Ecological Indicators 90 (2018) 38–46

Contents lists available at ScienceDirect

Ecological Indicators

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

Original Articles

Defining and measuring urban sustainability in Europe: A Delphi study on identifying its most relevant components

Jurian V. Meijering[a,*], Hilde Tobi[a], Kristine Kern[b,c]

[a] Biometris, Wageningen University & Research, Droevendaalsesteeg 1, P.O. Box 16, 6700 AA, Wageningen, The Netherlands
[b] Leibniz Institute for Research on Society and Space, Flakenstraße 29-31, 15537 Erkner, Germany
[c] Åbo Akademi University, Finland

ARTICLE INFO

Keywords:
Delphi method
Operationalization
Dimensions
Categories
Themes
Indicators

ABSTRACT

Urban sustainability rankings may be useful for urban planning. How urban sustainability is defined influences the results of urban sustainability rankings. Various efforts have been made to define the concept and to operationalize it into specific components (e.g. air quality, inequality, employment). Consequently, numerous different components are currently being used without agreement on which components are most relevant for defining and measuring urban sustainability. This study identified which components experts find most relevant for defining and measuring urban sustainability in a European context. The study thereby provides insight into what the concept actually entails. This may facilitate the development of future urban sustainability rankings. A European sample of 419 urban sustainability experts was invited to participate in a three-round Delphi study. In each round experts were asked to evaluate and comment on the relevance of various components of urban sustainability. The following seven components were identified as most relevant: air quality, governance, energy consumption, non-car transportation infrastructure, green spaces, inequality, and CO_2 emissions. Five of these components are part of the environmental dimension of urban sustainability, which suggests that urban sustainability is still perceived as mainly an environmental concept. Based on experts' evaluations of the components, weights could be established that reflect the relative relevance of each component for measuring urban sustainability. This study provides an expert-based framework in which urban sustainability is operationalized into several weighted components. This framework may be used by future developers of urban sustainability rankings to properly define the concept and to select appropriate indicators.

1. Introduction

With almost three-quarters of Europeans living in urban areas, Europe is among the most urbanized continents in the world (UN-Habitat, 2016). European cities are the engines of the European economy (European Commission, 2011) and in various ways they have been front-runners in the field of sustainable urban development (Newman and Kenworthy, 1999). However, European cities are also home to many problems such as unemployment, poverty, and environmental pollution to name only a few (European Commission, 2011). The European Union is therefore committed to making its cities more sustainable (European Commission, 2010). In 2008 the European Commission initiated the European Green Capital Award: a competition in which European cities are evaluated and ranked according to their environmental standards and commitment to future environmental improvement and sustainable development (European Commission, 2015). Since then, other institutions have also developed and published some form of urban sustainability ranking. Examples are the European Green City Index (Siemens, 2009) and the Sustainable Cities Index (Arcadis, 2015).

City rankings may be useful for urban governance, in particular urban planning and development (Besecke and Herkommer, 2007), but should also be used with caution because of methodological issues (Meijering et al., 2014). City rankings are often based on an indicator system. This means that the ranking attribute on which cities were finally ranked (e.g. urban sustainability) is operationalized into various indicators that each measure a specific aspect of the ranking attribute. For each city, data were collected on the indicators and then aggregated into a composite index value and corresponding rank number. Although rankings thus developed ought to reflect the performance of cities on the ranking attribute, they may be very sensitive to various methodological choices made, such as the techniques used to normalize and weigh indicators (Floridi et al., 2011; Jacobs et al., 2005; Lun et al., 2006).

* Corresponding author.
E-mail address: jurian.meijering@wur.nl (J.V. Meijering).

https://doi.org/10.1016/j.ecolind.2018.02.055
Received 6 June 2017; Received in revised form 24 January 2018; Accepted 26 February 2018
Available online 06 March 2018
1470-160X/

A fundamental choice concerns the definition of the ranking attribute. How urban sustainability is defined influences the selection of indicators and thereby ranking results (McManus, 2012). In this regard it is problematic that many different definitions of urban sustainability exist (for an overview see Huang et al. (2015)). Additionally, various concepts related to urban sustainability have been developed during the last couple of years. Bibliometric analyses of the trajectory of urban sustainability concepts revealed that the various concepts can be categorized into two clusters: one with an emphasis on eco-economic issues and one focusing more on socio-economic issues. Other concepts, such as low-carbon city or eco-city, seem to be hybrid forms which enrich the traditional sustainable city concept (de Jong et al., 2015; Fu and Zhang, 2017).

Since the report of the Brundtland Commission (World Commission on Environment and Development, 1987), it is widely accepted that sustainability in general and urban sustainability in particular consists of three pillars or dimensions: environmental, economic, and social sustainability (Hassan and Lee, 2015; Huang et al., 2015; Tanguay et al., 2010). Still, these three dimensions are very abstract and open to a wide range of interpretations. To help define and measure urban sustainability, various efforts have been made to divide the three traditional dimensions of urban sustainability into more specific components, also referred to as themes or categories (see for example Huang et al., 1998; Michael et al., 2014; Tanguay et al., 2010). So far, these efforts all ended up with a different mix of components. As a result, many different components of urban sustainability are currently being used without agreement on which components are most relevant for defining and measuring the concept.

Meijering et al. (2014) suggested that agreement on the definition and operationalization of urban environmental sustainability may be achieved by using the Delphi method: a structured data-collection method that aims to facilitate a group of experts in achieving agreement on a topic. The method has indeed been frequently used to develop definitions and operationalizations of various concepts such as 'team effectiveness' (Lohuis et al., 2013) and 'acute respiratory distress syndrome' (Ferguson et al., 2005). Therefore, the objective of the current study was to identify which components experts find most relevant for defining and measuring urban sustainability in a European context by means of the Delphi method. In doing so, the study provides insight into what the concept actually entails. This may help developers of future urban sustainability rankings to properly define the concept, select appropriate components, and in turn, find or develop corresponding indicators. The study was restricted to the European context as urban sustainability is a place-dependent concept (Hassan and Lee, 2015) and may thus be defined and measured differently in different parts of the world. Within this context an area may be defined as urban when at least 50% of the inhabitants live in high-density clusters (i.e. contiguous grid cells of 1 km^2 with a density of at least 1500 inhabitants per km^2 and a minimum population of 50.000; see Dijkstra and Poelman (2014) for full details).

