



INSTRUMENTS



Perceived danger during deployment: a Rasch validation of an instrument assessing perceived combat exposure and the witnessing of combat consequences in a war zone

Karen-Inge Karstoft^{a*}, Tine Nielsen^{b*} and Anni B. S. Nielsen^{a,c}

^aResearch and Knowledge Centre, the Danish Veteran Centre, Ringsted, Denmark; ^bDepartment of Psychology, University of Copenhagen, Copenhagen, Denmark; ^cThe Research Unit and Section of General Practice, Institute of Public Health, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

The potential stressors associated with military deployment are related to an increased risk of adverse mental health outcomes. Perceived exposure to combat has been found to be proportional to the severity of post-deployment posttraumatic stress disorder (PTSD). However, other perceived adversities during deployment, such as witnessing danger, distress, and hardship in the war zone, have been less systematically studied, but might play an equally substantial role for post-deployment mental health. The development and validation of scales that assess these related constructs are needed to distinguish their contribution to post-deployment risk of PTSD. We evaluated the validity of 10 items measuring perceived danger distributed to all deployed personnel with the Danish Defense since 1998. We hypothesize two scales: Exposure to Danger and Combat (EDC) and Witnessing Consequences of War (WCW). Two military cohorts deployed to Afghanistan in 2009 (Cohort 1, $N = 276$) and 2013 (Cohort 2, $N = 273$) were included. Questionnaire data was collected six months after homecoming, including deployment experiences and post-deployment reactions. We tested the construct validity of the 10 items of perceived danger with Rasch models (RM), focusing specifically on presence of subscales, and differential item functioning (DIF) across cohorts. We confirmed the existence of two separate subscales, EDCS and WCWS, both with adequate reliability. None of the subscales fitted a pure RM, but adequate fit was found for graphical log-linear RMs with evidence of DIF for the ECDS. However, adjusting the score to account for DIF had practically no effect, suggesting that the total non-adjusted mean score can be used in future cohort comparisons. Perceived exposure to combat and danger and witnessing consequences of war are related, but essentially distinct, concepts, each providing unique information about deployment adversities. Future studies should evaluate their shared and unique contribution to the risk of post-deployment PTSD.

Los estresores potenciales asociados al despliegue militar se relacionan con un mayor riesgo de presentar eventos adversos en salud mental

La percepción de exposición al combate ha resultado ser proporcional a la severidad del trastorno por estrés post-traumático (TEPT) posterior al despliegue militar. Sin embargo, otras adversidades percibidas durante el despliegue, tales como experimentar peligro, angustia, y dificultades en la zona de guerra, han sido menos sistemáticamente estudiadas, pero podrían jugar un rol sustancialmente equivalente para el impacto en la salud mental posterior al despliegue militar. Se requiere desarrollar y validar escalas que evalúen estos constructos relacionados para distinguir su contribución al riesgo de desarrollar TEPT posterior al despliegue. El objetivo es evaluar la validez de 10 ítems que miden el peligro percibido distribuido a todo el personal desplegado con la Defensa Danesa desde 1998. Se hipotetizan dos escalas; Exposición a Peligro y Combate (EDC) y Exposición a Consecuencias de Guerra (WCW). Se incluyó a dos cohortes militares, que se desplegaron en Afganistán el año 2009 (Cohorte 1, $N=276$), y el 2013 (Cohorte 2, $N=273$). Se recopiló datos a partir de cuestionarios seis meses después de regreso al país de origen, incluyendo experiencias dentro del despliegue y reacciones posteriores a éste. Se probó en las cohortes la validez de constructo de los 10 ítems de peligro percibido con modelos Rasch (RM), poniendo especial énfasis en la presencia de subescalas, e ítem de funcionamiento diferencial (DIF). Se confirmó la existencia de dos subescalas separadas, EDC y WCW, ambas con fiabilidad adecuada. Ninguna de las subescalas correspondía a un modelo Rasch puro, pero sí se encontró una relación adecuada para un modelo Rasch logarítmico-lineal con evidencia de DIF para EDC. No obstante, el ajustar el puntaje para explicar el IFD prácticamente no tuvo efectos,

ARTICLE HISTORY

Received 12 October 2017
Accepted 16 May 2018

KEYWORDS

Item response theory (IRT); Rasch model (RM); combat stress exposure; military personnel; psychometrics; military; post traumatic stress disorder (PTSD)

PALABRAS CLAVES

Teoría de la respuesta al ítem; Modelo Rasch; Exposición al estrés de combate; Personal militar; Psicometría; Militar; TEPT

关键词

项目反应理论; Rasch模型; 战争应激暴露; 军事人员; 心理测量学; 军事; PTSD

HIGHLIGHTS

- Using Rasch models we evaluate a measure of perceived danger distributed to Danish military personnel after deployment hypothesizing that it contains two subscales: perceived Exposure to Danger and Combat (EDCS) and witnessing consequences of war (WCWS).
- We find support for the two subscales both with adequate reliability but with evidence for cohort differential item functioning for the ECDS.
- Future studies should investigate these constructs in relation to risk of post-deployment mental health problems.

CONTACT Tine Nielsen ✉ Tine.Nielsen@psy.ku.dk 📍 Department of Psychology, University of Copenhagen, Øster Farimagsgade 2a, 1353 Copenhagen, Denmark

*Shared first authorship

📎 Supplemental data for this article can be accessed [here](#).

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

sugiriendo que la puntuación media total no ajustada puede ser utilizada en futuras comparaciones de cohortes. La exposición percibida al combate y el peligro y exposición a consecuencias de guerra son dos conceptos relacionados, pero esencialmente distintos, cada uno proveyendo información única sobre los efectos adversos del despliegue militar. Futuros estudios deberían evaluar su contribución compartida y única al riesgo de desarrollar TEPT posterior al despliegue militar.

入伍中的感知危险：对一个评估在战区感知到战争暴露以及目睹战争后果的工具进行Rasch验证

与入伍有关的潜在压力源也和不良心理健康结果风险增加有关。主观战争暴露已被发现与入伍后的创伤后应激障碍（PTSD）的严重程度成正比。然而，其他入伍期间遇到的负性环境（比如目击危险、苦恼和在战区的艰难处境）却较少被系统地研究，但这些可能对入伍后的心理健康发挥同样重要的作用。需要开发和验证评估这些相关结构的量表来区分他们对入伍后PTSD风险的影响。我们的目标是评估10个测量主观危险的题目的有效性，题目被分发给自1998年以来丹麦国防部所有入伍人员。我们假设有两个量表：暴露于危险和战争（EDC）；和目击战争的后果（WCW）。被试包括两个分别在2009年（队列1， $N = 276$ ）和2013年（队列2， $N = 273$ ）派往阿富汗的军事部队。问卷调查数据收集在士兵复员后的六个月，内容包括入伍经历和入伍后反应。我们使用Rasch模型（RM）验证了10个关于感知危险的题目的结构有效性，特别侧重验证分量表的存在以及跨队列的项目功能差异（DIF）。结果支持存在两个独立的分量表，EDCS和WCWS，都具有足够的信度。分量表都不符合纯RM模型，但是ECDS中发现图形对数线性（graphical log-linear）RM模型有足够的拟合度，并具有DIF。然而，调整分数对DIF几乎没有实际影响，说明未调整的总平均分未来可用于队列比较。感知到暴露于战争和危险以及见证战争后果是两个相关的但本质上不同的概念。每个概念都提供关于军队负性环境的独特信息。未来的研究应评估他们对入伍后PTSD风险的共同和独特贡献。

