



Universiteit
Leiden

The Netherlands

Measuring prison climate across contexts: lessons from administering the Prison Climate Questionnaire in the USA

Tiem, B. van; Nieuwbeerta, P.; Andersen, S.N.; Hyatt, J.M.; Palmen, J.M.H.

Citation

Tiem, B. van, Nieuwbeerta, P., Andersen, S. N., Hyatt, J. M., & Palmen, J. M. H. (2025). Measuring prison climate across contexts: lessons from administering the Prison Climate Questionnaire in the USA. *European Journal Of Criminology*, 22(5), 843-874.
doi:10.1177/14773708241290036

Version: Publisher's Version

License: [Creative Commons CC BY-NC 4.0 license](https://creativecommons.org/licenses/by-nc/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4283404>

Note: To cite this publication please use the final published version (if applicable).

Measuring prison climate across contexts: Lessons from administering the Prison Climate Questionnaire in the USA

European Journal of Criminology

2025, Vol. 22(5) 843–874

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14773708241290036

journals.sagepub.com/home/euc**Britte van Tiem** 

University of Pennsylvania, USA

Paul Nieuwbeerta

Leiden University, the Netherlands

Synøve N. Andersen

University of Oslo, Norway

Jordan M. Hyatt

Drexel University, USA

Hanneke Palmen

Leiden University, the Netherlands

Abstract

Prison climate surveys are uniquely positioned to identify how the quality of prison life differs both within and between institutions. However, much of this comparative potential remains unrealized, in part because of insufficient evidence that existing survey instruments are suitable for use in different contexts and that survey data can be reliably compared across contexts. In this paper, we explore the suitability of the Prison Climate Questionnaire (PCQ), originally developed in the Netherlands (NL), for use in the USA by assessing its factor structure, reliability, construct validity, and criterion validity using survey data from Pennsylvania ($N = 632$). We compare our findings

Corresponding author:

Synøve N. Andersen, University of Oslo, Department of Sociology and Human Geography, Postboks 1096 Blindern, 0317 Oslo, Norway.

Email: synovena@uio.no

with previously published psychometric results from the NL. Results of psychometric analyses show several striking similarities between the two countries, pointing to areas where the survey might be improved. While the PCQ shows potential for stand-alone use in the USA, further work would be required to use the tool in explicitly comparative research. We draw lessons from this collaboration to inform future efforts to develop standardized prison climate surveys more suitable for comparative analyses of prison climate in Europe and beyond.

Keywords

Comparative research, prison, prison climate, psychometrics, survey research

Introduction

Correctional environments can vary greatly in terms of their quality of life. People who have spent time in prison can often provide vivid descriptions of the meaningful differences across and within institutions. Differences in what it means to 'live' in a prison are likely even greater between countries. How do correctional environments differ in the eyes of incarcerated individuals? Over the past 50 years, researchers have administered surveys designed to measure constructs such as the 'quality of prison life' (Liebling and Arnold, 2004), 'prison climate' (e.g., Bosma et al., 2020; Schalast et al., 2008; van der Helm et al., 2011), 'correctional environments' (Moos, 1974), and 'prison performance' (Camp et al., 2002) to shed light on this question. While the constructs measured by these surveys have been defined in different ways, these surveys all employ composite scales to measure dimensions of the social and/or physical prison environment as experienced by incarcerated people, such as safety, personal relationships, and access to support and meaningful activity. Hereafter, we will use the term 'prison climate research' to refer to the body of work that is based on these surveys.¹

Prison climate research has typically focused on three main empirical inquiries. First, researchers have sought to document and describe the first-hand experiences of people who live or work in prison. This type of inquiry has both theoretical and ethical underpinnings: the scientific study of the inner life of totalitarian institutions has inspired criminological inquiry for centuries, and international human rights bodies continually stress the importance of how people deprived of their liberty are treated (Bouloukos and Dammann, 2001; van Zyl Smit, 2010). Second, prison climate research has compared prison climates across institutions, places, and times. Researchers have used prison climate instruments to compare prisons either at the aggregate level (Auty and Liebling, 2020; Liebling and Arnold, 2004; van Ginneken and Nieuwbeerta, 2020) or along specific dimensions such as the prison size (Johnsen and Granheim, 2011), prison security level (Long et al., 2011), whether prisons are publicly or privately run (Armstrong and MacKenzie, 2003; Camp et al., 2002; Crewe et al., 2015), or whether they are therapeutically oriented (Day et al., 2012; Schalast and Laan, 2017). Researchers have also started to compare scores on climate survey measures between countries (Crewe et al., 2022; Mjåland et al., 2021; Neubacher et al., 2021; Ross et al., 2008). Third, prison climate research has sought to better understand how experiences of prison environments relate to in-prison and post-release outcomes. For instance, better institutional climates are thought to be related to the effectiveness of health and educational programs in prison and

to affect both in-prison and post-release outcomes ranging from misconduct to well-being and recidivism (e.g., Auty and Liebling, 2020, 2024; Gonçalves et al., 2016; Harding, 2014; Schubert et al., 2012; van Ginneken and Palmen, 2022).

Prison climate research relies on the fundamental assumption that the survey scales we use are valid and reliable measures of the underlying latent constructs they intend to capture. In technical terms, this implies that data obtained when administering a given survey to different groups of respondents—in different contexts or at different times—demonstrate sound psychometric properties across administrations (i.e., internal consistency, factor structure, construct validity, and reliability). Such psychometric properties are best established through an iterative process, where items in an instrument are tested, adapted, and retested until the instrument properties are stable (Fauskanger et al., 2012; Smith, 2005). In practice, most existing prison climate surveys have undergone limited psychometric testing (Tonkin, 2016). If we wish to *compare* scores across contexts or groups, we also need formal evidence of measurement invariance. That is, we want to make sure that a comparison of scores across groups of respondents reflects true differences rather than measurement differences (Leitgöb et al., 2023). Like much comparative research in the social sciences, prison climate studies routinely compare scores across groups yet rarely formally test for measurement invariance (Leitgöb et al., 2023).² The fact that prison climate surveys tend to be designed for one context and then adapted to others—instead of being explicitly designed for comparative purposes—also means that they are less likely to achieve measurement invariance in practice (Fitzgerald and Jowell, 2010; Harkness, 2011; Leitgöb et al., 2023).

In this study, we add to the still nascent evidence on the psychometric properties of the Prison Climate Questionnaire (PCQ), a relatively new prison climate survey that was first used in the Netherlands (NL) in 2016 (Beijersbergen, 2016). The instrument has since been administered repeatedly to the full Dutch prison population,³ and several groups of researchers have started to administer it to prison populations in other countries in Europe and beyond (Johnsen et al., 2023; Maes et al., 2023; Ouaknine, 2023). To date, however, the psychometric properties of the PCQ have only been tested once, using data from the PCQ's first administration in the NL in 2017 (Bosma et al., 2020). We aim to make three main contributions to the literature. First, we conduct a psychometric analysis of the PCQ using data collected in one prison in Pennsylvania (PA) (USA) ($N=632$) and assess the instrument's suitability for use in this context. Second, we compare the results from PA to previously published psychometric results from the NL and draw on both shared and divergent psychometric patterns in the two countries to identify areas where the instrument can be improved. Third and finally, we link specific observations from this exercise to general guidance on survey design from the comparative survey literature. We hope that these insights might be used to inform the future development of standardized prison climate surveys more suitable for comparative analyses of prison climate in Europe and beyond.

Prison climate instruments, psychometric properties, and the PCQ

In the 1970s and 1980s, amidst early enthusiasm for the study of the human organizational environment (Schneider et al., 2013), researchers developed measures to assess

the 'climate' in prisons through surveys administered to incarcerated individuals (Moos, 1974; Saylor, 1984; Wright, 1985). One of the pioneering prison climate instruments, the Correctional Institutions Environment Scale (CIES) (Moos, 1974), was widely used until the early 2000s but has been largely abandoned in recent years after researchers flagged concerns about its psychometric properties (Liebling and Arnold, 2004; Tonkin, 2016).

In the half century that has passed since the first prison climate instrument was developed, researchers have developed several closely related survey instruments.⁴ Among all available instruments, the Measuring Quality of Prison Life (MQPL) instrument (Liebling et al., 2012) is one of the most widely used. Importantly, the MQPL was not conceptualized as a psychometric instrument, and research using or drawing on the MQPL has typically emphasized 'fused' approaches in which psychometric analysis is used as a supportive tool alongside qualitative approaches (Crewe et al., 2022; Neubacher et al., 2021). The instrument has been revised and adapted to suit the countries and prisons in which it is used (Neubacher et al., 2021). This makes the MQPL well positioned to capture local processes and meanings but reduces comparability across survey administrations.

In 2016, a review article by Matthew Tonkin identified 12 questionnaire-based measures of social climate, including the CIES, the MQPL, and 10 others. Tonkin highlighted that while prison climate instruments were routinely used by both researchers and practitioners, the psychometric properties of most of these prison climate instruments had not been sufficiently tested to 'justify their routine use' (Tonkin, 2016: 1376). Tonkin's (2016) review identified the Essen Climate Evaluation Schema (EssenCES) (Schalast et al., 2008) as the instrument with the most consistent empirical support for its psychometric qualities.⁵ As a short instrument of 15 items, the EssenCES assesses the *social* climate in prison: the instrument's three scales measure whether the climate is perceived as safe, supportive, and cohesive. Its items have been adjusted following psychometric testing and revised versions of the instrument have been retested in subsequent studies in a range of countries and populations (Day et al., 2011, 2012; Howells et al., 2009; Schalast and Laan, 2017; Siess and Schalast, 2017; Tonkin et al., 2012). The EssenCES was originally designed for forensic psychiatric wards, but it has been adapted for general prison populations, and there is evidence to suggest that the instrument's psychometric properties are stable across both these settings (Tonkin et al., 2012).

