Psicothema

Cutoff scores for workplace bullying: The Spanish Short-Negative Acts Questionnaire (S-NAQ)

José M. León-Pérez[1], Iván Sánchez-Iglesias[2], Alfredo Rodríguez-Muñoz[2], and Guy Notelaers[3]
[1] Universidad de Sevilla, [2] Universidad Complutense de Madrid, and [3] University of Bergen

Abstract

Background: Workplace bullying is considered a major social stressor at work. However, in the Spanish context, there is a lack of measures that allow researchers and practitioners to distinguish between non-targets and targets of workplace bullying. **Method:** This study reports the psychometric properties, factor structure, and cutoff scores for the Short-Negative Acts Questionnaire (S-NAQ) in a Spanish sample (N = 1,409). **Results:** The S-NAQ demonstrated adequate psychometric properties. Moreover, both Exploratory and Confirmatory Factor Analysis revealed that a unifactorial model of workplace bullying fit the data best. Two cutoff points for the S-NAQ were formulated using receiver operating characteristic curves to categorize respondents as "not exposed to workplace bullying" (scores below 15), "at risk of being bullied" (scores between 15 and 22), and "targets of workplace bullying" (scores above 22). **Conclusions:** These cutoff scores may help researchers and practitioners in diagnosing workplace bullying and designing intervention strategies.

Keywords: Workplace bullying, mobbing, cutoff scores, thresholds, ROC curve, prevalence.

Resumen

Puntos de corte para el acoso psicológico en el trabajo: el cuestionario breve de actos negativos (S-NAQ) en español. **Antecedentes:** el acoso psicológico en el trabajo es un estresor laboral severo de origen social. Sin embargo, en el contexto español, hay escasez de medidas que nos permitan diagnosticar de forma fiable si una persona está siendo objeto de acoso. **Método:** se informa de las propiedades psicométricas, la estructura factorial y los puntos de corte del Cuestionario Breve de Actos Negativos (S-NAQ) en una muestra española (N = 1.409). **Resultados:** el S-NAQ presenta unas buenas propiedades psicométricas. Los análisis factoriales exploratorios y confirmatorios revelan que la escala tiene una única dimensión. Los datos de las curvas ROC indican dos puntos de corte que nos permiten categorizar a las personas como "no expuestas a conductas de acoso" (puntuaciones menores de 15), "en riesgo de ser acosadas" (puntuaciones entre 15 y 22), o "víctimas de acoso" (puntuaciones mayores de 22). **Conclusiones:** estos puntos de corte pueden facilitar el diagnóstico de acoso y servir para diseñar intervenciones según cada categoría.

Palabras clave: acoso psicológico en el trabajo, mobbing, puntos de corte, curvas ROC, prevalencia.

Workplace bullying is considered one of the most important social stressors at work because it has severe negative consequences for employees' health and well-being (Einarsen & Nielsen, 2015; Hogh et al., 2019). Bullying at work refers to exposure to a range of repeated, continuing negative acts, directed towards one or more targets that typically end up unable to defend themselves (Einarsen, Hoel, Zapf, & Cooper, 2011). Furthermore, bullying victims at work find it difficult to defend themselves against these harassing and humiliating behaviors because they are usually in a position of inferiority with respect to the perpetrator/s and lack of support from their remaining co-workers (Bjørkelo, Einarsen, Nielsen, & Matthiesen, 2011; Samnani, 2013; Park, Bjørkelo, & Blenkinsopp, 2018).

Therefore, as several researchers have indicated (Samnani, 2013; Trépanier, Fernet, & Austin, 2015), workplace bullying can

be considered a severe stressful social situation that attacks people's basic needs (e.g., feelings of belonging to social groups) and may result in learned helplessness (i.e., perceptions of not being able to do anything to reverse the situation and, consequently, ending up not responding to bullying behaviors or following passive coping strategies). Thus, it is not surprising that bullying victims end up experiencing psychosomatic symptoms together with symptoms similar to those in post-traumatic stress disorder (for a review, see Hogh et al., 2019). At a behavioral level, it is common for people who suffer bullying at work to try to avoid the source of stress or the unpleasant circumstances they experience at work, which may result in more absenteeism and sick leaves compared to their non-exposed to bullying colleagues (Eriksen, Hogh, & Hansen, 2016; Magee, Gordon, Robinson, Caputi, & Oades, 2017).

Moreover, as previous studies have shown (Hogh et al., 2019; Rodríguez-Muñoz, Antino, & Sanz-Vergel, 2017), these negative consequences at the individual level also extend to the organizational level (e.g., lower performance and stressful work environment that makes witnesses also report a lower psychological well-being compared to their non-exposed colleagues), the family/private sphere (e.g., increased family conflicts that affect

negatively partners' well-being), or the society in general (e.g., increased costs to social security derived from job losses).

Despite this vast evidence about the detrimental consequences of workplace bullying for people's health and well-being, there is a lack of consensus about how to best measuring workplace bullying, and therefore labeling an individual as target of workplace bullying (Leon-Perez, Escartín, & Giorgi, 2019; Nielsen, Notelaers, & Einarsen, 2011; Zapf, Escartín, Einarsen, Hoel, & Vartia, 2011).

In response, this article offers a short instrument to measure workplace bullying, with cutoff scores that can be used to establish the risk of exposure to workplace bullying. This is the first study establishing cutoffs for identifying potential cases of bullying in Spanish working population.

Regarding workplace bullying measures, this concept is usually included in general epidemiological and sociological surveys. The European Working Conditions Survey (Eurofound, 2015) represents a good example of this approach. Drawing on this survey, the percentage of workers subjected to adverse social behavior in the European Union revolves approximately 16% of the working population. In the case of Spain, results from this survey indicated a prevalence rate around 10%, below the EU mean and quite far from the rates found for Denmark or The Netherlands (25% and 27%, respectively). In fact, there are numerous studies that have pointed out the subjectivity of the criteria adopted to classify a situation as bullying at work based on direct yes/no questions or self-labelling estimation methods (SLM: Nielsen et al., 2009, 2010). This can compromise cross-cultural comparisons because sociocultural factors may introduce some biases in participants' responses (for a review, see Leon-Perez et al., 2019).

