The EssenCES Measure of Ward Atmosphere: Mokken Scaling, Confirmatory Factor Analysis, and Investigating Patient-Level Characteristics

Jack Tomlin & Matthew Tonkin

Published online: 20 Oct 2022.

Submit your article to this journal ⌇

View related articles ⌇

View Crossmark data ⌇

Routledge
Taylor & Francis Group

The EssenCES Measure of Ward Atmosphere: Mokken Scaling, Confirmatory Factor Analysis, and Investigating Patient-Level Characteristics

Jack Tomlin[a] and Matthew Tonkin[b]

[a]School of Law and Criminology, University of Greenwich, London, UK; [b]School of Criminology, University of Leicester, Leicester, UK

ABSTRACT

Ward atmosphere is an important aspect of forensic mental health care. Positive perceptions have been linked to satisfaction during treatment, quality of life, autonomy, involvement in care, emotional expression and lower rates of aggression. The EssenCES is one of the most widely used measures of ward atmosphere. This study sought to add to the psychometric evidence base for the EssenCES and improve our understanding of how perceptions of ward atmosphere are associated with patient-level factors. N = 233 patients in English low, medium and high security hospitals completed the EssenCES, and data were collected on patient age, length of stay in current institution, level of security, ethnicity, Mental Health Act 1983 section, and mental health diagnosis. We used Mokken scaling, confirmatory factor analysis and multiple linear regression. Our analysis supports the three-factor structure of the EssenCES but signposts areas for improvement, specifically, revising and retesting items 10, 13 and 16. We found that Black, Asian and Minority Ethnic patients report lower Experienced Safety domain scores and that patients with a personality disorder diagnosis report lower Therapeutic Hold domain and EssenCES total scores, when controlling for other variables. We suggest future lines of research and situate our findings in the wider literature.

Introduction

Ward atmosphere is an important component of mental health care. As early as 1953 the World Health Organization considered that for inpatient mental health settings, ward atmosphere is 'the single most important factor in the efficacy of treatment given in a mental hospital' (World Health Organization, 1953). Numerous studies since then have attempted to delineate the relationship between this and treatment outcomes (Brunt & Rask, 2007). Though a difficult concept or cluster of concepts to operationalize, ward atmosphere is typically defined as the material, social, and emotional conditions of a given psychiatric ward or unit and the interaction between such factors (Schalast & Groenewald, 2009).

Despite different approaches to operationalization, these measures and qualitative studies identify common key features characteristic of ward atmosphere: the physical environment, relationships between and amongst staff and patients, safety and security, culture or ward-level ethos, attitudes of staff, mutual support and having a secure base (Doyle et al., 2017). van der Helm et al. (2011) propose that atmospheres lacking

positive aspects of these features are characterized by 'a grim and uninviting atmosphere (e.g., lack of safety and boredom) and high repression, including incremental rules, little privacy, and (frequent) humiliation', with these environments considered to be 'closed climates', ill-suited to supporting recovery (p. 161).

Ward atmosphere is as important for inpatient forensic mental health settings as in general psychiatric services. Arguably, given certain features of secure mental health services, ward atmosphere is comparatively more consequential. This is because forensic mental health hospitals of low, medium or high security are restrictive settings, wherein patients have highly regulated interactions with people, groups, organizations etc. outside their ward (Tomlin et al., 2018). Patients' length of stay in forensic settings can be long (years or decades) and may continue indefinitely (Hare Duke et al., 2018). Patients in forensic mental health settings thus spend considerable periods of time on wards with sustained and frequent interactions with peers and staff.

Studies of ward atmosphere in forensic settings link the construct to treatment and criminogenic outcomes, as well as patients' experiences of care. Positive experiences of ward atmosphere have been linked with better quality of life (O' Flynn et al., 2018; Tomlin et al., 2019); patient autonomy, involvement, emotional expression (Howells et al., 2009); and holistic approaches to care and violence reduction such as the Safewards model (Maguire et al., 2018). Negative experiences of ward atmosphere have been linked to feeling restricted in care (Tomlin et al., 2019), conflict, nervousness, and aggression (Howells et al., 2009). Drawing on research from prison settings, Auty and Liebling (2019) found that individuals staying in units with 'moral climates' (units characterized by decency, fairness, humanity, positive relationships with staff, and positive experiences with the use of authority) were less likely to reoffend after release.

Several studies have investigated which patient- or ward-level factors influence ward atmosphere. Gender, level of security, mental health diagnosis, behavioral disturbance, treatment motivation, treatment engagement and therapeutic alliance have all been found to significantly correlate with ward atmosphere (Dickens et al., 2014; Long et al., 2011). de Vries et al. (2016) highlight that given the multi-factorial nature of ward atmosphere, different patient-level characteristics predict these factors differentially. They found that historical risk factors measured with the HCR-20 and the 'interpersonal' domain of the Psychopathy Checklist-Revised (PCL-R) predicted 'Therapeutic Hold'; that the 'antisocial' domain of the PCL-R, and HCR-20 historical risk factors predicted 'Patient Cohesion'; and that 'Experienced Safety' was predicted by HCR-20 clinical risk factors. A more recent analysis of data from a group of long-term patients in forensic mental health settings (mean length of stay after index offense = 19.1 years) found that there was evidence to suggest that the Therapeutic Hold domain of the EssenCES was predictive of treatment readiness (Gaab et al., 2020). These studies suggest that ward atmosphere can have a dialectic relationship with these other (dynamic) variables (Tonkin, 2016).

Efforts to measure ward atmosphere have grappled with these conceptual challenges when seeking to develop valid and reliable instruments. The Ward Atmosphere Scale (Moos & Houts, 1968) and its forensic counterpart, the Correctional Institutions Environment Scale (CIES; Moos, 1975), were widely used but have been criticized for being too lengthy, outdated, and lacking internal consistency (Tonkin, 2016). The Measuring Quality of Prison Life (MQPL; Liebling

et al., 2014)) includes domains measuring 'Belonging', 'Prisoner social life', 'Power' 'Order', 'Safety' and 'Relationships' amongst others. However, the MQPL was developed and validated in prison units and has not yet been applied to secure psychiatric settings.

