Technological Forecasting & Social Change 103 (2016) 166–173

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

The effect of controlled opinion feedback on Delphi features: Mixed messages from a real-world Delphi experiment

Jurian V. Meijering *, Hilde Tobi

Research Methodology Group, Social Sciences Department, Wageningen University, Hollandseweg 1, 6706KN, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form 14 October 2015
Accepted 6 November 2015
Available online xxxx

Keywords:
Delphi experiment
Controlled opinion feedback
Drop-out
Opinion change
Agreement
Feedback perception

ABSTRACT

A real-world Delphi experiment was conducted to investigate the effect of two controlled opinion feedback conditions on the drop-out rate, experts' degree of opinion change, and the increase in the level of agreement among experts. Additionally, experts' perceived usefulness of feedback was explored. In the first and second Delphi round experts received a questionnaire which consisted of two sections. Within each section experts were asked to rate several items. In round 2, experts in one condition received feedback consisting of summary statistics and rationales (*S*&*R* condition), whereas experts in the other condition received rationales only (*R* condition). Results showed that drop-out of experts was greater in the *S*&*R* condition than in the *R* condition. No difference between conditions was found concerning experts' degree of opinion change. The increase in the level of agreement across the items in the second section of the questionnaire differed significantly between conditions. This difference was mainly due to a decrease in agreement in the *R* condition, suggesting that feedback of rationales may increase disagreement among experts. In round 3 experts preferred to receive both summary statistics and rationales, although they tended to perceive rationales as somewhat more useful than summary statistics.

1. Introduction

The Delphi method, originally developed by Dalkey and Helmer (1963), is a structured data-collection process that is often used to allow experts to achieve a certain level of agreement on a particular topic (Keeney et al., 2006). However, the method has various other uses such as maximizing the range of expert opinions (Landeta and Barrutia, 2011; van de Linde and van der Duin, 2011; Pätäri, 2010; Steinert, 2009; Banwell et al., 2005). Any Delphi study consists of at least two rounds. In each round experts are independently questioned about their opinion on the topic of interest by means of a standardized questionnaire. To avoid undue influence of dominant experts and group pressure, experts are anonymous and are not allowed to communicate with each other. Instead, the researcher provides controlled opinion feedback to the experts in the form of a summary of the results from the previous round. Based on this feedback, experts may choose to revise their opinion in the next round. A Delphi study usually ends when a desired level of agreement has been achieved or when a certain stability in experts' responses has been reached (Keeney et al., 2006; Linstone and Turoff, 1975; Dalkey, 1969; Landeta, 2006; Rowe and Wright, 1999).

Controlled opinion feedback is an essential part of the Delphi method. Still, no evidence based guidelines exist on how to provide feedback. As a result, Delphi studies differ in the kind of feedback provided. Typically, a distinction is made between summary statistics, which show the majority opinion, and rationales, which show why experts hold certain opinions (Rowe et al., 2005). A systematic review on the use of the Delphi method for selecting healthcare quality indicators, showed that most Delphi studies provided feedback consisting of summary statistics only (Boulkedid et al., 2011). Various researchers criticized this kind of feedback as being insufficiently informative and they proposed to feed back rationales as well (Murphy et al., 1998; Rowe et al., 1991; Bolger and Wright, 2011). Lately, some even suggested that feedback should solely consist of rationales to prevent experts from simply changing their opinion in the direction of the majority (Bolger et al., 2011).

Although the provision of feedback in the form of both summary statistics and rationales is advocated by some, there seems to be little empirical evidence in support of this claim (see Section 2.1 for a description of the few experiments that were found). In response to the debate about controlled opinion feedback, the current Delphi experiment aimed to investigate the effect of feeding back rationales with and without summary statistics on various Delphi features. These features include the drop-out rate, experts' degree of opinion change, and the increase in the level of agreement among experts. Additionally, experts' perception of the usefulness of feedback was explored.

* Corresponding author.
E-mail addresses: jurian.meijering@wur.nl (J.V. Meijering), hilde.tobi@wur.nl (H. Tobi).

http://dx.doi.org/10.1016/j.techfore.2015.11.008
0040-1625/

2. Theory and hypotheses

2.1. Research into the effect of controlled opinion feedback

Few experiments investigated the effect of different kinds of controlled opinion feedback. Generally, these experiments aimed to measure the effect of feeding back either summary statistics or rationales, sometimes in addition to summary statistics, on experts' degree of opinion change and their forecast accuracy (i.e. the correspondence between experts' judgments and a verifiable true value (Woudenberg, 1991)).

Concerning the effect of feedback on experts' degree of opinion change, experiments produced rather similar results: no significant difference in the degree of opinion change was found between study participants who received summary statistics and those who received rationales (Rowe et al., 2005; Rowe and Wright, 1996) or rationales in addition to summary statistics (Bolger et al., 2011). Concerning the effect of feedback on experts' forecast accuracy, experiments produced mixed results. Best (1974) showed that study participants who received rationales in addition to summary statistics were significantly more accurate on one out of two questions than those who received summary statistics only. Rowe and Wright (1996) found that the improvement in accuracy across rounds did not differ significantly between study participants who received rationales and those who received summary statistics. However, in a replication of the experiment Rowe et al. (2005) discovered that only those study participants who received summary statistics showed a significant improvement in accuracy across rounds. Finally, Bolger et al. (2011) found no significant difference in accuracy improvement between study participants who received summary statistics and those who received rationales in addition to summary statistics.

