



A comparison of preference elicitation methods for multi-criteria design decisions about resilient and sustainable buildings

Sahar Mirzaee¹ · David Fannon^{2,3} · Matthias Ruth⁴

Published online: 1 April 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Designing sustainable and resilient buildings is a complex process involving multiple actors and numerous, often conflicting, criteria to describe the social, environmental, and economic consequences of design decisions. Multi-Criteria Decision Making (MCDM) is a well-established branch of systematic decision science that may be used to narrow the choice set and balance decision criteria and objectives. MCDM methods rely on eliciting decision makers' (DM) criteria preferences through surveys—a critical, but often time-consuming and expensive process. This study compares two survey methods—Simple Multi-Attribute Rating Technique and Potentially All Pairwise RanKings of all possible Alternatives—used to capture DM's preferences for subjective criteria relevant to the design of resilient and sustainable buildings. Significant differences were found between the results of the two methods, indicating that building design criteria are evaluated differently in different contexts. However, most DMs reported that both surveys accurately captured their preferences, complicating conclusions about DM's actual preferences through single surveys. The design professionals surveyed also report lower cognitive load for the direct weighting method. Although this study does not report a preferred survey method, the findings suggest that great care must be taken when eliciting DM preferences, and when attempting to apply MCDM to building design decisions.

Keywords Multi-Criteria Decision Making (MCDM) · SMART · PAPRIKA · Decision-making cognitive load · Criteria preference techniques

1 Introduction

Multi-Criteria Decision Making (MCDM) is a branch of science that is particularly useful in environments where conflicting criteria and objectives impose immense cognitive burdens on decision makers (DMs). Decision making through MCDM breaks complex decisions into six steps: defining the problem; articulating the objective(s); identifying stakeholders; defining criteria to measure objectives; weighting the criteria; and ranking the decision alternatives

based on the objectives, criteria, and weights derived in previous steps (Triantaphyllou 2000). Unlike optimization approaches, MCDM works from a set of explicitly known feasible solutions. Furthermore, optimization demands a mathematical relationship between the criteria and the objective, which is almost impossible for many qualitative criteria (Gandibleux 2006). The design of resilient and sustainable buildings (RSB) can present conflicting—often qualitative—criteria and objectives, and generally proceeds through the iterative refinement of feasible solutions, suggesting it might be a good candidate for MCDM and motivating the present study. MCDM relies on experts' input to define and weight criteria, and the way these questions are asked can dramatically affect both the responses and the cognitive load on the DM as they balance the project-specific requirements within a context of hazards and environmental priorities. To support decision making, this paper compares two survey methods frequently used to establish weightings in the process of MCDM. The paper begins with a brief description of MCDM in Sect. 1, followed by discussion of sustainability, resilience, and decision making in the context of building

✉ David Fannon
d.fannon@neu.edu

¹ School of Public Policy and Urban Affairs, Northeastern University, Boston, MA, USA

² School of Architecture, Northeastern University, Boston, MA, USA

³ Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA

⁴ Department of Economics, University of Alberta, Edmonton, AB, Canada

design in Sect. 2. Section 3 introduces the method used in this study, and Sects. 4 and 5 present the results and conclusions, respectively.

2 MCDM background

This paper assumes that the first three steps of an MCDM—problem definition, articulation of objective(s), and identification of stakeholders—have been carried out already. This does not mean that these are trivial tasks with little bearing on the MCDM process and results. Rather, they may be considered pre-conditions for the analytical portion of MCDM and their influence on outcomes have been discussed in detail elsewhere (e.g., Wang and Lee 2009). The following two sections discuss the fourth step—criteria selection—and especially the fifth step of criteria weighting in detail.

2.1 Criteria selection

While there are many well-established methods for criteria identification, in all cases, criteria must exhibit four qualities (Phillips et al. 2017), namely be

- coherent with the overall objective
- measurable in some way
- presented on a common scale
- independent from one another.

In some instances, coherent criteria are readily available and easily measurable. For example, the criteria for a basic energy-retrofit project might include first cost and monthly energy use, each distinct from the other and both quantifiable using available tools and data. However, in many cases selecting or measuring the appropriate criteria presents a significant challenge. Subjective aspects like aesthetics or occupant comfort are not only difficult to measure, but also challenging to present on a common scale, and criteria independence is important to ensure that only the necessary criteria are included. Avoiding duplication reduces the complexity of the problem and facilitates analysis.

The present study selected criteria with the Delphi method. First used in 1948, the Delphi method aims to improve the use of expert predictions in policy-making (Woudenberg 1991) through interactions among a group of experts to agree on decision criteria (Martino et al. 1976). The process begins with a dialogue about criteria that most affect an objective of the problem, followed by an iterative commenting procedure; these repeated cycles of discussion and revision allow experts to update their reflections and arrive at group consensus. Delphi enables knowledge discovery and perspective sharing among experts with different disciplinary backgrounds (Pill 1971), and so the process

works well for complex problems that include considerable uncertainty and require multiple perspectives. However, the authority of researchers over the problem definition may cause them to select participants with a certain viewpoint, or to manage group dynamics, potentially introducing bias (Linstone and Turoff 2011). The lack of participant anonymity can sway individuals to change their positions, exacerbating these challenges. Nevertheless, the Delphi method affords results at a reasonable time and cost (William and Webb 1994) even for problems demanding judgment and subjective expertise; when causal relationships are hard to validate; or when the problem needs to consider diverse opinions within a large group of people (Powell 2003; Yang et al. 2012), a good fit for the present study.

2.2 Criteria weighting

After selecting the proper criteria of the problem, their relative importance or weights must be established as inputs for the alternative-ranking process. Criteria weight or value depends on the problem and context; for example, designing RSB in a desert might prioritize water conservation, while projects with scarce funding might value economic cost above other considerations. Weights also reflect different stakeholders' values. Criteria weights directly affect the final ranking, so assigning accurate values to criteria is one of the most crucial steps in MCDM. There are two general approaches to criteria weighting: subjective and objective. Objective methods use mathematical calculation to arrive at the weights, whereas subjective methods are based on DMs' opinions (Tzeng et al. 1998).

Objective methods are based on established mathematical procedures to assess or reconcile differences in the importance of individual criteria. They may require significant computational capacity, limiting the ability of DMs to revisit their judgments in a dynamic decision-making environment. Even in very technical cases, applications of objective methods are troublesome, as decisions nearly always involve diverse concerns (Maggino and Ruviglioni 2009). Objective methods tend to be more time-consuming, more expensive, and harder to execute than subjective approaches (Kamal 2012).

Subjective methods are based on DMs' making trade-offs among the criteria based on a subjective sense of relative importance, and the weights are easily adjusted if the objective of the problem changes. Although they demand little time or computational capacity, subjective methods incorporate human judgment—and the attendant imprecision and ambiguity—and must carefully attend to internal consistency (Önüt et al. 2009), to deliver accessible methods for criteria weighting with multifarious concerns. The idea of human judgment seems to align with the application of professional judgement made by architects and engineers

working in the built environment to address the interests of multiple stakeholders, for example client budgets and the health safety and welfare of the public.

2.3 Preference elicitation methods

Subjective weighting usually employs surveys to elicit DMs' preferences. Preference in this context means the relative importance of each criterion to the decision maker. MCDM does not recommend one method to establish criteria weighting over others, and similar decision-making problems employ different decision-making approaches. For example, a 2002 study to evaluate renewable energy strategies obtained criteria preferences directly (Polatidis and Haralambopoulos 2002), while another applied pairwise-comparison criteria weighting to a very similar problem in which the objective was to rank different renewable energy projects (San Cristobal 2011); neither study comments on the selection of method. This paper evaluates two subjective criteria preference elicitation methods—Simple Multi-Attribute Rating Technique (SMART) and Potentially All Pairwise RanKings of all possible Alternatives (PAPRIKA), to test their possible applicability to RSB design problems.

SMART is a straightforward weighting method that uses simple additive weight (SAW) technique to obtain weights for individual criteria (Edwards 1977; Edward and Barron 1994). After identifying the decision criteria, DMs assign the highest score to the most important criterion and score the rest of the criteria relative to that. Results are then normalized to unity to obtain criteria weights. This method is simple to administer and easy for DMs to complete, but in its abstraction does not have an analog to contingent and sometimes messy building design process.