Various other measures exist, but the most commonly used in forensic mental health contexts is the Essen Climate Evaluation Schema (EssenCES; Schalast et al., 2008). The EssenCES was developed in Germany and has been validated elsewhere. It has three domains: Patient Cohesion and Mutual Support, Therapeutic Hold, and Experienced Safety. Patient and staff versions exist, as do versions dedicated to prison settings, and gender-neutral versions. EssenCES is a 17-item instrument, with responses provided via a 5-point Likert scale. The EssenCES has demonstrated satisfactory psychometric properties, with its multi-factorial structure validated in UK samples using Exploratory and Confirmatory Factor Analysis (Howells et al., 2009; Milsom et al., 2014; Tonkin et al., 2012). Though these studies have generally supported the factorial structure of the EssenCES, some items have been found to cross-load (e.g., PC3: 'Most patients don't care about their fellow patients' problems' (Howells et al., 2009); and TH4: 'Often, staff seem not to care if patients succeed or fail in treatment' (Milsom et al., 2014)). The first of these two examples led to this item being revised (PC3 now reads: 'Patients care about their fellow patients' problems') and subsequently validated (Tonkin et al., 2012). All analyses to date have used Classical Test Theory (CTT) methods to assess the validity and internal reliability of the EssenCES (e.g., factor analysis).

More work is needed to investigate the psychometric properties of the EssenCES. Specifically, few studies have sought to validate the EssenCES in forensic samples, especially in low and medium secure settings. It is also problematic that studies have only used CTT methods. Although testing informed by CTT does produce useable results, it is arguable whether they are the most accurate methodology when working with ordinal, Likert-scale data, such as that collected in the EssenCES (Rusch et al., 2017). Some methodologists advocate that non-parametric item response theory (NIRT) approaches, such as Mokken scaling, are more suitable (van der Eijk & Rose, 2015). CTT methods, such as factor analysis, assume a linear relationship between test scores and true scores on a latent variable, investigate correlations and variance in item-responses (raw scores) to derive latent factors or components; NIRT approaches examine the non-

linear probabilistic relationship between an item and a latent construct, in other words the probability of a respondent endorsing or scoring highly on an item given their actual position on a latent construct (Watson et al., 2018). Finally, our understanding of the correlates of patient experiences of ward atmosphere is still developing. Further research into how patient-level factors are associated with ward atmosphere is needed to better our understanding of the construct, its subdomains, and its correlates. This can help clinicians structure/develop care and routine activities with patients in a way that maximizes positive experiences of ward atmosphere.

Aims

This study aimed to A) test the factorial structure of the EssenCES in a new sample across low, medium and high security levels and B) identify to what extent certain clinical, legal and demographic factors predicted ward atmosphere scores in a group of patients in forensic mental health settings. Specifically, we investigated whether age, length of stay in current institution, level of security, ethnicity, Mental Health Act 1983 section, and mental health diagnosis statistically predicted the EssenCES patient-rated total score and domains: Therapeutic Hold, Patient Cohesion, and Experienced Safety. We hypothesized that level of security and primary psychiatric diagnosis would be significant predictors, given past findings outlined in the introduction; we made no specific hypotheses in relation to the other patient-level predictor variables.

Materials and methods

This study is reported in accordance with the STROBE 'Checklist of items that should be included in reports of cross-sectional studies' (von Elm et al., 2007). This was a cross-sectional, observational study. Data collection took place between May 2018 and April 2019 as part of a wider project investigating the care experiences of patients in forensic mental health settings. Other questionnaires were used at the same time, the results of which are not relevant for this manuscript and are described elsewhere (Tomlin et al., 2019). Only data pertaining to ward atmosphere and patient-level characteristics are described.

Sampling frame and setting

The sampling frame for this study included all patients residing in inpatient (low, medium or high) secure

forensic mental health services in England. There were several exclusion criteria however: where patients were too unwell to participate, they had a primary diagnosis of a learning disability, or where patients lacked capacity to consent and participate. A National Health Service (NHS) approved translator was employed to help one participant understand the questionnaire items. The sampling strategy was stratified: we deliberately recruited participants from low, medium and high security settings; wards aimed at providing care for patients at different stages of their treatment journey (e.g., admission, treatment, rehabilitation, long-stay); wards supporting patients with different diagnoses e.g., mental illness and personality disorder; and wards for women and for men. The sampling strategy was also convenient: patients were not randomly selected, recruitment was contingent upon consent.

Recruitment

16 NHS Trusts (healthcare providers commissioned to provide services within specific geographical or specialist remits) were involved in the project. The research project, its aims and methods were first presented to senior management and clinical staff, and then to patients and ward staff in regular community meetings. Interested patients were given information sheets and at least 24 hours to consider their participation in the study. A meeting was arranged in person. A member of the research team, including local researchers from the NHS Clinical Research Network, sat with participants as they completed the questionnaires. Patients signed consent forms and were able to ask questions and were given a debrief after data collection. Capacity to consent was ascertained through discussion with a member of the participant's care team on the day of the meeting and by asking the participant to describe back to the researcher the purpose of the study, what it involves, and what happens after the meeting.

Ethics and consent

Ethical and procedural approvals were granted by the Leicestershire South Research Ethics Committee (REC:17/EM/0159), the University of Nottingham, and the National Health Service Health Research Authority (NHS HRA).

Data collection and variable coding

Patient- and hospital-level data were collected by research staff or members of the NHS Clinical Research

Network directly from patient hospital notes. These data included: age, length of stay in current institution, ethnicity, Mental Health Act 1983 section, mental health diagnosis/es, and level of security. Diagnoses were made and recorded in accordance with the International Classification of Diseases (ICD).