The PCQ was designed in 2016 after an extensive review of the literature on prison climate, including Tonkin's then-recent review (Beijersbergen, 2016), before it was further developed in 2017 (Bosma et al., 2020). The survey forms part of the Dutch Life in Custody (LIC) study, which is led by researchers at Leiden University with support from the Dutch Custodial Institutions Agency (in Dutch, the Dienst Justitiële Inrichtingen). The survey's official aim is to help the prison service monitor prison performance as well as to facilitate academic research on prison climate (Bosma et al., 2020). Three of the scales in the PCQ—prisoner relationships, prisoner–staff relationships, and safety—are closely related to the EssenCES' social climate dimensions. In addition, the PCQ also captures incarcerated people's perceptions and experience of other conditions in which they are confined, including individuals' access to and satisfaction with food, visits, healthcare, and rehabilitative and recreation activities. The scope of the PCQ more closely resembles that of the MQPL. Unlike the MQPL, however, it is explicitly

designed to be a psychometric instrument and uses little local, context-specific vernacular.⁶

The PCQ covers six primary domains of prison climate that are measured using 64 items in 14 scales. Each of the scales is composed of 3–8 items that are rated on a five-point Likert scale, with responses ranging from ‘strongly disagree’ to ‘strongly agree’. The conceptual background to the six domains in the survey—relationships in prison, safety and order, contact with the outside world, facilities (cell, shop, complaints), meaningful activities, and autonomy—is described in detail in van Ginneken et al. (2018).⁷ Scale items are presented in blocks to limit the demand the survey places on an individual’s reading capacity. A full list of scales and associated items in the prison climate scales of the original PCQ are provided in the second column of Table A1.⁸

Prison environments in the NL and PA

Before we address the two survey administrations in the NL and PA in more detail, we discuss some of the broader differences in prison conditions and correctional policy in the NL and PA that pertain most directly to our discussion of prison climate.

With 17.5 million inhabitants, the population of the NL is only slightly larger than that of the American state of PA, which has a population of 13 million.⁹ With an incarceration rate of 53.9 per 100,000 individuals in 2021, the NL had one of the lowest incarceration rates in Western Europe (Aebi et al., 2022). While annual estimates of PA’s jail, prison, and federal populations are harder to obtain, available data suggest that PA’s incarceration rate is at least 10 times that of the NL.¹⁰

Not only does the NL send fewer individuals to prison, but the country also imprisons people for shorter periods of time. In PA’s state prisons, 5336 individuals, or 16.8% of the total prison population, are serving a sentence of natural life (Kuba, 2021). The NL currently has about 43 life-sentenced incarcerated individuals, which corresponds to far less than 1% of its total prison population. In the NL, less than 10% of convicted individuals leaving prison have spent more than a year inside, and less than 5% have spent more than 2 years incarcerated. In contrast, it is the norm for individuals leaving a State Prison in PA to have spent more than 2 years behind bars.¹¹

Prison conditions in the NL and PA differ in numerous ways. PA’s 23 State Correctional Institutions (SCIs) have an official capacity of 43,957 beds, varying from around 1000 to over 3000 beds per institution.¹² In total, the NL’s 26 penitentiary institutions have around 8000 beds, varying from around 200 to 1000 beds per institution. Living units in PA’s SCIs are generally also much larger than in the NL, with some housing almost 300 individuals.¹³ The ratio of staff to incarcerated people varies but having one officer supervising 120 people on a single housing unit is not uncommon. In the NL, most units hold no more than 30 individuals and are staffed by at least two penitentiary workers.

In PA SCIs, double celling is the default among the general population; only individuals with select health conditions or security needs are allocated single cells. In contrast, most people in Dutch prisons live in their own cell.¹⁴ Where incarcerated individuals in the NL can wear their own clothes, correctional uniforms must be worn in PA. In both the NL and PA, incarcerated individuals are generally allowed four in-person visits a month for a minimum of 1 hour each, with physical contact limited to a brief hug at the start and

conclusion of the visit. Virtual visits are generally available when facility-wide scheduling permits. On average, incarcerated individuals in PA are allowed *more* time out of their cells, with many incarcerated people locked in their cells only at night and for the twice-daily headcount, whereas individuals in the NL are routinely locked up from 5 p.m. to 7:30 a.m.¹⁵ Furthermore, the resocialization programming in both PA and Dutch detention institutions includes drug and alcohol rehabilitation programming, education, work, cognitive behavioral therapy, and vocational training, but the specifics of these programs differ for different groups of detainees and across both countries.

While prisons in the two countries share many of the fundamental features of prison life, the substantial differences in correctional contexts mean that it is far from given that a survey instrument designed for the NL would have good psychometric properties in PA.

Administrations of the PCQ in the NL and PA

Administration in the NL (2017)

The PCQ was first used in the LIC study in the NL in 2017, when the survey was administered to the full Dutch prison population, including both men and women, pretrial detainees and convicted individuals, and regimes at all security levels, housed in 28 prisons and remand centers.¹⁶ Incarcerated individuals were handed a paper and pencil version of the questionnaire, and research assistants were available to help fill out the questionnaire where requested.¹⁷ A sample of 4538 respondents were surveyed, amounting to 64% of the Dutch prison population.

Bosma et al. (2020) published results from a psychometric analysis of the prison climate scales based on the LIC study held in 2017. They concluded that the PCQ's factor structure, reliability, and validity were satisfactory and mostly exceeded minimum expectations. The prison climate scales correlated in a theoretically conceivable manner, which provided evidence of construct validity.¹⁸ The analyses also verified that the data were, for the most part, structured as expected. In three cases, the factor structure differed from the originally envisioned scale structure: the procedural justice and staff-prisoner scales loaded on the same factor, items from the reintegration and meaningful activities scales loaded onto the same factor, and two items ('I enjoy receiving visits' and 'after receiving a visit, I feel good') split off from other items on the visits scale. In an analysis of criterion validity, the authors show that the prison climate scales explain a substantial share of the variation in overall satisfaction with the institution. Notably, however, the prison climate scales explained remarkably little variation in experienced sentence severity. It is an open empirical question whether the psychometric patterns observed in the Dutch analyses are particular to the 2017 data collection in the NL or whether similar psychometric patterns would be observed in different contexts or at different times. To shed light on this question, we now turn our attention to a 2022 administration of the PCQ in PA.

Administration in PA, USA (2022)

In early 2022, a PA-based research team adapted the PCQ for use in a PA SCI: a medium-security institution for men. The facility was built in the 1990s, has the capacity to house

just over 1100 individuals, and is generally known for its relatively extensive program offering. At the time of survey administration, the institution held just under 1000 individuals, spread across 14 housing units which differed in terms of staffing, target population, and structural layout. ‘General population’ units had two men living in each cell, typically in units of 128 beds. ‘Therapeutic communities’ provided residents with specialized drug programming in smaller units of 64 beds (Welsh, 2007). The ‘recovery unit’ housed men who had chosen to continue with a voluntary self-help program after completing their mandatory drug programming. Residents on the ‘transitional housing unit’ were to receive reentry-focused support. Men housed on the ‘honor block’ had earned access to additional privileges through good behavior, whereas privileges were reigned in on the ‘restrictive housing unit’ that was used for disciplinary purposes and for security-focused segregation. This prison, therefore, provided a diverse set of environments in which to test the psychometric properties of the PCQ within one American correctional facility.

The PA research team aimed to keep changes to the prison climate scales in the PCQ to a minimum in the interest of comparability. This, overall, was relatively straightforward: because the PCQ had previously been translated to English for non-Dutch speakers incarcerated in the NL, no translation from Dutch was necessary; the PCQ’s items were all relevant to prison climate in the PA context; and the short and simple wording of the items meant that few changes were needed to fit local vernacular. Nevertheless, a few minor changes were made to ensure relevance to the PA context and to meet the wider research needs of the PA research team. First, in the original PCQ, items on the scales for staff–prisoner relationships and procedural justice focus on a resident’s unit, whereas other items focus on the institution more generally. In the PA version of the survey, three additional scales were adapted to focus on an individual’s living unit.¹⁹ Second, some items were repositioned to facilitate survey readability. Third, the ‘settlement of complaints’ scale was replaced with a new set of items reflecting the functioning of the PA grievance system.²⁰ Note that because the complaints scale was substantively changed, this scale is left out of the analysis in this paper. This article thus focuses on 13 out of the 14 original prison climate scales in the PCQ.²¹

A draft version of the adapted PCQ was tested with 12 incarcerated people residing on one unit in the prison. Based on feedback from these individuals, words were changed in a few places without altering the item’s substantive meaning, for example, by changing ‘pastoral care’ to ‘religious services’. Most notably, the term ‘prisoners’ was replaced with ‘incarcerated people’. A full list of items in the original PCQ, alongside the adapted items and item positions, is provided in Table A1.

In the spring of 2022, researchers went to each unit in the late afternoon, while the majority of incarcerated people were in their cells to be counted. Every individual on the unit at that time was given a copy of the pen-and-paper instrument to complete while in their cell. For the small number of individuals who were working or at a visit during count time, surveys were left with cellmates, and their surveys were collected the next day. Participation was voluntary, and nominal compensation was provided to all respondents for their time, irrespective of survey completeness. Individuals provided their ID number on the survey to enable the payment of compensation and to facilitate administrative data linking.²²

Data and methods

Data sample (PA, USA 2022)

A total of 641 respondents completed a survey in the PA SCI, amounting to a 66% response rate.²³ Three respondents left all prison climate items on the survey blank. Another seven individuals had more than 10 missing answers. We drop these 10 individuals from all analyses, leaving us with a sample of 631 respondents. On 23 out of the 64 items in the PCQ, incarcerated individuals were provided with the option to indicate that they had 'no opinion',²⁴ for example, if they had no experience (yet) with the services the items pertained to. Forty-six percent of respondents used a 'no opinion' answer at least once.²⁵ Less than 1% of the data is missing in the sense that respondents provided no answer. Table A2 provides an overview of both missing data and 'no opinion' answers for each item on the questionnaire.

Survey data were linked to administrative data for all but 12 individuals who chose to participate anonymously. Table 1 lists sample characteristics from the PA respondents for whom administrative data were obtained, alongside comparable characteristics for the Dutch sample, where available.²⁶ Respondents in the PA sample differed from those of the Dutch sample on almost all characteristics except age. Most notably, but unsurprisingly given the differences in the criminal justice systems described above, respondents in PA had, on average, served much more time in prison at the time of completing the survey (7.81 years) than individuals in the NL (0.99 years). Even after excluding life-sentenced individuals from the PA sample, the disparity in time served remains large, at 5 years. The types of offenses for which individuals were detained differ in several ways. PA respondents were, for example, less likely to be in prison for property offenses than Dutch respondents (11% vs. 30%), and more likely to be in prison for violent offenses (53% vs. 42%). These differences at least in part reflect the fact that the Dutch population includes individuals in pretrial detention. PA respondents are also much more likely to be native-born (94% vs. 65%).

Statistical analyses

We conduct a reliability analysis, construct validity analysis, criterion validity analysis, and exploratory factor analysis on the PCQ data from PA. Our psychometric analyses mirror the analyses conducted in Bosma et al. (2020) wherever possible to facilitate comparisons across the two contexts.²⁷

The *reliability* of an instrument refers to its measurement precision and is generally thought of as a ratio of true score variance to observed score variance (Furr, 2017). One way to examine test reliability is by examining whether the items within a scale that are expected to measure the same construct in fact produce similar scores. The reliability of a test is then measured as the correlation between two parallel tests. When dealing with data from a single survey wave, researchers tend to think of scores on half the items of the scale as 'parallel' to the items on the other half of the scale. The intuition is that if both halves of the scale measure the same construct, then scores on one-half of the items should correlate highly with scores on the other half (De Vet et al., 2017). We

Table 1. Survey respondent and population characteristics.

