Seventh International Conference on Construction in the 21st Century (CITC-VII)
"Challenges in Innovation, Integration and Collaboration in Construction & Engineering"
December 19-21 2013, Bangkok, Thailand

Experience of using Delphi method in construction health and safety research

Justus Agumba
University of Johannesburg
jagumba@uj.ac.za

Innocent Musonda
University of Johannesburg
imusonda@uj.ac.za

Abstract

This paper discusses the Delphi method which is an inductive approach to research. It has been argued that Delphi method is an essential vehicle of reaching consensus in issues that cannot be resolved in a once off discussion. This paper reports on the experiences achieved by the authors in using the Delphi method among a group of experienced health and safety (H&S) experts in two separate case studies in the construction industry in South Africa. The first case study used three rounds of iteration whilst the second case study used four rounds of iteration. The authors argue that the Delphi method is a comprehensive method of attaining consensus on challenging issues of H&S in the construction industry. Furthermore the method requires proper communication to achieve the required results.

Keywords: Experience; Inductive approach; health and safety

1. Introduction

The Delphi method was first developed by the Rand Corporation. It is a structured process requiring experts on a subject to respond to non-leading unambiguous statements with the aim of achieving consensus (Holey *et al.,* 2007). The Delphi method can be useful in providing solutions to the many problems in the industry because it addresses the 'what can-if' kind of questions as opposed to the 'what is' kind of questions.

The Delphi method entails soliciting expert opinions on what would happen given a particular solution. This type of question can be addressed by methods such as the Delphi and or focus groups apart from experimental procedures. However experimentation most of the times is not feasible. A study using focus groups could be feasible except that there is a challenge of bringing experts to one destination and making them deliberate an issue or phenomena for long hours and probably for days. It is also considered costly and most of the times fail to attract the experts in view of the logistics. Furthermore, it is possible for biasness to creep due to peer pressure. Delphi method eliminates the biasness as members of the expert panel remain completely anonymous to each other and therefore no undue influence from other peers.

In addition, the Delphi method is considered to be a robust method for a rigorous query to experts. Unlike ordinary survey research, the Delphi's strength also lies in the iterative process (rounds of questioning) used which provide an opportunity for initial feedback, collation of feedback, and distribution of collated

feedback to participants for further review. This unique process requiring group communication is central to the strength of the Delphi (Gohdes and Crews, 2004).

2. Conducting the Delphi Study

2.1 Selection of Delphi panel members

One of the important tasks in conducting a Delphi study is to assemble a panel of experts. A Delphi study can be conducted with as few as eight panel members. Skulmoski, Krahn and Hartman (2007) suggested that 10 to 15 panellists could be sufficient if the background of the panellists was homogenous. A review by Rowe and Wright (1999) indicated that the size of a Delphi panel ranged from three to 80 in peer reviewed studies. Similarly, Okoli and Pawlowski (2004) and Skulmoski *et al.,* (2007) found that a panel size of about 10 to 18 members was adequate. Hallowell and Gambatese (2010) suggested that since most studies incorporate between eight and 16 panellists, a minimum of eight should suffice. The size of the panel should be dictated by characteristics of the study such as the number of available experts, the desired geographical representation and the capacity of the facilitator. Based on this discussion, and the fact that the Delphi method does not depend on the statistical power, but rather on group dynamics for arriving at consensus among experts, the case studies presented in this paper was considered to be adequate.

The selection of panellists is based on criterion sampling. Panellists are purposefully selected to apply their knowledge to a concept raised in the study based on the criteria that is developed from the research questions under investigation. A Delphi study does not depend on a statistical sample that attempts to be representative of any population. However, the Delphi technique is a group-decision mechanism requiring qualified experts who have a deep understanding of the issues to come up with solutions through consensus (Okoli *et al.,* 2004). Therefore, one of the most critical requirements for the Delphi technique to achieve the research objectives is the selection of qualified experts because it is the most important step in the entire Delphi process. It directly relates to the quality of the results generated. Therefore, careful selection of the panel of experts is the keystone to a successful Delphi study (Gohdes *et al.,* 2004; Hsu and Sandford, 2007).

2.2 The process of data collection in the Delphi technique

The Delphi method may involve at least three or four rounds of an iterative process with the view of achieving consensus between the panel members. A Delphi questionnaire is sent out electronically to all panel members who are then asked to take time and respond to the questions according to their ability and expertise. The experts may be requested to rate the, severity or impact; make a prediction; or make a judgement on an issue that has been presented before them. The ultimate goal is for panellists to reach consensus on all issues.