2. Methods

2.1. The Delphi method

The Delphi method, developed in the 1950s by Dalkey and Helmer (1963), consists of at least two rounds of data collection. In the first round experts are independently questioned about their opinion on the topic of interest, usually by means of a standardized questionnaire. To prevent group pressure and inadvertent influence of dominant individuals, experts participate anonymously and do not directly communicate with each other. Instead, the study moderator provides experts with so called controlled opinion feedback: a summary of the findings from the previous round. Based on this feedback experts are allowed to reconsider and change their opinion in the second round. This process continues until a pre-specified number of rounds has been completed, a certain level of agreement has been achieved, or experts' opinions have stabilized (Diamond et al., 2014; Hasson and Keeney, 2011; Linstone and Turoff, 1975). In the current Delphi study a sample of urban sustainability experts was questioned about which components are most relevant for defining and measuring urban sustainability in a European context. The number of Delphi rounds was pre-specified at three as it was expected, based on a previous comparable Delphi study (Meijering et al., 2015), that this would be sufficient to obtain the required data.

2.2. Expert sample

Urban sustainability experts were considered to be people whose work is related to urban sustainability as inferred from the institution they work for, their position within that institution, their job description, or work related activities (i.e. participating in urban sustainability conferences or projects). With regard to Delphi studies it is recommended to compile a heterogeneous panel of experts to assure the inclusion of a diverse range of views (Hussler et al., 2011; Powell, 2003). For the current study it was therefore decided to search for urban sustainability experts from four different types of institutions: academia, business, civil society (i.e. NGOs, non-profit, and community-based organisations that pursue charitable or member-oriented goals), and government. These four types of institutions have all been involved in the development of urban sustainability rankings. To illustrate, the European Smart Cities Ranking was developed by a team of researchers from the Vienna University of Technology, the Networked Society Index by Ericsson, the Sustainable Cities Index by the Forum for the Future, and The European Green Capital Award by the European Commission. In addition to searching for experts from four different types of institutions, it was decided to search for experts from various European countries.

A convenience sample was assembled from various sources. Several conferences on urban sustainability that took place in different European countries in 2013 or 2014 formed a first major source of experts. Initially, contributors (i.e. presenters and authors of accepted abstracts as mentioned in the conference program or proceedings) to the following three conferences were regarded as potentially suitable experts: The Sustainable City Conference 2014 (Siena, Italy), The Urban Sustainability and Resilience Conference 2014 (London, United Kingdom), and The PLEA Conference 2013 (on sustainable architecture and urban design, Munich, Germany). Because the three conferences mainly yielded experts from academia, additional experts were acquired from two conferences targeted at a more diverse audience: The Future Cities Forum 2014 (Munich, Germany) and The Reference Framework for European Sustainable Cities Conference 2013 (Brussels, Belgium). Names of potentially suitable experts were researched online to acquire additional background information (i.e. the institution they work for, their position within that institution, their e-mail address) and to verify whether they held a position in an institution located in a European country.

The Joint Programming Initiative Urban Europe formed another major source of experts. This program was established in 2010 by the European Commission and aims to "Enhance the capacities and knowledge on transition towards more sustainable, resilient and liveable urban developments" (Robinson et al., 2015). By means of two calls for proposals the program selected and funded 20 projects. The coordinators of these projects and their project partners as listed on the website were regarded as potentially suitable experts. Their names were researched online to acquire additional background information and to verify whether they held a position in an institution located in a European country.

Finally, by searching on the internet and talking to experts, many institutions were found that are active in the field of urban sustainability. For example, developers of urban sustainability rankings (e.g. Arcadis), partners of sustainable and smart city conferences (e.g.

Accenture), EU funded research projects (e.g. TRANSFORM), partnerships (e.g. Climate-KIC), and consortia (e.g. Amsterdam Institute for Advanced Metropolitan Solutions). These institutes were contacted by telephone to find out whether they had urban sustainability experts who were willing to participate in the Delphi study. In some cases names and contact details of one or more experts were directly acquired. In other cases a contact person agreed to forward a ready-made invitation to potentially suitable experts or to put the invitation in a newsletter or on a website. In this invitation the objective and procedure of the Delphi study was shortly explained. Furthermore, experts were asked to register their participation by sending an e-mail to the research team with the following information: their name, the name of the European country where they work, the name of the institution they work for, their job title, and how their work relates to urban sustainability. Experts who responded to the invitation were included in the sample.

Based on all sources a sample of 419 experts was assembled. This sample included experts from all four institution types, although there was an overrepresentation of experts from academia. Within each institution type, experts came from a wide range of fields. Experts from academia worked in departments such as Civil & Environmental Engineering, City & Regional Planning, and Human Geography to name only a few. Experts from business worked for different kinds of companies, such as Arcadis, Ericsson, and Bipolaire Arquitectos. Experts from civil society were few in number, but came for example from Climate-KIC, Energy Cities, and the World Future Council. Finally, experts from government worked for various local authorities as well as other governmental institutions such as the Joint Research Centre of the European Commission and the French Environment and Energy Management Agency. The experts worked in many different European countries (including Albania, Austria, Belgium, Czech Republic Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Switzerland, Turkey, and the United Kingdom). Of the 419 experts in the sample, 139 participated in the first round of the Delphi study.

2.3. Questionnaire development

The questionnaire for the first round of the Delphi study was developed based on several existing urban sustainability rankings. By entering a search query in three different search engines on the internet (Google, Bing, Yahoo) on two different computers, eight rankings were found that rank European cities on their sustainability or another closely related ranking attribute (see Table 1). The components and the corresponding indicators of these eight rankings were identified and put together in one list. This list was content analysed by categorizing components of different rankings (e.g. air pollution, ambient air quality) and labelling each category with a single component name (e.g. air quality). As such, a total of 38 components was identified of which 22 components recurred in at least three of the eight rankings. These 22 components were given a single-sentence explanation and presented in a concept version of the questionnaire. This concept was pre-tested using the cognitive interview approach (Willis, 2005) with three urban sustainability experts and one research methodologist from Wageningen University. Based on the pre-tests, the questionnaire and the list of components were reviewed and refined. One component (public service availability) was removed from the list because two urban sustainability experts noted that it was too comprehensive. The 21 components included in the final version of the questionnaire are listed in Appendix A.

Table 1
Overview of urban sustainability rankings used to construct the initial list of components.

Ranking	Initiator/developer	Edition
Cities of Opportunity	PricewaterhouseCoopers	2014
European Green Capital Award	European Commission	2016
European Green City index	Siemens/Economist Intelligence Unit	2009
European Smart Cities	Vienna University of Technology (Department of Spatial Planning)	2014
Networked Society City Index	Ericsson	2014
Sustainable Cities Index	Arcadis/Centre for Economics and Business Research	2015
The Smartest Cities in the World	Boyd Cohen	2014
The Sustainable Cities Index	Forum for the Future	2010

In the questionnaire experts were asked to express their agreement or disagreement with two statements about their work in relation to urban sustainability, using a 7-point scale ranging from 'strongly disagree' to 'strongly agree'. Next, experts were presented with the 21 components in a random order and were asked the following question: "Based on your expertise, how relevant are the following components for defining and measuring 'Urban Sustainability' in a European context?" For each component experts could give their answer on a 10-point scale, ranging from 1 'not relevant at all' to 10 'entirely relevant'. Subsequently, experts were asked to explain for up to three components (they had rated with at least an 8) why they evaluated them as relevant. Additionally, experts were invited to suggest up to three relevant components that they missed on the list of 21 components. The questionnaire was concluded with several background questions to verify whether experts were assigned to the correct institution type and European country.

