sustainability

MDPI

Article

A Sustainable Water Resources Management Assessment Framework (SWRM-AF) for Arid and Semi-Arid Regions—Part 2: Refining the Conceptual Framework Using the Delphi Technique

Badir S. Alsaeed [1,2,*], Dexter V. L. Hunt [1] and Soroosh Sharifi [1]

[1] Department of Civil Engineering, School of Engineering, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[2] Department of Civil Engineering, College of Engineering, Qassim University, Buraydah 52571, Saudi Arabia
* Correspondence: balsaeed@qu.edu.sa

Abstract: In the first paper of this two-part series on the development of a sustainable water resources management assessment framework (SWRM-AF), a conceptual framework for arid and semi-arid regions was developed. The framework, rigorously selected through an extensive literature review, consisted of two main parts: components and indicators. This second paper of the series utilizes the Delphi technique as a participatory method to refine the conceptual framework, working toward a final version. This technique employs an iterative questionnaire through which 60 expert stakeholders from the Gulf Cooperation Council (GCC) countries were invited to assess 24 indicators across four components—social, economic, environmental, and infrastructure—and to assign their respective weights. However, while 40 accepted the invitation to be involved, only 33 participated in the first round. This paper identified that 97% of stakeholders in the first round were familiar with sustainability and its three pillars. The final result of the two rounds showed a preference for indicators within the social, economic, and infrastructure components to carry equal weights (i.e., four indicators within each of the three components were assigned 25% weighting each). In contrast, stakeholders identified that indicators within the environmental component should be assigned different weights. Therein, *'Compliance of wastewater treatment plants with regulations'* and *'Carbon dioxide emissions from desalination sector'* scored the highest and lowest weightings, accounting for 24.2% and 14.3% of the available weighting, respectively. The validation process resulted in a framework of 17 indicators. Thus, the refined and final version of SWRM-AF is presented and ready for implementation. The next stage of the research, which will keep the audience engaged, is to apply the newly developed SWRM-AF to an arid country and evaluate its effectiveness.

Keywords: sustainability; water resources management; Delphi technique; questionnaire; water sustainable framework; framework; indicator; indicator-based; decision-making

Citation: Alsaeed, B.S.; Hunt, D.V.L.; Sharifi, S. A Sustainable Water Resources Management Assessment Framework (SWRM-AF) for Arid and Semi-Arid Regions—Part 2: Refining the Conceptual Framework Using the Delphi Technique. *Sustainability* **2024**, *16*, 4463. https://doi.org/10.3390/su16114463

Academic Editor: Jan Hopmans

Received: 12 February 2024
Revised: 21 May 2024
Accepted: 23 May 2024
Published: 24 May 2024

1. Introduction

The results of a systematic literature review (see [1]) showed that a number of sustainable frameworks developed for the assessment of water resource management (WRM) were solely based on searching the literature to find appropriate components and indicators [2–4] as the first stage. The review showed that, either for creation or validation purposes, stakeholder opinion should be sought [5–10].

Therein, these stakeholders, who are experts within the field (in this case, with respect to water), were used for screening, filtering, and validating the choices made initially by the researcher(s) [7,11]. This study aims to use these same processes to help refine the conceptual sustainable water resources management assessment framework (SWRM-AF) for arid and semi-arid regions (ASAR). Therein, the Delphi technique will be applied to seek consensus. The Delphi technique is suggested to be the following:

> *"... a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem."* [12] (p. 3)

Likewise, water resources management (WRM) and its sustainability measurement can be considered a complex problem to deal with and develop, requiring a participatory approach, as stated by the Dublin principles [13].

Initially, this technique is applied here with the aim of reaching a consensus through multiple rounds of questionnaire surveys [14]. As such, it is very much an alternative to face-to-face meetings, allowing opinions to be sought and anonymity to be protected [15]. This approach avoids peer pressure and enables opinions to be sought purely based on opinion and prior knowledge, particularly in the first round. Another advantage of this technique is that it can assist in obtaining expert opinions without having the experts gather at a pre-determined time and location [7]. Consequently, using this method will allow many experts from different backgrounds to provide validation for the SWRM-AF, making it both reliable and usable for decision-makers and the public as a whole.

In this research, the Delphi technique (See Section 2) is used to assist in reaching a consensus on the identification of final sets of both components and indicators and assigning their weights accordingly. Such an approach has previously been used successfully in a variety of domains (e.g., [16–19]), including refining the indicators of a water sustainability index or framework [7,11]. In this research, due consideration is required for the following:

- the Delphi technique (Section 2.1)
- identification of the selection of stakeholders (Section 2.2)
- the design of the questionnaires (Section 2.3),
- the distribution and collection of questionnaires (Section 2.4),

All of these aspects are further discussed in Section 2.

2. Application of Delphi Technique

To check whether the selected sets of components and indicators, which form the conceptual SWRM-AF, could match with locally conceived and context-specific demands, the participatory approach was chosen to assess these sets. Not only this but the weights of each elected indicator and component shall be assigned during this process. One of these approaches that can fulfil the two objectives is the Delphi technique. In short, the Delphi technique is a survey method used to help respondents make appropriate consensus-based decisions [20] or sometimes, like in this specific case for the SWRM-AF, to validate the approach. The following section will present more about this method and how it is being applied herein.

2.1. The Delphi Technique

2.1.1. Overview

The first aim for adopting the Delphi technique was to obtain the most reliable consensus among experts through the use of a questionnaire interspersed with sections appropriate for written feedback [21]. Numerous studies have adopted a questionnaire-based approach to achieve consensus among experts regarding a complex issue. Nonetheless, obtaining and incorporating knowledge from various specialists with diverse viewpoints is not easy [22], although applications such as Delphi can help overcome this difficulty [23]. The philosophy behind the technique is that obtaining an accurate and trustworthy assessment is ideally achieved by consulting a panel of experts, whose consensus should be accepted as the best estimate of the answer [24]. A consensus helps set priorities for a range of water-related planning purposes [25].

A proxy for measuring the effectiveness of the Delphi method in handling many issues and reaching a consensus in different scientific areas is to identify the number of studies that have applied it since it was introduced. To find the answer, a simple search was performed through the Scopus database to show the number of studies between 1969 and

2022 with the word "Delphi" within its Title/Abstract/Keywords. The results showed a total of 34,007 studies, starting from only 14 in 1969 and ending with 4110 in 2022. The main subject area was medicine, with 15,924 studies, followed by social sciences, engineering, business and management, and computer science, respectively. Hence, this variety of subjects can stand as a proxy to identify that this technique is a helpful and well-established method for both qualitative and quantitative research. The previous remark was also noticed by Faulkner and Valerio [26], who argued that the strength of this method resides in its capacity to simultaneously provide a framework for consultation while allowing for the evaluation of both qualitative and quantitative elements.

Another question might arise about when this technique should be adopted. There are two possible main conditions or justifications behind using this technique [12]:

1. When accurate data do not exist or are expensive to obtain; or
2. When subjective inputs are necessary for evaluation models to the extent that they become the primary parameters.

Thus, this method could reduce the effort and time needed to find precise data that might be unavailable or need to be replaced by subjective thoughts by consulting experts in that field. Another reason for using this technique is that gathering experts from various countries at one time and place is difficult and expensively prohibitive.

There are several features and advantages to using the Delphi technique. One of these is that this technique is considered a powerful tool that enables individuals to think about problems in more sophisticated ways than they often would [27]. Another is that all participants can freely give their opinions [28], which helps lessen the impact of some psychological elements, including time pressure, the bandwagon effect, peer pressure of majority opinion, fraudulent persuasion, and the unwillingness to renounce publicly declared ideas [29]. In addition, Delphi advocates contend that group decisions are of greater reliability than those made by a single person [30] and are more objective when experts and their intuitive perspectives are sought after on a specific topic [31]. Consequently, reliability and objectivity in the outputs of this technique can be ensured since it depends on the collective opinion of experts.

Despite these advantages, a critique of the Delphi technique argues that it is challenging to evaluate the correctness and dependability of a procedure based on judgment and opinion [32]. Likewise, it has been observed that this technique's output quality is not always accurate since the practices of selecting the expert panel sometimes might not be followed carefully [33]. However, it can be said that the argument countering these critiques is that choosing the Delphi panel should be based on both subject-matter competence and ensuring that panellists come from a variety of backgrounds with experience in the area being studied [34–36]. Hence, the outputs of this method would be more accurate and trustworthy if clear and sound criteria were assigned and applied correctly.

2.1.2. Application Approach

When applying the Delphi technique in the questionnaires reported herein, a series of Likert scales were used to evaluate identified components, indicators, and their weights. A Likert scale is a method that can help summarize ideas, experiences, and issues under investigation into a series of statements that can be rated and, then, ranked. As such, it overcomes the challenge of finding a reliable and valid way of quantifying subjective preferences into areas such as thought, feeling, and action [37].

While different point systems can be used, a 5- or 7-point Likert scale is considered the most popular [38]. Adopting coarser scales such as 3-point scales reduces reliability and validity [39]. Hence, it was decided to use the 5-point Likert scale with the Delphi technique in rating the questionnaires of this study. The 5-point scale typically ranges from (1) "Not important" to (5) "Very important", with 3 being neutral.

The previous rule applied to 73% of the questions in questionnaires except those regarding the weight of both components and indicators. The remaining questions started

with a *'yes'* or *'no'* answer about whether the participants agreed to consider assigning equal weighting or not, in which case the Likert scale was used to designate these accordingly.

Furthermore, since consensus among experts is the target of following the Delphi technique, it is essential to have clear criteria and know when and how this has been achieved. This is important for the robustness of the SWRM-AF and its adoption by stakeholders. While different statistical methods with other criteria have been assigned to measure and obtain consensus in some studies [19,40], using simple methods might be better to ensure the understanding and cooperation of several stakeholders thereby gaining their trust. For example, a coefficient of agreement (C_c) was used to determine the level of consensus (i.e., the vote of the majority of participants) [11,41,42]. This level can be estimated by Equation (1) [41]:

$$C_c = \left(1 - \frac{V_n}{V_t}\right) \times 100 \quad (1)$$

where C_c is the coefficient of agreement presented as a percentage; V_n is the number of experts disagreeing with the dominant direction; V_t is the total number of experts. Regarding the acceptable level of consensus in general, there were different opinions in the previous studies. For instance, Salmond (1994) [43] conducted a brief literature survey and revealed that consensus levels can range from 50% to 80%. Moreover, some studies decided that consensus can be achieved when participants have at least two-thirds or a percentage equal to or bigger than (i.e., $\geq$) 67% towards one option [7,44]. Others preferred $\geq$ 50% to obtain a consensus level [11,42], which can be understandable if the selection is among multi-options (i.e., more than two). However, in this study, the percentage suggested by Santos [41] to obtain a consensus level was followed, which is $\geq$60%, as a cut-off. This is because this level can work with two types of questions (i.e., those with two options and multi-options).

2.2. Identification of Water-Related Stakeholders

The identification of water-related stakeholders was one of the crucial tasks in the application of the Delphi method. A criterion was assigned to select this study's expert stakeholders:

- Work or expertise should belong to the water sector and/or sustainability;
- Experience should be related to GCC countries as the target;
- Application of expertise should be in ASAR.

This study identified stakeholders from six categories:

- Academics;
- Consultants;
- Government officials;
- Practitioners or engineers;
- Technical experts; and
- Community or not-for-profit organizations.

Unfortunately, the response rate from the last category was very low. As a collective, these categories aimed to have stakeholders from diverse backgrounds with genuine interests in WRM and its sustainability. During round one, about 60 stakeholders were targeted from all six categories.

