Journal of Criminal Justice 69 (2020) 101693

Contents lists available at ScienceDirect

Journal of Criminal Justice

journal homepage: www.elsevier.com/locate/jcrimjus

Climate consensus: A multilevel study testing assumptions about prison climate

Esther F.J.C. van Ginneken*, Paul Nieuwbeerta

Leiden University, the Netherlands

ARTICLE INFO

Keywords:
Prison climate
Institutional climate
Quality of prison life
Perceptions
Perceptual agreement

ABSTRACT

Background: Prison climate is commonly regarded as an important feature of a prison. However, the state of the art in the current literature is that the theoretical assumptions underlying the concept of prison climate are unclear, and empirical studies on prison climate lack precision.
Purpose: The aims of this paper are to provide (1) an overview of the general theoretical model, (2) a review of the empirical literature and its shortcomings in relation to the model, and (3) to test the relevant theoretical assumptions regarding the existence of shared and distinct prison climates.
Methods: This is done using high-quality data from the Life in Custody Study, a nation-wide survey study among men and women imprisoned in the Netherlands (N = 4538).
Results: Supporting evidence is found for the existence of a shared prison climate, which can distinguish between prisons and prison units. Yet, most variation rests with individual perceptions. Contextual effects of different dimensions of climate are identified for well-being and misconduct.
Conclusions: Distinct prison climates exist for prison units, and to a lesser extent prisons, and they can be detected with sufficiently large samples. Adequate methods are required to measure prison climate, including climate consensus, and study its effects.

1. Introduction

Prison climate has received a great deal of attention in the management and organization of prisons worldwide. To people who have stepped foot in several prisons, be it as incarcerated individual, visitor, employee, or researcher, it is intuitively clear that each prison has its own atmosphere:

> Reflecting collectively in the period following our fieldwork, we were confident that, were we to be deposited on a wing in any of the prisons in our study, we would very soon be able to sense whether it was publicly or privately managed. (Crewe, Liebling, & Hulley, 2014, p. 394)

Prison climate captures the idea that the perception of the environment matters: it can be defined as the *perceived* quality of conditions of imprisonment, including interpersonal, material, and organizational dimensions; the institutional equivalent of a personality (Moos, 1997). Differences in climate across prisons or prison units are considered highly relevant, since a better prison climate is assumed to be related to more positive behavior and well-being of incarcerated individuals. It is even considered one of the aspects that can make prisons more or less survivable (Liebling, 2011). What is more, prison climate is presumed to be related to release outcomes and may therefore even be seen as an instrument to reduce levels of recidivism (Auty & Liebling, 2019). Therefore, Prison Services worldwide view improving prison climate as an important intervention to improve well-being and behavior of incarcerated persons, and prison staff.[1]

Prison climate is not only regarded as important by practitioners and policymakers; it also is a central concept in criminological and prison literature. Many prison researchers, for example, have derived inspiration from Gresham Sykes' study of a maximum-security prison, in which he described the 'pains of imprisonment' and how the incarcerated men adapted and took on various 'argot roles' (Sykes, 1958). Prison climate embodies the idea that pains of imprisonment can vary,

* Corresponding author at: Institute of Criminal Law and Criminology, Leiden University, PO Box 9520, 2300 RA Leiden, the Netherlands.
E-mail address: e.f.j.c.van.ginneken@law.leidenuniv.nl (E.F.J.C. van Ginneken).

[1] The US Bureau of Justice statistics carries out a periodic survey of prison inmates (https://www.bjs.gov/index.cfm?ty=dcdetail&iid=488). The UK Ministry of Justice carries out a routine 'decency audit' in its prisons, using the MQPL survey (https://data.justice.gov.uk/prisons/prison-reform/mqpl-decency). The Dutch Prison Service carries out a period survey among prisoners into prison climate (https://www.dji.nl/themas-cijfers-en-publicaties/cijfers-en-publicaties/leefklimaat-vragenlijst-penitentiaire-inrichtingen.aspx).

https://doi.org/10.1016/j.jcrimjus.2020.101693
Received 1 April 2020; Received in revised form 30 April 2020; Accepted 30 April 2020
Available online 04 June 2020
0047-2352/ © 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

depending on the conditions of confinement. Over time, instruments have been developed to measure different aspects of prison climate (Liebling, assisted by Arnold, 2004; Liebling, Hulley, & Crewe, 2011; Moos, 1975; Wright, 1985), and the relationship between prison climate and well-being, misconduct, attitudes, and recidivism has been studied. For instance, positive individual perceptions of prison climate have been associated with better well-being and mental health (Gonçalves, Endrass, Rossegger, & Dirkzwager, 2016; Van Ginneken, Palmen, Bosma, & Sentse, 2019), lower misconduct and aggression (Bosma, Van Ginneken, Sentse, & Palmen, 2020; Long et al., 2011; Van der Helm, Stams, Van Genabeek, & Van der Laan, 2012), and lower rates of recidivism (Schubert, Mulvey, Loughran, & Losoya, 2012).

Considering that prison climate is such a central theoretical concept and has considerable practical sway, the state of the art of the empirical literature is surprisingly poor. A review of the literature shows, first, that empirical studies tend to make assumptions that are not explicitly stated; second, that these main theoretical assumptions are not properly tested; and third, that tests of the presumed consequences are carried out poorly, because these implicit assumptions are ignored. In sum, current research on the existence and consequences of prison climate is not adequately linked to a clear theoretical framework. With this paper, we build on the existing literature on prison climate, but aim to make progress by explicating what exactly is meant with the concept 'prison climate' and what the (often implicit) assumptions are. Furthermore, we make progress by empirically testing the theoretical assumptions in an optimal way. Therefore, the aims of this paper are to provide (1) an overview of the general theoretical model, by synthesizing the current theoretical literature, (2) a review of the empirical literature and a discussion of its main shortcomings in relation to the theoretical model, and (3) a test of the most relevant hypotheses that can be derived from the theoretical model, using unique high-quality empirical data and appropriate research methods. In doing so, we also aim to provide a blueprint for future studies on prison climate.

2. A theoretical model of prison climate

The origins of the focus on prison climate may be traced back to Donald Clemmer's and Gresham Sykes' prison studies, in which they described the prisoner community and adaptation to imprisonment (Clemmer, 1940; Sykes, 1958). Since then, scholars have further laid foundations of a theoretical model of prison climate and refined this model with their ideas (Liebling, assisted by Arnold, 2004; Moos, 1975; Saylor, 1984; Schalast, Redies, Collins, Stacey, & Howells, 2008; Toch, 1977; Van der Helm, Stams, & Van der Laan, 2011; Van Ginneken, Palmen, Bosma, Nieuwbeerta, & Berghuis, 2018; Wright, 1985). In this study, we refrain from giving a complete overview of the determinants and consequences of prison climate (but see, for example, Auty & Liebling, 2019; Bosma et al., 2020; Casey, Day, & Reynolds, 2016; Crewe, Liebling, & Hulley, 2011; Harding, 2014; Liebling, assisted by Arnold, 2004). Instead, we focus on the prison climate theoretical model itself.

A review of the literature on prison climate reveals that scholars differ in the concepts they use, the language and terminology they use, and the assumptions they make. Nevertheless, it appears that the underlying theoretical framework across all these studies is very similar, albeit often implicit. Without pretending that we can do full justice to the details of each prison climate study's theoretical framework, we can summarize the general theoretical framework as follows:

2.1. The first core idea is that prison climate reflects the subjective quality of prison life and is thus based on perceptions

One of the issues with prison climate is that it has many guises: it is referred to as institutional climate, atmosphere, perceived quality, therapeutic climate, prison (social) climate, and (moral) quality of prison life. Regardless of terminology, the defining feature of prison climate is that it is inherently *subjective*, reflecting how an environment is perceived. As such, it is different from objective indicators of quality and deprivation, such as levels of crowding, security, and time-out-of-cell.

Differences in (the perceptions of) prison climate of each prison (unit) are assumed to be determined by (1) differences in institutional characteristics of the prison (unit) and experiences, such as whether a prison is publicly or privately managed, daily programming, security, and staff characteristics and orientations; and (2), different individual characteristics and experiences of individuals incarcerated, including, for example, offense histories, sentences, demographic characteristics, and social network.

Due to the effects of different individual characteristics and experiences, it is unlikely that there is full perceptual agreement on prison climate within prison (units): a substantial component of variance in perceptions of the environment is expected, due to differences in individual experiences and sensibilities. While individuals incarcerated share some aspects of prison life, such as the environment, programming, and staff, their day-to-day experiences with each of these are unique. For example, they have personal interactions with members of staff, individual visit experiences, preferences, and vulnerabilities that influence activities and perceptions.

2.2. The second core idea is that prison climate is conceived of as a multidimensional construct

A review of the literature shows that almost all theoretical and empirical studies distinguish between distinct concepts and dimensions (Liebling, assisted by Arnold, 2004; Van Ginneken et al., 2018). The most prominent dimensions are peer relationships (*e.g.,* 'relationships', 'cohesion', and 'harmony'); relationships with staff (*e.g.*, 'support', 'professionalism', and 'therapeutic hold'); and safety. A few also measure autonomy (or 'freedom'), activity-related attitudes (*e.g.,* 'personal growth', 'well-being and development', and 'activity'), perceptions of material conditions and facilities, and satisfaction with organizational aspects such as handling of complaints.

