

Multidimensional Rasch Analysis of the Arabic Occupational Fatigue Exhaustion Recovery Subscales

Michael Edward Clinton, PhD, RN

Lina Sabbagh Younan, DNS, RN

Souha Fares, PhD

American University of Beirut, Beirut, Lebanon

Background and Purpose: Investigation of the psychometric properties of the Arabic version of the Occupational Fatigue Exhaustion Recovery (A-OFER) subscales. **Methods:** Partial credit model (PCM) analyses of 2037 Lebanese nurses' responses to the chronic fatigue (CF), acute fatigue (AF), and intershift recovery (IR) subscales. **Results:** The data were a better fit to the three-dimensional Rasch PCM; difference, $\chi^2 = 2199.3$, $df = 5$, $p = .01$; unidimensional Akaike information criterion (AIC) = 107355; multidimensional AIC = 105166. Rating categories were disordered and there was item dependence among negatively-keyed items. All items were free of subgroup bias. Inter-subscale correlations were contrary to expectations. **Conclusion:** The CF subscale has sufficient reliability for screening nurses in Lebanon and the Eastern Mediterranean region for low to moderate levels of occupational fatigue. The construct validity of the A-OFER requires further investigation.

Keywords: psychometrics; registered nurses; fatigue; Lebanon

An increasing number of international studies have focused on the relationship between occupational fatigue in nurses and patient safety, occupational health, and the long-term health consequences of work-related fatigue. Investigators have examined the prevalence and consequences of excessive work hours in nursing (Kunaviktikul et al., 2015; Logie & Geiger-Brown, 2017), psychological factors as predictors of fatigue (Rahman, Abdul-Mumin, & Naing, 2016), and the relationship between work-related fatigue and musculoskeletal disorders in emergency room and critical care nurses (Rahman, Abdul-Mumin, & Naing, 2017). Other investigators have examined fatigue among hospital nurses from a work systems perspective (Steege, Pasupathy, & Drake, 2018), fatigue among nurse leaders (Steege, Pinekenstein, Arsenault Knudsen, & Rainbow, 2017), and the interrelationship between fatigue and intentions to leave (Liu et al., 2016). Similar studies are required to address two research imperatives in Lebanon: the impact of nurse occupational fatigue on patient outcomes and the long-term health consequences of work-related fatigue among nurses.

Parsimonious, valid, and reliable measures of occupational fatigue are required to support programs of research focused on occupational fatigue among nurses in Lebanon and the Eastern Mediterranean region. The need for Arabic versions of suitable measures is particularly urgent. Consequently, the suitability of the Occupational Fatigue Exhaustion Recovery (OFER) (Winwood, Lushington, & Winefield, 2006) subscales for translation into Arabic was evaluated in a previous study (Clinton & Tchapanian, 2017). That study showed that the OFER items were suitable for translation into Arabic.

The purpose of this article is to report the psychometric characteristics of the Arabic version of the OFER subscales. We use OFER to refer to the English versions of the subscales and A-OFER to refer to the Arabic versions.

A national study was conducted in Lebanon to investigate the relationship between occupational fatigue and the organization of nursing work. Fatigue was measured with Arabic translations of the OFER subscales (Rella, Winwood, & Lushington, 2009; Winwood et al., 2006; Winwood, Winefield, & Lushington, 2006) approved by the developers. A preliminary study was conducted with the English versions of the subscales, which were administered to 366 registered nurses (RNs) at an academic medical center in Beirut (Clinton & Tchapanian, 2017). A sample of 3,057 nurses from all governorates of Lebanon were recruited for the national survey. The characteristics of the sample were reported previously (Clinton, Younan, & Fares, 2018). The reliability and validity of the A-OFER subscales were examined by fitting the data of 2,037 respondents to the unidimensional and a multidimensional partial credit model (PCM) (Masters, 1982).

METHODS

Design

The study was a secondary analysis of cross-sectional survey data collected in a national survey of work organization, scope of practice, fatigue, and muscular-skeletal injuries among nurses in Lebanon. The results of the wider study will be reported in forthcoming publications.

Sample

Data collected from 2,037 bedside nurses in public and private hospitals with at least 100 beds across all governorates in Lebanon (Akkar, Baalbek, Beirut, Beqaa, Mount Lebanon, North Lebanon, Nabatiyeh, and South Lebanon) were analyzed. All 2,037 respondents met the inclusion criteria for the study: working for current employer in a hospital with a minimum of 100 beds and facilities for 48 hour care postadmission for at least 1 year with the 3 most recent months in the same unit.

