



# Real-effort survey designs: Open-ended questions to overcome the challenge of measuring behavior in surveys

Caroline Fischer<sup>1</sup>

Based on data triangulation, open-ended questions can be used to overcome a typical problem with data collection using surveys: Human behavior can only be captured as stated or intended, but not as real behavior. In this study on knowledge sharing in the workplace, a quantitative measure of behavioral intention was accompanied by such a qualitative, open-ended measure of behavior. The latter was used as a proxy for real instead of stated behavior. This item was coded according to the effort a participant made in answering. It is assumed that the greater the effort put into answering the open-ended question, the more likely it is that the described behavior will be performed in reality. A factorial experimental design was used to analyze the effect of rewards on employees' knowledge-sharing behavior. As a within-subject design was used, participants had to answer three open-ended questions referring to different vignettes. A strong order effect appeared, leading to longer answers on average for the first vignette (baseline) compared to subsequent vignettes, independent of treatment. Therefore, this approach to operationalizing behavior in surveys might not be useful in within-subject designs. However, it can be used in between-subject comparisons when participants are asked to answer to a single vignette.

**Keywords** *survey design, survey experiment, real-effort design, human behavior, order effect*

Measuring human behavior is an ongoing challenge for social scientists (Schwarz & Oyserman, 2001). In ideal circumstances, human behavior should be observed either in the field or in the lab. However, field trips and lab experiments are often both time- and resource-consuming. Observing behavior can also be prohibited in some cases, for instance if the behavior occurs too infrequently to be observed in a timely manner, or is not possible for others to observe (Schwarz & Oyserman, 2001, p. 128). Additionally, in the field of public administration and other fields using sensitive target groups, it is often difficult to recruit participants for a lab experiment or to even be allowed to observe them in the field. If lab experiments are conducted, they usually rely on students as participants. However, student samples are not externally valid, especially when a special population is being studied, such as public employees.

Therefore, self-reported behavior in surveys and survey experiments is an often-used alternative (James et al., 2017, p. 120). Surveys are "quick and cheap" (Friborg & Rosenvinge, 2013, p. 1398) and can easily reach an intended population. However, these approaches to data collection come with other disadvantages, such as not measuring real behavior. One solution could be to integrate open-ended questions as a measure of answering effort, and thus a proxy for real behavior, into quantitative sur-

veys. This kind of method and data triangulation could help to overcome limitations of surveys related to the measurement of behavior.


Real-effort designs assume that greater effort in fulfilling a task makes it more likely that participants actually perform what they have stated (Dutcher et al., 2015). Fulfilling real-effort tasks are typically cognitively, creatively, or physically costly for participants (Charness et al., 2018). These costs are higher when more effort is put into fulfilling a task. In some studies, for example, participants are asked to count errors in real-effort tasks, thereby increasing the cognitive cost of the task (Andersen et al., 2018; Gneezy & List, 2006). Hence, it can be assumed that the more extensive, accurate, and fitting an answer to a question about future behavior is, the more likely it is that the described behavior will actually be performed in the future.

To test this approach to measuring behavior, a survey experiment was conducted on the effect of rewards on public employees' knowledge-sharing behavior. An open-ended measure of behavior accompanied a closed-ended question measuring behavioral intention. However, it turned out that differences in answers to the open-ended measure were determined by their order. Earlier answers were longer and more accurate, whereas subsequent answers were significantly shorter and, therefore, represented less effort. This unforeseen order ef-

<sup>1</sup> University of Potsdam

Received 29 September, 2020  
Accepted 12 July, 2021  
Published 10 Sep, 2021  
Issued

**Correspondence**  
University of Potsdam,  
August-Bebel-Straße 89, 14482  
Potsdam, Germany  
caroline.fischer.ii@uni-  
potsdam.de

**License**   
This article is licensed under the  
Creative Commons Attribution  
4.0 (CC-BY 4.0) license, which  
allows you to copy, redistribute,  
transform and build upon the  
article and all its contents in any  
medium or format for any  
purpose, provided that  
appropriate credit is given.