2.3. The third core idea is that prison climate is meaningful at the unit or prison level, in the sense that it reflects shared perceptions at these levels

This is a defining feature of prison climate: it is assumed to exist at the *meso*/macro level (and is more than simply the sum of individual perceptions). In other words, the prison literature implicitly assumes that there is some perceptual agreement among people on prison climate, which means that individuals in the same unit and prison agree more on the perceptions of prison climate than with individuals in other units and prisons.

There are different mechanisms that may explain perceptual agreement among individuals in the same unit or prison. Most obviously, agreement may be explained by the shared environment (*structural approach*). It may also be partly explained by individual characteristics that members of the population have in common, due to selective composition (*selection-attraction-attrition*, Joyce & Slocum Jr, 1984). For instance, high security prisons tend to have a population with a different background in terms of offense history, sentences, and in-prison behavior than lower security prisons. Both security level and population composition may contribute to perceptions of prison climate. Moreover, agreement may occur through perceptual formation, where individuals perceive a situation similarly, as well as perceptual adjustment resulting from social interaction (Joyce & Slocum Jr, 1984).

2.4. The fourth core idea is that prison climate affects individual well-being and behavior, above and beyond effects of individual characteristics of incarcerated persons

Prison climate is assumed to have important consequences on

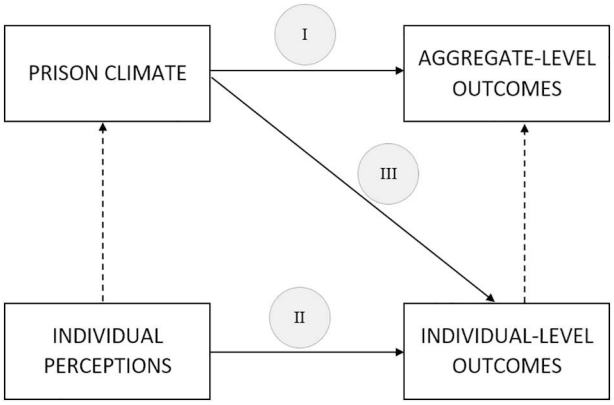


Fig. 1. Theoretical model of the levels of analysis of prison climate and its effects.

behavior and well-being, during and after imprisonment. This means it matters *where* individuals are incarcerated. So, a main feature of the theoretical literature on prison climate is that – in addition to effects of individual perceptions of prison climate – the shared perception of prison climate in a prison (unit) is assumed to have an independent effect on individual outcomes (relationship III in Fig. 1).

2.5. Prior empirical research on prison climate

In the past two decades, empirical research on (the determinants and consequences of) prison climate has made enormous progress. First, more adequate instruments to measure prison climate have been developed. Second, an increasing number of studies surveying large samples of incarcerated individuals have been carried out. Third, more studies have included multiple sites, *i.e.*, prisons and prison units. Fourth, multilevel analysis has been increasingly applied in prison climate and related research. Despite the strengths of existent research, there are also serious shortcomings. In particular, the current research and measures of prison climate are not adequately linked to the above discussed theoretical framework.

While most prior studies do recognize the first core idea of the prison climate theoretical model as outlined above, namely that prison climate reflects the subjective quality of prison life, they do not always measure perceptions of the conditions of confinement adequately. As revealed by Tonkin's (2016) review of existent surveys on prison climate, only few surveys meet the standards of sufficient reliability and validity. With adult prison populations, the most commonly used instruments are the Correctional Institutions Environment Scale (CIES; Moos, 1975), the Prison Environment Inventory (PEI; Wright, 1985), the Essen Climate Evaluation Schema (EssenCES; Schalast et al., 2008), and the Measurement of Quality of Prison Life (MQPL+; Liebling, assisted by Arnold, 2004). Of these measures, only the EssenCES has good convergent and divergent validity, support for the proposed factor structure, and good internal consistency (Tonkin, 2016). It has also been validated with Australian and English samples (Day, Casey, Vess, & Hulsy, 2011; Tonkin et al., 2012). The Prison Group Climate Instrument (PGCI; Van der Helm et al., 2011) has only been tested with young people in custody and individuals in forensic psychiatric institutions. Most recently, the Prison Climate Questionnaire (PCQ) was developed to overcome limitations of prior surveys and validated among the population of Dutch men and women in prison (Bosma et al., 2020). It was found to have good internal consistency, criterion validity, construct validity, and support was found for the proposed factor structure.

The second core idea that prison climate has a multidimensional structure is not strongly contested (see Liebling, assisted by Arnold, 2004; Tonkin, 2016; Van Ginneken et al., 2018). It is, however, difficult to distinguish the core dimensions of prison climate from the existent literature, given that only two of the above instruments have a well-supported factor structure that has been tested with adults in prison. It would be most helpful to use an instrument that does justice to the multidimensional nature of prison climate, but does not suffer from reliability or validity problems, such as multicollinearity. Valid instruments are important for establishing independent effects of each of the dimensions, and for enabling replication in other studies.

The third core idea that prison climate reflects shared perceptions (*i.e.,* some form of consensus) rather than individual perceptions alone, has been given almost no attention at all (but see Camp, 1999; Mulvey, Schubert, & Odgers, 2010). Disappointingly, despite the availability of multilevel analysis techniques, the current empirical literature does not make a proper distinction between individual and contextual levels of analysis. That is, the vast majority of studies that supposedly measure prison climate (and, more broadly, institutional climate, social climate, and therapeutic climate) using attitude surveys, have analyzed climate only on the individual level (Beech & Hamilton-Giachritsis, 2005; Eidhin, Sheehy, O'Sullivan, & McLeavey, 2002; Kupchik & Snyder, 2009; Schubert et al., 2012; Woessner & Schwedler, 2014). Moreover, some studies claim to discuss the effects of prison climate, but only study a single unit or prison – and thus only examine effects of individual perceptions (and not prison or unit climate) (Gonçalves et al., 2016; Skar et al., 2019; Van der Helm et al., 2012; Van der Helm, Beunk, Stams, & Van der Laan, 2014).

In addition, prior studies that have considered prison climate at the aggregate level did not properly test whether climate existed at the unit or prison level, and thus did not support the idea that the sum of perceptions of individual prisoners is a valid entity. For example, studies using the Measurement of Quality of Prison Life (Liebling, assisted by Arnold, 2004) have not tested to what extent the proportion of variance on each of the climate dimensions were attributable to the institution as opposed to the individuals within the institution (*i.e.,* the intraclass correlation, ICC[1]), nor have they reported the consistency of mean differences across institutions (ICC[2]) (Auty & Liebling, 2019; Crewe et al., 2011; Liebling & Ludlow, 2016; Ross, Diamond, Liebling, & Saylor, 2008; Ross, Liebling, & Tait, 2011). In the development of the Prison Environment Inventory (PEI), Wright (1985) touched upon this issue by examining to what extent six individually phrased items (*e.g.,* 'How do you feel today?') were able to explain variance in the scale scores about prison climate (*e.g.,* 'The guards ask inmates about their personal feelings'). He concluded that 4% to 14% of variance could be explained by individual differences (measured by questions such as 'How do you feel today?'). This study did not investigate the level of perceptual agreement among people in different prisons. To our knowledge, the only recent large-scale empirical study that provided such useful information, was a study with adolescents in juvenile justice facilities in the US (Mulvey et al., 2010). They found unusually high ICC(1)s, with an average of 0.23 (range 0.4–0.52), which means that a substantial portion of differences in individual ratings is attributable to between-institution differences. The consistency of institution means varied from a very low 0.08 to a high 0.85 (M = 0.52). These values should be interpreted with great caution, given that the median number of respondents in each institution was 2 (the ICCs were only calculated for the 111 institutions with 2 or more respondents).

The fourth core idea that prison climate affects important outcomes beyond individual characteristics has not been properly tested. The majority of studies have only examined relationships between individual perceptions and individual outcomes (relationship II in Fig. 1) (Bosma et al., 2020; Gonçalves et al., 2016; Schubert et al., 2012; Van der Helm et al., 2012). Some prison climate studies have examined aggregate-level predictors and linked these to aggregate-level outcomes (relationship I in Fig. 1), including aggression and treatment orientation of units in a secure hospital (Long et al., 2011), staff attitudes and institutional aggression (Tonkin et al., 2012), and recidivism rates per

prison (Auty & Liebling, 2019). So far, only very few studies have studied the effect of prison climate on the individual and *meso*/macro level simultaneously, relationship II and III in Fig. 1 (Van Ginneken et al., 2019; Wright, 1991). Yet, this type of multilevel analysis should arguably be the key focus of prison climate research, as it examines the contribution of contextual climate effects above and beyond individual-level effects.

To summarize, the current literature on prison climate has major shortcomings particularly in relation to the level of analysis: it is not yet clear to what extent prison climate is an individual-level *versus* a prison (unit) phenomenon. Moreover, the different dimensions are not consistently measured, and existent instruments are often not sufficiently validated. Most importantly, the existent research does not do justice to the central theoretical framework of prison climate, which presumes that climate exists at the *meso*/macro level and has a contextual effect on outcomes. These concerns echo debates that have already been rehearsed to some extent in literature on climate in schools and organizations (Schneider, Ehrhart, & Macey, 2013). In this body of literature, it is recognized that climate can exist at multiple levels, including the individual level when it refers to individual perceptions (*i.e.*, psychological climate) and the group- or organizational level (*i.e.,* organizational climate) when these individual climate perceptions are aggregated (James et al., 2008). However, "in order to aggregate psychological climate scores, a shared meaning must be demonstrated via perceptual agreement" (James et al., 2008, p. 16).

