-identified trauma therapists, Sprang and Craig (2015) found that overall, only 5.0% of their respondents were deemed to be at a high risk for CF. A possible explanation for these findings is that, as Stamm (2002) suggests, workers may develop CS instead of CF, for which CS refers to the pleasure one derives from helping others (Stamm, 2010). Indeed, even under stressful conditions, working with this clientele can be highly rewarding (Voss Horrell, Holohan, Didion, & Vance, 2011). For example, clinicians with higher levels of CS are also believed to be more resilient (Burnett & Wahl, 2015). In addition, CS has been shown to be negatively correlated with both STS and BO (Burnett & Wahl, 2015; Stamm, 2010).

The ProQOL scale is similar to other instruments currently available for measuring CF (Bride et al., 2007; Watts & Robertson, 2015). The first instrument ever developed to measure CF was the Compassion Fatigue Self-Test (Figley, 1999). This questionnaire was originally created by Figley (1999) based on his clinical knowledge and was designed to assess only CF (which he defined as PTSD developed due to indirect exposure) and BO. Initial investigations seemed to support the reliability and validity of the instrument (Figley, 1999). Further adjustments to this measure and the inclusion of a third subscale, which measured CS, led to the creation of a new instrument under a different name: Professional Quality of Life scale (ProQOL). This latest measure was updated several times and is now in its fifth revision (ProQOL-5).

Although Stamm (2010) reported some indicators of discriminant validity in her manual, there is currently very little publicly available data to support the validity of the ProQOL-5 (Bride et al., 2007; Hemsworth et al., 2018). Stamm (2010) reported small shared variance between CS and STS (r = .14; co-X = 2.0%; n = 1,187) and between CS and BO (r = −.14; co-X = 2.0%; n = 1,187) as well as a significant shared variance between BO and STS (r = .58; co-X = 34.0%; n = 1,187). In contrast, in their meta-analysis of the association between ProQOL’s BO and STS subscales, Cieslak and colleagues (2014) concluded that studies that used the ProQOL-4 or ProQOL-related measures found significantly stronger associations between the two constructs (BO and STS) than studies that used different instruments (R² = .55 vs. R² = .34). They concluded that the estimated overlap of 55.0% suggests the two concepts are likely indistinguishable under the CF framework. Furthermore, Stamm (2010) reported good internal consistency for the subscales, with Cronbach’s alpha values of .88 for CS, .75 for BO, and .81 for STS. Cieslak and colleagues (2014) reported alpha values ranging from .65 to .83 for BO and from .68 to .87 for STS in their meta-analysis. Alpha values for CS also appear to vary greatly from study to study. For example, some authors have reported an alpha value of .72 (Musa & Hamid, 2008), whereas others have reported alpha values of .83 (Søndena et al., 2013) and .90 (Smart et al., 2014). With regards to its internal structure, Sprang and Craig (2015) arrived at a final three-factor solution for the CF subscale—one for STS and two for BO—using an exploratory factor analysis with the ProQOL-4. They found the inverted items of the BO subscale were especially problematic (e.g., “I am happy”) as they did not load onto any of the other subscales. Unlike Cieslak and colleagues (2014), Sprang and Craig’s study provided support for a distinction between trauma-related symptoms and BO-related symptoms, but they nonetheless concluded, “the ProQOL-4 BO scale may need to be reevaluated for its validity” (Sprang & Craig, 2015, p.9). Hemsworth and colleagues (2018) also found these items problematic and suggested a complete revision of this subscale. Overall, it seems that the intended three-dimensional structure of the instrument cannot be easily reproduced. Recently, Stamm (2019) proposed that a composite score (i.e., combining all subscale scores into one general score) could theoretically be possible and clinically relevant. To the best of our knowledge, no researchers have attempted to do this thus far. In summary, although the scale has been widely used, there is currently not enough publicly available evidence to validate its use (Hemsworth et al., 2018).

There are relatively few independent investigations of the ProQOL’s validity (Heritage, Reese, & Hegney, 2018; Hemsworth et al., 2018). The few studies that used factor analyses were conducted in hopes of validating the ProQOL-4 (Lago & Codo, 2013; Musa & Hamid, 2008; Palestini, Prati, Pietrantoniti, & Cicognani, 2009; Shen, Yu, Zhang, &
Jiang, 2015; Sprang & Craig, 2015) and ProQOL 5 (Dang et al., 2015; Galiana, Arena, Oliver, Sanso, & Benito, 2017; Hemsworth et al., 2018; Samson, Lecovich, & Shvartzman, 2016) in languages other than English. For example, using exploratory factor analyses, Musa and Hamid (2008) found only two dimensions of the ProQOL-4, with one assessing CS and the other CF (i.e., a combination BO and STS), but several items had to be discarded. In order to arrive at a satisfactory final solution, most authors have had to make several changes to the instrument. Some of the changes to the structure have been relatively minor and made in hopes of remaining culturally relevant or to preserving internal consistency (Dang et al., 2015; Shen et al., 2015; Sprang & Craig, 2015). Some other changes, however, have been more problematic, such as switching items between the BO and STS dimensions to retain a three-dimensional structure (Lago & Codo, 2013; Palestini et al., 2009). The most recent analysis on the subject concluded that the CS scale had sound psychometric properties but that nine out of 20 items had to be dropped from the CF scale to remain empirically valid (Heritage et al., 2018).