Alternatively, based on a behavioral approach, some questionnaires like the EAPA-T (Escartín, Rodríguez-Carballeira, Gómez-Benito, & Zapf, 2010) or the NAQ-R (Einarsen, Notelaers, & Hoel, 2009) have been proposed for the diagnosis of workplace bullying. These questionnaires usually ask participants about the frequency they have been exposed to several negative acts at work without introducing the term bullying. Although these questionnaires may reduce the influence of some subjective biases (e.g., bullying awareness), several studies have concluded that bullying prevalence rates under this estimation method depend to a large extent on the number of negative behaviors that a person may suffer in order to classify a situation as bullying (i.e., exposure to one, two or more negative acts in a certain frequency, ranging from occasionally to daily, see Agervold, 2007). Furthermore, some authors have questioned the validity and reliability of these instruments for differentiating victims vs. non-victims of workplace bullying (Leon-Perez et al., 2014; Nielsen et al., 2009, 2010; Notelaers et al., 2006).

In response, Notelaers et al. (2006) proposed using a Latent Class Cluster (LCC) approach to establish different bullying groups based on patterns of association in the exposure to negative acts according to their conditional probabilities. However, given the methodological complexity of this analysis, Notelaers & Einarsen (2013) have more recently proposed using a Receiver Operating Characteristic (ROC) curve analysis for establishing cutoff points or thresholds for workplace bullying instruments (the NAQ-R in particular). This procedure has typically been applied in both clinical and research settings for assessing the discriminative ability of a screening tool. In workplace bullying, this ROC approach combines previous estimation methods: self-labelling question and negative acts exposure to establish cutoff scores that

allow differentiating bullying targets. As it has been pointed out by Notelaers & Einarsen (2013), stablishing cutoff scores may help practitioners to better assess the prevalence of bullying as well as the relative risk in certain working populations. Nowadays, ROC approach is spreading among scholars that aim to establish more accurate workplace bullying prevalence rates in their countries, such as Serbia (Petrović, Vukelić, & Čizmić, 2017), India (Gupta, Bakhshi, & Einarsen, 2017), Australia (Hutchinson, Bradbury, Browne, & Hurley (2017), or Denmark (Conway et al., 2018).

Nevertheless, no previous studies have adapted the short version of the Negative Acts Questionnaire (S-NAQ: Notelaers et al., 2018) and have established cutoff scores in Spain to date. Thus, the aim of this study is to examine the psychometric properties of the S-NAQ and provide cutoff scores in the Spanish context. In particular, this study focus on (a) analyzing the S-NAQ's internal consistency and reliability (Cronbach's alpha and omega coefficient); (b) examining its factor solution through exploratory and confirmatory factor analysis (EFA and CFA); and (c) exploring cutoff scores or thresholds with ROC curves analysis. Our findings may have interesting theoretical and practical implications for investigating/diagnosing bullying cases and/or designing tailored interventions according to the identified bullying groups.

Method

Participants

Data come from eight medium and large companies ($N = 1,409$ workers) settled in Andalusia (Spain) that voluntarily participated in psychosocial risk assessments (period 2016-2018). The participants were employed in the service sector (41.4%), healthcare sector (35.6%), or manufacturing sector (23%). As workplace bullying is a sensitive issue, limited socio-demographical data was gathered in some companies; therefore, only 740 participants reported their sex (38.5% women) and 673 their age (ranging between 21 and 68 years old; $M = 41.57$, $SD = 7.44$).

Instruments

Exposure to workplace bullying. This variable was measured with the short version of the Negative Acts Questionnaire-Revised (S-NAQ: Notelaers et al., 2018), which was translated into Spanish following a back-translation procedure (Muñiz, Elosua, & Hambleton, 2013). This questionnaire consists of nine specific negative behaviours measuring exposure to bullying within the last six months (see Table 2). Participants scored the frequency of each negative act according to the following response categories: 1-Never, 2-Rarely, 3-Monthly, 4-Weekly, and 5-Daily. The internal consistency measured with Cronbach's alpha was .85.

Self-reported workplace bullying. Participants read a global definition of bullying that have been used in previous studies conducted in Spanish (Leon-Perez et al., 2014) and indicated whether they had been bullied at work over the last six months according to such definition (1-Never, 2-Rarely, 3-Occasionnally or Monthly, 4-Weekly or Daily).

Psychological distress. This variable was measured with the available Spanish version of the General Health Questionnaire in its 12-item format (GHQ-12: Goldberg & Williams, 1988). This scale is a screening measure of psychological distress by asking whether the respondent has recently experienced a symptom or

behavior of psychological disturbance (e.g., "Have you recently… lost much sleep over worry?"). The scale points are described as follow: "less than usual", "no more than usual", "rather more than usual", "much more than usual". From the existing scoring methods, we used the so-called 'GHQ scoring method' (0-0-1-1) because is the method advocated by the test authors (Goldberg et al., 1997). The scale therefore gives a total score ranging from 0 to 12; a higher score indicates a greater degree of psychological distress. The internal consistency measured with Cronbach's alpha was .79.

Job satisfaction. This *ad hoc* variable was measured with a single Likert-scale item asking participants about their overall job satisfaction ("overall, how satisfied are you with your job?"). Response categories ranged from 1 ('very unsatisfied') to 5 ('very satisfied').

Sickness absences. Participants reported the number of days out of work due to sickness during the last year.

Procedure

Data collection was undertaken in the workplaces at working hours in presence of one assistant researcher. According to the American Psychological Association's (APA) Ethical Principles of Psychologists and Code of Conduct, participants were informed about the aim of the study and requisites for voluntary and confidential participation, and all participants gave their written informed consent. Participants placed their completed questionnaires in a sealed box to ensure the anonymity of their responses.