	Respondents		Population	
	PA (1)	NL (2)	PA (3)	NL (4)
Demographic characteristics				
Age	38.31	36.84	39.04	36.76
Male	1.00	0.95	1.00	0.95
Children	0.70	0.60	–	–
Has partner	0.32	0.59	–	–
Education: high school or higher	0.58	–	0.55	–
Country of birth (USA/NL)	0.94	0.65	–	0.61
White	0.61	–	0.61	–
Black	0.38	–	0.39	–
Offense type				
Violent	0.53	0.42	0.53	0.41
Property	0.11	0.30	0.10	0.32
Drugs	0.15	0.18	0.17	0.17
Sex	0.07	0.05	0.07	0.04
Public order	0.14	–	0.14	–
Other	–	0.06	–	0.06
Conditions of incarceration				
Time served (ex. lifers)	6.02	–	5.97	–
Time served (inc. lifers)	7.78	0.99	7.82	–
Double cell	0.87	0.21	–	–
Pretrial detention	0.00	0.38	0.00	0.36
Prison	1.00	0.35	1.00	0.38
Prison unit				
General population	0.54	–	0.55	–
Therapeutic community	0.21	–	0.19	–
Transitional housing unit	0.05	–	0.05	–
Honor block	0.10	–	0.11	–
Little Scandinavia	0.01	–	0.01	–
Restricted housing unit	0.03	–	0.03	–
Recovery unit	0.06	–	0.05	–
N	619	4538	973	6822

Note: Columns 1 and 2 list sample characteristics from the PA respondents for whom administrative data were obtained, alongside comparable characteristics for the NL sample, where available. Columns 3 and 4 list population characteristics for the PA and NL samples as available in administrative data. The last row lists the total sample size in each column. For some variables, data were not available for all individuals in the sample. In those cases, percentages are calculated over the individuals for whom data were available. The Dutch sample and population characteristics were calculated from Table 1 in Bosma et al. (2020) and from Table 1 in van Ginneken et al. (2018). PA: Pennsylvania; NL: the Netherlands.

report 95% confidence intervals for Spearman–Brown (SB) corrected split-half reliability estimates based on all possible such splits for all scales consisting of four or more items (Revelle and Condon, 2018). We also report Cronbach's alpha (CA), which estimates a lower-bound estimate of reliability based on characteristics of the covariances of the items in a scale.²⁸

Construct validity refers to the extent to which a measure adequately assesses the construct it purports to assess (Nunnally and Bernstein, 1994).²⁹ We assess the extent to which associations between scales in the PCQ match theoretical predictions about their expected associations. To do so, we estimate an interscale correlation matrix based on the correlations of scales formed from the covariance matrix of items, which we correct for attenuation. In addition, we analyze construct validity by examining differences in mean scale scores between various types of prison units.

Concurrent criterion validity refers to the extent to which a measure is empirically associated with relevant criterion variables that are assessed at the same time (Western and Rosenthal, 2003). We assess the *concurrent criterion validity* of the PCQ by conducting a series of linear regression analyses that examine to what extent prison climate scale scores vary with respondents' overall satisfaction with the institution as well as with how painful they experience their current sentence to be. The latter is measured on a scale composed of three items ('I experience my sentence as painful', 'My time here feels a lot like punishment', and 'This sentence is more painful than I anticipated'), with higher scores indicating sentences that are experienced as more painful. Overall satisfaction with the institution is measured with a single item, 'I am satisfied with this institution', where higher scores indicate higher satisfaction. The linear regression analyses include the prison climate scales and a range of control variables.³⁰

Finally, the *exploratory factor analysis* aims to uncover the underlying factors in the survey data. Exploratory factor analysis partitions item variance into common variance, which is accounted for by underlying latent factors, and unique variance, which is a combination of item-specific variance and random error. Mirroring Bosma et al. (2020), we use principal axis factoring and a direct OBLIMIN rotation method. Exploratory factor analyses require the researchers to specify the number of factors in the data. We obtain an estimate of the number of factors in the data through a parallel analysis procedure, which extracts factors until the eigenvalues of the real data are less than the corresponding eigenvalues of a random dataset of the same size (Hayton et al., 2004).

As discussed previously, 23 out of the 64 items in the PCQ provided respondents with a 'no opinion' answer option. Consequently, for the scales constructed on these items, this resulted in a relatively large proportion of 'missing' data. While 'no opinion' answer options accommodate individuals who recently arrived or have not used certain services, they can complicate psychometric analyses (Riegel et al., 2000). We use three strategies to deal with 'missing' data (including items where respondents chose 'no opinion' and items that respondents left blank).³¹ First, we impute missing data based on responses to the prison climate scales and 14 supplementary items on individual service use,³² using a random forest-based approach. Specifically, we use the MissForest algorithm, which outperforms most other imputation methods on datasets with mixed data types (Stekhoven and Buhlmann, 2012). This approach has the benefit of providing one dataset that can be used for all analyses. Second, because imputing missing data may

affect results, we also run all analyses on complete cases only ($N = 280$), for which no data are missing. Results from this smaller subsample can, however, not necessarily be generalized to the full population (Allison, 2010). Third, we report results from a pairwise deletion approach to missing data to ensure comparability with results reported in Bosma et al. (2020). Only in the exploratory factor analysis did results differ meaningfully across these three approaches, and we therefore display them in the main text. For all other analyses, we display results based on data in which missing data were imputed.³³

Results of psychometric analyses

We discuss the results from each of the psychometric analyses discussed above in turn. We pay particular attention to similarities and divergences with the findings of Bosma et al. (2020). Where appropriate, published results from Bosma et al. (2020) are reprinted here to facilitate comparisons between the two psychometric analyses.

Reliability analyses

The results of the reliability analyses are presented in Table 2, alongside the overall mean and standard deviation for each of the scales.³⁴ The last column in Table 2 shows 95% confidence intervals for the SB-corrected reliability estimates. A score above 0.70 is commonly seen as acceptable reliability, and scores generally easily exceeded this threshold, although the lower-bound estimate for the satisfaction with visits scale is substantially below this at 0.54. This result is in line with observations made by Bosma et al. (2020) and likely reflects the fact that two items on the visits scale ('I enjoy receiving visits' and 'after receiving a visit, I feel good') load onto a separate factor in both analyses (see Exploratory Factor Analysis, below). The values for the CA for almost all scales used in the PA survey are good, although the high values of α ($>.90$) on some scales suggest that some items may be redundant and thus that the number of items on these scales could be reduced (Tavakol and Dennick, 2011).

Construct validity analyses

Results of the construct validity analyses show that the observed associations between the prison climate scales are largely in line with theoretical expectations. The interscale correlation matrix, as presented in Table 3,³⁵ shows that all scales correlate positively. The scales on staff–prisoner relationships, procedural justice, the availability of meaningful activities, reintegration, and autonomy all correlate highly with each other. We note that the staff–prisoner relationships and procedural justice scales are nearly perfectly correlated, a finding that is in line with the fact that items on this scale map onto the same factor (see *Explanatory factor analysis*, below). Notably, the safety scale is not strongly correlated with any other scale, including scales with which one would theoretically expect a correlation, such as prisoner relationships and staff–prisoner relationships. Overall, these results closely mirror the relative magnitude of the associations found by Bosma et al. (2020).

Table 2. Descriptive statistics for prison climate scales and reliability analysis.

Scale	PA					NL			
	Items	Mean	SD	CA	SB CI	Mean	SD	CA	SB
Relationships									
01. Prisoner relationships	5	3.09	0.80	0.90	(0.84, 0.88)	3.44	0.71	0.86	0.82
02. Staff–prisoner relationships	4	2.59	1.02	0.91	(0.87, 0.93)	3.32	0.94	0.89	0.87
03. Procedural justice	4	2.58	1.07	0.92	(0.88, 0.95)	3.30	0.94	0.91	0.92
Safety									
04. Safety	5	3.73	0.83	0.80	(0.68, 0.86)	4.00	0.83	0.89	0.87
Contact with the outside world									
05. Satisfaction with visits	8	2.99	0.61	0.76	(0.54, 0.91)	2.94	0.72	0.79	0.65
06. Satisfaction with frequency of contact	3	2.55	1.05	0.83	–	2.84	1.06	0.82	0.70
Facilities									
07. Sleep quality	3	2.80	0.97	0.80	–	2.77	1.06	0.78	0.77
08. Quality of care	6	2.95	0.84	0.83	(0.7, 0.88)	3.30	0.91	0.89	0.87
09. Shop quality	3	2.58	0.95	0.87	–	2.39	0.97	0.90	0.86
Meaningful activities									
10. Satisfaction with activities	7	3.20	0.90	0.89	(0.81, 0.91)	3.12	0.87	0.86	0.85
11. Availability of meaningful activities	4	2.84	1.01	0.88	(0.87, 0.91)	2.27	0.96	0.91	0.91
12. Reintegration	4	3.10	1.17	0.91	(0.9, 0.92)	2.49	1.07	0.92	0.92
Autonomy									
13. Autonomy	4	3.19	0.92	0.84	(0.81, 0.87)	2.71	0.96	0.86	0.84

Note: Table lists the number of items on each scale, as well as three statistics for each scale for PA and NL: baseline means, standard deviations, and CA. For the PA results, we provide 95% confidence intervals for SB-corrected split-half reliability estimates based on all possible such splits for all scales consisting of four or more items (Revelle and Condon, 2018). Split-half estimates are only calculated for scales consisting of four or more items. Bosma et al. (2020) calculated split-half reliability using just a single split, so a single SB coefficient is displayed. PA: Pennsylvania; CA: Cronbach's alpha; SB: Spearman–Brown.

Differences in mean PCQ scale scores between various types of prison units, presented in Table 4, provide further evidence of construct validity. As expected, mean scale scores tend to vary positively with the level of privileges afforded on a unit.³⁶ Overall, the lowest scores are found for the restrictive housing unit, where individuals are housed in relative isolation as a sanction or during an active investigation for serious rule violations. Scores are highest for the honor block, which houses residents who have earned additional privileges. The greatest differences between units are found in the autonomy, relationships,

Table 3. Interscale correlation matrix.

