



# Collective Efficacy in Australian and German Neighborhoods: Testing Cross-Cultural Measurement Equivalence and Structural Correlates in a Multi-level SEM Framework

Dominik Gerstner<sup>1</sup> · Rebecca Wickes<sup>2</sup> · Dietrich Oberwittler<sup>1</sup>

Accepted: 28 January 2019  
© The Author(s) 2019

## Abstract

In neighborhood research, the concept of collective efficacy has been particularly successful in capturing social cohesion and behavioral expectations among residents. Research has spread beyond the U.S. where it originated, and many studies from different countries have shown that collective efficacy is related to structural disadvantage in similar ways and affects outcomes as crime, education or health. However, methodological issues about measurement and modeling persist, and no study has yet investigated the cross-cultural measurement equivalence of this scale. We close this gap using two recent neighborhood surveys from Australia and Germany with large samples of respondents ( $N = \text{ca. } 12,800$ ) and neighborhoods ( $N = \text{ca. } 440$ ) in four cities. We employ multilevel structural equation modeling to test for measurement equivalence of collective efficacy across countries and to model its association with concentrated poverty, ethnic diversity, and residential stability. We find that the measurement of collective efficacy is metrically equivalent in both countries, modeling two latent factors on the respondent level—the two components informal social control and social cohesion/trust—but only one latent factor on the neighborhood level. Considering the relationship between the key correlates of collective efficacy, we find broad similarities but also substantial differences across contexts and compared to U.S. research, particularly concerning the role of ethnic diversity which has a stronger diminishing effect in Germany than in Australia. Possible explanations for these differences are discussed.

**Keywords** Collective efficacy · Neighborhood effects · Measurement equivalence · Cross-cultural research · Multi-level SEM

---

✉ Dominik Gerstner  
d.gerstner@mpicc.de

<sup>1</sup> Department of Criminology, Max Planck Institute for Foreign and International Criminal Law, Günterstalstrasse 73, 79100 Freiburg I.Br., Germany

<sup>2</sup> School of Social Sciences, Monash University, Clayton Campus, Melbourne, VIC 3800, Australia

## 1 Introduction

Collective efficacy (CE) is one of the most influential theories in the ecology of crime scholarship. It represents a significant advance on social disorganization theory by shifting the focus from neighborhood structural characteristics and the presence of strong social ties to the willingness of community residents to prosocially respond to local problems of crime and disorder. In criminology, collective efficacy represents a *task-specific* community level mechanism of informal social control. Sampson et al. (1997) define collective efficacy as the linkage of mutual trust and a willingness to intervene in task specific ways to reduce crime and disorder. Although the presence of strong networks may facilitate informal social control, the theory of collective efficacy suggests that the informal regulation of unwanted behavior does not require strong ties or associations amongst community members (Sampson 2006). Instead a working trust and weak ties are sufficient to generate regulatory action. Thus collective efficacy represents a dynamic process for activating social ties to achieve desired outcomes across the urban landscape.

The association between collective efficacy and a range of social problems is well documented in the literature. Communities with high levels of collective efficacy have significantly lower levels of violence (Morenoff et al. 2001) and burglary (Zhang et al. 2007). Residents living in collectively efficacious communities report higher self-rated health (Browning and Cagney 2002; Franzini et al. 2005). The presence of collective efficacy also appears to mediate low parental monitoring as it relates to the timing of first intercourse (Browning et al. 2005), protects against the negative effects of neighborhood deprivation on children's behavior at school entry (Odgers et al. 2009), and increases the likelihood that women will formally or informally report instances of domestic violence (Browning 2004).

The link between collective efficacy and a range of social problems is found in international studies in both Western (Maimon et al. 2010; Mazerolle et al. 2010; Sampson and Wikström 2008), and non-Western countries (Zhang et al. 2007). The magnitude of the relationship between collective efficacy and crime, in particular violence, is also similar across different community contexts. For example, the variation in violence attributable to collective efficacy is nearly the same in three cities in three countries, despite significant differences in overall levels of violence (Sampson and Wikström 2008; Mazerolle et al. 2010).

Less clear is whether or not collective efficacy is influenced by similar neighborhood characteristics. There is solid evidence that social disadvantage is negatively associated with collective efficacy in neighborhoods across several countries. Yet the strength of this relationship varies (Jiang et al. 2007; Mazerolle et al. 2010; Sampson and Wikström 2008; Wikström et al. 2012). Secondly, ethnic minority presence and diversity is viewed as an impediment to collective efficacy by many scholars (Neal 2017; Putnam 2007; Sampson et al. 1997). Even more than with social disadvantage, the evidence from other countries is mixed and suggests that ethnic and racial concentration may not as critical to the development of collective efficacy in some countries when compared to US settings (Brunton-Smith et al. 2014; Hipp and Wickes 2017; Twigg et al. 2010; van der Meer and Tolsma 2014). Given the legacies of racial residential segregation and poverty (Sampson and Wilson 1995) and the different patterns of immigration in the US when compared to other advanced western democracies (Sydes 2017), this is unsurprising, yet it speaks to the important role that context plays in shaping the relationships necessary for promoting collective efficacy across different settings in difficult cities or even nations.

Despite the significant up-take of collective efficacy across different cultural contexts, there has been little attempt to examine and explain the cross-cultural similarities or differences in collective efficacy, and no attempt to ascertain whether similarities across research sites are due to the equivalence of the measurement of collective efficacy or if they are a function of country specific differences. We argue that assessing whether collective efficacy generalizes across countries has been overshadowed by a focus on the association between neighborhood structural characteristics, collective efficacy and a particular social problem. Indeed at the time of writing only two papers have considered the structural conditions across countries that might differentially influence collective efficacy (Sampson and Wikström 2008; Mazerolle et al. 2010). Although Sampson (2006: 161) has argued that “nothing in the logic of collective efficacy is necessarily limited to specific cities, the United States or any country for that matter”, this remains an empirical question.

In this paper, we advance the collective efficacy literature by progressing three important goals. First, we consider the measurement equivalence of collective efficacy across neighborhoods in four cities in Germany and Australia. For the first time, we assess whether the items used to measure collective efficacy in one country have the same capacity to measure collective efficacy in another. As Davidov et al. (2014, 2018) contend, in order to meaningfully compare relationships between given constructs cross-culturally, especially those that are subjective in nature, we must first ascertain if we are advancing equivalent measurements of these concepts. Thus we test the extent to which the measurement of collective efficacy shows configural, metric and scalar invariance on individual and neighborhood levels across Germany and Australia.

Next, we consider whether the key neighborhood characteristics associated with collective efficacy in the US (Sampson et al. 1997) similarly or differentially predict collective efficacy in German and Australian cities. Societal-level conditions and policies may influence the relationship between ethnic concentration, residential stability, disadvantage and collective efficacy. Germany and Australia are highly developed Western countries with similar GDP and comparable economic and social structure, yet their immigration histories differ substantially from each other and from the United States. While Germany is a European country with long historical roots and only very recently became the second largest receiving country of international migration in the world (United Nations 2016), Australia is an Anglo-Saxon “settler society” which from the start of its modern existence has been a country of immigration, resulting in very different demographic structures, and reflected in different policies and attitudes towards migration and ethnic diversity. These differences could be consequential for collective efficacy in urban neighborhoods in both countries. More practically, these sites are useful for assessing cross-cultural equivalence as the ongoing large-scale empirical studies in both countries have been similarly designed and provide measures that are directly comparable.

Finally, drawing on recent advances in the collective efficacy literature, we address the question of whether CE should be seen as one unified construct or as consisting of two or more distinct components. Whereas Sampson and his colleagues saw social cohesion and trust and informal social control as closely enough linked processes to justify their union to a single scale (Raudenbush and Sampson 1999), others have argued that these two constructs are related but should be considered distinct processes (Rhineberger-Dunn and Carlson 2009; Twigg et al. 2010; Wickes et al. 2013). We follow recent work of Dunn et al. (2015) to determine if collective efficacy represents a latent construct consisting of two (or even more) components or as a unidimensional and *collective* property of neighborhoods across the neighborhoods in Germany and Australia.

We draw on survey data from over 12,800 residents living in 436 neighborhoods across four cities in Germany and Australia and on administrative data to develop a unique pooled dataset of identical items. We then employ multilevel confirmatory factor analyses to test the measurement equivalence of collective efficacy and multilevel structural equation models to compare the similarities and differences in the individual and neighborhood level correlates of collective efficacy.

We find that the measurement of collective efficacy and its two components informal social control and social cohesion and trust are metrically equivalent in both countries. Considering the relationship between the key correlates of collective efficacy, we find broad similarities but also substantial differences across contexts and compared to U.S. research particularly concerning the role of ethnic diversity.

In what follows we provide a review of the collective efficacy literature with a focus on the international studies of collective efficacy. Here we are interested in identifying not only the patterns between collective efficacy and a range of neighborhood problems across different countries, but the specific measurement of collective efficacy in various sites and the operationalization of the predictors of collective efficacy. We then provide an overview of the two research sites and the neighborhood studies and detail our methodological approach which focusses on the assessment of cross-cultural validity of collective efficacy. We conclude with a summary of our results and recommendations for future research.

## 2 Literature Review

Sampson et al. (1997), Sampson and Raudenbush (1999) define collective efficacy (CE) as the capacity of neighborhoods to realize common goals (i.e. to live in safe and orderly environments) through the informal regulation of unwanted, deviant or criminal behavior. The concept of collective efficacy represents two related processes: The first is the capacity for informal social control (ISC), or the “willingness of local residents to intervene for the common good” (Sampson et al. 1997: 919). In the Project for Human Development in Chicago Neighborhoods (PHDCN), this aspect of CE was measured by five survey questions on the perceived likelihood that neighbors would “do something about it” in response to a range of neighborhood problems. The second process identified in the CE literature is social cohesion and trust (SCT) or as Sampson et al. (1997: 919) state, the “mutual trust and solidarity among neighbors”.