In all, although previous studies have found the ProQOL-5 to be of interest for examining CF and CS, some of the scale’s psychometric properties remain controversial, particularly its factor structure (Hemsworth et al., 2018). Specifically, whether the covariances among the 30 items are best explained by a single factor or by two or three correlated factors remains unclear. Moreover, relevant alternative factor models have never been tested, particularly a bifactor model postulating a general professional quality of life factor (i.e., a continuum score). Given the current popularity of the ProQOL-5, it is imperative to clarify doubts regarding the scale’s validity, particularly by comparing the adequacy of alternative factor structures.

The issues raised herein highlight important questions that have been thus far neglected by studies using the ProQOL-5 without prior validation. The aim of this study was to assess the construct validity of the ProQOL-5. First, the factor structure of the ProQOL-5 was assessed with confirmatory factor analyses (CFAs). Different alternative conceptually relevant models were tested (i.e., CFAs with one, two, and three factors; a second-order factor model; two bifactor models with two and three group factors). Second, the internal consistency and reliability were evaluated using omega coefficients, explained common variance, and percentage of uncontaminated correlations (Rodriguez, Reise, & Haviland, 2016). Third, given the resemblance of STS symptoms to PTSD symptoms, convergent validity was assessed using a validated measure of PTSD. Convergent validity was also examined with a validated measure of well-being and distress at work, as the ProQOL reflects positive and negative outcomes of caring for others as a profession (Stamm, 2010). Fourth, discriminant validity was assessed using a validated instrument of non-work-related traumatic events using a different time frame (e.g., childhood traumatic events).

Method

Participants

The sample included 310 child protection workers. All participants were French speaking and worked at the Montérégie’s Youth Center located south of Montreal, Canada. Of these respondents, 46.5% were educators working in residential settings, 10.0% were educators working in the community with families, and 43.5% were human relation agents tasked with evaluation and follow-up services in the community. Of the 310 participants, 84.5% were women (n = 262) and 15.5% were men (n = 38). The mean age of participants was 35.75 years (SD = 9.83), 67.3% of participants worked full time, and on average, altogether, the sample reported a mean of 9.84 (SD = 7.84) years of experience in youth protection services.

Procedure

The current study was part of a broader research project aimed at assessing the impact of workplace aggression on child protection workers’ adjustment. A randomized sample stratified by sex was recruited to achieve the purpose of this study. Participants who had not been in contact with the clientele in the year prior to survey completion and those on maternity leave were excluded from the recruiting process. Researchers were provided with a complete list of all child protection workers employed by the agency; they then sent invitations by email to those individuals who had been randomly selected to participate. The invitation email contained information on the study, the guarantee of anonymity, and a confirmation of the Montérégie’s Youth Center’s intention to have participants fill out the survey during working hours. Participants were asked to return their consent forms in order to receive a personal identification number that would give them access to the online questionnaire. Participants answered on a voluntary basis. Data collection spanned from November 2013 to July 2014. The valid response rate was 40.9%. Post-hoc comparisons were made between respondents and nonrespondents based on the institution’s human resources data with regards to age, sex, work environment (residential vs. field work), and years of experience; no statistically significant differences were found. Ethics approval was obtained from a local university and child protection agency (Centre Jeunesse de Québec–Institut Universitaire et Université de Montréal, #CJQ-IU-2013-10).

Measures

Professional quality of life. The ProQOL-5 (Stamm, 2009) is a self-report instrument comprising three different 10-item subscales (BO, STS, and CS). Each item assesses how frequently in the last 30 days a respondent has experienced symptoms. Items are rated on a 5-point scale (1 = never to 5 = very often). We used the French version of the ProQOL-5, which was translated and promoted by Stamm (2009) and has been used previously (Thomas, Billon, Chaumier, Barruche, &
PTSD. The French version of the Posttraumatic Stress Disorder Checklist (PCL-S; Ventureyra et al., 2001) is a 17-item self-report measure based on criteria from the fourth edition (text revision) of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) (Weathers et al., 1993). Respondents rate their symptoms in the past month on a Likert-scale ranging from 1 (not at all) to 5 (extremely). The PCL-S possesses good diagnosis validity as well as good psychometric properties (Blanchard, Jones-Alexander, Buckley, & Forneris, 1996; Weathers et al., 1993). The instrument presented good internal consistency within the present study, Cronbach’s $\alpha = .90$ (based on George & Mallery, 2003) and was used for convergent validity testing.