© Fischer 2021



fect made the qualitative data unusable for studying within-subject differences. Between-subject differences and differences between rotated questions did not suffer from this order effect.

This paper is organized as follows. Firstly, the state of research on the use of open-ended questions in surveys and real-effort experimental designs is described. Secondly, the results of the study are presented. Problems with the approach used to measure behavior with open-ended questions are highlighted. Subsequently, these issues are discussed, and recommendations are made for future research.

It is not possible to measure actual behavior in surveys and stated behavior is not always valid. Real-effort designs use real tasks in experiments to imitate real-world behavior and measure behavior more validly. This study combines both approaches, surveys and real-effort designs, and uses open-ended questions to measure effort in answering behavior as a proxy for actual instead of stated behavior. However, an order effect appears when this strategy is used to measure behavior in within-subject designs and, therefore, results are misleading.

### Open-ended questions in quantitative surveys as a measure of effort

Combining open-ended and closed-ended questions in a survey requires a pragmatic approach of method and data triangulation (Rossman & Wilson, 1985, p. 631). Quantitative and qualitative methods not only succeed one another, but are integrated within a single study. Due to this triangulation, "researchers are allowed to improve the accuracy of conclusions by relying on data from more than one method" (Rossman & Wilson, 1985, p. 632). Hence, the combination of qualitative and quantitative data is used here to offset weaknesses of the latter, draw on the strengths of the former, and enhance the validity of the findings (Bryman, 2006; Dewasiri et al., 2018, p. 105).

Several studies use this approach and compare answers to open-ended and closed-ended questions. Most of these studies report substantial discrepancies regarding answering behavior (Converse, 1984; Schwarz, 1999). For example, open-ended questions usually produce a more diverse set of answers. In a study reported by Schuman and Presser (1979), 60% of answers to an open question fell outside of the pre-coded answer categories of a related closed-ended question in the control group. Additionally, new types of responses occurred in a frequency that justified additional answer categories. In contrast, some answer options were less likely to be spontaneously produced in open-ended questions than they were to be chosen when explicitly offered as an option in closed questions (Schuman

& Presser, 1979, p. 365). Similarly, in a study on the quantity of alcohol consumption, Greenfield et al. (2006) found that the answers on the quantity of consumed alcohol of around 30% of respondents differed significantly between answers on open-ended and closed questions. They found that the combination of both an open-ended and a closed measure of the quantity of alcohol consumption was a stronger predictor of alcohol-related consequences than the individual measures (Greenfield et al., 2006).

The aim of studies using open-ended and closed-ended questions in combination is usually to measure attitudes, feelings, or past behavior. However, self-reports, especially on past behavior, are a complex cognitive task that are highly context-dependent, and data are often seen as unreliable (Schwarz & Oyserman, 2001, p. 128). Some authors state that it is easier for participants to understand questions and recall relevant behavior when closed-ended questions are used (Schwarz & Oyserman, 2001, p. 131). In contrast, it is usually agreed that open-ended questions are more suited to asking about sensitive or stigmatizing information or when more in-depth information is needed (Friborg & Rosenvinge, 2013).

However, I argue that the advantages of closed-ended questions only apply to questions on past instead of future behavior, where more emphasis lies on "editing" the answer. Hence, it is more important to get an accurate answer than a correct recall of past behavior. Accordingly, Singer and Couper (2017) suggest using open-ended questions in quantitative surveys more often to "encourage more truthful answers [...] [and use them] as an indicator of response quality" (p. 3).