Recruitment

The directors of nursing and chief executive officers of hospitals with inpatient facilities were contacted for permission to conduct the survey. The Order of Nurses in Lebanon helped distribute survey packages to the hospitals. Each package contained a consent document and the survey questionnaire. The American University of Beirut Social and Behavioral Sciences Institutional Review Board approved the study. Respondents confirmed voluntary informed consent by returning questionnaires in sealed envelopes to drop boxes in administrative offices at their hospitals.

Instrument

Available in English, French, and now Arabic, the 15-item OFER scale was developed to measure work-related fatigue (Winwood et al., 2006). A seven-category rating scale (0–6), where 0 = Strongly Disagree and 6 = Strongly Agree, is used to ensure sufficient sensitivity. Subscale scores for chronic fatigue (CF), acute fatigue (AF), and intershift recovery (IR) are converted to quotients. Winwood et al. (2006) noted that negatively-keyed items can result in errors caused by respondent carelessness and that scales that have only positive keyed items lead to artificial factor solutions and may lack unidimensionality. The OFER has 5 negatively-keyed items and 10 positively-keyed items.

The OFER is unique in measuring IR. Winwood et al. (2006) report that gender neutrality of the OFER was confirmed in a pilot study conducted on female nurses and male quarry workers. The IR subscale is hypothesized to correlate negatively with both the CF and AF subscales with a stronger inverse relationship between the IR and AF subscales. The reasoning behind these predictions is that persistent low recovery from high levels of IR is associated with higher levels of CF (Winwood et al., 2006). The explicit theoretical assumptions on which the hypothesized relationships are based are that acute occupational fatigue is mediated through work–time factors whereas CF is an evolving trait mediated by persistent low levels of IR (Winwood et al., 2006).

Translation. The second author translated the OFER into Arabic by following World Health Organization (2010) guidelines. A bilingual expert panel removed ambiguous and inconsistent phrases from the translation before an independent nurse academic, blind to the original instrument, back-translated the Arabic version into English. The second author then conducted cognitive interviews (Miller, Willson, Chepp, & Padilla, 2015) with bilingual RNs to make the Arabic version more consistent with the English version. A panel of expert nurses reviewed, edited, and approved the final version of the A-OFER subscales. Dr. Peter Winwood, the author of the OFER subscales, approved the Arabic versions and gave permission for them to be used in the national study.

DATA ANALYSIS

Model Selection for Data Analysis

Traditional approaches to psychometric evaluation of surveys and measurement instruments assume that raw scores and their linear transformations are interval-level data. Rasch models assume that a survey or rating scale measures participants on a single latent dimension (Wright & Masters, 1982). If data do not fit the Rasch model, the measure is not unidimensional, and measures of performance on the latent scale are inaccurate because they cannot be estimated precisely (Wu, Adams, & Wilson, 1997). Multidimensional statistical models exploit correlations between subscales and improve measurement precision by reducing standard errors (Wu, Adams, Wilson, & Haldane, 2007).

A multidimensional approach was suitable for analyzing the psychometric properties of the A-OFER scale because it measures related but distinct dimensions of occupational fatigue. The higher the correlations between subscales, the greater the measurement precision achieved by the multidimensional approach. If there are enough items in each subscale, composite unidimensional estimates are accurate enough for most research purposes. When subscales have fewer items than needed for acceptable reliability, the multidimensional approach provides more precise measures. The multidimensional approach has the added advantage of estimating correlations between latent dimensions

without measurement error. It is better to use the unidimensional Rasch model when latent dimensions are highly correlated.

The multidimensional random coefficients multinomial logit model (MRCMLM) (Wu et al., 2007) was the most appropriate model for this study. Liu, Wilson, and Paek (2008) formulate the MRCMLM with the following equation:

$$P(X_{nik} = 1 | \xi_n, \xi) = \frac{\exp \left[b'_{ik} \xi_n + a'_{ik} \xi \right]}{\sum_{k=1}^{ki} \xi \exp \left[b'_{ik} \xi_n + a'_{ik} \xi \right]}$$

