

Article

Content Validation through Expert Judgement of an Instrument on the Nutritional Knowledge, Beliefs, and Habits of Pregnant Women

Elisabet Fernández-Gómez ¹, Adelina Martín-Salvador ¹, Trinidad Luque-Vara ^{1,*},
María Angustias Sánchez-Ojeda ¹, Silvia Navarro-Prado ¹ and Carmen Enrique-Mirón ²

¹ Department of Nursing, Faculty of Health Sciences, Melilla Campus, University of Granada, Calle Santander s/n, 52001 Melilla, Spain; elisabetfdez@ugr.es (E.F.-G.); ademartin@ugr.es (A.M.-S.); maso@ugr.es (M.A.S.-O.); silnado@ugr.es (S.N.-P.)

² Department of Inorganic Chemistry, HUM-613 Research Group, Faculty of Health Sciences, Melilla Campus, University of Granada, Calle Santander s/n, 52001 Melilla, Spain; cenrique@ugr.es

* Correspondence: triluva@ugr.es; Tel.: +34-686-951-942

Received: 20 March 2020; Accepted: 15 April 2020; Published: 18 April 2020



Abstract: The aim of this study was to conduct content validation through expert judgement of an instrument which explores the nutritional knowledge, beliefs, and habits during pregnancy. This is a psychometric study in which 14 experts participated in the evaluation of each of the questionnaire items, which were divided into two blocks according to the characteristics of sufficiency, clarity, coherence, and relevance. Fleiss' κ statistic was used to measure strength of agreement. A pre-test with 102 participants was conducted to measure the degree of understandability of the instrument. The strength of agreement obtained for each of the dimensions was almost perfect. For each pair of experts, strength of agreement ranged between substantial and almost perfect. Sufficiency was the characteristic of the questionnaire that obtained the highest values in the two blocks, and was also the most statistically significant ($p < 0.001$). Coherence was the most statistically significant characteristic in the first block ($p = 0.030$). Clarity was the most statistically significant characteristic in the second block ($p = 0.037$). The wording of five of the twenty original items was corrected. The new version of the instrument attained a high degree of understandability. The results suggest that the instrument is valid and may therefore be applied.

Keywords: content validity; expert judgement; Fleiss' κ ; pregnant women; eating habits

1. Introduction

Studies involving women during pre-conception, pregnancy, and breastfeeding report inadequate food intake for their physiological state and highlight the real need to ensure proper maternal and fetal nutrition [1]. Inadequate maternal nutritional intake during pregnancy may lead to adverse outcomes in fetal development, such as negative metabolic effects on offspring [2]. Pregnant women's nutritional knowledge may influence their food intake [3]. This is why it is considered of paramount importance to be able to delve into the nutritional knowledge and beliefs pregnant women have, as intended in the present study.

There is currently no standard guideline for validating health-related measures. However, several criteria developed in the fields of psychology and education sciences are used. There is a growing need for the use of health-related measuring instruments in clinical and research practice. The methodology for the adaptation of instruments is not very well known among health professionals, which may explain the existence of incomplete instruments or word-for-word translations of existing instruments in the field. A number of relevant and essential skills and guidelines, which should be acquired

and implemented in a dynamic and continuous way, are therefore required for the validation of an instrument [4].

Reliable instruments are necessary, as are instruments that have been validated using construct validity, criterion validity, and content validity, which are the most widely used measures of validity [5]. Validity and reliability are two quality criteria that any measuring instrument must satisfy in order to be used by researchers in their studies. In order to validate health-related measuring instruments and ensure that they are reliable and valid, their psychometric properties must be subjected to a process of adaptation and validation. All of this is essential to determine the quality of the measurements produced by these instruments [4]. Content validity is defined as, “the degree to which elements of an assessment instrument are relevant to, and representative of, the targeted construct for a particular assessment purpose” [6] (p. 238). Instruments should therefore be of the highest quality, which would make it easier to obtain valid and reliable evidence [7]. In this sense, the usual way to assess the quality of an instrument is by consulting experts, which fundamentally consists of evaluating an instrument using a procedure known as expert judgement [8].

Content validation through expert judgement is defined by Escobar-Pérez & Cuervo-Martínez [9] (p. 29) as an informed opinion from individuals with a track record in the field who are regarded by others as qualified experts and who can provide information, evidence, judgements, and assessments. Evaluation through expert judgement consists of asking a number of individuals to make a judgement on an instrument or to express their opinion on a particular aspect [10]. Content validations are generally conducted either during the design of a test or for the validation of the translation and standardization of an instrument for use in a different culture. In both cases, the role of experts is fundamental in clarifying, adding, and/or modifying the necessary aspects [11].

In order to validate the content of an instrument through expert judgement, content validity and expert judgement must be conceptualized, an implementation procedure must be established, and statistical alternatives for data analysis must be provided in order to make decisions [9]. To ensure their proper implementation, various aspects should be taken into account, such as the criteria and strategies for the selection of experts, the appropriate number of experts, and the instruments involved. As mentioned in the previous paragraph, the role of experts is fundamental in this endeavor, as it allows for nuances to be added or modified, which is why the diversity of experts is important. For the selection of the judges, the aforementioned procedure is used. It is important to select individuals who are knowledgeable in the subject matter, either because of their academic background, work experience, or recognition in the community [9]. When selecting the judges, it is important to choose individuals who are knowledgeable about the subject, either because of their academic background or because of their work experience. The information provided by the judges may be collected individually, in groups, or using the Delphi method [10]. As part of the procedure, the experts assess the different dimensions using a numerical scale [12] while taking into consideration their reasoned opinions. This is useful to identify the weaknesses and strengths of the instrument [5]. As a result, the expert judgement includes both quantitative and qualitative aspects.

According to Cabero and Llorente [10], expert judgement as an evaluation strategy offers many advantages, such as the high quality of the judges' responses and the possibility of obtaining extensive information on the subject matter. This is a procedure whose correct performance is sometimes the only indicator of the content validity of a research instrument [9]. Therefore, using formal methods leads to more scientifically sound results, especially when placing particular emphasis on the characterization and selection of experts, on the use of a scale that facilitates quantitative assessments, and on the analysis of the results using appropriate statistical tests [12].

