Achieving Consensus Deal with Methodological Issues in the Delphi Technique

Zahra Goodarzi ¹, Enayat Abbasi ²* and Homayou Fardadian ³

Abstract

Delphi is a powerful technique used to seek answers to appropriate questions. The purpose of this paper is to provide an overview of the Delphi technique as a research method. This paper discusses the scientific merit of the Delphi technique by investigating on 41 studies of Journal of Agricultural Education from 1981 to 2013, and 2 studies of Journal of Agricultural Science and Technology. The results showed that there is no general agreement on using indexes in different rounds of Delphi technique; however, according to the frequencies of using indexes in different studies, the following suggestions are presented. The favourable number of panel of experts is between 10 to 20. Purposive sampling method is used for selecting the panel members. Usually a three rounds of Delphi method is used. One question is designed in round one. Mean and standard deviation indexes are used for passing from round two to round three and agreement level of 70 present is used for achieving expert's consensus in round three.

Keywords: consensus level, Delphi technique, methodological criticism

¹ PhD. Student of Agricultural Extension and Education, Tarbiat Modares University, Tehran, Iran
² Assistant Professor, Department of Agricultural Extension and Education, Tarbiat Modares University, Tehran, Iran
³ PhD. Department of Agricultural Extension and Education, Tarbiat Modares University, Tehran, Iran
* Corresponding author's email: Enayat.abbasi@modares.ac.ir
**INTRODUCTION**

Delphi method is a qualitative research method that provides a reliable group opinion using expert judgment (Landeta, 2006). A number of different types of ‘Delphi’ studies have been identified. Van Zolingen and Klaassen (2003) described four categorizations: Classical Delphi, policy Delphi, decisions Delphi, and fuzzy Delphi. The classical Delphi is characterised by five features including anonymity, iteration, controlled feedback, statistical group response, and stability in responses among those with expertise on a specific issue. The aim of policy Delphi in this situation is to generate policy alternatives by using a structured public dialogue, and not to reach stability in responses among those with expertise. The decision Delphi is used for decision making on social developments and reality is created by a group of decision-makers. Fuzzy Delphi is a combination of the traditional Delphi method with fuzzy set theory in order to address some of the ambiguity of the Delphi panel consensus (Ishikawa et al., 1993). The fuzzy Delphi is a more advanced version of the Delphi method in that it utilizes triangulation statistics to determine the distance between the levels of consensus within the expert panel.

Due to the nature of the Delphi design, there are some critical methodological issues that force the prudent researcher to view Delphi results with caution (Woudenberg, 1991).

Accordingly, it is difficult to show clear conclusions about paradigmatic assumptions underpinning all Delphi studies, and it is necessary to define a new framework for using this technique.

**What is Delphi method?**

Delphi method is a structured technique, originally developed as a systematic, interactive forecasting method which relies on a panel of experts (Brown, 1968; Delbecq, et al., 1975; Dalkey & Helmer, 1963; Linstone & Turoff, 1975; Sackman, 1974). According to the old rule that “two heads are better than one” (Dalkey, 1972), the Delphi is a structured group communication process (Delbecq et al., 1975; Linstone & Turoff, 1975; Powell, 2003), designed to obtain a consensus from a group of experts. The technique has the benefits of group decision making while preventing the limitations of group decision-making and undesirable interaction effects (Cline, 2000). The Delphi judgment is arrived through sequential questionnaires or ‘rounds’, interspersed with summary and feedback derived from previous panels responses (Delbecq et al., 1975; Linstone & Turoff, 1975; Witkin & Altschuld, 1995). Participants in Delphi panel are stimulated to produce new ideas, which they consider more suitable to solve a problem. The experts answer questionnaires in at least two or more rounds. After each round, the primary researcher provides a summary of the experts’ forecasts from the previous round and also the reasons that they provided for their judgments. Therefore, with responses obtained from other panel’s members, experts are encouraged to revise their earlier answers in light of the replies of other members of their panel. During this process, the range of separate answers will decrease and leads to the group coverage of the correct answer. Finally, the process is stopped after a predefined stop criterion (e.g. achievement of consensus, number of rounds, and stability of results).