2.3. Design of the Questionnaires

Whilst it was assumed that multiple rounds of questionnaires would be required to reach a consensus on the selected indicators and the weight of components and selected indicators, this was, in fact, achieved in two rounds.

A first draft of the questionnaire was prepared, which passed an ethical review from the University of Birmingham ethics team (Approval code: ERN_0969-Apr2023). The questionnaire was refined following a pilot study. This involved 12 specifically chosen post-graduate researchers and alums from the University of Birmingham who had worked in the

field of sustainability, WRM, or GCC countries to obtain their early feedback on the overall design and the questions' clarity. This pilot study also allowed for the measurement of the average time participants would need to finish the whole questionnaire. Subsequently, questions were either revised, removed, or added based on their valuable feedback. The estimated time to finish the questionnaire (25 to 35 min) was noted in the invitation sent to the main participants.

It is worth mentioning that the traditional application of Delphi during round one typically starts with open-ended questions to collect information in the initial round [45]. However, the modified method adopted herein features close-ended questions and is considered a more appropriate option if the information related to the targeted field exists [46]. Using close-ended questions has several advantages, such as saving time for both the creator (i.e., researcher) and respondents, encouraging more people to participate and complete all rounds, ensuring important statements are included and not ignored [45,47], and reducing the confusion and randomness that comes with an open-ended discussion [28].

Hence, most questions were close-ended since the conceptual framework was already available. As suggested previously, the majority of questions required answers based on a 5-point Likert scale, while the minority required *'yes'* or *'no'* answers. Very few questions required open-ended answers. Regarding the Likert scale, respondents are considered to support or be given high importance if they select 4 or 5 on the Likert scale and are deemed not to support or give low priority if they pick 1 or 2 on the Likert scale. The answer of 3 on the Likert scale is regarded as a neutral response.

In the first round, the questionnaire was divided into seven main sections, as summarized in Figure 1.

The details of each section in Figure 1 are as follows:

1. Introduction (including consent form in English and Arabic) and outline of the purpose of the questionnaire. If the participants are willing to continue at this stage, they need to provide their approval to the consent form to let them move to the next section.
2. Participants' background includes 5 questions about their demographics (see Section 2.2) and 4 questions about their general scientific background in relation to sustainability, assessment tools, and WRM.
3. A brief explanation is presented for the framework and its mechanism, followed by a question on a five-level Likert scale to see to which degree the provided explanation was clear to the participants.
4. Evaluation of the four components with one mandatory question about whether their weight should be equal. If not, another question based on a four-level Likert scale is provided to evaluate the relative importance of each component.
5. As it is the main section, there is a brief description defining the chosen six indicators under each respective component (i.e., environmental, economic, social, and infrastructure); participants are subsequently allowed to evaluate them based on a five-level Likert scale. The scale of 1 to 5 ranges from "Not important at all" to "Very important", with 3, "Moderately important", being neutral. In the description of the evaluation question of each indicator, it is mentioned to participants that indicator(s) with low evaluation are likely to be excluded in the final version of the framework. Next, a 'yes' or 'no' question is presented to see if the participant suggests adding any indicator(s). If yes, then an open-ended question is given. This section's last question is whether each set of indicators under each respective component should carry equal weight. If the answer to the previous question by the majority is "No", then the weight would be based on their prior evaluation of the remaining indicators (i.e., the ones that have high evaluation, so they are not excluded).
6. Participants are asked for feedback by way of two questions based on a five-level Likert scale regarding the use or support of the policy or decision-maker(s) to use any SWRM-AF in the future after going through this questionnaire. The scale of 1 to 5

ranges from "Very unlikely" to "Very likely", with 3 being neutral with the phrase "Neither likely nor unlikely".

7. In this final section, participants are asked two questions. The first one is about whether they are willing to participate in a second round of questionnaires. If the response is yes, then they are asked to send an email to confirm. The purpose of not only asking the participant to provide a preferred method to be contacted for a second round is to keep their identity unknown during the analysis of results. The second question concerns the final or additional comment(s) that participants want to add before submitting the answers to the questionnaire.

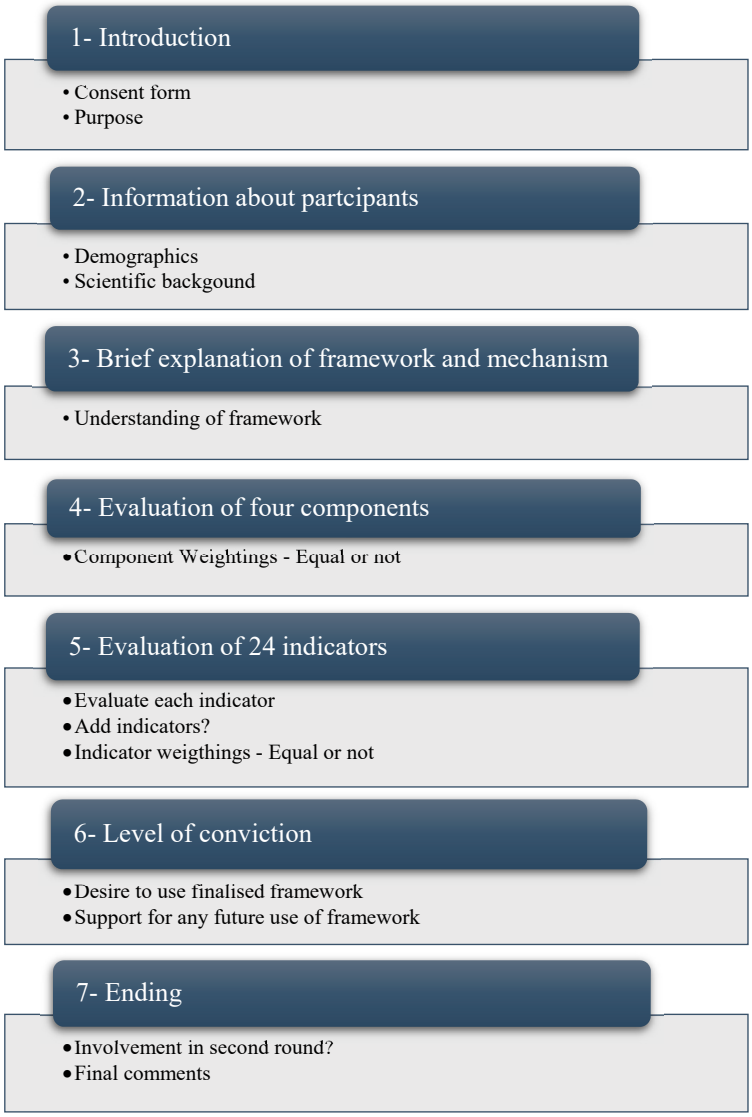

Figure 1. Seven main sections that constitute the questionnaire of the first round.

Based on the outcomes of the first round, the questionnaire for round two was created. This second questionnaire first provided a summary of the first round. Secondly, any question that did not achieve a consensus by 60% in the first round was asked again in order to achieve consensus around the majority's opinion. Finally, any comments or indicators that had been repeated at least three times were given to the participants to decide whether to consider it or not.

The design of the questionnaires in the second round started with another pilot study. The draft was distributed among the same group above but with a smaller number of participants from the first round. The feedback indicated that all questions in this round

were clear and straightforward. Moreover, the estimated time to finish the questionnaire of the second round was much less than the first, at a time range from 5 to 10 min.

2.4. Distribution and Receiving of Completed Questionnaires

The first step was to send a short invitation to each stakeholder, telling them briefly about the purpose of the questionnaire, asking them if they were willing to participate, and asking them what method they would prefer to be sent the link to the questionnaire through. Different techniques were used to send this invitation, such as emails, ResearchGate, WhatsApp, and social media accounts such as Twitter (now X) and LinkedIn. This starting step aligns with the preferred procedures applied in general surveys [48] and in the Delphi technique, in particular [45].

The first invitations were sent on the 14 June 2023. After receiving the acceptance of the experts to participate, the link to the questionnaire was sent to them based on their answers to the invitation letter through their preferred method (i.e., email, LinkedIn, ResearchGate, Twitter (now X), and WhatsApp). All questionnaires in this study were designed via Google Forms because of its simplicity and popularity among users.

The total number of respondents was 40 out of 60, and they were interested in filling out the questionnaire. However, some of them asked us to send a reminder to them on or after a specific date because they were busy during the time of the invitation.

The estimated time to receive the completed questionnaire was one month, which was broadly similar to the second round, except for one participant who submitted late because of his special conditions. Nonetheless, only 33 out of 40 experts actually fully completed and submitted the first questionnaire; thus, the return rate for this stage was 82.5%. This first-round response rate was higher than that of other research that used the Delphi approach, which frequently ranges from 30% to 50% [49]. Similarly, McMillan et al. (2016) [15] conducted a literature review on several previous studies to show the number of experts who agreed and completed the questionnaire later. The calculated results showed that the average response rate of experts who agreed and then completed the questionnaire was 70.76%. Therefore, it can be said that having an 82.5% response rate in the first round is higher than other studies, which could indicate an efficient application of this technique.

The participants in the first round were categorized as 13 academics, 9 consultants, 6 technical experts, 3 government officials, and 2 practitioners or engineers. It should be noted that a higher proportion of academic stakeholders (39.4%) could lead to analysis biases, such as the dominance of their viewpoint over that of other stakeholders. However, this can also be an advantage because the academic stakeholders were well known for having an in-depth understanding of water resource challenges in GCC countries, and some of them had participated in community-based water resource projects.

Regarding the gender of participants in the first round, 29 were males, 2 were females, and 1 chose not to say. From the responding cohort, and anecdotally based on the small piece of evidence presented here, it might be noted or suggested that fewer females (i.e., since only 6% participated) than males work or have experience in the water sector of GCC countries. Whilst more female stakeholders were invited to take part, response rates still were low.

In terms of education level, 23 respondents (69.7%) held PhD degrees, 5 held Master's degrees, and 5 held Bachelor's degrees. Hence, it can be concluded that all participants were very well-educated, which should give more credibility to the final results of this technique. The breakdown of countries the respondents were from is as follows: 24 from Saudi Arabia, 4 from Kuwait, 2 from Bahrain, 1 each from UAE, Oman, and Canada, respectively. Despite being asked, unfortunately, nobody participated from Qatar.

Regarding experience, 11 respondents (33.3%) had (more than 20 years) experience, 8 (24.2%) had (6 to 10 years), 5 (15.2%) had (16 to 20 years), 5 (15.2%) had (1 to 5 years), and 4 (12.1%) had (11 to 15 years). Hence, it can be seen that most of the participating experts (i.e., 84.8%) had a high experience that surpassed 5 years in the field, which gave their evaluation more credibility.

In the second round, questionnaires were sent to only 20 experts who showed their willingness to participate in another round, if needed. The decreased number of participants for different reasons between rounds is not rare [7,11,50–52]. Pan et al. [51] promoted the idea that the sample size needs to be as large as possible to account for subsequent drop-outs during new rounds while still being sufficiently small to guarantee that the participants are all specialists in their areas. Thus, the number of participants who expressed their desire to participate in the second round of this study (i.e., 20) can be considered sufficient.

However, only 19 participants out of 20 actually completed and submitted the second questionnaire. If we calculate the response rate based on the agreed experts from the first round, the return rate becomes high (i.e., 95%). Still, if the calculation is based on the first round participants (i.e., 33), the return rate becomes low (i.e., 57.6%). A flow chart illustrating the procedure of distributing invites and receiving completed questionnaires is presented in Figure 2.