Usually, the nutritional knowledge, beliefs, and habits of pregnant women are measured because their nutritional needs increase during this period. In addition to fulfilling the mother's needs, the needs of the fetus must also be covered [13,14]. Inadequate nutritional intake at this stage in life may have negative short- and long-term health consequences for both the mother and the child [2,15,16]. Pregnancy is widely regarded as a vulnerable nutritional stage. As a consequence, it is of the utmost

importance to identify any existing problems through validated instruments in order to be able to address them using suitable educational measures. Knowing which supplements are useful during pregnancy and how much weight to gain during this period is very important when it comes to introducing proper eating behaviors to pregnant women. Nutrition-related myths and the deficit of nutrition education put at risk the presence of essential foods in their diet and may lead to increased consumption of foods with low nutritional value and quality [17].

Behaviors are difficult to modify and are generally influenced by various environmental factors beyond personal control [18]. However, it has been shown that individuals can regulate their eating behaviors for different reasons [19]. Pregnancy may be one of these reasons, as it is a motivating period of time to acquire healthy eating patterns. The dietary behaviors of pregnant women have been found to be different compared to those of non-pregnant women [20].

Eating is not only a nutritional phenomenon, but also a cultural and social phenomenon where different beliefs influence the acceptance or rejection of certain foods [21]. Most religions establish rules about the intake of certain foods, which foods are to be considered pure or impure, the times established of fasting, etc. [22]. It should not be forgotten that nutritional interventions targeting pregnant migrants should also consider the symbolic nature of food [23].

Pregnancy is thus a highly vulnerable period in terms of nutrition, as well as a motivating stage in life for modifying eating behaviors, and pregnant women sometimes have poor nutritional knowledge. For these reasons, healthcare and teaching staff need a useful instrument for measuring their level of nutritional knowledge and their nutritional beliefs and habits.

Different studies aim to assess nutritional knowledge and beliefs during pregnancy and provide useful information on nutritional knowledge and beliefs [3,24]. However, these studies focus on conducting such assessments using methods which have already been validated. In contrast, the present study also presents the appropriate tools to validate a questionnaire, which is why this study uses a more innovative method.

Having this new tool available would be very useful for healthcare providers in identifying the gaps in nutritional knowledge and the unhealthy eating habits and misconceptions held by pregnant women from different cultures. This tool will thus make it possible to design nutritional education strategies to promote healthy eating behaviors while taking into account socio-cultural aspects [25,26], eating habits, and levels of education [27].

The objective of this study is to present the content validation process, using expert judgement, of an instrument for measuring two dimensions. The first dimension is nutritional knowledge and the nutritional beliefs and the second is eating habits of pregnant women from different cultures in the city of Melilla, Spain, where the crude birth rate was 15.95 births per 1000 inhabitants in 2018 [28] compared to 7.86 at the national level [29].

2. Materials and Methods

This is a descriptive, psychometric study on content validity through expert judgement which was conducted at the Melilla Campus of the University of Granada (Spain).

2.1. Sample

Convenience and intentional sampling were used. The participants were 14 doctors (PhD degree holders) from the University of Granada (Spain), with a mean work experience (in research and teaching) of 14.92 years (SD: 10.37 years) and academically trained in the fields of educational psychology, language and literature didactics, research and diagnosis methods in education, nursing, obstetrics and gynecology, and nutrition and food science (Table 1).

Table 1. Judges, field of expertise, academic training, and work experience.

| Judge | Field of Expertise/Academic Training | Work Experience (Years) |
|-------|---|-------------------------|
| 1 | Educational psychology | 22 |
| 2 | Educational psychology | 5 |
| 3 | Language and literature didactics | 11 |
| 4 | Language and literature didactics | 19 |
| 5 | Research and diagnosis methods in education | 9 |
| 6 | Research and diagnosis methods in education | 18 |
| 7 | Nursing | 32 |
| 8 | Nursing | 10 |
| 9 | Nursing | 14 |
| 10 | Nursing | 6 |
| 11 | Obstetrics and gynecology | 3 |
| 12 | Obstetrics and gynecology | 23 |
| 13 | Nutrition and food science | 35 |
| 14 | Nutrition and food science | 2 |

2.2. Instrument

Many pregnant women have no knowledge of the recommended guidelines for weight gain [30] or when to start taking folic acid [31]. Pregnant women generally have limited knowledge of dietary guidelines for eating healthily during pregnancy [32], which is why it is important for pregnant women to be aware of the issues raised in this study. With regard to the selection of questions in the questionnaire, it should be noted that the questions have been designed with the aim of assessing whether pregnant women are aware of the most relevant aspects of their nutrition and the proper development of their pregnancy. To this end, we have taken into account aspects dealt with in maternal education courses, criteria of interest included in dietary recommendations or food guides for an adequate nutritional status in pregnant women, and further consulted literature [2,14,16,17,33]. The content questions regarding nutritional knowledge, beliefs, and habits were selected after conducting a literature review using the PubMed and Web of Science databases, as well as guidelines within the framework of food institutions. After this search, 20 questions were proposed and scored according to four categories.

The questionnaire “Nutritional knowledge, beliefs, and habits during pregnancy” (NKBHP) consists of two parts or dimensions (nutritional knowledge and nutritional beliefs/habits) with 10 items each. Each item was assessed following the “Template for assessing content validity through expert judgement” developed by Escobar-Pérez and Cuervo-Martínez [9], which establishes four levels (“does not meet the criterion,” “low level,” “moderate level,” and “high level”) for each one of the characteristics assessed. These characteristics are sufficiency, clarity, coherence, and relevance (Table 2). The indicator “one” was assigned when the item did not conform to the category, up to indicator “four”, which was assigned when the item fully conformed to the category (only sufficiency was scored by dimension rather than by item). The experts’ qualitative observations for each of the twenty items that made up the initial instrument were also taken into account.