where $X_{nik} = 1$ if person n responds to item i in category k (otherwise $X_{nik} = 0$). Density functions for items, categories, and persons are estimated with X_{ni1} fixed to zero to identify the model. θ_n is a $d \times 1$ performance vector for respondent n ($1 \leq d$ [dimension]). b'_{ik} is a $1 \times d$ scoring vector for category k of item i ; ξ is a $p \times 1$ item parameter vector; and a'_{ik} is a $1 \times p$ vector that specifies the linear combinations of p elements of ξ for each response category. ξ is a fixed unknown parameter vector and θ_n is a random parameter vector (Liu et al., 2008). The elements of θ_n are assumed to follow a multivariate normal distribution (Liu et al., 2008). The MRCMLM uses marginal maximum likelihood (MML) to estimate parameters. Wu et al. (2007) explain the mathematics behind the MRCMLM and its use in estimating a range of multidimensional models including the Rasch rating scale model (RSM) (Andrich, 1978) and the PCM (Masters, 1982).

Unlike the unidimensional Rasch model, MRCMLM analysis assumes that a set of latent traits underlie individual responses to survey items. Since the A-OFER has three subscales, responses (levels of occupational fatigue) are modeled in three-dimensional latent space rather than as a point on a single latent dimension. "Performance level" in this study means self-reported occupational fatigue as measured by the A-OFER subscales. A respondent who selects a higher frequency rating category for an item has a higher "performance level" for that item than one who responds to the item in a lower frequency category.

RESULTS

Unidimensional Analysis

The unidimensionality of each of the three A-OFER subscales was explored with Winsteps v. 4.0.0 software (Linacre, 2017) using principal component analysis of residuals. The analysis was repeated for all 15 items in the A-OFER scale. Table 1 shows the eigenvalues for the Rasch dimension and first contrasts for the measures. The first contrast eigenvalues for the AF subscale and the A-OFER scale were higher than 2, which indicated multidimensionality in both scales. Using the criterion of 0.5 logits difference, there was no evidence of differential item functioning for any of the A-OFER items based on age, gender, marital status, nursing qualification (below vs. at or above university baccalaureate level), or workload (above or below the median patient-nurse ratio for the sample). The absence of differential item functioning indicated that the A-OFER items were independent of the subgroups of respondents in the sample and that the data met the requirement of measurement invariance.

TABLE 1. Eigenvalues for Rasch Dimensions and First Contrasts for A-OFER Scale and Subscales

Subscale/Scales	Eigenvalues	
	Rasch Dimension	First Contrast
CF	7.1	1.5
AF	7.0	3.1
IR	5.1	1.9
Scale: A-OFER (CF, AF, IR)	12.3	3.4

Note. AF = acute fatigue; CF = chronic fatigue; IR = intershift recovery.

Q_3 statistics were examined to identify item dependence. For a five-item scale, a Q_3 statistic or residual correlation between items of -0.25 is ideal (Yen, 1984, 1993). The Q_3 statistics for the A-OFER subscales ranged from -0.41 to -0.04 for the CF items, -0.54 to 0.53 for the AF items, and -0.13 to 0.23 for the IR items, which suggested item dependence in all three subscales. When residual correlations were examined for the 15 items in the A-OFER scale, Q_3 values ranged from -0.26 to 0.72 , which confirmed item dependence (ideal value -0.071). Item dependence was most noticeable for the negatively-keyed items in the A-OFER subscales.

The A-OFER subscales had negligible floor effects, but the CF subscale had a ceiling effect of 16.8% that indicated poor targeting and inability to measure precisely the highest levels of CF in the sample. The 15 items in the A-OFER scale had an operational range of approximately six logits. Figure 1 shows that the seven-category A-OFER rating scale