	1	2	3	4	5	6	7	8	9	10	11	12
Relationships												
01. Prisoner relationships												
02. Staff–prisoner relationships	.470											
03. Procedural justice	.413	.978										
Safety												
04. Safety	.189	.321	.306									
Contact with the outside world												
05. Satisfaction with visits	.280	.307	.286	.237								
06. Satisfaction with frequency of contact	.193	.246	.231	.175	.487							
Facilities												
07. Sleep quality	.291	.338	.289	.312	.361	.285						
08. Quality of care	.407	.433	.389	.270	.560	.275	.361					
09. Shop quality	.312	.215	.188	.099	.473	.189	.269	.528				
Meaningful activities												
10. Satisfaction with activities	.325	.484	.443	.246	.452	.402	.340	.502	.391			
11. Availability of meaningful activities	.384	.606	.571	.272	.337	.310	.324	.461	.238	.703		
12. Reintegration	.409	.650	.615	.248	.294	.259	.321	.433	.280	.552	.726	
Autonomy												
13. Autonomy	.347	.475	.478	.313	.295	.249	.233	.387	.248	.495	.577	.558

Note: The interscale correlations reported here were based on the correlations of scales formed from the covariance matrix of items and corrected for attenuation. Correlations greater than .5 are highlighted in bold.

Table 4. Differences in mean scores across units.

Scale	General	Therapeutic	Honor	Recovery	Restrictive	Transitional
Relationships						
01. Prisoner relationships	2.95 ^{hr}	3.16	3.38 ^{gi}	3.56 ^{gi}	2.57 ^{hrs}	3.41 ⁱ
02. Staff-prisoner relationships	2.39 ^{hs}	2.55 ^h	3.42 ^{gt}	2.93 ⁱ	2.04 ^{rs}	3.04 ^{gi}
03. Procedural justice	2.36 ^{hr}	2.53 ^h	3.50 ^{gttri}	2.87 ^{gh}	2.10 ^{hs}	3.05 ⁱ
Safety						
04. Safety	3.72	3.74	3.82	3.62	3.35	3.88
Contact with the outside world						
05. Satisfaction with visits	3.01	2.91	2.99	3.12	3.07	2.98
06. Satisfaction with frequency of contact	2.50	2.46	2.84	2.61	2.43	2.70
Facilities						
07. Sleep quality	2.76	2.91	2.87	2.89	2.38	2.82
08. Quality of care	2.87	3.08	2.87	3.24	2.77	3.10
09. Shop quality	2.49 ^t	2.87 ^{gh}	2.25 ^{tr}	2.82 ^h	2.72	2.69
Meaningful activities						
10. Satisfaction with activities	3.09 ^h	3.27	3.59 ^{gi}	3.30	2.76 ^{hs}	3.52 ⁱ
11. Availability of meaningful activities	2.62 ^{hr}	2.93 ^h	3.46 ^{gti}	3.36 ^{gi}	2.31 ^{hrs}	3.15 ⁱ
12. Reintegration	2.75 ^{thrs}	3.44 ^{gi}	3.74 ^{gi}	3.60 ^{gi}	2.40 ^{thrs}	3.80 ^{gi}
Autonomy						
13. Autonomy	3.13 ^{hi}	3.12 ^h	3.65 ^{gti}	3.34 ⁱ	2.41 ^{ghrs}	3.41 ⁱ

Note: Table lists mean scores on scales by type of housing unit, from left to right: the general population units, therapeutic communities, the honor block, the recovery unit, the restrictive housing unit, and the transitional housing unit. For details, see section "Administration in PA, USA (2022)" in the main text. We conduct Tukey's HSD test to identify significant differences between all pairwise comparisons of means. The superscripts indicate whether the group mean is different from the general population (g), therapeutic community (t), honor block (h), recovery unit (r), restricted housing unit (i), or transitional housing unit (s) means at the .05 significance level.

and activities domains. We observe less variation across units in the facilities domain and the domain that concerns contact with the outside world. This too is in line with expectations because scales in these domains concern centralized services, such as visits and medical care, which are the same for residents across housing units. It is notable that, like in the Dutch results, there is little variation in feelings of safety across units.³⁷

Criterion validity analyses

Results of tests of the PCQ's *criterion validity* are displayed in Table 5.³⁸ The dependent variables in columns 1 and 2 are an individual's overall satisfaction rating with the institution and their experienced sentence severity, respectively. Overall satisfaction with the

Table 5. Criterion validity.

	General opinion (1)	Subjective severity (2)
Prisoner relationships	0.092 (0.053)	0.065 (0.055)
Staff–prisoner relationships	0.294*** (0.085)	0.027 (0.088)
Procedural justice	0.031 (0.079)	−0.062 (0.082)
Safety	0.019 (0.048)	−0.188*** (0.050)
Satisfaction with visits	−0.083 (0.075)	0.077 (0.078)
Frequency of contact	0.044 (0.040)	−0.101* (0.041)
Sleep quality	−0.055 (0.042)	−0.130*** (0.044)
Quality of care	−0.023 (0.056)	−0.004 (0.057)
Shop quality	0.090 (0.048)	−0.071 (0.049)
Satisfaction with activities	0.085 (0.057)	−0.025 (0.059)
Availability of activities	0.271*** (0.057)	−0.078 (0.059)
Reintegration	0.087 (0.046)	0.006 (0.048)
Autonomy	−0.030 (0.050)	−0.069 (0.051)
Controls		
Age	0.004 (0.004)	−0.003 (0.004)
Foreign born	−0.023 (0.152)	0.230 (0.157)
Finished high school	0.073 (0.074)	0.107 (0.077)
Has partner	−0.030 (0.080)	−0.035 (0.083)
Has children	−0.018 (0.084)	0.080 (0.087)
Time served (months)	0.001 (0.000)	0.002*** (0.000)
Black	0.020 (0.086)	0.063 (0.089)

(Continued)

Table 5. (Continued)

	General opinion (1)	Subjective severity (2)
Shares a cell	-0.037 (0.121)	-0.016 (0.125)
Therapeutic community (ref = general population)	-0.310** (0.102)	0.039 (0.105)
Recovery (ref = general population)	-0.207 (0.160)	-0.184 (0.165)
Honor (ref = general population)	-0.141 (0.143)	0.164 (0.147)
Restrictive housing (ref = general population)	0.152 (0.215)	0.247 (0.223)
Transitional housing (ref = general population)	0.067 (0.172)	0.291 (0.178)
R^2	0.410	0.205
Adjusted R^2	0.383	0.168
Number of observations	589	590

Note: Results from linear regression analyses. The dependent variable in column 1 is respondent's overall satisfaction with the institution, measured with a single item, 'I am satisfied with this institution', where higher scores indicate higher satisfaction. The dependent variable in column 2 is a respondent's score on an 'experienced severity' scale of three items ('I experience my sentence as painful', 'My time here feels a lot like punishment', and 'This sentence is more painful than I anticipated'), with higher scores indicating sentences that are experienced as more painful. Scale scores are included in the regression as simple averages of the items in each scale. Standard errors in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.

institution is predicted by staff-prisoner relationships and the availability of activities. These two factors also predict overall satisfaction with the institution in the NL, although more factors are statistically significant in the Dutch context. Note that, in both contexts, staff-prisoner relationships emerge as a key predictor of satisfaction with the institution in the regression analyses, reinforcing the established notion that staff-prisoner relationships are particularly salient for the quality of prison life (Liebling and Arnold, 2004). Like in the Dutch results, demographics and sentence characteristics of respondents incarcerated in the PA facility are largely unrelated to overall satisfaction with the institution.³⁹

Notably, the scales that predict experienced sentence severity are different from the scales predicting an individual's overall opinion of the institution, with people who feel less safe and who report poorer quality of sleep experiencing their sentence as more severe in both contexts, while staff-prisoner relationships and the availability of activities, which most strongly predict individuals' overall opinion of the institution, are not significant predictors of experienced sentence severity. Time served significantly predicts experienced sentence severity in PA but not in the NL, which likely reflects the fact that Dutch sentences are much shorter. It is notable that the prison climate scales explain only a relatively small share of the variance in experienced sentence severity, with the prison climate scales increasing the adjusted R^2 from .07 to .17 (PA) and .05 to .12 (NL) compared with models that include control variables only. In contrast,

adding the prison climate scales to a model that included just control variables increased the adjusted R^2 of the model on overall satisfaction with the institution from .07 to .38 (PA) and from .06 to .54 (NL). These results confirm that the prison climate scales capture an ‘overall institution rating’ but cast doubt on the idea that the pains of imprisonment vary directly with conditions of confinement, at least as measured in this survey.

Exploratory factor analysis

Finally, the results from the exploratory factor analysis are shown in Table 6. The left-most column of this table lists the preassumed structure of the survey, specifying which scale the authors of the original PCQ hypothesized an item belonged to. Columns 1–6 present the factor that the item loaded onto alongside its associated factor loading, for three datasets—a dataset in which missing data has been imputed (columns 1 and 2), a dataset with complete cases only (columns 3 and 4), and a dataset in which we delete missing data pairwise (columns 5 and 6), which are most comparable with the results previously published in Bosma et al. (2020) (columns 7 and 8). To ensure that the reader can compare results with those previously published in Bosma et al. (2020), we display factor loadings for all items. Items with factor loadings lower than .40 are grayed out. Where items double-load on multiple scales, we retain the highest factor loading, listing the factor on which an item double-loaded in the superscript.

Table 6 shows that the factor mapping aligns with the originally envisioned scale structure for 9 out of the 13 scales in the PA data. Notably, it shows that all three scales that showed unexpected factor mappings in the Dutch survey (discussed in section “Administration in the NL (2017)” above), ‘rebel’ in almost identical ways in the PA results. First, like in the Dutch results, the satisfaction with procedural justice and staff–prisoner relationships items load on the same factor in all three datasets. Second, the same two items (‘I enjoy receiving visits’ and ‘after receiving a visit, I feel good’) that split off from the satisfaction with visits scale in the NL also consistently split off from this scale in the PA results. The PA results also point toward some overlap between the satisfaction with visits scale and the frequency of contact scale in the dataset that includes only complete cases. Third, the ‘availability of meaningful activities’ items and the reintegration scale overlap largely in the factor analysis based on complete cases (see columns 3 and 4).⁴⁰ This mirrors the Dutch results, in which these two scales also mapped onto the same factor. In the PA data, we further observe some overlap between the availability of meaningful activities scale and the satisfaction with activities scale, with the items about work and education double-loading onto both of these scales in two out of the three datasets. The fact that unexpected factor mappings in the Dutch survey appeared in almost identical ways in the PA results clearly points to areas where the factor structure of the PCQ could be improved.