The EssenCES questionnaire was used to measure patient experiences of ward atmosphere. Items 1 and 17 are excluded from the analysis as these are intended by the EssenCES developers to ease respondents into and out of the questionnaire by using words with positive valence (feeling), with the remaining 15 items being used to calculate a total ward atmosphere score (ranging from 0-60). The EssenCES consists of three domains, each with five items. These domains are: Therapeutic Hold, Experienced Safety, and Patient Cohesion. Total scores can be calculated for each of these domains, with domain scores ranging from 0-20. Responses were coded such that a higher score represents a greater amount of the measured construct e.g., overall ward climate and the three domains.

To facilitate the linear regression methods used in this analysis and to allow an assessment of the unique contributions of multiple predictor variables (instead of conducting a series of ANOVAs, t-tests etc.), the following variables are measured at the continuous level: age, length of stay in current institution; at the ordinal level: level of security = low, medium, and high; and dichotomously: ethnicity = black, Asian and minority ethnic or white, Mental Health Act 1983 section = forensic section or civil section, and mental health diagnosis = primary diagnosis of mental illness (ICD-10 F.2 diagnoses) or personality disorder (ICD-10 F.6 diagnoses). Very few participants were women. As we did not want to exclude them from the study, we have grouped both men and women together in the analysis.

Data analysis

IBM SPSS v27 was used for the descriptive and regression analyses; R was used to conduct the Mokken scaling and Confirmatory Factor Analysis (CFA). Data were assessed for missingness and normality. 1% of clinical, legal and demographic data were missing; these were treated in a pairwise fashion in the analysis. 0.4% of EssenCES response data were missing. As missing values were missing at random (MCAR; χ^2 = 95.932, df = 106, sig = 0.748; (Little, 1988)), missing values were imputed with SPSS's Automatic Imputation Method (Tabachnick & Fidell, 2013). Shapiro-Wilk statistic indicated that the EssenCES

total score was normally distributed (p=.084), but that the three domains and individual items were not. Descriptive statistics are presented in Table 1 as counts, frequencies and percentages.

Mokken scaling was used to test the factorial structure of the EssenCES questionnaire in an exploratory, inductive way. Mokken scaling is a NIRT approach to identifying latent constructs in polytomous (non-continuous, ordinal) questionnaire response data that is considered more appropriate than Classical Test Theory methods, such as Exploratory Factor Analysis (EFA) and Principal Components Analysis (PCA) (Hardouin et al., 2011). This method reports several Loevinger coefficients in the form of: inter-item coefficients (Hij), item-factor coefficients (Hi), and a total scale coefficient (H). Loevinger coefficients >0.3 are desirable (Hardouin et al., 2011; van der Ark, 2007). Mokken scaling holds several assumptions: unidimensionality (considered met as all items have been shown to measure 'ward atmosphere' in previous empirical studies), double monotonicity (considered met where 'crit' scores for items > 80), and local independence (considered met as no item responses are contingent on other responses) (Hardouin et al., 2011; van der Ark, 2007). The R package 'Mokken' was used (van der Ark et al., 2021). Some methodologists suggest that Mokken scaling should be conducted with samples N= >250. As our sample nearly obtains this (N = 233), we feel it justifiable to conduct Mokken scaling but acknowledge this limitation (Watson et al., 2018).

We then conducted Confirmatory Factor Analysis on the factorial structure derived from the Mokken scaling and compared the Goodness of Fit (GoF) statistics of this model to the Schalast three domain model of the EssenCES. In the Mokken-derived model items loading onto a factor with Hi coefficients >0.3 were fitted as exogenous variables onto the latent factors described in the Results section. In the Schalast model, the five items of each EssenCES domain were fitted as exogenous variables onto the three latent factors (Experienced Safety, Therapeutic Hold, Patient Cohesion). The internal consistency of the EssenCES total scale, domains, and Mokken scales are described with Cronbach's α, with scores >0.7 indicative of good internal consistency (Bland & Altman, 1997). Corrected Item-Total Correlation (CITC) scores are reported for all items in both models, with values >0.3 considered acceptable.

The R package 'lavaan' was used to perform the CFA (Rosseel et al., 2022). As the EssenCES data were non-normally distributed, the "MLM" estimator was

Table 1. Participants' demographic, clinical and legal profiles.

Variable	Frequency	%
Security Level		
Low	97	42
Medium	87	37
High	49	21
Total	233	100
Sex		
Male	223	96
Female	9	4
Missing	1	1
Total	233	100
Ethnicity		
White	160	69
Black & Caribbean, Asian, Mixed, Other	69	30
Missing	4	2
Total	233	100
Primary psychiatric diagnosis		
Personality disorder	37	16
Mental illness	140	60
Other/missing	56	24
Total	233	100
Mental Health Act 1983 Section[a]		
Civil section (s 3)	44	19
Forensic section (ss 37, 37/41. 41(5), 45(A), 47/49, 36, 48/49, 38 MHA)	186	81
Missing	3	1
Total	233	100

Notes: Percentages might not add to 100% as values have been rounded.
[a]Mental Health Act 1983 sections: s 3, involuntary treatment not ordered in relation to criminal legal proceedings; s 37, a 'hospital order' ordered by a judge after conviction for an offense, indefinite period; s 37/41, a 'hospital order' with restrictions which necessitate additional approvals from the Secretary of State/Ministry of Justice prior to, for example, leave or hospital transfer; s 41(5), patients whose restriction orders have come to an end; s 45(A), a 'hybrid order' or 'hospital direction' whereby an individual is sent to a secure hospital after conviction but can then be transferred to prison to serve the remaining period of their sentence; s 47/49, a 'prison transfer with restrictions' whereby someone whose mental health cannot be safely managed in prison is transferred to a secure hospital with restrictions from the Secretary of State/Ministry of Justice; s 36, courts can order someone be placed in secure hospital whilst awaiting or during trial in custody where it is suspected the individual might have a mental disorder; s 48/49, a 'prison transfer with restrictions' for individuals on remand awaiting sentencing.

used, which provides robust test statistics and standard errors, and the Satorra-Bentler scaled (mean adjusted) test statistic (Rosseel et al., 2022). The following GoF statistics are described: robust versions of the Satorra-Bentler correction, χ^2 the comparative fit index (CFI), the Tucker–Lewis index (TLI), Standardized Root Mean Square Residual (SRMR), and the root-mean-square error of approximation (RMSEA). Good fit is suggested with CFI and TLI values above .95, and SRMR below .08, and RMSEA below .06 (Tabachnick & Fidell, 2013). GoF statistics of the Mokken-derived and Schalast model were compared descriptively.