The significant theoretical and empirical innovation of collective efficacy is the focus on group level processes within neighborhoods that prevent unwanted behavior, in particular crime. The measurement of collective efficacy therefore moves beyond traditional psychometric indicators to capture the extent to which neighborhoods could mobilize resources effectively and remedy problems facing the collective. In short, collective efficacy represents an emergent and collective process of informal social control, which is more than just the sum of its constituent parts. In the PHDCN SCT is measured through a number of items that assess the degree to which residents share the same values, work together to solve local problems and are perceived as trustworthy. Sampson et al. (1997) consider the two aspects as closely related and maintain that “it is the linkage of mutual trust and the willingness to intervene for the common good that defines the neighborhood context of collective efficacy”. The very high correlation between the two components of CE on the neighborhood level lends empirical support for this view. In the Chicago study, the correlation between SCT and ISC on the neighborhood level was  $r = .80$  (Sampson et al. 1997:

920) or even  $r = .88$  with corrections for measurement error (Sampson and Raudenbush 1999: 620). Further, the level of variability in each of the components that was attributable to the level of neighborhood was largely the same leading them to conclude that “the two measures were tapping aspects of the same latent construct”. Thus, according to its proponents, the two components of CE are seen as inextricably linked.

Yet, other scholars have argued that the two components of CE should be seen as theoretically distinct and shaped differentially by respondents’ experiences and perceptions and by social processes within local communities (Silver and Miller 2004; Twigg et al. 2010). This view has received some support in empirical studies. In a re-analysis of the PHDCN community survey data, Rhineberger-Dunn and Carlson (2009) applied confirmatory factor analysis (CFA) (excluding two items from the original scale) which supported two distinct dimensions which were correlated .66 at the respondent level and .84 at the neighborhood level. Oberwittler and Wikström (2009) using the Peterborough Community Study applied principal component analysis to the original PHDCN items on the respondent level and found support for a two-factor solution for SCT and ISC. Twigg et al. (2010) using data from the British Crime Survey likewise found two distinct dimensions in a principal component analyses and some support for differential effects on ISC and SCT in multilevel models. A study of CE in Australia found differences in the predictors of ISC and SCT, lending further weight that there may be important distinctions in the measurement of CE that warrant further investigation (Wickes et al. 2013).

The upshot from a growing number of studies is that while CE is highly reliable neighborhood predictor of crime, the dimensionality of CE varies across different contexts and data sets. This may be a function of analysis. For example, some scholars have used more complex models which better reflect the dual but highly connected nature of CE. Kochel (2012) modelled CE as a second-order factor consisting of SCT and ISC as underlying constructs. Uchida et al. (2013) employed a “bi-factor” approach in which SCT and ISC both are distinct latent constructs and part of a unified dimension.

Both studies as well as some of the above mentioned, however, treat CE as an individual respondent-level construct which runs counter its collective conception and ignores the hierarchical data structure. More recent advances in statistical modelling have integrated the previously separate approaches of SEM and multilevel modelling, making it possible to run confirmatory factor analyses and more complex models in a multilevel framework. Discussing its use in cross-cultural survey research, Ruelens et al. (2017) maintain that evidence for diverging factorial structures at different levels can “enrich both theoretical and empirical research”. Yet very few studies have applied multilevel CFA to investigate latent constructs on individual and collective levels simultaneously. In neighborhood effects research, Dunn et al. (2015) used survey data from the L.A.FANS study and found the best fit for a model with two latent constructs (ISC and SCT) at the respondent level but only one overarching construct (CE) at the neighborhood level. They concluded that CE and other theoretical constructs “can have very different meanings at each level of analysis and are perhaps most appropriately studied at the neighborhood level as one overarching construct and not divided into its two dimensions”. Likewise, Ward et al. (2017) found that physical and social incivilities are distinguishable at the individual level but not at the neighborhood level.

The above mentioned differences in factor structures may, however, also be a function of context in which collective efficacy is being studied and the degree to which the theory of collective efficacy generalizes across place and culture. This especially concerns the equivalence of measurements and latent constructs which has evolved as a relevant issue in recent cross-cultural research. From the broader literature examining measurement

equivalence, there is significant evidence that while some theoretical concepts generalize across national contexts, others do not. For example, Billiet (2013) shows that ignoring questions of measurement equivalence leads to biased results when comparing religious involvement across cultures with Muslim- and non-Muslim-majorities. Schaap and Scheepers (2014) found that constructs measuring trust in public institutions are invariant within most European countries, except those with diverging historical experiences. In contrast, Costa et al. (2016) found that concepts of parental and family efficacy are equivalent when comparing Italy and Portugal. Similarly, a well-being index constructed by Žemojtel-Piotrowska et al. (2017) and measured in 26 countries was equivalent across all contexts. Additional examples can be found in an overview by Davidov et al. (2014). Yet, until now, there has been no empirical test of the cross-cultural measurement invariance of CE. We argue that this is a significant gap in the literature. The question of whether collective efficacy generalizes across contexts is an important one, and one that is necessary in order to meaningfully compare associations and effects of CE in different countries.

### 3 Sample and Measures

The data used in our analyses come from two community surveys on crime, crime perceptions and neighborhood social processes carried out in two German and two Australian cities. For Germany, we use the first wave of a longitudinal postal survey of 3907 respondents in 140 randomly selected neighborhoods (with an oversampling of more disadvantaged areas) in Cologne and Essen, two large cities in the Western part of Germany (41% response rate). The average population size in these neighborhoods was 2880 persons living in areas of 0.5 square kilometers on average. The Australian data comes from the ‘Australian Community Capacity Study’ (ACCS) which took place in two major Australian cities, Brisbane and Melbourne. The study uses a longitudinal design (up to now: 4 waves in Brisbane, 1 wave in Melbourne) and use CATI survey data (46% response rate). The Australian data consists of 8996 respondents in 297 neighborhoods.<sup>1</sup> The total average population size in these neighborhoods was 6644 persons living in an area of about 14.5 square kilometers on average. In this paper we use the wave 3 (Brisbane) and the wave 1 (Melbourne) survey data collected in 2010.

Surveys in Germany and Australia used a set of near identical items designed to advance the knowledge about the conditions and consequences of neighborhood social processes in multi-ethnic, urbanized societies, alongside questions on sociodemographic background. Neighborhood structural data were drawn from administrative statistics.

#### 3.1 Measures of Collective Efficacy

To measure collective efficacy (CE) we use six of the original items of the Chicago Study by Sampson et al. (1997) and one additional item. In both studies, three identical items measured ‘social cohesion and trust’ (SCT) and five identical items measured ‘informal

<sup>1</sup> The term “neighborhood” refers to subsections of our research sites. In Australia state suburbs are used. They are similar to census tracts in the U.S. context, though some of them might be larger than census tracts because they are not determined by population. In the German cities Cologne and Essen, neighborhoods are administrative units but often reflect the historical development of urban landscapes. They also differ in population size but are often smaller than U.S. census tracts.

social control' (ISC). The ISC-items asked if people in respondents' neighborhood would do something if "a group of local children were skipping school and hanging around on a street corner" (SKIP), "some children were spray painting graffiti on a local building" (GRAFFITI). Respondents were also asked "if there was a fight in front of your house and someone was being beaten or threatened, how likely is it that people in your community would break it up" (FIGHT) and "if somebody was getting mugged, how likely is it that people in your community would help this person?" (MUGGED). The SCT-items asked if the statements: "People around here are willing to help their neighbors." (HELP), "People in this neighborhood can be trusted." (TRUST) and "People in this neighborhood do not share the same values." (VALUES) pertain to people in respondents' neighborhood. The slightly differing original wordings for Australia and Germany can be found in the "Appendix" (Table 10).

The German questionnaire uses a four-point Likert-type scale with the answer options "strongly agree" to "strongly disagree" for the SCT items and "very likely" to "very unlikely" for the ISC items, whereas the Australian questionnaire uses a five-point Likert-type scale with the same answer options but adding "neither agree nor disagree" and "neither likely nor unlikely", respectively, as middle response options. This issue of even or uneven numbers of answer categories has been debated for decades (Sturgis et al. 2014). Survey research has hinted at heterogeneous answering pattern of respondents faced with or without the option of a neutral middle category, including choosing the middle category as a "face-saving don't know" (Sturgis et al. 2014) and a tendency of some respondents to answer more negatively in its absence (Weijters et al. 2010). Adelson and Betsy McCoach (2010) found very little difference in the internal consistency of an instrument comparing a four- and a five-point Likert-type scale. Based on extant research, we assume only small effects of the diverging number of answer categories which, however, would rather make it harder to achieve measurement equivalence. For the comparative analyses, the items were standardized to a minimum value of 0 (lowest agreement) and a maximum of 1 (highest agreement) in each country—similar to a percent of maximum possible (POMP) score (e.g. Cohen et al. 1999). Although it is recommended to treat variables with less than five answer categories as ordinal (Rhemtulla et al. 2012) we treated the items as metric due to pragmatic reasons: Group comparisons to test measurement invariance in a multi-level framework are not possible with ordinal indicators in Mplus. All models were estimated with Mplus 7.4 (Muthén and Muthén 2012; Muthén and Muthén 1985-2012) using maximum likelihood estimation with robust standard errors. In the multilevel confirmatory factor analysis (MCFA) full information maximum likelihood (FIML) estimation is used to handle missing data. In the multilevel model with covariates (MSEM) missing data is excluded listwise.<sup>2</sup>

### 3.2 Socio-Demographic Predictors of Collective Efficacy

Collective efficacy varies as a function of individual and neighborhood level characteristics. In our analyses, we are particularly interested in potential differences in degree and kind between the two countries.