For the second Delphi round a questionnaire was developed which was similar to the previous one. Again, experts were asked to evaluate the relevance of the 21 components for defining and measuring urban sustainability. This time, each component was accompanied by a summary of the findings from the first round. This summary consisted of a table with summary statistics showing the component's median evaluation, interquartile range, and the percentage of evaluations equal to or greater than 8, compared to the most and least relevant component. The summary also provided a short explanation of why experts evaluated the component as relevant. As an example, Fig. 1 shows the summary that accompanied the component 'green spaces'. Next, experts were asked to evaluate the relevance of six additional components which were added based on experts' suggestions given in the first round (see Appendix A). Finally, experts were once again invited to explain for up to four components (rated with at least an 8) why they evaluated them as relevant.

The questionnaire for the third and final Delphi round had a different set-up. First, an overall summary of the findings from the second round was presented. This summary consisted of two parts: (1) a table with summary statistics showing each component's median evaluation, interquartile range, and percentage of evaluations equal to or greater than 8, and (2) short explanations of why experts evaluated each component as relevant. Experts were asked to read the summary carefully and to select, from the list of 27 components, five components which are most relevant for defining and measuring urban sustainability. To acquire insight into the relative relevance of the five selected components, experts were also asked to distribute 100 points across these components. Finally, experts were invited to explain why they selected the five components as most relevant.

All three questionnaires were written in English, checked by a native English speaker, and programmed as web-surveys using the online survey builder Qualtrics (Qualtrics, 2015). In May 2015 an invitation e-mail was sent to all 419 experts in the sample. One week later experts received an e-mail with a link to the first online questionnaire. The e-mails with a link to the second and third questionnaire were sent in June and September 2015 respectively. In each round of the Delphi study experts were given up to three weeks to complete the questionnaire. Experts who did not respond received a maximum of two reminders.

Summary statistics of the component '**green spaces**' in comparison to the most and least relevant components:

	Median evaluation [1]	Interquartile range [2]	% of evaluations ≥ 8
Most relevant component	9	8 - 10	88%
Green spaces	9	8 - 10	81%
Least relevant component	6	5 - 8	28%

Note: experts evaluated components on a scale which ranged from 1 'not relevant at all' to 10 'entirely relevant'.
[1] When all evaluations of a component are ordered from lowest to highest, the median is the middle evaluation.
[2] The interquartile range contains the middle 50% of evaluations.

Summary of experts' explanations regarding why the component is relevant:
Green spaces provide many benefits for inhabitants of cities, such as an improved well-being and opportunities for recreation and community building. Furthermore, green spaces offer potential solutions for various problems related to among others climate change, water runoff, and the urban heat island effect. Green spaces are also relevant for nature and biodiversity within a city.

Based on your expertise, how relevant is the component 'green spaces' for defining and measuring 'Urban Sustainability' in a European context?
Please give your answer by using the scale which ranges from 1 'not relevant at all' to 10 'entirely relevant'.

	not relevant at all 1	2	3	4	5	6	7	8	9	entirely relevant 10	don't know
Green spaces: the amount of nature and parks within a city	○	○	○	○	○	○	○	○	○	○	○

Fig. 1. Screenshot of the questionnaire administered in the second Delphi round.

2.4. Analytic procedure

After the first and second round of the Delphi study the following statistics were calculated for each component: the median evaluation, the interquartile range, and the percentage of evaluations equal to or greater than 8. After the third round the percentage of experts that selected each component as most relevant was calculated as well as the total number of points attributed to each component.

To identify the most relevant components in the first and second Delphi round, a 95% confidence interval was calculated for each component around the percentage of evaluations equal to or greater than 8. A component was regarded as one of the most relevant when the lower limit of that 95% confidence interval was greater than 50%, implying that significantly more than half of all experts evaluated the component with at least an 8. To identify the most relevant components in the third round, a 95% confidence interval was calculated for each component around the percentage of experts that had selected it. A component was regarded as one of the most relevant when the lower limit of the 95% confidence interval was greater than 19%, the percentage of experts that would have selected the component based on chance alone. All confidence intervals were calculated using the Exact method (Newcombe, 1998).

The level of agreement among experts regarding the relevance of each component was estimated using the Strict Agreement index (denoted SA) (Meijering et al., 2013). This index is easy to interpret as it expresses the number of agreeing expert pairs as a proportion of the total number of possible expert pairs, whereby two experts only agree if they evaluated an item (i.e. component) using the same point on the rating scale. Originally, the SA was developed for a 5-point scale whereas in the current Delphi study a 10-point scale was used in the first and second round. Therefore, when calculating the SA in these two rounds the agreement criterion of the index was relaxed in such a way that two experts agree if their evaluations did not differ by more than one point on the rating scale. Theoretically, the SA can take on any value between 0 (none of the experts agreed with each other) and 1 (all experts agreed with each other). Preferably, the SA of the most relevant components as identified in this Delphi study is greater than 0.5 as this implies that more than half of all expert pairs agreed on the evaluation of these components.

The robustness of the results regarding the most relevant components was examined by testing whether there were significant differences between experts from different institutions and experts from different European regions. As most experts in the sample were from academia and the number of experts from business, civil society, and government was limited, a distinction was made between experts from academia and experts from other institutions. Charron et al. (2015) showed that countries in Western Europe (i.e. covering the UK, Ireland, the three Benelux countries, France, Germany, Austria and Switzerland) and Northern Europe generally score better than countries in Southern Europe and in Central and Eastern Europe (CEE) regarding the quality of governance, a concept that is related to countries' social and economic development as well as their environmental conditions. Moreover, in most city rankings (e.g. European Green City Index, European Green Capital Award) cities in Northern and Western Europe generally score better than cities in Southern Europe and in CEE. Therefore, a distinction was made between experts from Northern and Western Europe on the one hand (in the following called North-West Europe) and Southern Europe and CEE on the other (in the following called South-East Europe). In the first and second Delphi round differences between groups were tested using the Mann-Whitney test (denoted U), whereas in the third round differences were tested using Fisher's exact test (denoted F) (Lindgren, 1976). Because of multiple testing, only differences with a p-value less than 0.01 were considered significant.