Wolf et al. (2013) suggest that a CFA of a three-factor model with moderate factor loadings (.65) can be reliably performed with sample sizes of $N = 230$ or $N = 140$ where there are either 3-4 or 6 items per factor respectively. As we have $N = 233$, we propose that the CFA was adequately powered.

SPSS v 27 was used for the regressions. Four hierarchical regression models were computed, with EssenCES total and domain scores as outcome variables. In each model, level of security was entered first as a control variable given this is not an individual-level trait and then the remaining variables added simultaneously (following Dickens et al., 2014). Multivariate outliers were assessed with the Mahalanobis Distance statistic, and outliers removed where $p < 0.001$ (Tabachnick & Fidell, 2013). One multivariate outlier case was removed from the regressions ($\chi^2 = 36.71867$, $p < 0.001$). Multicollinearity was assessed via the tolerance values for each predictor, with 0.1 considered indicative of multicollinearity (Tabachnick & Fidell, 2013). No multicollinearity was observed (all tolerance values were 0.688 or higher). Independence of residuals of predictors was confirmed (Durbin-Watson values ranged from 1.57 to 1.98). Scatterplots were used to investigate homoscedasticity; these suggested predictors were homoscedastic in all regressions with the exception of the Therapeutic Hold regression model.

Two *post-hoc* power analyses using the G*Power software suggested that a linear multiple regression, fixed model, R^2 deviation from zero, $N = 170$ (after pairwise exclusion of missing values and recoding of dependent variables) with six predictor variables and f^2 effect sizes 0.03 (smallest effect size reported) and 0.1 (largest effect size reported), yielded power scores

of 0.33 and 0.88 (see Table 4) (Faul et al., 2007). This suggests that our analyses for the regression models with very small (e.g., <0.09) effect sizes should be seen as indicative, but sufficiently powered to detect effect sizes larger than this (e.g., >0.09).

Results

N = 241 patients consented to participate in the project. Data for eight participants were excluded from the analysis as these patients withdrew consent to participate or did not attempt to complete the study questionnaires in full. Findings are based on the N = 233 remaining participants. The mean age of the sample was 39.4 years (S.D. 10.8, N = 233) and the median length of stay in current setting was 19 months (Q1 = 9, Q2 = 3, N = 230). Patient profiles are described in full in Table 1. Table 2 describes the EssenCES total and domain mean scores, standard deviations, and normality scores.

The results of the Mokken scaling confirmed that the EssenCES does not measure a unidimensional structure; 8 of the 15 items would not load onto a unidimensional structure with Loevinger coefficients >0.3. Mokken scaling indicated that a two-factor structure derived from 12 items was most appropriate. The 3 excluded items belong to the Therapeutic Hold domain in the Schalast model. Mokken factor 1 consisted of an amalgamation of items from the Therapeutic Hold and Patient Cohesion domains. Mokken factor 2 replicated the Experienced Safety

domain of the Schalast EssenCES model. Both of these scales met the relevant assumptions of Mokken scaling. See Table 3 for a summary of these items, Hi and H coefficients.

Internal consistency (α) scores for the factors/domains in the Mokken-derived and Schalast models were satisfactory (≥0.7), as were CITC scores for items in both models (>0.3). Standardized loading coefficients for all items on all factors/domains for both the Mokken-derived and Schalast models are also satisfactory (>0.3). These results are depicted in Tables 3 and 4.

CFA analysis of the 12-item Mokken two-dimensional model indicated good model fit. GoF statistics were as follows: CFI = .930; TLI = .913; SRMR = 0.066; RMSEA = 0.083; Satorra-Bentler correction = 1.313; and $\chi^2 = 118.329$ (53), $p < .001$. CFA analysis of the 15-item Schalast three-dimensional model indicated good model fit. GoF statistics were as follows: CFI = .944; TLI = .933; SRMR = 0.077; RMSEA =

Table 2. EssenCES total and domain mean scores, standard deviations, and normality scores.

	Mean total score	S.D.	Shapiro-Wilk p-value
EssenCES total	35.2	10.4	.084
Therapeutic Hold	12.2	4.3	.001
Experienced safety	12.8	5.0	<.001
Patient Cohesion	10.1	5.2	<.001

Table 4. EssenCES item CFA standardized loadings, CITC scores by domain, and domain internal consistency (α) scores if item removed values.

Item	CFA Standardized loading	CITC	α if item removed
Therapeutic Hold			**α = 0.699**
EssenCES4	0.735	.603	.583
EssenCES7	0.711	.629	.578
EssenCES10	0.841	.340	.695
EssenCES13	0.807	.348	.696
EssenCES16	0.849	.384	.678
Experienced safety			**α = 0.814**
EssenCES3	0.666	.578	.786
EssenCES6	0.761	.674	.756
EssenCES9	0.821	.720	.742
EssenCES12	0.684	.601	.779
EssenCES15	0.499	.453	.821
Patient Cohesion			**α = 0.889**
EssenCES2	0.780	.677	.877
EssenCES5	0.793	.649	.885
EssenCES8	0.382	.783	.853
EssenCES11	0.429	.761	.859
EssenCES14	0.469	.792	.851

Table 3. Mokken-derived EssenCES two factor model with standardized coefficients, Hi, H, Crit, α if item removed, and CITC scores.