Data analysis

First, we established the psychometric properties and factorial structure of the S-NAQ. We calculated descriptive statistics and estimators for internal consistency. Also, univariate normality was tested using the Kolmogorov-Smirnov (K-S) test. Reliability of the S-NAQ was estimated using the Spearman-Brown formula (arranging the items by mean in order to compose the two halves). The sample was then randomly divided into two subsamples. Factor validity was assessed by EFA using one subsample and by CFA on the other one. Regarding the EFA, we used an unweighted least squares extraction (ULS) method because the hypothesis of multivariate normality was rejected: Mardia's (1970) estimates of multivariate kurtosis and skewness coefficients were high, 254.32 and 68.44, respectively (both ps < .001). Also, we used the K1 method to decide the number of factors to retain (Horn, 1965). In addition, we considered the following fit index to address the adjustment of data to the proposed models in the CFA (Chau, 1997; Schreiber, Nora, Stage, Barlow, & King, 2006): the chi square statistic to degrees of freedom ratio (χ^2/df), the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standard root mean square residual (SRMR).

Second, we used ROC curves to establish cutoff scores. In our case, we followed Notelaers and Einarsen (2013) recommendation of using two different cutoff scores that are able to detect both targets in a preliminary stage of bullying (i.e., lower threshold) and targets of severe bullying or victims (i.e., higher threshold). In doing so, we use two "gold standards" based on combinations of the self-labeling method (SLM) to measure workplace bullying and a measure of psychological distress (GHQ-12). Regarding the SLM, the lower threshold corresponds to responses "monthly or more often" (thus responses "never" and "rarely" were collapsed into 0 = no case, and responses "monthly", "weekly", and "daily" into 1 = case of bullying); whereas the higher threshold corresponds to responses "weekly or more often" (thus responses "monthly", are not considered a case of bullying). For the GHQ-12, Goldberg et al. (1997) suggest that the best cutoff score for a probable mental illness is 1/2 for the GHQ-scoring method. Thus, this cutoff score (i.e., scoring 2 or higher) is used in the present study as the lower threshold for the GHQ-12, which represents the percentile 75 in the Spanish general population (Rocha, Pérez, Sanz, Borrell, & Llandrich, 2011). In addition, we consider the score that represents the percentile 95 for the GHQ-12 higher threshold, which is scoring 7 or higher when using the GHQ-scoring method (i.e., cases in need of psychological treatment). In sum, the combination of both gold standards results in the following thresholds:

– Lower threshold (LT): Labelling oneself as being subjected to bullying behaviors at least "monthly", and scoring equal to or higher than 2 on the GHQ-12 following the GHQ-scoring method.
– Higher threshold (HT): Labelling oneself as being subjected to bullying behaviors at least "weekly", and scoring equal to or higher than 7 on the GHQ-12 following the GHQ-scoring method.

As we included two thresholds, we calculated two ROC curves (one for the lower threshold and another for the higher threshold). Each ROC curve offers a plot of the sensitivity (true positive rate) of an instrument versus its specificity (false-positive rate) for all possible cutoff scores. This area under the ROC curve (AUC) is an indicator of accuracy as it shows the ability of the instrument to discriminate between people with a certain condition (cases) and people without it. When the AUC is closer to 1 indicates high accuracy (Metz, 1978). As we gave the same priority to both the sensitivity and the specificity of the instrument, the "optimal" cutoff scores for the S-NAQ were determined by using the Youden index (Youden, 1950). Such index identifies the score where both Type I (false negatives) and Type II (false positives) errors reaches its minimum following the formula: $J = Sensitivity + (Specificity - 1)$. Last, we conducted an ANOVA to offer information about the predictive validity of the cutoff scores.

Finally, regarding data analysis software, the omega (ω) consistency estimator and its CI was computed using the MBESS R package version 3.3.2 (Kelley & Lai, 2012), and the CFA was carried out using the lavaan R package, version 0.6-3 (Rosseel, 2012). The rest of the statistical analyses were performed using SPSS 20, including macros for parallel analysis (O'Connor, 2000) and Mardia's multivariate analysis (DeCarlo, 1997).

Results

S-NAQ psychometric properties

The scores of the S-NAQ ranged from 9 to 45 ($M = 13.42$, $SD = 4.89$, $Mdn = 12.00$, $IQR = 4.00$) in a positively skewed distribution ($p < .001$). Table 1 displays descriptive statistics for the items. The nine items ranged from 1 to 5, and showed positive skewness and a leptokurtic distribution (ps < .001 in K-S test). Internal consistency

Table 1
S-NAQ Item Analysis

Item	M	SD	Skewness	Kurtosis	z_{K-S}	α_w
1	1.88	1.09	1.51	1.75	11.03	.834
2	1.59	0.85	2.04	4.97	11.76	.838
3	1.51	0.89	2.17	4.82	13.98	.817
4	1.27	0.59	2.87	10.88	17.00	.839
5	1.30	0.63	2.80	9.96	16.64	.831
6	1.46	0.69	1.78	4.19	14.02	.833
7	1.84	1.02	1.53	2.11	10.54	.826
8	1.44	0.85	2.32	5.44	15.32	.830
9	1.13	0.43	4.65	28.49	19.12	.847

Note: N = 1384. z_{K-S} = z score for Kolmogorov-Smirnov test. α_w = Change in Cronbach's alpha of the whole scale if the item is removed

and reliability estimators were acceptable (α = .849; ω = .860, 95% CI [.840, .880]; r_{xx} = .855, N = 1,384).

Regarding the factor structure, the KMO estimate (.887) and Bartlett's test of sphericity, $X^2(36)$ = 1649.17, $p < .001$, supported the adequacy of EFA with our data. Only one factor showed up for the S-NAQ, accounting for 35.58% of the total variance (see Figure 1). The item factor loadings ranged from .396 to .745, and communalities from .157 to .555 (see Table 2).

