

From Dependent to Independent Variable: A Critical Assessment of Operationalisations of ‘Welfare Stateness’ as Macro-Level Indicators in Multilevel Analyses

Katharina Kunißen¹ 

Accepted: 17 May 2018 / Published online: 23 May 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract This paper explores different ways to operationalise properties of the welfare state as explanatory variable in multilevel frameworks. Based on the observation that many common applications of welfare state measurements as independent macro-level variables lack standardised proceeding, differences between commonly used approaches (single indicators, regime typologies, and composite indices) are examined concerning their consequences for empirical results and their comparability. Each approach is first discussed regarding conceptual premises and practical applications in the literature. In a second step, differences are demonstrated empirically by performing several multilevel analyses using welfare attitudes as an exemplary dependent variable. The comparison shows that even slight differences in the operationalisation have an impact on the results and their explanatory contribution. Based on this, the paper offers possible points of departure for the development of more fitting operationalisations for the specific use as explanatory instruments. This includes criteria a measurement should meet and a stronger focus on capturing the relevant causal mechanisms assumed to be at work. The aim of this paper is thus twofold. On the one hand, it provides an overview of existing strategies and raises awareness to critical issues. On the other hand, it gives some initial pointers for a conceptual debate about how to turn characteristics of the welfare state into macro-level indicators, which can serve as independent variable.

Keywords Welfare state · Operationalisation · Social policy · Indicator · Multilevel

✉ Katharina Kunißen
katharina.kunissen@uni-mainz.de

¹ Institute of Sociology, Johannes Gutenberg-University Mainz, Jakob-Welder-Weg 12, 55128 Mainz, Germany

1 Introduction

The welfare state plays a central role in explaining a variety of social phenomena at the micro-level. Especially cross-cultural studies often include the welfare state as an explanatory variable because its function goes beyond merely representing a conglomerate of social rights: The arrangement of social policies actively shapes social stratification and its outcomes (e.g. Esping-Andersen 1990). Among such outcomes on the level of individuals are attitude formation (e.g. Jordan 2013; Eger and Breznau 2017), political participation (e.g. Schneider and Makszin 2014), political trust (e.g. Mattila and Rapeli 2018), well-being (e.g. Cruz-Martínez 2017; Schuck and Steiber 2017), the consequences of risk exposure (e.g. Angel and Heitzmann 2015) and much more. Many empirical studies dealing with these outcomes implement a multilevel design¹ in which properties of the welfare state serve as independent variables at the country level. However, there is no agreed-upon way to operationalise ‘welfare stateness’ as an indicator in such studies. Thus, scholars usually borrow instruments from literature, which examines welfare policies as a dependent variable. These instruments include welfare state typologies, single indicators and composite measures.

Since the demand for treating the welfare state as an independent variable is high, it surprises that the literature hitherto lacks comprehensive discussions of the extent to which existing measures *can* actually serve as suitable independent variables and the problems which may be associated with different operationalisations. In order to address these topics, it is necessary to have a short look at the debate surrounding the general measurement of different social systems.

Ever since more complex ways of empirically capturing social policy arrangements were introduced (Esping-Andersen 1990 certainly played an important role in this), there has been a lively and critical debate about how to appropriately measure differences between welfare states. This debate has addressed conceptual and empirical issues such as missing or underrepresented policy area like family policies (e.g. Orloff 1993), the addition of new countries or types to existing measures (e.g. Ferrera 1996), misspecifications in the literature (e.g. Scruggs and Allan 2006), differences in the conceptual and operational treatment of indicators (e.g. Wenzelburger et al. 2013) and much more.

More recently, the so-called *dependent variable problem* receives growing attention (e.g. Clasen and Siegel 2007). This methodological debate emerged as a by-product of a discussion about welfare state change and retrenchment (e.g. Pierson 1996). A key problem identified in this debate was the lack of a common understanding of what the object of research—the dependent variable—entails and how it should be measured (Green-Pedersen 2004). Until today, there is an ongoing discussion based on the repeated observation that different conceptual and operational strategies lead to different results (e.g. Kühner 2007; Bolukbasi and Öktem 2018).

In light of this existing debate on how to best conceptualise and measure features of the welfare state, one might wonder why we need an additional independent variable perspective instead of simply using the indicators proposed by literature addressing the dependent variable problem. Five arguments speak in favour of such an endeavour. First, there is no thorough account of how different conceptualisations affect explanatory power and informative scope when used as *independent* variables. Only recently, scholars start

¹ E.g. multilevel regression analysis or fixed effect models.

to voice concerns because existing measurements are treated as interchangeable options for the operationalisation of welfare policies as dependent as well as independent variables (Bolukbasi and Öktem 2018). Second, the existing methodological discussions mainly address the *macro-level*. In how far the proposed measures can be embedded in *macro-micro*-analyses remains unclear. Third, the exchange of *feasible recommendations* between general literature on the welfare state and research, which examines its outcomes, is highly underdeveloped. Systematic comparisons of varying strategies are rare and focus only on consequences of different ways to operationalise within one of the approaches and for singled out dependent variables (e.g. Bergqvist et al. 2013; Howell and Rehm 2009). Fourth, difficulties in choosing an appropriate independent variable are frequently expressed in the literature and ultimate selections often entail *compromises* (examples follow later on). Fifth, it has never really been discussed or tested, whether the existing indicators adequately capture theoretically assumed *mechanisms* in multilevel analyses of the outcomes of welfare policies, even though concerns are voiced sporadically (e.g. Pfau-Effinger 2005).

Based on these observations, this paper maps out critical issues from the perspective of scholars who are looking for an independent variable in multilevel analyses. For this purpose, single indicators, typologies, and composite indices are inspected more closely. First, all three strategies are discussed conceptually with an emphasis on sources of dissent. In the next step, popular operationalisations are compared in empirical analyses of cross-national survey data from the International Social Survey Programme (ISSP) and the European Social Survey (ESS) in order to illustrate the consequences of different conceptual choices. This is followed by a summary of critical issues and a discussion of possible points of departure for the development of more suitable and standardised operationalisations for the specific use as explanatory instruments.

2 Measuring Welfare Stateness: Popular Operationalisations and the Surrounding Debates

There are many ways of defining and conceptualising the welfare state. While early research mainly focussed on welfare state effort—in most cases represented by social spending—the literature nowadays agrees that social policy arrangements are captured more adequately by focussing on social rights of citizenship (e.g. Esping-Andersen 1999; Stephens 2010). Still, both conceptualisations can be found in empirical operationalisations of welfare stateness. In the following section, debates surrounding the most frequently used operationalisations—single indicators, typologies and composite indices—are briefly summarised.

2.1 The Single Indicator Approach

One very popular way of operationalising different welfare policies is to use single indicators representing important elements of the welfare state. They can be found in general discussions about measuring and classifying regimes² as well as in studies which include

² As indicators underlying the construction of typologies or as part of generosity measures.

characteristics of the welfare state as independent variables (e.g. Jæger 2006; Jordan 2013; Eger and Breznau 2017).