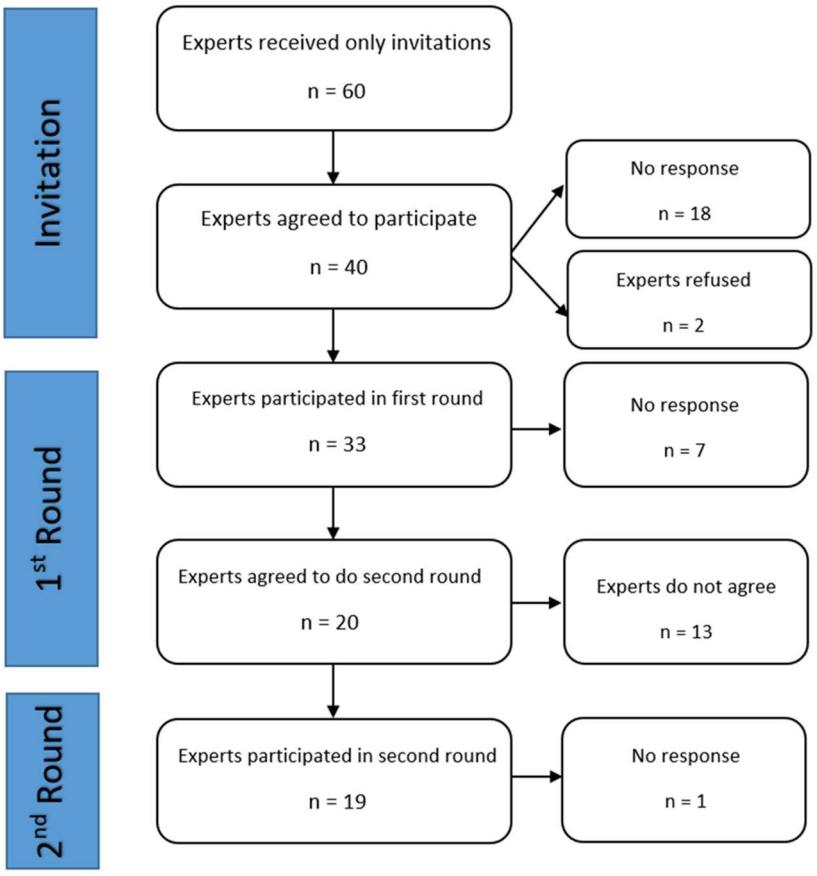

Figure 2. Questionnaire distribution and completion flow chart.

Moreover, the number of participants in the second round of each category was 12 academics, 4 consultants, 2 technical experts, and 1 government official. It can be seen that the academics are only missing one person from the first round, maybe because they understand the importance of surveys and their results to contribute knowledge. On the other hand, 89.5% of experts in this round were males, and 10.5% were females, while the education level was PhD for 78.9%, Master's for 5.3%, and Bachelor's degree for 15.8%. The countries the participant is working in now, or at least has good experience with, were Saudi Arabia (12 participants), Kuwait (4 participants), and 1 participant each for UAE, Oman, and Bahrain, respectively. Regarding the experience, 7 have (more than 20 years) experience, 7 have (6 to 10 years), 2 have (16 to 20 years), 2 have (11 to 15 years), and 1 has (1 to 5 years). Therefore, the participant group was made up of highly educated professionals with a high level of experience.

3. Results and Analysis

3.1. Round One

As stated previously, the primary purpose of this questionnaire was to let experts evaluate the initially selected set of components and indicators and their weights. In addition, experts were asked to suggest any crucial indicators that were not included under any components and to give pertinent feedback regarding the conceptual framework. This conceptual framework was presented in the first paper of this two-part series on the development of SWRM-AF for ASAR (see [53]).

Below is a summary of the responses for the remaining sections of the first questionnaire.

3.1.1. Introductory Questions—Participant Background

Before moving to the main questions, four introductory questions were posed to assess participants' general knowledge about sustainability, assessment tools, and WRM. These questions and the type of answers are illustrated in Table 1. The answers' rate and their distribution are shown in Figure 3.

Table 1. A brief survey about the scientific background of participants.

#	Question	Type of Answer
a	How familiar are you with sustainability and its three primary pillars?	five-level Likert scale (Not familiar to Extremely familiar)
b	How frequently do you use any sustainability assessment tools?	five-level Likert scale (Never to Every time)
c	How would you rate your knowledge of sustainability in the field of Water Resources Management (WRM)?	five-level Likert scale (Very poor to Excellent)
d	Are you currently using any Sustainable Water Resources Management Assessment Frameworks (SWRM-AFs)?	Yes or No

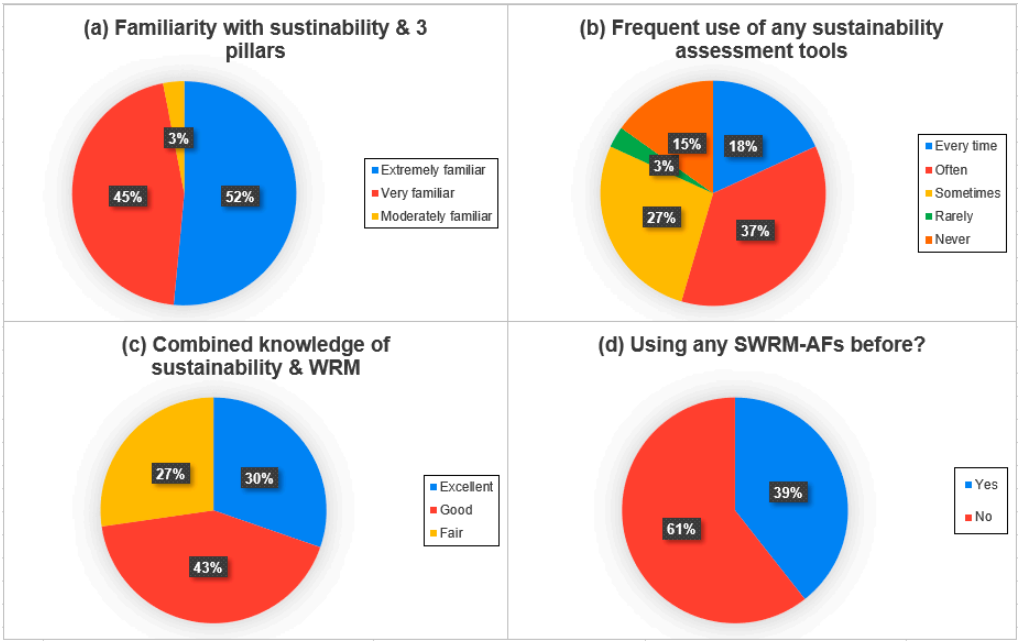

Figure 3. Distribution of answers to questions about the scientific background of participants.

From the answers to Question a (Figure 3a), it can be said that most (i.e., 97%) participants were either "Extremely familiar" or "Very familiar" with the principle of sustainability and its three pillars. In addition, 3% were moderately familiar, meaning no participants

were either "Slightly familiar" or "Not familiar". Similarly, Figure 3b shows that at least 82% had either sometimes, "often", or "frequently" used a sustainability assessment tool. In parallel, Figure 3c indicates that 73% had either good or excellent knowledge of sustainability and WRM, while only 27% had acceptable levels. The last two levels, "Poor" and "Very poor", were once again not selected for this question. Hence, these results strongly suggest that most of the participants had, in one way or another, some experience with (and, therefore, would be suitable for) helping achieve the aim of this study (i.e., to refine the SWRM-AF).

The last question in Table 1, a yes or no question, was provided to check any previous use for any SWRM-AFs. This question could be split into two questions depending on the answer. If the answer is no, no further inquiry is required, while if the answer is yes, another question appears about the name of this SWRM-AF. Figure 3d shows that 61% of respondents had never used any SWRM-AFs before, while 39% claimed they had used them. These participants were then able to state the type of SWRM-AF they had adopted, stated as follows:

- The national water strategy of the Kingdom of Saudi Arabia (KSA),
- The use of greywater, which can be classified as a sustainable intervention,
- The importance of groundwater (GW) management,
- The supply and demand management,
- The sustainability of the non-conventional water resources.

Whilst many of these suggestions cannot individually represent a whole framework, they can collectively be considered as an essential contributing part toward developing an SWRM-AF.

Additionally, the following water-related indices or tools were also stated:

- Water Quality Index (WQI), which treats the issue of water quality in particular.
- Stormwater Management Model (SWMM) for sustainable urban drainage systems as a type of SWRM-AF, while it is more of an evaluation tool to improve stormwater infrastructure.
- Integrated Water Resources Management (IWRM), which is more broad than an assessment tool.

Lastly, a minority of experts provided a range of actual SWRM-AFs (e.g., Water Poverty Index (WPI), Arab Water Sustainability Index (AWSI), and Abu Dhabi Water Index (ADWI)). One water expert was involved in developing a framework to help achieve sustainable development, but its name was not given since it was still in the conceptual stage. One water expert mentioned Mostadam and LEED, which are assessment frameworks but are more about green building principles. That said, they do include a number of key indicators on water. Hence, it appears that not all experts knew precisely the existence of (or have used) any of the SWRM-AFs, which is backed up by previous findings. This also validates the premise of this research.

3.1.2. Check the Understanding of the SWRM-AF

In the third section of the questionnaire (See Figure 1), a concise overview of the framework and its mechanism was presented. This overview included a visual example of the key parts that form indicator-based frameworks, along with the calculation direction. This explanation was crucial in ensuring participants could effectively evaluate subsequent sections containing the most important components and indicators.

Participants were subsequently asked to rate their understanding of an SWRM-AF, and the question was as follows: is the brief explanation provided clear to you? The answers were on a five-level Likert scale, which starts from the lowest "Strongly disagree" to the highest "Strongly agree".

Of all the participants, 42.4% strongly agreed that the explanation was clear, and 48.5% agreed. Only one person found the explanation unclear, while two others (9.1% com-

bined) chose the neutral answer. Based on this, it can be concluded that most participants (i.e., 90.9%, or 30 people) found the explanation clear and sufficient.

3.1.3. Evaluation of Components

While the main pillars of sustainability shall be included in any sustainable indicator-based frameworks, some doubts can arise about adding more components. However, it can be argued that the infrastructure is a crucial component if this framework is developed to assist in evaluating the sustainability of WRM, as it is in this study. Therefore, infrastructure was added as the fourth main category or component of the SWRM-AF.

Regarding how this is related to the questionnaire, a statement was given to the experts to inform them about the selection of the three pillars of sustainability and the addition of the infrastructure. This is very much supported by the findings of a systematic literature review undertaken by [1] as part of the development process of the SWRM-AF. This statement ended with a 'yes' or 'no' question to check whether these four components should carry the same weight during the calculation process. This process can give an indirect sign to measure the acceptability of adding this component. The majority (i.e., 54.5%) voted that all four components should carry the same weight. Moreover, 13 out of 15 (i.e., 45.5%) rated the infrastructure as requiring a higher weighting than the other components. Therein, five experts suggested it should be considered as "Very important", and eight suggested it was "Important". Therefore, these results could prove indirectly that having the infrastructure is essential. On the other hand, since the rate of both results is not equal to or more than 60%, which was selected as a cut-off to gain the consensus, this question needs to be repeated with its results to the experts in the second round to see whether some of them are willing to change their opinion.