Well-being and distress at work. The scale of Well-Being at Work (WBW) and Psychological Distress at Work (PDW; Gilbert, Dagenais-Desmarais, & Savoie, 2011) is a two-dimensional, culturally relevant, validated questionnaire designed to measure positive and negative outcomes at work. The WBW scale contains 25 items and the PDW scale contains 23. Responses are based on a 5-point Likert scale ranging from 1 (never) to 5 (almost always). Each subscale has three dimensions. The WBW comprises Peace of Mind (Cronbach’s $\alpha = .86$), Affective Engagement at Work (Cronbach’s $\alpha = .84$), and Social Harmony (Cronbach’s $\alpha = .82$; Gilbert et al., 2011). In contrast, the PDW combines Anxious and Depressive Mood (Cronbach’s $\alpha = .85$), Hostility (Cronbach’s $\alpha = .91$), and Emotional Disengagement from Work (Cronbach’s $\alpha = .88$; Gilbert et al., 2011). In the present study, the WBW composite score demonstrated excellent internal consistency, Cronbach’s $\alpha = .94$, whereas internal consistency was good for the Peace of Mind and Affective Engagement subscales, Cronbach’s $\alpha = .89$ and .82, respectively, and acceptable for the Social Harmony subscale, Cronbach’s $\alpha = .79$. The PDW composite scale demonstrated excellent internal consistency in the current sample, Cronbach’s $\alpha = .95$, whereas internal consistency was acceptable for the Hostility subscale, Cronbach’s $\alpha = .78$; excellent for the Anxious and Depressive Mood subscale, Cronbach’s $\alpha = .90$; and good for the Emotional Disengagement subscale, Cronbach’s $\alpha = .89$.

Lifetime traumatic event exposure. The French version of the Life Events Checklist (LEC) for DSM-IV (Jehel, Brunet, Patermiti, & Guelfi, 2005) is a 16-item self-report measure that assesses lifetime exposure to potentially traumatic events. Respondents are given a list of event types and choose from the following responses for each: happened to me, witnessed the event, happened to someone close, not sure, and does not apply. The LEC has demonstrated adequate psychometric properties as a measurement instrument of exposure to traumatic events as well as convergent validity with measures that assess varying levels of exposure to potentially traumatic events and psychopathology related to traumatic exposure (Gray, Litz, Hsu, & Lombardo, 2004). For the purpose of the main research project described in this study, researchers asked participants to limit their answers to non-work-related events (work-related events were assessed using a more precise scale). On average, participants reported having personally experienced 1.43 ($SD = 1.52$) potentially traumatic events, witnessing 1.31 ($SD = 1.66$), and hearing about 2.07 ($SD = 2.34$) during their lifetimes.

Data Analysis

Factor structure, internal consistency, and reliability. Four CFAs (i.e., each item loads in a single factor, with no cross loading allowed) were tested using Mplus (Version 8; Muthén & Muthén, 2017). Out of 310 respondents, six did not complete the French version of the ProQOL-5, an additional four participants skipped two items, and 33 individuals skipped only one item; thus, 267 respondents had no missing data. Missing data were handled with full information maximum likelihood, in which missing values are not replaced or imputed but rather handled within the analysis model. We tested CFA models with one, two, and three factors, allowing correlations between factors. Moreover, a three-factor model with a second-order factor of CF, which combined BO and STS (reflecting Stamm’s model [2010]), was also used. Bifactor modeling was also used to assess the multidimensionality of the ProQOL-5. Bifactor analysis has been shown to complement traditional dimensionality analyses by allowing researchers to examine the utility of forming subscales and by providing an alternative to nonhierarchical multidimensional models for scaling individual differences (Reise, Morizot, & Hays, 2007). A robust weight least square estimator (WLSMV) was used based on the polychoric correlation matrices. Specific factors of the bifactor models were restricted to be uncorrelated. Several fit indices were used to validate the adequacy of the model and data, as recommended by West, Taylor, and Wu (2012). The indices correspond to the standards usually required for this type of model, meaning Tucker–Lewis Index and comparative fit index (CFI) values greater than .950 (Hu and Bentler, 1999), a 90% confidence interval upper limit of less than .080 for the root mean square error of approximation (RMSEA), and a confidence interval of 90.0% (Steiger, 2007). The internal consistency of the final model was tested with omega reliability coefficients (omega, omegaS, omegaH, omegaHS; McDonald, 1999; Reise, Bonifay, & Haviland, 2013; Revelle & Zinbarg, 2009). Explained common variance (Reise, Scheines, Widaman, & Haviland, 2013) and percentage of uncontaminated correlations (Bonifay, Reise, Scheines, & Meijer, 2015; Reise et al., 2013) were also used to judge the dimensionality of the bifactor CFA model. According to Gorsuch (1983), the sample size ($n = 310$) was adequate for performing factorial analyses given his rule of 10 participants per item.
Convergent and discriminant validity. Pearson correlations were computed using R (Version 3.5) with the composite score of the PCL-S, the composite score for each subscale of the WBW, and the composite score of the PDW in order to test the convergent validity of the retained solution of the ProQOL. Scores from the PCL-S were only available for participants who had endorsed experiencing at least one event on the LEC (n = 71). Pearson correlations were computed with composite scores of each subscale of the LEC to assess discriminant validity. These analyses were also performed using the latent variables of the ProQOL-5 and the external variables.