3. Current study

This study tries to fill this gap in the current literature on prison climate by addressing the shortcomings using survey data collected among adult men and women incarcerated in the Netherlands. In doing so, we seek to answer the following research questions: (1) To what extent does each prison or prison unit have a distinct and shared prison climate? and (2) Is *shared* prison climate associated with different types of outcomes?

Aggregation of prison climate scores can only be meaningful if it measures elements to which most or all individuals in a unit or prison have the same exposure. Following this, it is expected that the level of perceptual agreement will be higher within units than within prisons, given that each prison unit in the Netherlands has its own daily programming and regime. Furthermore, staff assignment to each unit tends to be relatively stable. In the Netherlands, there are no separate institutions for individuals with pre-trial or convicted/sentenced status; instead, there are separate regimes within prisons for each of these populations. There are also further specialized regimes, namely: short-stay custody (*e.g.,* for individuals who violated sentence conditions), extra-care units (for vulnerable people due to mental health problems or the nature of their index offense), persistent offenders, terrorist regimes, and minimum security regimes. These regimes differ in terms of programming and out-of-cell time, as well as composition of the population and turnover (which is especially high in pre-trial regimes and short-stay custody). There may be contact between prisoners from different units during activities, but association (time-out-of cell not in activities) is spent on the unit. Overall, the conditions within units are fairly homogeneous, but within prisons there exists substantial heterogeneity. Nevertheless, there are also characteristics shared across an entire prison, most notably the building and management. We formulate the following hypotheses:

Hypothesis 1. There are measurable differences in prison climate across (a) prisons and (b) prison units;

Hypothesis 2. There is some perceptual agreement among incarcerated individuals (a) within the same prison and, (b) to a greater extent, within the same prison unit, regarding prison climate;

Hypothesis 3. A positive prison climate has a positive relationship with (a) psychological well-being and (b) subjective well-being, and a negative relationship with (c) victimization and (d) misconduct, beyond individual-level effects.

4. Methodology

4.1. Life in Custody Study

To address the formulated research questions, data are used from the Life in Custody Study, a nation-wide survey of adults (men and women) incarcerated in the Netherlands (see Van Ginneken et al., 2018) for detailed information about the study). Data for this study were collected between January and April 2017, using the Prison Climate Questionnaire (Bosma et al., 2020). The Life in Custody Study is a collaboration between the Dutch Prison Service and Leiden University, and serves the dual purpose of performance/service monitoring and improvement, and independent research on the experience of imprisonment.

4.2. Sample and procedure

Our aim was to maximize response rates through ensuring confidentiality and personal contact in the distribution and collection of questionnaires. A team of research assistants visited each prison for one week, to contact each person incarcerated and explain the study. Participation was voluntary and confidential. A small incentive was handed out to each person contacted, regardless of their decision to participate. We offered assistance with the completion of the questionnaire if desired. Participants had to give informed consent for the use of their data for research purposes and for the linking with administrative data. It was also possible to participate anonymously if participants did not want their survey responses to be linked to administrative data.

We invited all individuals who were incarcerated at the time of our visit, who could be approached, to participate in the study. A total of 6088 individuals were approached (86% of the incarcerated population at the time), of whom 4938 completed a questionnaire (81%).[2] There were 400 questionnaires that could not be used for research purposes, which means the final sample consisted of 4538 respondents. They were housed in 244 units across 28 prisons. The population size of most prisons in the Netherlands is between 150 and 500 individuals, while a few relatively large facilities hold between 500 and 800 individuals. The average capacity of units during the time of our survey was 41 beds (SD = 19, min. 7, max. 98) and 21% of our sample reported sharing a cell.[3] At the time of our survey, the average occupation rate of units was 90% and there was no overcrowding.

Compared to the total prison population, there was a slight overrepresentation of respondents who were born in the Netherlands, respondents convicted of drugs offenses, and people incarcerated in pre-trial regimes relative to people in regular prison regimes. There was a slight underrepresentation of respondents convicted of property offenses (see Van Ginneken et al., 2018) for test statistics). Overall, given that the sample comprised the majority of the population, with a reasonable representation of different groups, the sample allows for generalization to the Dutch adult prison population (excluding individuals

[2] We were unable to approach people with severe mental health problems, people held in segregation, or people who were unable to read or speak one of the languages in which the questionnaires were available (Dutch, English and Spanish). People in immigration detention or incarcerated in psychiatric penitentiary institutions were not included in the study.

[3] Large units may have been physically split, but we were unable to split them in the data due to combined staff assignment, for example. Based on our observations, units were typically between 12 and 48 beds in size.

Table 1
Descriptive statistics

		n	*M*	(*SD*)	Range
Level 1 variables	**Prison climate (individual perceptions)**				
	Autonomy	4400	2.71	(0.96)	1–5
	Peer relationships	4425	3.44	(0.71)	1–5
	Staff-prisoner relationships	4375	3.31	(0.89)	1–5
	Safety	4432	4.00	(0.83)	1–5
	Availability of meaningful activities	4389	2.27	(0.96)	1–5
	Overall institution rating	4413	2.92	(1.11)	1–5
	Dependent variables				
	Psychological well-being	4334	3.81	(0.99)	1–5
	Subjective well-being	4297	3.19	(1.06)	1–5
	Misconduct: yes	4083	0.15		0–1
	Victimization: yes	4302	0.25		0–1
	Personal characteristics				
	Age	4538	36.84	(11.74)	18–81
	Country of birth: the Netherlands	4322	0.35		0–1
	Sex: male	4538	0.95		0–1
	Time served (months)	4536	11.91	(21.91)	0–326
	Previous incarceration: no	4535	0.43		0–1
	Double cell: yes	4263	0.21		0–1
Level 2 variables	**Prison climate (unit aggregate)**				
	Autonomy	244	2.70	(0.40)	1.44–4.24
	Peer relationships	244	3.44	(0.25)	2.64–4.13
	Staff-prisoner relationships	244	3.31	(0.36)	2.37–4.51
	Safety	244	4.00	(0.24)	2.97–4.71
	Availability of meaningful activities	244	2.27	(0.37)	1.22–3.50
	Overall institution rating	244	2.92	(0.49)	1.50–4.29
	Institutional characteristics				
	Cell capacity of unit	242	40.99	(19.00)	7–98
	Unit occupancy rate	242	0.90	(0.13)	0.38–1
	Staff-prisoner ratio (FTE per prisoner)	240	0.27	(0.19)	0.11–3.06
	Staff ratio female to male (FTE)	244	0.21	(0.13)	0–1
	Staff work experience (years)	242	18.99	(3.39)	11.53–29.17
	Regime				
	Prison	244	0.35		0–1
	Pre-trial detention	244	0.38		0–1
	Minimum security	244	0.05		0–1
	Short-stay custody	244	0.11		0–1
	Extra care	244	0.06		0–1
	Persistent offenders (ISD)	244	0.05		0–1

Note. For dichotomous variables (with a Range from 0 to 1), the Mean should be interpreted as a proportion (*e.g.*, $M = 0.35$ for Prison means that 35% of participants were incarcerated in regular prison regimes).

in foreign national prisons and psychiatric penitentiary institutions).

4.3. Prison climate questionnaire

The Prison Climate Questionnaire (PCQ) consisted of 136 items; 11 scales covered the dimensions of prison climate (Autonomy, Safety, Relationships in prison, Contact with the outside world, Meaningful activities, and Facilities). There was also one item that asked for an overall rating of the institution. Additionally, scales were included to measure well-being, misconduct and victimization, and background variables. The prison climate scales comprised statements that were rated on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), with higher scores reflecting more positive attitudes. A detailed examination of the psychometric properties of the instrument can be found elsewhere (Bosma et al., 2020); this shows that the scales have good internal consistency with Cronbach's alphas ranging from 0.78 to 0.92, and that they load on separate but correlated factors. Table 1 displays descriptive statistics of relevant sample characteristics and the selected prison climate scales.

5. Measures

5.1. Prison climate

For the purpose of this article, a selection was made of scales that correspond to recurrent concepts in previous research on prison climate, as discussed in the literature review. For each scale, a mean score was calculated for participants who had answered more than half of the scale's items. Safety was measured with five items (*e.g.,* 'There are places in this institution where I feel unsafe') and has a Cronbach's alpha of 0.89. Relationships with staff and procedural justice was measured with eight items (*e.g.,* 'Staff members in this unit treat me fairly') and has a Cronbach's alpha of 0.94. Peer relationships was measured with five items (*e.g.,* 'The prisoners treat each other respectfully here') and has a Cronbach's alpha of 0.86. Availability of meaningful activities was measured with four items (*e.g.,* 'During the daily program I learn useful skills') and has a Cronbach's alpha of 0.91. Autonomy was measured with four items (*e.g.,* 'There is much I can decide for myself here') and has a Cronbach's alpha of 0.86. Finally, the one-item overall institution rating ('Generally speaking, I am satisfied with this institution') was also included.