PAPRIKA is a pairwise method that presents DMs with a series of hypothetical scenarios in which all but two criteria are held equal. Those two criteria have different values in the two scenarios, and the DM is required to assess the tradeoffs and select the more desirable scenario. Each time a DM makes a choice, PAPRIKA assigns points to the preferred criterion and updates the previous point assignments to determine which comparison to present next. It then continues the process until the final weights are calculated. PAPRIKA surpasses conventional pairwise methods which merely ask which of two criteria is more important and cannot include multiple levels for each criterion or scenario-based tradeoffs. Developed in 2004, PAPRIKA is relatively, appearing in academic literature from 2009 onward (Noseworthy et al. 2009). This study offers the first comparative analysis of PAPRIKA to other survey methods in the context of a real-world problem.

Donald Schön in his book *The Reflective Practitioner* suggests that when confronted with complex, unique, and uncertain situations, professionals engage in “reflective

conversations” that seek to fit a framework of generic knowledge and prior experience to the problem at hand. By repeatedly responding to the “back-talks” of this fitting exercise, professionals refine these frameworks, until the unique and uncertain situation comes to be understood (Schön 1983). Schön cites design decisions made by architects and engineers as examples of reflective practice, and the similarities with the repeated scenario comparisons and tradeoffs suggest PAPRIKA may be an authentic model for eliciting DMs' behavior in this domain.

2.4 Cognitive burden on DMs

Subjective MCDM can impose a heavy cognitive load on the DMs (Narasiman and Vickery 1988). Cognitive load theory postulates that the cognitive capacity of people's working memory is limited, and exceeding this threshold hampers performance—perhaps because of the presence of complexities and uncertainties (De Jong and Den Hartog 2010). As DMs' roles vary in different MCDM methods, it is logical to assume the cognitive load on DMs varies as well. Traditional MCDM methods require the DMs to understand and value each criterion relative to other criteria or make tradeoffs among them, while considering the importance of all criteria simultaneously. An acute awareness of the cognitive burden on DMs when evaluating the criteria motivates this study, with the hypothesis that a series of scenarios, although individually complex, may impose a smaller burden and better reflect the process of making design decisions and tradeoffs for RSB.

DMs are usually considered reliable sources of information in a complex environment, but because of their limited ability to comprehend all the complexities involved, psychological pressure on DMs can adversely affect MCDM (Larichev and Nikiforov 1987). Measuring the cognitive load on the DMs while engaging in the MCDM is difficult, and DMs may not have a coherent view about it (Pomerol and Barba-Romero 2012). A handful of studies have focused on the difficulty of surveying methods for DMs as well as accuracy of capturing their preference weights. Buchanan and Daellenbach (1987) used different survey techniques to collect preference weights, and also asked decision makers about the difficulty. In the context of cost minimization and labor force utilization in the production of electrical components of lamps, Buchanan and Daellenbach concluded that pairwise comparisons pose no problem for DMs. Similarly, Narasimhan and Vickery (1988) recorded preference weights for three criteria using SMART and pairwise comparisons and found no statistical difference in responses to the two methods. In contrast, in a study of production scheduling task, Wallenius (1975) found that users prefer unstructured methods to more sophisticated approaches. More recently, Aloysius et al. (2006) administered two surveys to 153

individuals, one using a direct technique similar to SMART, and the other one using pairwise comparisons. They focused on the participants' reflection on cognitive load and reported that users believe that pairwise comparisons cause more decisional conflicts, are less accurate, more effortful, and less desirable (Aloysius et al. 2006). However, the results in these fairly simple scenarios may not hold true in complex settings requiring more criteria and criteria levels, such as integrating sustainability and resilience in buildings, and the pairwise methods used were less sophisticated than PAPIKA. Whatever the influence of the survey instrument on cognitive load, it will of course also vary by DM based on knowledge and experience (Pomerol and Barba-Romero 2012).

2.5 Ranking methods

The last step of an MCDM brings together criteria weights and values to rank the various alternatives. The various methods for alternative rankings lie outside the scope of this paper, which refers the interested reader to other studies for more detail on the following approaches: Analytical Hierarchy Process, or AHP (e.g., Saaty 1980); Preference Ranking Organization Method for Enrichment Evaluation, or PROMETHEE (e.g., Brans 1986); The Elimination and Choice Translating Reality, or ELECTRE (e.g., Roy and Bouyssou 1986); and Technique for Order Preference by Similarity to Ideal Solutions, or TOPSIS (e.g., Yoon 1987). Instead of enumerating already well-documented similarities and differences among these methods, this study wishes to highlight the relevance that cognitive burden in decision making has on the process of eliciting criteria preferences.

3 Sustainability, resilience, and decision making in buildings

MCDM has been applied in many fields to improve decision making in complex environments, including building design (e.g., Hsieh et al. 2004; Flager et al. 2009; Balcomb and Curtner 2000), and is a promising method to integrate resilience and sustainability in building design decisions. While concern for the environmental, social, and economic performance of buildings, generally summed up under the label of *sustainability*, has become widespread in research and practice (e.g., Lützkendorf and Lorenz 2006; Zuo and Zhao 2014) with many scholars defining the term in detail (e.g., Pater and Cristea 2016; Galvic and Lukman 2007), the notion of building *resilience* is only just capturing the attention of decision makers (Laboy and Fannon 2016; Burroughs 2017). As with sustainability, a wide range of definitions of resilience exist, many of them stressing the need for systems to cope with disruption, maintain essential operations,

return to normal operations after the disruption has ended, and elevate to a state of advanced performance following some exogenous shock (Alexander 2013; Linkov et al. 2014; Mirzaee et al. 2018). In the context of buildings, Alexander (2006) argued that while building codes can regulate the design, construction, and maintenance of structures to protect occupants from disasters, the protection measures have not kept pace with the growing vulnerability of places with high risk to disasters. Common to both resilience and sustainability is the notion that there is no single criterion by which they can be evaluated, and that any application of these concepts for decision making must be cognizant of the boundary constraints—environmental, social, economic, and mental—within which alternative actions are compared to each other and judged.

Some prior work has applied MCDM to questions of building sustainability, including energy management (Kolokotsa et al. 2009) and selecting wall insulation to retrofit existing buildings (Ruzgys et al. 2014). Jin et al. used pairwise comparisons to capture DMs' preferences on ten sustainability criteria and then ranked ten green building technologies for existing buildings (Jin et al. 2016). Cegan et al. (2017) performed a broad literature review on the applications of MCDM in Environmental sciences and sustainability domain and found a rapid growth in the use of MCDM in such problems from 2000 to 2015. MCDM has recently been used to improve sustainability (Invidiata et al. 2018), and to find the best combination of design strategies to construct net zero energy buildings (Harkouss et al. 2018). To date, applications of MCDM to the burgeoning field of building resilience mostly focus on specific hazards and design responses, for example, identifying the best alternative for a structural seismic retrofit, whether concrete and masonry (Formisano and Mazzolani 2015), or steel-braced frames (Deierlein et al. 2011). These applications advance what has been called “engineering resilience,” which assumes an equilibrium model, and prioritizes the rapid return to function (Gunderson 2000; Holling 1996; Scheffer et al. 1993). While not wrong, engineering resilience tends towards narrow *optimization* (albeit with multiple objectives) rather than the broad incorporation of relevant stakeholders, objectives, and criteria characteristic of MCDM.

Although it can yield ranked lists of alternatives, MCDM does not employ an objective function nor provide a single, optimal solution for three reasons: (i) the decision alternatives are fuzzy and tend to change during the decision process; (ii) the decision maker is not a single person or entity, and members of a decision-making group may choose conflicting or even contradictory criteria; and (iii) the criteria preference weights elicited from DMs may be imprecise or badly formulated (Majumder 2015; Mota et al. 2013; Carlsson and Korhonen 1986; Zimmermann 1990). Real-world decisions about building design often exhibit all three

reasons. Fuzziness of the decision environment means the goals and/or decision criteria are fuzzy in nature and their boundaries are not sharply defined, although this does not necessarily mean that the system itself is fuzzy (Belman and Zadeh 1970). Although building for sustainability and resilience relies on objective performance data, the complex tradeoffs of balancing sustainability and resilience in buildings depend on human-in-the-loop interactions. Building projects inherently involve diverse stakeholders, sometimes with conflicting goals. Finally, eliciting accurate criteria preference weights from those DMs is challenging. The limited prior literature and our initial trials suggest that simple, direct ranking methods do not adequately capture the complexities of the decision space from diverse stakeholders. However, more nuanced comparative methods risk over-specification, while demanding considerable time and effort to develop and administer. Robust decision outcomes must address all three of these challenges, by reducing the fuzziness of the decision environment, identifying critical stakeholders, and eliciting accurate criteria preferences.