Results

Descriptive Analysis of the ProQOL Items

Each subscale of the ProQOL-5 was computed according to Stamm’s (2010) scoring method. As such, the bottom (25th) and top (75th) quartiles of the present sample reflect the quartiles obtained by Stamm (2010), based on data from multiple studies, throughout the development of the instrument. For CS, Stamm’s bottom and top thresholds are 44 and 57, respectively; the thresholds for the present sample were similar, at 44 and 56. For BO, Stamm suggested scores of 43 and 56 for the bottom and top thresholds; we obtained scores of 44 and 58. Finally, for STS, our bottom and top quartiles were the same as Stamm at 42 and 56 (for scales distributions see the Supplementary Tables).

Factor Structure

Table 1 shows the goodness-of-fit indices for the different factor models obtained. The fit indices of the CFA one-factor, two-factor, three-factor, and second-order factor (i.e., with a combined CF factor) models showed an unsatisfactory fit to the data (e.g., Models 1–4). The best fit to the data was obtained using the bifactor CFA (BCFA) solutions. The BCFA with three group factors had a satisfactory fit to the data, as did the BCFA with two factors. As shown in Table 1, the fit indices of these two models (Models 5 and 6) were essentially the same. Therefore, these two solutions were retained for further analyses.

For both BCFA models, the majority of items loaded better on the general factor than on the CS factor (k = 8 for the three-factor [BCFA-3] and two-factor [BCFA-2] models; see Table 2). All of the CS items on the general factor had loadings that were statistically significant and higher than 0.30. In the two models, Items Q02 and Q05 clearly did not load onto the general factor but did load significantly onto the group factor. For the BCFA-3, four items of the BO factor loaded more strongly onto the general factor, whereas six items loaded more strongly onto the BO factor. Only Item Q08 had a factor loading on the general factor less than 0.30. For the BCFA-3 (i.e., Model 6), all of the STS items loaded more strongly onto the STS factor than onto the general factor. Only Items Q11 and Q25 had loadings higher than 0.30 on the general factor. For the BCFA-2 (i.e., Model 5), four items of the CF factor loaded more strongly onto the general factor, whereas 16 loaded more strongly onto the CF factor. Of the CF items, nine had loading factors less than 0.30 onto the general factor. Figure 1 illustrates all models tested.

Internal Consistency and Reliability

Table 3 illustrates the internal consistency of the two BCFA solutions. Omega values, which reflect true score variance and are the latent variable analog to coefficient alpha (Rodriguez et al., 2016), were high for all factors. Thus, general factors and group factors for both models had high internal consistency. The high value of OmegaH for the general factor and its low value for group factors in both bifactor models suggested a unidimensional structure, with group factors reliably reacting stronger to the general factor than with their own group factors (Rodriguez et al., 2016). The higher coefficient of the STS suggests that the specific factor tends to provide valid individual differences over and above that of the general factor. Explained common variance of both models was of .50, indicating that the general factor accounted for half the common variance. Similarly, the percentage of uncontaminated correlation was around .45 for both models, suggesting that almost half of the correlation matrices reflected the general factor. Given the factor structure

Table 1

| Model | $\chi^2$ | df | CFI | TLI | RMSEA | 90% CI     |
|-------|--------|----|----|-----|-------|------------|
| Model 1: CFA 1 factor | 2,993 | 405 | .727 | .707 | .145 | [.140, .150] |
| Model 2: CFA 2 factors | 2,157 | 404 | .815 | .801 | .119 | [.115, .124] |
| Model 3: CFA 3 factors | 1,760 | 402 | .857 | .845 | .105 | [.100, .110] |
| Model 4: CFA 3 factors + 1 second-order factor (Compassion Fatigue)* | 2,063 | 403 | .825 | .811 | .116 | [.111, .121] |
| Model 5: BCFA 2 factors | 790 | 375 | .954 | .946 | .060 | [.054, .066] |
| Model 6: BCFA 3 factors | 785 | 374 | .954 | .947 | .060 | [.054, .066] |

Note. CFA = confirmatory factor analysis; BCFA = bifactor confirmatory analysis; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation.