Content analysis (Gray, 2004) was performed on answers given on open questions. Suggestions for additional components relevant for defining and measuring urban sustainability, were analysed by categorizing comparable suggestions. For each category that contained more than four suggestions a new component with a corresponding single-sentence explanation was formulated that captured the underlying suggestions as fully as possible. Appendix A shows the resulting

Table 2
Number of respondents in each round of the Delphi study.

Round	Total response	Response per institution type[a]				Response per European region[b,c]	
		Academia	Business	Civil society	Government	North-West	South-East
1	139	86	18	8	20	86	43
2	99	59	16	6	13	61	31
3	72	47	8	4	10	39	26

[a] Seven expert were not classifiable due to conflicting information.
[b] Based on Charron et al. (2015) with Switzerland categorized in North-West and Albania in South-East Europe.
[c] Ten experts were not classifiable as they stated to work on a European level or did not specify the correct country.

new components included in the second and third round of the Delphi study. Explanations on the relevance of each component were analysed by categorizing comparable explanations, labelling each category with a single sentence that captured the underlying explanations as fully as possible, and combining the labels into one short explanation stating why the component was considered to be relevant. Fig. 1 shows an explanation for the component ‘green spaces’.

3. Results

3.1. Response and drop-out rates

In the first round 33% of the 419 experts in the panel responded (see Table 2). Of these experts 93% agreed with the statement “It is important to me that my work relates to urban sustainability” and 86% agreed with the statement “In my work I spend most of my time on urban sustainability issues”, indicating that there was a high level of commitment among the respondents with regard to urban sustainability. In the second and third round 29% and 27% of the experts dropped-out respectively. In all three rounds more experts from academia than experts from other institutions (business, civil society, government) responded. Furthermore, more experts from North-West Europe than experts from South-East Europe responded. In the first round, North-West Europe included experts from Austria (n = 9), Belgium (n = 4), Denmark (n = 3), Finland (n = 4), France (n = 4), Germany (n = 12), Luxembourg (n = 2), the Netherlands (n = 25), Norway (n = 2), Sweden (n = 5), Switzerland (n = 2), and the United Kingdom (n = 14), while South-East Europe included experts from Albania (n = 1), Greece (n = 3), Italy (n = 20), Lithuania (n = 1), Poland (n = 5), Portugal (n = 5), Romania (n = 1), Spain (n = 5), and Turkey (n = 2).

3.2. The first Delphi round

In the first round the percentage of evaluations equal to or greater than 8 ranged between 28% (international embeddedness) and 88% (non-car transportation infrastructure) (see Table 3). Fourteen components were evaluated with at least an 8 by significantly more than half of all experts and were thus identified as most relevant for defining and measuring urban sustainability. Of these fourteen components the following ten also had a SA greater than 0.5: non-car transportation infrastructure, air quality, CO_2 emissions, energy consumption, green spaces, health, solid waste, climate resilience, waste water treatment, and water usage.

Significant differences were found between experts from academia and experts from other institutions regarding the relevance of the following seven components: non-car transportation infrastructure ($U = 1395.5$, $p < .01$), air quality ($U = 1495$, $p < .01$), CO_2 emissions ($U = 1429.5$, $p < .01$), energy consumption ($U = 1055.5$, $p < .01$), solid waste ($U = 1396$, $p < .01$), waste water treatment ($U = 1210.5$, $p < .01$), and water usage ($U = 1103$, $p < .01$). Experts from academia tended to evaluate these components as more relevant than experts from other institutions. Significant differences were also found between experts from North-West Europe and South-East Europe regarding the relevance of the components ‘air quality’ ($U = 2341$, $p < .01$) and ‘health’ ($U = 2407$, $p < .01$), Experts from South-East Europe tended to evaluate these components as more relevant than experts from North-West Europe.

Table 3
Percentage of expert evaluations ≥8 (95% Confidence Interval) and SA per component in the first and second Delphi round.

Component	Round 1 (n = 139)		Round 2 (n = 99)	
	% evaluations ≥ 8 (95% CI)	SA	% evaluations ≥ 8 (95% CI)	SA
Air quality	87 (80–92)[a]	0.64	89 (81–94)[a]	0.66
CO_2 emissions	87 (80–92)[a]	0.61	89 (81–94)[a]	0.58
Non-car transportation infrastructure	88 (81–93)[a]	0.61	89 (81–94)[a]	0.66
Energy consumption	83 (75–89)[a]	0.59	88 (80–94)[a]	0.68
Governance	n.a.	n.a.	88 (80–94)[a]	0.65
Green spaces	81 (74–87)[a]	0.58	87 (78–93)[a]	0.67
Health	82 (75–88)[a]	0.59	85 (76–91)[a]	0.63
Solid waste	81 (74–87)[a]	0.54	80 (70–87)[a]	0.57
Climate resilience	80 (72–86)[a]	0.55	78 (68–85)[a]	0.54
Waste water treatment	82 (74–88)[a]	0.55	77 (67–85)[a]	0.59
Water usage	82 (74–88)[a]	0.53	75 (65–83)[a]	0.58
Education	61 (53–70)	0.41	71 (61–80)[a]	0.51
Civic engagement	71 (62–78)	0.49	70 (60–79)[a]	0.53
Local resources	n.a.	n.a.	70 (60–79)[a]	0.51
Housing	66 (58–74)	0.48	68 (58–77)[a]	0.61
Inequality	73 (65–80)	0.48	68 (58–77)[a]	0.51
Employment	59 (50–67)	0.44	66 (55–75)[a]	0.56
Noise pollution	n.a.	n.a.	65 (54–74)	0.47
Safety	67 (59–75)	0.44	64 (54–73)[a]	0.54
Cultural capacity	n.a.	n.a.	59 (48–68)	0.50
Smart infrastructure	61 (52–69)	0.41	58 (48–68)	0.47
Biodiversity	n.a.	n.a.	56 (46–66)	0.46
Economic productivity	48 (40–57)	0.45	56 (45–66)	0.53
Urban microclimate	n.a.	n.a.	55 (44–65)	0.44
Business climate	47 (39–56)	0.43	38 (29–49)	0.55
International embeddedness	28 (21–36)	0.35	37 (27–47)	0.40
Entrepreneurship	35 (27–43)	0.42	36 (27–46)	0.45

[a] Components of which the percentage evaluations ≥8 is significantly greater than 50% and the SA is greater than 0.5.

Experts suggested 204 additional components relevant for defining and measuring urban sustainability. The content analysis yielded 43 different categories (including an ‘other’ category with unique suggestions), of which six contained more than four suggestions. For each of these six categories a component label and explanation was formulated, resulting in the following new components: biodiversity, cultural capacity, governance, local resources, noise pollution, urban microclimate (see Appendix A).