The purpose of this study was to demystify Delphi methodology and update knowledge in order to inform future debate. The objectives of the study were as follows:

1. Achieving favourable number of rounds and panellists.
2. Describe favourable indexes for reaching agreement in Delphi studies.
3. Describe favourable Delphi panellists’ level of agreement with the generated competency statements.
4. Describe how to achieve validity and reliability in Delphi studies.

**MATERIALS AND METHODS**

This research is a kind of content analysis, which gives an extensive review of all the studies that had employed the Delphi method and were published in the Journal of Agricultural Education from 1981 to 2013 as well as Journal of Agricultural Science and Technology (two studies). The researchers calculated the total
Achieving Consensus Deal with Methodological ... / Goodarzi et al.

number of research studies that had employed the Delphi method (43 studies). In each study, the researchers elicited all number of rounds, number of questions in the first round, the sampling method and size, as well as how to define consensus and validity and reliability of the questionnaire. After gathering this information, the result for each section was shown in a table and a summary of that section is presented under the table.

Methodological considerations

While there are different Delphi studies that report answers to specific questions about Delphi method, there are less studies dealt with methodological issues such as selecting the panel, survey administration, and other challenges as well as use in evaluation studies (H.-L. Hung et al., 2008). Some researchers give recommendations for improvement and more efficient use of the Delphi (Hasson et al., 2000; Keeney et al., 2006; Lang, 1994; Skulmoski et al., 2007). The following observations are based on a number of concerns founded in the literature.

Sample size/panel size

The size and participant dropout and selection of experts for the panel affect most Delphi studies. Panel selection is vital to the success of the study (Moore, 1987).

There is no established rule for determining the appropriate sample size (Williams & Webb, 1994). Like other research methods, the more participants, the better; however, Powell (2003) points out that the numbers of experts vary according to the nature of the problem and resources available to researchers and as would be expected with larger samples as well as more heterogeneous ones, the complexity of the research would tend to be higher. Gordon (1994) notes that most panels have 15–35 respondents; however, there have been studies with groups ranging from four to 345 experts. Witkin and Altschuld (1995) suggest that panels should be less than 50 in size with some occasionally being larger. Skulmoski et al. (2007) observe that a homogeneous group needs a smaller sample (10–15); however, heterogeneous ones (such as in an international study) may require up to several hundred subjects. The number of panel size and its frequency is summarized in Table 1.

As shown by the Table 1, although there is little agreement on the ideal number of panellists for a Delphi study, it seemed that the panel size between (10-20) is very common.

The sampling method

The fact that must be given consideration by researchers is that Delphi does not use a random sample which is representative of the target population; but rather, it employs ‘experts' (McKenna, 1994). A judicious and purposeful selection of experts is a critical factor to the reliability of data collected (Clayton, 1997). Skulmoski et al. (2007) recommend using purposive sampling with ‘snowballing’ for expert re-