*aThe second order is Compassion Fatigue, which comprises Burnout and Secondary Traumatic Stress as depicted by Stamm’s (2010) model.
Table 2
Standardized Factor Loadings for the Bifactor Confirmatory Factor Analysis (BCFA)-3 BCFA-2 Models of the Professional Quality of Life Scale (ProQOL)

| Item | BCFA 3-Factor | | BCFA 2-Factor | |
|------|---------------|------------------|------------------|
|      | G    | CS   | BO    | STS   | G    | CS   | BO    | STS   |
| Q03  | .559*** | .426*** |       |       | .559*** | .425*** |       |       |
| Q06  | .592*** | .301*** | .426*** | .112  | .593*** | .301*** | .860*** | .181** |
| Q12  | .860*** | .181**  |       |       | .860*** | .181**  | .476*** | .112  |
| Q16  | .476*** | .112   |       |       | .476*** | .112   | .476*** | .112  |
| Q18  | .875*** | .084   |       |       | .895*** | .084   |       |       |
| Q20  | .510*** | .562*** |       |       | .511*** | .562*** |       |       |
| Q22  | .546*** | .633*** |       |       | .547*** | .633*** |       |       |
| Q24  | .675*** | .595*** |       |       | .675*** | .595*** |       |       |
| Q27  | .773*** | .297*** |       |       | .773*** | .297*** |       |       |
| Q30  | .947*** | .058   |       |       | .947*** | .057   |       |       |
| Q01  | .646*** | .164*** |       |       | .645*** | .163*** |       |       |
| Q04  | .372*** | .396*** | .751*** | .102  | .371*** | .384*** | .194**  | .726*** |
| Q08 R|.191**  | .708*** | .477*** | .107  | .194**  | .688**  | .477*** | .688**  |
| Q10 R|.474*** | .101   | .488*** |       | .477*** | .688**  |       |       |
| Q15  | .489*** | -.101  | .488*** |       | .488*** | .688**  |       |       |
| Q17  | .663*** | .102   | .663*** |       |       | .102   |       |       |
| Q19 R|.445*** | .615*** | .456*** |       | .456*** | .606*** |       |       |
| Q21 R|.377*** | .583*** | .379*** |       | .379*** | .569*** |       |       |
| Q26 R|.416*** | .510*** | .417*** |       | .417*** | .497*** |       |       |
| Q29  | .521*** | -.290*** | .520*** |       | .520*** | -.281*** |       |       |

Note. \( n = 304 \). “R” denotes items that need to be inverted in order to represent positive QOL items. G = general factor (ProQOL); CS = Compassion Satisfaction; BO = Burnout; STS = Secondary Traumatic Stress; CF = Compassion Fatigue.

and internal consistency and reliability results, the BCFA-3 (i.e., three group factors) was retained as the final model for subsequent analysis. Furthermore, this solution was retained as it approximated Stamm’s (2009) original three-dimensional professional quality-of-life framework. As such, the general factor reflecting the ProQOL scale was computed and ranged from 30 to 150 (the Supplementary Materials detail the SPSS syntax of this scale). A low score represented a poor professional quality of life at work, and vice versa. Three subscales were also retained: (a) CS (score range: 10–50), on which a high score represents a high level of satisfaction; (b) BO (score range: 10–50), on which a high score represents a high level of BO symptoms; and (c) STS (score range: 10–50), on which high score represents a high level of STS.

Convergent Validity

Based on Cohen’s (1992) effect size thresholds, the correlations between the ProQOL’s general factor composite score and the PCL-S composite scores were moderate, \( r = -.427, p < .001 \). A High score on the PCL-S was associated with a lower score on the general factor. The correlations between the CS scale and the PCL-S were weak and negative. Correlations between BO and PCL-S were moderate and positive as were the correlations between the STS and PCL-S (see Table 4). The correlations between the ProQOL’s general factor and WBW and PDW composite scores were large, \( r = .694, p < .001 \) for the WBW and \( r = -.666, p < .001 \) for the PDW. As such, the ProQOL’s general factor was positively associated with the WBW and negatively associated with the PDW.
Correlations between the ProQOL’s general factor and subscales of the WBW were positive and ranged between medium and very large, whereas correlations between the ProQOL’s general factor and PDW were negative and ranged between medium and large. When composite scores were used, there was a large, positive association between the CS scale and the WBW and its three subscales; a negative, moderate association between the CS scale and the PDW and its subscales; and a
Table 3
Estimates of Reliability and Explained Common Variance

| No. of Items | Factors   | Content          | Omega (ω and ωS) | Omega H (ωH and ωHS) | ECV (%) | PUC |
|--------------|-----------|------------------|------------------|----------------------|---------|-----|
|              | Model 6: BCFA 3-Factor                           |                  |                  |                      |         |     |
| 29           | General   | General ProQOL   | .94              | .72                  | 49.7    | .45 |
| 14           | F1        | Compassion Satisfaction | .94              | .17                  | 9.7     |     |
| 12           | F2        | Burnout          | .84              | .21                  | 15.2    |     |
| 12           | F3        | Secondary Traumatic Stress | .88              | .81                  | 25.5    |     |
|              | Model 5: BCFA 2-Factor                           |                  |                  |                      |         |     |
| 29           | General   | General ProQOL   | .95              | .64                  | 50.3    | .46 |
| 18           | F1        | Compassion Satisfaction | .94              | .17                  | 9.8     |     |
| 11           | F2        | Compassion Fatigue | .91              | .59                  | 39.9    |     |