1. Introduction

Deployment to war zones entails a risk of exposure to threatening situations, such as direct combat and siege (Osório et al., 2017). Studies have shown that the intensity and frequency of combat exposure are often proportional to post-deployment severity of posttraumatic stress disorder (PTSD) (Fear et al., 2010; Xue et al., 2015). However, life in a war zone often includes stressful experiences not related directly to combat and not posing a direct threat, e.g. exposure to the adverse effects of battle such as witnessing distress among the locals and assaults on civilians (King, King, Vogt, Knight, & Samper, 2006; Polusny et al., 2014; Polusny et al., 2011). Such stressful experiences during deployment have also been found to be related to an increase in PTSD symptoms from before to after deployment (Osório et al., 2017; Polusny et al., 2011; Vasterling et al., 2010). As such, not only combat exposure but also exposure to the adverse effects of battle should be taken into account when considering war zone stress of importance for post-deployment mental health.

Given the relationships between combat exposure and deployment stress to PTSD, studies investigating these factors in a comparable way across cohorts and missions can reveal the degree to which deployment stressors account for differences in the risk of post-deployment PTSD. The last decades have seen a plethora of research on the toll of recent wars in Iraq and Afghanistan (Andersen, Karstoft, Bertelsen, & Madsen, 2014; Carlson et al., 2010; McDevitt-Murphy et al.,

2010; Ramchand et al., 2010; Schnurr et al., 2010; Shea, Vujanovic, Mansfield, Sevin, & Liu, 2010). Many studies have aimed at estimating the prevalence of PTSD following deployment and have reported greatly varying rates. In general, higher PTSD prevalences have been found for the war in Iraq (9.8–19.9%) than for the war in Afghanistan (Hoge, Auchterlonie, & Milliken, 2006; Hoge et al., 2004; Seal, Bertenthal, Miner, Sen, & Marmar, 2007). These differences might, to some degree, reflect different methodological choices such as time of assessment, instruments and case definition, but they might also reflect actual differences in the toll of wars and missions (Sundin et al., 2014).

From 1992 through 2016, approximately 32,000 Danish soldiers have been deployed to international missions with the Danish Defense, primarily to the Balkans, Iraq and Afghanistan. In 1998, the Danish Military Psychology Department commenced an ongoing data collection for which questionnaires were distributed to all deployed soldiers approximately six months after home coming. With an average response rate of 65%, this database currently contains data on approximately 21,000 individuals. The questionnaire, called *Psychological Reactions to International Missions* (PRIM), contains questions on deployment experiences as well as post-deployment reactions, with 18 items pertaining to perceived combat stress exposure (CSE). Throughout the data collection period, these 18 items have been used to assess CSE for screening purposes. However, when scrutinizing these 18 items (Table S1), it is clear that some are

Table 1. The 10-item perceived Danger-Injury scale.

During the deployment, did you experience:
1. Being threatened with a weapon?*
2. Being shot at?*
3. Being in areas with roadside bombs or mines?*
4. Passing areas with combat activities?*
5. Aggressive behaviour from the locals?*
6. Witnessing distress among the locals?^
7. Seeing dead people?^
8. Seeing wounded people?^
9. Being witness to assaults on civilians?^
10. Insufficient reinforcement or relief of your unit?*

Items marked * are hypothesized to belong to the Exposure to Combat and Danger subscale (ECDS); items marked ^ are hypothesized to belong to the Witnessing Consequences of War subscale (WCWS). All items had four response categories (1 = never, 2 = seldom, 3 = fairly often, 4 = almost daily).

semantically overlapping, inaccurately phrased or would have only extremely limited occurrence.

More recently, 10 items from the CSE scale have been selected to provide a more condensed measure of perceived danger during deployment (Berntsen et al., 2012) labelled the Danger-Injury (DI) scale. Importantly, this 10-item scale (Table 1) contains items that are directly related to perceived exposure to combat and danger as well as items that are related to witnessing the aftermath of battle such as injured or dead people and distress among local populations. The items directly related to exposure are to some extent similar to those in the widely used combat exposure scale (Keane et al., 1989) measuring only the dimension ‘combat experiences’. However, the distinction in DI between perceived exposure to combat and witnessing the consequences of war is more in line with another well-known self-reported combat exposure scale namely the Deployment Risk and Resilience Inventory (DRRI-2), with a distinction between combat experiences (perceived threat, perceived exposure) and post-battle experiences (King, King, Knight, & Samper, 2006; Vogt et al., 2013). Perceived DI was expressed as a total sum score, and thus based on an assumption of unidimensionality. Nevertheless, this assumption of unidimensionality has never been tested or published and it is unclear whether the scale is indeed unidimensional. Here, on the basis of examining each item, we hypothesize that, just like the DRRI-2, the DI consists of two separate scales: *exposure to danger and combat* and *witnessing the consequences of war*, respectively. Further, it is not known whether some items in the 10-item scale function differentially for subgroups of soldiers and may result in biased measurement for some groups. Thus, testing for differential item functioning (DIF) is indeed also relevant.

Recently, scales assessing symptoms of PTSD (Karstoft, Andersen, & Nielsen, 2017) and depression

(Karstoft, Nielsen, & Nielsen, 2017) from PRIM have been validated to enable the comparison of symptom burden across missions and cohorts. Hence, a thorough validation of a perceived danger scale is of great importance and will enable not only the comparison of stressful experiences during deployment across cohorts and missions, but also of how such experiences relate to post-deployment symptoms of PTSD and depression.

2. Objective

In this study, we aim to evaluate the 10 items from the perceived combat stress exposure scale which make up the perceived DI scale reported by Berntsen and colleagues (2012). We do so with the hypothesis that the measure contains two construct-wisely distinct subscales. Hence, we investigate the construct validity of the two proposed subscales, one addressing the degree of perceived exposure to danger and combat (EDC) and another addressing witnessing the consequences of war (WCW) and test the assumption of unidimensionality across these. Furthermore, we also test whether some items suffer from DIF.