	Standardized coefficient	Hi (S.E.)	H (S.E.)	Crit	CITC	α if item removed
Factor 1	-		**0.547 (0.032)**			**α = .881**
EssenCES2	0.735	0.543 (0.041)		0	.659	.865
EssenCES4	0.536	0.471 (0.047)		0	.561	.877
EssenCES5	0.720	0.565 (0.036)		0	.698	.860
EssenCES7	0.520	0.463 (0.052)		0	.544	.878
EssenCES8	0.838	0.610 (0.032)		0	.759	.852
EssenCES11	0.795	0.565 (0.040)		0	.693	.860
EssenCES14	0.843	0.608 (0.033)		0	.760	.851
Factor 2			**0.492 (0.037)**			**α =.814**
EssenCES3	0.665	0.475 (0.044)		0	.578	.786
EssenCES6	0.762	0.537 (0.042)		0	.674	.756
EssenCES9	0.821	0.564 (0.036)		0	.720	.742
EssenCES12	0.681	0.499 (0.046)		0	.601	.779
EssenCES15	0.501	0.382 (0.054)		0	.453	.821

Table 5. Linear multiple regression models of Schalast EssenCES total and domain scores.

Variable	Patient Cohesion		Experienced Safety		Therapeutic Hold		Total Scale	
	Model fit: $R^2 = .027$; F (6,172) = 0.759, p = .603, $F^2 = 0.028$		Model fit: $R^2 = .088$; F (6,172) = 2.659, p = .017, $F^2 = 0.097$		Model fit: $R^2 = .055$; F (6,172) = 1.623, p = .144, $F^2 = 0.058$		Model fit: $R^2 = .06$; F (6,172) = 1.754, p = .112, $F^2 = 0.064$	
	Unstandardised Coefficient	Sig.	Unstandardised Coefficient	Sig.	Unstandardised Coefficient	Sig.	Unstandardised Coefficient	Sig.
Security level[1]	-.576	.357	-.633	.281	-.297	.558	−1.483	.233
Age	-.010	.787	.035	.327	.014	.637	.039	.608
Length of stay[2]	.004	.760	-.010	.396	.007	.490	.001	.976
Ethnicity[3]	1.295	.143	−2.082	.013	-.615	.390	−1.448	.408
Primary psychiatric diagnosis[4]	.439	.677	1.918	.054	1.853	.031	4.345	.039
Legal section[5]	.665	.516	.497	.604	1.318	.113	2.488	.221

Notes. [1] = Security level (low, medium, high); [2] = Length of stay in current institution (months); [3] = Ethnicity (white or black, Asian and minority ethnic); [4] = Primary psychiatric diagnosis (personality disorder or mental illness); [5] = Legal section (civil or forensic); N = 170.

Table 6. Linear multiple regression models of Mokken two-dimensional EssenCES total and scale scores.

Variable	Mokken 1		Mokken 2 (same as ES)		Total Scale	
	Model fit: $R^2 = .031$; F (6,172) = 0.879, p = .512, $F^2 = 0.032$		Model fit: $R^2 = .088$; F (6,172) = 2.659, p = .017, $F^2 = 0.097$		Model fit: $R^2 = .051$; F (6,172) = 1.487, p = .186, $F^2 = 0.054$	
	Unstandardised Coefficient	Sig.	Unstandardised Coefficient	Sig.	Unstandardised Coefficient	Sig.
Security level[1]	-.925	.249	-.633	.281	−1.552	.159
Age	.000	.998	.035	.327	.034	.604
Length of stay[2]	.008	.590	-.010	.396	-.001	.951
Ethnicity[3]	1.162	.304	−2.082	.013	-.959	.536
Primary psychiatric diagnosis[4]	1.027	.447	1.918	.054	2.986	.108
Legal section[5]	1.305	.320	.497	.604	1.808	.314

Notes. [1] = Security level (low, medium, high); [2] = Length of stay in current institution (months); [3] = Ethnicity (white or black, Asian and minority ethnic); [4] = Primary psychiatric diagnosis (personality disorder or mental illness); [5] = Legal section (civil or forensic); N = 170.

0.062; Satorra-Bentler correction = 1.289; and χ^2 = 146.935 (87), $p < .001$.

The CFI, TLI and RMSEA GoF statistics were preferable in the Schalast model, SRMR was preferable in the Mokken-derived model. Based on this, we can conclude that on the whole, the Schalast model better explains variance in the data.

Regressions for the Schalast and Mokken-derived models are depicted in Tables 5 and 6. In relation to the Schalast model; the regressions demonstrate that the variables included in our models have little explanatory effect on the domain and total scores of the Schalast model of the EssenCES and models are generally non-significant. The R^2 values range from .027 to .088, and the F^2 effect sizes are notably small. The model with Experienced Safety as dependent variable was significant, the other models were not. In this model, ethnicity was the only significant predictor with black, Asian and minority ethnic patients rating the Experienced Safety domain two points lower on average than white patients. There was a trend toward diagnosis being a significant predictor, with patients diagnosed with mental illness giving higher Experienced Safety ratings. Primary psychiatric diagnosis was a significant predictor in the Therapeutic Hold and total scale models, with patients diagnosed with a mental illness reporting higher scores.

A similar conclusion can be made for the Mokken-derived model (see Table 6). R^2 values ranged from 0.031 to 0.088. Mokken scale 2 returns the same findings as the regression of the Experienced Safety domain analysis above as these factors/domains consist of the same items.

Discussion

Previous research emphasises the importance of ward atmosphere for inpatient psychiatric settings (Tonkin, 2016). Contemporary treatment philosophies and models highlight the role of community and collaboration in secure or criminal justice-related settings, e.g., the Safewards model and Therapeutic Communities (Bowers, 2014; Magor-Blatch et al., 2014). Our research has sought to add to the relatively limited body of evidence pertaining to ward atmosphere in secure settings by investigating the dimensionality of the EssenCES questionnaire and it's correlations with patient-level variables. This research, therefore, fills important gaps in the literature.