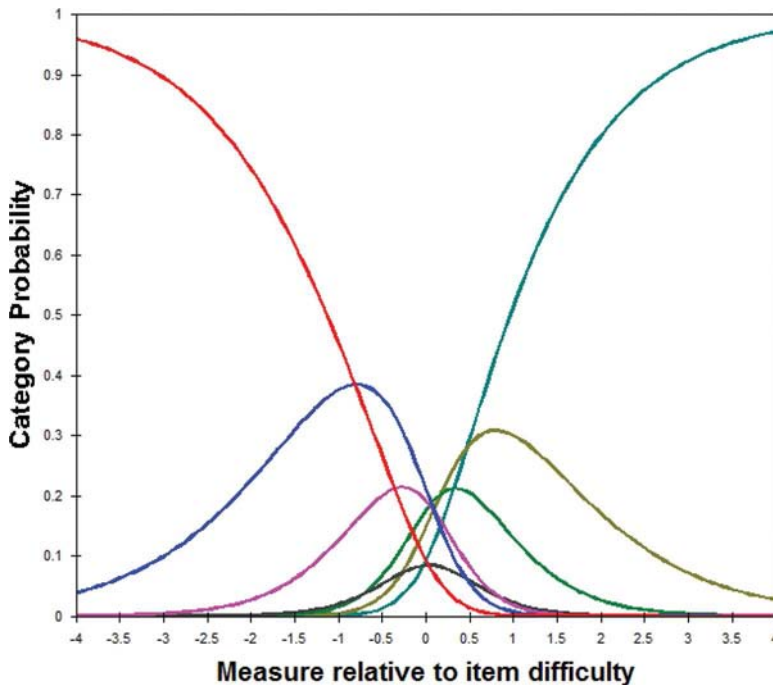


Figure 1. Item characteristic curves for A-OFER seven-category rating scale.

did not function as intended for this sample because responses in rating categories 2, 3, and 4 were never the most probable responses. The reported eigenvalues for first contrasts, Q_3 statistics, and weighted mean-square (MNSQ) outside the sample adjusted expected values of 0.96 to 1.04 (Smith, Schumacker, & Bush, 1998) indicated that the A-OFER did not satisfy the Rasch measurement assumptions of unidimensionality and local item independence. The failure of the scale to meet these basic assumptions of Rasch measurement indicated the need for multidimensional analysis.

Multidimensional Analysis

ConQuest v.2 software (Wu et al., 2007) was used to apply the MRCMLM by fitting the data to the PCM. The PCM was selected because it was a better fit for the A-OFER scale than the RSM ($\chi^2 = 746.12$ $df = 70$, $p < .01$). The PCM was appropriate as well because the A-OFER scale and subscales have ordered response categories. The three subscales in the A-OFER scale were estimated and standard fit statistics were computed for each item. Table 2 shows that items B and D in the CF subscale, items I and J in the AF subscale, and items L and M in the IR subscale had weighted MNSQ values outside their 95% confidence intervals. According to Wu et al. (2007), MNSQ values outside the 95% confidence intervals misfit the PCM model because they have an absolute T value of ≥ 2 . MNSQ

TABLE 2. Infit and Outfit Values for 15 A-OFER Items by Subscale

Subscale/Scales	MNSQ	
	Weighted (CI)	Unweighted (CI)
CF		
Item A	1.10 (0.90, 1.10)	0.92 (0.94, 1.06)
Item B	1.10 (0.93, 1.07)	1.08 (0.94, 1.06)
Item C	1.01 (0.93, 1.07)	0.98 (0.94, 1.06)
Item D	1.24 (0.93, 1.07)	1.25 (0.94, 1.06)
Item E	1.07 (0.93, 1.07)	1.01 (0.94, 1.06)
AF		
Item F	0.90 (0.92, 1.08)	0.89 (0.94, 1.06)
Item G	0.87 (0.90, 1.10)	0.77 (0.94, 1.06)
Item H	0.86 (0.93, 1.07)	0.81 (0.94, 1.06)
Item I	1.14 (0.96, 1.04)	1.16 (0.94, 1.06)
Item J	1.15 (0.95, 1.05)	1.20 (0.94, 1.06)
IR		
Item K	0.98 (0.95, 1.05)	1.05 (0.94, 1.06)
Item L	1.08 (0.94, 1.06)	1.09 (0.94, 1.06)
Item M	1.07 (0.95, 1.05)	1.18 (0.94, 1.06)
Item N	1.02 (0.95, 1.05)	1.01 (0.94, 1.06)
Item O	0.85 (0.95, 1.05)	0.85 (0.94, 1.06)

Note. AF = acute fatigue; CF = chronic fatigue; CI = confidence interval; IR = intershift recovery; MNSQ = mean-square.

values <1 are normally ignored because overfitting items are rarely a problem when calibrating item parameters. Except for items A, G, H, J, and M, the weighted MNSQ and unweighted MNSQ values were similar, which indicated that item parameters were not distorted by respondents with outlier scores. Since the departure of the misfitting items from the expected Rasch estimates was modest, all 15 A-OFER items were retained for further analysis.

Table 3 shows the item difficulty parameters and standard errors for the 15 A-OFER items. The difficulty parameters were calibrated on a latent dimension common to both the difficulty of the items and the amount of the latent trait required to respond to the items in the different rating categories. Since all the standard errors were small, the difficulty of the parameters was estimated with sufficient precision. The parameters shown with asterisks are constrained values required to identify the measurement model. By default, ConQuest v.4 constrains the value of the last item in each subscale to achieve model identification (Wu et al., 2007).