<table>
<thead>
<tr>
<th>Panel size</th>
<th>Results in</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 10</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>10-20</td>
<td>Ludwig &amp; Barrick, 1996; Boyd, 2003; Dyer &amp; Breja, 2003; Dyer et al., 2003; Martin et al., 2006; Roberts, 2006; Shinn et al, 2008; Warner &amp; Washburn, 2009; Harder et al., 2010; Rayfield &amp; Croom, 2010; Smalley &amp; Retallick, 2010; Franklin, 2011; Ramsey &amp; Edwards, 2011; Conner and Roberts, 2013.</td>
<td>14</td>
</tr>
<tr>
<td>20-30</td>
<td>Park &amp; Rudd, 2005; Simon et al., 2005; Mantooth &amp; Fritz, 2006; Myers &amp; Thompson, 2009; Nistler et al., 2011; Saucier et al., 2012; Wooten et al., 2013.</td>
<td>7</td>
</tr>
<tr>
<td>30-40</td>
<td>Varnadore &amp; Iverson,1991; Buriak &amp; Shinn,1993; Camp et al., 2000; Dobbins &amp; Camp,2003; Roberts &amp; Dyer, 2004; Trexler et al., 2006; Jenkins &amp; Kitchel,2009; Rasouli et al., 2009; Jenkins et al., 2010; Slusher et al., 2011; Namdar &amp; Sadighi, 2013.</td>
<td>9</td>
</tr>
<tr>
<td>≥ 40</td>
<td>Sutphin &amp; Newcomb,1983; Lawrence &amp; Mallilo,1989; Buriak &amp; Shinn,1989; Johnson &amp; Schumacher, 1989; Frick et al., 1991; Shih &amp; Gamon, 1997; Connors, 1998; Murphy&amp;Terry,1998; Mundt &amp; Connors,1999; Akers et al., 2003; Myers et al., 2005.</td>
<td>11</td>
</tr>
</tbody>
</table>
Purposive sampling (judgment, selective or subjective sampling) is a sampling method in which the researcher handpicks the cases to be included in the sample. Creswell (2005) defined purposive sampling as “a qualitative sampling procedure that researchers intentionally select individuals and sites to learn or understand the main phenomenon”.

Well-defined principles for selection objectives are needed; however, guidelines are in little supply (Keeney et al., 2006). Mitchell (1991) advise that it is important to avoid selection bias.

Different sampling methods are summarized in Table 2. As shown by the Table, Purposual sampling was used to select members for most studies’ expert panel.

<table>
<thead>
<tr>
<th>Sampling method</th>
<th>Results in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snowball sampling</td>
<td>-</td>
</tr>
<tr>
<td>Purposeful sampling</td>
<td>Sutphin &amp; Newcomb, 1983; Buriak &amp; Shinn, 1989; Fricket, 1991; Varnadore &amp; Iverson, 1991; Buriak &amp; Shinn, 1993; Ludwig &amp; Barrick, 1996; Shih &amp; Gamon, 1997; Connors, 1998; Murphy &amp; Terry, 1998; Akers et al., 2003; Boyd, 2003; Dyer &amp; Breja, 2003; Dyer et al., 2003; Dobbins &amp; Camp, 2003; Myers et al., 2005; Park &amp; Rudd, 2005; Martin et al., 2006; Mantooth &amp; Fritz, 2006; Trexler et al., 2006; Shinn et al., 2008; Rasouli et al., 2009; Myers &amp; Thompson, 2009; Jenkins &amp; Kitchel, 2009; Warner &amp; Washburn, 2009; Harder et al., 2010; Jenkins et al., 2010; Rayfield &amp; Croom, 2010; Smalley &amp; Retallick, 2010; Franklin, 2011; Nistler et al., 2011; Slusher et al., 2011; Conner and Roberts, 2013; Wooten et al., 2013; Namdar &amp; Sadighi, 2013.</td>
</tr>
<tr>
<td>Systematic sampling</td>
<td>-</td>
</tr>
<tr>
<td>Non sampling</td>
<td>Johnson &amp; Schumacher, 1989; Lawrence &amp; Mallilo, 1989; Mundi &amp; Connors, 1999; Roberts &amp; Dyer, 2004; Simon et al., 2005; Roberts, 2006; Saucier et al., 2012.</td>
</tr>
</tbody>
</table>