Our goodness of fit analysis suggests that the three-dimensional EssenCES structure is superior to a two-dimensional structure derived from Mokken scaling methods. As this analysis was conducted on the established 15-item, three-dimensional structure, rather than an initial pool of items this finding might not be very surprising. It does suggest that the EssenCES questionnaire should retain a three-factor model, however. This is an important finding as it suggests the measurement of these three dimensions, rather than the structure of the EssenCES might be improved. We found evidence to support the continued use of the Experienced Safety domain as it currently is; we found a high internal consistency score, and this domain appeared in both the CFA and Mokken scaling analyses.

Previous research has highlighted the need to revise some items in the Therapeutic Hold and Patient Cohesion domains and some changes have been made in previous studies, e.g. PC3 was revised as outlined in the introduction to this paper (Howells et al., 2009; Milsom et al., 2014). Our analysis would support possible further revisions of these domains, especially Therapeutic Hold. This domain only met the standard acceptable internal consistency threshold (>0.7) when rounded, and the three EssenCES items excluded from the Mokken-derived model came from this domain (items: 10, 13, 16), potentially explaining the two-factor structure indicated by the Mokken analysis. This finding might suggest that these three items could be reworded (Milsom et al., (2014) found that item 13 loaded onto two domains, indicating need for revision) or that the items in the Therapeutic Hold domain reflect a latent construct slightly different yet highly related to the ward atmosphere latent construct being tapped into by the two other EssenCES domains. These had better internal consistency scores and none of their items were dropped by the Mokken analysis.

Our findings do not enable us to answer precisely why this might be the case, though we can hypothesize several possible reasons which can be investigated in future research. One possibility is that patients might be thinking about therapeutic alliance in the context of relationships and interactions with individual staff, which are distinct from an assessment of collective ward atmosphere. A second possibility is that patients view interactions with staff through a 'us and them' lens, with the domains Experienced Safety and Patient Cohesion more closely aligning with the 'us', and the Therapeutic Hold domain reflecting the 'them'. Third, the three items excluded by the Mokken analysis (items: 10, 13, 16) might be too ambiguously worded creating a lack of conceptual clarity. Indeed, it might be that the original translation of these items from German to English could be improved, which would enhance the validity of data collected using the EssenCES.

Our research adds to the extant literature by investigating to what extent level of security, age, length of stay in current institution, ethnicity, primary psychiatric diagnosis and legal section were associated with ward atmosphere scores. Broadly speaking, these factors were not significantly linked to EssenCES domains or total scores. Our finding that level of security was not a significant predictor differs from the findings of Dickens et al. (2014). However, there were several exceptions when controlling for the effects of all these patient-level variables. Patients from black, Asian and minority ethnic backgrounds were more likely to report lower Experienced Safety scores than their counterparts, and patients with a primary diagnosis of mental illness were more likely to report higher Therapeutic Hold and total scale scores.

Our analyses found that ethnicity was a significant predictor of Experienced Safety scores, with black, Asian and minority ethnic patients rating this domain two points lower on average than white patients. This is an important and novel finding, as the relationship between social climate and ethnicity has not previously been examined (as far as the authors are aware). Our findings do, however, support the wider literature where black, Asian and minority ethnic individuals have reported significantly worse experiences than white individuals in secure settings, including forensic mental health (Hui, 2017) and prison settings (Chief Inspector of Prisons, 2020). Furthermore, these findings come within a wider context where black, Asian and minority ethnic individuals experience an increased risk of involuntary psychiatric care, longer stays within secure services and higher rates of readmission compared to white individuals (Arya et al., 2021). Given the established relationship between social climate, institutional behaviors (such as aggression and adjudications) and long-term treatment outcomes (e.g., reoffending), it is important that secure settings seek to understand and reduce the apparent disparity in experience between black, Asian and minority ethnic and white groups within forensic mental health settings. Failure to address this disparity might further fuel the disproportionately negative experiences and outcomes of black, Asian and minority ethnic groups who come into contact with the criminal justice and forensic mental health estates.

Primary psychiatric diagnosis was a significant predictor in the Therapeutic Hold and total scale regression models, with patients diagnosed with a mental illness reporting higher scores than those diagnosed with personality disorder. The challenges posed by individuals diagnosed with personality disorder in secure settings is well-known, with these individuals generally considered more difficult to manage, treat and interact with (e.g., see Freestone et al., 2015). It is perhaps unsurprising, therefore, that the current study found that individuals with personality disorder had more negative perceptions of the social climate than those with mental illness. An exception to this, however, was the lack of a significant relationship between primary psychiatric diagnosis and Experienced Safety scores in our study. This finding stands in contrast to the findings of Dickens et al. (2014), who found higher safety scores for those diagnosed with personality disorder. The reason for these disparate findings is unclear. Further research is needed to explore the lived experiences of individuals with a personality disorder diagnosis residing within forensic mental health settings to understand if/how their perceptions of social climate can be improved.

The literature suggests that risk-related, dynamic, mood and symptom-related factors might be better predictors of ward atmosphere than static demographic, clinical and legal factors. Indeed, behavioral disturbance, treatment motivation, treatment engagement, therapeutic alliance, treatment readiness, HCR-20 and PCL-R domain scores have been associated with EssenCES domains and total scores (de Vries et al., 2016; Dickens et al., 2014; Gaab et al., 2020; Long et al., 2011). As we did not measure these in our study, it is possible that our focus on static demographic, clinical and legal factors, is the reason for the low amount of overall variance in EssenCES domain and total scale scores explained.

Future research

Future research should focus on rewording items 10, 13 and 16 and trialing these to see if this leads to improvements in psychometric properties. Classical test theory and item-response theory methods can be used to evaluate the appropriateness of these items. Further research should investigate the relationships between perceptions of ward atmosphere and risk-related, dynamic, mood and symptom-related factors as previous studies indicate that patient experiences of care in general and ward atmosphere specifically

might be influenced by these factors more than static demographic, clinical and legal factors.