4 Method

This study grows out of an effort to develop an MCDM-based tool to support design decisions about sustainable and resilient buildings, a process that encompasses multiple criteria evaluated by multiple stakeholders. The present study compares methods to elicit preference weights, seeking reasonable accuracy and acceptable cognitive load. These results do not provide criteria values to increase resilience and sustainability of any *particular* project, but rather help design teams identify methods to collect such values. The focus on *design* decisions reduces the fuzziness of the decision space and the diversity of stakeholders, focusing only on decision makers who are design professionals from the Architecture, Engineering, and Construction (AEC) industry. Of course, other stakeholders (e.g., owners, occupants) may weigh the same criteria differently. As detailed below, preference weights from 42 North-American AEC-industry design professionals regarding ten resilience and sustainability criteria were collected using two different survey methods.

4.1 Criteria selection

The absence of a common definition for concepts like building resilience complicates the selection of decision-making criteria: indeed, finding a suitable definition of *resilience* may itself constitute a decision problem. Although there are many possible metrics for both sustainability and resilience, prior work identified commonalities and conflicts among resilience and sustainability standards and rating

systems (Philips et al. 2017). Building on this literature, the present study used the Delphi method with discussion and knowledge sharing among a panel of subject-matter experts to identify important criteria for resilience and sustainability in buildings. The panel experts included leading researchers and practitioners with expertise in structural engineering, architecture, environmental chemistry, LCA, high-performance buildings, urban resilience, and social sciences. This step of our study has been particularly important because people with diverse backgrounds often use the term *resilience* to mean different things so clearly establishing an appropriate criterion, if not a definition, is essential in a climate of increasing uncertainty (Laboy and Fannon 2016). The prompt, coupled with the particular expertise of this panel, led to identification of criteria in Table 1, which incorporated resilience criteria into the commonly used *Triple Bottom Line* (TBL) (Elkington 2013) categories, as well as adding a new “recovery” bottom line, resulting in a *Quadruple Bottom Line* (QBT). The criteria reflect the panel’s collective definition of resilience and sustainability, and acknowledge those criteria will be valued higher, lower, or not at all depending on stakeholder, context, and project. For example, criteria 9 and 10 show a bias towards an engineering definition of resilience; however, economic and socio-technical definitions of resilience emerge in criteria 3, 6, and 8. The criteria selection step is of great importance to the MCDM process, as a different set of criteria can shift the decision outcome. In this case, a constructive dialogue yielded ten criteria that cover the four categories of QBT as shown in Table 1.











4.2 Survey design

Two surveys were designed to quantify DM’s preference weights among the ten resilience and sustainability criteria. The first—using the SMART method—asked DMs directly value the criteria relative to each other. The second used the PAPRIKA method, in which DMs select their preference through a series of nearly identical scenarios pairs. Both surveys were developed and administered using online tools: Qualtrics Survey platform (Qualtrics, Provo, UT) for the SMART survey, 1000 Minds for PAPRIKA (Hansen and Ombler 2008).

Qualtrics is an online tool for data collection that supports multiple question types. As shown in Fig. 1, the SMART survey was administered using slider bars with scales from 0 to 100. The prompt asked decision makers to assign 100 points to the most important criterion and value the remaining nine criteria relative to that maximum.

1000minds.com is an online platform for PAPRIKA that elicits preferences from multiple actors by forcing them to choose between pairs of alternative solutions that slightly differ on specific criteria. The survey designer identifies

Table 1 Ten criteria for sustainability and resilience organized into categories and subcategories

Category	Sub-category	No.	Criterion	Criterion Type
Quadruple Bottom Line Resilient Triple Bottom Line	Economic	1	First costs	
		2	Operating and maintenance costs	
		3	Recovery cost	
	Environmental	4	Life cycle impact	
		5	Energy conservation	
	Social	6	Beyond-Code Safety	
		7	Occupant comfort and satisfaction	
		8	Impact on the neighboring community	
	Recovery	9	Functional portion of the building after a disaster	
		10	Time to recover after a disaster	

Key:  Economic Criterion |  Environmental Criterion |  Social Criterion |  Resilience Criterion

the quantitative and/or qualitative criteria of the problem, and at least two performance levels for each criterion. During the survey administration, the software automatically develops the necessary pairwise scenario comparisons needed to capture DMs' preferences based on the total decision space, i.e., the responses to each question govern which question is presented next.

Prior studies in the realm of social, behavioral, and psychological sciences evaluated the effect on responses when the same question is worded differently, and generally conclude that the way the researchers ask questions matters for the results (Moser and Kalton 2017; Walgrave et al. 2016). The primary requirement is *clarity*, such that the question has only one interpretation (Lee et al. 2016; Magelssen et al. 2016). To promote clarity in the PAPRIKA method, each criterion was associated with a uniquely phrased question and performance levels. This phrasing, shown in Table 2, avoided the redundancy and confusion from repeated performance levels such as “high, medium, low.”

With the ten criteria included in this study and at least two performance levels for each, comparing every criteria and level would require hundreds of pairwise comparisons. However, 1000 minds determines which pairwise comparisons are necessary to arrive at the final weights, thus optimizing to the fewest questions. This survey typically asked DMs approximately 40 questions. An example question is presented in Fig. 2.

4.3 An opinion-based survey

Complementing the first two surveys about criteria preference weights, participants were asked to complete a simple meta-survey about the experience based on a series of questions assessed using 5-point Likert-type scale. Likert scaling is a bipolar scaling system measuring positive and negative responses usually with neutral response in the middle. The survey is focused on four main dimensions presented in Table 3.

As shown in Table 3, self-reported cognitive load is used to compare the difficulty of the two methods. Web-based surveys often use self-assessed ratings of cognitive load to indicate participants' perception of cognitive load, not least because they are dramatically easier to collect than other measures (Paas et al. 2003). While self-reported values always demand careful consideration, prior studies have found participants can quantify their mental load (Gopher and Braune 1984), and that self-reported cognitive load is meaningfully correlated with actual cognitive load (Gimino 2002; Paas et al. 1994). In another example, Lin et al. (2013) asked participants to evaluate a task difficulty, and the participants were found to have over 74% accuracy on the self-reported cognitive load. Many studies successfully use self-reported mental effort on a Likert-type scale as a measure of cognitive load (e.g., Chang and Yang 2010; Kalyuga and Sweller 2005; Scharfenberg and Bogner 2013;

Please score the following criteria based on their importance to you when making choices about the building type you selected earlier. Assign **100 points to the most important**, and value the remaining criteria **relative** to that.

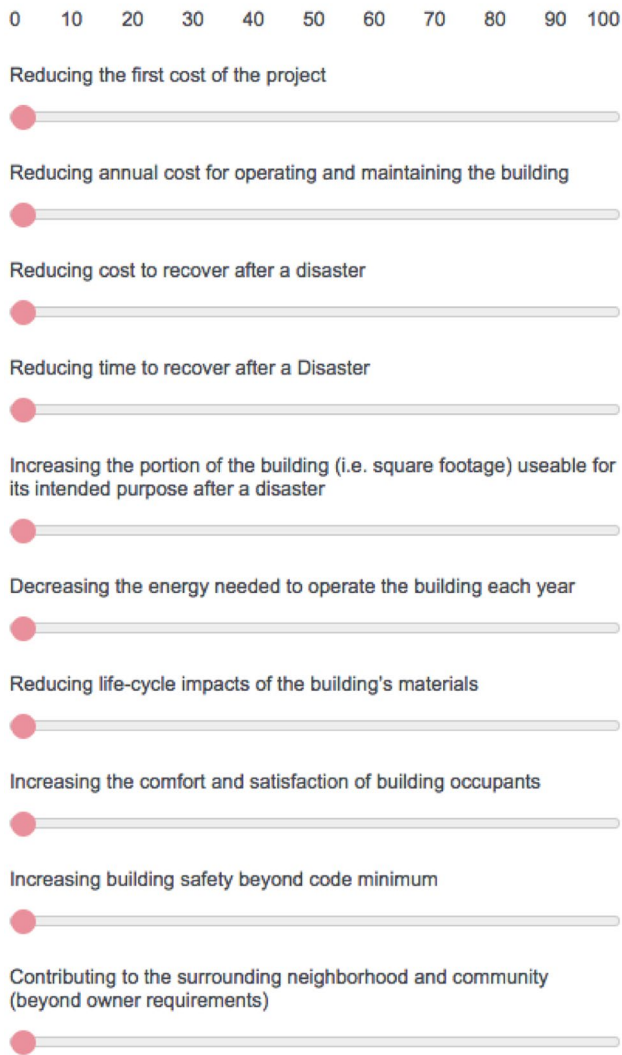


Fig. 1 SMART survey in Qualtrics environment showing the 0–100 scale and sliders

Antonenko and Neiderhauser 2010). Collectively, this body of prior work results indicates that self-assessment is a reliable and valid measure of the cognitive load (Chen et al. 2016; Ayres 2006).