### Table 3

**Number of Rounds Used in Delphi Studies**

<table>
<thead>
<tr>
<th>Number of rounds</th>
<th>Results in</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Sutphin &amp; Newcomb, 1983; Buriak &amp; Shinn, 1989; Johnson &amp; Schumacher, 1989; Ludwig &amp; Barrick, 1996; Mundi &amp; Connors, 1999; Akers et al., 2003; Boyd, 2003; Dobbins &amp; Camp, 2003; Myers et al., 2005; Park &amp; Rudd, 2005; Simon et al., 2005; Mantooth &amp; Fritz, 2006; Roberts, 2006; Trexler et al., 2006; Myers &amp; Thompson, 2009; Rasouli et al., 2009; Rayfield &amp; Croom, 2010; Smalley &amp; Retallick, 2010; Nistler et al., 2011; Robinson &amp; Edwards, 2011; Conner and Roberts, 2013; Wooten et al., 2013; Namdar &amp; Sadighi, 2013.</td>
</tr>
<tr>
<td>3</td>
<td>Connors, 1998; Camp et al., 2000; Dyer &amp; Breja, 2003; Dyer, Breja, &amp; Ball, 2003; Shinn et al., 2008; Kitchel &amp; Hains, 2010; Harder et al., 2010; Saucier et al., 2012.</td>
</tr>
<tr>
<td>4 or more</td>
<td></td>
</tr>
</tbody>
</table>
shorted to two or three rounds (Beech, 1997; Green et al., 1999).

The number of rounds in the modified technique may be decreased to as few as two if panelists have been provided with an event list, and if early group consensus is achieved (Snyder-Halpern, 2002).

Other authors have focused on participant’s burden as a problem and suggested that when the number of rounds exceeds four, the response rates can be very low (Linstone & Turoff, 1975).

Three rounds of iterations (as it is shown in Table 3) are commonly viewed as sufficient for arriving at a high-level of the agreement. Although the possibility of more than three rounds is offered, there is a need to balance time, cost and possible participant fatigue. As noted by McCampbell and Stewart (1992), most Delphi studies reach consensus at the third round. However, failing to achieve consensus on a majority of the items, a fourth round was initiated.

**Round one**

Round one of the classical Delphi starts with one or several open-ended questions, thereby allowing the panel members to enjoy great freedom in their responses. Round one is used to generate opinions, and the panel members are asked for their responses or comments about a subject (Keeney et al., 2006). Franklin and Hart (2007) expressed that researchers develop the initial questionnaire based on a perfect literature review. The questionnaire, thus, can be a summary of previous research theories and findings postulated by scholars. The first questionnaire gives a way to structure ideas around a set of common statements to panelists. Researchers use content analysis to identify the main themes from the open-ended questions of the first round (Powell, 2003) in order to form items for future researches (Keeney et al., 2006).

Open-ended questions are used to collect an array of views or issues to be addressed in later rounds. Using broad questions in the first round of a Delphi survey may discourage experts with time constraints to participate in a study, which was indicated by the dropout rate of some participants in the first Delphi study accommodate. Less broad survey questions should be considered to stimulate expert participation in a Delphi study. In Table 4 the number of questions in the first round and its frequency is presented. As shown in Table 4, in 22 studies, the Delphi started with only one question.