The most commonly chosen indicators are expenditure-based measures (Kvist 2011). Usually, this means including a variable on social spending as a percentage of GDP in one specific policy area (e.g. in the labour market, Schneider and Makszin 2014) or as an overarching measure (e.g. Steele 2015) of welfare effort. Spending indicators receive much criticism. One of the main arguments is that other areas of social policy—for instance entitlement criteria—are more important and that a focus on spending postulates a linearity of welfare efforts which is not given in reality (Esping-Andersen 1990) and disregards how multifaceted systems of social security actually are (Bonoli 1997). Furthermore, a high amount of social spending may signal a generous system but could also mean that more people depend on social benefits (Bergqvist et al. 2013). In the end, equally high spending may not necessarily mean that two countries actually provide similar benefits (Kvist 2011), and we cannot tell if higher or lower income groups profit more from redistribution (as already noted by Titmuss 1974). Such criticism led to a widespread consensus that spending is a problematic proxy for welfare stateness (a more differentiated discussion is given by Jensen 2011).

An alternative that is preferred by literature on welfare state retrenchment is to use net replacement rates (NRR) for individuals in a certain risk position as indicators of welfare generosity based on social rights. However, the calculation of replacement rates is still controversial and they vary depending on the source. This is discussed for example by Scruggs (2013) and Wenzelburger et al. (2013), who compare differences between replacement rates calculated in the Comparative Welfare Entitlement Dataset (CWED2, Scruggs et al. 2014) and the Social Citizenship Program (SCIP, later included in the Social Insurance Entitlement Dataset (SIED)).³ Recently, Bolukbasi and Öktem (2018) add that other non-replacement indicators—such as waiting days and qualification periods—are affected by the same problem and also differ depending on the data source because similar indicators are operationalised based on varying conceptual premises.

Using single indicators as *independent variables* in comparative research has advantages and disadvantages, both of which are visible in the existing literature. The two main disadvantages address their limited informative value on the one hand and the above noted deviations in the calculations on the other hand. In empirical studies, these disadvantages are often outweighed by the main advantage of this operationalisation: since a variety of international organisations such as the OECD and Eurostat offer extensive and regularly updated information on key indicators, data is easily accessible and available for a great number of countries.

A common way to overcome the problem of limited informative value is to use more than one indicator. There are many studies which draw upon a theoretically well-grounded selection of several single indicators representing relevant areas of the welfare state (e.g. Jæger 2006), give a detailed justification of why they choose a single indicator instead of another operationalisation (e.g. Jakobsen 2010; Visser et al. 2018), or examine single indicators along with other operationalisations (e.g. Jakobsen 2011). However, there are also studies, which only briefly elaborate on their selection. This is problematic because there is an obvious conceptual difference between using for instance replacement rates and social expenditure. Still, studies frequently forgo justifying their selection and instead only argue

³ Among other issues, differences in the replacement rates are due to the calculation of taxes and the referenced period of time in which a benefit is received.

that they would have liked to use an alternative (like a composite measure), which was not available for their sample of countries or time periods (e.g. Kulin and Meuleman 2015; Angel and Heitzmann 2015).

Regarding the second disadvantage, I have not yet encountered a study, which analyses the consequences of deviations between data sources when using single indicators as independent variables. Thus, I recommend that further research not only justifies why a specific indicator is chosen, but also discusses the sources of macro-level data in more detail and compares the selection to the referenced literature.

Concluding, using single indicators as proxies for welfare state differences in multilevel frameworks has limitations. Since no recommendations exist regarding which indicator to choose when modelling specific causal assumptions, the selection requires a well-grounded justification. In light of the mentioned divergent operationalisations, failing to do so may have consequences for results and their comparability with other studies using different measures or different data sources.

2.2 The Regime Approach

Esping-Andersen's (1990) "Three Worlds of Welfare Capitalism" (TWWC) has been a substantial contribution to the field. Here, he identifies three distinct types of welfare states in 18 OECD countries based on how social policies impact social stratification,⁴ de-commodification⁵ and the so-called public-private mix.⁶ He identifies a generous *Social-Democratic*, a status-oriented *Conservative*, and a market-oriented *Liberal* regime. This classification has inspired a remarkable body of literature and a critical and ongoing discussion regarding the number, composition and scope of regimes (comprehensive discussions are provided by Arts and Gelissen 2002; Ferragina and Seeleib-Kaiser 2011; Rice 2013; van Kersbergen and Vis 2015). As a result, research following Esping-Andersen's initial typology has introduced a great number of varying classifications. Before discussing the applicability of typologies as independent variables, it is important to look at sources of dissent between different classifications, which address conceptual as well as operational details.

When classifying typical arrangements of social policies, scholars have focussed on very different elements of the welfare state. While some focussed on *how much* a welfare state spends, others classified *how* social policies are organised and financed (Bambra 2007 and Bonoli 1997 discuss and combine both perspectives). Another lively debate surrounds the question *how many* welfare states exist. Popular additions to Esping-Andersen's typology include a Mediterranean (e.g. Ferrera 1996) and a post-socialist welfare regime (e.g. Castles and Obinger 2008).

Turning to the empirical operationalisation of such types, a great variety in both indicators and methods mirrors the varying conceptual considerations. While some studies base their classifications on expenditure (Kuitto 2011), others focus on benefit coverage and replacement rates (Ferrera 1996), or on a two-dimensional approach combining spending and funding of welfare provision (Bambra 2007; Bonoli 1997). Moreover there are those who add measures of economic insecurity (Menahem 2007) or stratification (Esping-Andersen 1990). These indicators are merged into typologies through very different

⁴ The extent to which social policies reproduce, increase or decrease social inequality.

⁵ The extent to which one's social protection is detached from the market.

⁶ The importance of different providers of welfare (state, family, and market).

analytical techniques and each methodological approach claims to shed light on aspects, which have been disregarded so far (e.g. certain indicators or countries).

Lastly, the country sample constitutes a considerable source of variation. The selection of countries, which underlies a typology often draws on pragmatic considerations like data availability (Ebbinghaus 2011). Thus, most studies only cover an arbitrary selection of countries and especially Central and Eastern European (CEE) states are highly under-represented. Apart from the oversight of countries, different samples may affect the classification itself because most approaches determine types based on proximity between cases. For instance, Esping-Andersen's (1990) classification is based on composite indices of decommodification and stratification where countries receive a score based on their deviation from the overall mean. However, mean and deviation vary depending on the included countries and are sensible to slight changes or miscalculations. Ironically, Esping-Andersen himself serves as an example for this.⁷ A similar argument applies to cluster analysis (e.g. Kuitto 2011; Castles and Obinger 2008), which groups countries based on the proximity between them. In light of these differences in conceptualisations and operationalisations, it does not surprise much that the number, title, and composition of regimes differ remarkably between typologies.

The lack of agreement on which typology suits best and which theoretical perspective is preferable is acknowledged in many studies using them as *independent variables*. Nonetheless, many of them still rely heavily on the regime approach—sometimes even with an apologetic reference to the need to circumvent a more detailed discussion of the scientific debate (e.g. Motel-Klingebiel et al. 2009). While regime typologies bear the advantage that they are easily operationalised as dummy variables, their main disadvantage is a practical one: the selection of countries in survey data (like the ESS) usually deviates from the countries covered by a typology. Thus, authors face a difficult conceptual choice having to either exclude unclassified countries or include them by combining classifications or extending them. Since cross-cultural analyses often aim at examining as many countries as possible, the second option is preferred. Such combination or extension often relies on instinct since the literature offers no clear recommendation on what to do in this situation and an abundance of different typologies. As a result, a *buffet strategy* evolved in which authors pick a combination “from the vast array of welfare state typologies” (Arts and Gelissen 2001: 285) that seems helpful for the envisioned purpose. There are many examples for such buffet-approaches (more recently Deeming and Jones 2015; dem Knesebeck et al. 2016; Arundel and Lennartz 2017; Schuck and Steiber 2017) and the proceeding often seems inspired more by practical considerations than by theoretical ones. As a result, many modifications not only entail adding countries that were not classified in whatever typology serves as a starting point, but also go along with uncommented reclassifications. In light of the existing debate on welfare state change, it furthermore seems problematic that many of the buffet-type studies still rely heavily on typologies from the 1990s and assume that those classifications (very prominent are Esping-Andersen 1990 and Ferrera 1996) are still valid and only require some additions or slight modifications.