Individual Correlates of Collective Efficacy: The age of respondents was measured in years, and possible curvilinear effects are represented by a quadratic term. Respondents'

<sup>2</sup> This affects 0.02% of the cases in the Australian data and 0.03% in the German data.

sex (female), migration background (foreign born) and the presence of children in the household were represented by dummy variables. Length of residence in neighborhood was measured by six categories ranging from “up to one year” to “20 years and more”. Both surveys provide comparable categories and the scale was treated as metric.

In order to control for household socio-economic status, the levels of education were standardized across countries via the International Standard Classification of Education (ISCED-97) (OECD 1999) and recoded into five categories with medium–high education (3A) as reference category. Univariate statistics are shown in Table 6 (“Appendix”). In Germany, economic status was measured via three questions on the income situation: One item asked how respondents make ends meet considering their monthly income, a second item asked whether respondents could pay a large bill, and a third asked about receiving social benefit payments during the last 12 months. Exploratory factor analysis based on polychoric correlation revealed a unidimensional construct, and a factor score was saved. In the Australian survey, economic status was measured by asking about the approximate annual household income, providing eight answer categories. As the share of missing values was too high (24%) for listwise deletion we applied expectation maximization (EM) algorithm to impute missing data.<sup>3</sup> In both countries, the resulting measure of SES was grouped into quintiles with medium status as reference group.

To examine the neighborhood structural drivers of collective efficacy we focus on the three core variables emerging from Shaw and McKay’s (1942) social disorganization theory and replicated in Sampson and colleagues’ (1997) collective efficacy research. Structural data are from the 2006 census in Australia and from register-based data for 2013 in Germany provided by the city statistical offices. *Socioeconomic disadvantage* was measured as the percentage of unemployed people, *residential (in)stability* was represented by the percentage of people living longer than 5 years at the same address, and *racial/ethnic heterogeneity* was measured as the percentage of people with foreign citizenship. The levels and ranges of these indicators differ between Australian and German neighborhoods. For example, the neighborhood unemployment rate, which is based on comparable definitions in the two countries, in Germany ranges between 2 and 28% with a mean of 10.4%, whereas Australian neighborhoods range between 0 and 8% with a mean of 3.4%. While the national unemployment rates were around 5% in both countries, the two German cities recorded unemployment rates well above the national average (9.1% in Cologne, 12.1% in Essen), whereas Melbourne and Brisbane roughly matched the national average (5.5% and 5.9%, respectively).<sup>4</sup>

Comparable indicators of ethnic diversity for the two countries are more difficult to find. As data on ethnic origin is not available in Germany, and data on migration background is not available for neighborhoods in Essen, we use the percentage of residents with foreign citizenship as a proxy variable which is available in all four cities. In Cologne where data on both citizenship and migration background are available, the bivariate correlation between the two indicators is  $r = .91$ . There are higher concentrations of foreign citizens

<sup>3</sup> Referring to Mazerolle et al. (2010) we estimated the value for missing responses on income using sex, age, education and country of birth. Imputation was repeated 20 times and data from 4 randomly selected realisations were used in MPLUS analyses. Results were almost identical and only one is reported. Alternatively, models with missing values on income as an extra category were estimated. Results were almost identical and not reported in this paper.

<sup>4</sup> National and city-level unemployment rates are taken from the Australian Bureau of Statistics (Regional Statistics by ASGS2016) and from the German Federal Employment Agency (Regional Statistics).



in the German cities compared to the Australian cities (mean 17% and 9%, respectively), yet this difference is due to a more reluctant naturalization policy in Germany (Tolley and Vonk 2016). A higher proportion of migrants remain foreign citizens for longer periods in Germany than in Australia, while the share of migrants is in fact larger in Australia than in Germany. Thus, whereas the absolute levels of foreign citizens do not reflect the underlying extent of ethnic diversity in the two countries, the relative distribution within cities can be used as an indicator of ethnic heterogeneity and intra-urban segregation. The focus of our analyses will be on the effects of between-neighborhood structural differences on CE.

## 4 Analytical Approach

The multilevel SEM framework blends two separate statistical approaches: multilevel analyses and SEM. The combination of these approaches allows for the modelling of collective social phenomena at group levels (as schools or neighborhoods) while at the same time accounting for measurement error, ideally resulting in better estimates (Lüdtke et al. 2008; Marsh et al. 2012). Collective efficacy is a concept that is supposed to capture neighborhood-level social processes and hence calls for statistical approaches geared towards the analysis of hierarchical data structures (Hox 1998). Ignoring the clustered nature of data in a single-level SEM may lead to estimation problems if the proportion of group-level variance as indicated by the intraclass correlation coefficient (ICC) exceeds 0.05 (Julian 2001), as it is true for the data under investigation (see below, Sect. 5).

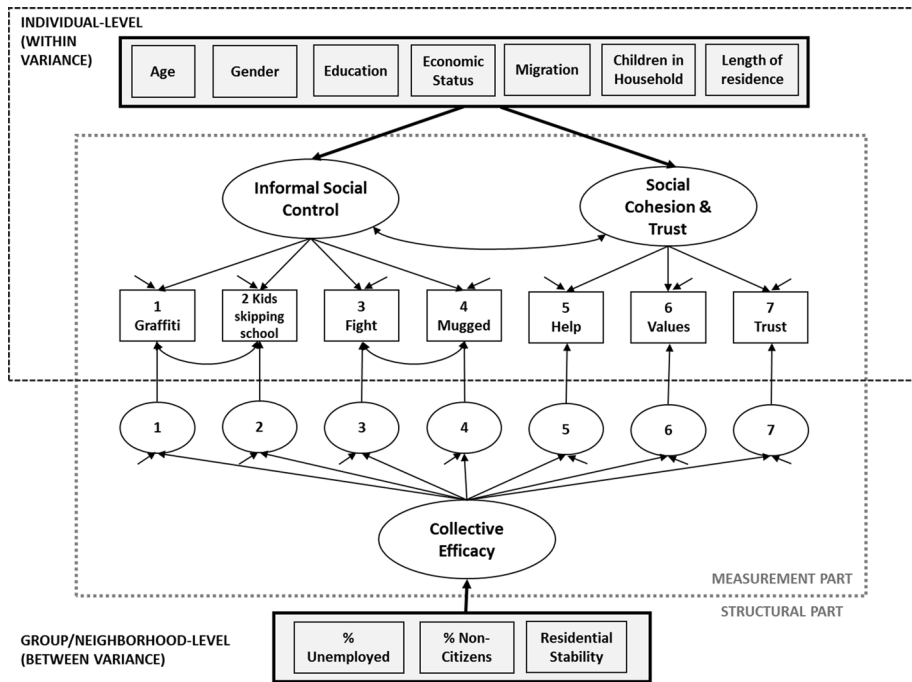
### 4.1 Multilevel Confirmatory Factor Analysis (MCFA)

In the first part of our analyses, we followed the approach of Dunn et al. (2015) analysis of collective efficacy. As our respondents (within-level  $W$ ) were clustered in neighborhoods (between-level  $B$ ), we apply multilevel confirmatory factor analysis (MCFA) to test our postulated models for each country as well as in purpose of testing measurement invariance across countries. As the study by Dunn et al. (2015) suggests, it is reasonable to hypothesize that two latent constructs, ISC and SCT, can be found on the within-level, but only one overarching construct, namely CE, on the between-level. This structure is shown in Fig. 1 (measurement part)<sup>5</sup> and the function to model the answers ( $y_{ij}$ ) of a person  $i$  in neighborhood  $j$  can be written like this:

$$y_{ij} = \mu_B + \Lambda_W \eta_W + \Lambda_B \eta_B + \varepsilon_B + \varepsilon_W$$

The vector  $\mu_B$  represents the overall means for the set of Items  $M$ .  $\Lambda_B$  is a matrix of factor loadings at the within-level and represents the relationships between the latent factors,  $\eta_B$ , and the manifest variables  $M$ .  $\Lambda_W$  describes the relationships between the group-level factor,  $\eta_W$ , and the random intercept indicators on the group-level.  $\varepsilon_B$  and  $\varepsilon_W$  are the residual errors at the within an between-level.

<sup>5</sup> We also conducted multilevel exploratory factor analysis (MEFA) to examine the fit of a model with only one latent factor on the within-level. The two factor model revealed a better fit. With multi-group MCFA a model with two within-level factors and two between-level factors was tested. The model fitted the data well but on the between-level the correlation between the latent factors was 0.995 in Germany and 0.982 in Australia—both factors are measuring the same. Results are not reported and can be requested from the authors.



**Fig. 1** Hypothesized multilevel factor structure of collective efficacy on individual and neighborhood level, including structural paths (simplified illustration)

Our first step examines the postulated factorial structure for each country separately. The purpose is to find a baseline model which fits the data best in relation to parsimony and meaningfulness (e.g. a residual covariance might be added to the model), whereas an exactly identical baseline model in both groups is not mandatory (Byrne 2012: 195). In a second step, we estimate the models for both countries simultaneously to test different levels of measurement invariance, following a commonly used step-up-approach (e.g. Brown 2015) in which different forms of invariance from less restrictive to more restrictive models are assessed. The central concern of this approach is to investigate the comparability of two measurement models across independent samples, i.e. from two different countries. The three main types of invariance are known as *configural*, *metric* and *scalar* invariance [see Davidov et al. (2014) and Jilke et al. (2015) for an overview]. *Configural* invariance implies an identical number of factors and identical pattern of factor loadings across groups. This indicates absence of construct bias, and allows that the constructs can be meaningfully discussed in both countries. *Metric* measurement invariance requires equal factor loadings between the manifest items and the latent variables across groups. Metric invariance implies that a unit change on the latent variable has the same meaning in all groups and makes it possible to compare e.g. relationships (unstandardized regression coefficients or covariances) between the latent constructs and other variables across groups. To compare raw scores of latent factors across groups, *scalar* invariance is required. Here the indicator intercepts are constrained to be equal across groups. This means that, given a latent mean of 0 in each group, respondents from different groups with the same value on the latent factor show the same means on the observed items.