As mentioned above, the other subsample was used to confirm the factor structure in a more restrictive CFA. In order to deal with non-normality issues, we used a maximum likelihood with robust standard errors estimation method (MLR) and the Satorra-Bentler statistic. The fit indices for the CFA model can be seen in Table 3. The observed indices did not reach the recommended threshold values on the χ^2/df, the CFI, and the TLI. However, they did on the RMSEA and its 90% confidence interval, and the SRMR. Moreover, the model showed significant positive item factor loadings ($p < .001$), ranging from .553 to .794 (see Figure 2).
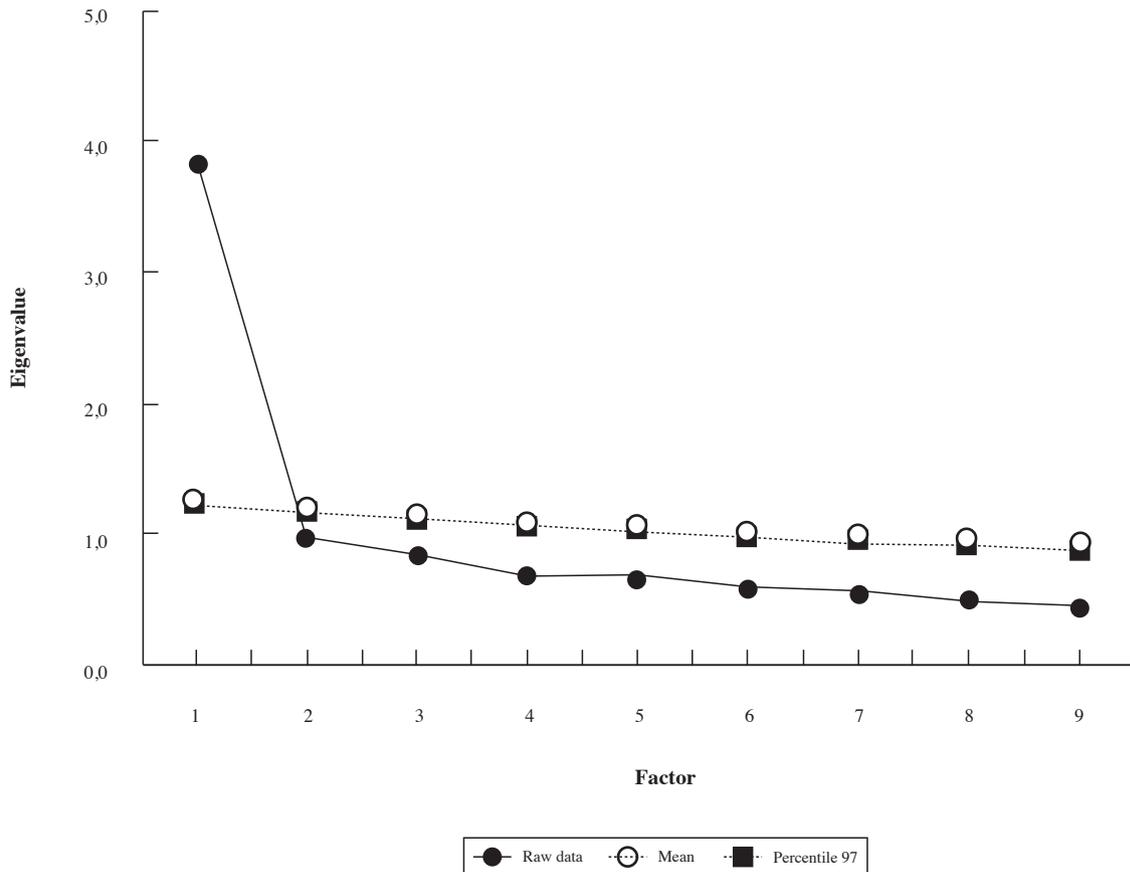

Figure 1. EFA scree plot. The plot displays empirical data eigenvalues, and mean and 95th percentile eigenvalues of 100 random samples in a parallel analysis. N = 692

Table 2
S-NAQ Results from Exploratory Factor Analysis

Item		Communality	Factor loading
1	Someone withholding information which affects your performance [Alguien le oculta información que afecta a su rendimiento]	.347	.589
2	Spreading gossip and rumours about you [Se difunden cotilleos y rumores sobre usted]	.292	.541
3	Being ignored or excluded [Ser ignorado, excluido, le "hacen el vacío"]	.555	.745
4	Having insulting or offensive remarks made about your person, attitudes or your private live [Ser insultado o recibir comentarios ofensivos sobre su persona (p.e. costumbres y orígenes), sus actitudes o su vida privada]	.287	.536
5	Being shouted at or being a target of spontaneous rage [Ser objeto de gritos y enfados espontáneos (o ataques de ira)]	.351	.593
6	Repeated reminders of your errors or mistakes [Se le recuerdan constantemente sus errores y equivocaciones]	.429	.655
7	Being ignored or facing a hostile reaction when you approach [Ser ignorado o percibir reacciones hostiles cuando se acerca]	.411	.641
8	Persistent criticism of your work and effort [Se le critica constantemente su trabajo y esfuerzo]	.373	.610
9	Practical jokes carried out by people you do not get along with [Ser objeto de bromas realizadas por personas con las que no te llevas bien]	.157	.396

Notes: ULS extraction method. Explained variance 35.58%.

Table 3
S-NAQ. Fit indices for CFA Model

		Value	RV
Fit index	χ^2/df	3.145	≤ 3.000
	RMSEA	.056	< .060 to .080
	RMSEA [90% CI]	[.047, .064]	< .060 to .080
	SRMR	.046	≤ .080
	CFI	.939	≥ .950
	TLI	.919	≥ .950

Notes: *n* = 692. RV: Recommended values (Chau, 1997; Schreiber et al., 2006)