4.4 Survey administration

As a pilot test, the surveys were administered to 10 researchers working in the field of building design and who were familiar with the principles of MCDM as well as concepts of sustainability and resilience. Results were reviewed with each respondent to solicit feedback about the survey design and survey method, as well as the language of questions. The

responses were used only to refine the survey prior to wider deployment and are not included in the analysis presented below.

After pilot testing, a sample of 250 adults over 18 years of age and with a professional role in the building industry was identified. There was no selection criterion related to gender, ethnicity, or socio-economic level, as these personal data were irrelevant to the research. The survey solicitation consisted of an email with a link to an intake-landing page, which provided information about the relevant human subject protocols and collected basic demographic data about industry role, experience level, and educational attainment. As part of the intake, DMs were prompted to consider a recent project as a method to frame the otherwise-abstract decision weighting process, lending some specificity as the cost of potentially greater variance. The landing page then linked to the SMART survey, at the completion of which users were presented the PAPRIKA survey, and finally presented the self-reporting meta-survey. The nature of the scenario-comparison questions precluded randomizing the otherwise-preferable randomization of the sequence of the two surveys, potentially introducing order-effects.

The surveys were distributed to 250 professionals in the AEC industry and 59 responded to both surveys, a 23.6% response rate. Of those respondents, 42 (72%) identified their role as design professionals (61% architects or designers, 11% engineers), the target population for this study about design decision making. The remaining respondents indicated roles as owners and developers, product manufacturing, education or other related industries, and were not included in the analysis below. The 42 design professionals averaged 12 years of experience, ranging from 1 year to 37 years.

5 Results and analysis

The following sections present two types of results: first are DM’s preference weights for the building sustainability and resilience criteria, which reveal the priorities of this sample of AEC design professionals. Recalling that the purpose of this study is to compare the two survey methods used to elicit such preferences, the second group of results analyzes the differences in weights between the two methods and the DM’s subjective responses about them.

5.1 Comparison of mean criteria weights for SMART and PAPRIKA surveys

Surveys of decision makers to establish criteria weights for MCDM often use either the average or median of responses, and for small samples with possible outliers, the median better describes central tendency. Therefore, this study

Table 2 Table of criteria, questions, and performance levels used in 1000 minds survey. Note the unique phrasing of the performance levels associated with each criterion

No.	Criteria	Question	Performance levels
1	First cost	First costs to design and construct the building are...	Higher than typical About typical Lower than typical
2	Operating and maintenance costs	My operating and maintenance costs...	I spend extra About normal I save some money
3	Recovery cost	Recovering after a disaster...	Costs a lot Is not too expensive
4	Material sustainability	Materials in the building are...	Unsustainable Typically sustainable Very sustainable
5	Occupant satisfaction and comfort	Occupants are...	Slightly uncomfortable or dissatisfied Generally comfortable and satisfied Very comfortable and highly satisfied
6	Energy conservation	Normal operation of the building...	Demands extra energy Demands about typical amount of energy Demands less energy than a typical building
7	Recovery time	After a disaster it takes...	A long time to recover to previous conditions A short time to recover to previous conditions
8	Functional portion of the building	Immediately after a disaster most of the building is...	Safe to evacuate, but not able to be used for its intended purpose Safe and fully functional for its intended purpose
9	Life safety	Life safety design...	Meets code requirements Exceeds code requirements
10	Contribution to the surrounding neighborhoods	Contribution to the surrounding neighborhoods and communities...	Satisfies owner's needs Satisfies the owner's needs while contributing to the surrounding neighborhood and community (e.g., providing public space, or serving as a disaster shelter)

Fig. 2 An example of a scenario-comparison question generated by 1000minds.com based on the criteria and performance levels

Imagine a decision about a building. Assuming everything else is equal, which of these two outcomes would you prefer?
(all else being equal)

First costs to design and construct the building are
About typical

Occupants are
Generally comfortable and satisfied

this one

this combination is impossible

OR

First costs to design and construct the building are
Lower than typical

Occupants are
Slightly uncomfortable or dissatisfied

this one

this combination is impossible

they are equal

« undo last decision

skip this question for now »

Table 3 Table of opinion-based survey questions and ranges of response options

Question	From (value 1)	To (value 5)
Ease of navigating through each survey	Not challenging at all	Extremely challenging
Satisfaction about how the questions reflected real-life tradeoffs	Not very realistically	Very realistically
The effort that was needed to complete each survey	Very easy	Very difficult
Confidence on the reflection of their actual preferences in each method	Not accurately at all	Very accurately

compares the median weights to test for significant differences between the two surveys, as summarized in Table 4. Six out of ten criteria show significant differences in preference between the two survey methods: first cost, operational cost, percent functional of the building, energy conservation, life cycle impact, and beyond-code safety. The remaining four criteria—recovery cost, recovery time, occupant comfort, and neighborhood impact—do not show statistically significant differences between survey methods. Energy and life cycle are valued low when they are asked about directly. However, DMs trade other criteria off for these two, valuing them more highly in scenario comparisons compared to direct questions. Beyond-code-safety and first cost are highly valued in direct questions; however, respondents value them less in scenario comparisons, and trade them away for other criteria. In general, when asked directly, respondents place high value on economic/financial criteria. However, when presented with the forced choice of scenarios, respondents trade-off profit for greater environmental sustainability and increased resilience. Neighborhood impact and occupant comfort, which are categorized as social criteria, appear highly consistent across respondents, while beyond-code safety, which is in the same category, is valued quite differently.

The percentage change describes the increase or decrease in median preference values as captured through PAPRIKA compared to the SMART median value as baseline.

Beyond-code safety and first cost showed the largest difference between survey methods, with changes of 6.2 points and 4.7 points, respectively, from the median collected with SMART to the median collected with PAPRIKA. The greatest change on a percentage basis was also beyond-code safety: its median value of preference in PAPRIKA is 81% less than its median value in SMART. The next largest change was for first cost, which showed a difference of 57% between the two methods.

Because the sample size is the same in all the cases, it is possible to compare the *t*- and *p*-values with each other. *T*-values are an indication of the difference in means in the two groups and measure the size of the difference relative to the variance of the data. In Table 4, the three largest absolute *t*-values correspond to the three lowest *p*-values, eliminating the possibility that these significant differences are based on chance. These three criteria are, first cost, beyond-code safety, and energy conservation.