3.1.4. Evaluation of Indicators

In the questionnaire, all participants were once again asked to provide an evaluation based on a five-level Likert scale (from "Not important at all" to "Very important") for each suggested indicator. The researcher assigned a point rating system to convert all these evaluations to a specific number, as can be seen in Table 2. Then, a yes or no question was given to see if any indicator(s) were missing and the experts would like to add any. If yes, another open-ended question would appear to let the participant add this indicator(s) and give it a rating from 1 to 5 based on the previous scale. Finally, a similar question to the one related to the weight of the components was asked to see whether these indicators under the same component should carry equal weight or not. If the answer is yes (by 60% of the participants), then nothing should be done by the researcher. If the answer is no, then the weight assigned to each indicator would be determined by their previous evaluation based on the point rating system (i.e., Table 2), and that would include only those that are not excluded. However, if both answers did not achieve a consensus rate of 60%, this question should be asked again in the second round with a summary of the results of the first round.

Table 2. Point rating system for indicators.

Rate	Meaning or Description	Equivalent Points
1	Not important at all	−2
2	Slightly important	−1
3	Moderately important	0
4	Important	1
5	Very important	2

Environmental Indicators

All six environmental indicators in the conceptual SWRM-AF and their abbreviations are shown in Table 3. The summation of the points and the final evaluations are provided in Figure 4. It can be seen that Env.Ind.1 and Env.Ind.3 received the lowest evaluations, respectively. Therefore, they were color-coded, highlighting the second lowest (pink) and

lowest points (orange), respectively, and earmarked for exclusion from the final version of SWRM-AF.

Table 3. Environmental indicators and their abbreviations.

#	Type of Environmental Indicator	Abbreviation
1	Brine discharge rate of desalination plants	Env.Ind.1
2	Quality of discharged wastewater	Env.Ind.2
3	Carbon dioxide emissions from the desalination sector	Env.Ind.3
4	Compliance of wastewater treatment plants with regulations	Env.Ind.4
5	Share of reusing of treated wastewater	Env.Ind.5
6	Share of using groundwater for drinking	Env.Ind.6

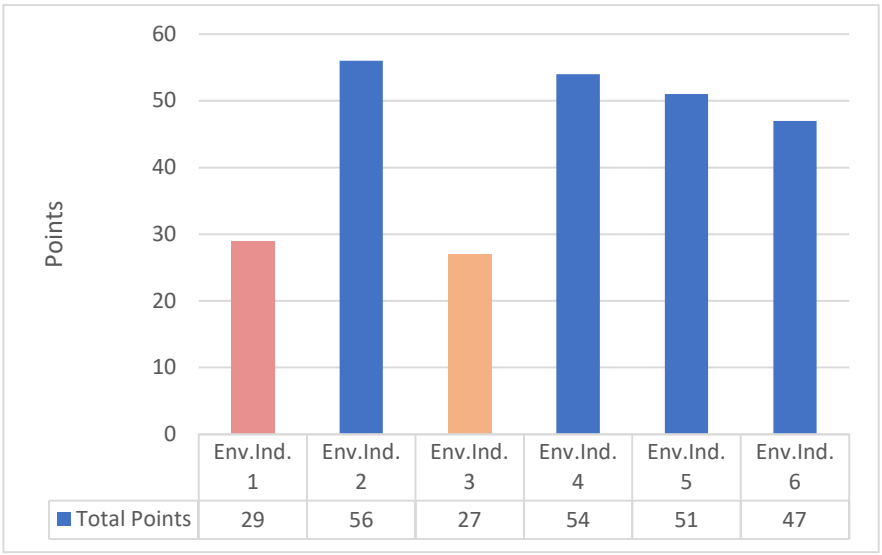

Figure 4. Collective evaluation of environmental indicators.

Regarding adding any environmental indicator(s), 51.5% said yes, while 48.5% said No. Those who selected yes were asked to add their suggested indicators accordingly. Some of these proposed indicators were related to the agriculture sector, while the targeted sector was the domestic sector, as was mentioned in the questionnaire introduction.

Others mentioned other indicators already included under another component, such as water quality and water stress, which is included under the social component. Similarly, rainwater harvesting was suggested by three experts; however, this was already incorporated as a part of the "Intervention acceptability" indicator under the social component. The main reason for these requests probably was that the order of the questions in the questionnaire was taken step by step. Therefore, the participants were not aware that these suggestions were already included.

Lastly, different views were given about greenhouse gas (GHG) emissions and their impacts on climate change. Three experts indicated this view explicitly, while others talked about energy consumption, electricity production, and the use of clean energy, which, together, could represent the same case. Furthermore, all these issues were represented by the third environmental indicator in Table 3 (i.e., carbon dioxide emissions from the desalination sector), since it has a massive impact on GCC countries and is considered a proxy for them. Despite the requests made, the collective evaluation of this particular indicator did not meet the desired standards and turned out to be the lowest. However, the researchers set a condition to ensure the minority's opinion was not ignored. If any issue or

indicator is repeated at least three times in any round, it must be brought to the table in the next round so that the majority can have the final say.

In the last inquiry, the topic at hand was whether or not the environmental indicators should be given equal weight in the calculation process. Experts' opinions were divided and, ultimately, a slim majority (51.5%) voted in favor of giving different weights, while 48.5% voted in opposition of choosing equal weights. However, since neither answer achieved a 60% consensus rate, the question was to be asked again in round two with a summary of first round results.

Social Indicators

The six social indicators and their abbreviations are shown in Table 4, and the summation of the point and the final evaluation for each indicator are provided in Figure 5. It is clear that Soc.Ind.3 and Soc.Ind.6 received the lowest evaluations and were color-coded, highlighting the lowest points (in pink) and the second-lowest points (in orange), which means they were excluded from the final version of SWRM-AF.

Table 4. Social indicators and their abbreviations.

#	Type of Social Indicator	Abbreviation
1	Per capita water consumption	Soc.Ind.1
2	Drinking water quality	Soc.Ind.2
3	Water stress	Soc.Ind.3
4	Water awareness	Soc.Ind.4
5	Intervention acceptability	Soc.Ind.5
6	Existence of participatory framework and guidelines	Soc.Ind.6

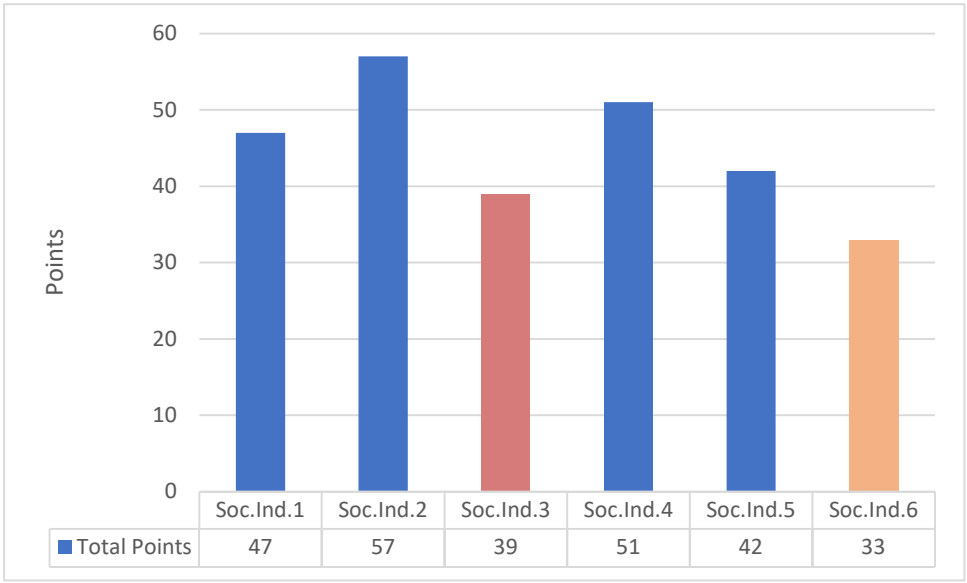

Figure 5. Collective evaluation of social indicators.

In contrast, the majority of the experts (78.8%) voted against adding any additional social indicators, implying that the current list is sufficient. However, the question of whether to assign equal weight to the indicators or not did not reach a consensus in the first round of voting. Out of the experts who voted, 57.6% were in favor of assigning equal weight, while 42.4% were against it. Therefore, this question must be re-evaluated in the second round of voting.

Economic Indicators

While the six economic indicators and their abbreviations are shown in Table 5, the summation of the points and the final evaluations are provided in Figure 6. It is evident that Eco.Ind.4 and Eco.Ind.5 received the lowest evaluations, respectively. Therefore, they were color-coded, highlighting the lowest points (in pink) and second lowest points (in orange), and excluded from the final version of SWRM-AF.

Table 5. Economic indicators and their abbreviations.

#	Type of Economic Indicator	Abbreviation
1	Water supply cost related to user's income	Eco.Ind.1
2	Unaccounted for water (water losses)	Eco.Ind.2
3	Water sector share in total public spend	Eco.Ind.3
4	Cost recovery of water supply utilities	Eco.Ind.4
5	Cost recovery of wastewater utilities	Eco.Ind.5
6	Pro-poor and pro-efficiency water fees	Eco.Ind.6

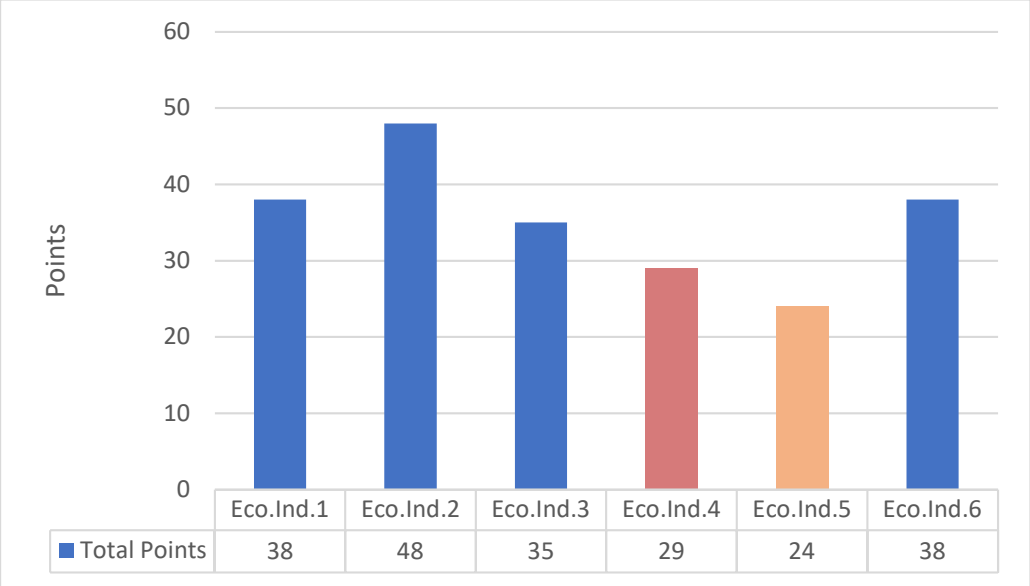

Figure 6. Collective evaluation of economic indicators.

The results are in for the next two questions. A total of 66.7% of the voters chose not to add any more economic indicators, which indicates that the current indicators are sufficient. Additionally, 63.6% voted to give equal weight to all economic indicators. This means both questions were resolved in the first round, and no further questions are needed in the second round.

Infrastructure Indicators

Table 6 shows the six infrastructure indicators and their abbreviations, while Figure 7 shows the results of their collective evaluations. It is evident that Inf.Ind.4 and Inf.Ind.5 were color-coded, highlighting the lowest points (in pink) and second lowest points (in orange), as they received the lowest evaluations and were excluded from the final SWRM-AF version.

Table 6. Infrastructure indicators and their abbreviations.