Real-effort experimental designs rely on the assumption that greater effort in fulfilling a task makes it more likely that participants actually perform what they have stated (Dutcher et al., 2015). This assumption is based on the costs connected to fulfilling a real-effort task: The higher the cost an individual is willing to invest in answering, the more likely it is that these costs will also be invested in performing the actual behavior. Results based on real-effort designs are more externally valid when "the cost function of that task shares important characteristics with the field task" (Dutcher et al., 2015, p. 3). In surveys and survey experiments, open-ended questions can be used to measure effort in answering a question and, ultimately, we can use this as a proxy to measure real instead of stated behavior. Hence, the effort is for the most part cognitive and creative, but typing answers might also cost physical effort (Charness & Grieco, 2018, p. 75). Such a real-effort design might be a better estimator of actual behavior because "simply choosing a number may not capture the field environment and the psychological forces involved in putting forth actual effort" (Charness et al., 2018, p. 75). Greater effort can be operationalized with

an extensive, accurate, and fitting answer. This is based on different assumptions:

1. One dimension of effort is the *length of time* during which cognitive resources are used (Christensen-Szalanski, 1980). Accordingly, the length of a written answer is an approximator of the duration of the effort. Similarly, the length of answers to open-ended questions are often used as indicators of data quality (Galesic & Bosnjak, 2009, p. 350).
2. *Accuracy and errors* can also be used to capture effort (Charness et al., 2018, p. 78). Such a procedure was used by Gneezy and List (2006), who conducted an experiment in which participants had to enter library data into a database. They counted the number of errors in these entries as a measure of effort. Similarly, Andersen et al. (2018) registered an experiment in public administration research using this approach to compare work effort between public and private organizations. They asked participants to transcribe handwritten timesheets and checked their accuracy.
3. Furthermore, participants may not only be asked to transfer existing information into a database but also to be *creative*. Similarly, Charness et al. (2018) asked participants in an experiment to write a story about a predefined topic or by using specified words, and used this story as a measure for the real effort put forth in creative tasks. As it is difficult to rate creativity, the details provided in such an answer can be used to measure effort.

In such settings, participants' skills and abilities may strongly confound the results (Charness et al., 2018, p. 82). Longer answers, for example, may indicate that some participants can write quickly. Unfitting answers may indicate that some participants are unable to understand the question or articulate themselves. Charness et al. (2018) recommend using larger samples to capture this treatment effect. Furthermore, within-subjects designs might be useful to overcome these between-subjects differences.

### Putting a real-effort survey design into practice: The effect of rewards on knowledge-sharing behavior

A survey experiment on knowledge-sharing behavior was conducted to analyze whether tangible or intangible incentives can foster that behavior. Knowledge sharing is the exchange of knowledge among individuals, teams, units, or organizations (Paulin & Suneson, 2012). It is the basis of and a subprocess in an organization's knowledge management. In this article, the term "knowledge sharing" describes the behavior of donating knowledge from one person to another or to a medium. Multiple determinants influence knowledge sharing. Among others, tangi-

ble and intangible rewards are considered to foster knowledge-sharing behavior. Tangible rewards are material incentives, such as financial bonuses, and usually enhance extrinsic motivation. In contrast, intangible rewards, such as praise from a colleague or supervisor, usually influences intrinsic motivation. On the one hand, it has been shown that monitoring (e.g., controlling with performance measures) and rewards increase the knowledge-sharing activity of employees in an organization (Wang et al., 2011; Witherspoon et al., 2013). On the other hand, Bock and Kim (2002)'s results suggest that expected rewards do not affect knowledge sharing.

A 2x3 factorial survey experiment was designed to observe the within-subjects and between-subjects effects of the rewards offered. The research design was preregistered on the Open Science Framework (Fischer, 2018). Data were collected from German public employees in the core administration and health sector ( $N = 623$ ) in 2018 using a self-administered questionnaire. As can be seen from Appendix A, most participants were female (61%). The mean age was 45 years, and participants had a tenure of 20 years on average.

Each participant was randomly assigned a set of three vignettes from a pool of six. Randomization was done after participants started to answer the survey, based on their respondent ID. The vignettes incorporated all independent variables (Appendix B). Each set contained vignettes on either explicit or implicit knowledge (between-subject design). The first vignette in each set was a baseline vignette, while the two following vignettes presented a tangible and an intangible reward for knowledge sharing in a randomly assigned order (within-subjects design).