**Table 1** Intraclass correlation coefficients (ICC)

Indicators	ICC <sup>a</sup>		
	Germany n = 3907	Australia n = 8996	Difference
People in this neighborhood can be trusted	0.091	0.050	0.041
People in this neighborhood do not share the same values	0.147	0.114	0.033
People around here are willing to help their neighbors	0.116	0.048	0.068
... children were skipping school and hanging around on a street corner	0.109	0.086	0.023
... youths were spraying graffiti at a building	0.167	0.083	0.084
... someone was being beaten or threatened	0.095	0.044	0.051
... somebody was getting mugged	0.099	0.045	0.054

<sup>a</sup>ICC = intraclass correlation coefficient, proportion of variance that is to neighborhood

Most of the studies on measurement invariance are using multigroup confirmatory factor analysis (MGCFAs) on a single-level (e.g. De Beuckelaer et al. 2007; Steinmetz et al. 2008; Vazsonyi and Belliston 2007; Zemojtel-Piotrowska et al. 2017). Due to the nested sampling design, our data has to be modeled in a multilevel latent variable model in multiple populations. Although some studies examine similar data structures (Mayer et al. 2014) there has to our best knowledge not been any published research focusing on neighborhood constructs. We follow the approach outlined in Muthén et al. (1997).

## 4.2 Multilevel Structural Equation Models (MSEM)

In the second part of our analyses we examine neighborhood-level correlates of collective efficacy using official (census) data while controlling for individual influences on the within-level. These covariates are either identical or very similar in both countries. Neighborhood-level structural predictors are transformed using square and square-root in order to reduce skewness. We do not test for measurement equivalence across populations in these models but rather explore the tendencies of individual and structural influences on collective efficacy and its components.

## 5 Results

Descriptive statistics show that the means of the items measuring collective efficacy are somewhat higher in Australia than in Germany whereas the standard deviation tends to be lower (see Table 6). This indicates a higher agreement with indicators of social cohesion and trust and informal social control in Australia. This is particularly true for the item asking if people in the neighborhood are willing to help their neighbors. A larger proportion of variance—measured by the intraclass correlation coefficient (ICC)—is between the neighborhoods in the German compared to the Australian cities (Table 1). The ICCs of the single items are between ca. 0.10 and 0.15 in Germany and between 0.05 and 0.10 in Australia, and in any case clearly above the threshold reported by Julian (2001). For comparison, Dunn et al. (2015) report ICCs in the range roughly between 0.08 and 0.25. Thus, respondents in the same neighborhood in Germany are

**Table 2** Unstandardized factor loadings in Germany and Australia, configural model (M1)

Indicator variables <sup>a</sup>	ISC (within)		SCT (within)		CE (between)	
	GER	AUS	GER	AUS	GER	AUS
	... someone was being beaten or threatened	1	1			0.654
... somebody was getting mugged	0.968	0.958			0.661	0.732
... youths were spraying graffiti at a building	0.911	0.996			0.753	0.897
... children were skipping school and hanging around on a street corner	0.834	0.986			1	1
People around here are willing to help their neighbors			1	1	0.743	1.184
People in this neighborhood do not share the same values (reverse coded)			0.458	0.537	0.699	0.642
People in this neighborhood can be trusted			0.919	1.02	0.659	0.662

<sup>a</sup>Correlation between ISC SCT on within level: GER 0.752/AUS 0.680, all factor loadings significant on  $p < 0.001$

more similar in their assessment of their shared environment than Australian respondents. This is most likely due to the smaller spatial units in Germany which tend to be more homogeneous (cf. Oberwittler and Wikström 2009). Despite differences in magnitude, the rank order patterns of ICCs of single items reveal similarities between the two countries. Tables 8 and 9 (see “Appendix”) report the bivariate correlations of items both on within- and between-level. In both countries, correlations are much higher on the between-level, reflecting the elimination of respondent-level variance (cf. Dunn et al. 2015).

Turning to the separate measurement models by country, modification indices pointed at shared error variances between GRAFFITI and SKIP as well as between FIGHT and MUGG for both countries. This seems reasonable as the first pair is related to delinquent but non-violent behavior of adolescents and the second pair is related to violent acts. Including these shared error variances improved the fit of both models. Based on the results of an exploratory factor analysis and unrestricted models the items with high and similar loadings were used as reference indicators. FIGHT is used as reference indicator for ISC and HELP serves as reference indicator for SCT on the individual level. GRAFFITI is used as reference indicator on the neighborhood level.

In a next step we estimated the model simultaneously in both countries (configural invariance, Model 1). Looking at the goodness of fit statistics (see Table 3), we see that the model fits the data well on the within- and between-levels [see Table 3, cutoff values CFI 0.95, RMSEA 0.08, SRMR-W and SRMR-B 0.08, in MCFA or MSEM the SRMR is split up to the within- and between-levels, there are uncertainties about the cutoff value available for SRMR-B, like other authors we rely on 0.08 (Dunn et al. 2015; Davidov et al. 2016)]. From this we can conclude that the assumption of configural equivalence holds, letting us affirm that we are measuring the same constructs in both countries. Table 2 shows the unstandardized factor loadings of Model 1. All factor loadings are significant and the coefficients reveal considerable similarities between both countries. The correlation of SCT and ISC on the within level was 0.75 in Germany and 0.68 in Australia, compared to 0.52 reported by Dunn et al. (2015).

**Table 3** Model fit invariance models

Statistic	M1 CONFIGURAL	M2 METRIC L2	M3 METRIC L1 L2	M4 SCALAR
$\chi^2$	291.953	374.966	376.731	1236.190
df	52	58	63	70
<i>p</i> value	0.000	0.000	0.000	0.000
Scaling correction factor	0.811	0.825	0.861	0.860
CFI	0.986	0.982	0.982	0.934
TLI	0.978	0.974	0.976	0.920
McDonald's NCI	0.991	0.988	0.988	0.956
RMSEA	0.027	0.029	0.028	0.051
SRMR W	0.020	0.020	0.020	0.020
SRMR B	0.033	0.040	0.040	0.158
$\Delta\text{-}\chi^2$ <sup>a</sup>		-76.65	-79.73	-824.63
df		6	11	18
<i>p</i> value		0.000	0.000	0.000
$\Delta\text{-CFI}$ <sup>b</sup>		0.004	0.004	0.052
$\Delta$ McDonald's NCI <sup>c</sup>		0.003	0.003	0.035
Models		2 vs. 1	3 vs. 1	4 vs. 1

Sample size: GER L1 (respondents),  $n=3907$ /L2 (neighborhoods)=140; AUS L1,  $n=8996$ /L2,  $n=297$

<sup>a</sup>Satorra–Bentler scaled Chi square difference test/<http://www.statmodel.com/chidiff.shtml>

<sup>b</sup> $\Delta\text{-CFI} \leq 0.01$  H0 (invariance) should not be rejected/Cheung and Rensvold (2002)

<sup>c</sup> $\Delta\text{-Mc} \leq 0.02$  H0 (invariance) should not be rejected/Cheung and Rensvold (2002)

In the next step we constrained the factor loadings on the between level to be equal across both countries (metric invariance on the between-level). Model 2 also shows a good fit (see Table 3). A comparison of recommended fit measures, e.g.  $\Delta\text{-CFI}$  ( $<0.01$ ) and  $\Delta\text{-NCI}$  ( $<0.02$ ), suggests that the overall fit of Model 2 does not deteriorate significantly compared to Model 1 (Cheung and Rensvold 2002). As both measures are based on Chi square, they are mainly driven by the within-level, but the restriction made in Model 2 affects the between-level. A recent simulation study (Hsu et al. 2015) supported this conclusion as SRMR-B was the only index which was sensitive to misspecifications on the between-level. According to this the SRMR-W does not change, the SRMR-B, however, rises by 0.007. We consider this difference as small enough to indicate invariance—although there is no cutoff value available (cf. Hsu et al. 2015).

In Model 3 the factor loadings on both levels are constrained to be equal. The model fit shows that our data also fits to this model, indicating that SCT and ISC on the individual level and CE on the neighborhood level are measuring the same construct in both countries, and thus meaningful comparisons of relationships between the latent constructs and other variables across groups are possible.  $\Delta\text{-CFI}$  and  $\Delta\text{-NCI}$  (compared to Model 1) are both smaller than the recommended cutoff values and SRMB-B remains the same than in Model 2.