Responses are received for each round of the Delphi. In order to establish consensus on each and every question, group medians are calculated for each response for each element. The group mean, median and mode can all be used as measures of central tendency in Delphi analysis (Hsu *et al.,* 2007). The group median is the preferred measure of central tendency because it indicates a general inclination of the group on a particular question. The median eliminates biasness and takes into consideration outlier responses. It makes the consensus notion more reasonable. Other measures of central tendency, such as the mean for example, may not reflect a reasonable central tendency because it does not consider frequencies.

Upon receipt of responses from the first round, group medians are computed for each question. In the second round, the same questionnaire is sent back to panellists individually with their own responses from

the first round with the group median responses included so that responses in the second round could be made taking into account the group median. In the second round expert panel members are asked to either maintain their original responses made in the first round, or they could change their initial response to either be in agreement with the group median or make a new rating altogether. The panellists, who have ratings of two units either above or below the group median on any one particular question, are requested to state the reasons for their dissenting opinion if at all they opted to stick to their rating. The stated reasons are sent to all panellists so that together with the calculated group medians in the second round, panellists could take cognisance of those comments in making their new ratings in the third round. Panellists are specifically requested to consider reasons from the outliers made in the second round in making their decisions in the third round.

After the second round, group medians and comments made by those with dissenting views are sent to panellists together with each panellist rating made in the second round for a third round of responses. Calculated group medians and comments are sent to all panellists and individual ratings are sent only to those that had made them for anonymity.

After the third or fourth round, group medians and the absolute deviations are again computed for this round. If calculations for the third round of the Delphi process indicate the desired level of consensus, then there is no need to proceed to the fourth round as there is no further value that could possibly be added to the degree of consensus that has been attained. Throughout the Delphi process, anonymity of panel members should be maintained to avoid undue influence on other members. The aspect of anonymity is crucial to the credibility of the Delphi process.

2.3 Analysis of data from Delphi

The analysis of data from the Delphi may involve an analysis to establish or confirm consensus on responses to the predetermined criteria. This involves determining the group median responses for each question. After each round of the Delphi, absolute deviations (D_i) about the group medians ($m(X)$) of each rating for every question may be calculated using equation 01. In addition to this, average (mean) absolute deviations (MAD) are calculated for every question. The mean absolute deviation is a calculated mean of all absolute deviations for all panellists about the median on each question. A further analysis may involve determining the statistical range in ratings by panellists on each question and the percentage of panellists with a similar opinion inclination on each and every question as in case study two.

$$\mathrm{D_i} = \left(x_i - m(x)\right) \qquad \textbf{Equation 01}$$

Where:

D_i = Absolute deviation

x_i = Panellist rating

$m(X)$ = Measure of central tendency

2.4 Establishing consensus from the Delphi process

Unlike questionnaire surveys which simply requests for the opinion of non-experts on a matter, the Delphi technique seeks to establish the opinion of experts on a particular matter. Therefore it is important in a Delphi study that consensus is reached on all questions. However, measuring or determining consensus is a highly contended subject in literature. It should in fact be mentioned that there is no

agreement in literature on how consensus can be said to have been attained regarding a set of opinions (Hsu *et al.,* 2007). Holey *et al.,* (2007) suggest that consensus is the same as agreement and that agreement can be determined by the following:

- the aggregate of judgements;
- a move to a subjective level of central tendency; or
- alternatively by confirming stability in responses with the consistency of answers between successive rounds of the study.

Consequently, in order to determine agreement or consensus, some researchers have used frequency distribution and the criterion of 60% responding to any given response category (Gohdes *et al.,* 2004). Other studies such as one conducted by Rayens and Hahn (2000), used means and standard deviations with a decrease in standard deviations between rounds indicating an increase in agreement. Coupled with the means and standard deviations, measures such as the inter-quartile deviation (IQD) have also been used to determine consensus (Rayens *et al.,* 2000) with smaller values of the inter-quartile range indicating higher degrees of consensus. In his study, Rayens *et al.,* (2000) included another criterion to determine consensus (in addition to the IQD) in order to achieve what he referred to as stability. The criterion to achieve consensus was that the IQD should equal to one unit for which more than 60% of respondents should have answered either generally positive or generally negative. Items which had an IQD =1, for which the percentage of generally positive or generally negative responses was between 40 and 60%, were determined to indicate a lack of consensus or agreement.
As for Holey *et al.,* (2007), they used the following criteria to determine consensus in their study:

- an increase in percentage agreements;
- convergence of importance rankings;
- increase in kappa values;
- a decrease in comments as rounds progress;
- a smaller range of responses; and
- smaller values of standard deviations

Although there is little consensus on how to determine consensus, it is apparent that there has to be some measure of central tendency about which a measure of dispersion of individual responses can be done. It is also apparent that agreement is said to be reached when responses congregate around the central tendency measure and if variability in responses is small. It is in fact agreeable from the practice that for consensus to be attained, there has to be a convergence of ideas and reasoning towards a subjective central tendency measure.