Limitations

It is important to note several limitations of the current study. First, this study utilized SPSS' Automatic Imputation Method to replace missing values rather than the scoring guidelines outlined in the EssenCES manual (Schalast & Tonkin, 2016). It might be fruitful to re-score the missing data using Schalast and Tonkin (2016) guidelines and re-run these analyses to determine any differences in findings. Second, the sample size of the Mokken analysis (N = 233) was slightly below the recommended minimum for Mokken scaling of 250 (Watson et al., 2018); and the post-hoc power calculation for the regressions suggest that these analyses should be repeated with larger samples. However, our sample is larger than some previous studies investigating ward atmosphere in forensic settings and thus makes an important contribution to the literature. Third, as noted in the Section titled "Data analysis", there were some issues in the analyses regarding the homoscedasticity of predictors in the regressions. Heteroscedasticity can impact on measures of statistical significance in regression, making them unreliable. Fourth, though most data were collected by researchers who did not conduct the statistical analyses, the lead author, JT, both collected and analyzed data, which could be a source of bias. Finally, too few women were recruited to conduct a comparison with men in this study, so future research should ensure a sufficient number of women are included to conduct a well-powered analysis of difference.

Conclusion

Ward atmosphere is an important element of inpatient psychiatric care. This study sought to investigate the psychometric properties of the EssenCES measure of ward atmosphere in forensic psychiatric settings using item-response theory and classical test theory methods. It also examined the relationship between patient perceptions of ward atmosphere and patient age, length of stay in current institution, level of security, ethnicity, Mental Health Act 1983 section, and mental health diagnosis. Our analysis supports the three-factor structure of the EssenCES but signposts areas for improvement, specifically, revising and retesting items 10, 13 and 16. We found that black, Asian and minority ethnic patients report lower

Experienced Safety domain scores and that patients with a personality disorder diagnosis reported lower Therapeutic Hold domain and EssenCES total scores, when controlling for other variables. We suggest that this is to some extent in line with other literature describing poorer outcomes and experiences for these groups e.g., longer involuntary treatment periods and higher rates of seclusion in UK secure settings (Wessely, 2018).

ORCID

Jack Tomlin http://orcid.org/0000-0002-7610-7918
Matthew Tonkin http://orcid.org/0000-0001-6913-7759

References

Arya, D., Connolly, C., & Yeoman, B. (2021). Black and minority ethnic groups and forensic mental health. *BJPsych Open*, *7*(S1), S123–S123. https://doi.org/10.1192/bjo.2021.357

Auty, K. M., & Liebling, A. (2019). Exploring the relationship between prison social climate and reoffending. *Justice Quarterly*, *37*(2), 358–381. https://doi.org/10.1080/07418825.2018.1538421

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, *314*(7080), 572–572. https://www.bmj.com/content/314/7080/572 https://doi.org/10.1136/bmj.314.7080.572

Bowers, L. (2014). Safewards: A new model of conflict and containment on psychiatric wards. *Journal of Psychiatric and Mental Health Nursing*, *21*(6), 499–508. https://doi.org/10.1111/jpm.12129

Brunt, D., & Rask, M. (2007). Ward atmosphere - The scarlet pimpernel of psychiatric settings? *Issues in Mental Health Nursing*, *28*(6), 639–655. https://doi.org/10.1080/01612840701354513

Chief Inspector of Prisons, H. (2020). *Thematic report by HM Inspectorate of Prisons Minority ethnic prisoners' experiences of rehabilitation and release planning A thematic review by HM Inspectorate of Prisons*. http://www.justiceinspectorates.gov.uk/hmiprisons/

de Vries, M. G., Brazil, I. A., Tonkin, M., & Bulten, B. H. (2016). Ward climate within a high secure forensic psychiatric hospital: perceptions of patients and nursing staff and the role of patient characteristics. *Archives of Psychiatric Nursing*, *30*(3), 342–349. https://doi.org/10.1016/j.apnu.2015.12.007

Dickens, G., Suesse, M., Snyman, P., & Picchioni, M. (2014). Associations between ward climate and patient characteristics in a secure forensic mental health service. *The Journal of Forensic Psychiatry & Psychology*, *25*(2), 195–211. https://doi.org/10.1080/14789949.2014.903505

Doyle, P., Quayle, E., & Newman, E. (2017). Social climate in forensic mental health settings: A systematic review of qualitative studies. *Aggression and Violent Behavior*, *36*, 118–136. https://doi.org/10.1016/j.avb.2017.06.008

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Freestone, M. C., Wilson, K., Jones, R., Mikton, C., Milsom, S., Sonigra, K., Taylor, C., & Campbell, C. (2015). The impact on staff of working with personality disordered offenders: A systematic review. *Plos One*, *10*(8), e0136378. https://doi.org/10.1371/journal.pone.0136378

Gaab, S., Brazil, I. A., de Vries, M. G., & Bulten, B. H. (2020). The relationship between treatment alliance, social climate, and treatment readiness in long-term forensic psychiatric care: An explorative study. *International Journal of Offender Therapy and Comparative Criminology*, *64*(9), 1013–1026. https://doi.org/10.1177/0306624X19899609

Hardouin, J.-B., Bonnaud-Antignac, A., & Sébille, V. (2011). Nonparametric item response theory using Stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, *11*(1), 30–51. https://doi.org/10.1177/1536867X1101100102

Hare Duke, L., Furtado, V., Guo, B., & Völlm, B. A. (2018). Long-stay in forensic-psychiatric care in the UK. *Social Psychiatry and Psychiatric Epidemiology*, *53*(3), 313–321. https://doi.org/10.1007/s00127-017-1473-y

Howells, K., Tonkin, M., Milburn, C., Lewis, J., Draycot, S., Cordwell, J., Price, M., Davies, S., & Schalast, N. (2009). The EssenCES measure of social climate: A preliminary validation and normative data in UK high secure hospital settings. *Criminal Behaviour and Mental Health : CBMH*, *19*(5), 308–320. https://doi.org/10.1002/cbm.745

Hui, A. M. L. (2017). The lived experiences of Black and Minority Ethnic patients within a high security hospital. *Health Education England*. https://nottingham-repository.worktribe.com/output/900991/the-lived-experiences-of-black-and-minority-ethnic-patients-within-a-high-security-hospital