Results differed depending on whether missing data was imputed (columns 1 and 2), deleted pairwise (columns 5 and 6), or restricted to complete cases (columns 3 and 4). This suggests that there are systematic differences between individuals who do and do

Table 6. Exploratory factor analysis.

Item	PA						NL	
	Imputed		Complete cases		Pairwise		Pairwise	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Domain: Relationships								
01. Prisoner relationships								
Incarcerated people on this unit treat each other respectfully	1	0.83	1	0.84	1	0.84	1	0.81
Incarcerated people on this unit are quickly accepted into the group	1	0.71	1	0.74	1	0.71	1	0.79
Incarcerated people on this unit are considerate of each other	1	0.83	1	0.84	1	0.84	1	0.83
Incarcerated people on this unit get along with each other	1	0.78	1	0.81	1	0.78	1	0.65
Incarcerated people on this unit help and support each other	1	0.78	1	0.83	1	0.78	1	0.80
02. Staff–prisoner relationships								
Staff on this unit help me if I have problems	2	0.73	2	0.74	2	0.74	2	−0.73
Staff on this unit are kind to me	2	0.83	2	0.83	2	0.82	2	−0.79
Staff on this unit are there to talk to if I feel worried or sad	2	0.79	2	0.84	2	0.77	2	−0.72
Staff on this unit motivate and encourage me to participate in activities	2	0.66	2	0.74	2	0.66	2	−0.63
03. Procedural justice								
Staff on this unit treat me fairly	2	0.87	2	0.89	2	0.87	2	−0.78
Staff on this unit explain their decisions to me	2	0.76	2	0.86	2	0.75	2	−0.69
Staff on this unit treat me with respect	2	0.87	2	0.82	2	0.87	2	−0.78
Staff on this unit give me a chance to express my views before they make decisions								
Domain: Safety								
04. Safety	2	0.75	2	0.78	2	0.75	2	−0.67
I feel safe in this institution	3	0.31	3	0.28	3	0.30	3	0.71
I sometimes feel threatened by incarcerated people	3	0.78	3	0.80	3	0.78	3	0.89
There are places in this building where I feel unsafe	3	0.73	3	0.72	3	0.72	3	0.89
I am afraid of some incarcerated people	3	0.86	3	0.89	3	0.86	3	0.89
I am afraid of some staff	3	0.54	3	0.63	3	0.55	3	0.69

(Continued)

Table 6. (Continued)

Item	PA						NL	
	Imputed		Complete cases		Pairwise		Pairwise	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Domain: Contact with the outside world								
05. Satisfaction with visits								
The visiting room is pleasant	4	0.54	4	0.62	4	0.51	4	0.79
My visitor and I can have enough physical contact during visits	4	0.82	4	0.77	4	0.78	4	0.82
The visiting hours are long enough	4	0.72	4	0.66	4	0.68	4	0.67
I have sufficient privacy during visiting hours	4	0.71	4	0.77	4	0.66	4	0.52
The staff treat my visitors nicely	4	0.62	4	0.46	4	0.55	4	0.51
The visiting hours are frequent enough	4	0.67	4	0.55	4	0.59	4	0.52
I enjoy receiving visits	5	0.89	5	0.89	5	0.89	5	-0.89
After receiving a visitor, I feel good	5	0.81	5	0.78	5	0.78	5	-0.89
06. Satisfaction with frequency of contact								
I am satisfied with how often... I can see my family, friends or partner here	6	0.81	4	0.44	6	0.80	6	0.89
I am satisfied with how often... I can see my child(ren) here	6	0.95	410	0.49	6	0.96	6	0.90
I am satisfied with how often... I can see my lawyer here	6	0.55	410	0.41	6	0.50	6	0.68
Domain: Facilities								
07. Sleep quality								
My sleep is often restless	7	0.77	7	0.74	7	0.77	7	0.81
My sleep is often disturbed	7	0.84	7	0.85	7	0.84	7	0.82
Due to poor conditions in this institution and/or my cell, I can't sleep well	7	0.59	7	0.60	7	0.60	7	0.80
08. Quality of care								
I can get medical care here if I want to	8	0.64	8	0.65	8	0.66	8	-0.77
Health problems are being taken care of adequately here	8	0.71	8	0.68	8	0.71	8	-0.88
I am satisfied with the work of the nurse	8	0.60	8	0.47	8	0.55	8	-0.88
I am satisfied with the work of the doctor	8	0.78	8	0.61	8	0.78	8	-0.88
I am satisfied with the work of the dentist	8	0.44	8	0.42	8	0.42	8	-0.65
I am satisfied with the work of the psychologist	8	0.46	8	0.40	8	0.44	8	-0.55

(Continued)

Table 6. (Continued)

Item	PA						NL	
	Imputed		Complete cases		Pairwise		Pairwise	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
09. Shop quality								
I am satisfied with the range of products in the commissary	9	0.72	9	0.72	9	0.72	9	0.84
I am satisfied with the prices in the commissary	9	0.85	9	0.86	9	0.83	9	0.78
I am satisfied with the quality of the products in the commissary	9	0.81	9	0.82	9	0.81	9	0.85
Domain: Meaningful activities								
10. Satisfaction with activities								
I am satisfied with the recreation activities	10	0.84	10	0.86	10	0.81	10	-0.43
I am satisfied with the sports	10	0.86	10	0.86	10	0.85	10	-0.59
I am satisfied with the library	10	0.58	10	0.57	10	0.58	10	-0.66
I am satisfied with my work in this institution	10	0.42	10	0.54	10	0.39	10	-0.51
I am satisfied with the education/ courses	10	0.42	10	0.53	10	0.40	10	-0.50
I am satisfied with the outdoor activity	10	0.69	10	0.71	10	0.68	10	-0.59
I am satisfied with the religious services	10	0.46	10	0.56	10	0.44	10	-0.60
11. Availability of meaningful activities								
The daily program is interesting enough	12	0.52	10	0.39	12	0.52	12	0.44
I learn useful skills here	12	0.75	11	0.53	12	0.75	12	0.59
I have enough to do here	12	0.57	11	0.47	12	0.59	12	0.43
The activities here help me to develop myself	12	0.71	11	0.51	12	0.72	12	0.62
12. Reintegration								
On this unit, I can prepare well for my return into society	11	0.67	11	0.74	11	0.65	12	0.63
On this unit staff encourage me to make plans for after release	11	0.71	11	0.64	11	0.70	12	0.56
On this unit I can get extra support to prepare for my return to society	11	0.81	11	0.77	11	0.80	12	0.60
On this unit I can learn things that help me to stay away from crime after release	11	0.81	11	0.81	11	0.81	12	0.65
Domain: Autonomy								
13. Autonomy								
On this unit there is much I can decide for myself	13	0.85	12	0.85	13	0.85	13	-0.80

(Continued)

Table 6. (Continued)

Item	PA						NL	
	Imputed		Complete cases		Pairwise		Pairwise	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
On this unit I can decide for myself on matters that are important to me	13	0.88	12	0.90	13	0.89	13	−0.75
On this unit I am encouraged to arrange matters myself	13	0.73	12	0.76	13	0.72	13	−0.51
On this unit I have enough freedom of movement	13	0.35	12	0.35	13	0.35	13	−0.55

Note: Table lists results from an exploratory factor analysis for three different datasets. Columns 1 and 2 list results for our core dataset, in which missing data are imputed using a random forest-based approach. Columns 3 and 4 list results based on complete cases only. Columns 5 and 6 list results based on a dataset in which missing data are deleted pairwise, which mirrors the approach used in Bosma et al. Factors are extracted using principal axis factoring and a direct OBLIMIN rotation method. We used a parallel analysis procedure to identify the number of factors in the data. The parallel analyses identify 13 factors in our core dataset and when we use pairwise deletions of missing data and 12 factors when we use complete cases only. Columns 7 and 8 list results as published in Table 2 of Bosma et al. (2020). Note that factor numbers differ from those published in Bosma et al. (2020) to facilitate comparisons with the PA results. Some items load only weakly onto any scale, so we display all items with values greater than .25 to ensure that we see what scales items load onto. Where items double-load at this value, we retain the highest factor loading, listing the factor on which an item double-loaded in the superscript.

not answer all questions.⁴¹ We will discuss this issue further in the section ‘Lessons for Comparative Prison Research: Towards Standardized Prison Climate Scales?’

Taken together, the results of these psychometric analyses show that the PCQ appears no less suitable for use in PA than it is in the NL. This is remarkable given the substantial differences in correctional contexts across the two countries and suggests that the survey captures an underlying essence of prison climate that many contexts share. Perhaps more strikingly, notable patterns in the Dutch psychometric analyses were consistently replicated in the analyses reported here. These shared patterns clearly point to areas where the structure of the survey can be improved. For example, items that did not load onto the correct scale in both surveys should be substantially revised or dropped altogether from future iterations of the survey.

Lessons for comparative prison research: Toward standardized prison climate scales?

This study adds to a small but growing literature that tests the psychometric properties of prison climate surveys outside of contexts for which they were originally designed.⁴² That the PCQ, administered with minimal adaptations, demonstrates good psychometric properties in the USA, where correctional environments are notably different from those

in the NL, is perhaps testament to the essential similarities of custodial environments across the Western world. Our findings are meaningful because they suggest that the PCQ is likely to be suitable for use in much of the USA as well as in Europe, where correctional environments generally differ less from those in the NL. These observations almost inevitably lead to questions about the potential for comparative research on prison climate that uses these surveys. To conduct meaningful comparisons, however, we need evidence of measurement invariance across contexts in addition to evidence of a survey's psychometric properties in those contexts. While researchers have started to compare prison climates across institutions, places, and times, they have generally done so without demonstrating formal evidence of measurement invariance. This is problematic because it means that we do not know whether a comparison of scores across groups of respondents reflects true differences or measurement differences (Leitgöb et al., 2023). The literature's foundation for comparative prison climate research, then, is much weaker.

Measurement invariance across contexts is much easier to achieve with surveys that are explicitly designed for comparative use, using items that maximize comparability (Fitzgerald and Jowell, 2010; Harkness, 2011; Leitgöb et al., 2023). Existing prison climate surveys—including the PCQ—were originally designed for use in a specific national context. While they have later been used in or adapted to other contexts, none have been explicitly designed and standardized for comparative use. Given that the three most prominent prison climate surveys have all been developed in Europe—the EssenCES was originally developed in Germany, the MQPL in the UK, and the PCQ in the NL—there seems to be much potential to learn from the existing literature to develop a pan-European prison climate survey with standardized scales intended for comparative use.