The differences in medians of the weights for each criterion as determined by the two surveys are illustrated in Fig. 3. The median of the criteria weight responses in PAPRIKA was subtracted from the median value of the SMART survey, so values greater than zero indicate survey subjects gave greater preference to those criteria when assigning values directly in the SMART than they evinced through their choices in scenarios. Conversely, negative values indicate criteria that DMs preferred when choosing

Table 4 Median preference values observed using each survey, their differences, and measures of statistical significance

Criteria	Median weights (normalized)				Statistical tests of difference		
	SMART	PAPRIKA	Difference	% Change	<i>t</i> value	<i>p</i> value	df
–							
First costs	0.082	0.035	0.047	– 57	4.97	4.61·10 ^{–6}	69.105
Operation and maintenance costs	0.124	0.083	0.041	– 33	3.82	2.62·10 ^{–4}	79.137
Recovery costs	0.074	0.081	– 0.007	1	1.95·10 ^{–3}	9.98·10 ^{–1}	78.484
Life cycle impacts	0.107	0.141	– 0.034	32	– 4.49	2.90·10 ^{–5}	67.082
Energy conservation	0.133	0.156	– 0.023	17	– 4.02	1.38·10 ^{–4}	72.881
Beyond-code safety	0.076	0.014	0.062	– 81	5.23	1.61·10 ^{–6}	71.746
Occupant comfort and satisfaction	0.129	0.139	– 0.01	7	– 1.06	2.94·10 ^{–1}	64.196
Impact on the neighboring community	0.093	0.083	0.010	– 11	1.07	2.75·10 ^{–1}	79.981
Portion of the building functional after a disaster	0.079	0.094	– 0.015	19	– 2.11	3.80·10 ^{–2}	83.205
Time to recover after a disaster	0.098	0.096	0.002	2	– 0.70	4.84·10 ^{–1}	74.064

Shaded criteria show statistically significant differences

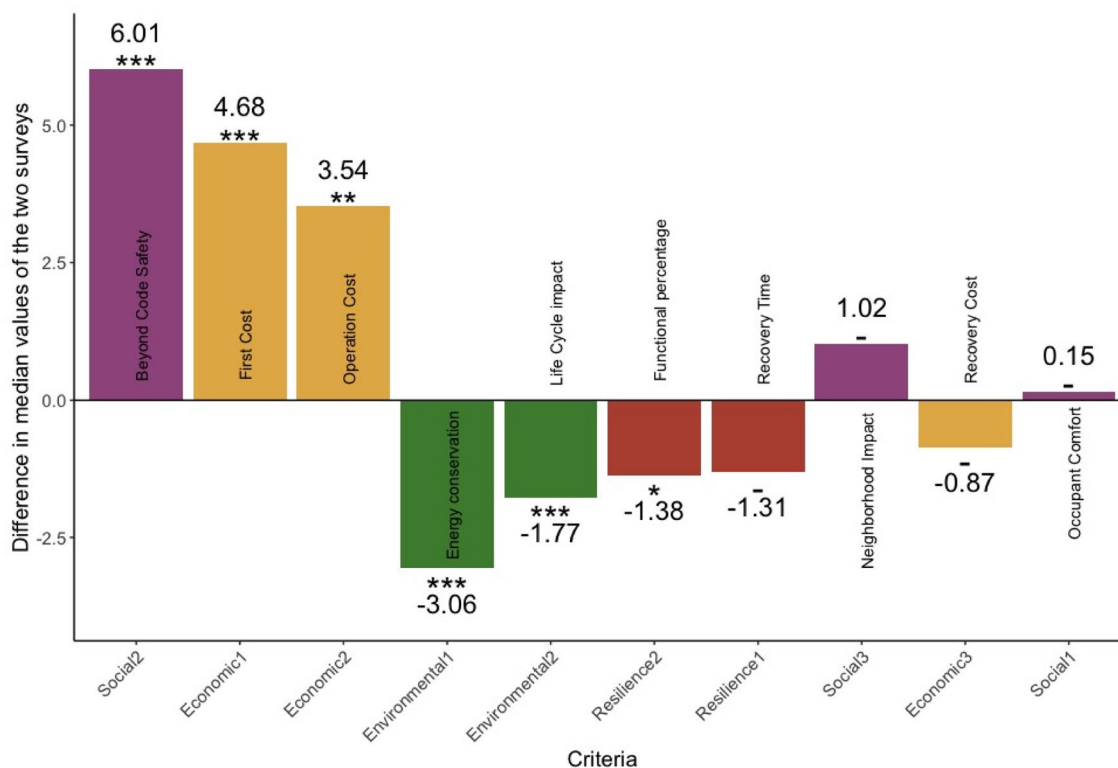


Fig. 3 Differences in median weight for each criterion from SMART and PAPRIKA surveys. Differences that are not statistically significant are marked with (–), while *, **, *** indicate significance at the

$p=0.5$, 0.001, and 0.0001 thresholds, respectively. The colored version of this figure indicates category groupings

between scenarios but had directly assigned lower scores. The significance levels of the differences are shown on each bar using star marks and the category of each criterion is both color-coded and named on each bar. The category-based color-coding reveals suggestive trends, namely that economic criteria seem to be valued more highly in direct assessment, while environmental and resilience criteria appear to be valued more highly when DMs are making tradeoffs in scenarios. These trends suggest that a user's valuation not only varies depending on the method of survey, but also that those changes in weighting may be non-uniform across various criteria, and perhaps criteria groups.

5.2 Comparison of preference value distributions for each criterion

As described above, respondents as a group valued criteria differently when asked to directly assign a score vs. when selecting those criteria in a scenario. To better understand the responses from individuals that result in these group patterns, each subject's scenario responses are plotted as a function of their direct weighting for each criterion. These plots are shown in Fig. 4. Each dot represents the opinion of one individual, with the weight in response to direct questions

on the horizontal axis and the weight determined through scenario-comparison choices on the vertical axis. A perfectly consistent respondent would have the same value on both axes.

Beyond revealing the consistency (or lack thereof) for individual subjects across the two surveys, these distributions also suggest criteria in which respondents are generally consistent, or for which a small number of dramatic changes overwhelm a generally consistent set of responses. To aid in this comparison, black lines represent a linear regression of the distribution, and the gray-toned area defines the 95% confidence interval. The white lines illustrate an idealized distribution in which all the DMs are perfectly consistent, providing the same preference about each criterion in both surveys. Although the graphs visually suggest some criteria are more consistent than others, the low R-squared values in all cases indicate that there is no strong trend in responses to compare to the ideal. Graphs in which most data points are concentrated above or below the white line also highlight those criteria—identified in the previous section—which generally score highly on one survey method but not the other, for example beyond-code life safety.

Respondents highly value first cost and beyond-code safety when they are asked about them directly but trade

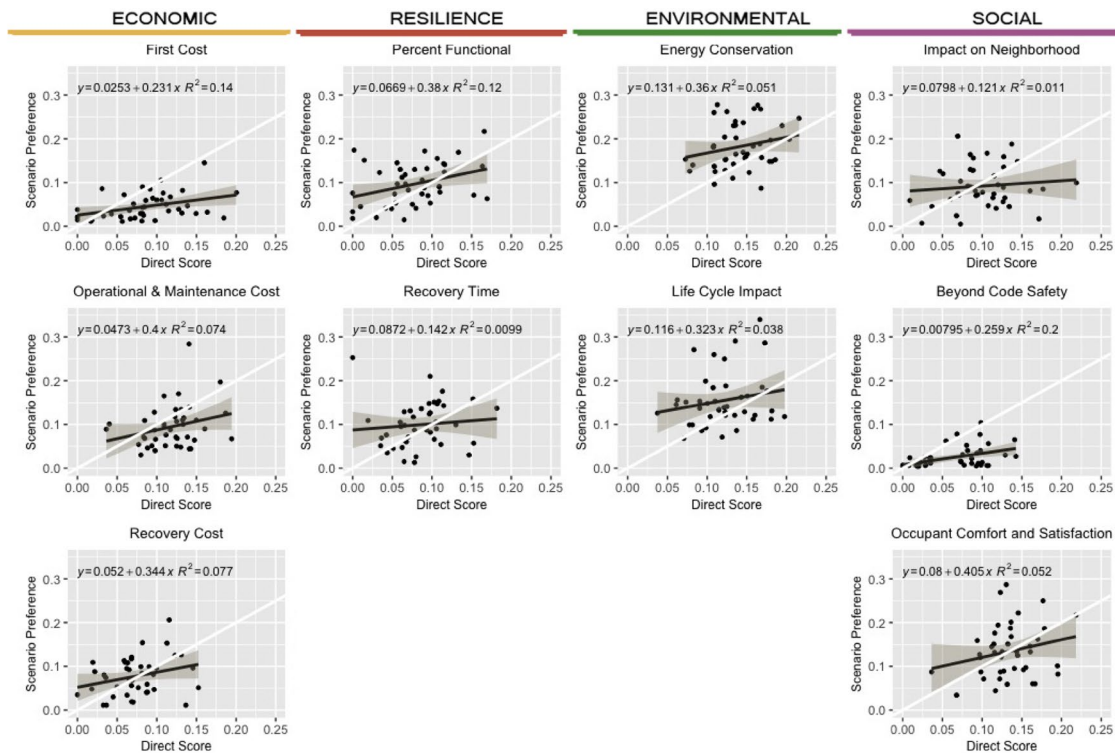


Fig. 4 Distributions of individual responses for each criterion, showing the scenario (PAPRIKA) weightings as a function of the direct (SMART) valuations. Perfectly consistent respondents would score each criterion identically in both surveys, yielding the hypothetical