3.3. The second Delphi round

In the second round the percentage of evaluations equal to or greater than 8 ranged between 36% (entrepreneurship) and 89% (non-car transportation infrastructure, air quality, CO_2 emissions) (see Table 3). Nineteen components were identified as most relevant of which 18 also had a SA greater than 0.5. These included, in addition to the ten components that already had a SA greater than 0.5 in the previous round, the following eight components: governance, education, civic engagement, local resources, inequality, housing, employment, and safety.

Differences between experts from academia and experts from other institutions regarding the relevance of the components were no longer statistically significant. Significant differences were found between experts from North-West Europe and South-East Europe regarding the relevance of the components 'green spaces' ($U = 1322.5$, $p < .01$), 'education' ($U = 1249$, $p < .01$), and 'noise pollution' ($U = 1324.5$, $p < .01$). Experts from South-East Europe tended to evaluate these components as more relevant than experts from North-West Europe.

3.4. The third Delphi round

In the third round none of the components was selected by a majority of experts (see Table 4). As a result, the high values of the SA are mainly due to a majority of expert that did *not* consider the components to be most relevant.

The following seven components were selected significantly more often as most relevant than what may be expected based on chance alone: air quality, governance, energy consumption, non-car transportation infrastructure, green spaces, inequality, and CO_2 emissions. These components also received the most points. Although the component 'air quality' was selected more often as most relevant than the component 'governance', it received less points. Likewise, the component 'non-car transportation infrastructure' was selected more often than 'green spaces', but also received less points.

Table 4
Percentage of experts that selected each component as most relevant (95% Confidence Interval), the SA and total number of points that was attributed to each component in the third Delphi round.

Component	Round 3 (n = 72)		
	% selected (95% CI)	SA	Number of points
Air quality	46 (34–58)[a]	0.50	667
Governance	40 (29–53)[a]	0.51	676
Energy consumption	40 (29–53)[a]	0.51	585
Non-car transportation infrastructure	33 (23–45)[a]	0.55	462
CO_2 emissions	32 (21–44)[a]	0.56	470
Inequality	32 (21–44)[a]	0.56	452
Green spaces	32 (21–44)[a]	0.56	446
Health	25 (16–37)	0.62	391
Climate resilience	25 (16–37)	0.62	372
Solid waste	22 (13–34)	0.65	232
Civic engagement	18 (10–29)	0.70	247
Local resources	17 (9–27)	0.72	265
Biodiversity	17 (9–27)	0.72	260
Education	14 (7–24)	0.76	185
Employment	13 (6–22)	0.78	164
Water usage	13 (6–22)	0.78	128
Economic productivity	11 (5–21)	0.80	195
Smart infrastructure	11 (5–21)	0.80	160
Waste water treatment	11 (5–21)	0.80	136
Housing	10 (4–19)	0.82	115
Noise pollution	8 (3–17)[b]	0.85	119
Safety	7 (2–15)[b]	0.87	145
Urban microclimate	7 (2–15)[b]	0.87	113
Entrepreneurship	6 (2–14)[b]	0.89	85
Cultural capacity	6 (2–14)[b]	0.89	60
Business climate	4 (1–12)[b]	0.92	50
International embeddedness	1 (0–7)[b]	0.97	20

Note: Based on chance alone, each component would have been selected by 19% of experts.
[a] Components which were selected significantly more often than expected based on chance alone.
[b] Components which were selected significantly less often than expected based on chance alone.

Experts provided various explanations of why they selected a component as most relevant. They stressed that air quality is relevant for the health, well-being, and quality of life of inhabitants: "*Air quality is related to health, which is important for people's wellbeing.*" They also pointed out that air quality is related to many other components of urban sustainability, such as biodiversity, green spaces, and CO_2 emissions.

With regard to the relevance of the component 'governance', experts explained that city governments determine the extent to which cities develop in a sustainable way: *"Governance is a key item in developing sustainability. Without political support, plans and projects cannot reach results."* City governments can support a sustainable development among others by adapting legislation, providing resources, involving stakeholders, as well as by planning and managing city development.

Experts selected the component 'energy consumption' as most relevant, because energy is needed to sustain life and because the use of energy impacts the environment. Cities' energy consumption is related to various environmental issues, such as CO_2 emissions, climate change, and the use of natural resources. It also determines cities' self-sufficiency and dependency on non-renewable energy. Experts stressed that renewable energy consumption is crucial for a sustainable urban development: *"The degree of renewability of the overall energy consumption is a crucial element to achieve sustainability."*

Experts explained that a non-car transportation infrastructure is most relevant, because it reduces the use and negative impacts of cars and improves the sustainability of urban mobility: "A *sustainable city should be measured by its mobility infrastructure and in particular to non-car mobility which is the most environment friendly.*" A non-car transportation infrastructure also improves the environment in terms of among others pollution, CO_2 emissions, energy and resource consumption, noise pollution, congestion, quality of life, and space use.

Regarding the relevance of the component 'CO_2 emissions', experts explained that CO_2 emissions impact the environment as a whole and affect climate change: *"The level of industrial activity in urban Europe is very high. These industrial activities generate a lot of CO_2 emissions, which contribute in no small way to climate change and ultimately affect urban sustainability."* As emitters of high levels of CO_2, cities need to mitigate and fight climate change. Experts also explained that CO_2 emissions are related to other urban sustainability issues, such as energy consumption, transportation, as well as inhabitant's health and quality of life.

Experts stated that inequality, as a social aspect of sustainability, is a relevant component that is often neglected: *"Societies that aim to be environmentally sustainable should not forget the social aspects of sustainability, of which reducing inequality is the most important."* Inequality creates many social problems like social exclusion as well as tensions and conflicts between different groups of people. A sustainable, functioning, and inclusive city needs to provide for all its inhabitants, among others by sharing benefits and offering equal opportunities.

Lastly, experts explained that green spaces are most relevant, because they improve a city's general quality of life as well as the well-being of its inhabitants: "*Green spaces in fact improve the well-being of inhabitants.*" They provide many benefits to a city in terms of among others biodiversity, air quality, health, and urban microclimate. Green spaces are also important for recreation, relaxation, and as meeting places for people. Finally, green spaces support and raise people's awareness regarding environmental topics.

In addition to the most relevant components, Table 4 also shows the seven components that were selected significantly less often than what may be expected based on chance alone: noise pollution, safety, urban micro climate, entrepreneurship, cultural capacity, business climate, and international embeddedness. The corresponding high values of the SA indicate that experts generally agreed that these components are *not* the most relevant for defining and measuring urban sustainability. The components also received the least points, except for the component ‘safety’. This component was selected by few experts who gave it relatively many points.

With regard to the percentage of experts that selected each component as most relevant, no significant differences were found between experts from academia and experts from other institutions. Likewise, no significant differences were found between experts from North-West and South-East Europe.