The multidimensional approach to item analysis increases measurement precision by calibrating item parameters on all latent dimensions simultaneously (Wu et al., 2007). Table 4 shows direct estimates of the variance-covariance and correlation matrices for the

TABLE 3. Item Difficulty Parameters and Standard Errors for 15 A-OFER Items by Subscale

Subscale/Scales	Item Difficulty (<i>SE</i>)
CF	
Item A	-0.691 (0.016)
Item B	0.418 (0.013)
Item C	0.128 (0.014)
Item D	0.292 (0.013)
Item E	-0.148 (0.029) ^a
AF	
Item F	-0.193 (0.009)
Item G	-0.510 (0.010)
Item H	-0.322 (0.009)
Item I	0.404 (0.008)
Item J	0.621 (0.019) ^a
IR	
Item K	-0.226 (0.009)
Item L	0.475 (0.010)
Item M	-0.182 (0.009)
Item N	0.247 (0.009)
Item O	-0.314 (0.019) ^a

Note. AF = acute fatigue; CF = chronic fatigue; IR = intershift recovery; *SE* = standard error.

^aconstrained values.

TABLE 4. Variances, Covariances, and Correlations for A-OFER Subscales

Scale and Subscales	1	2	3
OFER-1			
CF		<i>0.160</i>	<i>0.248</i>
AF	0.596		<i>0.020</i>
IR	0.760	0.250	

Note. AF = acute fatigue; CF = chronic fatigue; IR = intershift recovery. Correlations are in roman; covariance are in italics.

A-OFER subscales. The subscale correlations ranged from 0.25 for the AF and IR subscales to 0.76 for the CF and IR subscales. The 0.60 correlation of the CF and IR subscales implies that they had reasonable convergent validity and assessed a reasonably broad concept of fatigue.

The ConQuest software (Wu et al., 2007) reports reliability indices analogous to Cronbach's alpha. The expected a posteriori/plausible value ratio (EAP/PV) is an index of reliability obtained by dividing the variance of the expected a posteriori ability estimates by the total variance of person measures and indicates how well a set of items separates individuals along the latent measurement scale. The EAP/PV ratio of 0.82 for the CF subscale indicated that it had acceptable reliability (≥ 0.70) whereas that of 0.54 for the AF subscale indicated that it might not measure a single dimension of occupational fatigue. The EAP/PV ratio of 0.68 for the IR subscale indicated that it had acceptable internal consistency. The EAP/PV ratios for all three subscales reflect the weighted MNSQ values outside confidence intervals reported in Table 2. Maximum likelihood estimation (MLE) person separation values for the three A-OFER subscales ranged from a low 0.51 for the AF subscale to a moderate 0.63 for the IR subscale and indicated that the measures had low to modest ability to distribute respondents along the underlying latent dimensions (Wright & Stone, 1999). The item separation reliability of 0.99 estimated by the ConQuest software indicated that the 15 A-OFER items were completely separated along the latent trait and that every item contributed to the interpretation of occupational fatigue.

DISCUSSION

The results reported show that the A-OFER does not constitute a unidimensional measure of occupational fatigue. As reported by Winwood et al. (2006), the 15 OFER items consist in three separate subscales that measure CF, AF, and IR. The results of the unidimensional analysis confirmed that the A-OFER items were independent of respondent subgroup characteristics when used with this sample and, therefore, had generalizability (Messick, 1995). The results reported identified possible item dependence in all three subscales, most noticeable for the negatively-keyed items. For this sample, the seven-category A-OFER rating scale lacked structural validity (Messick, 1995) because it did not function as intended for this sample. For future studies, the seven categories can be collapsed into three-rating categories (1 = disagree, 2 = neutral, 3 = agree). Alternatively, future investigators could consider using dichotomized rating scale categories. Overall, the results of the unidimensional analysis confirmed the need for multidimensional analysis of the A-OFER items and

subscales. The standard errors for the item difficulty parameters were small enough to confirm measurement precision. Several items misfitted the multidimensional PCM, but their weighted MNSQ values were too small to justify deletion.