#	Type of Infrastructure Indicator	Abbreviation
1	Access to safe drinking water	Inf.Ind.1
2	Access to sanitation	Inf.Ind.2
3	Rate of water physical leakage	Inf.Ind.3
4	Share of desalinated water produced to domestic sector	Inf.Ind.4
5	Share of collected wastewater to domestic water	Inf.Ind.5
6	Age of water infrastructure	Inf.Ind.6

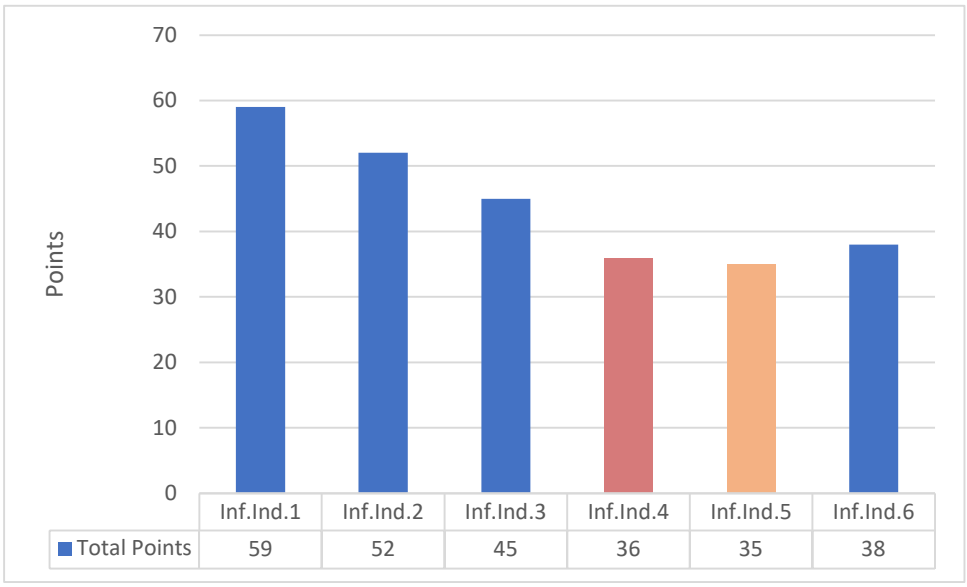

Figure 7. Collective evaluation of infrastructure indicators.

It was discovered in the following question that the majority of experts (75.8%) did not want to add any more infrastructure indicators, indicating that the current list is sufficient. However, there was no consensus in the first round of voting on whether the indicators should be given equal weight or not. The results were similar to the previous category (i.e., social indicators). Of those who voted, 57.6% favored giving equal weight, while 42.4% were against it. Therefore, this question must also be presented again in the second round of voting.

Level of Conviction

Before the final section of the questionnaire, two questions were asked to determine the extent to which participants were convinced or willing to support the use of SWRM-AFs. Both questions began with the phrase, "After participating in our questionnaire, how likely will you. . .". The two questions were based on a five-level Likert scale from 1 to 5, ranging from (very unlikely) to (very likely) with 3 being neutral. The first question focused on the participants' own adoption of SWRM-AFs, while the second question was about supporting the decision-makers in their area to use any SWRM-AFs. As illustrated in Figure 8a, an overwhelming 97% of experts expressed they would likely (36%) or very likely (61%) implement the use of SWRM-AF, with 100% stating they would support decision-makers to do the same: 21% likely and 79% very likely (Figure 8b). These results demonstrate a high level of conviction among experts regarding the benefits of using SWRM-AF either by themselves or by decision-makers in their area.

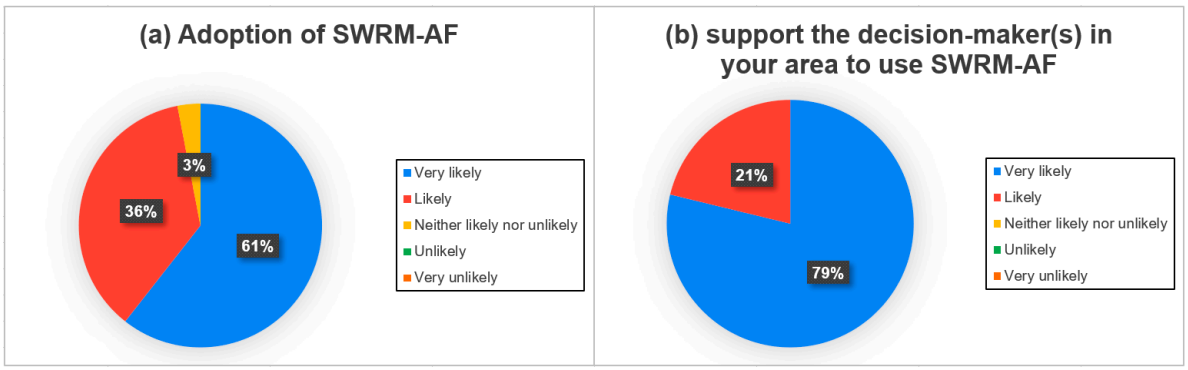

Figure 8. Level of conviction by experts (**a**) to use, or (**b**) to support the use of SWRM-AF.

3.1.5. Ending and Final Comments on Round One

In the last section of the questionnaire, there were two final questions. The first question asked participants if they were willing to participate in a second round if needed. Of all the respondents, 60.6% expressed their desire to participate in the second round, while 39.4% said they were uninterested. Then, another optional question was asked if the expert wanted to provide additional comments before submitting the first round questionnaire. Many participants were thankful and expressed their gratitude for having been involved in this process.

On the other hand, some experts added their feedback regarding the process or the idea behind the questionnaire itself. One respondent suggested the questionnaire took too much time and effort to translate from English to Arabic in order to make the answer as accurate as possible. Another suggested that if there was another full Arabic version of the questionnaire, that would attract more professionals who have experience, but the language is a barrier. One expert thought most of the questions were very deep and needed explanation and may not be helpful if answered by someone who does not understand them. This is true if the participant is not an expert in the field. Although the respondents to this questionnaire had good expertise, it was appropriate to take part in the questionnaire. The last critical comment was from an expert who suggested that the framework should be flexible, allowing modification according to the country and region from where it will be applied. This point is under consideration by the researcher for future applications.

3.2. Round Two

As stated previously, the primary objective of the second questionnaire was to inform experts about the results of the first round and to let them reconsider the questions that did not achieve a consensus rate of 60% or higher. Meanwhile, it is important in this round to see the opinion of the expert about adding any indicators suggested by different experts at least three times in the first round. Once all the issues have been resolved, the refinement stage will be considered complete, and the final version of SWRM-AF will be produced as an output of this process.

3.2.1. Evaluation of Components

After informing the participants about the results of the first round regarding the evaluation of components (refer to Section 3.1.3), they were first asked if they agreed with the majority's opinion from that round (i.e., an equal weight for all components, voted for by 54.5% of the participants). In this second round, 63.2% of the participants agreed with this, indicating that the required consensus rate had been achieved.

3.2.2. Evaluation of Indicators

Environmental Indicators

The four environmental indicators that received the highest evaluation in the first round were subsequently presented to participants. The only suggested indicator that shall be considered is the addition of a GHG emissions indicator since it was suggested explicitly by three experts and implicitly by another three. Before the question, a short explanation was provided about how large the CO_2 emissions generated by the desalination plants in the GCC countries are. An inquiry followed and was raised about whether to include carbon dioxide emissions from desalination as a fifth environmental indicator. The majority of respondents (78.9%) voted in favor, while the minority (i.e., 21.1%) voted against. As a result, a consensus was reached, and the decision was made to add this 5th indicator under the environmental component.

The next task involved consideration and allocation of weights since the first round did not achieve the consensus rate for equal or non-equal ratings. The answer in this round was clear, with the majority (i.e., 84.2%) supporting non-equal weighting. Hence, the consensus rate was achieved, and environmental indicators should not be treated equally.

As a result of this, participants were asked to rate each indicator in order to know their weights. This was based on a five-level Likert scale ranging from "Slightly important" to "Extremely important". The final list of environmental indicators and respective weights can be seen in Table 7. Therein, it can be seen that the central theme for indicators with higher weights was wastewater, and lower weights were given for the use of groundwater and GHG emissions. This might be because wastewater has had a visible and palpable impact on the environment for many years in the GCC countries, which made this issue require particular attention and fast treatments. In addition, since this weighting was based on a collective opinion of different experts, the difference between the weights was not that big, maybe except for the last indicator in Table 7.

Table 7. Final list of the environmental indicators and their weights (in hierarchy ordering).

Type of Environmental Indicator	Weight
Compliance of wastewater treatment plants with regulations	24.2%
Quality of discharged wastewater	23.8%
Share of reusing of treated wastewater	19.6%
Share of using groundwater for drinking	18.1%
Carbon dioxide emissions from the desalination sector	14.3%

Weight of Social and Infrastructure Indicators

The last two close-ended questions in the round two questionnaire were regarding the weight of the social and infrastructure indicators since several opinions during the first round did not achieve the targeted consensus rate of 60%. Using the same approach as in the previous section, a question was provided to see whether the participants would agree with the first round's collective result, where most experts agreed for equal weight but without reaching 60%.

In the current round, 63.2% of the experts agreed to give equal weight to the social indicators, while 68.4% voted for equal weight to the infrastructure indicators. Both these percentages are higher than the targeted consensus rate. Therefore, it can be said that the weighting issue for all indicators that form the SWRM-AF is resolved, where all indicators under three components, including the economic, require equal weights. In contrast, only the environmental indicators should have different weights, as illustrated in Table 7 above. The final weightings of the economic, infrastructure and social indicators are shown in Table 8.

Table 8. The final weightings of the economic, infrastructure and social indicators.

#	Components in (CAPS) and Indicators	Weighting
	ECONOMIC INDICATORS	
1	Water supply cost related to user's income	25%
2	Unaccounted for water (water losses)	25%
3	Water sector share in total public spend	25%
4	Pro-poor and pro-efficiency water fees	25%
	SOCIAL INDICATORS	
1	Per capita water consumption	25%
2	Drinking water quality	25%
3	Water awareness	25%
4	Intervention acceptability	25%
	INFRASTRUCTURE INDICATORS	
1	Access to safe drinking water	25%
2	Access to sanitation	25%
3	Rate of water physical leakage	25%
4	Age of water infrastructure	25%

3.2.3. Ending and Final Comments

In the last section of the second questionnaire, an open-ended question was given to gather additional feedback. The majority of comments were, in general, supportive of the developed SWRM-AF, while two comments highlighted were either highly critical ((1) below) or provided good insight ((2) below):

1. One participant suggested the selection of indicators can be linked to each other (i.e., as cause and effect), for example, the issue of the age of the infrastructure and water leakage. Thus, inadvertently, the framework would be weighted more towards these two (although they are related) than the other two.
2. Weight flexibility should be considered since the SWRM-AF is developed for different countries with different local contexts and local priorities within several time sets or terms.

While the first comment above could be significant, it might not always be true that the two indicators are related. Water leakage can happen because of other factors and problems, such as the material and diameter of the pipes [54], high pressure [55,56], improper fixing of pipelines, heavy traffic, and roots of trees [57]. On the other hand, the ageing of water infrastructure can cause different problems to the whole process other than leakage, such as impact on the water quality [58–60] and decreased water supply reliability [61,62]. The second comment above is both reasonable and logical. However, the final say here is (and should be) by the experts. Hence, some features of the SWRM-AF can be user-defined, allowing for future iterations of the SWRM-AF.

3.3. The Refined SWRM-AF

After going through the results of the questionnaires during the two rounds above, it can be said that the refinement stage is completed, and the final SWRM-AF can be introduced. The output of this stage is controlled and proceeded via the Delphi technique. Moreover, the final version of SWRM-AF includes four components that should be treated equally during the calculation process and 17 indicators, as shown in Figure 9. Indicators under the economic, social, and infrastructure components will have equal weights (Table 8), while only environmental indicators will have different weights, as illustrated above in Table 7.