3. Methods

3.1. Participants

We included military personnel from two different deployment cohorts: Cohort 1 deployed to Afghanistan in 2009 as part of the International Security Assistance Force (ISAF), and Cohort 2 also deployed to Afghanistan as part of ISAF but in 2013. These specific cohorts were selected since they, in parallel with PRIM, also provided data on validated questionnaires that can be used for evaluation of criterion validity and the cohorts have previously been used in other validity studies (Karstoft, Nielsen, et al., 2017; Karstoft, Andersen, et al., 2017). Furthermore, while both cohorts deployed to Afghanistan, they did so at different times with very different threat assessments and combat intensity. Hence, they are appropriate for the investigation of differences in perceived danger. We included everyone who, in addition to the central items of perceived danger, also had complete data on relevant exogenous variables: gender, rank and previous deployments. Based on these criteria we excluded 58 persons with incomplete data. The total *N* for the entire sample was then reduced from 607 to 549 (Cohort 1 = 276; Cohort 2 = 273). All participants were unique, i.e. no overlap in persons.

3.2. Instruments

3.2.1. Items assessing perceived exposure to danger and combat and witnessing the consequences of war

A list of the original 18 items of the perceived CSE scale from the PRIM questionnaire including response categories can be seen in the supplementary file (Table S1). Items included in the perceived DI scale are shown in Table 1 with indication of the items proposed to belong to the subscales of perceived *exposure to danger and combat* (EDC) and *witnessing consequences of war* (WCW) respectively. Items from the DI scale such as *being threatened with a weapon?* and *being shot at by the enemy?* have high semantic overlap with items in other known combat exposure scales (Keane et al., 1989). The same is the case for items such as *seeing wounded people, seeing dead people* and *witnessing distress among the locals*, which have a high semantic overlap with items from, for example, the *Aftermath of Battle*-subscale of the DRRI/DRRI-2 (Vogt et al., 2013). All items in the DI scale (Berntsen et al., 2012) had four response categories (1 = never, 2 = seldom, 3 = fairly often, 4 = almost daily).

3.2.1.1. Exogenous variables. For the purpose of examining measurement invariance (i.e. no DIF), a range of exogenous variables of potential importance for the assessment of exposure to danger and combat and witnessing the consequences of war was included. In that regard, we selected variables that have previously been found to be related to adverse mental health outcomes following deployment, since those might interact with perceived danger. Specifically, we included variables of previous deployments (yes/no), rank and gender. Studies hitherto have found a relatively consistent positive association between previous deployments and depression, as well as PTSD (Kline et al., 2010; Reger, Gahm, Swanson, & Duma, 2009; Xue et al., 2015). An association between lower rank and higher risk of post-deployment PTSD has also been found (Iversen et al., 2008; Xue et al., 2015). The same is true for female gender, which has consistently been found to be associated with higher risk of post-deployment PTSD (Xue et al., 2015). Finally, we included the deployment cohort itself as an exogenous variable, as it is crucial to uncover whether it is possible to make unconfounded comparisons of exposure to danger and combat and witnessing the consequences of war across the military cohorts when using the raw scale scores.

3.3. Analysis

When undertaking validity studies of questionnaire scales as the ones in the present study, an initial issue

is whether the scales in question are indeed reflective scales or formative scale. In reflective scales there is a causal relationship from the latent construct being measured to the items, such that the latent construct is what determines the responses to the items in a causal relationship (i.e. effect indicators), as latent construct cannot be measured directly (Bollen & Bauldry, 2011; Edwards, 2011). Thus, different measures of the same latent construct can have somewhat varying items, as long as the items are indicators for the construct in question. In a formative scale the direction of the causal relationship goes the other way. In such scales the items form the construct being measured (i.e. composite or formative indicators), and thus if you change the items then you change the construct. Hence the construct in formative scales is not a latent but a manifest one composed of specific elements (Bollen & Bauldry, 2011; Edwards, 2011).

In the present study, the scales used are not objective in the sense that they attempt to measure objective exposure to combat or objective consequences of combat, but rather the soldiers are reporting their perception of exposure to combat and how they perceive that the war has affected the locals and the area they live in. Not all soldiers report they have experienced high exposure to danger and combat – or often have witnessed consequences of war – even though they have been at the same team, participated in the same battle and/or have been at the same places. Some explanations for these individual differences may be different cognitive appraisals of threat and challenge, e.g. based on former war experience or their personality. For example, Caska and Renshaw (2013) in a cross-sectional study found that personality traits impacted soldiers' responses to war trauma indicating that such individual differences may be important factors influencing soldiers' responses when asked about traumatic events during deployment (Caska & Renshaw, 2013). Further, it is clear that in different types of war and conflict different 'exposures' might be present or possible, but we would still say that a conflict-specific scale is a measure of perceived exposure and thus the same latent construct, just with slightly varying indicators. Thus, the scales employed are clearly reflective, as the scales measure the soldier's perception of the degree to which they have been exposed to dangerous and psychologically taxing situations and phenomena, via a number of effect and not composite indicators.

4. Rasch measurement models

The Rasch model (RM) for dichotomous items (Rasch, 1960) is the simplest of the large family of item response theory models (IRT). In the present study, we used the partial credit model (PCM) (Masters, 1982) since the

items of the scale are ordinal. The PCM is simply a generalization of the RM to also include polytomous or ordinal items. Thus, the requirements for the PCM are the same as for the dichotomous RM, as are the properties of scales fitting a PCM. We therefore continue to use the term RM. The RM has five basic requirements for measurement (Kreiner, 2013; Mesbah & Kreiner, 2013): (1) Unidimensionality; that items of a scale measure a single underlying latent construct, (2) Monotonicity; that all items have an increasing probability of a high score with increasing values on the latent variable, (3) Local independence (or no local dependence; LD); that the responses to items are conditionally independent from one another given the latent variable, (4) Absence of differential item functioning (no DIF); that the responses to items are conditionally independent from exogenous variables given the latent variable, (5) Homogeneity; that the rank order of item parameters (item ‘difficulties’) is the same for all persons regardless of their level on the latent variable. Fulfilment of these requirements by a set of item responses means that the scale in question is criterion-relatedly construct valid according to Rosenbaum’s (1989) definition, and that the sum score is statistically sufficient for the latent variable. Sufficiency is desirable when one wishes to use the summed raw score of measurement scales, and this property distinguishes Rasch items from items fitting other IRT models (Nielsen, Kyvsgaard, Sildorf, Kreiner, & Svensson, 2017).

When fit to a RM is rejected, it is still possible to achieve close to optimal measurement, provided that the departures from the RM are in the form of uniform differential item functioning (uniform DIF) and/or uniform local dependence (uniform LD) between items (Kreiner & Christensen, 2007). ‘Uniform’ here simply refers to the fact that the way items depend either on exogenous variables, or other items, is the same for all levels of the latent variable. If this is the case, the uniform DIF or LD can be included and adjusted for in a so-called graphical loglinear Rasch model (GLLRM), which is simply an extension of the RM allowing precisely these departures. If a GLLRM includes only uniform LD, this does not affect the sufficiency of the sum score, but it does affect the reliability of the scale negatively to some degree. If a GLLRM includes uniform DIF, the sum score is no longer a sufficient statistic for the latent variable, as additional information on a person’s membership of any subgroups for which items function differentially is also needed. This may, however, be resolved by equating the score across subgroups to allow subsequent comparisons to be unconfounded by the DIF (Kreiner, 2007).

