International Journal of Forecasting 34 (2018) 216–224

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The effects of feeding back experts' own initial ratings in Delphi studies: A randomized trial

Jurian Vincent Meijering *, Hilde Tobi

Biometris, Wageningen University & Research, Wageningen, Netherlands

A B S T R A C T

This study examined the effects of feeding back experts' initial ratings on three Delphi outcome measures: (1) the percentage of items on which experts changed their opinion; (2) the degree to which experts changed their ratings towards the group response; and (3) the increase in the level of agreement among experts. Additionally, two conformity indices were developed. Within a real-world Delphi study, experts were randomly assigned to one of two conditions: either their initial ratings were included in feedback (IN) or excluded from feedback (EX). Results showed that experts in the EX condition changed their opinion relatively more often than experts in the IN condition. Results also suggested that experts in the EX condition changed their ratings to a greater degree towards the group response than experts in the IN condition. No difference between conditions was found regarding the increase in the level of agreement.

© 2018 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

The Delphi method, developed in the 1950s (Dalkey & Helmer, 1963), was generally defined by Linstone and Turoff (1975, p. 3) as: "a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem". They further explained that the method is particularly useful when the problem at hand "does not lend itself to precise analytical techniques but can benefit from subjective judgements on a collective basis" (p. 4). Since its public introduction in the 1960s, many different types of Delphi designs have been developed that usually aim to allow a group of experts to predict future events accurately or to achieve agreement on a particular topic (Hasson & Keeney, 2011), such as the dimensions of a concept (Zill, Scholl, Härter, & Dirmaier, 2015) or a list of quality criteria (Verhagen et al., 1998). Regardless of the type and aim, the Delphi method has several defining characteristics (Dalkey, 1969; Keeney, Hasson, & McKenna, 2006; Rowe & Wright, 1999). A Delphi study consists of at least two rounds of data collection. In each round, experts give their opinion on the topic of interest, commonly by rating a number of items (pre-selected or developed during a first Delphi round) using a standardized questionnaire. Experts do not communicate with each other directly, but instead receive so-called controlled opinion feedback: a summary of the findings from the previous round. In light of this feedback, experts may reconsider and change their opinion in the next round.

Although controlled opinion feedback is a crucial aspect of any Delphi study, there is still debate as to what information should be fed back. Often, the feedback solely consists of summary statistics, showing a measure of location (e.g. mean, median) and dispersion (e.g. standard deviation, interquartile range) for each questionnaire item. This type of feedback has been criticized because it may induce experts to simply conform their opinion to the group response, thus creating an artificial agreement (Woudenberg, 1991). Therefore, various researchers have recommended that studies also feed back rationales, to give experts insights into why items were rated in a certain way (Bolger, Stranieri, Wright, & Yearwood, 2011; Murphy et al., 1998; Rowe, Wright, & Bolger, 1991). However, despite these recommendations, only a few experiments into the effects of different types of controlled opinion feedback have been conducted, and most of these experiments have focussed on the effects of feeding back summary statistics, rationales, or some combination of the two, on experts' degree of opinion change, forecast accuracy, or level of

* Corresponding author.
E-mail address: jurian.meijering@wur.nl (J.V. Meijering).

https://doi.org/10.1016/j.ijforecast.2017.11.010
0169-2070/© 2018 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

agreement (Best, 1974; Bolger et al., 2011; Meijering & Tobi, 2016; Rowe & Wright, 1996; Rowe, Wright, & McColl, 2005). Some experiments into the effects of other types of feedback have also been conducted, such as feedback in the form of computer-generated if-then rules (Gowan & McNichols, 1993) and feedback regarding experts' individual percentile locations (showing each individual expert the percentage of experts in the panel that gave a lower rating on each item in the previous round; see Dalkey, Brown, & Cochran, 1970).