S-NAQ cutoff scores

Two ROC functions were calculated based on the two thresholds established for the S-NAQ to determine from what score someone is occasionally or severely bullied. First, results revealed that t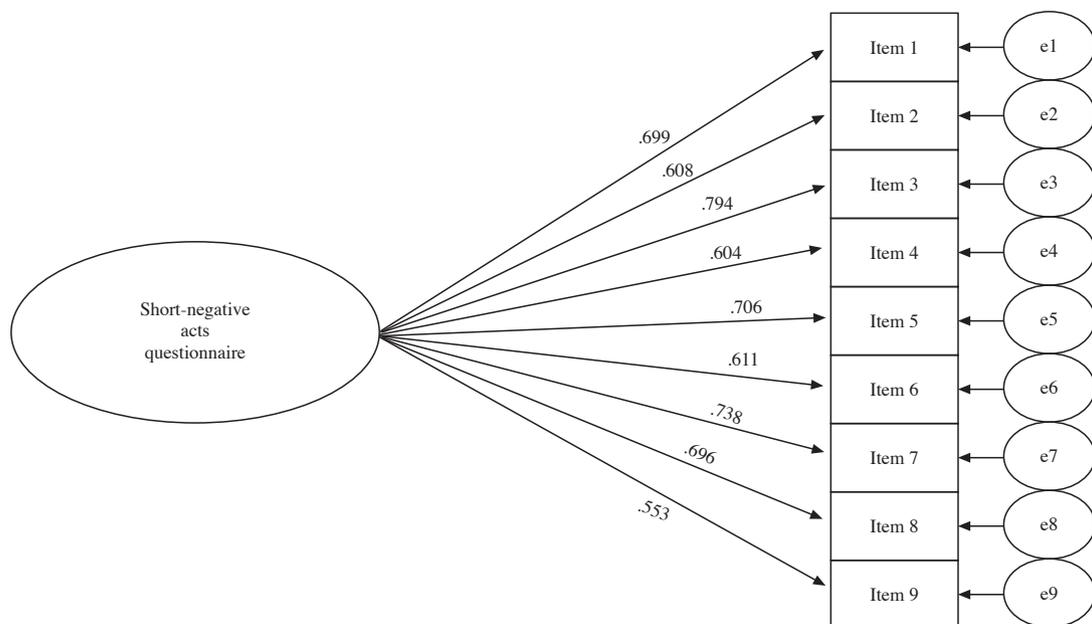he Area Under the Curve (AUC) for the lower threshold cutoff score (i.e., we used a gold standard that combines the responses to the definition of workplace bullying "monthly or more often" with a GHQ-12 score ≥ 2) was .944 (95% CI = .928, .960; *p* < .001); whereas the AUC for the higher threshold cutoff score (i.e., we used a gold standard that combines the responses to the definition of workplace bullying "weekly or more often" with a GHQ-12 score ≥ 7) was .975 (95% CI = .952, .997; *p* < .001). Both AUC values indicate at least 94% probability of making a correct diagnosis.

Next, the Youden index is used to determine the best better tradeoff between true negatives (NP) and true positives (TP) in both thresholds considering the raw sum score in the S-NAQ. Thus, the optimal cutoff scores are 15 for the lower threshold and 23 for the higher threshold (see Table 4). As can be seen in Table 5, using a sum score of 15 as an optimal cutoff point for a lower threshold means that 16.4% of the current sample was occasionally bullied and these workers are at risk of suffering workplace bullying. In addition, the cutoff score of 23, which is the higher threshold estimate, means that 7.1% of the current sample can be labelled

Figure 2. CFA for the unifactorial model. Standardized coefficients. All factor loadings are statistically significant (*p*s < .001)
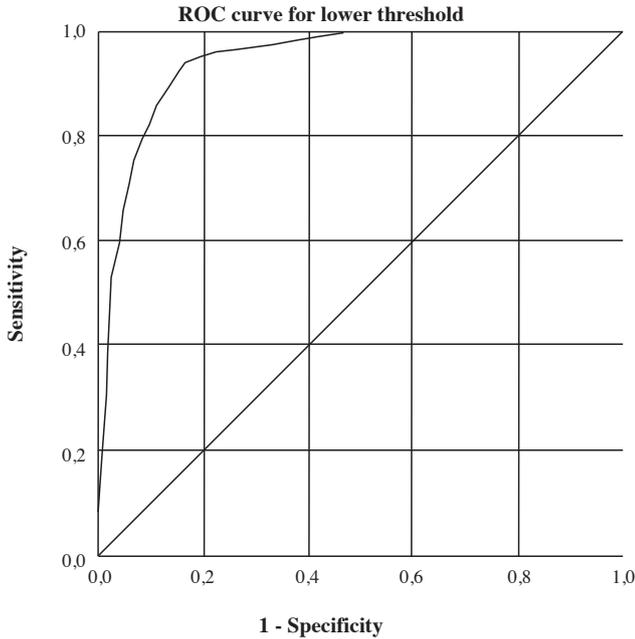
ROC curve for lower threshold

Figure 3. ROC curve for the lower threshold
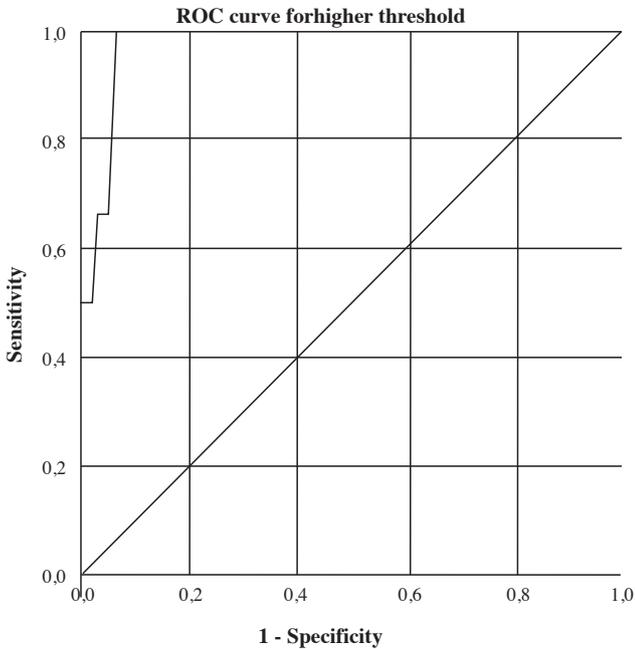
ROC curve forhigher threshold

Figure 4. ROC curve for the higher threshold

as targets or victims of workplace bullying. Therefore, results revealed that 76.3% can be labelled as non-exposed workers. For the lower threshold the chances to classify a respondent as being occasionally bullied when the respondent is not bullied is approximately 15%. Yet, the chance to miss a respondent that is at least occasionally bullied is close to zero. Regarding the higher threshold both erroneous classification decrease. The chance to label a respondent as a victim of bullying using the cut-off when the respondent is not a victim of bullying is only 6,6 % whereas the chance to not identify a victim with the cut-off score of 23 is zero.