Table 2. Categories and indicators used by the judges to validate the tool.

| Categories | Indicators |
|--|--|
| Sufficiency The items within the same dimension suffice to measure this dimension | The items are sufficient to measure the dimension |
| | The items measure some aspects of the dimension, but do not represent the full dimension |
| | A few items must be added in order to fully assess the dimension |
| | The items are insufficient |

Table 2. Cont.

| Categories | Indicators |
|---|--|
| Clarity The item can be understood easily, i.e., syntax and semantics are appropriate | The item is unclear |
| | The wording of the item requires several modifications or a very large modification in terms of meaning or word order |
| Coherence The item is logically related to the dimension or indicator it is measuring | Some of the terms in the item require very precise modificationsThe item is clear, with appropriate semantics and syntax |
| | The item bears no logical relationship to the dimension |
| | The item has a tangential relationship to the dimension |
| | The item has a moderate relationship to the dimension it is measuring |
| Relevance The item is essential or important, i.e., it must be included | The item is completely related to the dimension it is measuring |
| | The removal of the item would not affect the measurement of the dimension |
| | The item is somewhat relevant, but another item may be covering what this item is measuring |
| | The item is rather important |
| | The item is very relevant and should be included |

Source: adapted from Escobar-Pérez and Cuervo-Martínez [9] (p. 37).

2.3. Statistical Analysis

The SPSS Statistics 24.0 software was used for data analysis. The degree of agreement among the experts was determined using Fleiss' κ , as this is an analytical statistic that makes it possible to assess the degree of agreement among three or more raters who independently judge a series of items using an instrument with a certain number of ordinal categories [34,35]. The minimum value assumed by this coefficient is 0 and the maximum value is 1. The scale produced by Landis and Koch [36], which quantitatively expresses the strength of agreement among observers, was used for the interpretation of Fleiss' κ values (Table 3).

Table 3. Fleiss' κ values and strength of agreement [36].

| Fleiss' κ | Strength of Agreement |
|------------------|-----------------------|
| 0.00 | Poor |
| 0.1–0.20 | Slight |
| 0.21–0.40 | Fair |
| 0.41–0.60 | Moderate |
| 0.61–0.80 | Substantial |
| 0.81–1.00 | Almost perfect |

2.4. Procedure

The sample was selected using convenience (or affinity) sampling. The experts participated voluntarily and signed the informed consent form. All of them were experts in areas that could contribute to improving both the content, procedural, and wording aspects of the questionnaire. A cover letter was sent to the judges by email with acknowledgement of receipt alongside the questionnaire to be validated. This letter contained information on the main objective of this study and how to respond and assured the experts of the confidentiality of their data. In order for the experts to evaluate a certain number of items, both the amount of information and the way in which it is presented are important [37]. These aspects were therefore all taken into account. The researchers were available at all times to answer any questions the experts might have had. The judges sent their signed informed

consent forms by mail and were given one month to assess and rate the questionnaire online. All the experts who were sent a request agreed to participate. No reminders had to be sent to them, as they all responded within the deadline.

Once the experts had assessed the questionnaire, the resulting instrument was subjected to the standard pre-test using potential respondents in order to have information on how it would work in real life [10]. To this end, a dichotomous yes/no response was encoded in each of the items to measure their degree of understandability and applicability/feasibility. The sample consisted of 102 women of childbearing age from various cultures and religions (50% Muslim, 46% Christian, 3% Jewish, and 1% Hindu) from the city of Melilla (Spain). This distribution was based on a demographic study by the Union of Islamic Communities of Spain [38]. The questionnaires were administered at the healthcare centers and were completed on-site in person, and all the questionnaires were returned. The sampling was incidental, with the participation of women who were visiting the healthcare center for various health-related issues. The time taken by the participants to answer the questionnaire ranged from 5 to 10 min. This was assessed quantitatively. We simply collected the questionnaires and subsequently analyzed the responses. Degrees of understandability were classified as follows: high understandability (equal to or greater than 85%), medium understandability (from 80% to 85%), and low understandability (less than 80%).

2.5. Ethics

This research was conducted in compliance with the ethical principles set out in the Declaration of Helsinki. All participants were informed of the purpose of this study and participated voluntarily, having signed an informed consent form. The knowledge and approval by management of the Comarcal Hospital of Melilla, on which the Unit for Attention to Women depends, was assured.

3. Results

3.1. Content Validation by Expert Judgement

For the evaluation of the original instrument, the proportion of possible agreements occurring in each dimension was taken into account in the calculation of Fleiss’ κ . The magnitude of the strength of agreement was considered to be *almost perfect* for both dimensions by the set of judges, as shown in Table 4.

Table 4. Strength of agreement among judges for the dimensions of the original instrument.

| Dimensions | Fleiss’ κ | Strength of Agreement (Landis and Koch, 1977) |
|--------------------|------------------|---|
| Knowledge | 0.860 | Almost perfect |
| Beliefs and habits | 0.830 | Almost perfect |

The magnitude of the strength of agreement by pairs of experts was also analyzed. Values corresponding to “substantial” and “almost perfect” were found, as shown in Table 5.

Table 5. Agreement by pairs of experts.

| Dimensions | Fleiss’ κ —Agreement by Pairs of Experts | | | | | | | | | | | | | |
|--------------------|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| | 1–14 | 2–13 | 3–12 | 4–11 | 5–10 | 6–9 | 7–8 | 8–6 | 9–5 | 10–4 | 11–3 | 12–2 | 13–1 | 14–7 |
| Knowledge | 0.929 | 0.868 | 0.763 | 0.885 | 0.717 | 0.920 | 1 | 0.830 | 0.811 | 0.785 | 0.900 | 0.735 | 0.889 | 1 |
| Beliefs and habits | 1 | 0.984 | 0.711 | 0.700 | 0.732 | 0.833 | 0.931 | 0.846 | 0.744 | 0.714 | 0.706 | 0.708 | 0.949 | 1 |

In addition, the characteristics of the instrument regarding the indicators of sufficiency, clarity, coherence, and relevance were assessed using the ordinal measurement scale. A strength of agreement between “substantial” and “almost perfect” was found (Table 6), with “relevance” having the highest

values in both dimensions (0.890 in knowledge and 0.901 in habits) based on the degree of overall agreement among the judges.