Performance appraisals were used as the *tangible reward* treatment (participants were given this reminder: "You know that all shared information improves your performance appraisal"). Performance appraisals served as a proxy for later rewards because it seemed unrealistic in the public sector to offer bonuses or other tangible rewards directly based on a person's knowledge-sharing behavior. This decision should address the problem that vignette experiments are frequently criticized for being unrealistic and lacking external validity (Aguinis & Bradley, 2014, p. 361).

The *intangible reward* treatment was operationalized by offering explicit appreciation from co-workers without further illustrating how this appreciation would occur (participants were given this reminder: "You know, your co-workers appreciate your knowledge sharing"). A rather superficial description was chosen because every team shows appreciation in a different way (e.g., good team climate, respectful interactions) and the vignette was thought to be not too restrictive.

*Knowledge-sharing behavior* (KSB) was measured with an open-ended question as a proxy for real

**Table 1** Descriptive Statistics of KSI After Treatment

Variable	N	M	SD	Min	Max
KSI (vig. 1)	319	4.23	.80	1	5
KSI (vig. 2)	318	4.31	.81	1	5
KSI (vig. 3)	318	4.27	.83	1	5
KSI (vig. 4)	310	4.29	.78	1	5
KSI (vig. 5)	307	4.24	.83	1	5
KSI (vig. 6)	309	4.29	.78	1	5
KSI (without treatment)	629	4.26	.79	1	5
KSI (intangible reward)	625	4.27	.82	1	5
KSI (tangible reward)	627	4.28	.80	1	5

**Table 2** Descriptive Statistics of KSB After Treatment

Variable	N	M	SD	Min	Max
KSB (vig. 1)	261	2.75	.91	1	5
KSB (vig. 2)	239	2.36	.70	1	5
KSB (vig. 3)	243	2.47	.77	1	5
KSB (vig. 4)	286	2.66	.86	1	5
KSB (vig. 5)	262	2.40	.68	1	5
KSB (vig. 6)	257	2.46	.71	1	5
KSB (without treatment)	547	2.70	.89	1	5
KSB (intangible reward)	501	2.38	.69	1	5
KSB (tangible reward)	500	2.47	.74	1	5

behavior. The item reads: "If you decided to share your knowledge, please briefly describe how exactly you will share your knowledge." Thus, participants were not asked to actually share their knowledge, but it was assumed that the cost function of answering the open-ended question on KSB was similar to the cost function of performing the described behavior in reality, for instance, when individuals share knowledge by writing an e-mail to a co-worker.

Additionally, *knowledge-sharing intention* (KSI) was measured in a closed-ended way to compare results to the verbatim responses. A measure of KSI was formed by taking the mean of two items, which were rated on 5-point Likert scales. The items read: "I will share this knowledge with my co-workers" and "I would like to share this knowledge with my co-workers." The former item represents the future-oriented and behavioral part of intention and the latter represents the motivational part of intention. Both items were based on scales on KSI used by, for example, Bock and Kim (2002) and Lin and Suneson (2007). Both were adapted to the experimental context. The items correlated highly (depending on the vignette:  $b = .66-.96, p < .001$ ) and were therefore compiled into one index for further analysis.

To code the open-ended answers, three dimensions were defined preliminarily, based on the literature, to rate the quality of the answer: the description's length, accuracy, and fit with the vignette. Two dimensions were added inductively during the coding process. Firstly, some participants answered the question with few words and therefore did not score in the quantity dimension, but were considered separately from participants who did not answer because they at least put some effort into answering the question. Secondly, some participants answered very elaborately in terms of length and content, and this effort was assigned a bonus point. Each dimension was coded as a dummy variable and an additive index was formed, ranging from zero to five.

Coding was done by two independent raters.<sup>1</sup> Interrater reliability was calculated using Gwet's Agreement Coefficient (AC; Gwet, 2014). Due to the fact that there was only partial agreement in the ratings of some dimensions (see Appendix C), the raters discussed differences in coding and specified coding rules. In a second step, both raters agreed on a rating. This "negotiated" rating was used for further analysis.

## Results and discussion of issues emerging from the open-ended question approach

Table 1 and Table 2 give a short overview of the descriptive statistics for the dependent variables. Examples of verbatim answers to the open-ended question on KSB and their coding are provided in Appendix D.