5.2. Well-being

Two scales measure different dimensions of well-being: psychological health and subjective well-being. This distinction acknowledges the idea that well-being in an affective sense (including life satisfaction and happiness) can vary even in the absence of psychological distress of symptoms of mental health problems, such as depression and anxiety (Van Ginneken et al., 2019). Psychological health was measured with a mean score on the *Kessler Screening Scale of Psychological Distress* (K6, Kessler et al., 2002), which includes six items measured on a 5-point

scale (*e.g.*, 'During the past week, about how often did you feel worthless?'). This validated scale can be used as a screening tool for serious mental illness and has good internal consistency in the present study (α = 0.91). The scale was recoded so that higher scores reflect better psychological health. Subjective well-being, which reflects the affective dimension of well-being, was measured with the sub-scale 'emotional well-being' from the *Mental Health Continuum-Short Form* (Lamers, Westerhof, Bohlmeijer, Ten Klooster, & Keyes, 2011). It consists of 3 items measured on a 5-point scale (*e.g.,* 'During the past month, about how often did you feel happy?'). The scale is validated and has good internal consistency in the present study (α = 0.82).

5.3. Misconduct

A dichotomous measure was created to reflect whether someone had reported that they had committed physical violence (towards a fellow incarcerated person or a member of staff), property damage, theft, or had been in possession of contraband items in the last two months (1 = any misconduct, 0 = no misconduct). Verbal misconduct was excluded.

5.4. Victimization

A dichotomous measure was created to reflect whether someone had reported that they had been victim of physical violence (by a fellow incarcerated person or a member of staff), property damage, or theft in the last two months (1 = victimized, 0 = not victimized).

6. Analytical strategy

In order to establish the level of perceptual agreement on various dimensions of prison climate at the unit and prison level, four aggregation indices were selected from the literature on organizational climate and multilevel analysis: $r_{wg(j)}$,AD_M, ICC(1), and ICC(2) (see Appendix I for a detailed overview). Each of these conveys unique information and has advantages and disadvantages. While $r_{wg(j)}$ and AD_M can be calculated for each unit or prison separately, ICC(1) and ICC(2) communicate something about the sample as a whole.

First, it was calculated to what extent individuals in the same unit and prison agree on their rating of each of the climate scales. Perceptual agreement, or climate strength, reflects the level of consensus among members of a unit or organization in terms of absolute ratings on climate scales (*i.e.,* the extent to which ratings of different individuals within a unit are interchangeable). Perceptual agreement is most commonly measured using $r_{wg(j)}$ (James et al., 2008; James, Demaree, & Wolf, 1984). The index $r_{wg(j)}$ compares actual agreement to chance agreement (by comparing observed within-group variances to a specified null distribution, see Cohen, Doveh, & Nahum-Shani, 2009). It also takes into account that scores of 4 and 5 are more in agreement than scores that are further apart, such as 2 and 5. A difficulty with this index is that an appropriate null distribution must be specified (*i.e.,* agreement that can be expected due to chance, taking into account possible response bias). Typically, values of 0.70 and higher are considered sufficient to justify aggregation (Ginsburg & Oore, 2016). An advantage of $r_{wg(j)}$ is that it can be used as a variable in analyses to reflect climate strength within a unit or prison, as it can be calculated for each unit. Less commonly, the average absolute deviation from the mean (AD_M) is used to demonstrate the degree of interrater (dis)agreement (Cohen et al., 2009); the lower the AD_M, relative to the range of the scale, the higher the within-unit agreement. The average absolute deviation from the item means (AD_M) does not require the specification of a null distribution. Furthermore, it can easily interpreted as it yields an estimate of interrater agreement in the metric of the original scale (Burke, Finkelstein, & Dusig, 1999).

Second, the intraclass correlation (ICC[1]) reflects the proportion of variance that can be explained by group membership, and can be regarded as a combined measure of interrater agreement and reliability (LeBreton & Senter, 2008). Even a small amount of shared variance (ICC[1]) can enable detection of aggregate relationships that are not evident in lower-level data, if the groups are large enough (Bliese, 2000). There is no recommended cut-off value for ICC(1), but an early review of prior studies identified a median ICC(1) value of 0.12 in studies on organizational climate (James, 1982). As long as the ICC(1) is significantly different from zero, multilevel analysis is the preferred method of analysis for climate research (Bryk & Raudenbush, 1992). Even if scores are not aggregated to the group/organizational level, one should take into account the nested nature of the data. The LR test was used to determine if the fit of an empty nested model was superior to a simpler linear model.

Finally, the reliability of the unit and prison means were considered. It is generally thought that high climate strength is associated with stronger and more valid relationships between climate and outcomes, partly due to more reliable means (Schneider et al., 2013). ICC(2) is an index of group-mean reliability, which is a function of sample size and between-group variance relative to within-group variance. Values of 0.70 and higher are normally considered adequate (Bliese, 2000). The higher the ICC(2), the more stable the aggregated mean, and the more precisely the data is able to capture mean differences across units or prisons.

In order to examine if there are meaningful relationships between aggregate-level variables and individual outcomes, multilevel analyses were conducted for the selected prison climate variables and four dependent variables: psychological well-being, subjective well-being, victimization, and misconduct. The analyses were conducted in two steps: first, each dependent variable was regressed on each aggregated prison climate variable separately; second, each dependent variable was regressed on each aggregated unit-level prison climate variable with its individual-level counterpart.

Statistical analyses were conducted using Stata Version 15 (StataCorp, 2017). The *ira* (SSC) procedure (Graf-Vlachy, 2017) was used to calculate $r_{wg(j)}$ and AD_M. Between- and within-group variances to calculate ICC(1) and ICC(2) were generated with *xtmixed*. Multilevel regression analyses were conducted for testing Hypothesis 3, using *xtmixed* (continuous dependent variables) and *xtmelogit* (dichotomous dependent variables). For the purpose of these analyses, all independent variables (unit- and individual-level) were centered on their grand mean.

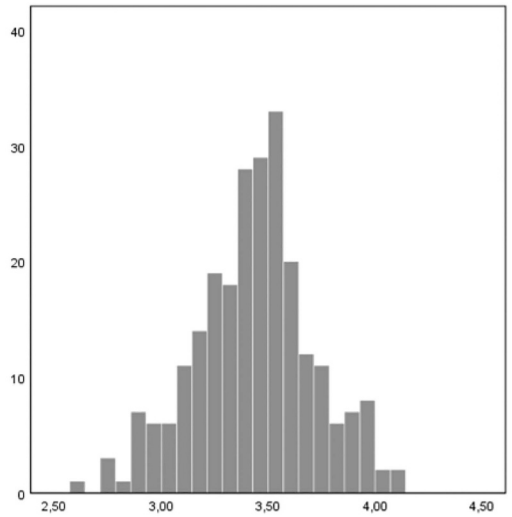
7. Results

In order to test our hypotheses, we first examine the distribution of aggregated prison climate scores. Second, we report a variety of indices to establish perceptual agreement and reliability of the aggregated scores. Third, we investigate the relationship between prison climate and four outcome variables.

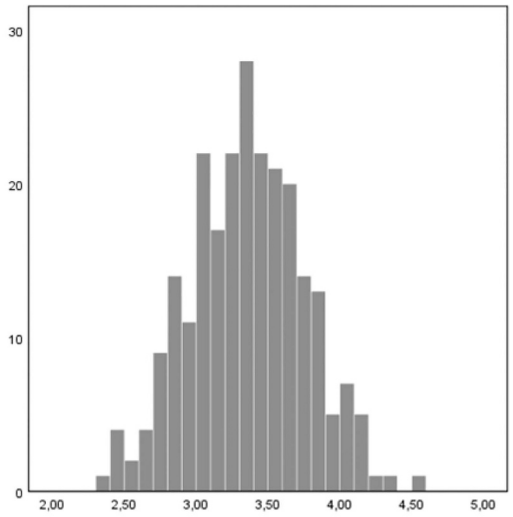
7.1. Shared and distinct prison climates

Inspection of the prison climate scores in the present study shows a normal distribution of unit means (see Fig. 2a-f). This confirms Hypothesis 1a that there is variation in mean scores, and thus in climate perceptions for prison units. The distribution of scores for the prison level (see Fig. 3a-f) revealed variation as well, but within smaller ranges for each of the domains (Hypothesis 1b). This suggests that it may be meaningful to compare prison climate between units, and to a lesser extent between prisons, but the following section looks in more detail at the within-unit/prison agreement and reliability of group-means.