Validity

The correlations between the A-OFER subscales shown in Table 4 can be compared with those reported by Winwood et al. (2006). The correlation of 0.60 between the CF and AF subscales found in this study was close to that of 0.61 reported by Winwood et al. (2006). However, contrary to the negative correlations of -0.53 for the CF and IR subscales and of -0.61 for the AF and IR subscales reported by Winwood et al. (2006), both correlations were positive in this study. The unexpected positive correlations between the three subscales brings into question the substantive validity (Messick, 1995) of the A-OFER subscales because, contrary to Winwood et al. (2006) hypothesis, IR did not mediate occupational fatigue in this sample. There are three reasons why the correlations between the A-OFER subscales were positive: miscoding prior to data analysis, respondent inattention to the wording of the A-OFER items, and invalidity of the Winwood et al. (2006) IR hypothesis.

Miscoding was ruled out when two of the authors rechecked the data file independently and confirmed that all items had been properly coded. The second possibility was examined by reversing the coding of the negatively-keyed items in the IR subscale and repeating the multidimensional PCM analysis. The results of the repeated analysis indicated that respondent inattention or carelessness could not be ruled out because the IR subscale correlated negatively with both the AF and CF subscales. However, contrary to Winwood et al. (2006) prediction, the negative correlation of the IR and CF subscales of -0.48 was weaker than that of -0.86 for the IR and AF subscales. The strength of the negative associations among the A-OFER subscales leaves open the possibility that Winwood et al. (2006) assumptions about the relationships among the IR, AF, and CF subscales are invalid. Consequently, more studies are required to assess the substantive validity of the A-OFER subscales.

In addition to structural validity, substantive validity, and generalizability, Messick (1995) identified three other general standards for assessing the construct validity of psychological and educational tests: content validity, external validity, and consequential validity. Winwood et al. (2006) describe the extensive work done to finalize the OFER items and to establish the external validity of the CF, AF, and IR subscales. Messick's (1995) consequential aspect of construct validity is not mentioned by Winwood et al. (2006). The results reported above confirmed that the A-OFER is free of cultural bias when used with Arabic-speaking nurses working in acute hospitals.

This study did not establish the distributive justice of the A-OFER subscales because the reported ceiling effect prevented the CF subscale from measuring occupational fatigue among respondents with the highest levels. This limitation can be addressed by adding items to the A-OFER CF subscale. With items better targeted to nurses with high levels of CF, the A-OFER CF subscale will be a suitable screening instrument for occupational fatigue, but to fulfill the requirement of distributive justice investigators need to establish the external validity of the A-OFER scales independently of the results reported by Winwood et al. (2006).

The A-OFER measures fatigue as a trait and, therefore, as an attribute of individual nurses, but occupational fatigue can be conceptualized as an attribute of healthcare systems. Future studies with the A-OFER are required that investigate the relationships among

patient acuity, organization of nursing work, nursing workload, and occupational fatigue. Without such studies, the extent to which distributive justice is achieved when interpreting A-OFER scores cannot be established. Conversely, measures are required that assess the extent to which fatigue is related to lifestyle and outside-of-work activities, otherwise the assumption that occupational fatigue is associated solely with workplace factors cannot be evaluated. Comprehensive studies that measure the impact of fatigue on leisure time and lifestyle fatigue on work will achieve distributive justice for health-sector employers as well as for nurses and other health workers.

The reported lack of structural validity of the A-OFER is unrelated to the translation of the OFER items. In a previous study OFER subscales administered in English (Clinton & Tchapanian, 2017), similar rating category disordering was evident. The seven-category rating scale is clearly too nuanced for samples recruited in Lebanon irrespective of whether the OFER subscales are administered in English or Arabic. Possible solutions are to use a three-category rating scale (disagree, neutral, agree) or to dichotomize the rating scale. However, either solution will reduce raw score variance and measurement sensitivity. Therefore, we confirm our previous advice that investigators should train respondents in using seven-category rating scales and in responding to negatively-keyed items before administering OFER scales in English or in translation in Lebanon and the Eastern Mediterranean region (Clinton & Tchapanian, 2017).