It was rarely tested how different typologies affect results if treated as independent variables. Bergqvist et al. (2013) provide one of the few overviews using the example of health inequality as dependent variable. In their re-analysis of 34 studies employing regime typologies as independent variable they found not only considerable differences in the kind of

⁷ A miscalculation in the decommodification score for several countries led to a misclassification of several countries (e.g. discussed by Bamba 2006).

typology used and the amendments made to classifications but also in the results. Since different associations with health were even found within identical typologies, they conclude that the main problem is not the theoretical and empirical conception but the general use of welfare regimes as an explanation for health inequality. However, they examined studies, which draw on different data sources and apply different methods of analysis. Thus, it should be tested if their finding holds true if these aspects were kept constant.

Concluding, regime typologies may be a great tool for classifying different policy arrangements. Nevertheless, they rarely fit the country sample in cross-national survey data leading scholars to retreat to combinations and reclassifications. In light of the severe conceptual and operational variations underlying different classifications, such proceeding seems highly problematic. It is thus important to test the consequences of different classification much more thoroughly.

2.3 The Composite Index Approach

Composite indices and scores measuring welfare commitment are comparatively rare. Throughout the literature there are scattered attempts to devise such measures (in an early version e.g. Castles and McKinlay 1979). In more recent approaches, the two indices which underlie Esping-Andersen's (1990) TWWC typology have been a major influence. Especially his Decommodification Index has been replicated, updated and revised (e.g. Bambra 2005; Scruggs and Allan 2006; Scruggs 2014; Kuitto 2016). Noteworthy are furthermore the attempts by Segura-Ubierno (2007) and Cruz-Martinez (2014), who devise multidimensional measures of welfare state arrangements for Latin American countries. However, these proposals have not been adapted for European samples so far. Other composite measures in the literature either take a more specific perspective (e.g. on defamilialisation, Lohmann and Zagel 2016) or a more general one which goes beyond characteristics of social policies and includes overall features of governance (e.g. the Social Policy Index⁸). The main sources of dissent within the index approach include the operationalisation and country sample.

To name some examples for differing operationalisations: Castles and McKinlay (1979) devise an index of welfare commitment based on educational expenditure, transfer payments and infant mortality, Esping-Andersen's (1990) Decommodification Index includes replacement rates, extent and duration of individual contribution, waiting periods and insurance coverage, and Menahem (2007) combines insurance coverage and replacement rates with disposable income. Besides these obvious differences in the choice of indicators, there are also differences when it comes to weighting procedures and modes of standardisation. The Benefit Generosity Index in the Comparative Welfare Entitlement Dataset—an updated and slightly modified version of Esping-Andersen's decommodification Index—z-standardises the underlying variables (Scruggs 2014). In contrast, Esping-Andersen's original version using data from the Social Citizenship Program gives countries a value between one and three for each underlying indicator representing levels of generosity and adds them up. Furthermore, Esping-Andersen only superficially justifies why some indicators are given more weight than others (discussed among others by Bambra 2006). However, as Wenzelburger et al. (2013) point out, not just the modes of combining indicators

⁸ Naren Prasad at the United Nations Research Institute proposed the SPI for Social Development in 2006. No final version of the index has been published apart from a research proposal, which has been picked up by other researchers (e.g. Garcés Ferrer et al. 2016).

vary, the underlying indicators themselves may differ as well depending on the data source (as discussed in the preceding section on single indicators).

The second source of variation within the approach is closely linked to the first. The measures introduced above all rely on mean values and deviations from that mean and are thus very sensible to the underlying country sample. If the composition of countries changes, these values will most likely change as well (as discussed in the case of typologies). This affects the comparability of results and it impairs stretching composite measures to further countries. A way to overcome this problem, which I did not encounter in the literature so far, would be to refrain from standardisations based on mean and deviation. An alternative could be a benchmark approach, which standardises based on the highest existing occurrence of a given indicator in a meaningful population. Such a population could for instance consist of the entire European Union or all OECD member states. In this case, the standardised numbers would indicate how close a country is to an existing frontrunner (for instance the highest existing replacement rate) and they could be used independently of the country sample.

Composite indices are perhaps the most desired but least implemented *independent variables*. They promise the multidimensionality of typologies while maintaining the metric scale and variation of single indicators. However, the number of existing measures is very limited and the most popular ones are only available for a limited selection of countries and points in time. This shortcoming is often stated as a reason for having to resort to a less desirable alternative (e.g. Angel and Heitzmann 2015; Kulin and Meuleman 2015).

Concluding, composite measures represent very promising tools for capturing welfare stateness. However, since the most comprehensive ones cover only a small number of countries, their usefulness as independent variables is very limited at this point.

3 An Illustration: The Welfare State and Differences in Welfare Attitudes

In the following section, the discussed operationalisations are tested empirically with an emphasis on illustrating the advantages and disadvantages mentioned before. In this empirical test, welfare attitudes serve as exemplary dependent variable on the individual level to illustrate the consequences of differing operationalisations. Welfare attitudes are among the more popular dependent variables in the relevant literature. The main assumption entails that attitudes towards social policies are shaped by the institutional context—in this case the welfare state—individuals are embedded in (Arts and Gelissen 2001; Svallfors 1997). It is hypothesised that generous and universal social policies following social-democratic principles generate political support and positive attitudes towards the welfare state (Jaime-Castillo 2013; Roosma et al. 2014) while redistribution-based and targeted policies increase conflicts between beneficiaries and contributors, leading to disapproval of welfare policies (Jordan 2013). However, the empirical tests of this *policy feedback* hypothesis produce mixed results and various studies cannot confirm such a linear relationship between generosity and support (Jæger 2009; Jakobsen 2011). One reason for this may be that different operationalisations of welfare policies have been tested—including different typologies and single indicators. While typologies may fail to grasp subtle differences between welfare states (Jordan 2013), single indicators could be correlated with other macroeconomic indicators and thus may have no independent effect once other variables are controlled (Jæger 2013 suspects this in the case of social expenditure). Due to these divergent findings and the ongoing discussion, welfare attitudes present a good example of a micro-level outcome,

which may be explained differently depending on the conceptualisation of welfare state-ness in an analysis.

3.1 Data and Method

The following analyses use data from the fourth wave of the European Social Survey (ESS 2008) and the International Social Survey Programme (ISSP Research Group 2017). These two datasets were chosen for several reasons. First, they both include questions addressing attitudes towards the welfare state. Second, the data was collected during a similar period of time (mainly 2008 and 2009), which means that the same macro-level indicators can be used in both analyses. Third, both datasets are frequently used in comparative research on how welfare attitudes are shaped by different welfare state arrangements (more recently Kulin and Meuleman 2015; Steele 2015; Eger and Breznau 2017). Fourth, using ESS and ISSP data represents a common situation in which the researcher has no influence on the country selection. Lastly, the comparison between the two datasets will allow to determine—at least partly—the reliability of findings.