4. Discussion

4.1. Study outcomes

In this study the Delphi method was used to identify which components experts find most relevant for defining and measuring the ranking attribute urban sustainability in a European context. The following seven components were identified as most relevant and may therefore be regarded as central for defining and measuring urban sustainability: air quality, governance, energy consumption, non-car transportation infrastructure, green spaces, inequality, and CO_2 emissions. Remarkably, none of these components was selected as most relevant by a majority of experts. This lack of agreement reflects the ambiguity surrounding the definition and measurement of urban sustainability, which has also been discussed in the literature (Ameen et al., 2015; Tanguay et al., 2010).

Five of the seven components that were identified as most relevant are part of the environmental dimension of urban sustainability (air quality, energy consumption, non-car transportation infrastructure, green spaces, and CO_2 emissions). Although sustainability was initially considered to be mainly an environmental concept (Ameen et al., 2015), researchers have stressed the importance of especially the social dimension and to some extent also the economic dimension of urban sustainability (Ameen et al., 2015; Dempsey et al., 2011; Hassan and Lee, 2015; Lorr, 2012; Michael et al., 2014). Nonetheless, in this Delphi study only the social component ‘equality’ and none of the economic components ended up among the most relevant components. These findings suggest that urban sustainability is still perceived as mainly an environmental concept.

The results of this Delphi study also suggest that most environmental policy sub-fields such as water pollution, waste management, and biodiversity have become somewhat less relevant in favour of other issues such as climate mitigation, renewable energy, transportation and green spaces. The only exception here is air pollution which is a classical field of environmental policy that is considered as the biggest environmental health risk and is thus still placed high on the political agenda (see for example the EU’s Air Quality Directive 2008/50/EC or major events such as the international conference on “Clean Air for European Cities” which took place in 2015 in Berlin). It appears that social issues, in particular those related to health and inequality, have become more relevant, though not yet as relevant as environmental issues. On the other hand, it seems that economic issues are still not considered as relevant as social or environmental issues. The component ‘governance’, which was included in the Delphi study based on experts’ suggestions, may be difficult to place under one of the three dimensions. Whereas some researchers placed it among the social components (Tanguay et al., 2010), others included it as a separate fourth dimension (Shen et al., 2011) or see the three dimensions as embedded in a framework of governance (Petschow et al., 2005).

4.2. Limitations

The first round of the Delphi study started with a list of components based on various existing urban sustainability rankings. Some of these rankings may be regarded as leaning heavily on the environmental dimension of urban sustainability. This could have resulted in bias in the final results if experts had not been invited to suggest additional components. Many experts accepted the invitation to make suggestions and based on these suggestions the list was complemented with six components.

A convenience sample of European urban sustainability experts was assembled. To reduce selection bias, selection criteria and different search strategies were applied. Obviously, this did not result in similarly-sized expert strata. The relatively large number of academics does not mean the results are biased or not representative per se; after all we do not know the composition of the total population of European urban sustainability experts. In the sample no lasting significant differences between expert groups could be identified. Nonetheless, one may ask whether or not similar results would have been obtained with a different sample composition or a different sampling strategy. As the study was explicitly restricted to a European context, caution is advised when using the results for defining and measuring urban sustainability in other parts of the world (see Science Communication Unit, 2015).

A low response rate and high drop-out rates are well-known limitations of the Delphi method (Hung et al., 2008; Keeney et al., 2006) that may also be a cause for concern in this study. In the first Delphi round 33% of the invited experts completed the questionnaire, which is comparable to the average response rate of ordinary web-surveys (34%) (Shih and Fan, 2008). Considering that experts were invited to participate in not one, but three questionnaire rounds, the response rate may be regarded as satisfactorily. Whether the drop-out rates in the second (29%) and third (27%) Delphi round are favourable or unfavourable in comparison to other Delphi studies is difficult to determine as many Delphi studies do not report them (Boulkedid et al., 2011).

4.3. Implications

The results of this study may serve to develop a conceptual framework in which urban sustainability is conceived as a multi-dimensional concept (i.e. environment, society, economy, and perhaps also governance), with each dimension consisting of several specific components that are measurable by means of indicators. Such a framework, also referred to as a domain-issue-based (Maclaren, 1996) or theme-oriented (Huang et al., 2015) framework, would then contain at least the seven components which were identified as most relevant.

Selecting indicators for each component remains a daunting task as a large pool of indicators is available (Huang et al., 2015). Systematic procedures and criteria for selecting a parsimonious list of urban sustainability indicators are available (see for example Maclaren, 1996; Tanguay et al., 2010). Indicators for some components could already be selected based on existing work (e.g. Tanguay et al., 2010), such as ‘the average annual concentration of PM_{10} particles in designated residential zones’ (for the component ‘air quality) and ‘the percentage of users of mass-transit versus other means of transport’ (for the component ‘non-car transportation infrastructure’). Alternatively, indicators may be selected by means of experts (Giovannini et al., 2008). Perhaps here lies another opportunity for using the Delphi method, for example to identify indicators for the component ‘governance’. It would also be interesting to replicate the current study in other parts of the world to find out which relevant components are place-dependent and which components are universally relevant for defining and measuring urban sustainability.

5. Conclusion

Meijering et al. (2014) suggested to use the Delphi method to define and operationalize urban environmental sustainability. This study showed that the Delphi method indeed seems to be useful for this purpose. Although high levels of agreement among experts were not always obtained, the method did prove useful for identifying the seven most relevant components for defining and measuring urban sustainability in a European context. These components are: air quality, governance, energy consumption, non-car transportation infrastructure, green spaces, inequality, and CO_2 emissions. Remarkably, five of the seven components are part of the environmental dimension of urban sustainability, which suggests that the experts perceived urban sustainability as mainly an environmental concept (and not so much a social or economic concept). Overall, this study provides the basis for an expert-based framework in which urban sustainability is operationalized into several weighted components. This framework may be used by future developers of urban sustainability rankings to properly define the concept and to select or develop appropriate indicators.