The results reported for the A-OFER confirm the findings of our previous study (Clinton & Tchapanian, 2017). The CF subscale has high ceiling effects and respondents to the CF and all three OFER subscales tend to overestimate occupational fatigue at lower and higher levels irrespective of whether the OFER subscales are administered in English or Arabic. The disattenuated correlations between the A-OFER subscales in Table 2 were noticeably different from those reported in our previous study (Clinton & Tchapanian, 2017). The correlation for the CF and AF subscales were lower than that reported for when we administered the OFER questionnaire in English to a different sample of nurses in Lebanon ($0.56 < 0.77$). The positive correlation of the IR subscale with the CF and AR subscales reported for the current study confirms our previous finding that OFER subscale correlations are positive when used with nurses in Lebanon (Clinton & Tchapanian, 2017). This finding requires further investigation after training respondents in the use of negatively-keyed items before administering the A-OFER subscales in Lebanon and the region because it brings into question the validity of the hypothesized relationship among the IR and CF and AR scales.

Reliability

Compared with the estimates reported for the reliability of the OFER subscales reported for the previous study (Clinton & Tchapanian, 2017), the MLE values were lower for the A-OFER CF ($0.58 < 0.65$) and the AF ($0.51 < 0.73$) scales and higher for the IR subscale ($0.63 > 0.48$). However, guidance on whether a measurement scale has sufficient reliability is arbitrary (Taber, 2018). The rule of thumb of regarding measures as sufficiently reliable if they have internal consistency of ≥ 0.7 is widely accepted, but lower values have been used for low-stakes studies and higher values are typically required for clinical measures (Taber, 2018).

The reliability required for measures of occupational fatigue depends on the purpose of measurement. Reliability of 0.65 may be sufficient for screening purposes, but internal consistency of ≥ 0.70 is desirable for initial studies investigating the impact of interventions aimed at reducing occupational fatigue (Nunnally & Bernstein, 1994). The

Spearman–Brown formula was used to identify that two items need to be added to the CF and AF subscales and one item to the IR subscale to achieve reliability of 0.65. If reliability of 0.70 is preferred, four items need to be added to the CF subscale, seven items to the AF subscale, and two items to the IR subscales. The ceiling effect of the CF subscale can be corrected by adding items that measure higher levels of occupational fatigue.

Limitations

The cross-sectional design of the national survey prevented investigation of the predictive validity of the A-OFFER subscales. Self-completion of the survey by respondents without the presence of a member of the research team produced response errors that might have been avoided. The failure of the study to validate Winwood et al. (2006) theoretical assumptions about the interrelationships among the A-OFFER subscales is due either to misunderstanding of the negatively-keyed items or lack of empirical support for the hypothesized relationships. The better explanation cannot be identified from this study. Longitudinal studies are required to examine whether there is empirical support for the theoretical assumptions on which the A-OFFER is based after providing training in using seven-category rating scales and responding to positively-keyed and negatively-keyed items.

CONCLUSION

The A-OFFER subscales have potential for screening nurses in Lebanon and the Eastern Mediterranean region for occupational fatigue. The seven-category rating scale did not function as intended with this sample and respondents were inattentive to negatively-keyed items. Although the substantive validity of the subscales was not supported by this study, the A-OFFER items were found to have good generalizability. The hypothesized relationships among the A-OFFER subscales could not be validated in this study. Further studies are required to validate and establish the reliability of the A-OFFER subscales. The CF subscale requires additional items to measure nurses with the highest levels of occupational fatigue. The AF subscale needs additional items to increase its reliability. Respondents need prior training in the use of seven-category rating scales and in responding to negatively-keyed items. The presence of a trained research assistant to administer the A-OFFER subscales and respond to questions may assist respondents to overcome problems with negatively-keyed items. The convergent and divergent validity of the A-OFFER subscales remain to be established in Arabic-speaking countries. Studies are required to investigate the external validity and consequential validity of the A-OFFER subscales. The CF scale is suitable for screening nurses in Lebanon and Eastern Mediterranean region for low to moderate levels of chronic occupational fatigue.

REFERENCES

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573. doi:10.1007/BF02293814
- Kunaviktikul, W., Chitpakdee, B., Akkadechanunt, T., Sirakamon, S., Wichaikhum, O., Nantsupawat, A., & Supamane, T. (2015). Nurses' extended work hours: Patient, nurse and organizational outcomes. *International Nursing Review*, 62(3), 386–393. doi:10.1111/inr.12195