To ensure that the examined population is suitable for the proposed analysis and covers comparable units of analysis, the sample is reduced to respondents from countries, which are member states of the European Union or strongly associated with it.⁹ Thus, 21 countries covered by both datasets are included.¹⁰

The *dependent variable* is a question regarding government responsibility for aiding unemployed people. This particular aspect of attitudes towards the generosity of benefits is covered in a comparable—albeit not identical—manner in both datasets. In the ESS, respondents answered the question “*how much responsibility do you think governments should have to ensure a reasonable standard of living for the unemployed?*” on an eleven-point scale ranging from “*should not be governments’ responsibility at all*” to “*should be entirely governments’ responsibility*”. In the ISSP, respondents indicated on a five-point scale to what extent they agreed with the statement “*the government should provide a decent standard of living for the unemployed*”.

The analyses focus on *independent variables* on the country-level. Since the main surveying period of both datasets was in late 2008 and early 2009, those indicators rely foremost on data from 2008. The only exception is SCIP/SIED data, which is available in five-year intervals and was therefore taken from 2005. Furthermore, since the dependent variable addresses attitudes towards generosity in the field of unemployment, macro-level indicators, which represent unemployment policies were chosen, whenever possible.

Four single indicators are tested: *overall social expenditure as percentage of GDP* (Eurostat 2018a), *social expenditure in the field of unemployment policies* (Eurostat 2018b), and two versions of *net replacement rates for unemployed average production workers*, which stem from different data sources and are based on slightly varying operationalisations (CWED2 and SCIP/SIED).

Since there are no typologies covering all analysed countries, two different *buffer-typologies* are included. The first version uses Esping-Andersen’s classification as a starting point and adds a Southern type following Ferrera (1996). The CEE countries were all joined in an Eastern-European group by applying classifications used among others

⁹ EU member states (in 2008) plus Norway and Switzerland.

¹⁰ AT, BE, BG, CH, CZ, DE, DK, EE, ES, FI, FR, GB, HU, LT, LV, NO, PL, PT, SE, SI, SK.

in analyses by Roosma et al. (2014) and Bambra et al. (2014). This leaves Cyprus (only included in additional analyses), which was classified as Southern following Castles and Obinger (2008). The second buffet-typology differs from the first in the classification of two countries, which represent ambiguous cases. Switzerland is classified as Liberal (instead of Conservative) following Obinger and Wagschal (1998) and Ferragina et al. (2013) and Austria is assigned to the Social-Democratic type instead of the Conservative one, which is supported by Arts and Gelissen (2001).

As a composite measure, I include the *Welfare Generosity Score* that is provided in the CWED2 dataset. Since it covers only a small selection of countries and none of the CEE states, I added a few missing indicators¹¹ and updated the index following Scruggs' (2014) instructions so that it now covers all 21 countries in the main analysis. The correlation of my version with the unemployment generosity score already provided in the dataset is very in high (0.98) for the 12 countries that are shared by CWED2, ISSP and ESS.

Furthermore *unemployment rate* is controlled in all models, as is often done in analyses of welfare attitudes (Jæger 2013; Arikan and Ben-Nun Bloom 2015; Eger and Breznau 2017).

Testing the different operationalisations within each of the two surveys should help to illustrate differences while reducing potential bias stemming from varying survey periods and country samples.

The empirical tests are based on *multilevel analyses* (MLA). In the last decades this method has become increasingly popular in comparative research because it takes into account the hierarchical structure of cross-cultural data in which individuals are nested in national contexts. Multilevel analysis is able to estimate variance components on the level of individuals and contexts (in this case countries) simultaneously. This leads to a more correct estimation of standard errors and reduces the risk of ecological or individualistic fallacies, which can arise when results on either level are translated to the other. Furthermore, it enables us to estimate the effects of independent variables on the micro- and macro-level in the same analysis (for a more detailed description see Snijders and Bosker 2012).

3.2 Results

The following two tables report the results of multilevel analyses using the two different data sources. Both versions show very similar intraclasscorrelation coefficients (ICC) in the random-intercept-only model (model 0): in both datasets, about 10 percent of the variation in attitudes towards the role of government can be attributed to the country-level.

Looking at the coefficients, many similarities can be found in the ESS (Table 1) and ISSP (Table 2) data, which indicate a certain robustness of the findings. In both analyses, *overall social expenditure* is negatively associated with wanting a strong role of government in the field of unemployment policies and explains a considerable amount of variation between countries (model 1). *Social expenditure in the field of unemployment policies* (model 2) points in the same direction, even though this effect is only significant in the ISSP analysis. Respondents from countries with higher social expenditure thus want less government responsibility for providing a decent standard of living for the unemployed.

¹¹ Missing data on coverage of unemployment insurance was added from the SIED (Social Policy Indicator Database (SPIN) 2015).

Table 1 Government responsibility for providing standard of living for unemployed (ESS 2008). Source: ESS4-7, own calculations

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Intercept	6.88 (.16)	8.11 (.79)	6.62 (.48)	5.70 (.68)	5.64 (.79)	7.24 (.47)	7.03 (.50)	6.24 (.88)
Single indicators								
SocEx (overall)		-.06 (.03)						
SocEx (unempl.)			-.27 (.23)					
NRR (SCIP/SIED)				1.59 (.90)				
NRR (CWED)					1.70 (1.14)			
Buffet 1 (ref. soc-dem.)								
Conservative								
Liberal						-.94 (.40)		
Southern						-1.20 (.65)		
Eastern						.36 (.66)		
Buffet 2 (ref. soc-dem.)								
Conservative								
Liberal							-.83 (.51)	
Southern							-.90 (.51)	
Eastern							.48 (.72)	
Composite measure							.01 (.38)	
BGI unempl.								.03 (.06)
Control variables								
Unemployment rate		.03 (.07)	.08 (.08)	.04 (.07)	.03 (.07)	-.01 (.08)	.00 (.09)	.05 (.08)
Model fit								
Variance (individuals)	4.76	4.76	4.76	4.76	4.76	4.76	4.76	4.76
Variance (countries)	.54	.42	.49	.46	.48	.33	.36	.52
R ² (countries)		.22	.08	.15	.11	.38	.32	.03
N	40,779 (individuals), 21 (countries)							

Multilevel analysis (FIML); standard errors in parentheses; ICC in Model 0 (Random Intercept Only); .10; bold numbers indicate significance ($p \leq .10$); R² (countries) following Raudenbush and Bryk (2012); buffet-typology 1: Soc-Dem.: DK, FI, SE, NO; Cons.: AT, BE, CH, DE, FR; Lib.: GB; South.: ES, PT; East.: BG, CZ, EE, HU, LT, LV, PL; buffet-typology 2: Soc-Dem.: DK, FI, SE, NO, AT; Cons.: BE, DE, FR; Lib.: GB; CH; South.: ES, PT; East.: BG, CZ, EE, HU, LT, LV, PL, SI, SK; BG/ Benefit Generosity Index

Table 2 Government responsibility for providing standard of living for unemployed (ISSP 2009). *Source:* ISSP, own calculations