3. Case studies

3.1 Delphi Case Study One

The Delphi case study No. 01 sought to obtain as an output a conceptual model of the environment and client influence on construction project health and safety (H&S) performance. The external environment and the client H&S culture have a certain level of influence on project H&S performance. Literature informs that improving H&S culture could lead to improvement in H&S performance. What was however not clear in the literature was specifically the level or extent of impact of client H&S culture on project H&S performance. An attempt was made to infer client's influence by relating total recordable injury rates (TRIR) to client involvement (Huang and Hinze, 2006). However, it was noted that accidents, incidents, injury or even fatality records are not a reliable and certainly not the only measure of H&S performance. These are lagging indicators and the TRIR is also a lagging indicator. The H&S culture of an organisation and other leading indicators are now the accepted measure of H&S performance.

An alternative measure of the impact of client influence was therefore necessary in order to establish not only whether clients do have influence on H&S performance but also the extent or level of this influence. Such a study would ordinarily call for an experimental research. However, this kind of research was not feasible and practical considering the time frame, ethical issues and the willingness of would be participants.

The Delphi method was therefore considered in part, the most suitable method to achieve the general objective of determining the impact of client's influence on project H&S performance. Therefore, the specific objectives of the Delphi study entailed establishing the impact of the:

- **D1.** external environment factors on client H&S performance;
- **D2.** client H&S culture on H&S consideration (hazard identification, risk analysis, assessment and mitigation) throughout the project life cycle;
- **D3.** client H&S culture on contractor H&S performance (top management);
- **D4.** client H&S culture on designer H&S performance; and
- **D5.** client H&S culture on the overall project H&S performance.

Achieving the above objectives, resulted in the following outcomes:

1. factors of client H&S culture, contractor and designer H&S performance, and external environment that were of critical significance to H&S performance improvement; and
2. a conceptual client-centred model on H&S performance improvement.

Consequently, the data that needed to be collected was the ratings of impact of factors of client H&S culture on construction project H&S performance. In addition, ratings were also made on the likelihood of H&S elements or programmes being implemented on a construction project as a result of client's influence. This data was obtained through the use of questionnaires. Experts were asked to complete the questionnaires and reach consensus on the rated likelihoods and severity (impact) of various H&S factors. The process involved a three round iterative process with the main aim of getting experts to reach consensus on the questions raised in the questionnaires. Experts were also encouraged to give reasons for their dissenting views.

The decision on consensus was not based on one measure as criteria, however several other measures congruent to most literature were necessary and formed the criteria to determine consensus. These included:

1. the median as the measure of central tendency;
2. percentage of panellists with a generally positive or negative rating on a question not to be less than 60%;
3. the mean absolute deviation to be used as the measure of dispersion about the central tendency measure;
4. the calculated average (mean) of the absolute deviation (MAD) not to be more than one unit; and
5. the range in responses not to be more than four units.

After three rounds of the Delphi process, the mean absolute deviation was found to be -0.09 which was less than one unit. The value was negative because it was below the measure of central tendency - the median. The percentage of panellists leaning either generally positive or negative was 91.6% up from 91.0% and 78% from the second and first rounds respectively. This was an indication of a movement towards a convergence of ideas and reasoning. The range in responses between the highest rating and the lowest rating was 3.50. This value was lower than the upper limit of 4.0. These results are summarised in Table 1 and Figure 1 and 2. With those values after the third round, it was determined that consensus had been reached and therefore there was no need to proceed to round four.