With this ambitious long-term goal in mind, we limit ourselves here to a modest contribution. In this section, we briefly discuss three lessons from the Dutch/PA collaboration that could inform efforts to design new prison climate instruments or adapt existing ones for use in comparative prison climate research. In doing so, we link specific observations about the PCQ to general guidance from the comparative survey literature as well as to broader literatures on survey measures of organizational climate.

First, items of differential relevance to different subgroups within and across contexts can reduce the comparability of items and scales. The PCQ includes items that are relevant only to subgroups who have used specific services or who have participated in certain activities, alongside more generic items that can be answered independently of prior service use. While everyone can answer the question 'I can get medical care here if I want to', only some people can answer the question 'I am satisfied with the work of the psychologist'. The PCQ accommodates respondents who had no experience with specific services or activities (yet) by providing a 'not applicable' answer option. We have already discussed how the choice of method for dealing with such and other missing data can meaningfully affect study conclusions. In comparative contexts, these issues deepen. Different prisons offer different services, and the institutional and cultural barriers to accessing commonly offered services also differ across settings. By implication, a much larger share of people will have something to say about the work of the psychologist where psychology services are routinely provided than in settings where

such services are much harder to access. What appears to be much stronger or more diverse opinions about psychology services can, in such instances, reflect differences in access rather than opinions. In other cases, issues of differential relevance arise in more subtle ways. An item like 'staff encourage me to make plans for after release' is much more likely to invoke a 'no opinion' option in PA than in NL because there are many more people who have no prospect of being released (ever or in the near future). Differential relevance can thus lead to differential non-response levels, which in turn can introduce bias (Couper and de Leeuw, 2002). Importantly, these issues are particularly difficult to deal with when items with differential relevance to different subgroups are combined within the same scale. To resolve this problem, comparative survey researchers sometimes combine a common 'core' of items that are assumed to be universally relevant (regardless of individual characteristics, prior service use, or national context), with items in optional modules that are relevant to specific subgroups only or are only used for a country-specific population (Harkness, 2011).

Second, what parts of prison life happen at the level of the housing unit varies across contexts, creating difficulties with items that explicitly refer to a respondent's housing unit. Some items/scales in the original PCQ specifically ask about elements of climate at the institutional level, whereas others refer explicitly to the unit on which a respondent resides. Researchers have long recognized the importance of housing units (or treatment groups) within prisons (Saylor, 1984; van der Helm et al., 2011), and differences between institution- and unit-level provisions are what prompted item-level references to 'the prison' or 'your unit' in the PCQ. While such distinctions are also salient in PA, what is primarily provided at the level of a housing unit, and what is organized centrally for all individuals housed in the facility, varies between the two countries. As discussed in section "Administration in PA, USA (2022)" above, these differences prompted the PA research team to adapt three further scales to focus on the unit, thus introducing differences in the two surveys. We note that the appropriateness of asking questions at the unit versus the prison level may also vary both across and within prisons *within the same country*. Items included in standardized prison climate scales intended for comparative use should thus avoid such context-specific leveling distinctions. As above, where unit-level provisions are of particular interest, they could be included in optional modules.

Third, our exercise highlighted a lack of clarity about the level of analysis in the PCQ. Researchers of organizational climate in other institutions like schools and hospitals distinguish between psychological climate and organizational climate as two conceptually distinct areas of study (Schneider et al., 2013). Researchers interested in the former typically study individual experiences of climate⁴³; researchers studying climate as an attribute of an organization, however, are typically interested in aggregating climate features to the organizational level and therefore use items that refer to the level of aggregation (Schneider et al., 2013). The PCQ mixes items that refer to attributes of the prison with items that tap into individual-level perceptual and psychological variation. For example, the items 'I enjoy receiving visits' and 'After receiving a visitor, I feel good' concern how individuals *feel* about visits, whereas the remaining six items focus on the quality of the visiting facilities and the nature of visit policies. This may explain why these items split off from the other items on visitation in both the NL and PA.

While correctional managers can ensure that visiting environments are pleasant and that visiting policies are accommodating, visits are highly emotionally charged events in prison (Cochran and Mears, 2013; de Jong et al., 2022; Siennick et al., 2013; Turanovic and Tasca, 2019), and how one feels after a visit is likely to depend on many personal factors—not least because the nature of an incarcerated person's relationship to their visitors will vary. The use of items that tap into individual-level variation may also explain why the safety scale correlated so weakly with other scales and why safety scores hardly varied across units. By asking respondents whether they are 'afraid' or 'feel threatened', the items on the safety scale tap into feelings of vulnerability and fear, which research has documented may vary as much with individuals' past experiences as with facility-level measures of safety (Edgar et al., 2003; Mulvey et al., 2010). Thus, when scales mix items that tap into individual-level variation with items that tap into facility-level variation, observed differences may reflect both differences in organizational functioning and population differences that are outside of the control of prison management.

Discussion

In this article, we have discussed the lessons learned from using the PCQ in one PA SCI. We showed that the PCQ's factor structure, reliability, and validity were good in this context and that the survey appears no less suitable for the PA context than for the Dutch one. We also showed that the psychometric properties of the survey were remarkably similar in the NL and PA. Specifically, the fact that all three scales that showed unexpected factor mappings in the Dutch survey 'rebelled' in almost identical ways in the PA results, clearly points to areas where the factor structure of the PCQ could be improved. The close replication of psychometric patterns in the two contexts adds weight to suggestions in Bosma et al. that select survey items should be revised or deleted. Future iterations of the PCQ should be revised in accordance with the findings in this article, and future research should continue to test the psychometric properties of the PCQ in other settings.

A key limitation of this study is that, because of formal restrictions on data sharing with members outside of the core research teams, we have been unable to pool the datasets from NL and PA to conduct a direct empirical comparison of the two sets of psychometric results. Instead, we have compared the psychometric analysis conducted on the PA data to results from a previously published study (Bosma et al., 2020). This has prevented us from conducting a formal test for measurement invariance across the two contexts. We note, however, that given our discussion in the section 'Lessons for Comparative Prison Research: Towards Standardized Prison Climate Scales?', such measurement invariance is more likely when surveys are explicitly designed for comparative use. Future research should intend to pool datasets, which would enable the calculation of statistical point estimates on differences between two sets of results.

This research has contributed to a small but growing literature that suggests that prison climate surveys are well positioned to measure prison climates in a range of contexts. Given the sizeable differences between the Dutch and PA correctional contexts, the findings presented here suggest that the PCQ is likely suitable for stand-alone use in most

prison environments in Europe, which tend to more closely resemble the Dutch context. These observations also buttress a more general point: that two prisons that at first instance look very different still share a distinct identity as *prisons*, ‘a prison is a prison and feels like a prison according to most prisoners’ (Neubacher et al., 2021: 5). Life within many prison facilities in Europe, the USA, and other developed countries follows the cadence of a daily regime that offers some mix of work, education, treatment, care, and recreation within a framework in which both the control of movement and maintenance of safety takes center stage. Nevertheless, prisons that are very similar in terms of population, architecture, resourcing, and functions can differ meaningfully in their quality of life (Liebling, 2011). Prison climate surveys, then, have the potential to shed light on both what prisons share and what differences between them meaningfully affect life inside them.

To realize the *comparative* value of prison climate instruments, however, we need more than psychometric tests alone; we would ideally use a survey that is explicitly designed and standardized for comparative use. Given that the three prison climate surveys that are most actively used—the MQPL, the EssenCES, and the PCQ—all originated in Europe, the prospect of developing a pan-European prison climate survey that draws on lessons from the existing literature seems particularly promising.⁴⁴ Drawing on our experience of administering this survey in two contexts, we have discussed some general lessons that we hope will be useful for researchers intending to design such comparative survey instruments of prison climate. Specifically, we highlighted examples of differential item relevance across contexts and technical issues in scale construction arising from the level-of-analysis issues. A thoroughly tested European prison climate survey suitable for comparative analyses would create myriad opportunities for future research. Much can be learned, for example, from simply comparing the relative positioning of scale scores across contexts. Comparing the effects of similar interventions and policy changes across contexts with different ‘baseline’ climates could also help answer long-standing questions about the environmental requirements for such interventions to thrive.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Time spent by Britte van Tiem, Jordan M. Hyatt, and Synøve N. Andersen on this research was supported by a grant from Arnold Ventures.

ORCID iD

Britte van Tiem  <https://orcid.org/0000-0003-2723-1469>

Supplemental material

Supplemental material for this article is available online.

Notes

1. Note that this is conceptually distinct from the literature on ‘prison culture’, which has focused on the nature of behavioral rules and expectations in prison (Young et al., 2023).
2. For an exception, see Tonkin et al. (2012) who conducted a multiple group analysis to examine the fit of their model to data from prison versus hospital settings and for residents versus staff.
3. The PCQ was used as part of the Life in Custody Study, which was fielded in 2017, 2019, 2022, and 2024. The next administration is scheduled for 2026.
4. Several existing reviews provide helpful overviews of the nature of these instruments (Tonkin, 2016; van der Helm et al., 2011) and their underlying theoretical models (van Ginneken and Nieuwbeerta, 2020).
5. Two other questionnaires that had some initial promising psychometric results at the time of Tonkins’ (2016) review were the Forensic Satisfaction Scale and the Prison Group Climate Instrument, but these two tools are aimed at forensic health-care settings and therapeutic group settings, respectively. To the best of our knowledge, these tools have not been adapted for wider prison settings.
6. The survey was explicitly designed using generic language so as to be suitable for use in the full variety of Dutch custodial environments, ranging from units serving pretrial populations to those serving primarily repeat offenders or populations in need of extensive mental health support.
7. More information about the PCQ and requests for using the questionnaire can be addressed to Dr. Hanneke Palmen at j.m.h.palmen@law.leidenuniv.nl.
8. In addition, the PCQ includes a one-item overall institution rating (‘Generally speaking, I am satisfied with this institution’) and a three-item scale that captures how individuals experience the severity of their sentence. The survey also asks about individual characteristics not readily available through administrative data. Finally, the prison climate scales in the PCQ are supplemented by rotating supplementary modules measuring other concepts such as visit experiences, cognitive ability and autonomy, and reintegration. Unless otherwise indicated, references to the PCQ in this article solely refer to the items included in the prison climate scales.
9. Population figures are based on 2022 estimates from the United States Census Bureau.
10. As of the end of 2021, the Pennsylvania Department of Corrections alone held 37,303 individuals in its 23 SCIs, making for an incarceration rate of 286.9 individuals per 100,000 state residents in 2021 (Kuba, 2021). Aggregate jail population estimates are not published every year, but in 2019, PA held 31,790 individuals across 73 jails in 2019 (National Institute of Corrections, 2019). Federal facilities in PA held another 9,343 inmates (BOP statistics from February 16, 2023).
11. Detailed comparisons of prison population composition are complicated due to differences in the way officially published data are constructed. Figures cited here are based on a comparison of release data from two reports from 2017 for the NL (Table 2.13, de Looft et al., 2018) and PA (Table 29, Krausse et al., 2017).
12. At the end of 2021, prisons operated at 83% of operational capacity (Kuba, 2021).
13. Treatment and other specialized units, however, can house as few as 15 people.
14. De Looft et al. (2018) note that 3073 cells were fit for double occupancy in 2017, out of a total of 10,223 cells, although only 1460 individuals actually shared a cell in September 2017. Twenty percent of respondents in the administration of the PCQ in the NL in 2017 reported to live in a double cell.
15. Regimes and privileges often vary substantially. For example, incarcerated persons on good behavior in the NL get 11 additional hours out of cell, including to participate in an