### Validating of the first round question

<table>
<thead>
<tr>
<th>Number of questions in the first round</th>
<th>Results in</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lawrence &amp; Mallilo, 1989; Fricke et al., 1991; Connors, 1998; Mundt &amp; Connors, 1999; Dyer et al., 2003; Dyer &amp; Breja, 2003; Roberts &amp; Dyer, 2004; Myerson et al., 2005; Roberts, 2006; Martin et al., 2006; Mantooth &amp; Fritz, 2006; Trexler et al., 2006; Shinn, et al., 2008; Rasouli et al., 2009; Warner &amp; Washburn, 2009; Myers &amp; Thompson, 2009; Harder et al., 2010; Nistler et al., 2011; Ramsey &amp; Edwards, 2011; Slusher et al., 2011; Saucier et al., 2012; Conner and Roberts, 2013; Wooten et al., 2013; Namdar &amp; Sadighi, 2013.</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Shih &amp; Gamon, 1997; Park &amp; Rudd, 2005; Rayfield &amp; Croom, 2010.</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Buriak &amp; Shinn, 1989; Akers et al., 2003; Boyd, 2003; Jenkins &amp; Kitchel, 2009; Jenkins et al., 2010; Franklin, 2011; Smalley &amp; Retaillick, 2011.</td>
<td>7</td>
</tr>
<tr>
<td>4 or more</td>
<td>Murphy &amp; Terry, 1998; Camp et al., 2000.</td>
<td>2</td>
</tr>
</tbody>
</table>
The questions of the first round were validated for content regarding their appropriateness for the objectives of the studies by external experts that could be a panel of faculty and graduate students or a jury of agricultural educators (Akers et al., 2003; Dobbins & Camp, 2003; Dyer & Breja, 2003; Dyer et al., 2003; Myers et al., 2005) or a panel of internal expert was asked to validate the round one questions (Buriak & Shinn, 1989; Johnson & Schumacher, 1989; Lawrence & Mallilo, 1989; Ludwig & Barrick, 1996; Connors, 1998; Dyer & Breja, 2003; Dyer, Breja, & Ball, 2003; Roberts & Dyer, 2004; Myers, Dyer, & Washburn, 2005; Park & Rudd, 2005; Fritzschke, & Ball, 2006; Roberts, 2006; Trexler et al., 2006; Myers & Thompson, 2009; Warner & Washburn, 2009; Smalley & Retallick, 2010; Slusher et al., 2011; Franklin, 2011; Saucier et al., 2012; Wooten, Rayfield & Moore, 2013; Namdar & Sadighi, 2013).

Data analysis

The instruments that was used in the second and third rounds contained items on which a predetermined level of consensus was not achieved during the panel of the previous round. The researchers determined a priori that only those competencies receiving the percent level of agreement would be used for the inclusion in the investigation.

In round 2, the jury was asked to rate their strength of agreement for each statement on a Likert-type scale. Those statements that received a five or six points (agree or strongly agree in a six point Likert type) from at least two-thirds of the jury responding in round 2 were kept for the third round. This would ensure a true consensus of the entire group. With a mean cut–off score, one could have a high mean score, yet have one or more panellists mark half of a Likert scale or below, which does not indicate agreement of an item to be included as, in this case, a quality indicator. Thus, this use of the method is a more stringent approach to item selection.

How to reach an agreement

Because of the disparate nature of the panel, the lack of a clear agreement on how to define consensus in the Delphi presented a minor challenge. In practical Delphi studies, investigators should be more transparent about their choice of agreement index and report the value of the selected index within every round. Hassan et al., (2000) argued that statistical aggregation of responses to scaled items are measures of central tendency like mean, median, and mode, and dispersion like standard deviation and inter quartile range.

In a systematic review of the literature on Delphi method, different descriptive statistics were used. These statistics included mean, median, mode, percentages for each event, ranks, upper and lower quartile ranges, regression weights or induced (if-then) rules, as well as the statistical average of points for each factor. Stone, Fish, and Busby (2005) suggested analysing Delphi data using median and inter quartile ranges to identify rates of group agree-

<table>
<thead>
<tr>
<th>Index</th>
<th>Results in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/mean &amp; standard deviation</td>
<td>21</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>2</td>
</tr>
<tr>
<td>Frequencies, percentages, and ranks</td>
<td>10</td>
</tr>
<tr>
<td>Inter-quartile range</td>
<td>0</td>
</tr>
<tr>
<td>Cronbach’s alpha</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5

The Indexes for Reaching Agreement in Delphi Studies
ment for consensus. The use of inter-quartile ratings (IQR) provides the researcher with information “… about the variability in the data without being affected by extreme scores” (Sprenkle & Piercy, 2005). The usual indexes of reaching agreement are shown in Table 5. As can be seen, comparing different indexes mean and standard deviation was more used.

The percent of agreement required for round two in Delphi studies was more than 70% that is shown in Table 6.

The purpose of round 3 is to begin the process of developing consensus among the jury. Used similar benchmarks for consensus in round 2. The percent of consensus required in round three in Delphi studies was more than 60 as is shown in Table 7.

If 100% of the respondents had chosen “agree”, it would have been included as a quality indicator.