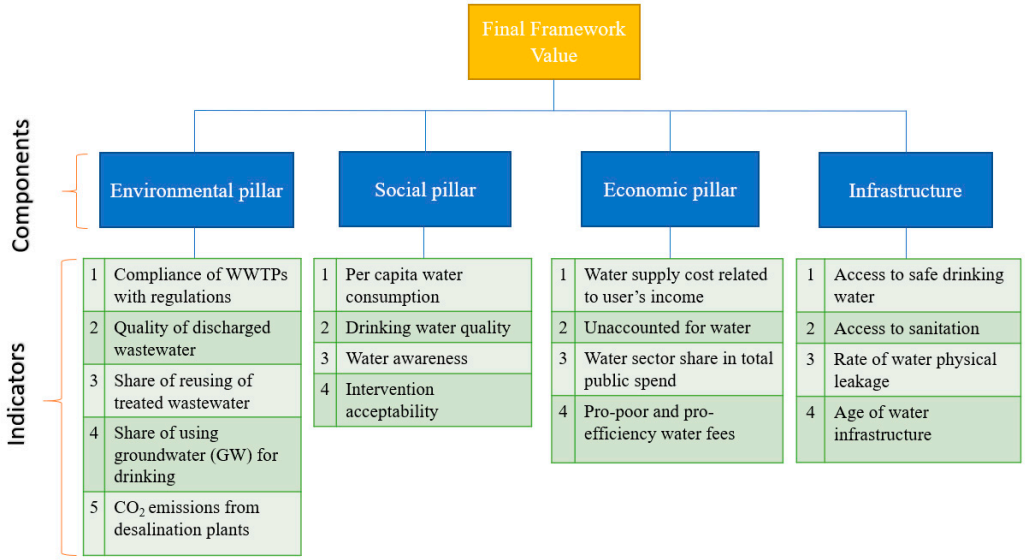

Figure 9. The final version of SWRM-AF after the application of the Delphi technique.

Regarding additional information on the SWRM-AF that has been used to develop the SWRM-AF, the reader is directed to the systematic literature review undertaken by the authors [1]. Each indicator's benchmarks or metrics are provided in Part 1 of this paper [53].

The SWRM-AF application scale for ASAR and particularly GCC countries is assumed to be national. The normalization method, which is another main element in the formation of an indicator-based water sustainability framework (IBWSF) [1], is mainly the categorical scaling method. At the same time, few indicators require the continued re-scaling method. While the weighting scheme is mixed between equal and non-equal based on the opinions of stakeholders, the aggregation technique used here to make the summation among components and indicators is the arithmetic technique. Finally, the final framework (or index) value is a number or percentage from 0 to 100.

The initial design for the SWRM-AF as a tool is produced using an Excel spreadsheet. At the same time, the user is required to enter a specific whole number, which is any number from 0 to 5 only. These numbers have particular colours similar to the traffic light system (red to amber to green) and are called scores, representing specific evaluations or levels of each indicator. While the indicator with an evaluation value of 0 to 1 usually means the worst option and is located in the red area, the evaluation value of 5 is the best level and is located in the green area. Also, an equivalent value from 0 to 100 is given for each score during the calculation process, where zero is 0, 1 is equivalent to 20, 2 is equivalent to 40, 3 is equivalent to 60, 4 is equivalent to 80, and 5 is equivalent to 100.

Then, the final evaluation for each group of indicators under one component and the final framework value that all of them result in from the final normalization and aggregation process are what the user of SWRM-AF can expect to see as an output. These results are presented in a radar chart. Examples of these outputs of each group of indicators and all components are shown in Figures 10 and 11. Finally, it can be seen in Figure 11 that the final framework value is equal to 66.46 (out of 100), which represents the sustainability level of the domestic sector, for example, based on the included indicators and components of the SWRM-AF.

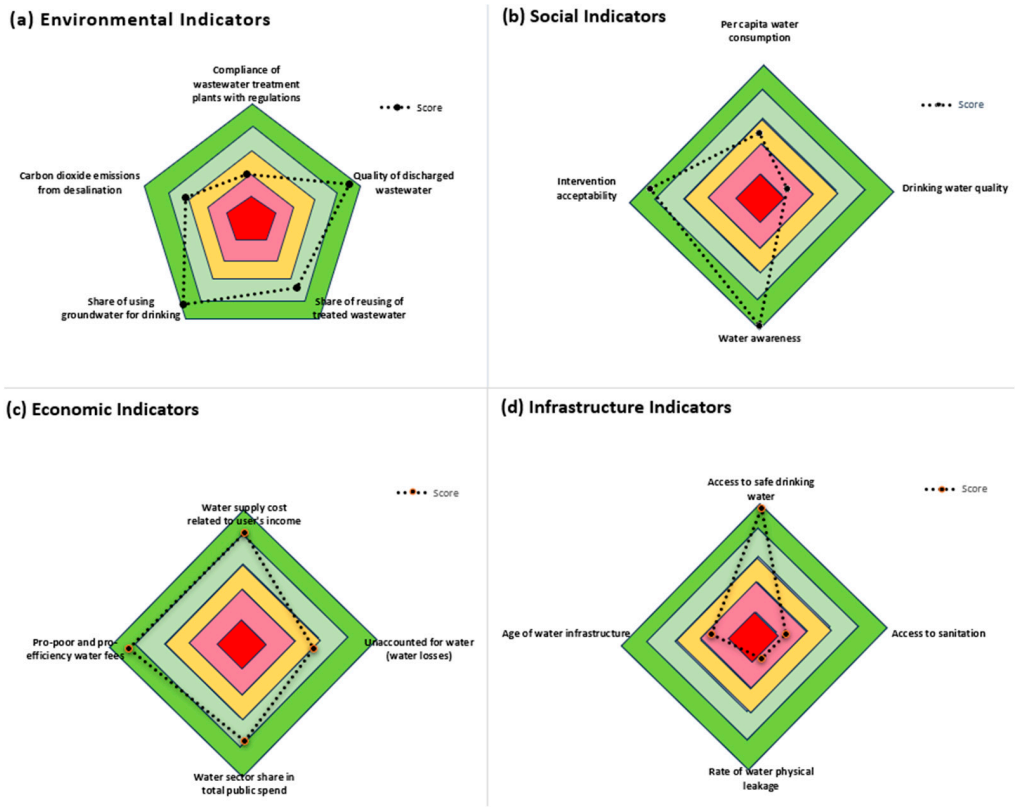

Figure 10. Example for the visual output of SWRM-AF for each group of indicators.

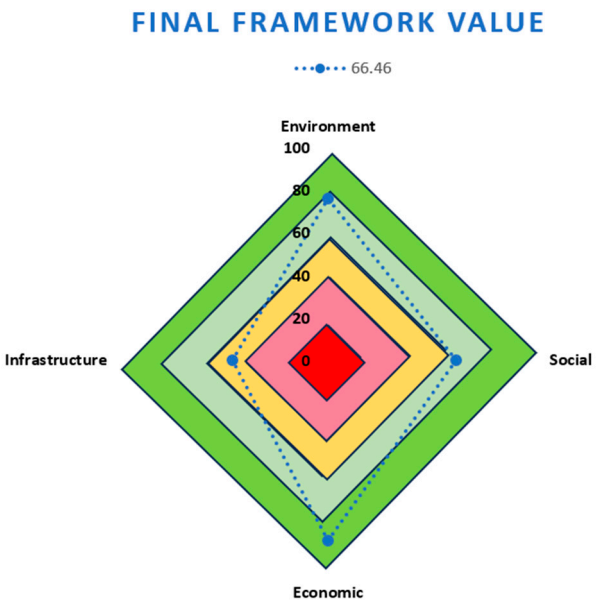

Figure 11. Example for the final visual output of SWRM-AF for all components.

4. Discussion

This research sought to refine the conceptual SWRM-AF by applying a participatory method represented by the Delphi technique. A further discussion about this method, the reason for selecting this technique, and the design philosophy of the questionnaires are provided in Section 4.1. Section 4.2 illustrates the possible shortfalls of this research.

4.1. Reasons behind Selection and Design

The first discussion point is whether the participatory approach, which is the base of this study, is mandatory (and significant) or optional (and unnecessary). One method to find the answer is to look back at the previous indicator-based water sustainable framework. A comparison between 19 frameworks in the literature conducted by [1] can be used here. Among the researchers who developed these frameworks, 78.9% have made a connection with several stakeholders in one way or another to assist in one or both main missions that include indicators' selection and/or assigning weights for them. The majority (i.e., 73.7%) have been involved in the selection process of indicators, while only 42.1% have a leading role in giving or assigning specific weights to indicators and/or components. This low percentage in assigning weights was because most other researchers decided to use equal weighting as their primary assumption for all indicators and components. It is not inappropriate to assume that the reason behind this is likely that assigning specific weights can be administratively expensive. Nevertheless, this overall high percentage (i.e., 78.9%) in adopting the participatory approach to develop these frameworks can indicate its importance in accomplishing such a mission.

Another question might be posed about how the Delphi technique can serve the validity of the SWRM-AF. The simple answer is that it can be divided into two parts.

Firstly, it includes the opinion of experts from different backgrounds, who are simultaneously stakeholders with a genuine interest in seeking and pursuing the sustainability of WRM. Therefore, the Delphi technique can ease the way to share and combine their pure scientific opinions without any pressure since their identities are anonymous to each other. Then, their consensus would represent a quantum leap in the field for ASAR, which can provide a realistic method of measurement striving towards improved sustainability credentials.

Secondly, using this technique as a validation method or tool has a proven track record in various areas. For example, this includes, but is not limited to, the following: the manufacturing industry [63], the role of advanced practice nursing [64], web engineering field [65], digital competence in higher education [66], construction and project complexity indicators [67], and ecological models [68]. In their studies, each of the previous examples emphasized the role of the Delphi technique in the validation stage and as a validation tool. Moreover, it was found in a study about the sustainable building criteria, which applies multiple consensus methods, that the results of the Delphi method are more significant in validating the gathered and reviewed data [69]. Therefore, the adoption of this technique to validate the conceptual SWRM-AF can be justified based on all these scientific findings and views.

Another aspect that warrants further discussion is concerning the overall design philosophy of the questionnaires. The preliminary design of the first questionnaire, followed by the initial draft, was much longer. This is because there was a need to cover every aspect with a specific and direct question. In other words, there was a desire to make the questionnaire as comprehensive as possible as the first objective. On the other hand, the second objective was to not exceed 40 questions, following the assumption that each question would require one minute to be read and answered. However, some criticism was given regarding the length of the questionnaire during the first pilot test. This increased awareness of the risk concerning receiving few responses with less engagement if significant reductions were not considered. Therefore, several questions were subsequently merged with each other to decrease the required time to complete the questionnaire.

Regarding the timing, some studies claim that the ideal time to finish a questionnaire should not surpass 10 min to gain the maximum response rate [70] and to reduce drop-out rates and the effects of tiredness [71,72]. However, while this aim is suitable for short surveys, adhering to this rule with a lengthy nature questionnaire became unrealistic. Therefore, based on the results of the first pilot test, the time suggested to complete the first questionnaire was between 25 and 35 min, and that was communicated to each participant before they started the questionnaire. This timing was somehow in line with a study

about the average time to complete a lengthy questionnaire found to be between 24 and 33 min [70].