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	
Intercept	2.90 (.07)	3.44 (.35)	2.91 (.19)	2.57 (.30)	2.69 (.35)	3.17 (.17)	3.05 (.20)	3.14 (.37)	
Single indicators									
SocEx (overall)		-.02 (.01)							
SocEx (unempl.)			-.22 (.09)						
NRR (SCIP/SIED)			.56 (.39)						
NRR (CWED)					.35 (.50)				
Buffet 1 (ref. soc-dem.)									
Conservative								-.43 (.15)	
Liberal								-.63 (.24)	
Southern								.18 (.24)	
Eastern								.01 (.14)	
Buffet 2 (ref. soc-dem.)									
Conservative								-.33 (.20)	
Liberal								-.35 (.20)	
Southern								.29 (.29)	
Eastern								.12 (.15)	
Composite measure									
BGI unempl.								-.02 (.03)	
Control variables									
Unemployment rate		-.01 (.03)	.03 (.03)	-.00 (.03)	-.00 (.03)	-.03 (.03)	-.03 (.04)	-.00 (.03)	
Model fit									
Variance (individuals)	.91	.91	.91	.91	.91	.91	.91	.91	
Variance (countries)	.10	.08	.07	.09	.09	.04	.06	.09	
R ² (countries)		.15	.22	.09	.02	.53	.40	.03	
N	24,713 (individuals), 21 (countries)								

Multilevel analysis (FIML); standard errors in parentheses; ICC in Model 0 (Random Intercept Only); .10; bold numbers indicate significance ($p \leq .1$); R² (countries) following Raudenbush and Bryk (2012); buffet-typology 1: Soc-Dem.: DK, FI, SE, NO; Cons.: AT, BE, CH, DE, FR; Lib.: GB, South.: ES, PT; East.: BG, CZ, EE, HU, LT, LV, PL; buffet-typology 2: Soc-Dem.: DK, FI, SE, NO, AT; Cons.: BE, DE, FR; Lib.: GB; CH; South.: ES, PT; East.: BG, CZ, EE, HU, LT, LV, PL, SI, SK; BG/ Benefit Generosity Index

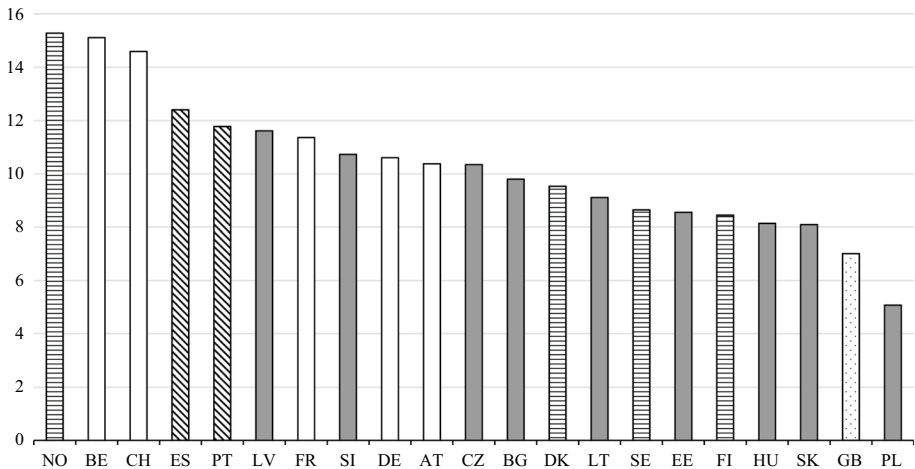


Fig. 1 Unemployment generosity index by country; colours indicate membership in regime according to buffet typology 1; data: CWED2; colouring of pillars indicate membership in regimes according to buffet-typology 1: stripes (horizontal): Soc.-Dem., white: Cons., stripes (diagonal): South., grey: East., dotted: Lib

The two different *unemployment replacement rates* (models 3 and 4) produce slightly differing results. In the ESS analysis, only the version provided by the SCIP/SIED data produces a significant and positive effect, while the CWED2 version is insignificant. In the ISSP analysis neither of the rates exhibit significant effects. Still, this shows that varying data sources should at least be discussed—especially if results are compared with studies using indicators from a different data source. In this analysis, generous benefits in case of unemployment tend to lower support for government responsibility in the field but this effect does not appear to be very robust. Apart from this, the opposed directions of the effects compared to the spending indicators correspond to the prevalent finding that welfare effort and welfare generosity represent very different parts of the welfare state (as outlined in chapter 2.1).

The two *buffet-typologies* (models 5 and 6) consistently show that people living in Liberal welfare states, which are assumed the least generous, are significantly less in favour of government responsibility than those in inclusive Social-Democratic welfare states. Furthermore, the first typology (model 5) also reveals a significantly lower preference for state responsibility in countries belonging to the Conservative type. This effect disappears in the second buffet-typology (model 6) with the different classification of Austria and Switzerland and it indicates that a potential bias due to slightly differing combinations and extensions of existing typologies (as suspected in chapter 2.2) should be taken seriously.

Interpreting these results, the two typologies seem to point into the direction of the policy feedback hypothesis: living in a Social-Democratic welfare state seems to increase support for government action—at least compared to Liberal regimes. On the other hand, the insignificant effect of the Generosity Index (model 7) undermines this finding. Since this index is based on many of the indicators, Esping-Andersen used to construct his initial typology it should at least roughly indicate patterns that correspond to the TTWC typology or one of the succeeding classifications. However, this is hardly the case (Fig. 1). Instead, a ranking of generosity scores shows no clear clusters of countries that correspond to the typologies I used in the analyses, the TWWC or in fact any other typology.

Table 3 Comparison of regime typologies: government responsibility for providing standard of living for unemployed (ESS 2008). *Source:* ESS4–7, own calculations

	Model 0	Model 1	Model 2
Intercept	6.98 (.15)	7.26 (.40)	7.05 (.39)
Buffer 1			
Scandinavian (DK, FI, SE, NO)		Ref.	
Conservative (BE, DE, FR, NL, CH, AT)		−.93 (.37)	
Liberal (GB, IE)		−.80 (.50)	
Southern (CY, ES, GR, PT)		.62 (.45)	
Eastern (BG, CZ, EE, HU, LV, PL, RO, SI, SK, LT)		−.06 (.35)	
Buffer 2			
Scandinavian (DK, FI, SE, NO, AT)			Ref.
Conservative (BE, DE, FR, NL)			−.79 (.41)
Liberal (GB, IE, CH)			−.68 (.43)
Southern (CY, ES, GR, PT)			.76 (.45)
Eastern (BG, CZ, EE, HU, LV, PL, RO, SI, SK, LT)			.09 (.35)
Model fit			
Variance (individuals)	4.65	4.65	4.65
Variance (countries)	.58	.32	.34
R ² (countries)		.45	.41
N	49,608 (individuals); 26 (countries)		

Multilevel analysis (FIML); standard errors in parentheses; ICC in Model 0 (Random Intercept Only): .11; bold numbers indicate significance ($p \leq .10$); R² (countries) following Raudenbush and Bryk (2012); unemployment rate controlled for in all models

In addition to these findings, further analyses (Table 3) show that if the same two buffet typologies are tested in a slightly bigger country sample¹² the result turns out quite differently. Here, Liberal countries no longer differ significantly from Social-Democratic ones. Instead, Conservative welfare states now consistently show significantly less support for government action than the latter, while respondents from countries belonging to the Southern type show significantly more support for state responsibility—even though this effect is only found for the second typology. This finding is somewhat problematic because although it may seem obvious that different country samples may produce different results, samples in secondary analyses of survey data like the ESS will always vary from wave to wave. Thus, even if scholars use the same typologies, the differing samples will still hinder the comparability of results with previous research. Of course, the same argument holds true for every kind of indicator and analysis. Still, typologies exhibit a sense of homogeneity among the members of a category, which may tempt to underestimate the problem.