The descriptive statistics already show that when open-ended questions (*knowledge-sharing behavior*) were presented later in the survey, there were more missing answers, whereas such a high rate of missing answers was not observed with the closed-ended question (*knowledge-sharing intention*). Furthermore, in the second and third open-ended questions, participants often referred to their previous answers (e.g., "see above", "as just described") or answered in a significantly shorter way. Means for *knowledge-sharing behavior* were, therefore, significantly smaller in the second and third vignettes than in the first, which did not measure *knowledge-sharing intention*.

Data were analyzed using Wilcoxon signed-rank

<sup>1</sup>I want to thank Lisa-Sophia Preller for serving as the second coder of this study. Both raters were asked to use different coding strategies, to avoid effects of the coding procedure: (1) one rater coded the cases one by one in order to have an eye on the complete answer, (2) another rater coded each dimension across all cases one by one in order to have an eye on the comparability of the ratings. The fifth dimension (high effort in answering) was not coded by both raters but added afterward and coded by a single rater, therefore no interrater reliability was calculated here.

**Table 3** Correlation Matrix of KSI and KSB. *Note.*  $N$  baseline = 547,  $N$  intang. Reward = 501,  $N$  tang. Reward = 500. \* $p < .05$ . \*\* $p < .01$ .

	KSI without treatment	KSI appreciation	KSI achievement
KSB without treatment	.120**		
KSB intang. reward		.052	
KSB tang. reward			.083

tests (within-analysis) and Wilcoxon rank-sum tests (between-analysis) because the data were not normally distributed. Using the closed-ended measure of knowledge-sharing *intention* as the dependent variable led to a slightly greater intention to share when explicit knowledge was shared and a benefit was offered ( $M = 4.31$ ,  $SD = 0.80$ ) than without an offered benefit ( $M = 4.23$ ,  $SD = 0.79$ ;  $z = 3.23$ ,  $p < .001$ ,  $d = 0.09$ , bootstrapped 95% CIs [0.03, 0.16]).

However, using knowledge-sharing *behavior* as the dependent variable led to contradictory results, which might have been caused by a methodological problem. This issue was already identified while coding the data and inspecting the descriptive statistics: There was an order effect leading to longer answers on average for the first presented vignette compared to subsequent vignettes, independent of treatment. This result fits with the literature, as, for example, Galesic & Bosnjak (2009, p. 357) showed that open-ended questions asked later in a questionnaire were associated with shorter answers.

Accordingly, correlations between the stated knowledge-sharing *intention* and the coded answers related to knowledge-sharing *behavior* were not pronounced (Table 3). While knowledge-sharing *intention* and *behavior* were significantly but still only moderately correlated without an incentive treatment, the correlation was weaker and insignificant when incentives were induced. As the baseline vignette was always presented as the first vignette, this result supports the assumption of an order effect.

Analyzing knowledge-sharing *behavior* as the dependent variable, Wilcoxon signed-rank tests indicate significant differences between the treatment groups and the control group (explicit knowledge:

appreciation vs. control group:  $z = -5.86$ ,  $p < .001$ ; achievement vs. control group:  $z = -4.40$ ,  $p < .001$ ; implicit knowledge: appreciation vs. control group:  $z = -6.83$ ,  $p < .001$ ; achievement vs. control group:  $z = -4.89$ ,  $p < .001$ ). Contradicting the hypothesized relationships, the data show that knowledge-sharing *behavior* is significantly higher in the control group, and thus in the answer to the first vignette (baseline without treatment).

However, the comparison of the two treatments might not suffer from this methodological problem as they were presented in a randomly rotated order, thus either as the second or third vignette. Comparing the two treatments regardless of the kind of knowledge shared yields a significant difference. The tangible reward treatment ( $M = 2.47$ ,  $SD = 0.74$ ) triggered more knowledge-sharing behavior than the intangible reward treatment ( $M = 2.38$ ,  $SD = 0.69$ ;  $z = 2.40$ ,  $p = .017$ ,  $d = 0.10$ , bootstrapped 95% CIs [0.02, 0.18]).