5. Item analysis

The item analyses of the EDCS and the WCWS were conducted using the same overall strategy. First, we tested fit of the item responses to the RM. If this was rejected, then we proceeded to catalogue the

departures and subsequently to test the fit of the item responses to a GLLRM adjusting for the departures discovered.

The overall model fit (i.e. comparison of item parameters in low and high scoring groups), as well as overall test of no DIF, were tested using Andersen’s (Andersen, 1973) Conditional Likelihood Ratio test (CLR). The fit of individual items was tested using both conditional infit and outfit statistics (Kreiner & Christensen, 2013; Kreiner & Nielsen, 2013) and by comparing the observed item-rest-score correlations with the expected item-rest-score correlations under the model (i.e. the specified RM or GLLRM) (Kreiner, 2011). The presence of DIF and LD in GLLRMs was tested by conditional tests of independence using Goodman-Kruskal gamma coefficients for the conditional association between items and exogenous variables (presence of DIF) or between item pairs (presence of LD) given the rest-scores (Kreiner & Christensen, 2004). We specifically tested for the presence of DIF relative to Cohort (Cohort 2 vs Cohort 1), Previous deployment (yes, no), Rank (private, non-commissioned officer, officer) and Gender (male, female). Evidence of overall fit and no DIF was rejected if this was not supported by evidence of item fit and lack of evidence of both DIF and LD for individual items. The Benjamini-Hochberg procedure was used to adjust for false discovery rate due to multiple testing, when appropriate (Benjamini & Hochberg, 1995). We used $p < .05$ to imply statistical significance for all tests but, as recommended by Cox et al. (1977), this value was not used as a deterministic decision criterion.

For the analysis of unidimensionality, we calculated the expected correlation of the two subscales under the assumption that they measured one and the same latent variable (i.e. that they comprised a single latent construct) and compared these with the observed correlations from the data (Horton, Marais, & Christensen, 2013). We applied a Monte Carlo approach for exact p -values.

Reliability was estimated using Hamon and Mesbah’s (2002) Monte Carlo method, as this takes into account any LD in a GLLRM and adjusts the reliability accordingly. The reliabilities reported here are thus to be interpreted as Cronbach’s alpha, adjusted downwards due to the LD. Targeting, or the degree to which the study population was outside the target range, was assessed graphically by so-called item maps as well as by two indices. Item maps plot the distribution of the person parameters against the distribution of the item thresholds, and thus it can be evaluated whether the majority of persons in the study population are included in the range of item parameters. As the assessment of item maps is purely visual, we also calculated two indices to evaluate targeting of theta, as described in Kreiner and

Table 2. Global tests of DIF and global tests of fit to RM and GLLRM for the Exposure to Danger and Combat Scale and Witnessing Consequences of War Scale.

Groups defined by	RM			GLLRM ^a		
	CLR	df	p	CLR	df	P
A: Exposure to Danger and Combat Scale						
Low-high scores ⁺	49.6	17	< .001	41.9	44	.562
Cohort	147.0	17	< .001	32.4	20	.039*
Previous deployments	18.5	17	.357	32.1	38	.736
Rank	47.9	34	.057	108.0	88	.073
Gender	23.0	17	.149	57.6	44	.082
Groups defined by	RM ^b			GLLRM ^c		
	CLR	df	p	CLR	df	P
B: Witnessing Consequences of War Scale						
Low-high scores ⁺	3.1	10	.980	10.5	16	.839
Cohort	20.5	10	.025*	29.6	16	.020*
Previous deployments	16.0	10	.100	19.6	16	.237
Rank	22.9	20	.292	35.4	32	.311
Gender	22.3	10	.014*	26.3	16	.050

RM = Rasch model. GLLRM = Graphical loglinear Rasch model. CLR = Conditional Likelihood Ratio test. df = Degrees of freedom.

⁺ Global test of fit as a test of homogeneity in item parameters for persons with low and high scores.

* $p > .05$ after correction for false discovery rate due to multiple testing using the Benjamini-Hochberg procedure.

^aThe model assumes that items 2, 3 and 4 are locally dependent and affected by DIF relative to cohort and that item 5 is affected by DIF relative to previous missions and cohort.

^bRasch model rejected despite the global test results, due to strong evidence of local response dependence between items 7 and 8 in the analyses of the partial association between items given the total score over the other items ($lr = 75.84$, $df = 9$, $p < .0001$).

^cThe model assumes that items 7 and 8 are locally dependent.

Christensen (2013): the test information target index, which is the mean test information divided by the maximum test information for theta, and the root mean squared error (RMSE) target index, which is the minimum standard error of measurement divided by the mean standard error of measurement for theta. Both indices should preferably have a value close to one. Additionally, we estimated the target of the observed score and the standard error of measurement of the observed score (SEM). All item analyses were conducted using the DIGRAM software package (Kreiner & Nielsen, 2013).

6. Results

6.1. Model and item fit

The Exposure to Danger and Combat Scale (EDCS) did not fit a RM, as both the global test of fit rejected fit ($p < .001$) and strong overall evidence of Cohort-DIF was found ($p < .001$) (Table 2, RM column). Further tests of local independence showed evidence of several instances of both local dependence between items and DIF (results not shown). Thus, subsequently, the EDCS was found to fit a rather complex GLLRM that accounted for several instances of local dependence and DIF. Not unexpectedly, local dependence was found between items 2 and 4 (2: *Being shot at* and 4: *Passing areas with combat activities*) and items 3 and 4 (3: *Being in areas with roadside bombs or mines* and 4: *Passing areas with combat*

activities). DIF was present for items 2, 3, 4 and 5 (5: *Aggressive behaviour from the locals*) relative to cohort: For items 2 and 3, soldiers from Cohort 1 were more likely to report having experienced being shot at and passing through areas with improvised explosive devices or landmines than soldiers from Cohort 2, independently of their level of exposure. The reverse was the case for items 4 and 5, where soldiers from Cohort 2 were more likely to report having passed through areas with combat and with having experienced aggressive behaviours from the local population than were soldiers from Cohort 1. Furthermore, DIF was present for item 5 relative to previous missions, as soldiers deployed on missions previous to the current one were less likely to report having experienced aggressive behaviour from locals than soldiers not previously deployed, no matter their level of exposure.

The Witnessing Consequences of War Scale (WCWS) also did not fit an RM, even though all the global tests accepted fit (p -values $> .05$) (Table 2, RM column), as the further tests of local independence did find evidence of local dependence between two items (results not shown). Thus again, the WCWS was found to fit a simple GLLRM, accounting for local dependence between items 7 and 8 (i.e. 'did you see dead people' and 'did you see injured or wounded people'). No evidence of DIF was found for the WCWS.

All individual items fit the GLLRMs for the respective subscales (Table 3). For example, for the first EDC item (*Being threatened with a weapon*), the observed item-rest-score correlation ($\gamma = .63$) and the expected item-rest-score correlation under the GLLRM for the EDCS ($\gamma = .65$) were found to be equal ($p > .05$).