Finally, a one-way ANOVA was conducted to compare self-reported job satisfaction and sickness absences in each of these three bullying categories. There were significant differences on both job satisfaction $[F(2,1373) = 195.61, p = .01]$ and sickness absences $[F(2,1342) = 13.48, p = .01]$ for the three groups. Post hoc comparisons using the Tukey HSD test revealed that employees in the non-exposed group reported higher job satisfaction ($M = 3.79; SD = .82$) than their colleagues in both the risk-of-bullying group ($M = 2.92; SD = 1.10$) and in the victims group ($M = 2.27; SD = 1.11$). Also, employees in the non-exposed group ($M = 5.15; SD = 24.04$) and the risk-of-bullying group ($M = 8.27; SD = 23.68$)

	Raw score	Sensitivity	Specificity	Youden index
Table 4 S-NAQ Cutoff Scores				
Lower	–	–	–	–
	11	1.000	0.305	0.305
	12	1.000	0.500	0.500
	13	0.977	0.651	0.628
	14	0.961	0.768	0.730
Threshold	**15**	**0.938**	**0.838**	**0.776**
	16	0.884	0.873	0.757
	17	0.853	0.895	0.748
	18	0.806	0.911	0.717
	19	0.767	0.927	0.694
	–	–	–	v
Higher	–	–	–	–
	19	1.000	0.865	0.865
	20	1.000	0.883	0.883
	21	1.000	0.901	0.901
	22	1.000	0.917	0.917
Threshold	**23**	**1.000**	**0.933**	**0.933**
	24	0.667	0.945	0.612
	25	0.667	0.958	0.624
	26	0.667	0.966	0.633
	27	0.500	0.977	0.477
	–	–	–	–

		Non-exposed	At risk	Targets	Total
Table 5 Final percentages of true and false scores with the distinct cutoff scores					
		Sum < 15	**>= 15 Sum <23**	**Sum >= 23**	
Lower threshold	Self-labelling less than occasionally and GHQ-12 < 2	TN 76.0%	FP 12.5%	FP 2.1%	90.6%
	Self-labelling occasionally or more and GHQ-12 >= 2	FN 0.6%	TP 3.9%	TP 5.0%	9.4%
Higher threshold	Self-labelling less than weekly and GHQ-12 < 7	TN 76.5%	TN 16.4%	FP 6.6%	99.6%
	Self-labelling weekly or more and GHQ-12 >= 7			TP 0.4%	0.4%
Total		76.5%	16.4%	7.1%	100.0%

did not significantly differ on sickness absences. However, both groups reported significantly less sickness absences than their colleagues in the victims group (M = 19.85; SD = 51.04).

Discussion

Workplace bullying is a growing topic of interest because it has detrimental consequences for employee health and well-being. However, there is a lack of reliable instruments to measure bullying at work and differentiating victims versus non-victims. In response, the main goal of this study is establishing cutoff scores for diagnosing workplace bullying based on participants' responses to a 9-item short version of the Negative Acts Questionnaires (S-NAQ).

We used the S-NAQ because it is a short instrument that captures the core of bullying at work into a single factor: humiliating negative acts regarding the tasks a person needs to accomplish at work, unwanted negative acts concerning someone's private life, and social isolation at work (Notelaers et al., 2018). Also, the S-NAQ does not refer explicitly to the term mobbing or bullying, and therefore overcome potential responses biases due to sociocultural factors (Giorgi, Leon-Perez, & Arenas, 2015). However, as this instrument has not previously adapted to Spanish, we examined its psychometric properties. Our results showed that the S-NAQ presents an excellent internal consistency and reliability. Moreover, both EFA and CFA models showed an adequate fit to a single factor model, matching the unifactorial structure found by Notelaers et al. (2018).

Regarding cutoff scores or thresholds for bullying diagnosing, a ROC curve analysis was conducted with two gold standards over the total sum score of the S-NAQ. The first gold standard (SLM) captures the subjectivity of the bullying phenomenon. In line with stress models, it seems necessary that the target of bullying behaviors interpret them as harmful and directed to him/her (Leon-Perez et al., 2014). In a similar vein, the second gold standard (GHQ) focuses on the negative consequences of workplace bullying on employee health and well-being. Thus, we opted for the GHQ-12 because it is a well-known screening instrument for transitory experiences of psychological distress that can result in long-term psychiatric disorders.

The combination of gold standards into ROC curves over the S-NAQ indicated two cutoff scores: 15 and 23, which means that workers scoring less than 15 in the S-NAQ can be considered non-exposed to bullying at work; those scoring between 15 and 22 are at high risk of being bullying victims or may be immerse in a bullying process; whereas those scoring 23 or higher can be considered targets of workplace bullying. In addition, when examining both the AUC of the ROC curves and the ANOVA on job satisfaction and sickness absences, results revealed that the S-NAQ is an accurate instrument to discriminate between bullying victims versus non-victims.

This study has also some limitations associated to the design and sample. Perhaps the main limitation is that our sample is not representative of the Spanish working population. Therefore, future studies should replicate our results in other samples or use a representative sample of the general working population in other to obtain more accurate cutoff scores and normative data. In this regard, future research may consider replicating our results by examining bullying cutoff scores with different gold standards that include either other questionnaires for assessing psychological well-being or more objective instruments rather than only self-

report measures (e.g., clinical interviews instead the GHQ-12 or peer nominations instead the self-labelling question). Finally, future studies should analyze sex differential item functioning in the perception of bullying at work (Gómez-Benito, Sireci, Padilla, Hidalgo, & Benítez, 2018).