These results show that open-ended measures in surveys are exposed to strong order and fatigue effects. Open-ended questions presented later in the survey result in more missing answers, shorter answers, or answers referring to previous statements. Therefore, behavior reflecting minimal instead of sacrificing answers was observed. Hence, these results could not be used in this study to analyze within-subject differences. However, the open-ended measure could be used as a proxy for real behavior in analyzing between-subjects differences and within-subjects comparisons between rotated vignettes (tangible and intangible rewards).

## Conclusion: Recommendations for future research

Using open-ended questions to measure effort as a proxy for real instead of stated behavior was not useful in a within-subjects research design. Due to an order effect, open answers could not serve as a reliable estimator of the likelihood of actual behavior. However, an advantage of verbatim answers is that they can still be used to identify behavioral patterns and modes of KSB. In further research, this strength of qualitative data should be taken into account more instead of merely quantifying qualitative data.

While this approach toward operationalizing behavior in surveys might not be useful in within-subject designs, it can be used in between-subject comparisons if participants are asked to answer to a single vignette. However, as participants' knowledge and competencies might influence answers on questions designed as real-effort tasks, attention must be paid to the sampling strategy when solely using between-subjects designs. When a baseline answer for an individual is missing or not taken into account, which is the idea behind a within-subjects design,

## Original Purpose

This study aimed at testing a more valid way to study human behavior in surveys. I tried to overcome the limited validity of self-reported behavior in surveys. I wanted to adapt experimental real-effort tasks to a survey design by using open-ended questions and analyzing the effort participants put into answering them.



an individual cannot be matched to their own standard. The order effect that occurred in this study was not expected by the author in that magnitude. It is known in the literature that, especially regarding open-ended questions, questions asked later in a questionnaire are associated with fatigue and shorter answers (Galesic & Bosnjak, 2009). However, it was expected that questions directly succeeding each other within a survey would not suffer that strongly from such an order effect. Additionally, the survey was rather short and took participants between 10 and 15 minutes to complete. The author expected that fatigue effects were more likely to occur in longer surveys. Also, the study's pretest with students and public sector professionals did not reveal such order and fatigue effects. Hence, this outcome might be due to characteristics of participants from online panels, who might be motivated to answer a questionnaire in an efficient and quick manner.

Due to the state of research on order effects in surveys, the treatment vignettes were randomized. However, I decided to exclude the baseline vignette from this randomization to avoid misunderstandings among the participants and to ensure the vignettes and questions were presented in a logical order. Future studies should, however, be aware of the occurrence of order effects in such an experimental setting, especially when data are collected with a sample of merely extrinsically motivated participants. To prevent order effects, between-subjects designs could be used instead of relying on within-subjects analyses. Apart from that, a combination of open-ended and closed questions allows one to detect order effects and should be included in case of uncertainty. Using both measures together also gives the option to analyze at least a part of the dependent variable and, therefore, minimizes risks of failure based on poor research designs.

Further research in this area may want to further consider the above-mentioned order and fatigue effects in answering open questions in survey experiments and how these might be prevented. Studies could, for example, experiment with rotating all vignettes instead of only rotating those with treatments to avoid an order effect. However, if the baseline vignette is not presented first, participants may have difficulties understanding the vignettes.

Furthermore, open-ended questions might be distributed throughout the survey instead of directly after one another in order to prevent participants from referring to earlier answers. However, questions asked in between may induce other treatments and thereby affect the answers. By reporting the failure of the research strategy used in this study, this article intends to serve as a step in further testing the possibility of using open-ended questions as valid measures of effort in surveys and survey

experiments.