Summarising the results, the negative effect of social expenditure (overall and in the field of unemployment) on attitudes opposes the policy feedback hypothesis at first glance while net replacement rates and typologies show a tendency to support it. However, the indicators produce very unstable results and small modifications influence the significance of effects severely. Based on this, it would be very difficult to answer *why* attitudes differ.

¹² Based on ESS data: NL, IE, CY, GR, RO are added.

Regardless—and fortunately very much in line with the aim of this paper – the analysis reveals interesting sources of bias. Discussing and finding ways of avoiding these issues may help standardise proceedings.

4 Discussion

In the second chapter of the paper various sources of dissent within each approach are identified, all of which are visible in the subsequent empirical test.

Limited informative value and differing data sources are critical issues within the *single indicator* approach. Even though it may appear trivial to say that replacement rates and social expenditure address singled-out and very different aspects of the welfare state, both are still used as independent variables in analyses of welfare attitudes. The literature does not seem to offer a guideline recommending a standardised selection of suitable indicators and advisable combinations as well as data sources. The latter leads over to the second issue. The analyses reveal small variations in the results and their significance depending on the data source. This indicates a potential bias, which should be examined in more detail.

The *regime approach* is characterised by differences in the underlying conceptual and operational premises. As the empirical example shows, different classifications can affect the results—and there are many other classifications in the literature, which have not been tested in this paper and may produce even more divergent results. Furthermore, the differing country samples in survey data prove to be a highly problematic issue. More research is needed in order to test how much combination and extension a typology can take before results are no longer comparable.

Lastly, the *composite index approach* is very difficult to assess. Since comprehensive examples of this approach are only available for a small number of countries, they need to be extended to bigger country samples. However, the inclusion of more countries—meaning foremost CEE countries—proves to be quite unfruitful. There are many issues, which may be critical when trying to include CEE states in existing measurements. For instance, de jure and de facto benefit generosity in those countries might not coincide entirely, labour market participation differs systematically from older welfare states, atypical employment may be more common, and much more. A comprehensive discussion is given by Kuitto (2016) who extends Esping-Andersen's version of the Decommodification Index to CEE countries and raises these and more important issues.

Several practical recommendations can be made at this point. First, the different operationalisations should not be treated as interchangeable options – neither within nor between approaches. They have different conceptual premises and thus allow different interpretations. If possible, the selection of an indicator should be based on maximising comparability and should not be justified only by a lack of alternatives. Second, data sources should receive more attention. This directly applies to single indicators and indirectly to typologies and composite measures, because they are based on such single indicators. Third, combining and extending typologies should be avoided or follow clear theoretical justifications. Arbitrarily picking and blending classifications from the literature may impair comparability of results quite severely. Fourth, the frequent exclusion of Central and Eastern European countries is dated and obstructive to comparative analyses of social phenomena in Europe and beyond. If the existing indicators do not fit the character of the welfare state in those nations, more attention should be paid to finding proxies, which work for old and new welfare states alike.

Despite these problems, differences between welfare states reflect very important features of modern democracies. The lack of a reliable, easily available and applicable instrument should lead neither to making unsatisfactory compromises nor to excluding the welfare state from the analysis. Thus, it is important to explore what kind of an instrument is actually needed by scholars looking for an independent variable. Based on the previous discussion, I recommend two objectives, which could serve as starting points for a fruitful discussion. First, a more in-depth examination of what a measurement intended as explanatory instrument must entail and in how far it may deviate from existing approaches is needed. Second, there has to be a detailed theoretical and conceptual discussion of the mechanisms, which are hypothesised when exploring the outcome of different welfare state arrangements.

The problems identified in this paper already help to substantiate the first objective because they reveal obstructive issues, which can be avoided. Following the preceding discussion, the main criteria of a suitable explanatory variable should be *clarity*, *availability*, and *comparability*. In other words, it should be clear what information an indicator is based on and why it is a good proxy for the explanans. The indicator should be available for a big enough sample in order to facilitate replications and it has to be comparable to other research.

Strictly speaking, neither typologies nor composite measures fit these premises—at least not in their present form. In both cases the lack of *availability* for a big enough population is rather obvious. Moreover, they also lack *clarity* because their operationalisation aims at capturing the multidimensionality of welfare states and are thus based on a variety of indicators. In the case of composite measures, this combination may average out and thereby mask important outliers (Kvist 2011), while the broad categories of typologies may represent much more than just welfare state policies (like political cultures, economic and democratic development et cetera). As a result, neither of the two operationalisations allow determining, which specific part of the operationalisation is at work if an effect is observed.

This leaves single indicators as perhaps the most fruitful way to operationalise welfare policies as independent variables. Still, while *availability* is much better in this case, *clarity* and *comparability* are not a given. Social expenditure for instance is far from being a precise indicator. As argued in chapter 2.1, high social spending can represent very different things. Furthermore, data sources have to be addressed.

Regarding the second objective, I suggest a closer look at potential *dependent* variables in order to get a clearer picture of the hypothesised mechanisms. It is not enough to assume that ‘the welfare state’ influences an outcome. A key question is *why* this should be the case and *how* the mechanism may work. The answer to both questions does not come from the independent variable, but from the dependent one. This means that different dependent variables require different operationalisations of welfare stateness. Even though many studies reflect on their selection, others do so only very briefly or not at all—especially if the welfare state functions as one out of many explanatory variables or even just a control variable. In order to standardise proceedings and increase comparability of results, there has to be more conceptual work proposing standardised ways to capture the mechanisms underlying different dependent variables.

Returning to the example of welfare attitudes, this course of action is exemplified in Fig. 2. The hypothesis stated that attitudes are a result of policy feedback. The underlying mechanism implies a process of evaluation. To test the assumed affect, we thus require indicators, which contribute to opinion-formation because individuals are likely aware of them. Indicators like waiting and contribution periods, which are integral parts of

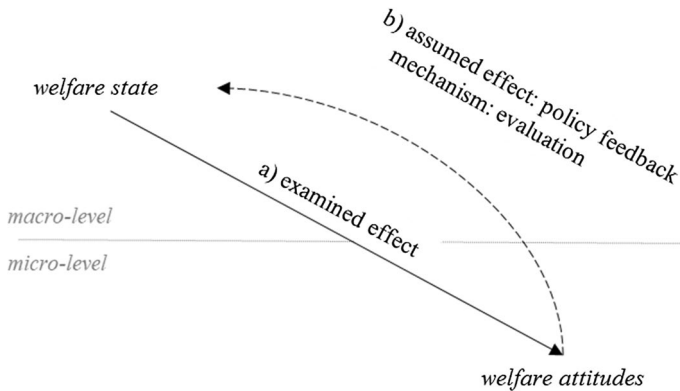


Fig. 2 Relationship between welfare state and welfare attitudes

composite measures and many typologies, do not fall under that category because only a small part of the population will know these details. However, respondents will have at least a basic knowledge of the generosity of benefits (e.g. replacement rates) making this a much better indicator.