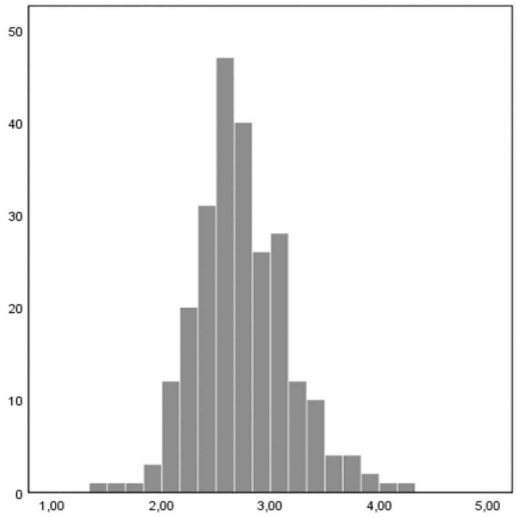
Various indices were inspected to determine (a) the level of agreement among individuals in the same unit or prison, (b) the clustering of variance at the unit and prison level, and (c) the reliability of the mean



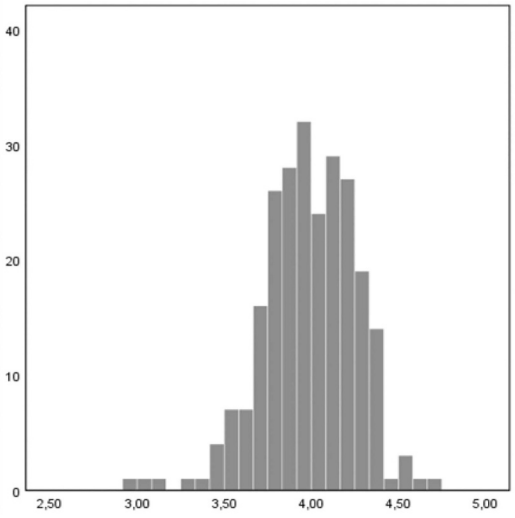
a: Distribution of unit means for 'Peer relationships'



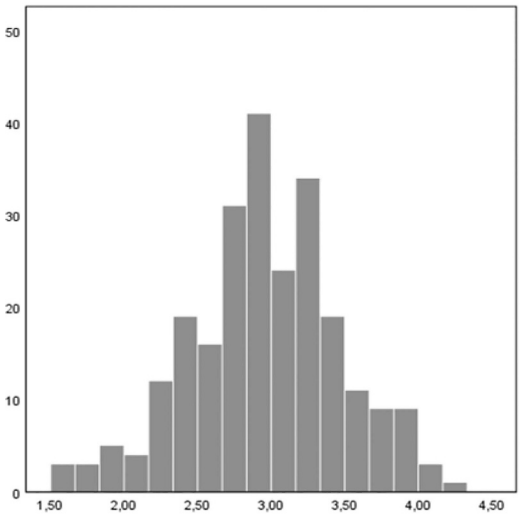
b: Distribution of unit means for 'Staff-prisoner relationships'



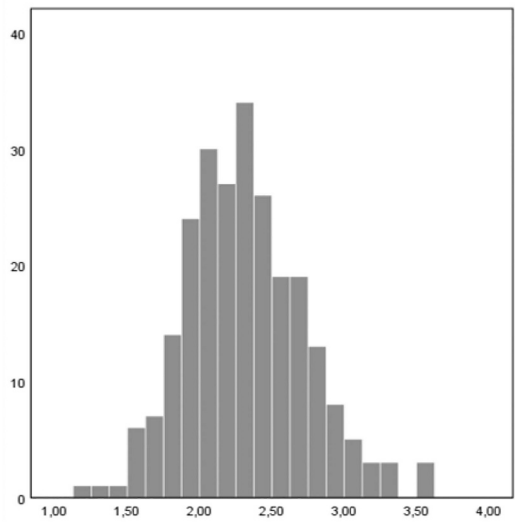
c: Distribution of unit means for 'Autonomy'



d: Distribution of unit means for 'Safety'

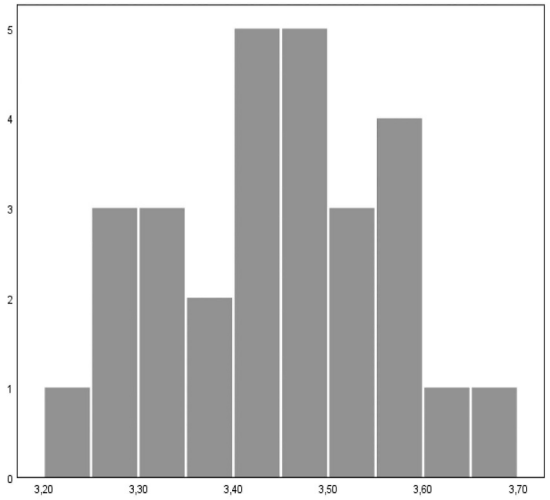


e: Distribution of unit means for 'Total institution rating'

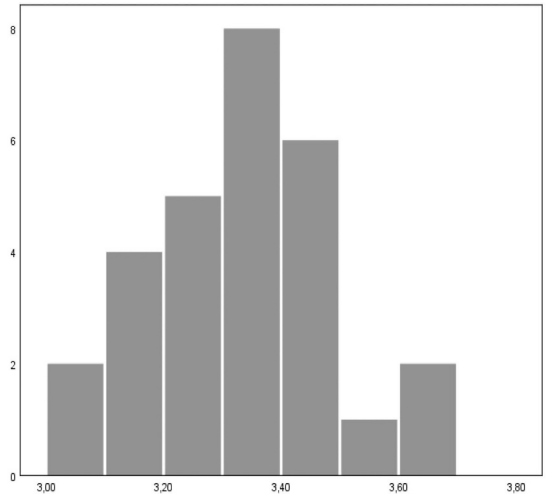


f: Distribution of unit means for 'Meaningful activities'

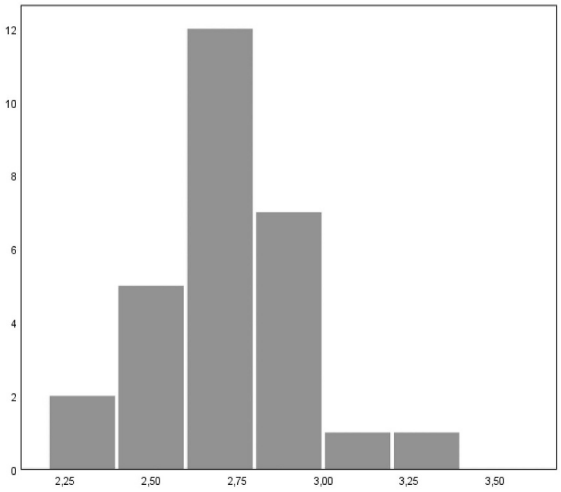
Fig. 2. a Distribution of unit means for 'Peer relationships'. b: Distribution of unit means for 'Staff-prisoner relationships'. c: Distribution of unit means for 'Autonomy'. d: Distribution of unit means for 'Safety'. e: Distribution of unit means for 'Total institution rating'. f: Distribution of unit means for 'Meaningful activities'.



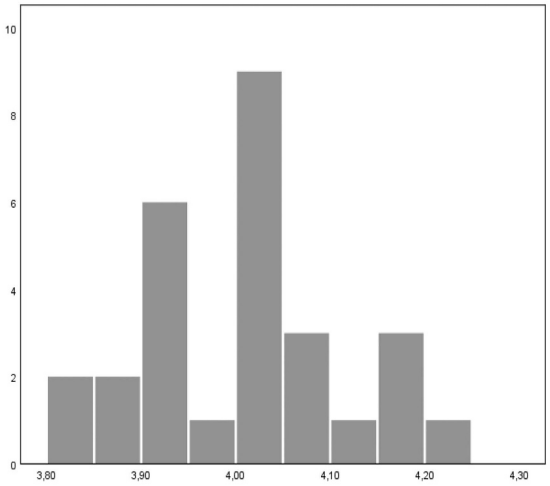
a: Distribution of prison means for 'Peer relations'



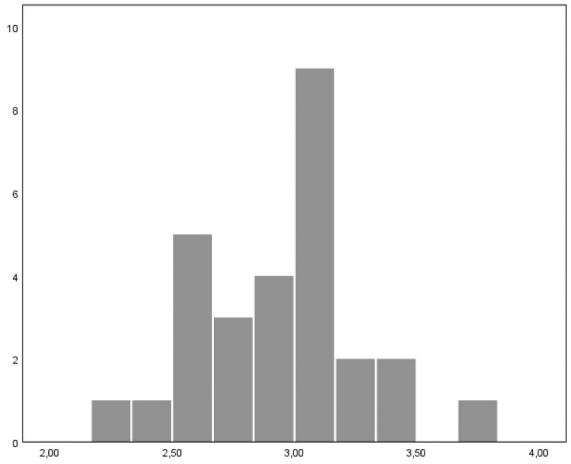
b: Distribution of prison means for 'Staff-prisoner relations'



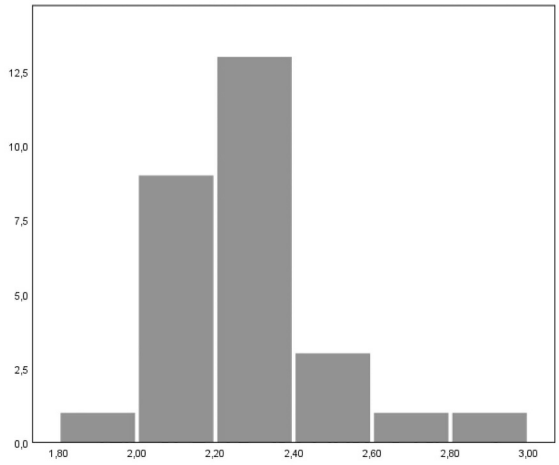
c: Distribution of prison means for 'Autonomy'



d: Distribution of prison means for 'Safety'



e: Distribution of prison means for 'Overall institution rating'



f: Distribution of prison means for 'Meaningful activities'

Fig. 3. a Distribution of prison means for ‘Peer relations’. b: Distribution of prison means for ‘Staff-prisoner relations’. c: Distribution of prison means for ‘Autonomy’. d: Distribution of prison means for ‘Safety’. e: Distribution of prison means for ‘Overall institution rating’. f: Distribution of prison means for ‘Meaningful activities’.