4.2. Shortfalls of This Research

This paper depends solely on the Delphi technique to refine and validate the framework. However, some researchers argue that since this technique was criticized because of a few weaknesses, such as its subjectivity and time-consuming nature [73], it is better to be combined with another method or theory, such as the fuzzy theory [74,75] and the analytic hierarchy process (AHP) method [76,77]. Hence, while using another method might enhance the final output, this did not apply to this study.

Another remark should be given to the small sample size that was included in this study. While the initial goal was to reach at least 100 experts, it was somehow difficult to reach out to many of these experts because their contact information was hard to find. Furthermore, some of these experts have old emails on the internet that were not being used anymore based on the (email delivery failure) automatic reply that was returned when the invitation was sent. Thus, they were not counted among the total 60 experts who received at least the invitation to participate. Still, 20 did not respond at all, even to the invitation letter (maybe it went to their junk mail, or they were busy or not interested). At the same time, only 33 completed the whole questionnaire of the first round out of the 40 experts who showed interest in participating. This can be considered a weakness of the results of this research.

On the other hand, this study was conducted with a particular emphasis on measuring the SWRM of ASAR. As a result, it is unclear how the framework might be applied to other regions. Meanwhile, despite being validated through multiple research rounds, the framework still needs to be tested and used in real-world scenarios.

5. Summary and Conclusions

Since it was recommended that a suitable participatory approach be applied to complete the development of the conceptual SWRM-AF, the Delphi technique was used. The purpose of this method was to contribute to refining and validating the initial framework designed for ASAR, particularly in the GCC countries. Hence, this paper illustrates the details of the application of the Delphi technique for bringing chosen experts (and stakeholders) to a consensus rate of 60% on the previously determined components, indicators, and their weights. This process consists of two rounds of survey questionnaires and allows the expert stakeholders to provide their opinions by evaluating, adding, or removing mainly the selected indicators. The first round included 33 participants, while only 19 completed the second round. The main results after the end of the Delphi questionnaire application are as follows:

- 97% of stakeholders in the first round were familiar with sustainability and its 3 main pillars;
- All four components (i.e., environmental, social, economic, and infrastructure) shall carry the same or equal weight;
- The previously determined indicators shall be reduced from 24 to 17 indicators;
- Three sets of sub-indicators (i.e., for social, economic, and infrastructure components) shall carry the same or equal weight (i.e., four indicators within each of the three components were assigned 25% weighting each);
- In contrast, only the set of environmental indicators is assigned with different weights. Therein, *'Compliance of wastewater treatment plants with regulations'* and *'Carbon dioxide emissions from desalination sector'* scored the highest and lowest weightings, accounting for 24.2% and 14.3% of the available weighting, respectively.

Thus, it can be concluded that the final version of SWRM-AF is ready to be used on the national scale. The next step, the SWRM-AF, will now be applied to different case studies in the domestic sector of the GCC countries, serving as the foundation for suggestions to

water authorities and decision-makers for appropriate plans for managing water resources in each country.

Overall, it appears that the use of the Delphi technique is practical as a validation tool. Not only this, but it could help reach consensus among experts, which can contribute to the widespread actual application of the subject of the study. This was noticed during the answers to questions about the level of conviction when the majority supported the adoption of this new framework, and all participants said they would like to see the decision-makers using it. Hence, this can indicate that GCC countries, represented by their experts, are eager and strive for the sustainability of their WRM.

Author Contributions: Conceptualization, B.S.A.; methodology, B.S.A.; investigation, B.S.A.; resources, B.S.A.; data curation, B.S.A.; writing—original draft preparation, B.S.A.; writing—review and editing, B.S.A., D.V.L.H. and S.S.; visualization, B.S.A.; supervision, D.V.L.H. and S.S.; project administration, D.V.L.H. All authors have read and agreed to the published version of the manuscript.

Funding: D.V.L.H. received the financial support of the UK EPSRC under grant EP/J017698/1 (Transforming the Engineering of Cities to Deliver Societal and Planetary Wellbeing, known as Liveable Cities) and EP/P002021 (From Citizen to Co-innovator, from City Council to Facilitator: Integrating Urban Systems to Provide Better Outcomes for People, known as Urban Living Birmingham).

Institutional Review Board Statement: The project has been considered in line with the University of Birmingham's research ethics processes, and on the basis of the information that has been provided, it is understood that while the project does involve human participants, it raises no substantial research ethics issues, and therefore, no further ethics review is required (Application number: ERN_0969-Apr2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Acknowledgments: B.S.A. gratefully acknowledges the volunteer help of the post-graduate researchers and alums from the University of Birmingham to enhance the questionnaire during the pilot study. In addition, special thanks are provided to all the GCC's experts who spent their valuable time giving their scientific opinions to refine and validate our framework. Finally, B.S.A. appreciates the financial support that Qassim University in Saudi Arabia gave him during his doctoral studies.

Conflicts of Interest: The authors declare no conflicts of interest.

Nomenclature

Abbreviations

ADWI	Abu Dhabi Water Index
ASAR	Arid and Semi-Arid Regions
AWSI	Arab Water Sustainability Index
GHG	Greenhouse Gas
GW	Groundwater
IBWSF	Indicator-Based Water Sustainability Framework
IWRM	Integrated Water Resources Management
KSA	Kingdom of Saudi Arabia
SWMM	Stormwater Management Model
SWRM-AF	Sustainable Water Resources Management Assessment Framework
WPI	Water Poverty Index
WQI	Water Quality Index
WRM	Water Resources Management

Notations

C_c	coefficient of agreement presented as a percentage
V_n	number of experts disagreeing with the dominant direction
V_t	total number of experts

References

1. Alsaeed, B.S.; Hunt, D.V.L.; Sharifi, S. Sustainable Water Resources Management Assessment Frameworks (SWRM-AF) for Arid and Semi-Arid Regions: A Systematic Review. *Sustainability* **2022**, *14*, 15293. [CrossRef]
2. Chaves, H.M.L.; Alipaz, S. An Integrated Indicator Based on Basin Hydrology, Environment, Life, and Policy: The Watershed Sustainability Index. *Water Resour. Manag.* **2007**, *21*, 883–895. [CrossRef]
3. Gain, A.K.; Giupponi, C.; Wada, Y. Measuring Global Water Security towards Sustainable Development Goals. *Environ. Res. Lett.* **2016**, *11*, 124015. [CrossRef]
4. Silva, J.d.; Fernandes, V.; Limont, M.; Dziedzic, M.; Andreoli, C.V.; Rauen, W.B. Water Sustainability Assessment from the Perspective of Sustainable Development Capitals: Conceptual Model and Index Based on Literature Review. *J. Environ. Manag.* **2020**, *254*, 109750. [CrossRef]
5. Sullivan, C.A.; Meigh, J.R.; Giacomello, A.M. The Water Poverty Index: Development and Application at the Community Scale. *Nat. Resour. Forum* **2003**, *27*, 189–199. [CrossRef]
6. Policy Research Initiative. *Canadian Water Sustainability Index (CWSI): Project Report*; Policy Research Initiative: Ottawa, ON, Canada, 2007; ISBN 978-0-662-45045-0.
7. Juwana, I.; Perera, B.J.C.; Muttil, N. A Water Sustainability Index for West Java—Part 2: Refining the Conceptual Framework Using Delphi Technique. *Water Sci. Technol.* **2010**, *62*, 1641–1652. [CrossRef]
8. Rogers, B.C.; Dunn, G.; Hammer, K.; Novalia, W.; de Haan, F.J.; Brown, L.; Brown, R.R.; Lloyd, S.; Urich, C.; Wong, T.H.F.; et al. Water Sensitive Cities Index: A Diagnostic Tool to Assess Water Sensitivity and Guide Management Actions. *Water Res.* **2020**, *186*, 116411. [CrossRef]
9. Bahar, N.S.; Noor, Z.Z.; Aris, A.; Kamaruzaman, N.A.B. An Indicator Framework Approach on Manufacturing Water Assessment towards Sustainable Water Demand Management. *J. Environ. Treat. Technol.* **2020**, *8*, 875–883.
10. Ben-Daoud, M.; Moumen, A.; Sayad, A.; Elbouhadioui, M.; Essahlaoui, A.; Eljaafari, S. Indicators of Integrated Water Resources Management at the Local Level: Meknes as a Case (Morocco). In Proceedings of the E3S Web of Conferences, Kenitra, Morocco, 2 February 2021; Volume 234.
11. Crispim, D.L.; Pimentel Da Silva, G.D.; Fernandes, L.L. Rural Water Sustainability Index (RWSI): An Innovative Multicriteria and Participative Approach for Rural Communities. *Impact Assess. Proj. Apprais.* **2021**, *39*, 320–334. [CrossRef]
12. Linstone, H.A.; Turoff, M. *The Delphi Method: Techniques and Applications*; Addison-Wesley Educational Publishers Inc.: Boston, MA, USA, 1975; ISBN 978-0-201-04293-1.
13. ICWE. The Dublin Statement on Water and Sustainable Development. In Proceedings of the International Conference on Water and the Environment, Dublin, Ireland, 26–31 January 1992.
14. Hanafin, S. *Review of Literature on the Delphi Technique*; National Children's Office: Dublin, Ireland, 2004; pp. 1–51.
15. McMillan, S.S.; King, M.; Tully, M.P. How to Use the Nominal Group and Delphi Techniques. *Int. J. Clin. Pharm.* **2016**, *38*, 655–662. [CrossRef]
16. Lee, G.; Jun, K.-S.; Chung, E.-S. Integrated Multi-Criteria Flood Vulnerability Approach Using Fuzzy TOPSIS and Delphi Technique. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 1293–1312. [CrossRef]
17. de Brito, M.M.; Evers, M.; Höllermann, B. Prioritization of Flood Vulnerability, Coping Capacity and Exposure Indicators through the Delphi Technique: A Case Study in Taquari-Antas Basin, Brazil. *Int. J. Disaster Risk Reduct.* **2017**, *24*, 119–128. [CrossRef]
18. Ahmad, N.; Zhu, Y.; Ibrahim, M.; Waqas, M.; Waheed, A. Development of a Standard Brownfield Definition, Guidelines, and Evaluation Index System for Brownfield Redevelopment in Developing Countries: The Case of Pakistan. *Sustainability* **2018**, *10*, 4347. [CrossRef]
19. Casanovas-Rubio, M.d.M.; Armengou, J. Decision-Making Tool for the Optimal Selection of a Domestic Water-Heating System Considering Economic, Environmental and Social Criteria: Application to Barcelona (Spain). *Renew. Sustain. Energy Rev.* **2018**, *91*, 741–753. [CrossRef]
20. Cantrill, J.A.; Sibbald, B.; Buetow, S. The Delphi and Nominal Group Techniques in Health Services Research. *Int. J. Pharm. Pract.* **1996**, *4*, 67–74. [CrossRef]
21. Dalkey, N.; Helmer, O. An Experimental Application of the Delphi Method to the Use of Experts. *Manag. Sci.* **1963**, *9*, 458–467. [CrossRef]
22. Hwang, G.-H.; Chen, J.-M.; Hwang, G.-J.; Chu, H.-C. A Time Scale-Oriented Approach for Building Medical Expert Systems. *Expert Syst. Appl.* **2006**, *31*, 299–308. [CrossRef]
23. Hugé, J.; Le Trinh, H.; Hai, P.H.; Kuilman, J.; Hens, L. Sustainability Indicators for Clean Development Mechanism Projects in Vietnam. *Environ. Dev. Sustain.* **2010**, *12*, 561–571. [CrossRef]
24. Campbell, S.M.; Cantrill, J.A. Consensus Methods in Prescribing Research: Consensus Methods in Prescribing Research. *J. Clin. Pharm. Ther.* **2008**, *26*, 5–14. [CrossRef]
25. Uhl, N.P. Using the Delphi Technique in Institutional Planning. *New Dir. Instit. Res.* **1983**, *1983*, 81–94. [CrossRef]
26. Faulkner, B.; Valerio, P. An Integrative Approach to Tourism Demand Forecasting. *Tour. Manag.* **1995**, *16*, 29–37. [CrossRef]
27. Pollard, R.; Tomlin, M.E. The Use of Expert Teachers to Improve Education. *Education* **1995**, *116*, 3–9.
28. Franklin, K.K.; Hart, J.K. Idea Generation and Exploration: Benefits and Limitations of the Policy Delphi Research Method. *Innov. High. Educ.* **2007**, *31*, 237–246. [CrossRef]
29. Helmer, O.; Rescher, N. On the Epistemology of the Inexact Sciences. *Manag. Sci.* **1959**, *6*, 25–52. [CrossRef]