## References

- Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, 17(4), 351–371. <https://doi.org/10.1177/1094428114547952>
- Andersen, S., James, O., & Jilke, S. (2018). Does work effort for public versus private organizations differ? Evidence from an online work task experiment: Pre-registration [Retrieved from]. <https://www.socialscienceregistry.org/trials/3361>
- Bock, G. W., & Kim, Y. G. (2002). Breaking the myths of rewards: An exploratory study of attitudes about knowledge sharing. *Information Resources Management Journal*, 15(2), 14–21.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97–113. <https://doi.org/10.1177/1468794106058877>
- Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149, 74–87. <https://doi.org/10.1016/j.jebo.2018.02.024>
- Charness, G., & Grieco, D. (2018). Creativity and incentives. *Journal of the European Economic Association*, 134(1), 48. <https://doi.org/10.1093/jeea/jvx055>
- Christensen-Szalanski, J. J. (1980). A further examination of the selection of problem-solving strategies: The effects of deadlines and analytic aptitudes. *Organizational Behavior and Human Performance*, 25(1), 107–122. [https://doi.org/10.1016/0030-5073\(80\)90028-8](https://doi.org/10.1016/0030-5073(80)90028-8)
- Converse, J. M. (1984). Strong arguments and weak evidence: The open/closed questioning controversy of the 1940s. *Public Opinion Quarterly*, 48(1B), 267–282. <https://doi.org/10.1093/poq/48.1B.267>
- Dewasiri, N. J., Weerakoon, Y. K., & Azeez, A. A. (2018). Mixed methods in finance research. *International Journal of Qualitative Methods*, 17(1), 160940691880173. <https://doi.org/10.1177/1609406918801730>
- Dutcher, G., Salmon, T., & Saral, K. J. (2015). Is “real” effort more real? *SSRN Electronic Journal. Advance online publication*. <https://doi.org/10.2139/ssrn.2701793>
- Fischer, C. (2018). Fostering knowledge sharing behavior: Preregistration. <https://osf.io/r5jws/>
- Friborg, O., & Rosenvinge, J. H. (2013). A comparison of open-ended and closed questions in the prediction of mental health. *Quality & Quantity*, 47(3), 1397–1411. <https://doi.org/10.1007/s11135-011-9597-8>
- Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73(2), 349–360. <https://doi.org/10.1093/poq/nfp031>
- Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, 74(5), 1365–1384. <https://doi.org/10.1111/j.1468-0262.2006.00707.x>
- Greenfield, T. K., Nayak, M. B., Bond, J., Ye, Y., & Midanik, L. T. (2006). Maximum quantity consumed and alcohol-related problems: Assessing the most alcohol drunk with two measures. *Alcoholism, Clinical and Experimental Research*, 30(9), 1576–1582. <https://doi.org/10.1111/j.1530-0277.2006.00189.x>

- Gwet, K. L. (2014). Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters (4th).
- James, O., Jilke, S. R., & Ryzin, G. G. (Eds.). (2017). *Experiments in public management research: Challenges and contributions*. Cambridge University Press.
- Lin, H. F., & Suneson, K. (2007). Effects of extrinsic and intrinsic motivation on employee knowledge sharing intentions'. *Journal of Information Science*, 33(2), 135–49.
- Paulin, D., & Suneson, K. (2012). Knowledge transfer, knowledge sharing and knowledge barriers—three blurry terms in km. *The Electronic Journal of Knowledge Management*, 10(1), 81–91.
- Rossman, G. B., & Wilson, B. L. (1985). Numbers and words. *Evaluation Review*, 9(5), 627–643. <https://doi.org/10.1177/0193841X8500900505>
- Schuman, H., & Presser, S. (1979). The open and closed question. *American Sociological Review*, 44(5), 692. <https://doi.org/10.2307/2094521>
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105. <https://doi.org/10.1037/0003-066X.54.2.93>
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *The American Journal of Evaluation*, 22(2), 127–160. [https://doi.org/10.1016/S1098-2140\(01\)00133-3](https://doi.org/10.1016/S1098-2140(01)00133-3)
- Singer, E., & Couper, M. P. (2017). Some methodological uses of responses to open questions and other verbatim comments in quantitative surveys. *Methods, Data, Analyses*, 11(2), 1–19. <https://doi.org/10.12758/mda.2017.01>
- Wang, S., Noe, R. A., & Wang, Z. M. (2011). Motivating knowledge sharing in knowledge management systems. *Journal of Management*, 40(4), 978–1009. <https://doi.org/10.1177/0149206311412192>
- Witherspoon, C. L., Bergner, J., Cockrell, C., & Stone, D. N. (2013). Antecedents of organizational knowledge sharing: A meta analysis and critique. *Journal of Knowledge Management*, 17(2), 250–277. <https://doi.org/10.1108/13673271311315204>