If however another exemplary topic were chosen, the argument could be very different. For instance when explaining the risk of poverty, the individual perception and evaluation of social policies is irrelevant. Here the organisation and especially the functioning of welfare policies seems more important—regardless of whether or not the majority of individuals are actually aware of them (e.g. waiting periods or benefit duration). Collecting and systematising these mechanisms and offering suitable indicators for their test, which meet comprehensible criteria should receive much more attention.

5 Conclusion

This paper identifies several problems associated with operationalising ‘welfare stateness’ as an independent variable in macro–micro analyses. A global issue is too much reliance on measures borrowed from literature, which never intended such use in the first place. The conceptual discussion revealed many sources of dissent within strategies and the empirical illustration suggests that their impact on results deserves more attention.

Thus, the central message of this paper is that existing strategies to including welfare state differences as independent variables should be treated much more cautiously. More discussion on the subject and feasible recommendations are needed and it seems very likely that the search for adequate indicators should entail a stronger separation from general comparative welfare state research. Two main objectives for further research are proposed. First, selected independent variables should fulfil criteria such as *clarity*, *availability*, and *comparability*. Second, focussing on the dependent variables and hypothesised macro–micro *mechanisms* seems to be a good point of departure for determining which indicators are useful when explaining specific objects of research. Discussing and substantiating the proposed objectives, may help finding a more standardised way of operationalising welfare stateness in multilevel analyses.

Acknowledgements I am very grateful to Gunnar Otte, Lena Zimmer, my colleagues at the Working Group of Social Stratification Research in Mainz (especially Debora Eicher and Theresa Wieland), and the anonymous referees for valuable comments on an earlier version of this paper.

References

- Angel, S., & Heitzmann, K. (2015). Over-indebtedness in Europe: The relevance of country-level variables for the over-indebtedness of private households. *Journal of European Social Policy*, 25(3), 331–351.
- Arikan, G., & Ben-Nun Bloom, P. (2015). Social values and cross-national differences in attitudes towards welfare. *Political Studies*, 63(2), 431–448.
- Arts, W., & Gelissen, J. (2001). Welfare states, solidarity and justice principles: Does the type really matter? *Acta Sociologica*, 44(4), 283–299.
- Arts, W., & Gelissen, J. (2002). Three worlds of welfare capitalism or more? A state-of-the-art report. *Journal of European Social Policy*, 12(2), 137–158.
- Arundel, R., & Lennartz, C. (2017). Returning to the parental home: Boomerang moves of younger adults and the welfare regime context. *Journal of European Social Policy*, 27(3), 276–294.
- Bambra, C. (2005). Cash versus services: ‘Worlds of welfare’ and the decommodification of cash benefits and health care services. *Journal of Social Policy*, 34(2), 195–213.
- Bambra, C. (2006). Decommodification and the worlds of welfare revisited. *Journal of European Social Policy*, 16(1), 73–80.
- Bambra, C. (2007). ‘Sifting the wheat from the chaff’: A two-dimensional discriminant analysis of welfare state regime theory. *Social Policy & Administration*, 41(1), 1–28.
- Bambra, C., Lunau, T., van der Wel, K. A., Eikemo, T. A., & Dragano, N. (2014). Work, health, and welfare: the association between working conditions, welfare states, and self-reported general health in Europe. *International Journal of Health Services*, 44(1), 113–136.
- Bergqvist, K., Yngwe, M. Å., & Lundberg, O. (2013). Understanding the role of welfare state characteristics for health and inequalities: An analytical review. *BMC Public Health*, 13(1), 1–20.
- Bolukbasi, H. T., & Öktem, K. G. (2018). Conceptualizing and operationalizing social rights: Towards higher convergent validity in SCIP and CWED. *Journal of European Social Policy*, 28(1), 86–100.
- Bonoli, G. (1997). Classifying welfare states: a two-dimension approach. *Journal of Social Policy*, 26(3), 351–372.
- Castles, F., & McKinlay, R. D. (1979). Does politics matter: An analysis of the public welfare commitment in advanced democratic states. *European Journal of Political Research*, 7(2), 169–186.
- Castles, F. G., & Obinger, H. (2008). Worlds, families, regimes: country clusters in European and OECD area public policy. *West European Politics*, 31(1–2), 321–344.
- Clasen, J., & Siegel, N. A. (Eds.). (2007). *Investigating welfare state change: The “dependent variable problem” in comparative analysis*. Cheltenham: Edward Elgar.
- Cruz-Martínez, G. (2014). Welfare state development in Latin America and the Caribbean (1970s–2000s): Multidimensional welfare index, its methodology and results. *Social Indicators Research*, 119(3), 1295–1317.
- Cruz-Martínez, G. (2017). Welfare state development, individual deprivations and income inequality: A Cross-country analysis in Latin America and the Caribbean. *Social Indicators Research*, 134(3), 955–979.
- Deeming, C., & Jones, K. (2015). Investigating the macro determinants of self-rated health and well-being using the European social survey: Methodological innovations across countries and time. *International Journal of Sociology*, 45(4), 256–285.
- Ebbinghaus, B. (2011). Case selection in medium-N comparative welfare state analysis. *Qualitative & Multi-Method Research. APSA Section Newsletter* (Fall), pp. 15–20.
- Eger, M. A., & Breznau, N. (2017). Immigration and the welfare state: A cross-regional analysis of European welfare attitudes. *International Journal of Comparative Sociology*, 30, 002071521769079.
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge, UK: Polity Press.
- Esping-Andersen, G. (1999). *Social foundations of postindustrial economies*. Oxford, New York: Oxford University Press.
- European Social Survey. (2008). Round 4 (ESS 4).
- Eurostat. (2018a). *Social expenditure in 2008 (% GDP): Last updated: March 2018*.
- Eurostat. (2018b). *Social expenditure unemployment (% GDP): Last updated: March 2018*.
- Ferragina, E., & Seeleib-Kaiser, M. (2011). Welfare regime debate: Past, present, futures? *Policy & Politics*, 39(4), 583–611.