7. Unidimensionality

Following the establishment of fit to GLLRMs for both subscales, we proceeded to test the null-hypothesis that these in fact measured one and the same latent variable, as opposed to the proposed two separate latent constructs. Both the asymptotic test and the Monte Carlo

Table 3. Item fit statistics for the Exposure to Danger and Combat Scale and Witnessing Consequences of War Scale.

Items	Observed γ	Expected γ	p
A: Exposure to Danger and Combat Scale			
1	.63	.65	.93
2	.75	.75	.97
3	.76	.76	.97
4	.78	.78	.92
5	.59	.61	.52
10	.68	.64	.27
B: Witnessing Consequences of War Scale			
6	.67	.68	.93
7	.78	.79	.69
8	.79	.79	.80
9	.68	.65	.41

Item-rest-score correlations for the respective graphical loglinear Rasch models. Infit and outfit statistics for item fit showed the same results and are therefore not shown here.

test clearly rejected unidimensionality (observed gamma between subscale scores .636, expected gamma between subscale scores .702, s.e. = .019, asymptotic $p < .001$, exact $p < .0001$). We therefore suggest that the DI domain measured with the DI scale is composed of the two qualitatively different, but highly correlated latent constructs of Exposure to Danger and Combat and Witnessing Consequences of War, respectively. Hence, we also suggest that reducing the two scales to a single total scale and one total score would be invalid. We have provided the results of the tests of overall fit to the RM for the original 10-item perceived DI scale (fit was rejected), as well as the result of the overall tests of fit to the common 10-item GLLRM resulting from combining the two subscale GLLRMs (fit was rejected) in the supplement (Table S2).

8. Reliability and targeting

Targeting of the EDCS and the WCWS differed for the two scales. The best targeting of the EDCS was found for soldiers from Cohort 1, with only a small variation among those who had previously been deployed and those who had not. For Cohort 2, the targeting of the EDCS was not very good, again with a small variation dependent on previous deployment (Figure S2). With regard to average test information provided by the EDCS items, this was lowest for Cohort 2 (40% and 45% of the maximum obtainable information, respectively), and highest for Cohort 1 (72% and 83% of the maximum obtainable information, respectively; Table 4). The targeting of the WCWS across all soldiers was not optimal (Figure S1), as also shown by the average test information provided by the WCWS items at only 58% of the maximum obtainable information (Table 4). The reliability of the EDCS and the WCWS were at satisfactory levels for all groups of soldiers, though with small variations in the case of the EDCS (Table 4).

9. The effect of DIF in the perceived exposure to danger and combat subscale

The EDCS items 2, 3, 4 and 5 were found to function differentially relative to Cohort, and item 5 was also relative to previous deployment (see above). The effect of this DIF on the EDCS score was, however,

found to be limited when making comparisons of sum score in the two cohorts or in the groups of soldiers previously deployed and not previously deployed, respectively, as the bias is very small in both cases (Table 5). A DIF-equation table with the values adjusting for the DIF across the entire score range with the group of Cohort 2 previously deployed to allow unconfounded comparison, is provided in the supplement Table S3).

10. Patterns of information from two scales compared to one overall scale

To further examine how individuals score on the ECDS and the WCWS relative to each other, we plotted the two against each other; using bins to illustrate the number of individuals with each score combination (see Figure 1). Further, as illustrated by different colours, we plotted individuals respective to a total score of above or below 20 on the original DI-scale. As seen in Figure 1, most individuals do not score equivalently on the two subscales. Furthermore, groups are clearly distinguished in their responses on the two subscales based on their DI-score. Most importantly, Figure 1 shows that most individuals who score high (20 points or above) on the original 10-item DI-scale score high on EDCS and low on WCWS or vice versa. This illustrates that the proposed subscales provide not only more accurate scores, but also a more nuanced pattern of information on exposure, than the original perceived DI scale.

Table 5. Effect of differential item function relative to previous deployment and cohort on the sum score of the Exposure to Danger and Combat Scale.

Previous deployment	Observed		Adjusted		Bias
	Mean	se	Mean	se	
Yes	10.66	0.18	10.70	0.18	-0.04
No	11.55*	0.20	11.44*	0.21	0.10
Cohort					
Cohort 1	12.03	0.18	12.05	0.19	-0.02
Cohort 2	9.75**	0.17	9.69**	0.16	0.06

Besides correction of the mean score due to the DIF in question (Previous deployment and Cohort), we also corrected for cohort DIF when considering previous deployment DIF and vice versa to arrive at the most accurate adjusted mean score.

* $p < .01$, ** $p < .001$, both testing equality of means

Table 4. Targeting and reliability of the Exposure to Danger and Combat Scale and Witnessing Consequences of War Scale.

Groups defined by DIF	Theta				Sum score				
	Target	Mean	Test inf.	Target index	RMSE target index	Target	Mean	Mean SEM	Reliability
Exposure to Danger and Combat Scale									
Cohort 1, previous deployments	1.68	-1.69	0.724		0.836	18.75	11.60	1.32	0.86
Cohort 1, no previous deployments	1.09	-1.10	0.828		0.898	17.65	12.63	1.44	0.77
Cohort 2, previous deployments	0.19	-2.91	0.398		0.633	15.59	9.65	1.17	0.83
Cohort 2, no previous deployments	0.17	-2.68	0.446		0.671	15.82	9.94	1.24	0.72
Witnessing Consequences of War Scale									
All	0.95	-2.00	0.580		0.783	11.07	7.06	1.03	0.76

RMSE = root mean squared error of the estimated theta score. SEM = standard error of measurement of the observed score.

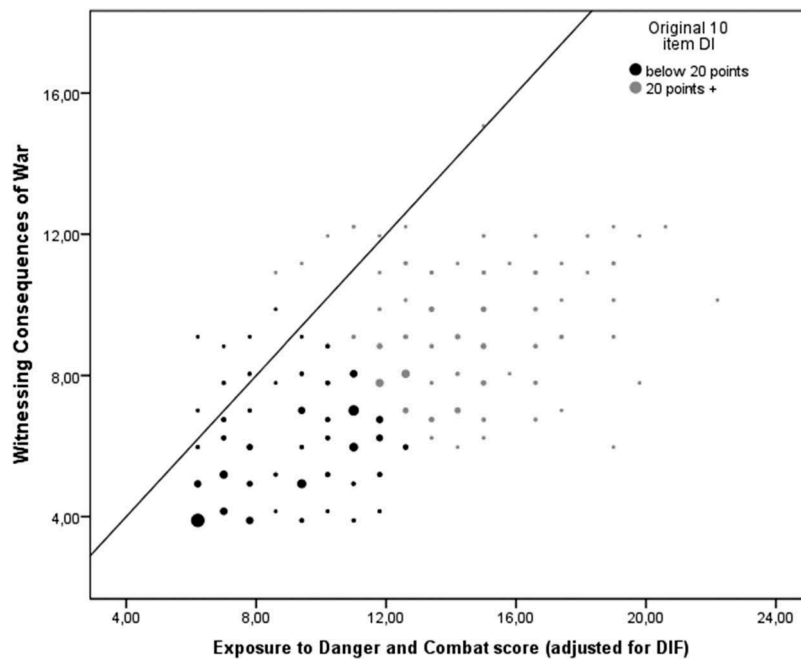


Figure 1. Comparing Exposure to Danger and Combat Scale and Witnessing Consequences of War Scale scores with the original 10-item perceived DI score.