Acknowledgements

We would like to acknowledge the following experts who participated in all three rounds of this Delphi study and gave their valuable opinions: Adriana Galderisi, Agnes Franzen, Ahmed Khoja, Albert Edman, Ana Poças Ribeiro, Angeliki Chatzidimitriou, Anicenta Bubak, Anne-Francoise Marique, Anónio Ramos, Antonio Gagliano, Arie Voorburg, Aurore Cambien, Bastiaan Kees Zoeteman, Blanca Pedrola Vidal, Carlos Borrego, Carmelina Cosmi, Carolina Mateo Cecilia, Conny Weber, Corrado Diamantini, Diogo Alarcão, Dominic Stead, Dominique van Ratingen, Donát Rabb, Emma Terama, Enrico Pisoni, Eva Pangerl, Fabiana Morandi, Francesca Pagliara, Frans van der Reep, Gerald Krebs, Giulia Sonetti, Greg Keeffe, Haldun Süral, Hyunjung Lee, Ian Skinner, Irina Aura Istrate, Jan Dictus, Janine Hogendoorn, Jaques Teller, Jeroen Nagel, Kornilia Maria Kotoula, Kwasi Gyau Baffour Awuah, Lorenzo Chelleri, Luca Bertolini, Luca Coscieme, Lucjan Goczol, Marcin Czyz, Margaretha Breil, Maria Loloni, Mariano Gallo, Marina Rigillo, Marleen Lodder, Martine Laprise, Martyna Surma, Max Grünig, Monica Salvia, Niels van Geenhuizen, Paul Swagemakers, Peter Bröde, Riikka Holopainen, Romano Fistola, Rosa Anna La Rocca, Sophie Jongeneel, Timothy Lee, Tobias Emilsson, Walter Unterrainer, Wim de Haas, as well as five other experts who wished to remain anonymous.

We would also like to acknowledge Charlotte Walther and Theo Xenakis for the many hours they put in searching and contacting urban sustainability experts.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A

Overview of components evaluated by experts.

Components based on a content analysis of eight existing urban sustainability rankings

- *Air quality:* the extent to which the air in a city contains pollutants
- *Business climate:* the extent to which a city is suitable for doing business, for example in terms of taxes, regulations, and corruption
- *Civic engagement:* the extent to which a city's inhabitants are being engaged in urban policy and politics
- *Climate resilience:* the extent to which a city is resilient to the potentially harmful effects of climate change
- *CO_2 emissions:* the amount of a city's carbon dioxide emissions
- *Economic productivity:* the extent of a city's economic productivity
- *Education:* the education level of a city's inhabitants
- *Employment:* the extent to which the inhabitants of a city are economically (in)active
- *Energy consumption:* the amount of renewable and non-renewable energy that is consumed within a city
- *Entrepreneurship:* the amount of business start-ups within a city
- *Green spaces:* the amount of nature and parks within a city
- *Health:* the health of a city's inhabitants
- *Housing:* the cost and quality of housing in a city
- *Inequality:* the extent of differences between groups of inhabitants, for example in terms of income, access to education, and political participation
- *International embeddedness:* the extent to which a city participates in international networks and hosts international events
- *Non-car transportation infrastructure:* the size, quality, and use of a city's non-car transportation infrastructure (for trains, the metro, buses, cycling, walking)
- *Safety:* the amount of violent crimes within a city and the extent to which inhabitants feel safe
- *Smart infrastructure:* the extent to which technology is used within a city to improve among others government services, commuter traffic, and the energy efficiency of buildings
- *Solid waste:* the amount of solid waste that is produced and recycled within a city
- *Wastewater treatment:* the amount of wastewater that is collected and treated within a city
- *Water usage:* the amount of water that is used within a city as a result of water consumption and leakages in the water distribution system

New components based on a content analysis of experts' suggestions

- *Biodiversity:* the diversity of plant and animal species in a city
- *Cultural capacity:* the availability of cultural facilities and activities in a city
- *Governance:* the extent to which the organization and political composition of the city government supports a sustainable urban development
- *Local resources:* the extent to which a city produces and uses local resources such as food and energy
- *Noise pollution:* the extent to which inhabitants of a city are exposed to bothersome noise
- *Urban microclimate:* the extent to which the climate in a city is comfortable for its inhabitants, for example in terms of temperature and humidity

References

Ameen, R.F.M., Mourshed, M., Li, H., 2015. A critical review of environmental assessment tools for sustainable urban design. Environ. Impact Assess. Rev. 55, 110–125.

Arcadis, 2015. Sustainable Cities Index. Retrieved from: http://www.sustainablecitiesindex.com/.

Besecke, A., Herkommer, B., 2007. Schönste Stadt – erfolgreichste Stadt – lebendigste Stadt, Sinn und Unsinn von Städterankings. Universitätsverlag der Technischen Universität Berlin, Berlin.

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., Alberti, C., 2011. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 6 (6), e20476. http://dx.doi.org/10.1371/journal.pone.0020476.

Charron, N., Dijkstra, L., Lapuente, V., 2015. Mapping the regional divide in Europe: a measure for assessing quality of government in 206 European regions. Social Indic. Res. 122 (2), 315–346.

Dalkey, N., Helmer, O., 1963. An experimental application of the Delphi method to the use of experts. Manage. Sci. 9 (3), 458–467. http://dx.doi.org/10.1287/mnsc.9.3.458.

De Jong, M., Joss, S., Schraven, D., Zhan, C., Weijnen, M., 2015. Sustainable-smart-resilient-low carbon-eco-knowledge cities; making sense of a multitude of concepts promoting sustainable urbanization. J. Clean. Prod. 109, 25–38. http://dx.doi.org/10.1016/j.jclepro.2015.02.004.

Dempsey, N., Bramley, G., Power, S., Brown, C., 2011. The social dimension of sustainable development: defining urban social sustainability. Sustainable Dev. 19 (5), 289–300.

Diamond, I.R., Grant, R.C., Feldman, B.M., Pencharz, P.B., Ling, S.C., Moore, A.M., Wales, P.W., 2014. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J. Clin. Epidemiol. 67 (4), 401–409. http://dx.doi.org/10.1016/j.jclinepi.2013.12.002.

Dijkstra, L., Poelman, H., 2014. A Harmonized Definition of Cities and Rural Areas: The New Degree of Urbanization (Regional working paper 01/2014, produced for the European Commission DG Regional and Urban Policy). Retrieved from: http://ec.europa.eu/regional_policy/sources/docgener/work/2014_01_new_urban.pdf.

European Commission, 2010. Making Our Cities Attractive and Sustainable. How The EU Contributes to Improving The Urban Environment. Publications Office of the European Union, Luxembourg.

European Commission, 2011. Cities of Tomorrow – Challenges, Visions, Ways Forward. Publications Office of the European Union, Luxembourg.

European Commission, 2015. Will Your City be the European Green Capital in 2018? Publications Office of the European Union, Luxembourg.

Ferguson, N.D., Davis, A.M., Slutsky, A.S., Stewart, T.E., 2005. Development of a clinical definition for acute respiratory distress syndrome using the Delphi technique. J. Crit. Care 20 (2), 147–154. http://dx.doi.org/10.1016/j.jcrc.2005.03.001.

Floridi, M., Pagni, S., Falorni, S., Luzzati, T., 2011. An exercise in composite indicators construction: assessing the sustainability of Italian regions. Ecol. Econ. 70 (8), 1440–1447. http://dx.doi.org/10.1016/j.ecolecon.2011.03.003.

Fu, Y., Zhang, X., 2017. Trajectory of urban sustainability concepts: a 35-year bibliometric analysis. Cities 60 (Part A), 113–123. http://dx.doi.org/10.1016/j.cities.2016.08.003.