A related issue concerns the feedback on each expert's own initial ratings from the previous round. Although feeding back this information has been recommended (Keeney et al., 2006; Murphy et al., 1998), a systematic review of Delphi studies regarding the selection of healthcare quality indicators by Boulkedid, Abdoul, Loustau, Sibony, and Alberti (2011) showed that only a minority of studies (39%) reported the feedback on experts' own initial ratings. According to Boulkedid et al. (2011), providing feedback on initial ratings is necessary because it informs experts about their position relative to the rest of the group, thus assisting them in making decisions in future Delphi rounds. However, feeding back experts' own initial ratings may also have disadvantages. Bolger and Wright (2011) stated that opinion change in Delphi studies is less than it could be because people tend to discount advice from others and favour their own initial opinion. This so-called egocentric discounting may be reinforced by feeding back experts' own initial ratings, thereby impeding experts from changing their opinion and achieving agreement. Surprisingly, no experiments into the effects of feeding back experts' own initial ratings could be found.

The current paper seeks to help to fill the identified knowledge gap by providing insight into the effects of feeding back experts' own initial ratings on three outcome measures: (1) the percentage of questionnaire items on which experts changed their opinion (i.e., by giving different ratings in the second round), (2) the degree to which experts changed their ratings towards the group response, and (3) the increase in the level of agreement among experts. Although the second outcome measure has not been used in any of the previously-mentioned experiments, it may shed some light on the debated issue of conformity. In addition, it may also be seen as an important intermediate step for achieving agreement in Delphi studies. An index was sought that estimates the degree to which experts in Delphi studies changed their ratings towards the group response presented in the controlled opinion feedback, and as no such index could be found, this paper also describes the development and application of two proposed conformity indices. Forecast accuracy was not used as an outcome measure, because the current Delphi experiment did not include a forecasting task, but a policy formation task in which experts give their subjective opinions on a particular topic (Rowe & Wright, 1996). It is impossible to measure the accuracy of experts' opinions, as there are no true values. Alternatively, it is important to measure the extent to which experts were able to achieve an agreement.

A hypothesis was formulated and tested for each of the three outcome measures. Considering the possibility that feedback on initial ratings reinforces egocentric discounting (Bolger & Wright, 2011), it might be expected that experts who received their own initial ratings would change their opinion relatively less often and change their ratings towards the group response to a lesser degree than experts who did not receive their own initial ratings. As no empirical evidence in support of this expectation could be found, the following two hypotheses were tested against the null hypothesis (no difference):

H_1: there is a difference between experts who received their own initial ratings and those who did not regarding the percentage of questionnaire items on which they changed their opinion.

H_2: there is a difference between experts who received their own initial ratings and those who did not regarding the extent to which they changed their ratings towards the group response.

If feedback on initial ratings impedes experts from changing their ratings towards the group response, it may also be expected that it will impede them from achieving agreement. In the absence of any empirical evidence, the following hypothesis was tested against the null hypothesis (no difference):

H_3: the increase in the level of agreement among experts who received their own initial ratings differs from the increase in the level of agreement among experts who did not receive their own initial ratings.

2. Materials and methods

2.1. Context of experiment

The current experiment was conducted within a real-world Delphi study that aimed to assist a panel of urban sustainability experts to achieve agreement as to the components (e.g., air quality, inequality, entrepreneurship) that are most relevant for defining and measuring urban sustainability in a European context. A total of three rounds of data collection were conducted, of which rounds one and two are relevant with regard to our experiment.

2.2. Expert sample

A convenience sample was compiled that consisted of European urban sustainability experts from four different types of institutions: academic, business, civil society (i.e., NGOs, non-profit, and community-based organisations that pursue charitable or member-oriented goals), and government. Most of the experts' names were obtained from the programs and proceedings of several urban sustainability conferences that were held in Europe in 2013 and 2014 (e.g., The Sustainable City Conference 2014) and projects funded by the Joint Programming Initiative Urban Europe, a program established by the European Commission that aims to create attractive, sustainable and economically viable urban areas (Robinson et al., 2015). In addition, various institutions and projects that were active in the field of urban sustainability (e.g. Arcadis, Climate-KIC, Amsterdam Institute for Advanced Metropolitan Solutions) were found by searching on the internet and talking to experts. These institutes and projects were contacted by telephone and asked whether they had urban sustainability experts who were willing to participate in the Delphi study. All in all, a final sample consisting of 419 experts from 26 European countries was obtained.

Download English Version:

https://daneshyari.com/en/article/7408116

Download Persian Version:

https://daneshyari.com/article/7408116

Daneshyari.com