The statistical significance threshold for the results was set at $p < 0.05$, with a 95% confidence interval for all cases. Agreement on the characteristic of “relevance” was found to be statistically significant ($p < 0.001$) for both dimensions. Agreement on “sufficiency” was also found to be statistically significant ($p < 0.001$) for the dimension of nutritional beliefs and habits.

Table 6. Fleiss’ κ and statistical significance of the characteristics of the original instrument.

| Dimensions | Characteristics | Fleiss’ κ | p |
|--------------------|-----------------|------------------|--------|
| Knowledge | Sufficiency | 1 | 0.001 |
| | Clarity | 0.805 | 0.002 |
| | Coherence | 0.795 | 0.030 |
| | Relevance | 0.890 | <0.001 |
| Beliefs and habits | Sufficiency | 1 | <0.001 |
| | Clarity | 0.780 | 0.037 |
| | Coherence | 0.847 | 0.007 |
| | Relevance | 0.901 | <0.001 |

These results, together with the qualitative observations and recommendations made by the judges on the items included in the two dimensions, made it possible to keep the original number of items at 20. However, the wording of five of the items was amended, resulting in the final validated instrument.

3.2. Measurement of Applicability: Pre-Test

The final validated instrument was administered to a total of 102 women of childbearing age to determine the percentage of comprehensibility of the dimensions and their corresponding items. The degree of comprehensibility of the instrument was found to be in the highest range, at 99.7%, as shown in Table 7.

Table 7. Percentages of comprehensibility of the dimensions and their items in the final version of the validated instrument.

| Dimensions | Items | Degree of Comprehensibility (%) | |
|--------------------|--|---------------------------------|----|
| | | Yes | No |
| Knowledge | Maximum weight gain | 98 | 2 |
| | Initiation of folic acid consumption | 100 | 0 |
| | What folic acid is for | 100 | 0 |
| | Iron | 99 | 1 |
| | What iron is for | 100 | 0 |
| | Instructions from healthcare personnel | 100 | 0 |
| | What to do during pregnancy | 100 | 0 |
| | Fiber during pregnancy | 100 | 0 |
| | Salt during pregnancy | 99 | 1 |
| | Fluids | 98 | 2 |
| Beliefs and habits | Awareness | 100 | 0 |
| | Number of meals and times | 100 | 0 |
| | Influence | 100 | 0 |
| | Time spent eating | 100 | 0 |
| | Most hungry | 100 | 0 |
| | Diet | 100 | 0 |
| | Good eating habits | 100 | 0 |
| | Importance of food | 100 | 0 |
| | Harmful foods | 100 | 0 |
| | Beneficial foods | 100 | 0 |

4. Discussion

Nutrition and health professionals and researchers need valid and reliable behavioral measures that are appropriate for use in a variety of community settings [39]. The validation of an instrument is an on-going and dynamic process that becomes more consistent the more psychometric properties are determined for that particular instrument in different contexts and populations. Validation will also be determined by the type and purpose of the instrument. In this case, where the aim is to collect factual information related to the knowledge and practices of certain subjects, content validity by experts takes priority [4,40].

The content validity of an instrument refers to the degree to which this instrument covers an adequate sample of the contents it is intended to cover, without omissions, oversights, or imbalances [41]. However, an instrument does not have to cover in detail each of the areas that make up a concept, as this would result in an overly large instrument. The instrument must therefore contain a representative sample of domains and possible issues relating to the concept of interest [42]. The twenty items of the questionnaire presented here include the most relevant aspects for determining the nutritional knowledge, beliefs, and habits of pregnant women from different cultures.

Even though ensuring the content validity of an instrument may seem to be time consuming and costly in terms of human resources, it deserves greater attention when developing a valid assessment instrument [43]. The evaluation technique of expert judgement can be very useful for the validation of diagnostic instruments but requires the correct selection of experts [10].

Determining the number of experts that should be involved in the content validation is one of the main difficulties to be addressed, as there is no widespread consensus on this [11]. Availability and level of knowledge on the subject matter of the research are some of the criteria used to establish the sufficient number of experts [10]. However, the appropriate number of experts will depend on the method used. Some methods are designed to measure agreement between two judges [37]. Other methods require a higher number of experts, between 7 and 30 [44,45]. Rubio et al. [46] propose a range of 6 to 20 experts and establish that using a greater number of experts may generate more information on the measure in question. In general, many authors recommend more than 10 experts [12,47–49]. As a result, a total of 14 experts were selected for this study.

With respect to experience, it is recommended that at least two of the judges be measurement and evaluation experts [9]. The current study includes two experts in the field of research and diagnosis methods in education.

The selection of experts is another important consideration. There are various procedures for this, such as structured procedures including selection criteria (e.g., graphical biographies and competence coefficients), and unstructured procedures without selection filters, e.g., the closeness or affinity of the researchers to the judges [10]. The latter procedure, the closeness (or affinity) of the researchers to the experts, has been used in this research.

In this study, experts were selected on published criteria while considering a procedure that ensures the assertiveness of their assessments. The criteria were the following: the judges' experience in issuing judgements and decision-making; their academic and scientific reputation; their willingness and motivation to collaborate; their objectivity; their compliance with what has been established [9]; and their ability to perform the question classification techniques required to validate the content [50]. Following this procedure, experts complying with these characteristics were sought to prevent introducing content bias in the analysis of the data.

The quality of the results in a study using expert judgement is strongly related to the experts selected. Therefore, using a good selection procedure is of paramount importance [51]. Ténière-Buchot [52] reports that there are three types of experts: tactical experts, conciliatory experts, and communicative experts. Tactical experts are selected on the basis of their experience and knowledge of the subject matter. Conciliatory experts are selected for their objectivity and common sense. Communicative experts are the experts who are most involved in the study. In this case, tactical experts or specialists

were included, since, according to Ténrière-Buchot [52], specialists in the field ensure a higher scientific quality of the study.