Table 1: MAD, range and percentage agreeing values

Round	MAD	Range in ratings	Generally leaning positive or negative
1	-1.34	5.80	78.0%
2	-0.09	3.90	91.0%
3	-0.09	3.50	91.6%

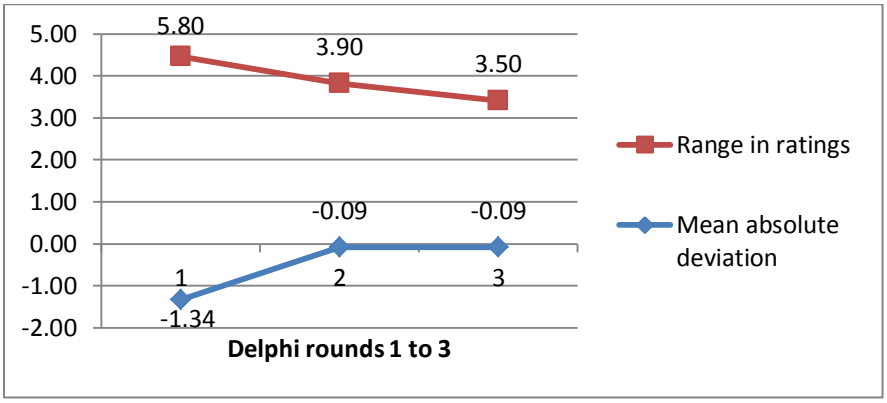

Figure 1: Plot of MAD and range to determine consensus

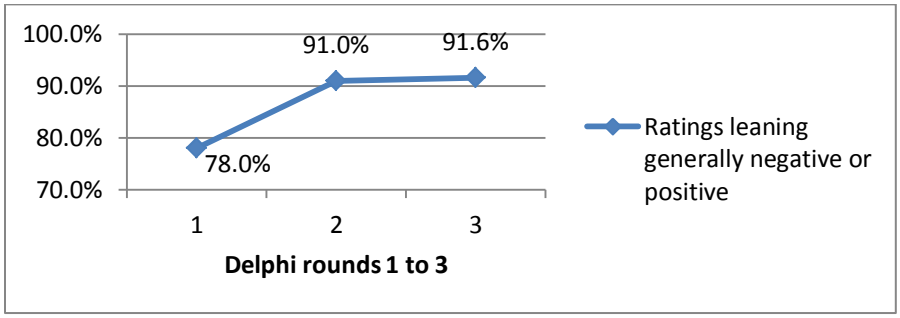

Figure 2: Percent of panellists with similar inclination on an issue

3.2 Delphi Case Study Two

The Delphi case study No. 02 sought to obtain as an output a conceptual model of the H&S practices on construction project health and safety (H&S) performance. Literature informs that H&S practices constitute the H&S culture of the organization which could lead to improvement in H&S performance. What was however not clear in the literature was specifically the important and the impact of H&S practices on project H&S performance. An attempt was made to infer H&S practices influence by relating total recordable accidents (Mohammed, 1999). However, it has been noted that accidents, incidents, injury or even fatality records are not a reliable and certainly not the only measure of H&S performance. These are lagging indicators and the recordable accident is also a lagging indicator. The H&S culture of an organisation and other leading indicators are currently an accepted measure of H&S performance according to Hinze, *et al.*, (2013) and Grabowiski *et al.*, (2010).

The Delphi method was therefore considered in part, the most suitable method to achieve the general objective of determining the important H&S practices and their perceived impact on project H&S

performance. Therefore, the specific objectives of the Delphi study entailed establishing the importance and impact of the:

- **D1.** Upper management commitment and involvement practices in H&S on H&S performance;
- **D2.** Employee involvement and empowerment practices in H&S on H&S performance
- **D3.** Informal and formal written communication practices in H&S on H&S performance
- **D4.** Informal and formal verbal communication practices in H&S on H&S performance
- **D5.** H&S resources practices on H&S performance
- **D6.** H&S training practices on H&S performance
- **D7.** H&S planning practices on H&S performance
- **D8.** H&S policy practices on H&S Performance
- **D9.** Project supervision practices on H&S performance; and
- **D10.** Appointment of H&S staff practices on H&S performance

Achieving the above objectives, resulted in the following outcomes:

1. Eleven practices of upper management commitment and involvement, five for employee involvement and empowerment, one practice for informal and formal verbal communication, two practices for H&S resources, three practices for H&S training, three practices for H&S planning and six practices for project supervision were critical to H&S performance improvement; and
2. A refined conceptual H&S performance improvement model was then proposed.

The decision on consensus was not based on one measure as criteria, however several other measures congruent to most literature were necessary and formed the criteria to determine consensus. These included:

3. The median as the measure of central tendency for both important and impact scale had to complement each other. The panellist group median for consensus was set at 9 to 10 for importance scale indicating very important and 90% to 100% for the impact scale indicating major impact;
4. 50% and more of respondents had to rate the H&S practice 9 to 10 in the importance scale indicating very important and 90% to 100% in the median impact scale indicating major impact. The two scales had to complement each other.