30. Murry, J.W., Jr.; Hammons, J.O. Delphi: A Versatile Methodology for Conducting Qualitative Research. *Rev. High. Educ.* **1995**, *18*, 423–436. [CrossRef]
31. Lang, T. An Overview of Four Futures Methodologies. *Manoa J. Fried Half-Fried Ideas* **1995**, *7*, 1–43.
32. Woudenberg, F. An Evaluation of Delphi. *Technol. Forecast. Soc. Chang.* **1991**, *40*, 131–150. [CrossRef]
33. Harrison, E.F. *The Managerial Decision-Making Process*, 5th ed.; Houghton Mifflin Boston: Boston, MA, USA, 1999; ISBN 0-395-90821-3.
34. Masser, I.; Foley, P. Delphi Revisited: Expert Opinion in Urban Analysis. *Urban Stud.* **1987**, *24*, 217–225. [CrossRef]
35. Rowe, G.; Wright, G.; Bolger, F. Delphi: A Reevaluation of Research and Theory. *Technol. Forecast. Soc. Chang.* **1991**, *39*, 235–251. [CrossRef]
36. Briedenhann, J.; Butts, S. Application of the Delphi Technique to Rural Tourism Project Evaluation. *Curr. Issues Tour.* **2006**, *9*, 171–190. [CrossRef]
37. Joshi, A.; Kale, S.; Chandel, S.; Pal, D.K. Likert Scale: Explored and Explained. *Curr. J. Appl. Sci. Technol.* **2015**, *7*, 396–403. [CrossRef]
38. Malhotra, N.K.; Peterson, M. *Basic Marketing Research: A Decision-Making Approach*, 2nd ed.; Pearson/Prentice Hall: Upper Saddle River, NJ, USA, 2006; ISBN 0-13-154865-4.
39. Dawes, D.J. Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales. *Int. J. Mark. Res.* **2008**, *50*, 61–104. [CrossRef]
40. de Brito, M.M.; Evers, M.; Almoradie, A.D.S. Participatory Flood Vulnerability Assessment: A Multi-Criteria Approach. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 373–390. [CrossRef]
41. Santos, A.C. O uso do método Delphi na criação de um modelo de competências. *Rev. De Adm.* **2001**, *36*, 25–32.
42. Santiago, L.S.; Dias, S.M.F. Matriz de indicadores de sustentabilidade para a gestão de resíduos sólidos urbanos. *Eng. Sanit. Ambient.* **2012**, *17*, 203–212. [CrossRef]
43. Salmond, S.W. Orthopaedic Nursing Research Priorities: A Delphi Study. *Orthop. Nurs.* **1994**, *13*, 31–45. [CrossRef] [PubMed]
44. Alexandrov, A.V.; Pullicino, P.M.; Meslin, E.M.; Norris, J.W. Agreement on Disease-Specific Criteria for Do-Not-Resuscitate Orders in Acute Stroke. *Stroke* **1996**, *27*, 232–237. [CrossRef]
45. Hsu, C.-C.; Sandford, B.A. Minimizing Non-Response in the Delphi Process: How to Respond to Non-Response. *Pract. Assess. Res. Eval.* **2007**, *12*, 17.
46. Kerlinger, F.N. *Foundations of Behavioral Research*, 2nd ed.; Holt, Rinehart and Winston: New York, NY, USA, 1973; ISBN 0-03-085462-8.
47. McCampbell, W.H.; Stewart, B.R. Career Ladder Programs for Vocational Educators: Desirable Characteristics. *J. Vocat. Educ. Res.* **1992**, *17*, 53–68.
48. Salant, P.; Dillman, D.A. *How to Conduct Your Own Survey*; Wiley: Hoboken, NJ, USA, 1994; ISBN 0-471-01267-X.
49. Wright, J.; Giovinazzo, R. Delphi: A Support Tool for the Prospective Planning. *Cad. Pesqui Adm.* **2000**, *1*, 54–65.
50. Green, H.; Hunter, C. The Environmental Impact Assessment of Tourism Development. *Perspect. Tour. Policy* **1992**, *36*, 29–47.
51. Pan, S.Q.; Vega, M.; Vella, A.J.; Archer, B.H.; Parlett, G.R. A Mini-Delphi Approach: An Improvement on Single Round Techniques. *Prog. Tour. Hosp. Res.* **1996**, *2*, 27–39. [CrossRef]
52. Dean, B.; Barber, N.; Schachter, M. What Is a Prescribing Error? *BMJ Qual. Saf.* **2000**, *9*, 232–237. [CrossRef] [PubMed]
53. Alsaeed, B.S.; Hunt, D.V.L.; Sharifi, S. A Sustainable Water Resources Management Assessment Framework (SWRM-AF) for Arid and Semi-Arid Regions—Part 1: Developing the Conceptual Framework. *Sustainability* **2024**, *16*, 2634. [CrossRef]
54. Saghi, H.; Aval, A.A. Effective Factors in Causing Leakage in Water Supply Systems and Urban Water Distribution Networks. *Am. J. Civ. Eng.* **2015**, *3*, 60. [CrossRef]
55. Marunga, A.; Hoko, Z.; Kaseke, E. Pressure Management as a Leakage Reduction and Water Demand Management Tool: The Case of the City of Mutare, Zimbabwe. *Phys. Chem. Earth Parts A/B/C* **2006**, *31*, 763–770. [CrossRef]
56. Pérez, R.; Puig, V.; Pascual, J.; Quevedo, J.; Landeros, E.; Peralta, A. Methodology for Leakage Isolation Using Pressure Sensitivity Analysis in Water Distribution Networks. *Control Eng. Pract.* **2011**, *19*, 1157–1167. [CrossRef]
57. Ali, H.; Choi, J. A Review of Underground Pipeline Leakage and Sinkhole Monitoring Methods Based on Wireless Sensor Networking. *Sustainability* **2019**, *11*, 4007. [CrossRef]
58. Clark, R.M.; Rossman, L.A.; Wymer, L.J. Modeling Distribution System Water Quality: Regulatory Implications. *J. Water Resour. Plan. Manag.* **1995**, *121*, 423–428. [CrossRef]
59. Al-Jasser, A.O. Chlorine Decay in Drinking-Water Transmission and Distribution Systems: Pipe Service Age Effect. *Water Res.* **2007**, *41*, 387–396. [CrossRef]
60. Cooper, A.; Fortuna, A.; Ahuja, S. Investigating the Missing Link: Effects of Noncompliance and Aging Private Infrastructure on Water-Quality Monitoring. In *Separation Science and Technology*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 11, pp. 329–339. ISBN 978-0-12-815730-5.
61. Ugarelli, R.; Di Federico, V. Towards Sustainability of Urban Water Networks: Addressing Management Issues with Innovative Approaches. *Irrig. Drain. Syst. Eng.* **2012**, *1*, 1000e103. [CrossRef]
62. Kerwin, S.; Adey, B.T. Optimal Intervention Planning: A Bottom-Up Approach to Renewing Aging Water Infrastructure. *J. Water Resour. Plan. Manag.* **2020**, *146*, 04020044. [CrossRef]

63. McLean, R.S.; Antony, J.; Garza-Reyes, J.A.; Samadhiya, A. A Continuous Improvement Implementation Framework for Manufacturing Companies: A Delphi Study-Based Approach for Development and Validation. *Int. J. Qual. Reliab. Manag.* **2023**, *40*, 2222–2246. [CrossRef]
64. Chang, A.M.; Gardner, G.E.; Duffield, C.; Ramis, M.-A. A Delphi Study to Validate an Advanced Practice Nursing Tool. *J. Adv. Nurs.* **2010**, *66*, 2320–2330. [CrossRef] [PubMed]
65. Torrecilla-Salinas, C.J.; De Troyer, O.; Escalona, M.J.; Mejías, M. A Delphi-Based Expert Judgment Method Applied to the Validation of a Mature Agile Framework for Web Development Projects. *Inf. Technol. Manag.* **2019**, *20*, 9–40. [CrossRef]
66. Mengual-Andrés, S.; Roig-Vila, R.; Mira, J.B. Delphi Study for the Design and Validation of a Questionnaire about Digital Competences in Higher Education. *Int. J. Educ. Technol. High. Educ.* **2016**, *13*, 12. [CrossRef]
67. Kermanshachi, S.; Dao, B.; Shane, J.; Anderson, S. Project Complexity Indicators and Management Strategies—A Delphi Study. *Procedia Eng.* **2016**, *145*, 587–594. [CrossRef]
68. MacMillan, D.C.; Marshall, K. The Delphi Process—An Expert-Based Approach to Ecological Modelling in Data-Poor Environments. *Anim. Conserv.* **2006**, *9*, 11–19. [CrossRef]
69. Chan, P. An Empirical Study on Data Validation Methods of Delphi and General Consensus. *Data* **2022**, *7*, 18. [CrossRef]
70. Cape, P.; Phillips, K. *Questionnaire Length and Fatigue Effects: The Latest Thinking and Practical Solutions*; Survey Sampling International: Shelton, CT, USA, 2015; 17p.
71. Revilla, M.; Ochoa, C. Ideal and Maximum Length for a Web Survey. *Int. J. Mark. Res.* **2017**, *59*, 557–565.
72. Hunt, D.V.L.; Shahab, Z. Sustainable Water Use Practices: Understanding and Awareness of Masters Level Students. *Sustainability* **2021**, *13*, 10499. [CrossRef]
73. Pill, J. The Delphi Method: Substance, Context, a Critique and an Annotated Bibliography. *Socio-Econ. Plan. Sci.* **1971**, *5*, 57–71. [CrossRef]
74. Tseng, M.-L.; Li, S.-X.; Lin, C.-W.R.; Chiu, A.S. Validating Green Building Social Sustainability Indicators in China Using the Fuzzy Delphi Method. *J. Ind. Prod. Eng.* **2023**, *40*, 35–53. [CrossRef]
75. Huang, J.; Koopialipoor, M.; Armaghani, D.J. A Combination of Fuzzy Delphi Method and Hybrid ANN-Based Systems to Forecast Ground Vibration Resulting from Blasting. *Sci. Rep.* **2020**, *10*, 19397. [CrossRef] [PubMed]
76. Shen, L.; Yang, J.; Jin, X.; Hou, L.; Shang, S.; Zhang, Y. Based on Delphi Method and Analytic Hierarchy Process to Construct the Evaluation Index System of Nursing Simulation Teaching Quality. *Nurse Educ. Today* **2019**, *79*, 67–73. [CrossRef]
77. Nguyen, T.A.V.; Tucek, D.; Pham, N.T. Indicators for TQM 4.0 Model: Delphi Method and Analytic Hierarchy Process (AHP) Analysis. *Total Qual. Manag. Bus. Excell.* **2023**, *34*, 220–234. [CrossRef]