The subsequent question whether latent means can be compared across countries is tested in Model 4 (scalar equivalence) in which factor loadings (within- and between-level)

and the item intercepts are constrained to be equal across countries. With item intercepts constrained equal the model fit deteriorates significantly. Although the overall fit measures are acceptable (CFI=0.934, RMSEA=0.051)  $\Delta$ -CFI and  $\Delta$ -NCI are both above the critical cutoff values. The most remarkable deterioration can be observed for the SRMR-B. The value changes from 0.040 in model 3 to 0.158 in model 4, way beyond the cutoff criterion of 0.08. This indicates that comparing the latent means of CE between Germany and Australia would lead to biased results. A potential remedy might be a partial scalar model (Byrne 2012: 254; cf. Zemojtel-Piotrowska 2017) or explaining the invariance with the aid of external variables (Davidov et al. 2012). De Beuckelaer et al. (2007) state that it is not uncommon and in line with previous research that scalar equivalence is not achieved in cross-cultural research and this might be a result of differences in response styles (e.g. acquiescence bias). In the current analysis, mean structures might be distorted by different interview modes (mail vs. CATI survey) and different numbers of answer categories (4 vs. 5 categories). But as our primary comparative interest is on associations and not mean estimates, metric invariance is a sufficient requisite to go ahead with the second concern of this paper which focusses on the structural (neighborhood) correlates of collective efficacy. In research about factorial invariance there is a discussion about unequal sample sizes biasing results. Considering the findings and recommendations reported in Yoon and Lai (2018), random subsamples were drawn from the Australian data to run our models with balanced sample sizes. With about 4000 respondents on the individual level and 140 clusters on the neighborhood level in both groups, the results are identical and our results do not seem to be biased by unequal sample sizes (results available on request).

## 5.1 The Correlates of Collective Efficacy

Having established the metric measurement invariance of collective efficacy, we turn to a comparison of the associations of this scale with individual (within-level) and neighborhood (between-level) attributes across countries. For this purpose we extend our measurement model (Model 3) to include predictors at both levels. As outlined above, we focus on those variables known from the extant literature as particularly important for explaining the within- and between-neighborhood variation of collective efficacy (see Fig. 1). As it is common practice in survey-based studies of neighborhood-level social processes, individual predictors primarily serve to control for the uneven socio-demographic composition of respondents across neighborhoods and thus help to adjust the neighborhood-level estimates of collective efficacy, whereas the neighborhood (between-) predictors represent the collective structural correlates which are of theoretical interest (Raudenbush and Sampson 1999). We allow the structural paths coefficients to vary across groups. The results of the multilevel structural equation model are shown in Table 4. For each country we present completely standardized coefficients (separately for each country) which indicate the direction, significance and relative size of effects (Brown 2015: 116ff).

Looking at the respondent-level predictors first, the main finding here is that they show only weak effects—all standardized coefficients ( $\beta$ ) are around or below .10—and contribute little to the explanation of ISC and SCT. In Germany, 8.6% of the within-level variance of SCT and 4.6% of the within-level variance of ISC are explained by individual-level predictors. In Australia, these shares of explained variance are 2.3% and 6.7%, respectively. The intercorrelation between both latent constructs is hardly affected by controlling for composition (see Tables 2, 4). This result supports the notion that residents' perceptions of

**Table 4** Completely standardized path coefficients and z-values (model with imputed income AUS), measurement part omitted

	Variable	Germany		Australia	
		Estimate	Est./S.E.	Estimate	Est./S.E.
Within level	Age	0.047	1.647	-0.104***	-4.689
ISC ON	Age (squared)	0.040	1.806	-0.106***	-7.063
	Female	0.033	1.585	0.082***	5.604
	Economic 1st Quintile (Ref 3rd)	-0.100***	-3.71	-0.041	-1.705
	Economic 2nd Quintile (Ref 3rd)	-0.029	-1.027	-0.014	-0.684
	Economic 4th Quintile (Ref 3rd)	0.044	1.772	0.051***	3.017
	Economic 5th Quintile (Ref 3rd)	0.071***	2.807	0.057***	3.057
	Education no/1A/other (Ref 3A)	-0.025	-1.046	-0.015	-0.927
	Education 2A (Ref 3A)	-0.048	-1.705	-0.005	-0.261
	Education 3B (Ref 3A)	-0.004	-0.154	0.001	0.069
	Education 5A, 6 (Ref 3A)	0.041	1.591	-0.025	-1.353
	Not born in Germany/Australia	0.012	0.506	-0.026	-1.664
	Children in household	0.107***	4.297	0.080***	4.487
	Length of residence in NBH	0.06*	2.184	0.040***	2.199
Within level	Age	0.184***	6.99	0.112***	6.043
SCT ON	Age (squared)	0.013	0.613	0.006	0.396
	Female	0.022	1.333	0.075***	5.335
	Economic 1st Quintile (Ref 3rd)	-0.147***	-5.892	-0.050*	-2.299
	Economic 2nd Quintile (Ref 3rd)	-0.002	-0.06	-0.005	-0.29
	Economic 4th Quintile (Ref 3rd)	0.065**	2.593	0.027 <sup>#</sup>	1.783
	Economic 5th Quintile (Ref 3rd)	0.097***	4.27	0.031 <sup>#</sup>	1.799
	Education no/1A/other (Ref 3A)	-0.030	-1.591	0.023	1.636
	Education 2A (Ref 3A)	-0.092**	-3.187	0.013	0.689
	Education 3B (Ref 3A)	-0.044*	-1.938	0.027	1.644
	Education 5A,6 (Ref 3A)	0.010	0.409	0.057***	3.05
	Not born in Germany/Australia	0.027	1.149	0.016	1.056
	Children in household	0.096***	4.772	0.072***	4.409
	Length of residence in NBH	0.016	0.66	0.062***	3.876
Within level	SCT WITH ISC	0.745***	34.886	0.684***	33.132
Between level	% Unemployed <sup>a</sup>	-0.503***	-7.266	-0.565***	-8.913
CE ON	% Foreign citizenship <sup>a</sup>	-0.371***	-4.857	-0.255***	-4.215
	% Resident > 5 years <sup>b</sup>	0.193***	3.534	-0.102	-1.552
R-square	ISC	0.048	4.976	0.066	7.914
	SCT	0.086	7.415	0.024	5.283
	CE	0.895	28.473	0.467	8.09
CHI2		1472.399			
CFI		0.936			
TLI		0.914			
RMSEA		0.028			
SRMR-W		0.02			
SRMR-B		0.052			

Sample size: GER L1 (respondents), n = 3775/L2 (neighborhoods) = 140; AUS L1, n = 8798/L2, n = 297

\*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ ; <sup>#</sup> $p < 0.1$ <sup>a</sup>Square root transformed<sup>b</sup>Squared

**Table 5** Changes in collective efficacy from good to bad neighborhoods based on unstandardized coefficients (compared to indicator item scaling 0 to 1)

	Changes in CE from neighborhoods with low (percentile 5) to high (percentile 95) values <sup>a</sup>		
	% Unemployed	% Foreign citizenship	% Resident > 5 years
Germany	-0.18 (-0.053)	-0.13 (-0.038)	0.06 (0.02)
Australia	-0.13 (-0.044)	-0.07 (-0.02)	n.s.

Percentiles computed by city

<sup>a</sup>Unstandardized path coefficients in parentheses; significant at  $p < 0.001$

shared neighborhood characteristics are very concurrent and largely independent of their age, gender, and socio-economic status (cf. Oberwittler and Wikström 2009).

In detail, a higher economic status and the presence of children in the household are associated with assessing both SCT and ISC more positively in both countries, whereas women and people living longer in the neighborhood have a more positive assessment in Australia only. Age has a positive effect on SCT in both countries and a negative effect on ISC in Australia only. It is interesting to note that migration status does not affect the assessment of ISC nor SCT in either country while the same variable has significant negative effects in both countries on the neighborhood level—which we now turn our attention to.

As explained above, we model a single latent variable—collective efficacy—encompassing both ISC and SCT on the neighborhood level. The neighborhood-level correlates are of substantial interest for the analyses of structural conditions which support or hinder collective efficacy. The largest standardized coefficients of neighborhood structural predictors are  $\beta = -.57$  in Australia and  $\beta = -.50$  in Germany hinting at much stronger effects than individual respondents' characteristics. We find that structural predictors are responsible for 47% of the variance in CE between neighborhoods in Australia and 90% of the variance in Germany. To put these findings in perspective, Sampson et al. (1997) reported that the same three structural variables accounted for 70.3% of the variance of CE between neighborhood clusters in Chicago, and Oberwittler and Wikström (2009) reported 74% explained variance of CE between Super-Output Areas in Peterborough/UK. Thus, CE in Australian cities seems considerably less strongly determined by socio-economic structure than in German cities or in Chicago, which is also reflected by the fact that only the unemployment rate has a strong negative influence on CE in Australia ( $\beta = -.57$ ), while ethnic diversity has only a moderate effect ( $\beta = -.26$ ) and residential stability has no effect at all, whereas in Germany both the unemployment rate and the share of foreign citizens have strong negative effects ( $\beta = -.50$  and  $-.37$ , respectively), and residential stability has a significant positive effect ( $\beta = .19$ ). In line with numerous studies on the effects of ethnic diversity on local social capital, we find in both countries that social disadvantage has a stronger eroding effect on CE than ethnic diversity (Ivarsflaten and Stromsnes 2013; Laurence 2011; Scheepers et al. 2013; van der Meer and Tolsma 2014; Wikström et al. 2012).