trendline with a slope of 1 and intercept of 0, as shown in white. A best-fit linear trend for the actual responses is shown in black, with the shaded zone indicating the 95% confidence interval. The universally low R^2 values indicate poor fit and low consistency

them away for other criteria when presented in scenario comparisons. The logic holds true for operating cost, although not as intensely. The converse behavior is also present; respondents report less concern about energy conservation when they are asked to value the sustainability criteria directly. However, the same respondents often choose the more sustainable option in scenarios, trading away other criteria to improve energy conservation. For other criteria, the data are quite obscure and responses inconsistent; none of the linear regressions representing the response data for other criteria fit perfectly, and there are many outliers where the same criterion is weighted very differently in the two surveys.

Figure 5 shows box plots of the distribution of responses to all ten criteria in both SMART and PAPRIKA surveys. These plots illustrate how central trends used for MCDM may not reflect the complexity of preference weights across two surveys. For example, the mean value for Life Cycle assessment in the scenario comparisons aligns with the fourth quartile of the direct measurements, because the long tail of high values in PAPRIKA indicate a handful respondents are highly valuing this criterion through that survey method. This is visible in Fig. 4 as well, because while most results fall close to the trend line, a few responses well above

the line pull the linear fit away from the idealized curve, perhaps because direct rating tends to encourage central tendency bias. In another cases, such as beyond-code safety, responses are skewed towards lower values when evaluated in scenarios. Overall, the results show that DMs do not consistently weigh the criteria in the two surveys.

5.3 Comparison of self-reported accuracy and cognitive load

In addition to comparing the criteria-weighting methods based on responses to the two surveys, the third meta-survey collected respondents’ subjective experience of the two methods. Their responses to the four questions about the questions are illustrated in Fig. 6.

Figure 6a shows that the DMs find navigating through scenario-comparison questions harder than the direct weighting. Figure 6b shows that DMs find the direct questions more realistic compared to scenario comparisons. Figure 6c illustrates that the DMs exert more effort to answer the scenario-comparison questions. Finally, Fig. 6d shows that, in DMs’ opinion, both methods capture their preferences equally well.

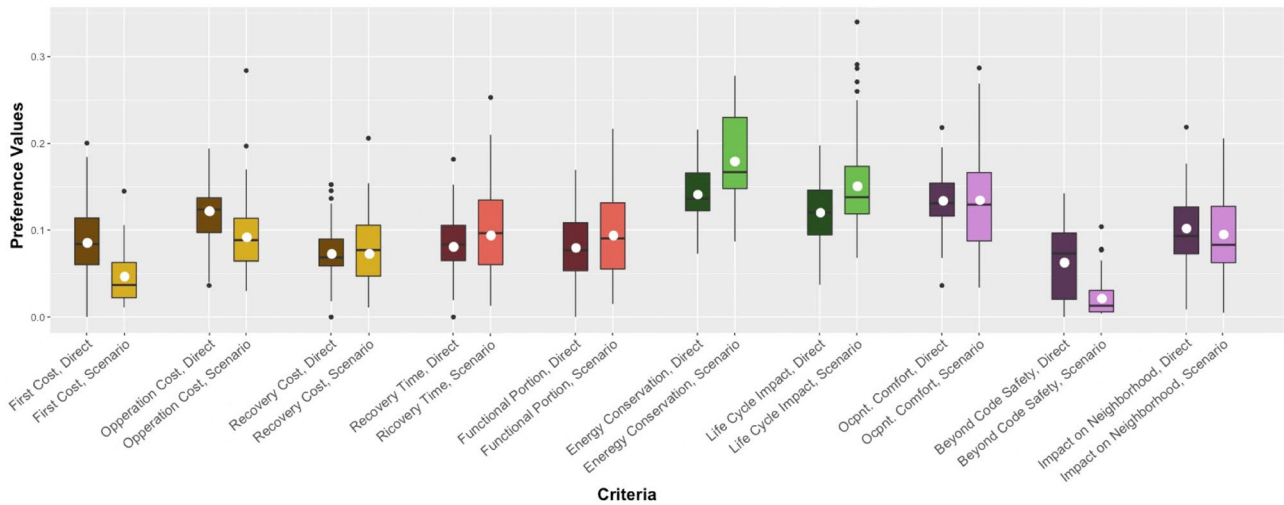


Fig. 5 Box plots of the distributions of responses to all ten criteria in both SMART (marked here as direct) and PAPIKA (marked here as Scenarios). The boxes show the first and third quartiles as well as the median value as a horizontal line and mean values with a white

circle. Points beyond the whisker represent outliers beyond 1.5 times the interquartile range above the upper quartile and below the lower quartile

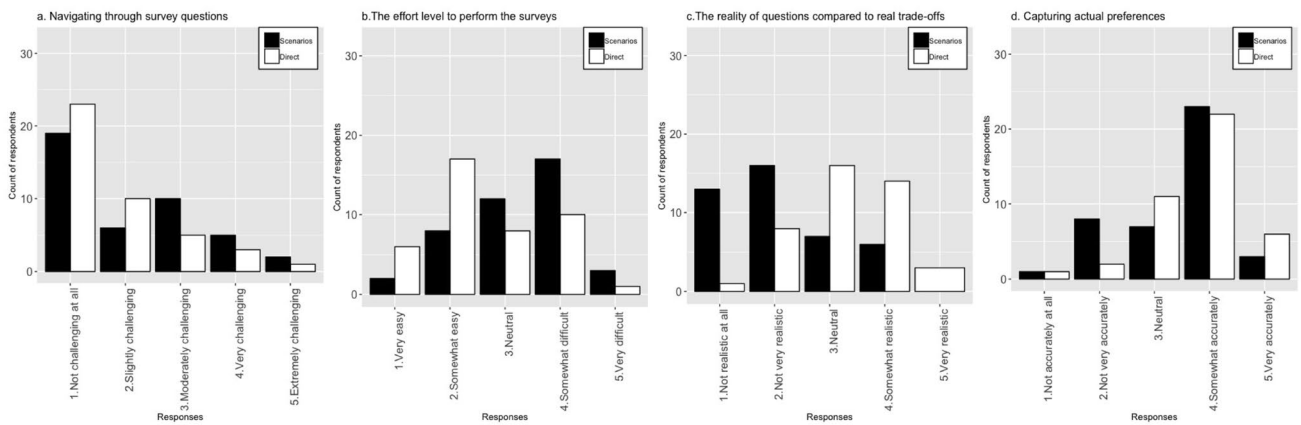


Fig. 6 Respondents’ self-reported assessment of the survey methods. From the left, graph **a** shows the difficulty of navigating through scenario questions versus navigating through direct questions. Graph **b** shows the responses to the level of effort needed to answer questions

in scenarios versus direct questions. Graph **c** shows the responses to how realistic the tradeoffs are in scenarios versus direct questions. Graph **d** shows the DM’s opinion on how accurately the two methods reflect their preference weights

Table 5 The mean and median value of the DMs’ self-reported assessment of survey methods

Question	Mean values		
	Direct	Scenarios	Difference
Navigation difficulty	1.79	2.17	−0.38
Level of effort	2.60	3.26	−0.66
Reflection of realistic tradeoffs	3.24	2.15	1.09
Accurately capturing preferences	3.71	3.45	0.56

Mean values were calculated for each Likert response, and are shown in Table 5. The numeric results suggest that, on average, DMs believe the direct questions better reflect the tradeoffs present in real projects, demand less effort, are easier to navigate, and more accurately capture their preferences. These findings are consistent with Aloysius et al. in that DMs find direct measurement of criteria less confusing, less difficult, and more accurate (Aloysius et al. 2006). These findings also align with Kotteman and Davis (1998),

in that DMs prefer less-sophisticated weighting methods. However, in the context of building design and construction, the assessment is quite surprising given that decisions about building design seldom ask for numerical or even ranked preferences across multiple criteria simultaneously. Instead, DMs must evaluate complex, multifarious alternatives, which would seem to be better simulated using the PAPRIKIA method. The high scores for accuracy across both methods are particularly surprising given the low consistency in responses by most DMs discussed in the previous two sections.