- Linacre, J. M. (2017). Winsteps® (Version 4.0.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved from <http://www.winsteps.com>
- Liu, O. L., Wilson, M., & Paek, I. (2008). A multidimensional Rasch analysis of gender differences in PISA mathematics. *Journal of Applied Measurement, 9*(1), 18.
- Liu, Y., Wu, L. M., Chou, P. L., Chen, M. H., Yang, L. C., & Hsu, H. T. (2016). The influence of work-related fatigue, work conditions, and personal characteristics on intent to leave among new nurses. *Journal of Nursing Scholarship, 48*(1), 66–73. doi:10.1111/jnu.12181
- Logie, A., & Geiger-Brown, J. (2017). Do RNs in British Columbia work excessive hours? A registry data study. *Journal of Nursing Regulation, 7*(4), 52–60. doi:10.1016/S2155-8256(17)30022-4
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. doi:10.1007/BF02296272
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749. doi:10.1037//0003-066X.50.9.741
- Miller, K., Willson, S., Chepp, V., & Padilla, J. L. (2015). *Cognitive interviewing methodology*. Hoboken, NJ: Wiley.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Rahman, H. A., Abdul-Mumin, K., & Naing, L. (2016). A study into psychosocial factors as predictors of work-related fatigue. *British Journal of Nursing, 25*(13), 757–763. doi:10.12968/bjon.2016.25.13.757
- Rahman, H. A., Abdul-Mumin, K., & Naing, L. (2017). Psychosocial work stressors, work fatigue, and musculoskeletal disorders: Comparison between emergency and critical care nurses in Brunei Public Hospitals. *Asian Nursing Research, 11*(1), 13–18. doi:10.1016/j.anr.2017.01.003
- Rella, S., Winwood, P. C., & Lushington, K. (2009). When does nursing burnout begin? An investigation of the fatigue experience of Australian nursing students. *Journal of Nursing Management, 17*(7), 886–897. doi:10.1111/j.1365-2834.2008.00883.x
- Smith, R. M., Schumacker, R. E., & Bush, J. M. (1998). Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement, 2*, 66–78.
- Steege, L. M., Pasupathy, K. S., & Drake, D. A. (2018). A work systems analysis approach to understanding fatigue in hospital nurses. *Ergonomics, 61*, 148–161. doi:10.1080/00140139.2017.1280186
- Steege, L. M., Pinekenstein, B. J., Arsenault Knudsen, É., & Rainbow, J. G. (2017). Exploring nurse leader fatigue: A mixed methods study. *Journal of Nursing Management, 25*(4), 276–286. doi:10.1111/jonm.12464
- Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273–1296. doi:10.1007/s11165-016-9602-2
- Winwood, P. C., Lushington, K., & Winefield, A. H. (2006). Further development and validation of the Occupational Fatigue Exhaustion Recovery (OFER) scale. *Journal of Occupational and Environmental Medicine, 48*(4), 381–389. doi:10.1097/01.jom.0000194164.14081.06
- Winwood, P. C., Winefield, A. H., & Lushington, K. (2006). Work-related fatigue and recovery: The contribution of age, domestic responsibilities and shiftwork. *Journal of Advanced Nursing, 56*(4), 438–449. doi:10.1111/j.1365-2648.2006.04011.x
- World Health Organization. (2010). *Process of translation and adaptation of instruments*. Retrieved from https://www.who.int/substance_abuse/research_tools/translation/en/
- Wright, B., & Stone, M. (1999). *Measurement essentials* (2nd ed.). Wilmington, DE: Wide Range.

- Wright, B. D., & Masters, G. N. (1982). *Rasch scale analysis: Rasch measurement*. Chicago, IL: MESA Press.
- Wu, M. L., Adams, R.J., & Wilson, M.R. (1997). ConQuest: Multi-Aspect Test Software [computer program]. Camberwell: Australian Council for Educational Research.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). ConQuest: Generalised item response modelling software. Camberwell, Australia: Australian Council for Educational Research.
- Yen, W. M. (1984). Effect of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*, 125–145.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187–213.

Disclosure. The authors have no relevant financial interest or affiliations with any commercial interests related to the subjects discussed within this article.

Acknowledgments. We are grateful to Dr. Peter Winwood for his permission to translate the OFER scale into Arabic for our national survey data. The Arabic version of the OFER scale is available from Dr. Peter Winwood, School of Psychology, Social Work and Social Policy, University of South Australia: Peter.Winwood@unisa.edu.au

Correspondence regarding this article should be directed to Michael Clinton, PhD, RN, American University of Beirut, Hariri School of Nursing, Faculty of Medicine, PO Box 11-0236, Riad El-Solh, Beirut 1107 2020, Lebanon. E-mail: mc42@aub.edu.lb

Copyright of Journal of Nursing Measurement is the property of Springer Publishing Company, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.