Despite the limitations inherent to the study design, our findings have interesting implications for theory and practice. Regarding theory, this ROC approach overcomes previous flaws in assessing workplace bullying, such as: (a) providing cutoff scores based on statistical criteria and well-established gold standards in the literature rather than more or less arbitrary criteria previously adopted by researchers (Argevold, 2007); and (b) using methodological tools that are user-friendly and provide clear cutoff scores rather than advanced techniques with results difficult to interpret and therefore losing practical applicability (Leon-Perez et al., 2014; Notelaers et al., 2006). Regarding practice, the main implication revolves around the use of the cutoff scores in occupational health and safety (OHS) for both preventing workplace bullying and improving employees' health and well-being. In that sense, having three groups fits with the preventive instruments that adopt a "traffic light" model in which red means elevated risk of bullying (in our case workers scoring equal or higher than 23), yellow means moderated risk of bullying (in our case workers scoring between 15 and 22), and green is the most favourable condition as it means having a low risk of bullying at work (in our case workers scoring lower than 15). According to this classification based on your score in the S-NAQ, different intervention measures should be taken into account or implemented. For example, for those in the low-risk condition, organizations should emphasize risk prevention measures and health promotion initiatives. The second condition implies early detection measures and further investigation of the case as it should be managed only after having a clearer picture of what is going on. In doing so, although showing good psychometric properties, the S-NAQ is a self-report measure that should be complemented with other information sources (e.g., HR records on performance, absenteeism, sick leaves; qualitative interviews with key people in the organization). In other words, for employees in this condition, organizations may conduct an investigation that should end up in a formal conflict management procedure and/or in changing some working conditions and procedures. Finally, those already targets of workplace bullying may have their health compromised to some extent; therefore, workers in the third condition or bullying targets may benefit from secondary and tertiary measures that try to restore the situation and prevent further health impairment.

In conclusion, workplace bullying is considered a social stressor that endangers health impairment and productivity losses. Thus, this study provides simple cutoff scores for establishing an organizational diagnosis with only 9 items that may offers an initial picture of the potential bullying cases in a certain workplace, which can be very useful to take further actions based on empirical rather than theoretical criteria. Moreover, this measure and its associated cutoff scores can be used to establish the bullying prevalence rate in more general populations.

Acknowledgements

References

Agervold, M. (2007). Bullying at work: A discussion of definitions and prevalence, based on an empirical study. *Scandinavian Journal of Psychology*, 48(2), 161-172. doi:10.1111/j.1467-9450.2007.00585.x

Bjørkelo, B., Einarsen, S., Nielsen, M. B., & Matthiesen, S. B. (2011). Silence is golden? Characteristics and experiences of self-reported whistleblowers. *European Journal of Work and Organizational Psychology, 20*(2), 206-238. doi:10.1080/13594320903338884

Chau, P. Y. (1997). Reexamining a model for evaluating information center success using a structural equation modeling approach. *Decision Sciences, 28*, 309-334. doi:10.1111/j.1540-5915.1997.tb01313.x

Conway, P. M., Høgh, A., Nabe-Nielsen, K., Grynderup, M. B., Mikkelsen, E. G., Persson, R., ... & Hansen, Å. M. (2018). Optimal cut-off points for the Short-Negative Act Questionnaire and their association with depressive symptoms and diagnosis of depression. *Annals of Work Exposures and Health, 62*(3), 281-294. doi:10.1093/annweh/wxx105

DeCarlo, L. T. (1997). Mardia's multivariate skew (b1p) and multivariate kurtosis (b2p) [SPSS macro]. Retrieved from http://www.columbia.edu/~ld208/

Einarsen, S., & Nielsen, M. B. (2015). Workplace bullying as an antecedent of mental health problems: a five-year prospective and representative study. *International Archives of Occupational and Environmental Health, 88*(2), 131-142. doi:10.1007/s00420-014-0944-7

Einarsen, S., Hoel, H., & Notelaers, G. (2009). Measuring exposure to bullying and harassment at work: Validity, factor structure and psychometric properties of the Negative Acts Questionnaire-Revised. *Work & Stress, 23*(1), 24-44. doi:10.1080/02678370902815673

Einarsen, S., Hoel, H., Zapf, D., & Cooper, C. L. (2011). The concept of bullying and harassment at work: the European tradition. In S. Einarsen, H. Hoel, D. Zapf, & C. L. Cooper (Eds.), *Bullying and harassment in the workplace: Developments in theory, research, and practice* (pp. 3-40). London: CRC Press.

Eriksen, T. L. M., Hogh, A., & Hansen, Å. M. (2016). Long-term consequences of workplace bullying on sickness absence. *Labour Economics, 43*, 129-150. doi:10.1016/j.labeco.2016.06.008

Escartín, J., Rodríguez-Carballeira, Á., Gómez-Benito, J., & Zapf, D. (2010). Development and validation of the workplace bullying scale EAPA-T. *International Journal of Clinical and Health Psychology, 10*(3), 519-539.

Eurofound (2015). *European Working Conditions Survey*. Data available at: https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys

Giorgi, G., Leon-Perez, J. M., & Arenas, A. (2015). Are bullying behaviors tolerated in some cultures? Evidence for a curvilinear relationship between workplace bullying and job satisfaction among Italian workers. *Journal of Business Ethics, 131*(1), 227-237. doi: 10.1007/s10551-014-2266-9

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*(1), 191-197. doi:10.1017/S0033291796004242

Goldberg, D., & Williams, P. (1988). *General health questionnaire (GHQ)*. Swindon, Wiltshire, UK: nferNelson.

Gómez-Benito, J., Sireci, S., Padilla, J. L., Hidalgo, M. D., & Benítez, I. (2018). Differential Item Functioning: Beyond validity evidence based on internal structure. *Psicothema, 30*, 104-109.

Gupta, R., Bakhshi, A., & Einarsen, S. (2017). Investigating workplace bullying in India: psychometric properties, validity, and cutoff scores of negative acts questionnaire–revised. *Sage Open, 7*(2), 1-12. doi:10.1177/2158244017715674.