The size (Bins) of the dots shows the proportion of respondents having a given combination of responses to the two sub-scales. The superimposed black line illustrates where the scores on the two subscales are equivalent.

11. Discussion

In the current study, we used RMs to evaluate the construct validity of two proposed subscales within a 10-item perceived DI scale. The two subscales aimed at assessing perceived exposure to danger and combat (EDCS) and witnessing consequences of war (WCWS), respectively.

Overall, we found that while none of these two subscales fit a simple RM, both could be fitted to GLLRMs. The GLLRMs took into account local dependence between items and DIF of items relative to exogenous variables. After establishing the fit of the two subscales to GLLRMs, we tested whether these were in fact one unidimensional scale. This was clearly rejected. Based on these overall findings, we conclude: (1) the DI scale does *not* consist of a unidimensional construct and (2) both ECDS and WCWS are essentially valid and objective scales (Kreiner & Christensen, 2007).

When taking a closer look at the two subscales, it is clear that targeting is uneven in the EDCS. More specifically, this subscale was better targeted for soldiers from Cohort 1 than for soldiers from Cohort 2. Within both cohorts, the best targeting was found for soldiers not previously deployed on missions. Since Cohort 1 was deployed during a time with more intensive combat than Cohort 2, this suggests that the EDCS provides most information for soldiers deployed to more combat-intense missions and deployed for the first time. It is, however, important to note that even if targeting is unequal due to items functioning differentially across groups of soldiers and one group has less than satisfactory

targeting, this is somewhat balanced by the fact that targeting is best for the soldiers most at risk of adverse psychological consequences of deployment (i.e. the soldiers experiencing more combat; Xue et al., 2015). The targeting of the WCWS was low, but acceptable, for soldiers from both cohorts and for those with and without previous deployments, respectively.

With regard to the reliability of the scales, both the EDCS and the WCWS have satisfactory levels of reliability for all groups of soldiers, with only slightly lower reliabilities for soldiers who had not previously been deployed. This is a crucial finding in relation to the clinical use of the EDCS and the WCWS for risk screening of individual soldiers, which is an important use of both the EDCS and the WCWS (Kimberlin & Winterstein, 2008).

A central element when trying to establish whether a measure can be used to compare different groups or samples (in our case: deployment cohorts) is the concept of DIF. For the ECDS, no less than four items functioned differentially relative to cohort, namely items 2, 3, 4 and 5. Two of these items (2: *Being shot at* and 3: *Being in areas with roadside bombs or mines*) were more frequently endorsed in Cohort 1 independently of the level of EDC, whereas the reverse was the case for the remaining two items (4: *Passing areas with combat activities* and 5: *Aggressive behaviour from the locals*), i.e. these were more frequently endorsed in Cohort 2 independently of the level of EDC. This potentially poses a problem, when the aim is to arrive at a measure for comparing the exposure to combat across deployment cohorts – and more cohorts than included in the present study. To obtain an

estimation of the effect of the bias in the scale score cause by DIF, we tested differences between cohorts using the observed and adjusted ECDS scores. We found that the differences in total scale scores, whether comparing using the adjusted scores or not, were minimal, and that the difference between cohorts was significant at the same level. Hence, when using the EDCS score as a practical means of assessing perceived exposure to danger of combat, the DIF will not make a noteworthy difference on the total EDCS score. Moreover, the finding that Cohort 1 score significantly higher than Cohort 2 on the EDCS adds to the validity of the scale, as the threat assessment for Cohort 1 was known a priori to be more severe than for Cohort 2.

Taking an even closer look at the discovered DIF, it appears that item 5: *Experiencing aggressive behaviour from the locals* not only functions differentially due to what might be considered 'objective' circumstances of different missions (i.e. that on some missions more aggressive behaviour is experienced). This item also functions differentially dependent on whether the soldiers have been deployed on previous mission or not. Specifically, first-time deployers are more likely to report having experienced aggressive behaviour from the locals than their colleagues who have been deployed before, regardless of their EDCS scores. As for cohort DIF, we tested the practical implications of previous deployment DIF and found that, for the total scale score, again the differences were minimal, whether using the adjusted scores or not, and that differences in EDCS scores between single and multiple deployers were significant at the same level, regardless of the score used. The finding that first time deployers score significantly higher on the EDCS than soldiers who have been deployed previously adds to the validity of the scale. Our hypothesis is that deployment experience to some degree might make one less prone to perceiving the behaviour of locals as aggressive. Indeed, a study of 'cultural stress' in the war zone has found the perception of locals to change with the number of deployments (Azari, Dandeker, & Greenberg, 2010).

The visual presentation of ECDS and WCWS scores compared to each other and to the original 10-item DI score clearly illustrates that (1) individuals typically do not score the same on the two subscales, (2) that individuals can be clearly grouped on subscale responses based on high or low DI-score and (3) that most individuals who score high on DI also score relatively high on the ECDS and relatively low on the WCWS and vice versa. The last point is an especially crucial one to make: considering only total scores of perceived exposure to danger and distress and the witnessing of it will not capture potentially important differences in how such perceived exposure or witnessing related to adverse outcomes of deployment. Indeed, our results imply that while being exposed to danger and combat and witnessing the consequences of war in a warzone might be highly correlated subscales, they are *not* one construct and they might

potentially have different implications for adverse outcomes following deployments. As such, failing to distinguish between the two will obscure the results of for example predictive analyses or, in practice, screening procedures. A few previous studies have examined different facets of combat, threat and other adverse experiences during deployment and their relations to PTSD-symptomatology and have found these facets relate differently to PTSD (Caska & Renshaw, 2013; Fontana & Rosenheck, 1999; King, King, Gudanowski, & Vreven, 1995; Osório et al., 2017; Renshaw, 2011). Expanding this line of research, future analyses based on this or other data sets including measures of war zone stress stemming from combat and other adversities should strive to separate the perceived effects of combat exposure from that of the witnessing of danger and distress and the consequences of war. As suggested by Osório et al. (2017), there could be a valuable future clinical implication of doing so since knowledge on how different kinds of exposures during deployment may be related to certain PTSD clusters could help clinicians when deciding what kind of initiatives that could possibly be minimizing the effects of the trauma-exposure. Osório et al. found in an analysis of combat experiences and their relationship to PTSD among Afghanistan deployed military personnel that larger exposure to violent combat predicted symptoms of numbing and re-experiencing; nearness to wounding or death experiences predicted re-experiencing and anxious-arousal symptoms; and exposure to explosive device predicted anxious-arousal symptoms (Osório et al., 2017). The knowledge of different effects of war zone stressors on PTSD was also found in a study of male and female Vietnam Veterans, e.g. the factor 'malevolent environment' had a higher impact than other measured factors. Furthermore, exposure to combat did not have a direct effect on PTSD but exert an indirect influence in that it depended on perceived threat (King et al., 1995). The latter was also found in a study of Middle East deployed soldiers; the association of combat experienced with PTSD was influenced by perceived threat but perceived threat had no impact on the relation between post-battle experiences and PTSD (Renshaw, 2011).