6 Conclusions

This study evaluated DM's preferences among decision criteria as revealed through two MCDM survey techniques and compared the two survey methods. Forty-two design professionals who participated in the project indicated their preferences for ten resilience and sustainability criteria using both the direct scoring SMART technique, and the scenario-comparison-based PAPRIKA method. For both individuals and the group, criteria weights varied significantly depending on the method used to elicit preference, and these differences would inevitably affect the final ranking of alternatives in MCDM. Respondents also assessed the effort, cognitive load, and estimated the accuracy of the two methods, reporting both as accurate but preferring the direct method. Overall the results indicate substantial challenges in selecting a method to elicit criteria preferences for MCDM in problems of resilience and sustainability in buildings.

The method of elicitation affects the evaluation of building design criteria. Economic criteria consistently scored highly in direct ranking, suggesting a primacy of financial considerations when viewed simultaneously with all other criteria. However, confronted with choices between two alternative scenarios, DM's choose greater sustainability and resilience even at increased economic cost, resulting in higher values for environmental and social criteria. It is impossible to discern motivations from the bare data, but one possible explanation for these results is that the direct ranking reflects design professionals' perception of their clients' priorities, while the indirect results from scenario choices reveal unconscious personal preferences or professional obligations to the public, albeit at some cognitive load. Alternatively, the scenarios may represent the psychological response to an urgent choice, while the SMART ranking represents an abstracted, top-down plan.

It is not clear whether or how the differences in preferences observed in these data affect the actual design process, complicating any application of MCDM to building design. Applying MCDM to real decisions requires understanding the decision process in that context. Although designed to

emulate the kind of decisions made in practice, the scenario comparisons remain abstract hypothetical models of real decisions. Even still, the scenario-comparison survey presents situations that direct questions simply cannot, reducing the cognitive load but also affecting the results. While difficult, additional research to determine criteria preferences empirically from the design of real buildings would validate the relationship between preference values and design decision making in the complex contingency and complexity of practice.

Regardless of method or context, the validity of criteria weightings is difficult to determine, as there is no inherently true or actual value. The difference in self-assessed accuracy between the two methods is very small, indicating that respondents are confident that both surveys reflect their actual preferences. This result contrasts sharply with significantly different results for the median value of six of ten preference values captured through the two surveys and the low reliability of individual scores. These findings challenge the DMs' confidence in their responses, and complicate efforts to collect meaningful criteria weights for use in MCDM.

Subjective criteria collection requires humans in-the-loop, rather than humans-as-inputs. To avoid influencing the meta-survey, study subjects were not provided the weights calculated by PAPRIKA. We hypothesize that the lack of feedback and increased survey time frustrated some respondents, because the nature of this study precluded individuals from reflecting on any inconsistencies in their answers. Further, the findings show that in complex decision environments with more than a handful of criteria, cognitive load may affect DMs' responses. An iterative procedure could address these limitations, but simply presenting the calculated preference criteria for respondents to modify would effectively replicate the direct survey. We propose presenting final ranked alternatives to the DMs and allowing them to *indirectly* evaluate their preferences based on insight into the *outcome* of their preference values. This could be coupled with a collaborative or benchmarking approach might improve the quality of preference weights collected.

The sample in this study is neither sufficiently large nor representative to support conclusions about design professionals' preferences as a population. Future work to increase the sample size and diversity would enable such findings, perhaps also identifying inter-population differences, for example based on role. Further analysis of intra-personal differences and testing for personal consistency would also yield promising insights in the development and application of decision support tools. Testing these tools in practice on specific projects would offer unique and situated insight into the validity, reliability, and utility of these preference elicitation methods in the context of the design of RSB. Applying of MCDM to the design of buildings requires further study

and careful design to support decisions that increase the resilience and sustainability of the built environment.

Acknowledgements This research is based on work supported by the National Science Foundation grant number CMMI #1455450. The authors wish to thank Dr. Mehrdad Sasani, Dr. Matthew Eckelman, Matthew Joyner, Robert Phillips, and Luke Troup for their iterative and constructive feedback.

References

- Alexander D (2006) Globalization of disaster: trends, problems and dilemmas. *J Int Aff* 59(2):1–22
- Alexander DE (2013) Resilience and disaster risk reduction: an etymological journey. *Nat Hazard Earth Syst Sci* 13(11):2707–2716
- Aloysius JA, Davis FD, Wilson DD, Taylor A, Ross, Kottemann JE (2006) User acceptance of multi-criteria decision support systems: the impact of preference elicitation techniques. *Eur J Oper Res* 169(1):273–285
- Antonenko PD, Niederhauser DS (2010) The influence of leads on cognitive load and learning in a hypertext environment. *Comput Hum Behav* 26(2):140–150
- Ayres P (2006) Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learn Instr* 16(5):389–400
- Balcomb JD, Curtner A (2000) Multi-criteria decision-making process for buildings. In: Energy conversion engineering conference and exhibit, 2000. (IECEC) 35th Intersociety (Vol 1, pp 528–535). IEEE
- Brans J-P (1986) L'élaboration d'instruments d'aide à la décision. Nadeau, Raymond et Maurice Landry, 183–213
- Buchanan JT, Daellenbach HG (1987) A comparative evaluation of interactive solution methods for multiple objective decision models. *Eur J Oper Res* 29(3):353–359
- Burroughs S (2017) Development of a tool for assessing commercial building resilience. *Proc Eng* 180:1034–1043
- Carlsson C, Korhonen P (1986) A parametric approach to fuzzy linear programming. *Fuzzy Set Syst* 20(1):17–30
- Cegan JC, Filion AM, Keisler JM, Linkov I (2017) Trends and applications of multi-criteria decision analysis in environmental sciences: literature review. *Environ Syst Decis* 37(2):123–133
- Chang C-C, Yang F-Y (2010) Exploring the cognitive loads of high-school students as they learn concepts in web-based environments. *Comput Educ* 55(2):673–680
- Chen F, Zhou J, Wang Y, Yu K, Arshad SZ, Khawaji A, Conway D (2016) Robust multimodal cognitive load measurement. Berlin: Springer, Ahmad
- De Jong J, Den Hartog D (2010) Measuring innovative work behaviour. *Creat Innov Manag* 19(1):23–36
- Deierlein G, Krawinkler H, Ma X, Eatherton M, Hajjar J, Takeuchi T et al (2011) Earthquake resilient steel braced frames with controlled rocking and energy dissipating fuses. *Steel Constr* 4(3):171–175
- Edwards W (1977) How to use multiattribute utility measurement for social decisionmaking. *IEEE Trans Syst Man Cybern* 7(5):326–340
- Edwards W, Barron FH (1994) SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. *Organ Behav Hum Decis Process* 60(3):306–325
- Elkington J (2013) Enter the triple bottom line. In: The triple bottom line. Routledge, London, pp 23–38
- Flager F, Welle B, Bansal P, Soremekun G, Haymaker J (2009) Multidisciplinary process integration and design optimization of a classroom building. *J Inform Technol Constr* 14(38):595–612
- Formisano A, Mazzolani FM (2015) On the selection by MCDM methods of the optimal system for seismic retrofitting and vertical addition of existing buildings. *Comput Struct* 159:1–13
- Gandibleux X (ed) (2006) Multiple criteria optimization: state of the art annotated bibliographic surveys, vol 52. Springer, Berlin
- Gimino AE (2002) Factors that influence students' investment of mental effort in academic tasks: a validation and exploratory study
- Glavič P, Lukman R (2007) Review of sustainability terms and their definitions. *J Clean Prod* 15(18):1875–1885
- Gopher D, Braune R (1984) On the psychophysics of workload: why bother with subjective measures? *Hum Factors* 26(5):519–532
- Gunderson LH (2000) Ecological resilience: in theory and application. *Annu Rev Ecol Syst* 31(1):425–439
- Hansen P, Ombler F (2008) Decision support system and method: Google Patents
- Harkouss F, Biwole PH, Fardoun F (2018) Passive designs and renewable energy systems optimization of a net zero energy building in Embrun/France. Paper presented at the IOP conference series: earth and environmental science
- Holling CS (1996) Engineering resilience versus ecological resilience. *Eng Ecol Constraints* 31(1996):32
- Hsieh T-Y, Lu S-T, Tzeng G-H (2004) Fuzzy MCDM approach for planning and design tenders selection in public office buildings. *Int J Project Manag* 22(7):573–584
- Invidiata A, Lavagna M, Ghisi E (2018) Selecting design strategies using multi-criteria decision making to improve the sustainability of buildings. *Build Environ* 139:58–68
- Jin Z, Eikevik TM, Nekså P, Hafner A (2016) Investigation on CO2 hybrid ground-coupled heat pumping system under warm climate. *Int J Refrig* 62:145–152
- Kalyuga S, Sweller J (2005) Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educ Technol Res Dev* 53(3):83–93
- Kamal MM (2012) Shared services: lessons from private sector for public sector domain. *J Enterp Inform Manag* 25(5):431–440
- Kolokotsa D, Pouliezos A, Stavrakakis G, Lazos C (2009) Predictive control techniques for energy and indoor environmental quality management in buildings. *Build Environ* 44(9):1850–1863
- Laboy M, Fannon D (2016) Resilience theory and praxis: a critical framework for architecture. *Enquiry* 13(1):39–52
- Larichev OI, Nikiforov AD (1987) Analytical survey of procedures for solving multicriteria mathematical programming problems (MMPP) toward interactive and intelligent decision support systems. Springer, Berlin. pp 95–104
- Lee CD, Goldman SR, Levine S, Magliano J (2016) Epistemic cognition in literary reasoning. *Handbook of epistemic cognition*, 165–183
- Lin FR, Yaffe K, Xia J, Xue QL, Harris TB, Purchase-Helzner E, Health ABC Study Group F et al (2013) Hearing loss and cognitive decline in older adults. *JAMA Intern Med* 173(4), 293–299
- Linkov I, Bridges T, Creutzig F, Decker J, Fox-Lent C, Kröger W, Nathwani J et al (2014) Changing the resilience paradigm. *Nat Clim Chang* 4(6):407
- Linstone HA, Turoff M (2011) Delphi: a brief look backward and forward. *Technol Forecast Soc Chang* 78(9):1712–1719
- Lützkendorf T, Lorenz DP (2006) Using an integrated performance approach in building assessment tools. *Build Res Inform* 34(4):334–356
- Magelssen M, Supphellen M, Nortvedt P, Materstvedt LJ (2016) Attitudes towards assisted dying are influenced by question wording and order: a survey experiment. *BMC Med Ethics* 17(1):24
- Maggino F, Ruviglioni E (2009) Obtaining weights: from objective to subjective approaches in view of more participative methods in the construction of composite indicators. In: Proceedings NTS: new techniques and technologies for statistics, pp 37–46