Hogh, A., Clausen, T., Bickmann, L., Hansen, A. M., Conway, P. M., & Baernholdt, M. (2019). Consequences of workplace bullying for individuals, organizations and society. In P. D'Cruz, E. Noronha, G. Notelaers, & C. Rayner (Eds.), *Concepts, Approaches and Methods. Handbooks of Workplace Bullying, Emotional Abuse and Harassment*. Singapore: Springer.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. doi:10.1007/BF02289447

Hutchinson, M., Bradbury, J., Browne, G., & Hurley, J. (2017). Determining the optimal cut-off scores for the Workplace Bullying Inventory. *Nurse Researcher, 25*(3), 46-50. doi:10.7748/nr.2017.e1543

Kelley, K., & Lai, K. (2012). MBESS: MBESS. R package version 3.3.2. [Computer software]. Retrieved from http://CRAN.R-project.org/package=MBESS

Leon-Perez, J. M., Escartín, J., & Giorgi, G. (2019). The presence of workplace bullying and harassment worldwide. In P. D'Cruz, E. Noronha, G. Notelaers, & C. Rayner (Eds.), *Concepts, Approaches and Methods. Handbooks of Workplace Bullying, Emotional Abuse and Harassment*. Singapore: Springer.

Leon-Perez, J. M., Notelaers, G., Arenas, A., Munduate, L., & Medina, F. J. (2014). Identifying victims of workplace bullying by integrating traditional estimation approaches into a latent class cluster model. *Journal of Interpersonal Violence, 29*(7), 1155-1177. doi:10.1177/0886260513506280

Magee, C., Gordon, R., Robinson, L., Caputi, P., & Oades, L. (2017). Workplace bullying and absenteeism: The mediating roles of poor health and work engagement. *Human Resource Management Journal, 27*(3), 319-334. doi:10.1111/1748-8583.12156

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530. doi:10.1093/biomet/57.3.519

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8*(4), 283-298. doi:10.1016/S0001-2998(78)80014-2

Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). International Test Commission guidelines for test translation and adaptation. *Psicothema, 25*(2), 151-157. doi: 10.7334/psicothema2013.24

Nielsen, M. B., Matthiesen, S. B., & Einarsen, S. (2010). The impact of methodological moderators on prevalence rates of workplace bullying. A meta-analysis. *Journal of Occupational and Organizational Psychology, 83*(4), 955-979. doi:10.1348/096317909X481256

Nielsen, M. B., Notelaers, G., & Einarsen, S. (2011). Measuring exposure to workplace bullying. In S. Einarsen, H. Hoel, D. Zapf, & C. L. Cooper (Eds.), *Bullying and harassment in the workplace: Developments in theory, research, and practice* (pp. 149-174). London: CRC Press.

Nielsen, M. B., Skogstad, A., Matthiesen, S. B., Glasø, L., Aasland, M. S., Notelaers, G., & Einarsen, S. (2009). Prevalence of workplace bullying in Norway: Comparisons across time and estimation methods. *European Journal of Work and Organizational Psychology, 18*(1), 81-101. doi:10.1080/13594320801969707

Notelaers, G., & Einarsen, S. (2013). The world turns at 33 and 45: Defining simple cutoff scores for the Negative Acts Questionnaire–Revised in a representative sample. *European Journal of Work and Organizational Psychology, 22*(6), 670-682. doi:10.1080/1359432X.2012.690558

Notelaers, G., Einarsen, S., De Witte, H., & Vermunt, J. K. (2006). Measuring exposure to bullying at work: The validity and advantages of the latent class cluster approach. *Work & Stress, 20*(4), 289-302. doi:10.1080/02678370601071594

Notelaers, G., Van der Heijden, B., Hoel, H., & Einarsen, S. (2018). Measuring bullying at work with the short-negative acts questionnaire: identification of targets and criterion validity, *Work & Stress, 33*(1), 58-75. doi:10.1080/02678373.2018.1457736.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-402. doi:10.3758/BF03200807

Park, H., Bjørkelo, B., & Blenkinsopp, J. (2018). External Whistleblowers' Experiences of Workplace Bullying by Superiors and Colleagues. *Journal of Business Ethics*. Advance online publication. doi:10.1007/s10551-018-3936-9.

Petrović, I. B., Vukelić, M., & Čizmić, S. (2017). Rocking at 81 and rolling at 34: ROC cut-off scores for the negative acts questionnaire–revised in Serbia. *Frontiers in Psychology, 7*, 2058-2070. doi:10.3389/fpsyg.2016.02058

Rocha, K. B., Pérez, K., Sanz, M. R., Borrell, C., & Llandrich, J. O. (2011). Propiedades psicométricas y valores normativos del General Health Questionnaire (GHQ-12) en población general española [Psychometric properties and normative values of the General Health Questionnaire (GHQ-12) in Spanish population]. *International Journal of Clinical and Health Psychology, 11*(1), 125-139.

Rodríguez-Muñoz, A., Antino, M., & Sanz-Vergel, A. I. (2017). Cross-domain consequences of workplace bullying: A multi-source daily diary study. *Work & Stress*, *31*(3), 297-314. doi:10.1080/02678373.2017.1330782

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1-36. doi:10.18637/jss.v048.i02

Samnani, A. K. (2013). The early stages of workplace bullying and how it becomes prolonged: The role of culture in predicting target responses. *Journal of Business Ethics*, *113*(1), 119-132. doi:10.1007/s10551-012-1286-6

Schreiber, J. B. Nora, A., Stage, F. K., Barlow, E. A., & King. J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research, 99*, 323-338. doi:10.3200/JOER.99.6.323-338

Trépanier, S. G., Fernet, C., & Austin, S. (2015). A longitudinal investigation of workplace bullying, basic need satisfaction, and employee functioning. *Journal of Occupational Health Psychology*, *20*(1), 105-116. doi:10.1037/a0037726

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32-35.

Zapf, D., Escartín, J., Einarsen, S., Hoel, H., & Vartia, M. (2011). Empirical findings on prevalence and risk groups of bullying in the workplace. In S. Einarsen, H. Hoel, D. Zapf, & C. L. Cooper (Eds.), *Bullying and harassment in the workplace: Developments in theory, research, and practice* (pp. 75-106). London: CRC Press.