Our study is not without limitations. First, we rely on two cohorts, both deployed to Afghanistan, to test DIF. More cohorts, preferably deployed to different war zones, might have strengthened the arguments made here. However, the cohorts in question deployed to Afghanistan at times when the situation, in terms of conditions, threat and tasks, was very different. Hence, it is reasonable to assume that DIF between these two cohorts would also be seen between additional cohorts. Second, the construction and selection of the original perceived DI items was performed by military psychologists and other researchers, and there is no record on theoretical or empirical justification on item generation or inclusion. However, comparison with other scales for

assessing combat exposure and other adverse events during deployment (King, King, Vogt, et al., 2006) suggests a substantial semantic overlap.

Despite these limitations, this study brings new valuable information on how best to assess war-related factors of importance for the risk of adverse mental health effects following deployment. To the best of our knowledge, only one prior study (Carvalho, Cunha, Pinto-Gouveia, & Da Motta, 2014) has applied RM with the purpose of validating a combat exposure measure, and no studies have done so comparing two or more cohorts for DIF. As such, we consider that our findings are an important step forward. For future studies, we suggest that more cohorts are included for studying cohort DIF, that increased attention is paid to the differences between perceived exposure to danger and combat vs witnessing consequences of war, and that these two facets of war experiences are studied in relation to potential adverse mental health outcomes following deployment.

Acknowledgments

The authors wish to thank all soldiers who filled out the PRIM questionnaire one or more times following their deployments with the Danish Military. Furthermore, we wish to thank Ann Karen Christensen and her colleagues at the Military Psychology Department for their help in developing and implementing the PRIM questionnaire in 1998.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123–140.
- Andersen, S. B., Karstoft, K.-I., Bertelsen, M., & Madsen, T. (2014). Latent trajectories of trauma symptoms and resilience: The 3-year longitudinal prospective USPER study of Danish veterans deployed in Afghanistan. *The Journal of Clinical Psychiatry*, 75(9), 1001–1008.
- Azari, J., Dandeker, C., & Greenberg, N. (2010). Cultural stress: How interactions with and among foreign populations affect military personnel. *Armed Forces & Society*, 36(4), 585–603.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Berntsen, D., Johannessen, K. B., Thomsen, Y. D., Bertelsen, M., Hoyle, R. H., & Rubin, D. C. (2012). Peace and war: Trajectories of posttraumatic stress disorder symptoms before, during, and after military deployment in Afghanistan. *Psychological Science*, 23(12), 1557–1565.
- Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, 16(3), 265–284.

- Carlson, K. F., Nelson, D., Orazem, R. J., Nugent, S., Cifu, D. X., & Sayer, N. A. (2010). Psychiatric diagnoses among Iraq and Afghanistan war veterans screened for deployment-related traumatic brain injury. *Journal of Traumatic Stress*. doi:10.1002/jts.20483
- Carvalho, T., Cunha, M., Pinto-Gouveia, J., & Da Motta, C. (2014). Development of exposure to combat severity scale of the combat experiences questionnaire (CEQ). *Journal of Anxiety Disorders*, 28(8), 938–946.
- Caska, C. M., & Renshaw, K. D. (2013). Anxiety, stress & coping: An personality traits as moderators of the associations between deployment experiences and PTSD symptoms in OEF/OIF service members. *Anxiety, Stress, & Coping*, 26(1), 36–51.
- Cox, D. R., Spjøtvoll, E., Johansen, S., van Zwet, W. R., Bithell, J. F., & Barndorff-Nielsen, O., et al. (1977). The role of significance tests [with discussion and reply]. *Scandinavian Journal of Statistics*, 4(2), 49–70.
- Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, 14(2), 370–388.
- Fear, N. T., Jones, M., Murphy, D., Hull, L., Iversen, A. C., Coker, B., ... Wessely, S. (2010). What are the consequences of deployment to Iraq and Afghanistan on the mental health of the UK armed forces? A cohort study. *The Lancet*, 375(9728), 1783–1797.
- Fontana, A., & Rosenheck, R. (1999). A model of war zone stressors and posttraumatic stress disorder. *Journal of Traumatic Stress*, 12(1), 111–126.
- Hamon, A., & Mesbah, M. (2002). Questionnaire reliability under the Rasch model. In Mesbah M., Cole B. F., Lee M. L. T. (Eds.). *Statistical methods for quality of life studies*. (pp. 155–168). Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-1-4757-36250_13.
- Hoge, C. W., Auchterlonie, J. L., & Milliken, C. S. (2006). Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *JAMA: The Journal of the American Medical Association*, 295(9), 1023–1032.
- Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, 351(1), 13–22.
- Horton, M., Marais, I., & Christensen, K. B. (2013). Dimensionality. In Christensen KB, Kreiner S, Mesbah M. (Eds.). *Rasch Models in Health* (pp. 137–158). ISTE and John Wiley & Sons, Inc.
- Iversen, A. C., Fear, N. T., Ehlers, A., Hacker Hughes, J., Hull, L., Earnshaw, M., ... Hotopf, M. (2008). Risk factors for post-traumatic stress disorder among UK Armed Forces personnel. *Psychological Medicine*, 38(4), 511–522.
- Karstoft, K.-I., Andersen, S. B., & Nielsen, A. B. S. (2017). Assessing PTSD in the military: Validation of a scale distributed to Danish soldiers after deployment since 1998. *Scandinavian Journal of Psychology*, e1–e9. doi:10.1111/sjop.12360
- Karstoft, K.-I., Nielsen, A. B. S., & Nielsen, T. (2017). Assessment of depression in veterans across missions: A validity study using Rasch measurement models. *European Journal of Psychotraumatology*, 8(1), 1326798.
- Keane, T. M., Fairbank, J. A., Caddell, J. M., Zimering, R. T., Taylor, K. L., & Mora, C. A. (1989). Clinical evaluation of a measure to assess combat exposure. *Psychological Assessment*, 1(1), 53–55.
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65(23), 2276–2284.