Table 2
Agreement and reliability indices

	Unit				Prison			
	$r_{wg(j)}$	AD_M	ICC(1)	ICC(2)	$r_{wg(j)}$	AD_M	ICC(1)	ICC(2)
Safety	0.81	0.73	0.03*	0.37	0.84	0.73	0.01*	0.60
Staff-prisoner relationships	0.85	0.76	0.12*	0.71	0.86	0.84	0.02*	0.73
Peer relationships	0.88	0.66	0.07*	0.59	0.89	0.71	0.02*	0.75
Meaningful activities	0.74	0.82	0.10*	0.66	0.75	0.88	0.04*	0.86
Autonomy	0.72	0.84	0.13*	0.74	0.70	0.93	0.05*	0.88
Overall institution rating	0.48	0.80	0.15*	0.76	0.43	0.86	0.08*	0.93

Note. $r_{wg(j)}$ = within-group rater agreement (> 0.70 is considered acceptable); AD_M = average deviation from the mean (*M* range 1–5); ICC(1) = between-group variance to total variance; ICC(2) = group-mean reliability (> 0.70 is considered acceptable).
* LR test Prob≥ chibar2 < *0*.05.

scores. The index of perceptual agreement, $r_{wg(j)}$, has average values of 0.70 and higher for each of the prison climate domains (see Table 2) on both the unit and prison level, which is generally considered acceptable to justify aggregation.[4] The exception is the single item 'Overall institution rating', which has an average value of 0.48 for the unit level and 0.43 for the prison level. This can be explained by the fact that an aggregated score based on a single item has greater measurement error. *AD* scores show that the average absolute deviation from the mean is less than 1 for all scales (with a range from 1 to 5), on the unit and prison level (see Table 2).

The intraclass correlations (ICC[1]) show that there is more clustered variance on the unit level than on the prison level for each of the scales (see Table 2). The clustered variance for Safety is minimal at the unit-level (0.03) and prison level (0.01), possibly because there is little variance on this scale more generally (it is concentrated around high safety scores). While the indices for perceptual agreement were lowest for 'Overall institution rating', this single item has the highest ICC(1) value at the unit-level (0.15) and prison-level (0.08). This means that at least some of the variance in prison climate perception is clustered at the unit (and to a lesser extent, the prison) level.

The ICC(2) scores for our data suggest that aggregated prison climate scores reliably reflect differences for most domains (again, except safety) between prisons (see Table 2). The group-mean reliability is lower for the unit-level, which may be explained by the larger group size of prisons. Thus, while there is more clustered variance at the unit-level, the group-mean reliability is higher at the prison-level. Furthermore, 'Overall institution rating' has the lowest $r_{wg(j)}$ but the highest ICC(1) and ICC(2) values, which does suggest that there are grounds to justify aggregation based on the agreement and reliability indices.

Overall, then, an inspection of the various agreement and reliability indices suggests that 1) individuals within units and prisons agree more with each other than with individuals from other units and prisons; 2) there is greater agreement among individuals within the same unit than within the prison as a whole; 3) our data is able to capture mean differences across units and prisons with acceptable to high precision for most prison climate scales; and 4) most variance in prison climate ratings is concentrated on the individual level, as opposed to the unit or prison level. This means that Hypothesis 2 is also supported, and that more agreement is found for prisoners in the same unit (Hypothesis 2a) than in the same prison (Hypothesis 2b).

7.2. Outcomes of prison climate

The results of the multilevel regression analyses (see Table 3) show that there are significant relationships between nearly all aggregated unit-level prison climate variables and psychological well-being, subjective well-being, victimization, and misconduct.[5] With the addition of the individual-level counterparts, many of these significant relationships disappear and instead, they become significant on the individual level. This is not surprising, given that it was established previously that most of the variance is at the individual level. Yet, particularly for misconduct, significant unit-level relationships remain for some of the prison climate variables (*i.e.,* safety, peer relationships, and relationships with staff). For example, the lower the average level of safety experienced on a unit, the more misconduct is reported by individuals in this unit. From these exploratory analyses, it can be concluded that meaningful relationships exist between aggregate prison climate scores and psychological well-being (Hypothesis 3a), subjective well-being (Hypothesis 3b), and misconduct (Hypothesis 3d). No relationship was found between prison climate and victimization, beyond individual-level effects (Hypothesis 3c).

8. Discussion

While prison climate is a frequently employed concept in prison research, a few key assumptions had not yet been systematically examined. We concluded that existent research and instruments were poorly linked to the theoretical model on prison climate. Therefore, this study researched two main questions: first, whether prisons and prison units have distinct and shared prison climates; and second, whether prison climate has meaningful relationships with various outcomes, above and beyond individual-level effects. Based on an empirical analysis of a nation-wide survey of individuals incarcerated in the Netherlands, it can be concluded that shared and distinct prison climates exist and can be detected with a sample large enough. In other words, there is evidence that some level of consensus on prison climate exists. In line with our expectations, given the more homogeneous circumstances in units *versus* prisons, we found a greater amount of clustered variance on the unit-level. However, it was also found that, by far, most variance on prison climate variables was concentrated at the individual level, which suggests that prison climate appears to be shared only to a small extent; for the remainder it can be best conceptualized as individual perceptions. Furthermore, the different dimensions of prison climate had strong relationships with various outcomes, although many of these effects were absorbed by the individual perceptions when these were added to the model.

This is the first study with incarcerated men and women to consider whether there is perceptual agreement in their perceived conditions of imprisonment. The average proportion of variance in prison climate that could be attributed to the unit was 10% (without the prison-level taken into account) and to the prison 4% (without the unit-level taken

[4] A uniform null distribution was chosen, given that there was no *a priori* expectation about the distribution of scores.

[5] The choice was made to report unit-level (rather than prison-level) effects, given that prior results suggested that more meaningful relationships can be expected at this level.

Table 3
Relationships between the aggregated prison climate variables and various dependent variables

	Model 1								Model 2							
	Psychological well-being		Subjective well-being		Victimization		Misconduct		Psychological well-being		Subjective well-being		Victimization		Misconduct	
Unit level	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Safety	0.70***	0.07	0.58***	0.08	−0.91***	0.17	−1.10***	0.21	0.30***	0.07	0.23**	0.08	−0.22	0.18	−0.90***	0.22
Relationships with staff	0.10	0.05	0.20***	0.06	−0.56***	0.11	−0.75***	0.15	−0.11*	0.06	−0.06	0.06	0.00	0.13	−0.35*	0.16
Peer relationships	0.33***	0.08	0.33***	0.08	−0.84***	0.16	−1.26***	0.21	0.20*	0.08	0.06	0.08	−0.31	0.18	−0.93***	0.22
Meaningful activities	0.22***	0.05	0.31***	0.05	−0.43***	0.12	−0.57***	0.15	0.05	0.05	0.11*	0.05	−0.08	0.13	−0.27	0.16
Autonomy	0.30***	0.05	0.35***	0.05	−0.35***	0.11	−0.44**	0.14	0.03	0.05	0.04	0.05	0.06	0.12	−0.22	0.15
Overall institution rating	0.21***	0.04	0.23***	0.04	−0.36***	0.09	−0.46***	0.11	0.02	0.04	0.01	0.04	0.09	0.00	−0.17	0.12
Individual level																
Safety									0.40***	0.02	0.35***	0.02	−0.70***	0.05	−0.15**	0.05
Relationships with staff									0.22***	0.02	0.27***	0.02	−0.57***	0.05	−0.41***	0.05
Peer relationships									0.22***	0.02	0.29***	0.02	−0.57***	0.05	−0.34***	0.07
Meaningful activities									0.18***	0.02	0.19***	0.02	−0.34***	0.04	−0.30***	0.05
Autonomy									0.26***	0.02	0.30***	0.02	−0.41***	0.04	−0.22***	0.05
Overall institution rating									0.19***	0.01	0.22***	0.02	−0.37***	0.04	−0.46***	0.11

Note. All prison climate variables (unit and individual-level) are grand mean centered. All prison climate variables were entered in separate regression analyses (model 1: aggregated variables only, model 2: aggregated + individual variables).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

into account). The unit-level variance is similar to the median proportion of variance of 12% as found in research on organizational climate (James, 1982). It is, however, much lower than the estimated proportion of prison-level variance between 86 and 96%, suggested in an early US-based study on prison climate (Wright, 1985) and the mean proportion of 22% in the study of juvenile justice facilities in the US (Mulvey et al., 2010). It likely that these were overestimations due to methodological limitations (*e.g.,* low numbers within each institution), but it may also be due to differences in the questions asked and country of study. An important conclusion that can be drawn from this, is that survey-based studies on prison climate and related concepts should 1) establish the extent to which perceptions are shared across units or prisons and 2) control for the nested structure of the data, because there is dependency in the observations of members of units and prisons. This echoes concerns of scholars that have previously advocated for the use of multilevel methods in prisons research, particularly related to misconduct (Camp, Gaes, Langan, & Saylor, 2003; Wooldredge, Griffin, & Pratt, 2001), although this is still not as widely applied as it should be.

The findings on perceptual agreement raise interesting points for a theoretical discussion regarding the construct of prison climate. On the one hand, 4–10% of shared variance could be considered relatively small; on the other hand, these are still proportions of variance that cannot be ignored. It was also demonstrated that even the smallest proportion of variance at the unit level (namely, safety) still had a significant relationship with well-being and misconduct, independent of its individual counterpart. It would be premature to make strong claims about the meaningful existence of prison climate at a group level, based on this study and a dataset limited to one source (a survey among incarcerated individuals) and one country. Nevertheless, based on our findings, we can confidently conclude that there is significant variation in *individual* perceptions of prison climate.