Liebling, A., Hulley, S., & Crewe, B. (2014). Conceptualising and measuring the quality of prison life. In D. Gadd, S. Karstedt & S. F. Messner (Eds.), *The SAGE handbook of criminological research methods* (pp. 358–372). SAGE Publications Ltd. https://doi.org/10.4135/9781446268285.n24

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of

the *American Statistical Association*, 83(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Long, C. G., Anagnostakis, K., Fox, E., Silaule, P., Somers, J., West, R., & Webster, A. (2011). Social climate along the pathway of care in women's secure mental health service: Variation with level of security, patient motivation, therapeutic alliance and level of disturbance. *Criminal Behaviour and Mental Health : CBMH*, 21(3), 202–214. https://doi.org/10.1002/cbm.791

Magor-Blatch, L., Bhullar, N., Thomson, B., & Thorsteinsson, E. (2014). A systematic review of studies examining effectiveness of therapeutic communities. *Therapeutic Communities: The International Journal of Therapeutic Communities*, 35(4), 168–184. https://doi.org/10.1108/TC-07-2013-0024

Maguire, T., Ryan, J., Fullam, R., McKenna, B., M, T., R, J., & F, R. (2018). Evaluating the introduction of the safewards model to a medium- to long-term forensic mental health ward. *Journal of Forensic Nursing*, 14(4), 214–222. https://doi.org/10.1097/JFN.0000000000000215

Milsom, S. A., Freestone, M., Duller, R., Bouman, M., & Taylor, C. (2014). Factor structure of the Essen Climate Evaluation Schema measure of social climate in a UK medium-security setting. *Criminal Behaviour and Mental Health : CBMH*, 24(2), 86–99. https://doi.org/10.1002/cbm.1878

Moos, R. H. (1975). *Evaluating correctional and community settings*. Wiley.

Moos, R. H., & Houts, P. S. (1968). Assessment of the social atmospheres of psychiatric wards. *Journal of Abnormal Psychology*, 73(6), 595–604. https://doi.org/10.1037/h0026600

O' Flynn, P., O' Regan, R., O' Reilly, K., & Kennedy, H. (2018). Predictors of quality of life among inpatients in forensic mental health: Implications for occupational therapists. *BMC Psychiatry*, 18(1), 16. https://doi.org/10.1186/s12888-018-1605-2

Rosseel, Y., Jorgensen, T., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2022). *Package "lavaan."*

Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management*, 54(2), 189–203. https://doi.org/10.1016/j.im.2016.06.005

Schalast, N., & Groenewald, I. (2009). Ein Kurzfragebogen zur Einschätzung des sozialen Klimas im Strafvollzug. Erste Befunde auf Abteilungen des Regelvollzugs und der Sozialtherapie. In *Drogen - Sucht - Kriminalität* (pp. 329–352). https://www.deutsche-digitale-bibliothek.de/item/7AIPEHDIUWM4JFDYUIDCPGL7NH3DU6L5

Schalast, N., Redies, M., Collins, M., Stacey, J., & Howells, K. (2008). EssenCES, a short questionnaire for assessing the social climate of forensic psychiatric wards. *Criminal Behaviour and Mental Health : CBMH*, 18(1), 49–58. https://doi.org/10.1002/cbm.677

Schalast, N., & Tonkin, M. (2016). *The Essen climate evaluation schema EssenCES: A manual and more*. Hogrefe Publishing.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tomlin, J., Bartlett, P., & Völlm, B. (2018). Experiences of restrictiveness in forensic psychiatric care: Systematic review and concept analysis. *International Journal of Law and Psychiatry*, 57, 31–41. https://doi.org/10.1016/j.ijlp.2017.12.006

Tomlin, J., Völlm, B., Furtado, V., Egan, V., & Bartlett, P. (2019). The forensic restrictiveness questionnaire: Development, validation, and revision. *Frontiers in Psychiatry*, 10, 1–10. https://doi.org/10.3389/fpsyt.2019.00805

Tonkin, M. (2016). A review of questionnaire measures for assessing the social climate in prisons and forensic psychiatric hospitals. *International Journal of Offender Therapy and Comparative Criminology*, 60(12), 1376–1405. https://doi.org/10.1177/0306624X15578834

Tonkin, M., Howells, K., Ferguson, E., Clark, A., Newberry, M., & Schalast, N. (2012). Lost in translation? Psychometric properties and construct validity of the English Essen Climate Evaluation Schema (EssenCES) social climate questionnaire. *Psychological Assessment*, 24(3), 573–580. https://doi.org/10.1037/a0026267

van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, 20(11), 1–19.

van der Ark, L. A., Koopman, L., Hendrik Straat, J., & van den Bergh, D. (2021). *Package "Mokken."*

van der Eijk, C., & Rose, J. (2015). Risky business: Factor analysis of survey data–assessing the probability of incorrect dimensionalisation. *PLoS One*, 10(3), e0118900. https://doi.org/10.1371/journal.pone.0118900

van der Helm, P., Stams, G. J., & van der Laan, P. (2011). Measuring group climate in prison. *The Prison Journal*, 91(2), 158–176. https://doi.org/10.1177/0032885511403595

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLoS Medicine*, 4(10), e296. https://doi.org/10.1371/JOURNAL.PMED.0040296

Watson, R., Egberink, I. J. L., Kirke, L., Tendeiro, J. N., & Doyle, F. (2018). What are the minimal sample size requirements for mokken scaling? An empirical example with the Warwick- Edinburgh Mental Well-Being Scale. *Health Psychology and Behavioral Medicine*, 6(1), 203–213. https://doi.org/10.1080/21642850.2018.1505520

Wessely, S. (2018). *Modernising the Mental Health Act: Increasing choice, reducing compulsion. Final report of the Independent Review of the Mental Health Act 1983* (Issue December). www.gov.uk/dhsc

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73(6), 913–934. https://doi.org/10.1177/0013164413495237

World Health Organization. (1953). *Expert Committee on Mental Health : Third report [of a meeting held in Geneva, 24-29 November 1952]*.