Note. N = 304. BCFA = bifactor confirmatory factor analysis; ProQOL = Professional Quality of Life scale; ECV = explained common variance; PUC = percentage of uncontaminated correlations.

negative, moderate association between the BO subscale and the WBW and its subscales. Similar associations were found between the BO subscale and PDW and its three subscales when composite scores were used, and the same patterns were found with the STS scale. Similar associations between the ProQOL’s general factor and its latent variables were found in terms of composite scores of the aforementioned instruments (see Supplementary Table 2).

Discriminant Validity

Additionally, Table 4 also shows the Pearson correlations between the ProQOL’s general factor, CS, BO, and STS composite scores and the three selected scales of the LEC. None of the correlations was statistically significant. Similar associations were found between the ProQOL’s general factor and its latent variable scores and LEC composite scores (see Supplementary Table 2).

Discussion

The main objective of this study was to examine the construct validity of the ProQOL scale (Stamm, 2010) using a representative sample of child protection workers. A CFA failed to confirm the adequacy of the three-factor structure proposed by Stamm (2010). However, a bifactor model postulating a structure with a general factor in addition to independent dimensions...
(i.e., CS, BO, and STS scales) showed an acceptable fit of the data. We also found that the general factor accounted for almost half of the common variance of the ProQOL, which suggests unidimensionality of the measurement instrument while also allowing researchers and clinicians to consider each subscale separately. In particular, the STS scale seems to provide valid scaling of individual differences beyond that of the general factor. Still, our analyses tend to suggest that a general factor of professional quality of life explained the largest share of the variance among the ProQOL. The composite score of the bifactor model remained well correlated with the PCL-S, WBW, and PDW, supporting the scale’s convergent validity. The scale’s discriminant validity was supported by its lack of association with the LEC. Several nuances, however, should be emphasized in order to guide future research and the use of the ProQOL.

Bifactor modeling allowed for the validation of a structure comprising a general factor and three group factors (CS, BO, and STS). Therefore, this analysis supports the construct validity of the ProQOL and the theoretical underpinnings of the scale, albeit with a slightly different use than originally intended. Indeed, the results suggest that the instrument may be better suited to reflecting general professional quality of life via a general factor than reflecting separate constructs using scores from the three subscales, as was originally intended (Stamm, 2010). However, the higher loadings of items on the BO and STS subscales suggest that in addition to considering a ProQOL’s general factor, these subscales provide something attributable to their specific contribution, which means they could be used on their own. Although combining CS and CF scores on a continuum runs contrary to the scale’s original intended use, doing so validates the theory proposed by Stamm (2010). Indeed, BO and STS remain two different concepts that become the opposite of CS when merged. On her website, Stamm (2019) conceded, “BO and STS might be collapsed given that they are both statistically and theoretically correlated based on depression-type symptoms, but they retain unique values that separate the two, primarily the element of fear associated with STS.” Thus, the findings of the current study confirm Stamm’s (2010) theory but suggest changes to the way her concepts are measured.

As such, the present study may answer previous questions as to the dimensionality of the ProQOL (Cieslak et al., 2014; Hemsworth et al., 2018; Sprang & Craig, 2015; Watts & Robertson, 2015). Specifically, this study bolsters what was reported by Mathieu (2007), who suggested that the CF framework must be seen as a continuum rather than three different subconcepts measured separately. In the same vein, Geoffrion et al., 2016 argued that the compassion framework should cover a response continuum of work-related stress ranging from fatigue to satisfaction, thereby reflecting the counterbalance of negative and positive outcomes. In other words, these authors refuted the idea that a helper can score simultaneously on CS and CF scales and suggested that these two concepts are not different but rather two opposite poles of the same construct. In all, the present findings validate the unidimensionality of the ProQOL-5, accompanied by specific residuals over and above the unique variance of the general dimensions.

Regarding internal consistency, the ProQOL’s general factor and the three group factors were acceptable, which support its construct validity. Our results, therefore, only partially support previous findings reported by Hemsworth and colleagues (2018) and Heritage and colleagues (2018), who validated the CS and STS scales but called for a review of the BO subscale. Once again, this may help shed light on the mixed results concerning the internal consistency of the three scales of Stamm’s proposed ProQOL (Cieslak et al., 2014). Moreover, it demonstrates that a composite score of the 30 items (i.e., the total for all three subscales) may be computed to assess professional quality of life among helpers. To summarize, the results of the ProQOL’s retained bifactor structure depicted a latent structural model with a single general trait reflecting the target construct the instrument was designed to assess, as well as several orthogonal group factors that represent subdomain constructs and are presumed to arise due to clusters of content similar items (Rodriguez et al., 2016, p. 223).