From the experiments mentioned above it may be concluded that the advantages of feeding back rationales have not been convincingly demonstrated in terms of increased opinion change and forecast accuracy. This may be due to the design characteristics of the experiments: study participants consisted of university students and staff who had to answer rather trivial questions. As such, high quality rationales were perhaps not elicited. Several researchers criticized the oversimplification of these so called laboratory Delphi experiments, dismissing them as largely inappropriate (Rowe and Wright, 1999; Rowe et al., 1991).

Based on a review of numerous studies Woudenberg (1991) concluded that feeding back summary statistics induces conformity to the majority opinion. This is particularly troublesome as the original idea of providing controlled opinion feedback is that it counteracts group pressure. Furthermore, pressure to conformity may impede experts to achieve a genuine agreement (Woudenberg, 1991; Hung et al., 2008). Bolger et al. (2011) concluded that study participants tended to ignore feedback of rationales and merely used summary statistics to change their opinion. Therefore, they suggested eliminating any information concerning the majority opinion from feedback and solely present rationales. Empirical confirmation for this suggestion is lacking.

The current study took the reviewed literature into account by conducting an experiment within a real-world Delphi study in which actual experts were asked their opinion on issues that were relevant to them. Experts were presented with either the recommended feedback consisting of summary statistics and rationales (*S*&*R* condition) or the feedback as suggested by Bolger et al. (2011) consisting of rationales only (*R* condition).

2.2. Hypotheses on the effect of controlled opinion feedback on Delphi features

The feedback given in the *S*&*R* and *R* condition may influence Delphi features in several different ways. First, the two feedback conditions may have a different effect on the drop-out rate. Drop-out of experts is recognized as a serious methodological issue in Delphi studies (Keeney et al., 2006; Landeta, 2006; Hung et al., 2008; Powell, 2003). Nevertheless, no study could be found that examined the effect of different feedback conditions on the drop-out rate. Therefore, the following hypothesis was tested against the null-hypothesis (no difference):

H_1. There is a difference between the *S*&*R* and *R* condition concerning the drop-out rate.

Second, the two feedback conditions may have a different effect on experts' degree of opinion change. The Delphi experiments mentioned earlier found no significant difference in the degree of opinion change between study participants who received summary statistics and those who received rationales, whether or not in addition to summary statistics (Rowe et al., 2005; Bolger et al., 2011; Rowe and Wright, 1996). The current study differs from these experiments in two important ways. First, feedback was manipulated by providing rationales either with or without summary statistics. Second, the experiment was conducted in a real-world setting. Because the effect of feedback on experts' degree of opinion change has not been investigated in such a real-world experiment, the following hypothesis was tested against the null-hypothesis (no difference):

H_2. There is a difference between the *S*&*R* and *R* condition concerning experts' degree of opinion change.

Third, the two feedback conditions may have a different effect on the level of agreement among experts. This is of particular importance, because the current Delphi study involved not a forecasting task, but a policy formation task (Rowe and Wright, 1996, p. 75): a task "(…) where subjective opinions and views are sought because objective optimal solutions are difficult to specify". In a policy formation task experts try to find common grounds concerning for example the indicators needed to measure a particular concept, the guidelines to be incorporated in a new protocol, or the content of a future research agenda. In such tasks true values do not exist. Consequently, determining the accuracy of experts' opinions is impossible. Alternatively, it is essential to determine the level of agreement among experts.

Although many Delphi studies involve a policy formation task, remarkably little is known about the effect of different feedback conditions on the level of agreement among experts. Gowan and McNichols (1993) showed that experts who received feedback in the form of computer-generated if-then rules achieved a greater level of agreement than experts who received either of two kinds of summary statistics. While it is generally assumed that in a Delphi study the level of agreement among experts increases across rounds, no evidence could be found indicating that the increase in the level of agreement would be greater in one of the two conditions under study. Therefore, the following hypothesis was tested against the null-hypothesis (no difference):

H_3. There is a difference between the *S*&*R* and *R* condition concerning the increase in the level of agreement among experts.

Finally, experts may perceive feedback as more or less useful. Although some research examined experts' satisfaction with the Delphi method as a whole (see for example Boje and Murnighan (1982)) no study could be found that specifically investigated experts' perception of the usefulness of feedback. Finding out how experts perceived feedback may help to explain other Delphi features. For example, a limited increase in the level of agreement across rounds may be expected when experts perceived the feedback as rather useless. Therefore, experts' perception of the usefulness of the feedback as provided in the *S*&*R* and *R* condition, as well as experts' perception of the usefulness of summary statistics and rationales as separate feedback components, was further explored.