Another aspect to consider in content validation through expert judgement is the amount of time given to judges to make their judgement [5]. In this study, a one-month deadline was established for each of the judges to analyze the weaknesses and strengths of the instrument and submit their opinion online.

Content validation requires the participation of both researchers and members of the target population [40]. The current study involved both experts who are linked to the research field and to the methodological aspects of the instrument, as well as potential members of the target population, since it was women of childbearing age who underwent the pre-test procedure. Recently published studies use this procedure for the content validation of instruments. In these studies, in addition to the expert phase, the instrument designed is subsequently subjected to a pre-test where a focus group assesses each item for clarity taking into account the level of understandability of the instrument [53]. Similarly, in a study by Bernal-García et al. [54] a pre-test was subsequently conducted to measure the degree of understandability of the instrument, which turned out to be high-ranking, as in this study.

Regarding the statistical analyses to calculate the agreement between the judges, the κ statistic and Kendall's coefficient are the most widely used [9]. In this study, the κ statistic was used, as it provides quantifiable methods to assess judgments on content and has the additional possibility of eliminating random chance agreement [55]. The κ statistic can be used to assess the degree of agreement at the individual level [39], although this is not how it is used in this study.

Given that there were multiple raters in this study, the Fleiss' κ statistic was used, as it is based on the agreement between different pairs of raters, which increases the accuracy of the results [35,56–58], unlike the weighted Cohen's κ statistic, which is used for nominal variables in the case of two raters [59].

For the questionnaire analyzed, the Fleiss' κ statistic yielded an "almost perfect" strength of agreement for each dimension and a strength of agreement between "substantial" and "almost perfect" for pairs of experts. Similar results were obtained by Bernal-García et al. [54]. However, in this study, the strength of agreement by pairs of experts attained a somewhat lower level, between "moderate" and "almost perfect".

The rating of the relevance, clarity, simplicity, and ambiguity of items using four-point scales is something that has been going on for years for content validation [33,43,60,61]. The same parameters were considered for this study. However, instead of simplicity, sufficiency was considered, as indicated in a study by Escobar-Pérez and Cuervo-Martínez [9].

In healthcare research, many relevant results and variables of interest are abstract concepts known as theoretical constructs. The use of valid and reliable instruments to measure such constructs is an essential component of the quality of research [62].

Several studies enquire whether pregnant women have received information on healthy eating habits during pregnancy [3,24,63,64]. This question has been included in the present questionnaire to obtain this information, which is considered to be important when assessing nutritional knowledge.

Other studies determine dietary knowledge and beliefs during pregnancy using food consumption surveys [65–67]. In the case of the present study, the aim is to validate a questionnaire based on nutritional knowledge and beliefs and not just based on food consumption with a quantitative approach.

There are numerous methods for measuring food-related aspects. However, almost all of them focus on studying food consumption from a quantitative approach. If this quantitative information is combined with the assessment of eating behaviors, such as nutritional beliefs and habits, the result would be a more complete study of the eating process in the different subjects, which could facilitate making nutritional educational recommendations [68].

Rather than in designing new tools, there is now interest in establishing indices that provide information on specific behavioral patterns associated with eating habits and socio-cultural habits, as well as on nutrients and foods consumed [69]. Culture is therefore an important factor to take into account when it comes to understanding eating behaviors [70,71].

As for the limitations of this study, it is worth noting that the participants in the pre-test were women of childbearing age and not pregnant women. As a consequence, this test cannot be interpreted in its real context, and as these were quantitative questions, no in-depth qualitative data could be obtained from these participants.

With respect to content validity through expert judgement, there are aspects that the researchers cannot control for, such as the complexity or degree of difficulty of the task. It should be noted that even when a test receives a very good rating from the experts, it must be continually reviewed and improved [9]. Furthermore, this process requires considerable attention and formal methods for the selection of experts, for the use of a scale that facilitates quantitative assessments, and for the analysis of the results using relevant coefficients. In general, validation processes through expert judgement prove to be demanding and time consuming.

5. Conclusions

Although content validity is subjective, it can add objectivity to the study by using statistics such as Fleiss' κ , which is very useful for measuring agreement between experts and thus to be able to validate the instrument correctly. Complementing this with the pre-test procedure described also facilitates the determination of the degree of comprehensibility of the final instrument. Understanding the need for and the process of conducting content validation studies is important for healthcare professionals and researchers. Having a guide in place may prove very helpful. We believe the objective of validating a measurement instrument has been met. This instrument may be used in different populations of pregnant women to determine their nutritional knowledge, beliefs, and habits using new psychometric tests. These tests will increase the validity of the instrument and favor comparisons between different populations in Spain and in other Spanish-speaking countries. This in turn will establish educational measures to ensure adequate eating behaviors among pregnant women and thus prevent negative health consequences for both the mother and the future baby. It is therefore considered of great utility to have reliable tools available that may be targeted at different cultures, since culture influence eating behaviors of the population in general and of pregnant women in particular.