Convergent and Discriminant Validity

The statistically significant association between the ProQOL’s general factor and group factors and the PCL-S was consistent with the association between these two concepts. Indeed, the ProQOL was theoretically designed to measure, in part, symptoms similar to those experienced with PTSD (Stamm, 2010). Furthermore, the ProQOL’s general factor and group factor composite scores converged with the WBW composite scale and its three subscales, showing a positive association between professional quality of life and psychological well-being at work. On the other hand, the statistically significant negative associations between the ProQOL and the PDW and its three subscales demonstrated that the ProQOL, also designed to measure BO symptoms, showed some agreement with an instrument designed to measure psychological distress at work. In all, our study supports the scale’s convergent validity.

The absence of associations between the scales of the LEC (for non-work–related events) and the ProQOL’s general factor and group factor composite scores support its discriminant validity. The same applies to the absence of associations between the scales of LEC and the ProQOL general factor. More specifically, it suggests that the main aim of the ProQOL, which is to assess professional quality of life, can be reached, as it does not correlate with potentially traumatic experiences encountered by an individual or an acquaintance outside of the professional sphere. In other words, these tests show that the ProQOL does not measure quality of life outside of work or work-related stress encountered prior to the assessment time frame, thus reinforcing its validity as a construct.

These findings support the construct validity of the ProQOL scale in its bifactor structure. Therefore, it appears that this
instrument is an appropriate tool to assess helpers’ professional quality of life, and, more specifically, that of child protection workers. Moreover, a general factor reflecting a continuum that ranges from low professional quality of life (i.e., CF) to high professional quality of life (i.e., compassion satisfaction) is an asset for clinicians in the measurement of work-related quality of life as it is easier to grasp, use, and calculate than a measurement consisting of multiple scores across multiple scales. Indeed, a composite score of the 30 items can be easily computed and interpreted (see the Supplementary Materials). The use of a continuum ranging from CF to CS will undoubtedly help clinicians increase the accuracy of their clinical assessments. Yet, if needed, the bifactor structure still allows for individual scoring of CS, BO, and STS, as proposed by Stamm (2010). Therefore, the proposed continuum is easier to understand and could potentially help improve the accuracy of assessments.

Despite its clear contributions, this study has some limitations that deserve mention. First, despite the use of a random sample, this sample is only representative of workers in child protection. Results could vary by work setting or professional identity (e.g., psychiatric nurses vs. community workers). These results should be replicated with other populations of helpers. Second, as with any self-reported measurement instrument, social desirability may have affected the response to different items (Moorman & Podsakoff, 1992). Third, given its cross-sectional design, this study could not assess the fluctuations across time of the scores or test–retest validity. Fourth, criterion validity could not be assessed, as we had no variables that allowed the examination of outside criteria that would influence results of the ProQOL composite score. Fifth, although the concepts of CF and CS are culturally appropriate for this sample, only the French-language version of the ProQOL was validated. Finally, although our bottom and top score quartiles were very similar to those proposed by Stamm (2010), it is important to replicate these results using the English and French versions and to compare the construct validity between the two versions with a measurement invariance analysis.

References
Berzoff, J., & Kita, E. (2010). Compassion fatigue and countertransference: Two different concepts. Clinical Social Work Journal, 38, 341–349. https://doi.org/10.1007/s10615-010-0271-8
Blanchard, E. B., Jones-Alexander, J., Buckley, T. C., & Forneris, C. A. (1996). Psychometric properties of the PTSD Checklist (PCL). Behaviour Research and Therapy, 34, 669–673. https://doi.org/10.1016/0005-7967(96)00033-2
Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. Structural Equation Modeling: A Multidisciplinary Journal, 22, 504–516. https://doi.org/10.1080/10705511.2014.938596
Bride, B. E., Radley, M., & Figley, C. R. (2007). Measuring compassion fatigue. Clinical Social Work Journal, 35, 155–163. http://doi.org/10.1007/s10615-007-0091-7
Brislin, R. W. (1980). Translation and content analysis of oral and written material. Handbook of Cross-Cultural Psychology, 2, 349–444.
Mathieu, F. (2007). Running on empty: Compassion fatigue in health professionals. *Rehab & Community Care Medicine, 4*, 1–7. http://compassionfatigue.org/pages/RunningOnEmpty.pdf

McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology, 65*, 131–149. https://doi.org/10.1111/j.2044-8325.1992.tb00490.x

Musa, S. A., & Hamid, A. A. (2008). Psychological problems among aid workers operating in Darfur. *Social Behavior and Personality: An International Journal, 36*, 407–416. https://doi.org/10.2224/sbp.2008.36.3.407

Muthén, L. K., & Muthén, B. O. (2008). Mplus (Version 5.1). Los Angeles, CA: Muthén & Muthén.