- King, D. W., King, L. A., Gudanowski, D. M., & Vreven, D. L. (1995). Alternative representations of war zone stressors: Relationships to posttraumatic stress disorder in male and female Vietnam veterans. *Journal of Abnormal Psychology, 104*(1), 184–195.
- King, L. A., King, D. W., Knight, J., & Samper, R. E. (2006). Deployment risk and resilience inventory: A collection of measures for studying deployment-related experiences of military personnel and veterans. *System, 18*(2), 89–120.
- King, L. A., King, D. W., Vogt, D. S., Knight, J., & Samper, R. E. (2006). Deployment risk and resilience inventory: A collection of measures for studying deployment-related experiences of military personnel and veterans. *Military Psychology, 18*(2), 89–120.
- Kline, A., Falca-Dodson, M., Sussner, B., Ciccone, D. S., Chandler, H., Callahan, L., & Losonczy, M. (2010). Effects of repeated deployment to Iraq and Afghanistan on the Health of New Jersey Army National Guard Troops: Implications for military readiness. *American Journal of Public Health, 100*(2), 276–283.
- Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology, 59*(3), 268–298.
- Kreiner, S. (2011). A note on item-restscore association in Rasch models. *Applied Psychological Measurement, 35* (7), 557–561.
- Kreiner, S. (2013). The Rasch Model for dichotomous items. *Rasch Models in Health, 5*–26. doi:10.1002/9781118574454.ch1
- Kreiner, S., & Christensen, K. B. (2004). Analysis of local dependence and multidimensionality in graphical log-linear Rasch models. *Communications in Statistics-Theory and Methods, 33*(6), 1239–1276.
- Kreiner, S., & Christensen, K. B. (2007). Validity and objectivity in health-related scales: Analysis by graphical loglinear Rasch models. In von Davier, & Carstensen (Eds.). *Multivariate and mixture distribution Rasch models* (pp. 329–346). Springer.
- Kreiner, S., & Christensen, K. B. (2013). Person parameter estimation and measurement in Rasch models. *Rasch Models in Health, 63*–78. doi:10.1002/9781118574454.ch4
- Kreiner, S., & Nielsen, T. (2013). *Item analysis in DIGRAM 3.04: Part I: Guided tours*. Copenhagen: Department of Biostatistics, University of Copenhagen.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.
- McDevitt-Murphy, M. E., Williams, J. L., Bracken, K. L., Fields, J. A., Monahan, C. J., & Murphy, J. G. (2010). PTSD symptoms, hazardous drinking, and health functioning among U.S.OEF and OIF veterans presenting to primary care. *Journal of Traumatic Stress. doi:10.1002/jts.20482*
- Mesbah, M., & Kreiner, S. (2013). The Rasch model for ordered polytomous items. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.). *Rasch Models in health*. London: Wiley.
- Nielsen, J. B., Kyvsgaard, J. N., Sildorf, S. M., Kreiner, S., & Svensson, J. (2017). Item analysis using Rasch models confirms that the Danish versions of the DISABKIDS® chronic-generic and diabetes-specific modules are valid and reliable. *Health and Quality of Life Outcomes, 15*(1), 44.
- Osório, C., Jones, N., Jones, E., Robbins, I., Wessely, S., & Greenberg, N. (2017). Combat experiences and their relationship to Post-traumatic stress disorder symptom clusters in UK military personnel deployed to Afghanistan. *Behavioral Medicine, 1*–10. doi:10.1080/08964289.2017.1288606
- Polusny, M. A., Erbes, C. R., Murdoch, M., Arbisi, P. A., Thuras, P., & Rath, M. B. (2011). Prospective risk factors for new-onset post-traumatic stress disorder in National Guard soldiers deployed to Iraq. *Psychological Medicine, 41*(4), 687–698.
- Polusny, M. A., Kumpula, M. J., Meis, L. A., Erbes, C. R., Arbisi, P. A., Murdoch, M., ... Johnson, A. K. (2014). Gender differences in the effects of deployment-related stressors and pre-deployment risk factors on the development of PTSD symptoms in National Guard Soldiers deployed to Iraq and Afghanistan. *Journal of Psychiatric Research, 49*, 1–9.
- Ramchand, R., Schell, T. L., Karney, B. R., Osilla, K. C., Burns, R. M., & Caldarone, L. B. (2010). Disparate prevalence estimates of PTSD among service members who served in Iraq and Afghanistan: Possible explanations. *Journal of Traumatic Stress, 23*(1), 59–68.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: The Danish Institute of Educational Research.
- Reger, M. A., Gahm, G. A., Swanson, R. D., & Duma, S. J. (2009). Association between number of deployments to Iraq and mental health screening outcomes in US army soldiers. *The Journal of Clinical Psychiatry, 70*(9), 1266–1272.
- Renshaw, K. D. (2011). An integrated model of risk and protective factors for post-deployment PTSD symptoms in OEF/OIF era combat veterans. *Journal of Affective Disorders, 128*(3), 321–326.
- Rosenbaum, P. (1989). Criterion-related construct validity. *Psychometrika, 54*, 625–633. doi:10.1007/BF02296400
- Schnurr, P. P., Kaloupek, D., Sayer, N., Weiss, D. S., Cohen, J., Galea, S., & Weaver, T. L. (2010). Understanding the impact of the wars in Iraq and Afghanistan. *Journal of Traumatic Stress, 23*(1), 3–4.
- Seal, K. H., Bertenthal, D., Miner, C. R., Sen, S., & Marmar, C. (2007). Bringing the war back home: Mental health disorders among 103,788 US veterans returning from Iraq and Afghanistan seen at Department of Veterans Affairs facilities. *Archives of Internal Medicine, 167*(5), 476–482.
- Shea, M. T., Vujanovic, A. A., Mansfield, A. K., Sevin, E., & Liu, F. (2010). Posttraumatic stress disorder symptoms and functional impairment among OEF and OIF national guard and reserve veterans. *Journal of Traumatic Stress. doi:10.1002/jts.20497*
- Sundin, J., Herrell, R. K., Hoge, C. W., Fear, N. T., Adler, A. B., Greenberg, N., ... Bliese, P. D. (2014). Mental health outcomes in US and UK military personnel returning from Iraq. *The British Journal of Psychiatry: The Journal of Mental Science, 204*(3), 200–207.
- Vasterling, J. J., Proctor, S. P., Friedman, M. J., Hoge, C. W., Heeren, T., King, L. A., & King, D. W. (2010). PTSD symptom increases in Iraq-deployed soldiers: Comparison with nondeployed soldiers and associations with baseline symptoms, deployment experiences, and postdeployment stress. *Journal of Traumatic Stress, 23* (1), 41–51.
- Vogt, D., Smith, B. N., King, L. A., King, D. W., Knight, J., & Vasterling, J. J. (2013). Deployment risk and resilience inventory-2 (DRRI-2): An updated tool for assessing psychosocial risk and resilience factors among service members and veterans. *Journal of Traumatic Stress, 26* (6), 710–717.
- Xue, C., Ge, Y., Tang, B., Liu, Y., Kang, P., Wang, M., & Zhang, L. (2015). A meta-analysis of risk factors for combat-related PTSD among military personnel and veterans. *PLoS One, 10*(3), e0120270.