Loughlin and Moore (1979) recommended that consensus would be at least 51% of agreement among respondents. Ulshak (1983) suggests 80%, and Green (1982) desires at least 70%. Mitchell (1991) views 75% as the lowest level. Since limiting the number of rounds could prevent total consensus, 75% agreements were chosen as the consensus level. It means that if 75% or less agreed an item should be included as a quality indicator, that item was dismissed as a possible quality indicator and removed from the study.

Table 6
Percent of Agreement To be Needed for Round Two in Delphi Studies

<table>
<thead>
<tr>
<th>Percent of agreement needed for round 2</th>
<th>Results in</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥25</td>
<td>Frick et al., 1991.</td>
<td>1</td>
</tr>
<tr>
<td>≥50</td>
<td>Burriak &amp; Shinn, 1989; Johnson &amp; Schumacher, 1989; Shih &amp; Gamon, 1997; Mantooth &amp; Fritz, 2006; Fritzsche, &amp; Ball, 2006; Ramsey &amp; Edwards, 2011.</td>
<td>6</td>
</tr>
<tr>
<td>≥60</td>
<td>Sutphin &amp; Newcomb,1983; Varnadore &amp; Iverson.1991; Connors,1998; Murphy &amp; Terry,1998; Mundt &amp; Connors,1999; Camp et al., 2000; Trexler et al., 2006; Shinn et al., 2008; Conner and Roberts, 2013.</td>
<td>9</td>
</tr>
<tr>
<td>≥70</td>
<td>Lawrence, L &amp; Mallilo, 1989; Akers et al., 2003; Boyd, 2003; Dyer &amp; Breja, 2003; Dyer et al.,2003; Myers et al.,2005; Jenkins &amp; Kitchel,2009; Myers &amp; Thompson, 2009; Warner &amp; Washburn, 2009; Kitchel, &amp; Hains, 2010; Smalley &amp; Retallick, 2010; Franklin, 2011; Nistler et al., 2011; Slusher et al., 2011; Nistler et al., 2011.</td>
<td>13</td>
</tr>
<tr>
<td>≥80</td>
<td>Ludwig &amp; Barrick,1996; Roberts &amp; Dyer, 2004; Park &amp; Rudd, 2005; Simon et al., 2005; Roberts, 2006; Rasouli et al., 2009; Rayfield &amp; Croom, 2010; Wooten et al., 2013.</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 7
Percent of Consensus Needed in Round Three in Delphi Studies

<table>
<thead>
<tr>
<th>Percent of consensus</th>
<th>Results in</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥50</td>
<td>Shih &amp; Gamon, 1997; Mantooth &amp; Fritz, 2006.</td>
<td>2</td>
</tr>
<tr>
<td>≥60</td>
<td>Sutphin &amp; Newcomb,1983; Murphy &amp; Terry,1998; Connors,1998; Mundt &amp; Connors,1999; Camp et al.,2000; Boyd, 2003; Dobbins &amp; Camp, 2003; Myers, Dyer, &amp; Washburn, 2005 Fritzsche, &amp; Ball, 2006; Trexler et al., 2006; Shinn et al., 2008; Nistler et al., 2011.</td>
<td>12</td>
</tr>
<tr>
<td>≥70</td>
<td>Akers et al., 2003; Dyer &amp; Breja, 2003; Dyer et al., 2003; Jenkins &amp; Kitchel, 2009; Myers &amp; Thompson, 2009; Warner &amp; Washburn, 2009; Kitchel, &amp; Hains, 2010; Harder et al., 2010; Smalley &amp; Retallick, 2010; Franklin, 2011; Ramsey &amp; Edwards, 2011; Slusher et al., 2011; Namdar &amp; Sadighi, 2013.</td>
<td>12</td>
</tr>
<tr>
<td>≥ 80</td>
<td>Ludwig &amp; Barrick, 1996; Roberts &amp; Dyer, 2004; Simon et al., 2005; Park &amp; Rudd, 2005; Roberts, 2006; Rasouli et al., 2009; Rayfield &amp; Croom, 2010; Conner and Roberts, 2013; Wooten et al., 2013.</td>
<td>8</td>
</tr>
</tbody>
</table>
of the study arise from the value led nature of feedback and the instability of responses and consensus. These areas are, in turn, influenced by the number of experts, their average expertise and the average inter-correlation of their judgments. Content validity of the questionnaire can be determined by piloting. Few 'Delphi' researchers report undertaking pilot tests before implementation (e.g. Akers, Vaughn, & Haygood, 2003; Simon et al., 2005), it is unclear how many pilot tests should be undertaken when using this method. For example, should there be one for every round or only one for the initial round? (Keeney et al., 2001).