After four rounds of the Delphi process consensus was achieved for each H&S practice. This was achieved when the median importance and median impact were above 9.00 and 90.00% respectively. A further analysis was that 50% of respondents should rate the practices above the median importance of 9.00 and median impact of 90.00%. It was determined that consensus had been reached after similarity of results in round three and four of Delphi survey.

4. Findings- Experiences in Conducting the Studies

In the two case studies presented in this paper, the Delphi study formed the initial stage of two larger studies whose objective was to develop models about specific solutions to improve H&S performance in the construction industry at project level. The Delphi studies were therefore useful in developing the conceptual models which were later validated using structural equation modelling.

The Delphi technique was found to be an invaluable tool to augment what literature presents in formulating a theory. It was also found to be a necessary step in conducting a rigorous investigation to complex questions.

However, conducting a Delphi study requires a careful management of both the data and the panel members. As a result, the Delphi study is much more demanding for the researcher compared to say ordinary questionnaire survey. In the ordinary questionnaire survey, data is only input once while for the Delphi data may be input at least three times. Further, the analysis is also undertaken at least three times.

Coupled with the above, a careful scrutiny of comments by panel members must be undertaken. With such amount of data and communication, it is also possible to mix up the email communication resulting in sending emails to panellists who were not intended to receive them.

The Delphi study by its nature is a rigorous iterative process and therefore it tends to involve long questionnaires and therefore equally demanding on the part of the experts. The authors found therefore that the researchers must spend a considerable amount of time communicating with the participants and keeping them motivated to continue with the study. This is however an easy task. In the Delphi case study one, out of an initial 15 panellists, 11 of them went through to the last round. Whereas in case study two out of the initial 20 panellists, 16 of them completed the four rounds.

Notwithstanding the difficulties and demands on both the researcher and the experts, the Delphi study was found to be an essential tool in seeking solutions to the problems that the construction industry has for example. It is not enough to merely report on the status of issues in the industry but what is required is to offer solutions to the problems which are in fact well known.

5. References

Gohdes, S.L.W. and Crews, B.T. (2004). "The Delphi technique: A research strategy for career and technical education". *Journal of career and technical education*, Vol. 20, No. 2, pp 55-67.

Grabowski, M., You, Z., Song, H., Wang, H. and Merrick, R.W.J. (2010). "Sailing on Friday: Developing the Link between Safety Culture and Performance in Safety Critical Systems". *IEEE Transactions on Systems Management and Cybernetics- Part A: Systems and Humans*, Vol. 40, No. 2, pp 263-284.

Hallowell, R.M. and Gambatese, A.J. (2010). "Qualitative research: application of the Delphi method to CEM research". *Journal of Construction Engineering and Management*, Vol. 136, No. 1, pp 99-107.

Hinze, J., Thurman S. and Wehle A. (2013). "Leading indicators of construction safety performance". *Safety Science*, Vol. 51, No. 1, pp 23-28.

Holey, E.A., Feeley, J.L., Dixon, J. and Whittaker V.J. (2007). "An exploration of the use of simple statistics to measure consensus and stability in Delphi studies". *BMC medical research methodology*, Vol. 7, No. 52, pp 1-10.

Hsu, C-C. and Sandford, A.B. (2007). "Minimizing non-response in the Delphi process: How to respond to non-response". *Practical Assessment, Research and Evaluation*, Vol. 12, No. 17, pp 62-78.

Huang, X. and Hinze, J. (2006). "Owner's role in construction safety". *Journal of Construction Engineering and Management*, Vol. 132, No. 2, pp 164-173.

Mohammed, S. (1999). "Empirical investigation of construction safety management activities and performance in Australia". *Safety Science*, Vol. 33, No. 3, pp129-142.

Okoli, C. and Pawlowski, S. D. (2004). "The Delphi method as a research tool: an example, design considerations and applications". *Information and Management*, Vol. 42, No. 1, pp 15-29.

Rayens, M.K. and Hahn, E.J. (2000). "Building consensus using policy Delphi method". *Policy, politics and nursing practice*, Vol. 1, No. 4, pp 308-315.

Rowe, G. and Wright, G. (1999). "The Delphi technique as a forecasting tool: Issues and analysis". *International Journal of Forecasting*, Vol. 15, No. 4, pp 353-375.

Skulmoski, J.G., Hartman, T.F. and Krahn, J. (2007). "The Delphi method for graduate research". *Journal of Information Technology Education*, Vol. 6, pp 1-21.