Giovannini, E., Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. Organisation for Economic Cooperation and Development (OECD), Paris.

Gray, D.E., 2004. Doing Research in The Real World. Sage Publications, London.

Hassan, A.M., Lee, H., 2015. The paradox of the sustainable city: definitions and examples. Environ. Dev. Sustain. 17 (6), 1267–1285.

Hasson, F., Keeney, S., 2011. Enhancing rigour in the Delphi technique research. [The Delphi technique: Past, present, and future prospects]. Technol. Forecast. Social Change 78 (9), 1695–1704.

Huang, L., Wu, J., Yan, L., 2015. Defining and measuring urban sustainability: a review of indicators. Landscape Ecol. 30 (7), 1175–1193.

Huang, S.-L., Wong, J.-H., Chen, T.-C., 1998. A framework of indicator system for measuring Taipei's urban sustainability. Landscape Urban Plann. 42 (1), 15–27.

Hung, H.L., Altschuld, J.W., Lee, Y.F., 2008. Methodological and conceptual issues confronting a cross-country Delphi study of educational program evaluation. Eval. Program Plann. 31 (2), 191–198. http://dx.doi.org/10.1016/j.evalprogplan.2008.02.005.

Hussler, C., Muller, P., Rondé, P., 2011. Is diversity in Delphi panelist groups useful? Evidence from a French forecasting exercise on the future of nuclear energy. [The Delphi technique: past, present, and future prospects]. Technol. Forecast. Social Change 78 (9), 1642–1653.

Jacobs, R., Goddard, M., Smith, P.C., 2005. How robust are hospital ranks based on composite performance measures? Med. Care 43 (12), 1177–1184.

Keeney, S., Hasson, F., McKenna, H., 2006. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. J. Adv. Nursing 53 (2), 205–212. http://dx.doi.org/10.1111/j.1365-2648.2006.03716.x.

Lindgren, B.W., 1976. Statistical Theory. Collier Macmillan Publishers, London.

Linstone, H.A., Turoff, M., 1975. The Delphi Method: Techniques and Applications (Vol. 29). Addison-Wesley Reading, MA.

Lohuis, A.M., van Vuuren, M., Bohlmeijer, E., 2013. Context-specific definitions of organizational concepts: defining 'team effectiveness' with use of the Delphi technique. J. Manage. Organ. 19 (6), 706–720. http://dx.doi.org/10.1017/jmo.2014.10.

Lorr, M.J., 2012. Defining urban sustainability in the context of North American cities. Nat. Cult. 7 (1), 16–30.

Lun, G., Holzer, D., Tappeiner, G., Tappeiner, U., 2006. The stability of rankings derived from composite indicators: analysis of the “IL Sole 24 Ore” quality of life report. Social Indic. Res. 77 (2), 307–331. http://dx.doi.org/10.1007/s11205-005-4505-z.

Maclaren, V.W., 1996. Urban sustainability reporting. J. Am. Plann. Assoc. 62 (2), 184–202. http://dx.doi.org/10.1080/01944369608975684.

McManus, P., 2012. Measuring urban sustainability: the potential and pitfalls of city rankings. Aust. Geogr. 43 (4), 411–424.

Meijering, J.V., Kampen, J.K., Tobi, H., 2013. Quantifying the development of agreement among experts in Delphi studies. Technol. Forecast. Social Change 80, 1607–1614. http://dx.doi.org/10.1016/j.techfore.2013.01.003.

Meijering, J.V., Kern, K., Tobi, H., 2014. Identifying the methodological characteristics of European green city rankings. Ecol. Indic. 43, 132–142.

Meijering, J.V., Tobi, H., van den Brink, A., Morris, F., Bruns, D., 2015. Exploring research priorities in landscape architecture: An international Delphi study. Landscape Urban Plann. 137, 85–94.

Michael, F.L., Noor, Z.Z., Figueroa, M.J., 2014. Review of urban sustainability indicators assessment – case study between Asian countries. Habit. Int. 44, 491–500.

Newcombe, R.G., 1998. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat. Med. 17 (8), 857–872. http://dx.doi.org/10.1002/(sici)1097-0258(19980430)17:8<857::aid-sim777>3.0.co;2-e.

Newman, P., Kenworthy, J., 1999. Sustainability and Cities: Overcoming Automobile Dependence. Island press, Washington, D.C.

Petschow, U., Rosenau, J.N., Weizsäcker, E.U., 2005. Governance and Sustainability: New Challenges for States, Companies and Civil Society. Greenleaf Publications.

Powell, C., 2003. The Delphi technique: myths and realities. J. Adv. Nursing 41 (4), 376–382. http://dx.doi.org/10.1046/j.1365-2648.2003.02537.x.

Qualtrics [computer software], 2015. Retrieved from: http://www.qualtrics.com.

Robinson, D., Bylund, J., Coutard, O., Finnveden, G., Hooimeijer, P., Kabisch, S.J.R., 2015. Transition Towards Sustainable and Liveable Urban Futures: The Strategic Research and Innovation Agenda of JPI Urban Europe. JPI Urban Europe.

Science Communication Unit, 2015. Science for Environment Policy: Indicators for Sustainable Cities (In-depth Report 12, Produced for the European Commission DG Environment). Retrieved from: http://ec.europa.eu/environment/integration/research/newsalert/pdf/indicators_for_sustainable_cities_IR12_en.pdf.

Shen, L.-Y., Jorge Ochoa, J., Shah, M.N., Zhang, X., 2011. The application of urban sustainability indicators – a comparison between various practices. Habit. Int. 35 (1), 17–29. http://dx.doi.org/10.1016/j.habitatint.2010.03.006.

Shih, T.-H., Fan, X., 2008. Comparing response rates from web and mail surveys: a meta-analysis. Field Methods 20 (3), 249–271. http://dx.doi.org/10.1177/1525822x08317085.

Siemens, 2009. European Green City Index. Siemens AG, München.

Tanguay, G.A., Rajaonson, J., Lefebvre, J.F., Lanoie, P., 2010. Measuring the sustainability of cities: an analysis of the use of local indicators. Ecol. Indic. 10 (2), 407–418. http://dx.doi.org/10.1016/j.ecolind.2009.07.013.

UN-Habitat, 2016. World Cities Report 2016: Urbanization and Development – Emerging Futures. Retrieved from: http://wcr.unhabitat.org/wp-content/uploads/2017/02/WCR-2016-Full-Report.pdf.

Willis, G.B., 2005. Cognitive Interviewing. A Tool for Improving Questionnaire Design. Sage Publications, Thousand Oaks.

World Commission on Environment and Development, 1987. Our Common Future (Brundtland Report). Oxford University Press, Oxford.