Ludwig and Starr (2005) point that “the validity of a Delphi study depends not on the number of participant survey, but rather on the expertise of the panel who participate”. Content and Face validity of the initial instrument was confirmed through a panel of experts (e.g. Ludwig & Barrick, 1996; Dyer et al., 2003; Jenkins & Kitchel, 2009; Kitchel, & Hains, 2010; Mundt & Connors, 1999; Mantooth & Fritz, 2006; Rayfield & Croom, 2010; Saucier et al., 2012; Robinson, & Edwards, 2011; Varnadore & Iverson, 1991) These authors argue for setting specific guidelines for each area, so the reliability of the study (or whether a replication of the study would give the same results with a different panel) can be judged.

Estimates of the internal consistency reliability of each questionnaire obtained using Cronbach’s alpha analysis. In often studies, researchers follow Dalky (1969). He found that when the size of the jury was greater than 13, mean correlations were greater than 0.80, satisfying questions of process reliability (e.g. Boyd, 2003; Burik & Shinn, 1989; Dyer & Breja, 2003; Franklin, 2011; Harder et al., 2010; Fritzsche, & Ball, 2006; Myers et al., 2005; Ramsey & Edwards, 2011; Rayfield & Croom, 2010; Roberts & Dyer, 2004; Saucier et al., 2012; Small & Retallick, 2010; Shinn et al., 2008; Robinson, & Edwards, 2011; Murphy & Terry, 1998; Warner & Washburn, 2009). Given the nature of the Delphi technique, additional types of validity and reliability estimates were not appropriate for the instrument (Dalkey et al., 1972).

CONCLUSIONS AND RECOMMENDATIONS

By searching through and reviewing the literature, the researchers were able to confirm that the Delphi method continues to be used and is a valid method for forecasting and supporting decision-making. There are inherent characteristics or weaknesses in the methodology or its application that have not been completely corrected. Delphi does not call for expert panels to be representative samples for statistical purposes. Representativeness is assessed on the qualities of the expert panel rather than its numbers. The Delphi method showed satisfactory reliability and validity indexes. The method should be used judiciously and only after careful preparation. Measurements of the main trend were obtained. The experiences provided evidence of the present and potentiality of the Delphi method in the areas of input for quantitative models by means of expert opinion. Based on the findings of the present study, the researchers recommend a Delphi study with the size ranging between 10-20, the purposeful sampling method for selecting the panel of expert, three rounds, one open-ended question in the first round, using Mean/Mean and Standard Deviation for reaching agreement in round two and 60% or more for consensus in round three for future studies.

ACKNOWLEDGEMENTS

In particular, we want to thank the editors of IJAMAD, for their valuable help and suggestions and for providing stimulating feedback that helped us to reinforce our arguments.

REFERENCES


staffs view the likely development of two community mental health centers over the course of the next two years. *Journal of Advanced Nursing*, 25(2), 331-338.


Achieving Consensus Deal with Methodological ... / Goodarzi et al.


agricultural education programs. *Journal of Agricultural Education, 47*(1), 100-109.


How to cite this article:

URL: [http://ijamad.iaurasht.ac.ir/article_540498_e5b9a16dbf24954c307c2228c2786d01.pdf](http://ijamad.iaurasht.ac.ir/article_540498_e5b9a16dbf24954c307c2228c2786d01.pdf)