Osłowsky, J. D., Putnam, F. W., & Lederman, J. C. S. (2008). How to maintain emotional health when working with trauma. *Juvenile and Family Court Journal, 59*, 91–102. https://doi.org/10.1111/j.1755-6988.2008.00023.x

Palestini, L., Prati, G., Pietrantoni, L., & Cicognani, E. (2009). La qualità della vita professionale nel lavoro di soccorso: Un contributo alla validazione italiana della Professional Quality of Life Scale (ProQOL) [Professional quality of life in rescue work: Supporting the validation of the Italian version of the Professional Quality of Life Scale (ProQOL)]. *Psicoterapia Cognitiva E Comportamentale, 15*, 205–227.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*, 129–140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31. https://doi.org/10.1007/s11136-007-9183-7

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*, 5–26. https://doi.org/10.1177/0016496612449831

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glib: Comments on Sijtsma. *Psychometrika, 74*, 145–154. https://doi.org/10.1007/s11336-008-9102-z

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*, 137–150. http://doi.org/10.1037/met0000045

Samson, T., Ieiovich, E., & Shvartzman, P. (2016). Psychometric characteristics of the Hebrew version of the Professional Quality-of-Life Scale. *Journal of Pain and Symptom Management, 52*, 575–581. https://doi.org/10.1016/j.jpainsymman.2016.03.019

Shen, J., Yu, H., Zhang, Y., & Jiang, A. (2015). Professional quality of life: A cross-sectional survey among Chinese clinical nurses. *Nursing & Health Sciences, 17*, 507–515. https://doi.org/10.1111/nhs.12228

Simms, L. J., & Watson, D. (2007). The construct validation approach to personality scale construction. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 240–258). New York, NY: Guilford Press.

Smart, D., English, A., James, J., Wilson, M., Daratha, K. B., Childers, B., & Magera, C. (2014). Compassion fatigue and satisfaction: A cross-sectional survey among U.S. healthcare workers. *Nursing & Health Sciences, 16*, 3–10. https://doi.org/10.1111/nhs.12068

Søndeenaa, E., Lauvrud, C., Sandvik, M., Nonstad, K., & Whittington, R. (2013). Resilience and professional quality of life in staff working with people with intellectual disabilities and offending behavior in community based and institutional settings. *Health Psychology Research, 1*(1). https://doi.org/10.4081/hpr.2013.e3

Sprang, G., & Craig, C. (2015). An inter-battery exploratory factor analysis of primary and secondary traumatic stress: Determining a best practice approach to assessment. *Best Practices in Mental Health, 11*, 1–13.

Stamm, B. H. (2002). Measuring compassion satisfaction as well as fatigue: Developmental history of the Compassion Satisfaction and Fatigue Test. In C. R. Figley (Ed.), *Treating compassion fatigue* (pp. 107–119). New York, NY: Brunner-Routledge.

Stamm, B. H. (2009). *Professional Quality of Life: Compassion Satisfaction and Fatigue Version 5 (ProQOL)* Retrieved from: https://proqol.org/uploads/ProQOL-5_French.pdf

Stamm, B. H. (2010). The concise ProQOL manual. Pocatello, ID: ProQOL.org. Retrieved from http://ProQOL.org/uploads/ProQOL_Concise_2ndEd_12-2010.pdf

Stamm, B. H. (2019). *How the ProQOL is used in research*. Retrieved from https://proqol.org/Research.html

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*, 893–898. https://doi.org/10.1016/j.paid.2006.09.017

Thomas, P., Billon, R., Chaumeri, J. A., Barruche, G., & Thomas, C. H. (2014). Psycho-social hazards for staff in geriatrics and geriatric psychiatry. *Open Journal of Psychiatry, 4*(02), 91–98. https://doi.org/10.4236/ojpsych.2014.42013

Ventureyra, V. A. G., Yao, S. -N., Cottraux, J., Note, I., & De Mey-Guillard, C. (2001). The validation of the Posttraumatic Stress Disorder Checklist Scale in posttraumatic stress disorder and nonclinical subjects. *Psychology and Psycosomatics, 51*, 47–53. https://doi.org/10.1159/000049343

Voss Horrell, S. C., Holohan, D. R., Didion, L. M., & Vance, G. T. (2011). Treating traumatized OEF/OIF veterans: How does trauma treatment affect the clinician? *Professional Psychology: Research and Practice, 42*, 79–86. https://doi.org/10.1037/a0024163

Watts, J., & Robertson, N. (2015). Selecting a measure for assessing secondary trauma in nurses. *Nurse Researcher, 23*, 30–35. https://doi.org/10.7748/nr.23.2.30.s7

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility*. Paper presented at the International Society for Traumatic Stress Studies annual meeting, San Antonio, TX.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In Hoyle, R. (Ed.) *Handbook of structural equation modeling* (pp. 209–231). New York, NY: Guilford Press.