- Majumder M (2015) Multi criteria decision making. In: Impact of urbanization on water shortage in face of climatic aberrations. Springer, Berlin, pp 35–47
- Martino JP, Harold A, Turoff M (1976) The Delphi method: techniques and applications: Linstone, Harold A., and Murray Turoff. Advanced Book Program. Addison-Wesley, Advanced Book Program, Reading
- Mirzaee S, Ruth M, Fannon D (2018) Reconciling diverse perspective of decision makers on resilience and sustainability. In: Advanced introduction to ecological economics. Edward Elgar, Cheltenham
- Moser CA, Kalton G (2017) Survey methods in social investigation. Routledge, London
- Mota P, Campos AR, Neves-Silva R (2013) First look at MCDM: choosing a decision method. *Adv Smart Syst Res* 3(2):25–30
- Narasimhan R, Vickery SK (1988) An experimental evaluation of articulation of preferences in multiple criterion decision-making (MCDM) methods. *Decis Sci* 19(4):880–888
- Noseworthy T, De Coster C, Naden R (2009) Priority-setting tools for improving access to medical specialists. Poster presentation. Paper presented at the 6th health technology assessment international annual meeting, Singapore
- Önüt S, Kara SS, Işık E (2009) Long term supplier selection using a combined fuzzy MCDM approach: a case study for a telecommunication company. *Expert Syst Appl* 36(2):3887–3895
- Paas FGWC, Van Merriënboer JGG (1994) Variability of worked examples and transfer of geometrical problemsolving skills: a cognitive-load approach. *J Educ Psychol* 86(1):122
- Paas F, Renkl A, Sweller J (2003) Cognitive load theory and instructional design: recent developments. *Educ Psychol* 38(1):1–4
- Pater LR, Cristea SL (2016) Systemic definitions of sustainability, durability and longevity. *Proc Soc Behav Sci* 221:362–371
- Phillips R, Troup L, Fannon D, Eckelman MJ (2017) Do resilient and sustainable design strategies conflict in commercial buildings? A critical analysis of existing resilient building frameworks and their sustainability implications. *Energy Build* 146:295–311
- Pill J (1971) The Delphi method: substance, context, a critique and an annotated bibliography. *Socioecon Plan Sci* 5(1):57–71
- Polatidis H, Haralambopoulos D (2002) Normative aspects of renewable energy planning; a participatory multi-criteria approach. The International Society for Ecological Economics. Ecological economics working paper series
- Pomerol J-C, Barba-Romero S (2012) Multicriterion decision in management: principles and practice, vol 25. Springer, Berlin
- Powell C (2003) The Delphi technique: myths and realities. *J Adv Nurs* 41(4):376–382
- Roy B, Bouyssou D (1986) Comparison of two decision-aid models applied to a nuclear power plant siting example. *Eur J Oper Res* 25(2):200–215
- Ruzgys A, Volvačiovas R, Ignatavičius Č, Turskis Z (2014) Integrated evaluation of external wall insulation in residential buildings using SWARA-TODIM MCDM method. *J Civ Eng Manag* 20(1):103–110
- Saaty TL (1980) The analytic hierarchy process: planning, priority setting, resources allocation. McGraw, New York. p 281
- San Cristóbal JR (2011) Multi-criteria decision-making in the selection of a renewable energy project in Spain: the VIKOR method. *Renew Energy* 36(2):498–502
- Scharfenberg F-J, Bogner FX (2013) Teaching gene technology in an outreach lab: students' assigned cognitive load clusters and the clusters' relationships to learner characteristics, laboratory variables, and cognitive achievement. *Res Sci Edu* 43(1):141–161
- Scheffer M, Hosper SH, Meijer ML, Moss B, Jeppesen E (1993) Alternative equilibria in shallow lakes. *Trends Ecol Evol* 8(8):275–279
- Schön DA (1983) The reflective practitioner: how professionals think in action, vol 1. Basic books, New York
- Triantaphyllou E (2000) Multi-criteria decision making methods. In: Multi-criteria decision making methods: a comparative study. Springer, Boston, pp 5–21
- Tzeng GH, Chen TY, Wang JC (1998) A weight-assessing method with habitual domains. *Eur J Oper Res* 110(2):342–367
- Walgrave S, Wouters R, Ketelaars P (2016) Response problems in the protest survey design: evidence from fifty-one protest events in seven countries. *Mobilization* 21(1):83–104
- Wallenius J (1975) Comparative evaluation of some interactive approaches to multicriterion optimization. *Manag Sci* 21(12):1387–1396
- Wang T-C, Lee H-D (2009) Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert Syst Appl* 36(5):8980–8985
- Woudenberg F (1991) An evaluation of Delphi. *Technol Forecast Soc Chang* 40(2):131–150
- Yang AL, Huang GH, Qin XS, Fan YR (2012) Evaluation of remedial options for a benzene-contaminated site through a simulation-based fuzzy-MCDA approach. *J Hazard Mater* 213:421–433
- Yoon K (1987) A reconciliation among discrete compromise solutions. *J Oper Res Soc* 38(3):277–286
- Zimmermann H-J (1990) Decision making in ill-structured environments and with multiple criteria. In: Readings in multiple criteria decision aid. Springer, Berlin, pp 119–151
- Zuo J, Zhao Z-Y (2014) Green building research-current status and future agenda: a review. *Renew Sustain Energy Rev* 30:271–281