To illustrate these effects, Table 5 shows how CE varies along the range of neighborhood structural conditions in Australian and German cities. Contrasting highly advantaged neighborhoods with low unemployment rates at the 5th percentile with the most disadvantaged neighborhoods at the 95th percentile in each country—thus focusing on intra-urban differences irrespectively of country level differences, CE (measured on a scale from 0 to 1) is estimated to be 0.18 lower in German neighborhoods and 0.13 lower in Australian



**Table 6** Descriptive statistics of individual-level variables

	Germany			Australia		
	Obs.	Mean or %	SD	Obs.	Mean or %	SD
People in this neighborhood ... can be trusted (min 0/max 1)	3834	0.65	0.25	8939	0.70	0.21
People in this neighborhood do not share the same values	3746	0.43	0.27	8894	0.46	0.26
People around here are willing to help their neighbors	3817	0.60	0.24	8980	0.76	0.20
... children were skipping school and hanging around on a street corner	3832	0.43	0.29	8890	0.55	0.33
... youths were spraying graffiti at a building	3849	0.63	0.32	8947	0.77	0.27
... someone was being beaten or threatened	3854	0.59	0.28	8927	0.70	0.30
... somebody was getting mugged	3858	0.69	0.26	8920	0.74	0.25
Age	3911	53.0	16.05	8911	51.7	15.57
% Female	3918	56.3		8997	61.4	
% Not born in Germany/Australia	3851	18.3		8952	26.3	
% Children in household	3906	33.9		8955	36.2	
Length of residence in neighbourhood: up to 1 year	267	6.84		224	2.51	
1–2 years	271	6.95		321	3.60	
2–5 years	714	18.30		1559	17.46	
5–10 years	744	19.07		2073	23.22	
10–20 years	769	19.71		2454	27.49	
20 years and more	1137	29.14		2296	25.72	
Total	3902	100		8927	100	
% Education no/1A/other	300	7.8		270	3.0	
% Education 2A	996	26.0		1526	17.1	
% Education 3B	787	20.6		1660	18.6	
% Education 3A	642	16.8		2072	23.2	
% Education 5A,6	1101	28.8		3404	38.1	
Total	3826	100.00		8932	100.00	
1st SES quintile	809	20.75		3291	36.58	

Table 6 (continued)

	Germany			Australia		
	Obs.	Mean or %	SD	Obs.	Mean or %	SD
2nd SES quintile	812	20.83		1632	18.14	
3rd SES quintile	767	19.68		1531	17.02	
4th SES quintile	988	25.35		1003	11.15	
5th SES quintile	522	13.39		1540	17.12	
Total	3898	100		8997	100	

neighborhoods. With regard to ethnic diversity, again, comparing neighborhoods with very low percentages of foreign citizens at the 5th percentile to neighborhoods with very high percentages at the 95th percentile, CE varies by 0.13 in Germany but only 0.07 in Australia (Table 5). Residential stability increases CE by 0.06 in Germany but shows no significant effect in Australia.

Thus, structural conditions generally influence CE more strongly in German than in Australian cities. In both countries, the effects of social disadvantage are relatively stronger than the effects of ethnic diversity. However, the latter effect is more pronounced in German than in Australian cities, indicating that ethnic diversity is a stronger impediment to CE in urban neighborhoods in Germany than in Australia.

## 6 Discussion

Research on CE has proliferated considerably since its conception two decades ago and is used across different cultural contexts to investigate neighborhood influences on various outcomes as crime, health, or educational achievement. Yet, important methodological issues remain unexplored, particularly those concerning cross-cultural research. This paper contributes to the extant literature on CE by analyzing its dimensionality and cross-cultural equivalence, and by comparing its socio-demographic covariates, all in a multi-level framework, using survey data from two Australian and two German cities. To our knowledge for the first time, this study employed multilevel confirmatory factor analysis (MCFA) to test the cross-cultural equivalence of CE, showing the same structure (configural invariance) in Australia and Germany. As one recent study from the US (Dunn et al. 2015), we found two latent factors on the respondent level—Informal social control (ISC) and social cohesion/trust (SCT)—but only one latent factor on the neighborhood level—CE. In addition we found equal factor loadings across countries on the within- as well as on the between-level (metric invariance). On this basis, it is possible to compare relationships between the latent factors and other covariates across groups—e.g. the correlates of collective efficacy on the neighborhood level. This marks an important improvement with respect to cross-national research on collective efficacy. However, we could not confirm scalar equivalence in our models, and thus latent factor means cannot be compared across countries. Related to our empirical findings, there remain uncertainties of model fit in MCFA and MSEM. Common fit measures are less-than-ideal for multilevel CFA/SEM e.g. SRMR-B does not offer cut-off criteria. Furthermore our study has some limitations concerning measurement invariance: First, recent research about factorial invariance suggests that the selection of reference indicators can bias the results (Johnson et al. 2009). This problem can be solved by iterative procedures (ibid) or by using data-based specification searches (Yoon and Millsap 2007). Future research on cross-cultural equivalence of collective efficacy might focus on this issue.

Second, scalar non-equivalence is detected but not explained. Nonetheless, as our paper can be seen as a first effort to examine the cross-cultural measurement equivalence of collective efficacy this gives a connection point for future research.

In a second step multilevel structural equation modeling (MSEM) was employed to examine the major correlates of collective efficacy following classic social disorganization theory. On the within-level individual predictors showed broad similarities between the

two countries. Some individual predictors (e.g. education, age) showed differential effects on SCT and ISC, substantiating the concept of two separate latent dimensions on the within-level. On the between-level, a large proportion of variance was explained by neighborhood characteristics in both countries, yet the share of explained variance was twice as large in Germany than in Australia. Compared to previous research pursued in cities in the US and UK, CE seems to be more independent of neighborhood structural conditions in Australian cities (47% explained neighborhood-level variance) but even more dependent in German cities (90% explained variance). In the absence of systematic comparisons based on larger samples of cities, it is difficult to relate these differences to certain geographic or socio-economic characteristics of urban landscapes. However, we see a possible lead for an explanation in the vastly different ranges of socio-economic conditions in the Australian and German cities in our study: The highest neighborhood concentration of unemployment was just 8% in our sample of Australian but 28% in the sample of German neighborhoods. One sixth of all neighborhoods in the two German cities under investigation had unemployment rates of 14% or higher. This clearly indicates the existence of numerous areas of concentrated disadvantage where a sizeable share of residents are affected by poverty (Kronauer and Siebel 2013; Musterd et al. 2006). In contrast, though the ACCS survey sites capture areas experiencing significant disadvantage, it is possible that the segregation of poverty differs across the two countries. Understanding the segregation of poverty between the sites and how this influences neighborhood processes and crime will be an important next step in our research.

Secondly, even controlling for concentrated poverty which tends to coincide with the share of minority residents, ethnic diversity has a relatively stronger eroding effect on CE in German than in Australian neighborhoods. Again, we do not offer a ready explanation for this difference. Only few studies have investigated the effects of ethnic diversity on neighborhood social capital cross-nationally (Fieldhouse and Cutts 2010; Koopmans and Schaeffer 2015; Piekut and Valentine 2017; Uslaner 2011). Cross-national studies face the challenge of finding equivalent indicators. The use of foreign citizenship as indicator of ethnic diversity in the current study (which was prompted by data restrictions) is less than ideal. In general terms, foreign citizenship indicates a subgroup of migrants who have been in the host country for shorter periods, who face higher hurdles to naturalization or who are less interested in adopting the host citizenship; hence, a subgroup of migrants who may be assumed to be less well integrated in the host society than naturalized migrants. While the definition of native versus foreign citizenship is technically the same in both countries, its empirical validity varies with different national policies of naturalization. Germany historically had a more restrictive naturalization policy than Australia (Koopmans and Michalowski 2017), resulting in different proportions of foreign versus native citizens among the migrant population. To what extent different indicators of ethnic diversity may show different effects remains a question for future research.

We see the results of our analyses as an opportunity to develop possible routes of explanation. The role of societal, macro-level conditions in shaping local effects of ethnic diversity on social capital remains largely unexplored. Yet, the notion that long-term patterns of migration and national migration policies influence social processes of integration and interethnic relations down to the neighborhood level does not seem far-fetched. Australia and Germany represent very distinct migration regimes (Belot and Hatton 2012; Boucher and Gest 2015). Germany is a populous European country which until the end of the nineteenth century has seen waves of mass emigration before it slowly and reluctantly accepted its role as a major receiving country of international migration (Meyers 2004). Both the labor immigration flows from Turkey and South Europe during the 1960s and 1970s as

well as more recent and predominantly humanitarian migration flows of asylum seekers and civil war refugees from the Balkans, the Near/Middle East and Africa were characterized by non-selective policies and the prevalence of unskilled migrants. In combination with the lack of pro-active integration policies, this resulted in relatively wide gaps in socio-economic success between the native and large segments of the migrant populations including second- and third-generation descendants (Diehl 2016). As an Anglo-Saxon “settler society”, Australia’s existence and growth into a prosperous modern society has been dependent on migration from the start (Pincus and Hugo 2012). However, Australian migration policy has for a long time preferred immigrants from Anglo-Saxon and European countries and remains selective favoring skilled migrants as well as very restrictive towards humanitarian immigration. As one indicator of national differences, 40% of the foreign-born residents in Australia but only 24% of these in Germany are highly educated (Belot and Hatton 2012), and the national unemployment rate of foreign-born men by far exceeds that of native-born men in Germany (8.3% vs. 4.8%) but is even slightly lower in Australia (5.6% vs. 6.3%) (2014, OECD Stat).<sup>6</sup> Different national policy contexts may thus account a more successful integration of migrants into the Australian compared to the German society, which in turn may explain different magnitudes of effects of ethnic diversity on neighborhood social capital as one important dimension where social relations between native and migrant populations crystallize. Societal integration and intergroup relations are of course shaped by more than just economic performance and include legal, cultural, linguistic as well as attitudinal dimensions which are beyond the scope of this paper. Yet, the current results suggest that societies differ in the way in which socio-economic conditions including ethnic diversity impact collective social capital in urban neighborhoods. Future research should use cross-cultural approaches to investigate in more detail the embeddedness of local social processes in macro-level contexts.

**Acknowledgements** Open access funding provided by Max Planck Society.

**Funding** The authors acknowledge a grant by the ‘Go8 Australia – Germany Joint Research Co-operation Scheme’ (DAAD: No. 57140243). The ‘Australian Community Capacity Study’ was supported by the Australian Research Council (RO700002; DP1093960 and DP1094589). The German study received funding from the German Federal Ministry of Education and Research (FKZ 13N12980).