Author Contributions: Conceptualization, E.F.-G.; Formal analysis, C.E.-M.; Methodology, C.E.-M.; Software, M.A.S.-O. and S.N.-P.; Supervision, C.E.-M.; Visualization, A.M.-S., T.L.-V., M.A.S.-O. and S.N.-P.; Writing—original draft, E.F.-G.; Writing—review & editing, E.F.-G., A.M.-S., T.L.-V. and C.E.-M. All authors reviewed and confirmed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Cuervo, M.; Sayon-Orea, C.; Santiago, S.; Martínez, J. Dietary and health profiles of Spanish women in preconception, pregnancy and lactation. *Nutrients* **2014**, *6*, 4434–4451. [[CrossRef](#)]
2. Morrison, J.L.; Regnault, T.R.H. Nutrition in Pregnancy: Optimising Maternal Diet and Fetal Adaptations to Altered Nutrient Supply. *Nutrients* **2016**, *8*, 342. [[CrossRef](#)] [[PubMed](#)]
3. Lee, A.; Newton, M.; Radcliffe, J.; Belski, R. Pregnancy nutrition knowledge and experiences of pregnant women and antenatal care clinicians: A mixed methods approach. *Women Birth*. **2018**, *31*, 269–277. [[CrossRef](#)] [[PubMed](#)]
4. Carvajal, A.; Centeno, C.; Watson, R.; Martínez, M.; Sanz Rubiales, Á. ¿Cómo validar un instrumento de medida de la salud? *An. Sist. Sanit. Navar.* **2011**, *34*, 63–72. [[CrossRef](#)] [[PubMed](#)]
5. Galicia, L.A.; Balderrama, J.A.; Edel, R. Validez de contenido por juicio de expertos: Propuesta de una herramienta virtual. *Apertura* **2017**, *9*, 42–53.
6. Haynes, S.N.; Richard, D.C.S.; Kubany, E.S. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychol. Assess.* **1995**, *7*, 238–247. [[CrossRef](#)]
7. González, C.G.Z.; Aguilera, P.C. Instrumentos de evaluación: ¿qué piensan los estudiantes al terminar la escolaridad obligatoria? *Perspect. Educ.* **2014**, *53*, 57–72. [[CrossRef](#)]

8. Sireci, S.G. The Construct of Content Validity. *Soc. Indic. Res.* **1998**, *45*, 83–117. [[CrossRef](#)]
9. Escobar-Pérez, J.; Cuervo-Martínez, A. Validez de contenido y juicio de expertos: Una aproximación a su utilización. *Av. Med.* **2008**, *6*, 27–36.
10. Cabero, J.; Almenara, J.C. La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información y comunicación (TIC). *Eduweb* **2013**, *7*, 11–22.
11. Garrote, P.R.; Rojas, M.D.C. La validación por juicio de expertos: Dos investigaciones cualitativas en Lingüística aplicada. *Rev. Nebrija Lingüísti. Apl. Enseñ. Leng.* **2015**, *18*, 124–139. [[CrossRef](#)]
12. Juárez-Hernández, L.G.; Tobón, S. Análisis de los elementos implícitos en la validación de contenido de un instrumento de investigación. *Rev. Espac.* **2018**, *39*, 23.
13. Das, J.K.; Salam, R.A.; Thornburg, K.L.; Prentice, A.M.; Campisi, S.; Lassi, Z.S.; Koletzko, B.; Bhutta, Z.A. Nutrition in adolescents: Physiology, metabolism, and nutritional needs. *Ann. N. Y. Acad. Sci.* **2017**, *1393*, 21–33. [[CrossRef](#)] [[PubMed](#)]
14. Most, J.; Dervis, S.; Haman, F.; Adamo, K.B.; Redman, L.M. Energy Intake Requirements in Pregnancy. *Nutrients* **2019**, *11*, 1812. [[CrossRef](#)]
15. Langley-Evans, S.C. Nutrition in early life and the programming of adult disease: A review. *J. Hum. Nutr. Diet.* **2015**, *28*, 1–14. [[CrossRef](#)]
16. Hu, Z.; Tylavsky, F.A.; Kocak, M.; Fowke, J.H.; Han, J.C.; Davis, R.L.; LeWinn, K.Z.; Bush, N.R.; Sathyanarayana, S.; Karr, C.J.; et al. Effects of Maternal Dietary Patterns during Pregnancy on Early Childhood Growth Trajectories and Obesity Risk: The CANDLE Study. *Nutrients* **2020**, *12*, 465. [[CrossRef](#)]
17. Chakona, G.; Shackleton, C. Food Taboos and Cultural Beliefs Influence Food Choice and Dietary Preferences among Pregnant Women in the Eastern Cape, South Africa. *Nutrients* **2019**, *11*, 2668. [[CrossRef](#)]
18. Woolf, S.H.; Purnell, J.Q. The Good Life: Working Together to Promote Opportunity and Improve Population Health and Well-being. *JAMA* **2016**, *315*, 1706–1708. [[CrossRef](#)]
19. Guertin, C.; Pelletier, L.; Pope, P. The validation of the Healthy and Unhealthy Eating Behavior Scale (HUEBS): Examining the interplay between stages of change and motivation and their association with healthy and unhealthy eating behaviors and physical health. *Appetite* **2020**, *144*, 104487. [[CrossRef](#)]
20. Verbeke, W.; De Bourdeaudhuij, I. Dietary behaviour of pregnant versus non-pregnant women. *Appetite* **2007**, *48*, 78–86. [[CrossRef](#)]
21. Contreras, J. Alimentación y religión. *Humanit. Humanid. Med.* **2007**, *16*, 1–31.
22. Amérigo, F. La problemática de la alimentación religiosa y de convicción en los centros educativos. *Rev. Derecho Polit.* **2016**, *97*, 141–178. [[CrossRef](#)]
23. Hunter-Adams, J.; Rother, H.A. Pregnant in a foreign city: A qualitative analysis of diet and nutrition for cross-border migrant women in Cape Town, South Africa. *Appetite* **2016**, *103*, 403–410. [[CrossRef](#)] [[PubMed](#)]
24. Okesene-Gafa, K.; Chelimo, C.; Chua, S.; Henning, M.; McCowan, L. Knowledge and beliefs about nutrition and physical activity during pregnancy in women from South Auckland region, New Zealand. *Aust. N. Z. J. Obstet. Gynaecol.* **2016**, *56*, 471–483. [[CrossRef](#)] [[PubMed](#)]
25. Díaz-Méndez, C.; García-Espejo, I. Eating out in Spain: Motivations, sociability and consumer contexts. *Appetite* **2017**, *119*, 14–22. [[CrossRef](#)] [[PubMed](#)]
26. Quintero-Angel, M.; Mendoza, D.M.; Quintero-Angel, D. The cultural transmission of food habits, identity, and social cohesion: A case study in the rural zone of Cali-Colombia. *Appetite* **2019**, *139*, 75–83. [[CrossRef](#)]
27. Daniels, S.; Glorieux, I. Convenience, food and family lives. A socio-typological study of household food expenditures in 21st-century Belgium. *Appetite* **2015**, *94*, 54–61. [[CrossRef](#)]
28. Spanish National Institute of Statistics [Instituto Nacional de Estadística]. *Estadísticas Territoriales. Melilla. IDB. Tasa Bruta de Natalidad*; INE: Madrid, Spain, 2018.
29. Spanish National Institute of Statistics [Instituto Nacional de Estadística]. *Indicadores Demográficos Básicos. Tasa Bruta de Natalidad. Datos Provisionales 2018*; INE: Madrid, Spain, 2018.
30. Shub, A.; Huning, E.Y.S.; Campbell, K.J.; McCarthy, E.A. Pregnant women’s knowledge of weight, weight gain, complications of obesity and weight management strategies in pregnancy. *BMC Res. Notes* **2013**, *6*, 278. [[CrossRef](#)]
31. Conlin, M.L.; MacLennan, A.H.; Broadbent, J.L. Inadequate compliance with periconceptional folic acid supplementation in South Australia. *Aust. N. Z. J. Obstet. Gynaecol.* **2006**, *46*, 528–533. [[CrossRef](#)]