These findings also shed new light on the importation *vs.* deprivation debate that has been well-rehearsed in the literature on adjustment to imprisonment (Crewe, 2016). On the one hand, we offer a methodological justification for conceptualizing an aspect of experienced deprivation at the unit or prison level by aggregating prison climate scores. On the other hand, we show that the effects at the aggregate level can be expected to be minimal, at least in the Netherlands, as the variance clustered at this level is small compared to the variance at the individual level. Nevertheless, we also show that there is variance in unit and prison means with respect to prison climate, and that there are meaningful relationships between aggregate-level prison climate variables and well-being and misconduct. Regardless of the direction of these effects, this underscores that subjective indicators of the quality of prison life can signal problems with health or behavior in prison, which can have far-reaching consequences. Further research should systematically examine the structural and individual determinants of prison climate.

This is the first study to examine the unique contribution of aggregated prison climate ratings on different dimensions separately, in combination with their individual-level counterparts. It can be concluded that prison climate has (correlational) effects, but that these should not be overstated. The purpose of this study was to give an indication whether there are effects – further research should explore the nature and mechanisms of these effects in greater detail. With the current design, it is not possible to establish direction for the reported effects, given that the variables were derived from the same survey. Theoretically, it is plausible that each of these relationships have bidirectional effects; for instance, an average score or tendency of psychological well-being on a unit may affect the prison climate experienced on the same unit, but prison climate may also affect the psychological well-being of members on a unit. Likewise, the perceived relationship with staff may contribute to misconduct on a unit, but misconduct may also affect the perceived relationship with staff.

It is important that similar studies in other countries replicate this study, because the results cannot be generalized to prisons elsewhere. It may be the case, for example, that there is less clustered variance in the Netherlands on the unit/prison level due to a relatively uniform prison experience compared to other countries. There are various aspects that may contribute to a unique Dutch prison climate. First, the Netherlands has a relatively small number of prisons (28 at the time of this study) and these prisons are all public prisons. This means that they have (more or less) the same provisions with respect to family visits, working hours, time in cell, educational facilities, and recreative activities. More diversity in prison climates may be expected in countries with more prisons and the involvement of private companies in prison management (see, for example, Crewe et al., 2014). Second, given that the majority of people incarcerated in the Netherlands serve short sentences, there may be less time for a shared climate to take shape than in countries where there is a larger population serving longer sentences, particularly if population turnover in units or prisons is low. Third, there is less variation in the Netherlands than in (some) other countries

in the security levels of regimes and associated restrictions, such as time out of cell. It may be expected that the prison climate in a super maximum facility in the US is different from a lower security facility, especially if there is no interaction among incarcerated individuals in the former. More generally, it would be helpful if future studies take into account the 'permeability' of a prison, by including measures of turnover of the prison population, but also a measure of contact time among prisoners in the same unit and across different units.

There are a few noteworthy limitations of the current study. First, in this study, we did not investigate the determinants of prison climate and the causal mechanism of perceptual agreement on the unit and prison level. The source of agreement may have been shared conditions (*structural*), as well as selective composition (*selection-attraction-attrition*). That is, it may be the case that respondents rated units and prisons differently due to variation in experienced conditions, regardless of their individual characteristics. Alternatively, it may be the case (and indeed, this is also likely) that certain units house people with certain characteristics and needs, which may also affect prison climate, such as sentence status and mental health. More research is needed to shine a light on the determinants and formation of prison climate.

Second, more thorough investigations are necessary to determine the (causal) relationship between prison climate and behavior and well-being of individuals in prison, including incarcerated persons and members of staff. As was stated earlier, a longitudinal design would be required to establish the possible direction of effects. Additionally, a more detailed design could shed a light on the exact contribution of prison climate to misconduct, including frequency and types of misconduct. A longitudinal design would also be needed to study the determinants of prison climate, including shared conditions and selective composition. Important potential 'shared conditions' to consider, would be staff culture and correctional orientations, prison design features, and daily program (see Molleman & Leeuw, 2012, for support that correctional orientation matters). A final limitation is that we solely relied on survey data from incarcerated individuals to develop a measure of prison climate. Arguably, a more comprehensive measure of prison climate would also include observations, interview data, and (survey) data from members of staff. This would also allow for an investigation of the extent to which staff perceptions of prison climate overlap with perceptions of incarcerated individuals.

A strong methodological recommendation that can be derived from our study is that at least the ICC(1) should be reported in perceptions research in prisons, to establish the degree of clustered variance on the unit and/or prison level, as this reflects both perceptual agreement and reliability of means to some extent. The other indices that were explored in this study can be used to give more detailed information and establish possible reasons for high or low ICC(1) values. Moreover, $r_{wg(j)}$ would be a suitable measure to distinguish between units or prisons in terms of levels of climate consensus and explore its determinants or effects. For example, high climate consensus may indicate high cohesiveness among individuals residing in a unit. Climate consensus and its effects have not been previously explored in relation to prison climate.

Another important implication of our study is that it appears that questionnaires that measure perceptions can reliably identify differences in prison climate at the unit and prison level. While our study is, at this point, unable to say how these differences may be explained, it could still help administrators identify units or prisons that score lower or higher than average on specific dimensions, especially if a prison climate survey is used in combination with other indicators, such as staff working conditions, incidents, and inspections (see Molleman, 2014, for guidance on performance measurement with dissimilar population characteristics). Our findings suggest that the Prison Climate Questionnaire is a useful and reliable tool to measure and therefore monitor important aspects of prison climate and differences among units and institutions. Further research should investigate the determinants of prison climate, and to what extent these can be explained by population composition (*e.g.*, sentence characteristics), objective indicators of deprivation (*e.g.,* security level, time-out-of-cell, and crowding), social interaction, management characteristics, and staff culture. This information is necessary in order to determine to what extent prison climate can be influenced by policy and management decisions. Eventually, if this type of research is replicated in other countries, it would be possible to identify best (and worst) practices internationally and even determine if national-level policies have an impact on prison climate.

Appendix A. Overview of aggregation indices with interpretation and calculation

Index	Interpretation	Calculation
$r_{wg(j)}$	Level of agreement of members within a unit, with 1 indicating perfect agreement and 0 indicating pure chance agreement. This score can be calculated for each unit and averaged across the sample. The literature indicates that scores of 0.70 and higher can be seen as sufficient agreement to justify aggregation. A different formula is used to calculate agreement for a 1-item measure.	$r_{wg(j)} = \frac{J\left[1-\left(\frac{\overline{s_{xj}^2}}{\sigma_{EU^2}}\right)\right]}{J\left[1-\left(\frac{\overline{s_{xj}^2}}{\sigma_{EU^2}}\right)\right]+\left(\frac{\overline{s_{xj}^2}}{\sigma_{EU^2}}\right)}$ $r_{wg(j)}$ = within-group interrater agreement for judges' mean scores based on J essentially parallel items $\overline{s_{xj}^2}$ = the mean of the observed variances on the J items σ_{EU^2} = the variance on Xj that would be expected if all judgements were due exclusively to random measurement error Source: James et al. (1984)
AD_M	Average absolute deviation from the unit mean, in the original scale metric. The higher this value, the more individual scores diverge from the unit mean. Note that average scores are affected by outliers, especially in small samples. A score of 0.90 on a scale from 1 to 5 indicates that, on average, members of a group score within 1 point on either side (lower or higher) of the group mean.	The average deviation from the mean for an item j: $AD_{M(j)} = \frac{\sum_{n=1}^{N} \lvert x_{jk} - \bar{x}_j \rvert}{N}$ N = the number of observations (number of deviations from an item) x_{jk} = the kth judge's score on item j $\bar{x}_j$ = the arithmetic mean of the judges' scores on item j The average deviation for the scale: $AD_{M(j)} = \frac{\sum_{j=1}^{J} AD_{M(j)}}{J}$ Source: Burke et al., 1999

ICC(1)	The ratio of between-group variance to total variance: The higher this ratio, the more variance may be attributed to between-group differences. In contrast, low scores indicate that variance can be mostly attributed to individual and idiosyncratic characteristics or experiences. In other words, to what extent are individuals in the same group more similar in their ratings compared to individuals in different groups?	$ICC(1) = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)}$ τ_{00} = between-group variance σ^2 = within-group variance Source: Bliese, 2000
ICC(2)	The reliability of the group means: The extent to which group means are able to capture consistently and with precision differences across groups, given the sample's average group size and variance attributable to group differences (ICC[1]). Scores can vary between 0 and 1, and scores of 0.70 and higher are normally considered to indicate good reliability. Low scores indicate that mean scores are unstable and cannot reliably differentiate between groups on a specific dimension, either due to high within-group variance (*i.e.,* low agreement) and/or small group sizes.	$ICC(2) = \frac{k(ICC(1))}{1 + (k-1)ICC(1)}$ k = average group or unit size Source: Bliese, 2000

Authors' note

The Life in Custody study was funded by the Dutch Custodial Institutions Agency (DJI) and Leiden University. The opinions, findings, and conclusions expressed in this article are those of the authors and do not necessarily reflect those of the DJI. The authors wish to thank the DJI for their support with the administration of the survey.

References

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

Beech, A. R., & Hamilton-Giachritsis, C. E. (2005). Relationship between therapeutic climate and treatment outcome in group-based sexual offender treatment programs. *Sexual Abuse: A Journal of Research and Treatment, 17*(2), 127–140.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein, & S. W. J. Kozlowski (Eds.). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bosma, A. Q., Van Ginneken, E. F. J. C., Palmen, H., Pasma, A., Beijersbergen, K. A., & Nieuwbeerta, P. (2020). A new instrument to measure prison climate: The psychometric quality of the prison climate questionnaire. *The Prison Journal*. https://doi.org/10.1177/0032885520916819.