## Compliance with ethical standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

**OpenAccess** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix

See Tables 6, 7, 8, 9 and 10.

<sup>6</sup> <http://stats.oecd.org/Index.aspx?DatasetCode=STLABOUR>, retrieved 12 February 2018.

**Table 7** Descriptive statistics of neighborhood structural variables in Germany and Australia

	Mean	SD	Min	Max
% Unemployed				
GER (n = 140)	10.4	5.9	1.8	28.3
AUS (n = 296)	3.4	1.2	0.0	7.9
% Foreign citizenship				
GER	16.9	11.6	1.2	67.7
AUS	9.4	5.2	1.7	38.3
% Resident > 5 years				
GER	58.7	9.2	31.2	76.3
AUS	52.7	11.1	14.1	76.9

**Table 8** Correlations ML within level (lower triangular part Germany, upper part Australia)

	(1) VALUES	(2) TRUST	(3) HELP	(4) SKIP	(5) GRAF-FITI	(6) FIGHT	(7) MUGG
(1) VALUES	1	0.197	0.146	0.102	0.116	0.077	0.096
(2) TRUST	0.212	1	0.424	0.211	0.204	0.190	0.222
(3) HELP	0.262	0.607	1	0.199	0.226	0.206	0.256
(4) SKIP	0.194	0.314	0.328	1	0.427	0.177	0.193
(5) GRAF-FITI	0.206	0.311	0.298	0.480	1	0.249	0.292
(6) FIGHT	0.195	0.345	0.385	0.359	0.365	1	0.446
(7) MUGG	0.200	0.382	0.411	0.335	0.380	0.631	1

**Table 9** Correlations ML between level (lower triangular part Germany, upper part Australia)

	(1) VALUES	(2) TRUST	(3) HELP	(4) SKIP	(5) GRAF-FITI	(6) FIGHT	(7) MUGG
(1) VALUES	1	0.918	0.741	0.793	0.931	0.691	0.902
(2) TRUST	0.987	1	0.894	0.843	0.925	0.722	0.924
(3) HELP	0.962	0.983	1	0.945	0.879	0.763	0.904
(4) SKIP	0.955	0.950	0.957	1	0.931	0.714	0.895
(5) GRAF-FITI	0.954	0.933	0.904	0.958	1	0.751	0.940
(6) FIGHT	0.994	0.991	0.970	0.966	0.944	1	0.880
(7) MUGG	0.982	0.993	0.964	0.937	0.917	0.991	1

**Table 10** Items measuring collective efficacy, original wording

Informal social control	
Germany	Stellen Sie sich bitte die folgenden Ereignisse vor, die so oder so ähnlich in Ihrem Wohngebiet passieren könnten. Was glauben Sie: Würden die Leute aus Ihrem Wohngebiet etwas dagegen unternehmen? Leute in meinem Wohngebiet würden etwas dagegen unternehmen, wenn ...
SKIP	... Kinder aus der Nachbarschaft die Schule schwänzen und auf der Straße herumhängen
GRAFFITI	... Jugendliche Graffitis an ein Gebäude sprühen
FIGHT	Wenn ein Streit vor meinem Haus ausbricht und jemand geschlagen oder bedroht wird, würden Leute in meinem Wohngebiet einschreiten und dazwischen gehen
MUGGED	Wenn jemand überfallen und ausgeraubt wird, würden Leute in meinem Wohngebiet dieser Person helfen
Australia	For each of the following statements, please respond with very likely, likely, neither likely nor unlikely, unlikely or very unlikely
SKIP	If a group of community children were skipping school and hanging around on a street corner, how likely is it that people in your community would do something about it?
GRAFFITI	If some children were spray painting graffiti on a local building, how likely is it that people in your community would do something about it?
FIGHT	If there was a fight in front of your house and someone was being beaten or threatened, how likely is it that people in your community would break it up?
MUGGED	If somebody was getting mugged, how likely is it that people in your community would help that person?
Social cohesion and trust	
Germany	Welche der folgenden Aussagen trifft Ihrer Ansicht nach auf die Leute in Ihrem Wohngebiet zu?
HELP	Die Leute hier helfen sich gegenseitig
TRUST	Man kann den Leuten in der Nachbarschaft vertrauen
VALUES	Die Leute hier haben keine gemeinsamen Werte
Australia	For each of the following statements, please indicate whether you strongly agree, agree, neither agree nor disagree, disagree or strongly disagree
HELP	People in this community are willing to help their neighbours
TRUST	People in this community can be trusted
VALUES	People in this community do not share the same values

## References

- Adelson, J. L., & Betsy McCoach, D. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological Measurement, 70*(5), 796–807.
- Belot, M. V. K., & Hatton, T. J. (2012). Immigrant selection in the OECD. *The Scandinavian Journal of Economics, 114*(4), 1105–1128.
- Billiet, J. (2013). Quantitative methods with survey data in comparative research. In P. Kennett (Ed.), *A handbook of comparative social policy* (2nd ed., pp. 264–300). Cheltenham: Edward Elgar.
- Boucher, A., & Gest, J. (2015). Migration studies at a crossroads: A critique of immigration regime typologies. *Migration Studies, 3*(2), 182–198.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research. Methodology in the social sciences* (2nd ed.). New York: The Guilford Press.

- Browning, C. R. (2004). The span of collective efficacy: Extending social disorganization theory to partner violence. *Journal of Marriage and Family*, 64(4), 833–850.
- Browning, C. R., & Cagney, K. A. (2002). Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting. *Journal of Health and Social Behavior*, 43(4), 383–399.
- Browning, C. R., Leventhal, T., & Brooks-Gunn, J. (2005). Sexual initiation in early adolescence: The nexus of parental and community control. *American Sociological Review*, 70(5), 758–778.
- Brunton-Smith, I., Jackson, J., & Sutherland, A. (2014). Bridging structure and perception: On the neighbourhood ecology of beliefs and worries about violent crime. *British Journal of Criminology*, 54(4), 503–526.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus. Basic concepts, applications, and programming (multivariate applications series)*. New York: Routledge.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
- Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research*, 34(3), 315–346.
- Costa, M., Faria, L., Alessandri, G., & Caprara, G. V. (2016). Measuring parental and family efficacy beliefs of adolescents' parents: Cross-cultural comparisons in Italy and Portugal. *International Journal of Psychology*, 51(6), 421–429.
- Davidov, E., Dülmer, H., Cieciuch, J., Kuntz, A., Seddig, D., & Schmidt, P. (2016). Explaining measurement nonequivalence using multilevel structural equation modeling. *Sociological Methods and Research*. <https://doi.org/10.1177/0049124116672678>.
- Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, 43(4), 558–575.
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40(1), 55–75.
- Davidov, E., Muthén, B., & Schmidt, P. (2018). Measurement invariance in cross-national studies: challenging traditional approaches and evaluating new ones. *Sociological Methods and Research*, 47(4), 631–636.
- De Beuckelaer, A., Lievens, F., & Swinnen, G. (2007). Measurement equivalence in the conduct of a global organizational survey across countries in six cultural regions. *Journal of Occupational and Organizational Psychology*, 80(4), 575–600.
- Diehl, C. (2016). Migration und Integration in der Bevölkerungssoziologie. In Y. Niephaus, M. Kreyenfeld, & R. Sackmann (Eds.), *Handbuch Bevölkerungssoziologie* (pp. 462–479). Wiesbaden: Springer.
- Dunn, E., Masyn, K., Johnston, W., & Subramanian, S. V. (2015). Modeling contextual effects using individual-level data and without aggregation: an illustration of multilevel factor analysis (MLFA) with collective efficacy. *Population Health Metrics*, 13(1), 1–11.
- Fieldhouse, E., & Cutts, D. (2010). Does diversity damage social capital? A comparative study of neighbourhood diversity and social capital in the US and Britain. *Canadian Journal of Political Science/Revue canadienne de science politique*, 43(2), 289–318.
- Franzini, L., Caughy, M., Spears, W., & Eugenia Fernandez Esquer, M. (2005). Neighborhood economic conditions, social processes, and self-rated health in low-income neighborhoods in Texas: A multilevel latent variables model. *Social Science and Medicine*, 61(6), 1135–1150.
- Hipp, J. R., & Wickes, R. (2017). Violence in urban neighborhoods: A longitudinal study of collective efficacy and violent crime. *Journal of Quantitative Criminology*, 33(4), 783–808.
- Hox, J. (1998). Multilevel modeling: When and why? In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 147–154). Berlin: Springer.
- Hsu, H.-Y., Kwok, O.-M., Lin, J. H., & Acosta, S. (2015). Detecting misspecified multilevel structural equation models with common fit indices: A Monte Carlo study. *Multivariate Behavioral Research*, 50(2), 197–215.
- Ivarsflaten, E., & Stromsnes, K. (2013). Inequality, diversity and social trust in norwegian communities. *Journal of Elections, Public Opinion and Parties*, 23(3), 322–342.
- Jiang, S., Lambert, E., & Wang, J. (2007). Correlates of formal and informal social/crime control in China: An exploratory study. *Journal of Criminal Justice*, 35(3), 261–271.
- Jilke, S., Meuleman, B., & Van de Walle, S. (2015). We need to compare, but how? Measurement equivalence in comparative public administration. *Public Administration Review*, 75(1), 36–48.
- Johnson, E. C., Meade, A. W., & DuVernet, A. M. (2009). The role of referent indicators in tests of measurement invariance. *Structural Equation Modeling*, 16(4), 642–657.



- Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, 8(3), 325–352.
- Kochel, T. R. (2012). Can police legitimacy promote collective efficacy? *Justice Quarterly*, 29(3), 384–419.
- Koopmans, R., & Michalowski, I. (2017). Why do states extend rights to immigrants? institutional settings and historical legacies across 44 countries worldwide. *Comparative Political Studies*, 50(1), 41–74.
- Koopmans, R., & Schaeffer, M. (2015). Statistical and perceived diversity and their impacts on neighborhood social cohesion in Germany, France and the Netherlands. *Social Indicators Research*, 125(3), 853–885.
- Kronauer, M., & Siebel, W. (Eds.). (2013). *Polarisierte Städte: Soziale Ungleichheit als Herausforderung für die Stadtpolitik*. Frankfurt: Campus.
- Laurence, J. (2011). The effect of ethnic diversity and community disadvantage on social cohesion: A multi-level analysis of social capital and interethnic relations in UK communities. *European Sociological Review*, 27(1), 70–89.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multi-level latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13(3), 203–229.
- Maimon, D., Browning, C. R., & Brooks-Gunn, J. (2010). Collective efficacy, family attachment, and urban adolescent suicide attempts. *Journal of Health and Social Behavior*, 51(3), 307–324.
- Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., et al. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, 47(2), 106–124.
- Mayer, A., Nagengast, B., Fletcher, J., & Steyer, R. (2014). Analyzing average and conditional effects with multigroup multilevel structural equation models. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2014.00304>.
- Mazerolle, L., Wickes, R., & McBroom, J. (2010). Community variations in violence: The role of social ties and collective efficacy in comparative context. *Journal of Research in Crime and Delinquency*, 47(1), 3–30.
- Meyers, E. (2004). *International immigration policy: A theoretical and comparative analysis*. New York: Palgrave Macmillan.
- Morenoff, J. D., Sampson, R. J., & Raudenbush, W. S. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, 39(3), 517–558.
- Musterd, S., Murie, A., & Kesteloot, C. (Eds.). (2006). *Neighbourhoods of poverty, urban social exclusion and integration in Europe*. London: Macmillan.
- Muthén, M. (1985–2012). Mplus Los Angeles, CA.
- Muthén, B., Khoo, S.-T., & Gustafsson, J.-E. (1997). *Multilevel latent variable modeling in multiple populations*. Unpublished manuscript, retrieved from [http://www.academia.edu/download/41460205/Article\\_074.pdf](http://www.academia.edu/download/41460205/Article_074.pdf).
- Muthén, B., & Muthén, L. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Neal, Z. P. (2017). Taking stock of the diversity and sense of community debate. *American Journal of Community Psychology*, 59(3–4), 255–260.
- Oberwittler, D., & Wikström, P.-O. (2009). Why small is better: advancing the study of the role of behavioral contexts in crime causation. In D. Weisburd, W. Bernasco, & G. N. Bruinsma (Eds.), *Putting crime in its place* (pp. 35–59). New York: Springer.
- Ogders, C. L., Moffitt, T. E., Tach, L. M., Sampson, R. J., Taylor, A., Matthews, C. L., et al. (2009). The protective effects of neighborhood collective efficacy on British children growing up in deprivation: A developmental analysis. *Developmental Psychology*, 45(4), 942–957.
- OECD. (1999). *Classifying educational programmes: Manual for ISCED-97 implementation in OECD countries*. Paris: OECD Paris.
- Piekat, A., & Valentine, G. (2017). Perceived diversity and acceptance of minority ethnic groups in two urban contexts. *European Sociological Review*, 32(3), 339–354.
- Pincus, J., & Hugo, G. (2012). *A greater Australia: Population, policies and governance*. Melbourne: Committee for Economic Development of Australia.
- Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century. *Scandinavian Political Studies*, 30(2), 137–174.
- Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology*, 29(1), 1–41.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under sub-optimal conditions. *Psychological Methods*, 17(3), 354–373.

- Rhineberger-Dunn, G. M., & Carlson, S. M. (2009). Confirmatory factor analyses of collective efficacy and police satisfaction. *Journal of Crime and Justice*, 32(1), 125–154.
- Ruelens, A., Meuleman, B., & Nicaise, I. (2017). Examining measurement isomorphism of multilevel constructs: The case of political trust. *Social Indicators Research*. <https://doi.org/10.1007/s11205-017-1799-6>.
- Sampson, R. J. (2006). How does community context matter? Social mechanisms and the explanation of crime rates. In P.-O. H. Wikström & R. J. Sampson (Eds.), *The explanation of crime: Context, mechanisms and development* (pp. 31–60). Cambridge: Cambridge University Press.
- Sampson, R. J., & Raudenbush, S. W. (1999). Systematic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology*, 105(3), 603–651.
- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328), 918–924.
- Sampson, R. J., & Wikström, P.-O. (2008). The social order of violence in Chicago and Stockholm neighborhoods: A comparative inquiry. In S. N. Kalyvas, I. Shapiro, & T. Masoud (Eds.), *Order, conflict, and violence* (pp. 97–119). New York: Cambridge University Press.
- Sampson, R. J., & Wilson, W. J. (1995). Toward a theory of race, crime, and urban inequality. In H. John & R. D. Peterson (Eds.), *Crime and inequality* (pp. 37–56). Stanford, CA: Stanford University Press.
- Schaap, D., & Scheepers, P. (2014). Comparing citizens' trust in the police across European countries: An assessment of cross-country measurement equivalence. *International Criminal Justice Review*, 24(1), 82–98.
- Scheepers, P., Schmeets, H., & Pelzer, B. (2013). Hunkering down as disruption of community cohesion: Municipal-, neighbourhood- and individual-level effects. *Procedia - Social and Behavioral Sciences*, 72, 91–106.
- Shaw, C. R., & McKay, H. D. (1942). *Juvenile delinquency and urban areas: A study of rates of delinquents in relation to differential characteristics of local communities in American cities*. Chicago: University of Chicago Press.
- Silver, E., & Miller, L. L. (2004). Sources of informal social control in Chicago neighborhoods. *Criminology*, 42(3), 551–584.
- Steinmetz, H., Schmidt, P., Tina-Booh, A., Wiczorek, S., & Schwartz, S. H. (2008). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality and Quantity*, 43(4), 599–616.
- Sturgis, P., Roberts, C., & Smith, P. (2014). Middle alternatives revisited: How the neither/nor response acts as a way of saying 'I don't know'? *Sociological Methods and Research*, 43(1), 15–38.
- Sydes, M. (2017). Revitalized or disorganized? Unpacking the immigration-crime link in a multiethnic setting. *Journal of Research in Crime and Delinquency*, 54(5), 680–714.
- Tolley, E., & Vonk, M. (2016). *Multiculturalism policy index: Immigrant minority policies (revised edition)*. Kingston: School of Policy Studies, Queen's University. Retrieved from [http://queensu.ca/mcp/sites/webpublish.queensu.ca.mcpwww/files/files/immigrantminorities/evidence/ImmigrantMinoritiesEvidence2016web\(1\).pdf](http://queensu.ca/mcp/sites/webpublish.queensu.ca.mcpwww/files/files/immigrantminorities/evidence/ImmigrantMinoritiesEvidence2016web(1).pdf).
- Twigg, L., Taylor, J., & Mohan, J. (2010). Diversity or disadvantage? Putnam, Goodhart, ethnic heterogeneity, and collective efficacy. *Environment and Planning A*, 42(6), 1421–1438.
- Uchida, C. D., Swatt, M. L., Solomon, S. E., & Varano, S. (2013). *Neighborhoods and crime: Collective efficacy and social cohesion in Miami-Dade County (final report submitted to the National Institute of Justice)*. Silverspring: Justice & Security Strategies.
- United Nations, Department of Economic and Social Affairs. (2016). International migration report 2015. Highlights, New York.
- Uslaner, E. (2011). Trust, diversity, and segregation in the United States and the United Kingdom. *Comparative Sociology*, 10, 221–247.
- van der Meer, T., & Tolsma, J. (2014). Ethnic diversity and its effects on social cohesion. *Annual Review of Sociology*, 40(1), 459–478.
- Vazsonyi, A. T., & Belliston, L. M. (2007). The family → low self-control → deviance. A cross-cultural and cross-national test of self-control theory. *Criminal Justice and Behavior*, 34(4), 505–530.
- Ward, J. T., Link, N. W., & Taylor, R. B. (2017). New windows into a broken construct: A multilevel factor analysis and DIF assessment of perceived incivilities. *Journal of Criminal Justice*, 51, 74–88.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236–247.
- Wickes, R., Hipp, J. R., Sargeant, E., & Homel, R. (2013). Collective efficacy as a task specific process: examining the relationship between social ties, neighborhood cohesion and the capacity to respond

- to violence, delinquency and civic problems. *American Journal of Community Psychology*, 52(1–2), 115–127.
- Wikström, P.-O. H., Oberwittler, D., Treiber, K., & Hardie, B. (2012). *Breaking rules. The social and situational dynamics of young people's urban crime (Clarendon studies in criminology)*. Oxford: Oxford University Press.
- Yoon, M., & Lai, M. H. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling*, 25(2), 201–213.
- Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling*, 14(3), 435–463.
- Žemojtel-Piotrowska, M., Piotrowski, J. P., Ciecuch, J., Adams, B. G., Osin, E. N., Ardi, R., et al. (2017). Measurement invariance of personal well-being index (PWI-8) across 26 countries. *Journal of Happiness Studies*, 18(6), 1679–1711.
- Zhang, L., Messner, S. F., & Liu, J. (2007). A multilevel analysis of the risk of household burglary in the city of Tianjin, China. *British Journal of Criminology*, 47(6), 918–937.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.