32. Lee, A.; Belski, R.; Radcliffe, J.; Newton, M. What do Pregnant Women Know about the Healthy Eating Guidelines for Pregnancy? A Web-Based Questionnaire. *Matern. Child Health J.* **2016**, *20*, 2179–2188. [[CrossRef](#)]
33. Bukenya, R.; Ahmed, A.; Andrade, J.M.; Grigsby-Toussaint, D.S.; Muyonga, J.; Andrade, J.E. Validity and Reliability of General Nutrition Knowledge Questionnaire for Adults in Uganda. *Nutrients* **2017**, *9*, 172. [[CrossRef](#)] [[PubMed](#)]
34. Falotico, R.; Quatto, P. Fleiss' kappa statistic without paradoxes. *Qual. Quant.* **2015**, *49*, 463–470. [[CrossRef](#)]
35. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Med.* **2012**, *22*, 276–282. [[CrossRef](#)]
36. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
37. Pedrosa, I.; Suárez-Álvarez, J.; García-Cueto, E. Evidencias Sobre La Validez De Contenido: Avances Teóricos Y Métodos Para Su Estimación/Content Validity Evidences: Theoretical Advances and Estimation Methods. *Acción Psicol.* **2013**, *10*, 3–18. [[CrossRef](#)]
38. Unión de Comunidades Islámicas de España. *Estudio Demográfico de la Población Musulmana. Explotación Estadística del Censo de Ciudadanos Musulmanes en España Referido a Fecha 31/12/2018*; UCIDE: Madrid, Spain, 2019.
39. Mainvil, L.A.; Horwath, C.C.; McKenzie, J.E.; Lawson, R. Validation of brief instruments to measure adult fruit and vegetable consumption. *Appetite* **2011**, *56*, 111–117. [[CrossRef](#)]
40. Pallas, J.M.A.; Villa, J.J. *Métodos de Investigación Clínica Y Epidemiológica*; Elsevier Health Sciences: Amsterdam, The Netherlands, 2019.
41. Streiner, D.L.; Norman, G.R.; Cairney, J. *Health Measurement Scales: A Practical Guide to Their Development and Use*; Oxford University Press: Oxford, UK, 2015. [[CrossRef](#)]
42. Alarcon, M.A.M.; Muñoz, N.S. Medición en salud: Algunas consideraciones metodológicas. *Rev. Méd. Chile* **2008**, *136*, 125–130. [[CrossRef](#)]
43. Yaghmaie, F. Content validity and its estimation. *J. Med. Educ.* **2003**, *3*. [[CrossRef](#)]
44. Urrutia, M.; Barrios, S.; Gutiérrez, M.; Mayorga, M. Métodos óptimos para determinar validez de contenido. *Educ. Med. Super.* **2014**, *28*, 547–558.
45. Varela-Ruiz, M.; Díaz-Bravo, L.; García-Durán, R. Descripción y usos del método Delphi en investigaciones del área de la salud. *Investig. Educ. Med.* **2012**, *1*, 90–95.
46. Rubio, D.M.; Berg-Weger, M.; Tebb, S.S.; Lee, E.S.; Rauch, S. Objectifying content validity: Conducting a content validity study in social work research. *Soc. Work Res.* **2003**, *27*, 94–104. [[CrossRef](#)]
47. García, A.; Antúnez, A.; Ibáñez, S.J. Análisis del proceso formativo en jugadores expertos: Validación de instrumento. *Rev. Int. Med. Cienc. Act. Fís. Deporte* **2016**, *16*, 157–182. [[CrossRef](#)]
48. Hyrkäs, K.; Appelqvist-Schmidlechner, K.; Oksa, L. Validating an instrument for clinical supervision using an expert panel. *Int. J. Nurs. Stud.* **2003**, *40*, 619–625. [[CrossRef](#)]
49. Jiménez, J.; Salazar, W.; Morera, M. Diseño y validación de un instrumento para la evaluación de patrones básicos de movimiento. *Eur. J. Hum. Mov.* **2013**, *31*, 87–97.
50. Garrido, M.E.; Romero, S.; Ortega, E.; Zagalaz, M.L. Designing and validation of a questionnaire on parents for children in sport. *J. Sport Health Res.* **2011**, *3*, 59–70.
51. Blasco, J.E.; López, A.; Mengual-Andrés, S. Validación mediante el metodo Delphi de un cuestionario para conocer las experiencias e interés hacia las actividades acuáticas con especial atención al Winsurf. *Ágora Educ. Fís. Deporte* **2010**, *12*, 75–94.
52. Ténrière-Buchot, P.F. *Décision, expertise, arbitraire et transparence: Éléments d'un développement durable. Le Courrier de l'Environnement de l'INRA*; Institut National de la Recherche Agronomique Délégation Permanente à l'Environnement: Paris, France, 2001; pp. 41–52.
53. Da Silva, A.F.; Coelho de Almeida, R.C.; Assunção, R.B.; Zandonadi, R.P. Good Practices in Home Kitchens: Construction and Validation of an Instrument for Household Food-Borne Disease Assessment and Prevention. *Int. J. Environ. Res. Salud Pública* **2019**, *16*, 1005. [[CrossRef](#)]
54. Bernal-García, M.I.; Salamanca, D.R.; Perez, N.; Quemba, M.P. Validez de contenido por juicio de expertos de un instrumento para medir percepciones físico-emocionales en la práctica de disección anatómica. *Educ. Med.* **2018**. [[CrossRef](#)]
55. Wynd, C.A.; Schmidt, B.; Schaefer, M.A. Two Quantitative Approaches for Estimating Content Validity. *West. J. Nurs. Res.* **2003**, *25*, 508–518. [[CrossRef](#)]