- evening program two nights a week and to receive additional visits (Elbers et al., 2021). In PA, prisons often have ‘honor blocks’ where incarcerated persons on good behavior are allowed more time out of their cell. Both correctional systems have segregation conditions in which time out of cell is more limited.
16. Prior to sentencing, the Dutch Prison Service holds individuals in remand sentences, whereas in PA they are held alongside regular populations in local jails. In the NL, suspects can first be detained prior to conviction for a maximum period of 90 days, and this term can be extended twice by a maximum of 90 days.
 17. Extensive details about the nature of the 2017 administration can be found in van Ginneken et al. (2018).
 18. The authors noted, however, that the safety and sleep quality scales correlated surprisingly weakly with other scales.
 19. These are the autonomy, prisoner relationships, and reintegration scales.
 20. This decision was taken mostly because the original PCQ asked respondents to skip these questions if they had not previously submitted a complaint; the PA version of the questionnaire eliminated this routing and included items that could be answered regardless of prior usage of the grievance system.
 21. As in the NL, the items in the prison climate scales were supplemented by several additional items and scales, which included new items as well as items drawn from the Dutch supplementary modules.
 22. Note that Tonkin (2016) recommended that residents should be allowed to respond anonymously based on findings from Moos (1974) that social climate was rated more negatively when respondents could remain anonymous. In our study, residents were allowed to remain anonymous but could only be paid for their participation if they provided their ID. More individuals may have chosen to remain anonymous if this did not entail missing out on compensation for the survey. We cannot formally test how this affects responses.
 23. The response rate is calculated over total number of individuals (973) who were officially resident in the surveyed units on the date the survey was conducted in those units.
 24. In the original PCQ, these answer options were labeled as ‘not applicable’. The PA-based research team was concerned this term might not be understood and so opted for labeling this answer option with ‘no opinion’.
 25. Patterns in ‘no opinion’ answers align closely with expectations about how often individuals use certain services. Individuals indicate that they have ‘no opinion’ much more frequently for types of activities that not all incarcerated individuals participate in, such as religious services, work, education, and visits.
 26. The Dutch sample characteristics are taken from Table 1 in Bosma et al. (2020) and from Table 1 in van Ginneken et al. (2018).
 27. In some cases, we have used improved approaches. We flag wherever our approaches differ meaningfully.
 28. Note that many psychometric studies, including Bosma et al. (2020), report CA as a measure of internal consistency. Leading psychologists, however, advise against its use in this way given that internal consistency is a poorly defined concept about which CA provides little information (Sijtsma, 2009). CA is a function of the sum of the interitem covariances and tells us something about the average degree of interrelatedness between items when all item covariances are positive.
 29. Note that there is some disagreement among psychometricians about the definition of validity and whether it can be measured at all (Borsboom and Molenaar, 2015).
 30. Note that Bosma et al. control for regime type in their criterion validity analysis, whereas we control for the type of housing unit on which an individual resides.

31. Note that in many cases, a 'no opinion' answer is not really 'missing' and instead signals that a variable is not defined. For example, for individuals who have never set foot in the visiting room, there is no 'real' value to be imputed on an item like 'the visiting room is pleasant'.
32. These questions are: 'Do you have children?' 'How long have you been in this prison/on this unit' 'In the past month, how often did you use the recreation activities/sports facilities/library/yard time/religious services/education/courses?' 'In the past month, how often did you see the doctor/nurse/dentist/psychologist?' and 'In the past month, how often did you receive a visit?'
33. Results from the other two analyses are available from the authors upon request.
34. Compare with Bosma et al. (2020), Table 1.
35. Compare with Bosma et al. (2020), Table 3.
36. Compare with van Ginneken et al. (2018), Table 2, which presents differences in scale scores between different regime types, including police detention, pretrial detention, extra care units, minimum security regimes, and regimes for persistent offenders. These results are discussed further in Bosma et al. (2020).
37. Note that the results in Table 2 show that the average absolute score on the safety scale is, by some distance, the highest of all scale scores in absolute terms in both PA (3.68) and the NL (4.00).
38. Compare with Bosma et al. (2020), Table 4.
39. The only factor that significantly predicted overall satisfaction with the institution in the NL was country of birth. While country of birth does not hold predictive value in PA, we suspect that results may have been different had there not been such substantial under-response of the foreign-born in the PA sample due to the unavailability of a survey in Spanish. We also observe that respondents on therapeutic communities are significantly less satisfied with the institution than respondents that are housed on other units. This is harder to compare with the results reported in Bosma et al. (2020), who control for regime type rather than the type of housing unit.
40. The items do load on different scales in the two other datasets. Note that the PA research team adapted the items on the reintegration scale to focus specifically on the unit, whereas the items on the availability of meaningful activities scales are asked in generic terms. This may have generated greater differences between scores on the two scales in the PA data.
41. Items with a high prevalence of 'no opinion' answers tend to have lower factor loadings than items with less missing data. Notably, the item on how often individuals can see their children, which had high missing data, displays a low factor loading in the complete case analyses, and an unusually high factor loading in analyses where the data is imputed or deleted pairwise.
42. As highlighted in the introduction, several studies have shown that the EssenCES has good psychometric properties across contexts. While those working with the MQPL instrument have taken a different methodological approach, those adapting the MQPL to different contexts have tended to conclude that many of the items, when appropriately reworded and contextualized, are relevant in different contexts (Crewe et al., 2022; Neubacher et al., 2021; Mjåland et al., 2021).
43. A good example of prison climate research focused on individual experiences of climate is Hans Toch's *Living in Prison* (1977), which explicitly focuses on individual climate preferences.
44. The development of a parallel staff survey—just like the MQPL is complemented by the Staff Quality of Life survey—would moreover provide a meaningful complement to the data from incarcerated people as well as further opportunities for validation. The LIC study 2017 included a parallel staff survey, but the questionnaire was not closely linked to the PCQ. The LIC study 2024 does include a staff survey that includes many of the same items as the PCQ as administered under detainees.

References

- Aebi MF, Cocco E, Molnar L, et al. (2022) *SPACE I—2021—Council of Europe Annual Penal Statistics: Prison Populations*. Strasbourg: Council of Europe.
- Allison PD (2010) Missing data. In: Wright JD and Marsden PV (eds) *Handbook of Survey Research*, 2nd ed. Bingley: Emerald, 631–657.
- Armstrong GS and MacKenzie DL (2003) Private versus public juvenile correctional facilities: Do differences in environmental quality exist? *Crime & Delinquency* 49(4): 542–563.
- Auty KM and Liebling A (2020) Exploring the relationship between prison social climate and reoffending. *Justice Quarterly* 37(2): 358–381.
- Auty KM and Liebling A (2024) What is a ‘good enough’ prison? An empirical analysis of key thresholds using prison moral quality data. *European Journal of Criminology* 21(5): 725–753.
- Beijersbergen K (2016) *Ontwikkeling van de Leefklimaat Vragenlijst Penitentiaire Inrichtingen*. Leiden: Leiden University. Available at: <https://www.dji.nl/documenten/publicaties/2016/01/01/eindrapport-ontwikkeling-van-de-leefklimaat-vragenlijst-penitentiaire-inrichtingen> (accessed 24 July 2024).
- Borsboom D and Molenaar D (2015) Psychometrics. In: Smelser NJ and Baltes PB (eds) *International Encyclopedia of the Social & Behavioral Sciences*. 2nd ed. Amsterdam: Elsevier, 418–422. Available at: https://brill.com/display/book/edcoll/9789004481916/B9789004481916_s031.xml (accessed 12 March 2024).
- Bosma AQ, van Ginneken E, Palmen H, et al. (2020) A new instrument to measure prison climate: The psychometric quality of the prison climate questionnaire. *The Prison Journal* 100(3): 355–380.
- Bouloukos AC and Dammann B (2001) The United Nations and the promotion of prison standards. In: Van Zyl Smit D and Dünkel F (eds) *Imprisonment Today and Tomorrow*. The Hague: Kluwer Law International, 756–774. Available at: https://brill.com/display/book/edcoll/9789004481916/B9789004481916_s031.xml (accessed 21 August 2023).
- Camp SD, Gaes GG, Klein-Saffran J, et al. (2002) Using inmate survey data in assessing prison performance a case study comparing private and public prisons. *Criminal Justice Review* 27(1): 26–51.
- Cochran JC and Mears DP (2013) Social isolation and inmate behavior: A conceptual framework for theorizing prison visitation and guiding and assessing research. *Journal of Criminal Justice* 41(4): 252–261.
- Couper MP and de Leeuw E (2002) Nonresponse in cross-cultural and cross-national surveys. In: Harkness JA, Van De Vijver FJR and Mohler PPH (eds) *Cross-Cultural Survey Methods*, 1st ed. Hoboken, NJ: Wiley-Interscience, 157–177.
- Crewe B, Ievins A, Larmour S, et al. (2022) Nordic penal exceptionalism: A comparative, empirical analysis. *The British Journal of Criminology* 63(2): 424–443.
- Crewe B, Liebling A and Hulley S (2015) Staff-prisoner relationships, staff professionalism, and the use of authority in public- and private-sector prisons. *Law & Social Inquiry* 40(02): 309–344.
- Day A, Casey S, Vess J, et al. (2011) Assessing the social climate of Australian prisons. Trends & issues in crime and criminal justice No. 427. Canberra: Australian Institute of Criminology. Available at: <https://researchonline.jcu.edu.au/47702>, <https://doi.org/10.52922/ti260013> (accessed 24 July 2024).
- Day A, Casey S, Vess J, et al. (2012) Assessing the therapeutic climate of prisons. *Criminal Justice and Behavior* 39(2): 156–168.
- De Jong EWA, Palmen H, Ramakers AAT, et al. (2022) Unraveling the black box of prison visitation: Incarcerated individuals’ and visitors’ conversations and feelings during visitation hour. *Crime & Delinquency* 70(6–7): 001112872211253.