- Ferragina, E., Seeleib-Kaiser, M., & Tomlinson, M. (2013). Unemployment protection and family policy at the turn of the 21st century: A dynamic approach to welfare regime theory. *Social Policy & Administration*, 47(7), 783–805.
- Ferrera, M. (1996). The 'Southern model' of welfare in social Europe. *Journal of European Social Policy*, 6(1), 17–37.
- Garcés Ferrer, J., Ródenas Rigla, F., & Vidal Figueroa, C. (2016). Application of Social Policy Index (SPI) amended in three OECD countries: Finland, Spain and Mexico. *Social Indicators Research*, 127(2), 529–539.
- Green-Pedersen, C. (2004). The dependent variable problem within the study of welfare state retrenchment: Defining the problem and looking for solutions. *Journal of Comparative Policy Analysis: Research and Practice*, 6(1), 3–14.
- Howell, D. R., & Rehm, M. (2009). Unemployment compensation and high European unemployment: A reassessment with new benefit indicators. *Oxford Review of Economic Policy*, 25(1), 60–93.
- ISSP Research Group. (2017). *International social survey programme: Social inequality IV - ISSP 2009. GESIS Data Archive, Cologne. ZA5400 Data file Version 4.0.0*. Mannheim: ISSP Research Group.
- Jæger, M. M. (2006). Welfare regimes and attitudes towards redistribution: The regime hypothesis revisited. *European Sociological Review*, 22(2), 157–170.
- Jæger, M. M. (2009). United but divided: Welfare regimes and the level and variance in public support for redistribution. *European Sociological Review*, 25(6), 723–737.
- Jæger, M. M. (2013). The effect of macroeconomic and social conditions on the demand for redistribution: A pseudo panel approach. *Journal of European Social Policy*, 23(2), 149–163.
- Jaime-Castillo, A. M. (2013). Public opinion and the reform of the pension systems in Europe: The influence of solidarity principles. *Journal of European Social Policy*, 23(4), 390–405.
- Jakobsen, T. G. (2010). Public versus private: The conditional effect of state policy and institutional trust on mass opinion. *European Sociological Review*, 26(3), 307–318.
- Jakobsen, T. G. (2011). Welfare attitudes and social expenditure: Do regimes shape public opinion? *Social Indicators Research*, 101(3), 323–340.
- Jensen, C. (2011). Less bad than its reputation: Social spending as a proxy for welfare effort in cross-national studies. *Journal of Comparative Policy Analysis: Research and Practice*, 13(3), 327–340.
- Jordan, J. (2013). Policy feedback and support for the welfare state. *Journal of European Social Policy*, 23(2), 134–148.
- Kühner, S. (2007). Country-level comparisons of welfare state change measures: another facet of the dependent variable problem within the comparative analysis of the welfare state?: ESPA net/JESP Doctoral Researcher Prize Essay. *Journal of European Social Policy*, 17(1), 5–18.
- Kuitto, K. (2011). More than just money: Patterns of disaggregated welfare expenditure in the enlarged Europe. *Journal of European Social Policy*, 21(4), 348–364.
- Kuitto, K. (2016). Measuring welfare entitlement generosity in transitional welfare states: The case of post-communist countries in central and Eastern Europe. *Social Indicators Research*, 136, 1–22.
- Kulin, J., & Meuleman, B. (2015). Human values and welfare state support in Europe: An east-west divide? *European Sociological Review*, 31(4), 418–432.
- Kvist, J. (2011). One-size-fits-all? Measurement issues in medium-N comparative welfare analysis. *Qualitative & Multi-Method Research. APSA Section Newsletter*, 9(Fall), 20–25.
- Lohmann, H., & Zagel, H. (2016). Family policy in comparative perspective: The concepts and measurement of familization and defamilization. *Journal of European Social Policy*, 26(1), 48–65.
- Mattila, M., & Rapeli, L. (2018). Just sick of it?: Health and political trust in Western Europe. *European Journal of Political Research*, 57(1), 116–134.
- Menahem, G. (2007). The decommodified security ratio: A tool for assessing European social protection systems. *International Social Security Review*, 60(4), 69–103.
- Motel-Klingebiel, A., Romeu Gordo, L., & Betzin, J. (2009). Welfare states and quality of later life: Distributions and predictions in a comparative perspective. *European Journal of Ageing*, 6(2), 67–78.
- Obinger, H., & Wagschal, U. (1998). Drei Welten des Wohlfahrtsstaates? Das Stratifizierungskonzept in der clusteranalytischen Überprüfung. In S. Lessenich & I. Ostner (Eds.), *Welten des Wohlfahrtskapitalismus: Der Sozialstaat in vergleichender Perspektive* (Vol. 40, pp. 109–135)., Theorie und Gesellschaft Frankfurt: Campus.
- Orloff, A. S. (1993). Gender and the social rights of citizenship: The comparative analysis of gender relations and welfare states. *American Sociological Review*, 58(3), 303–328.
- Pfau-Effinger, B. (2005). Culture and welfare state policies: reflections on a complex interrelation. *Journal of Social Policy*, 34(1), 3–20.
- Pierson, P. (1996). The new politics of the welfare state. *World Politics*, 48(02), 143–179.

- Raudenbush, S. W., & Bryk, A. S. (2012). *Hierarchical linear models: Applications and data analysis methods* (2nd ed., Vol. 1)., Advanced quantitative techniques in the social sciences Thousand Oaks: Sage.
- Rice, D. (2013). Beyond welfare regimes: From empirical typology to conceptual ideal types. *Social Policy & Administration*, 47(1), 93–110.
- Roosma, F., van Oorschot, W., & Gelissen, J. (2014). The preferred role and perceived performance of the welfare state: European welfare attitudes from a multidimensional perspective. *Social Science Research*, 44, 200–210.
- Schneider, C. Q., & Makszin, K. (2014). Forms of welfare capitalism and education-based participatory inequality. *Socio-Economic Review*, 12(2), 437–462.
- Schuck, B., & Steiber, N. (2017). Does intergenerational educational mobility shape the well-being of young Europeans?: Evidence from the European Social Survey. *Social Indicators Research*, 37, 307.
- Scruggs, L. (2013). Measuring and validating social program replacement rates. *Journal of European Public Policy*, 20(9), 1267–1284.
- Scruggs, L. (2014). *Social welfare generosity scores in CWED 2: A methodological genealogy: CWED working paper 01, February 2014*.
- Scruggs, L., & Allan, J. P. (2006). Welfare-state decommodification in 18 OECD countries: A replication and revision. *Journal of European Social Policy*, 16(1), 55–72.
- Scruggs, L., Jahn, D., & Kuitto, K. (2014). *Comparative welfare entitlements dataset 2. version 2014-03*.
- Segura-Ubierno, A. (2007). *The political economy of the welfare state in Latin America. The political economy of the welfare state in Latin America: Globalization, democracy, and development*. Cambridge: Cambridge University Press.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multi-level modeling* (2nd ed.). Los Angeles: Sage.
- Social Policy Indicator Database (SPIN). (2015). *Social Insurance Entitlements Dataset (SIED)*.
- Steele, L. G. (2015). Income inequality, equal opportunity, and attitudes about redistribution. *Social Science Quarterly*, 96(2), 444–464.
- Stephens, J. D. (2010). The social rights of citizenship. In F. G. Castles, S. Leibfried, J. Lewis, H. Obinger, & C. Pierson (Eds.), *The Oxford handbook of the welfare state* (pp. 511–525). Oxford: Oxford University Press.
- Svallfors, S. (1997). Worlds of welfare and attitudes to redistribution: A comparison of eight western nations. *European Sociological Review*, 13(3), 283–304.
- Titmuss, R. M. (1974). *Social policy. An introduction: Edited by Brian Abel-Smith and Kay Titmuss*. London: George Allen & Unwin Ltd.
- van Kersbergen, K., & Vis, B. (2015). Three worlds' typology: Moving beyond normal science? *Journal of European Social Policy*, 25(1), 111–123.
- Visser, M., Gesthuizen, M., & Scheepers, P. (2018). The crowding in hypothesis revisited: New insights into the impact of social protection expenditure on informal social capital. *European Societies*, 20, 1–24.
- von dem Knesebeck, O., Vonneilich, N., & Kim, T. J. (2016). Are health care inequalities unfair? A study on public attitudes in 23 countries. *International Journal for Equity in Health*, 15(1), 61.
- Wenzelburger, G., Zohlnhöfer, R., & Wolf, F. (2013). Implications of dataset choice in comparative welfare state research. *Journal of European Public Policy*, 20(9), 1229–1250.