56. Álvarez, D.A.R.; Díaz, L.C. Validity and Reliability of the Spanish Version of the Technological Competency as Caring in Nursing Instrument. *Investig. Educ. Enferm.* **2017**, *35*, 154–164. [[CrossRef](#)]
57. Garcia-Esteve, L.; Torres, A.; Navarro, P.; Ascaso, C.; Imaz, M.L.; Herreras, Z.; Valdés, M. Validation and comparison of four instruments to detect partner violence in health-care setting. *Med. Clin.* **2011**, *137*, 390–397. [[CrossRef](#)] [[PubMed](#)]
58. Kempen, T.G.H.; Hedström, M.; Olsson, H.; Johansson, A.; Ottosson, S.; Al-Sammak, Y.; Gillespie, U. Assessment tool for hospital admissions related to medications: Development and validation in older patients. *Int. J. Clin. Pharm.* **2019**, *41*, 198–206. [[CrossRef](#)] [[PubMed](#)]
59. Marasini, D.; Quatto, P.; Ripamonti, E. Assessing the inter-rater agreement for ordinal data through weighted indexes. *Stat. Methods Med. Res.* **2016**, *25*, 2611–2633. [[CrossRef](#)] [[PubMed](#)]
60. Dongare, P.A.; Bhaskar, S.B.; Harsoor, S.; Kalaivani, M.; Garg, R.; Sudheesh, K.; Goneppanavar, U. Development and validation of a questionnaire for a survey on perioperative fasting practices in India. *Indio J. Anaesth.* **2019**, *63*, 394–399. [[CrossRef](#)]
61. Emmanuel, A.; Clow, S.E. A questionnaire for assessing breastfeeding intentions and practices in Nigeria: Validity, reliability and translation. *BMC Pregnancy Childbirth* **2017**, *17*, 174. [[CrossRef](#)]
62. Kimberlin, C.L.; Winterstein, A.G. Validity and reliability of measurement instruments used in research. *Am. J. Health Syst. Pharm.* **2008**, *65*, 2276–2284. [[CrossRef](#)]
63. De Jersey, S.J.; Nicholson, J.M.; Callaway, L.K.; Daniels, L.A. An observational study of nutrition and physical activity behaviours, knowledge, and advice in pregnancy. *BMC Pregnancy Childbirth* **2013**, *13*, 115. [[CrossRef](#)]
64. Aktaç, S.; Sabuncular, G.; Kargin, D.; Gunes, F.E. Evaluation of Nutrition Knowledge of Pregnant Women before and after Nutrition Education according to Sociodemographic Characteristics. *Ecol. Food Nutr.* **2018**, *57*, 441–455. [[CrossRef](#)]
65. Sam, C.H.; Skeaff, S.; Skidmore, P.M. A comprehensive FFQ developed for use in New Zealand adults: Reliability and validity for nutrient intakes. *Public Health Nutr.* **2014**, *17*, 287–296. [[CrossRef](#)]
66. Wang, K.; Xie, Y.; Wang, D.; Bishop, N.J.; Tooker, E.M.; Li, Z. Socioeconomic correlates of adherence to mineral intake recommendations among pregnant women in north China: Findings from a cross-sectional study. *Asia Pac. J. Clin. Nutr.* **2020**, *29*, 127–135. [[CrossRef](#)]
67. Tayyem, R.F.; Allehdan, S.S.; Alatrash, R.M.; Asali, F.F.; Bawadi, H.A. Adequacy of Nutrients Intake among Jordanian Pregnant Women in Comparison to Dietary Reference Intakes. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3440. [[CrossRef](#)] [[PubMed](#)]
68. Márquez-Sandoval, Y.F.; Salazar-Ruiz, E.N.; Macedo-Ojeda, G.; Altamirano-Martínez, M.B.; Bernal-Orozco, M.F.; Salas-Salvador, J.; Vizmanos-Lamotte, B. Diseño y validación de un cuestionario para evaluar el comportamiento alimentario en estudiantes mexicanos del área de la salud. *Nutr. Hosp.* **2014**, *30*, 153–164. [[CrossRef](#)] [[PubMed](#)]
69. Gil, Á.; Martínez de Victoria, E.; Olza, J. Indicators for the evaluation of diet quality. *Nutr. Hosp.* **2015**, *31*, 128–144. [[CrossRef](#)] [[PubMed](#)]
70. Rouche, M.; de Clercq, B.; Lebacq, T.; Dierckens, M.; Moreau, N.; Desbouys, L.; Godin, I.; Castetbon, K. Socioeconomic Disparities in Diet Vary According to Migration Status among Adolescents in Belgium. *Nutrients* **2019**, *11*, 812. [[CrossRef](#)]
71. Ngongalah, L.; Rankin, J.; Rapley, T.; Odeniyi, A.; Akhter, Z.; Heslehurst, N. Dietary and Physical Activity Behaviours in African Migrant Women Living in High Income Countries: A Systematic Review and Framework Synthesis. *Nutrients* **2018**, *10*, 1017. [[CrossRef](#)]