Bosma, A. Q., Van Ginneken, E. F. J. C., Sentse, M., & Palmen, H. (2020). Examining prisoner misconduct: A multilevel test using personal characteristics, prison climate, and prison environment. *Crime & Delinquency, 66*(4), 451–484.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models in social and behavioral research: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods, 2*(1), 49–68.

Camp, S. D. (1999). Do inmate survey data reflect prison conditions? Using surveys to assess prison conditions of confinement. *The Prison Journal, 79*(2), 250–268.

Camp, S. D., Gaes, G. G., Langan, N. P., & Saylor, W. G. (2003). The influence of prisons on inmate misconduct: A multilevel investigation. *Justice Quarterly, 20*(3), 501–533.

Casey, S., Day, A., & Reynolds, J. (2016). The influence of incarceration length and protection status on perceptions of prison social climate. *Criminal Justice and Behavior, 43*(2), 285–296.

Clemmer, D. (1940). *The prison community.* New York, NY: Holt, Rinehart and Winston.

Cohen, A., Doveh, E., & Nahum-Shani, I. (2009). Testing agreement for multi-item scales with the indices rWG(J) and ADm(J). *Organizational Research Methods, 12*(1), 148–164.

Crewe, B. (2016). The sociology of imprisonment. In Y. Jewkes, J. Bennett, & B. Crewe (Eds.). *Handbook on prisons* (pp. 77–100). (2nd ed.). Abingdon: Routledge.

Crewe, B., Liebling, A., & Hulley, S. (2011). Staff culture, use of authority and prisoner quality of life in public and private sector prisons. *Australian & New Zealand Journal of Criminology, 44*(1), 94–115.

Crewe, B., Liebling, A., & Hulley, S. (2014). Heavy-light, absent-present: Rethinking the 'weight' of imprisonment. *The British Journal of Sociology, 65*(3), 387–410.

Day, A., Casey, S., Vess, J., & Hulsy, G. (2011). Assessing the social climate of Australian prisons. *Trends & Issues in Crime and Criminal Justice, no. 427*. Canberra: Australian Institute of Criminology.

Eidhin, M. N., Sheehy, N., O'Sullivan, M., & McLeavey, B. (2002). Perceptions of the environment, suicidal ideation and problem-solving deficits in an offender population. *Legal and Criminological Psychology, 7*(2), 187–201.

Ginsburg, L., & Oore, D. G. (2016). Patient safety climate strength: A concept that requires more attention. *BMJ Quality and Safety, 25*(9), 680–687.

Gonçalves, L. C., Endrass, J., Rossegger, A., & Dirkzwager, A. J. (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), 91–101.

Graf-Vlachy, L. (2017). IRA: Stata module to calculate rwg and related interrater agreement indices. *Statistical Software Components S458422*. Boston, MA: Boston College Department of Economics.

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.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*(2), 219–229.

James, L. R., Choi, C. C., Ko, C. H. E., McNeil, P. K., Minton, M. K., Wright, M. A., & Kim, K. I. (2008). Organizational and psychological climate: A review of theory and research. *European Journal of Work and Organizational Psychology, 17*(1), 5–32.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*(1), 85–98.

Joyce, W. F., & Slocum, J. W., Jr. (1984). Collective climate: Agreement as a basis for defining aggregate climates in organizations. *Academy of Management Journal, 27*(4), 721–742.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*(6), 959–976.

Kupchik, A., & Snyder, R. B. (2009). The impact of juvenile inmates' perceptions and facility characteristics on victimization in juvenile correctional facilities. *The Prison Journal, 89*(3), 265–285.

Lamers, S. M. A., Westerhof, G. J., Bohlmeijer, E. T., Ten Klooster, P. M., & Keyes, C. L. M. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHCSF). *Journal of Clinical Psychology, 67*(1), 99–110.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815–852.

Liebling, A. (2011). Moral performance, inhuman and degrading treatment and prison pain. *Punishment & Society, 13*(5), 530–550.

Liebling, A., & assisted by Arnold, H. (2004). *Prisons and their moral performance: A study of values, quality and prison life.* Oxford: Oxford University Press.

Liebling, A., Hulley, S., & Crewe, B. (2011). Conceptualising and measuring the quality of prison life. In D. Gadd, S. Karstedt, & S. Messner (Eds.). *The sage handbook of criminological research methods* (pp. 358–372). London: Sage Publishing.

Liebling, A., & Ludlow, A. (2016). Suicide, distress and the quality of prison life. In Y. Jewkes, B. Crewe, & J. Bennett (Eds.). *Handbook on prisons* (pp. 224–245). London: Routledge.

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

Molleman, T. (2014). Performance measurement in the Dutch prison system. *Methodological guidance for public sector performance assessment (doctoral dissertation)*. Utrecht: Utrecht University.

Molleman, T., & Leeuw, F. L. (2012). The influence of prison staff on inmate conditions: A multilevel approach to staff and inmate surveys. *European Journal on Criminal Policy and Research, 18*, 217–233.

Moos, R. H. (1975). *Evaluating correctional and community settings.* New York, NY: John Wiley.

Moos, R. H. (1997). *Evaluating treatment environments: The quality of psychiatric and substance abuse programs.* New Brunswick, NJ: Transaction Publishers.

Mulvey, E. P., Schubert, C. A., & Odgers, C. A. (2010). A method for measuring organizational functioning in juvenile justice facilities using resident ratings. *Criminal Justice and Behavior, 37*(11), 1255–1277.

Ross, M. W., Diamond, P. M., Liebling, A., & Saylor, W. G. (2008). Measurement of prison social climate: A comparison of an inmate measure in England and the USA. *Punishment & Society, 10*(4), 447–474.

Ross, M. W., Liebling, A., & Tait, S. (2011). The relationships of prison climate to health service in correctional environments: Inmate health care measurement, satisfaction and access in prisons. *The Howard Journal of Criminal Justice, 50*(3), 262–274.

Saylor, W. G. (1984). *Surveying prison environments.* Washington, DC: Federal Bureau of Prisons.

Schalast, N., Redies, M., Collins, M., Stacey, J., & Howells, K. (2008). EssenCES, a short questionnaire for assessing the social climate of forensic psychiatric wards. *Criminal Behaviour and Mental Health, 18*, 49–58.

Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annual Review of Psychology, 64*, 361–388.

Schubert, C. A., Mulvey, E. P., Loughran, T. A., & Losoya, S. H. (2012). Perceptions of institutional experience and community outcomes for serious adolescent offenders. *Criminal Justice and Behavior, 39*, 71–93.

Skar, M., Lokdam, N., Liebling, A., Muriqi, A., Haliti, D., Rushiti, F., & Modvig, J. (2019). Quality of prison life, violence and mental health in Dubrava prison. *International*

Journal of Prisoner Health, 15(3), 262–272.
StataCorp (2017). *Stata statistical software: Release 15.* College Station, TX: StataCorp LLC.
Sykes, G. (1958). *The society of captives.* Princeton, NJ: Princeton University Press.
Toch, H. (1977). *Living in prison: The ecology of survival.* New York, NY: Free Press.
Tonkin, M. (2016). A review of questionnaire measures for assessing the social climate in prisons and forensic psychiatric hospital settings. *International Journal of Offender Therapy and Comparative Criminology, 60*, 1376–1405.
Tonkin, M., Howells, K., Ferguson, E., Clark, A., Newberry, M., & Schalast, N. (2012). Lost in translation? Psychometric properties and construct validity of the English Essen climate evaluation Schema (EssenCES) social climate questionnaire. *Psychological Assessment, 24*(3), 573–580.
Van der Helm, P., Beunk, L., Stams, G. J., & Van der Laan, P. (2014). The relationship between detention length, living group climate, coping, and treatment motivation among juvenile delinquents in a youth correctional facility. *The Prison Journal, 94*(2), 260–275.
Van der Helm, P., Stams, G. J., & Van der Laan, P. (2011). Measuring group climate in prison. *The Prison Journal, 91*, 158–176.
Van der Helm, P., Stams, G. J., Van Genabeek, M., & Van der Laan, P. (2012). Group climate, personality, and self-reported aggression in incarcerated male youth. *The Journal of Forensic Psychiatry & Psychology, 23*(1), 23–39.
Van Ginneken, E. F. J. C., Palmen, H., Bosma, A. Q., Nieuwbeerta, P., & Berghuis, M. L. (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 Ginneken, E. F. J. C., Palmen, H., Bosma, A. Q., & Sentse, M. (2019). Bearing the weight of imprisonment: The relationship between prison climate and well-being. *Criminal Justice and Behavior, 46*(10), 1385–1404.
Woessner, G., & Schwedler, A. (2014). Correctional treatment of sexual and violent offenders: Therapeutic change, prison climate, and recidivism. *Criminal Justice and Behavior, 41*(7), 862–879.
Wooldredge, J., Griffin, T., & Pratt, T. (2001). Considering hierarchical models for research on inmate behavior: Predicting misconduct with multilevel data. *Justice Quarterly, 18*(1), 203–231.
Wright, K. N. (1985). Developing the prison environment inventory. *Journal of Research in Crime and Delinquency, 22*(3), 257–277.
Wright, K. N. (1991). A study of individual, environmental, and interactive effects in explaining adjustment to prison. *Justice Quarterly, 8*(2), 217–242.