- de Loeff M, van de Haar van Gemmert N, et al. (2018) *DJI in Getal 2013-2017*. The Hague: Ministerie van Justitie en Veiligheid.
- de Vet HCW, Mokkink LB, Mosmuller DG, et al. (2017) Spearman–Brown prophecy formula and Cronbach’s alpha: Different faces of reliability and opportunities for new applications. *Journal of Clinical Epidemiology* 85: 45–49.
- Edgar K, O’Donnell I and Martin C (2003) *Prison Violence: Conflict, Power and Victimization*. London: Willan. Available at: <https://www.taylorfrancis.com/books/9781317829102> (accessed 17 March 2023).
- Elbers JM, Ginneken E, Boone M, et al. (2021) Straffen en belonen in detentie. *Tijdschrift Voor Criminologie* 63(3): 263–291.
- Fauskanger J, Jakobsen A, Mosvold R, et al. (2012) Analysis of psychometric properties as part of an iterative adaptation process of MKT items for use in other countries. *ZDM Math Educ* 44(3): 387–399.
- Fitzgerald R and Jowell R (2010) Measurement equivalence in comparative surveys: The European Social Survey (ESS)-from design to implementation and beyond. In: Harkness JA, Braun M, Edwards B, Johnson TP, Lyberg L, Mohler PP, Pennell B-E and Smith TW (eds) *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*. John Wiley & Sons, 485–495. Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470609927.ch26> (accessed 23 June 2025).
- Furr RM (2017) *Psychometrics: An Introduction*, 3rd ed. Los Angeles: Sage Publications.
- Gonçalves LC, Endrass J, Rossegger A, et al. (2016) A longitudinal study of mental health symptoms in young prisoners: Exploring the influence of personal factors and the correctional climate. *BMC Psychiatry* 16(1): 91.
- Harding R (2014) Rehabilitation and prison social climate: Do ‘what works’ rehabilitation programs work better in prisons that have a positive social climate? *Australian & New Zealand Journal of Criminology* 47(2): 163–175.
- Harkness JA (2011) Comparative survey research. In: Leeuw ED, Hox JJ and DA D (eds) *International Handbook of Survey Methodology*. London: Routledge, 56–77. Available at: <https://www.taylorfrancis.com/books/9780203843123> (accessed 25 June 2023).
- Hayton JC, Allen DG and Scarpello V (2004) Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods* 7(2): 191–205.
- Howells K, Tonkin M, Milburn C, et al. (2009) The EssenCES measure of social climate: A preliminary validation and normative data in UK high secure hospital settings. *Criminal Behaviour and Mental Health* 19(5): 308–320.
- Johnsen B and Granheim PK (2011) Prison size and quality of life in Norwegian closed prisons in late modernity. In: Ugelvik T and Dullum J (eds) *Penal Exceptionalism? Nordic Prison Policy and Practice*. London: Routledge, 199–214. <https://www-taylorfrancis-com.ezproxy.uio.no/chapters/edit/10.4324/9780203813270-15/prison-size-quality-life-norwegian-closed-prisons-late-modernity-berit-johnsen-per-kristian-granheim2> (accessed June 25, 2023).
- Johnsen B, Pape H, Fransson E, et al. (2023) *Arkitektur og livskvalitet i Modell 2015 fengsler: En undersøkelse av soningsklimaet i standardiserte fengselsbygg*. Forskningsrapport. KRUS. Available at: <https://hdl.handle.net/11250/3074438> (accessed June 25, 2023).
- Krause M, Kuba R and Lategan D (2017) *Annual statistical report*. Pennsylvania Department of Corrections. Available at: <https://www.pa.gov/content/dam/copapwp-pagov/en/cor/documents/resources/statistics/reports-and-dashboards/2017%20Annual%20Statistical%20Report.pdf> (accessed 24 July 2024).
- Kuba R (2021) *Annual statistics report*. Pennsylvania Department of Corrections. Available at: <https://www.pa.gov/content/dam/copapwp-pagov/en/cor/documents/resources/statistics/reports-and-dashboards/2021%20Annual%20Statistical%20Report.pdf> (accessed 24 July 2024).

- Leitgöb H, Seddig D, Asparouhov T, et al. (2023) Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research* 110: 102805.
- Liebling A (2011) Moral performance, inhuman and degrading treatment and prison pain. *Punishment & Society* 13(5): 530–550.
- Liebling A and Arnold H (2004) *Prisons and Their Moral Performance: A Study of Values, Quality, and Prison Life*. Oxford: Oxford University Press.
- Liebling A, Crewe B and Hulley S (2012) Conceptualising and measuring the quality of prison life. In: Gadd D, Karstedt S and Messner SF (eds) *The SAGE Handbook of Criminological Research Methods*. London: Sage Publications, 358–372. Available at: <http://methods.sagepub.com/book/sage-hdbk-criminological-research-methods> (accessed 2 February 2022).
- Long CG, Anagnostakis K, Fox E, et al. (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* 21(3): 202–214.
- Maes EE, Robert L, Goossens E, et al. (2023) Victimization in prison. A study of victimisation and prison climate dimensions in Belgian prisons—KU Leuven. Available at: https://kuleuven.limo.libis.be/discovery/fulldisplay/lirias4089037/32KUL_KUL:Lirias (accessed 24 August 2023).
- Mjåland K, Laursen J, Schliehe A, et al. (2021) Contrasts in freedom: Comparing the experiences of imprisonment in open and closed prisons in England and Wales and Norway. *European Journal of Criminology* 20(5): 1641–1662.
- Moos RH (1974) *The Correctional Institutions Environment Scale Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Mulvey EP, Schubert CA and Odgers CA (2010) A method for measuring organizational functioning in juvenile justice facilities using resident ratings. *Criminal Justice and Behavior* 37(11): 1255–1277.
- National Institute of Corrections (2019) <https://nicic.gov/resources/nic-library/state-statistics/2019/pennsylvania-2019>.
- Neubacher F, Liebling A and Kant D (2021) Same problems, different concepts and language: What happens when prison climate research goes on a journey? *European Journal of Criminology* 20(4): 1446–1463.
- Nunnally JC and Bernstein IH (1994) *Psychometric Theory*. New York City, NY: McGraw-Hill.
- Ouaknine (2023). Prison Climate in Israel. In: *Paper presented at the workshop of the Working group on Prison Research of the European Society of Criminology*, Ashkelon Isreal, March 27 2023.
- Revelle W and Condon DM (2018) Reliability. In: Irwing P, Booth T and Hughes DJ (eds) *The Wiley Handbook of Psychometric Testing*. Hoboken, NJ: Wiley, 709–749. Available at: <https://onlinelibrary.wiley.com/doi/10.1002/9781118489772.ch23> (accessed 23 July 2024).
- Riegel B, Carlson B and Glaser D (2000) Development and testing of a clinical tool measuring self-management of heart failure. *Heart & Lung* 29(1): 4–15.
- Ross MW, Diamond PM, Liebling A, et al. (2008) Measurement of prison social climate: A comparison of an inmate measure in England and the USA. *Punishment & Society* 10(4): 447–474.
- Saylor WG (1984) *Surveying prison environments*. Federal Bureau of Prisons. Available at: https://www.bop.gov/resources/research_projects/published_reports/cond_envir/oresaylor2.pdf (accessed 24 July 2024).
- Schalast N and Laan JM (2017) Measuring social climate in German prisons using the Essen climate evaluation schema. *The Prison Journal* 97(2): 166–180.

- Schalast N, Redies M, Collins M, et al. (2008) EssenCES, a short questionnaire for assessing the social climate of forensic psychiatric wards. *Criminal Behaviour and Mental Health* 18(1): 49–58.
- Schneider B, Ehrhart MG and Macey WH (2013) Organizational climate and culture. *Annual Review of Psychology* 64(1): 361–388.
- Schubert CA, Mulvey EP, Loughran TA, et al. (2012) Perceptions of institutional experience and community outcomes for serious adolescent offenders. *Criminal Justice and Behavior* 39(1): 71–93.
- Siennick SE, Mears DP and Bales WD (2013) Here and gone: Anticipation and separation effects of prison visits on inmate infractions. *Journal of Research in Crime and Delinquency* 50(3): 417–444.
- Siess J and Schalast N (2017) Psychometric properties of the Essen Climate Evaluation Schema (EssenCES) in a sample of general psychiatric wards. *Archives of Psychiatric Nursing* 31(6): 582–587.
- Sijtsma K (2009) On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74(1): 107–120.
- Smith GT (2005) On construct validity: Issues of method and measurement. *Psychological Assessment* 17(4): 396–408.
- Stekhoven DJ and Buhlmann P (2012) MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics (Oxford, England)* 28(1): 112–118.
- Tavakol M and Dennick R (2011) Making sense of Cronbach's alpha. *International Journal of Medical Education* 2: 53–55.
- Toch H (1977) *Living in Prison: The Ecology of Survival*. Florence, MA: Free Press.
- 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.
- Tonkin M, Howells K, Ferguson E, et al. (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.
- Turanovic JJ and Tasca M (2019) Inmates' experiences with prison visitation. *Justice Quarterly* 36(2): 287–322.
- van der Helm P, Stams GJ and van der Laan P (2011) Measuring group climate in prison. *The Prison Journal* 91(2): 158–176.
- van Ginneken EFJC and Nieuwbeerta P (2020) Climate consensus: A multilevel study testing assumptions about prison climate. *Journal of Criminal Justice* 69: 101693.
- van Ginneken EFJC and Palmen H (2022) Is there a relationship between prison conditions and recidivism? *Justice Quarterly* 40(1): 106–128.
- van Ginneken EFJC, Palmen H, Bosma AQ, et al. (2018) The life in custody study: The quality of prison life in Dutch prison regimes. *Journal of Criminological Research, Policy and Practice* 4(4): 253–268.
- van Zyl Smit D (2010) Regulation of prison conditions. *Crime and Justice* 39(1): 503–563.
- Welsh WN (2007) A multisite evaluation of prison-based therapeutic community drug treatment. *Criminal Justice and Behavior* 34(11): 1481–1498.
- Western D and Rosenthal R (2003) Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology* 84(3): 608–618.
- Wright KN (1985) Developing the prison environment inventory. *Journal of Research in Crime and Delinquency* 22(3): 257–277.
- Young JTN, Meyers TJ and Morse SJ (2023) What is “prison culture”? Developing a theoretical and methodological foundation for understanding cultural schema in prison. *Criminology: An Interdisciplinary Journal* 61(3): 421–448.