## Appendix A

**Table 4** Sample Description

Variable	N	M	SD	Min	Max
<i>age</i>	609	45.10	10.29	22	80
<i>female</i>	618	0.61	0.49	0	1
<i>supervisor</i>	615	0.29	0.45	0	1
<i>tenure</i>	615	20.20	11.23	0	47
<i>fixed-term employm.</i>	613	0.05	0.23	0	1

## Appendix B

**Table 5** Vignette plan

	Explicit knowledge	Implicit knowledge
<b>Without treatment</b>	<b>Vignette 1</b> <i>During a daily routine at your workplace, you gathered information from several sources. This could also improve the work of your co-workers. Please decide whether you will share this information with your co-workers. (N = 319)</i>	<b>Vignette 4</b> <i>During a daily routine at your workplace, you had an experience that improved your work process. This experience could also help your co-workers. Please decide whether you will share your knowledge with your co-workers. (N = 310)</i>
<b>Intangible reward</b>	<b>Vignette 2</b> <i>During a daily routine at your workplace, you gathered information from several sources. This could also improve the work of your co-workers. You know that your co-workers appreciate your knowledge sharing. Please decide whether you will share this information with your co-workers. (N = 319)</i>	<b>Vignette 5</b> <i>During a daily routine at your workplace, you had an experience that improved your work process. This experience could also help your co-workers. You know that your co-workers appreciate your knowledge sharing. Please decide whether you will share your knowledge with your co-workers. (N = 307)</i>
<b>Tangible reward</b>	<b>Vignette 3</b> <i>During a daily routine at your workplace, you gathered information from several sources. This could also improve the work of your co-workers. You know that all shared information improves your performance appraisal. Please decide whether you will share this information with your co-workers. (N = 317)</i>	<b>Vignette 6</b> <i>During a daily routine at your workplace, you had an experience that improved your work process. This experience could also help your co-workers. You know that all shared information improves your performance appraisal. Please decide whether you will share your knowledge with your co-workers. (N = 309)</i>



## Appendix C

**Table 6** Initial Interrater Reliability Open Answers on Knowledge-Sharing Behavior (KSB)

Vignette order			Gwet's AC	Cohen's Kappa	Extent of Agreement
Vignette 1	First vignette (v28)	answer provided	.9856***	.7430***	substantial
		quantity of the answer	.4674***	.4586***	moderate
		accuracy of the answer	.7522***	.6225***	moderate
		fit with vignette	.9090***	.5805***	moderate
Vignette 1	First vignette (v34)	answer provided	1.0000***	1.0000***	perfect
		quantity of the answer	.4943***	.3995***	fair
		accuracy of the answer	.8426***	.6480***	substantial
		fit with vignette	.8726***	.5010***	perfect
Vignette 2	Second vignette (v30)	answer provided	.9753***	.8308***	perfect
		quantity of the answer	.5041***	.3075***	fair
		accuracy of the answer	.7646***	.3530***	substantial
		fit with vignette	.9170***	.7075***	perfect
Vignette 2	Third vignette (v38)	answer provided	.9918***	.9332***	perfect
		quantity of the answer	.7475***	.4645***	substantial
		accuracy of the answer	.9087***	.3802*	perfect
		fit with vignette	.7888***	.4652***	substantial
Vignette 3	Third vignette (v32)	answer provided	.9642***	.8022***	perfect
		quantity of the answer	.5584***	.3438***	moderate
		accuracy of the answer	.7813***	.4196***	substantial
		fit with vignette	.8627***	.6546***	substantial
Vignette 3	Second vignette (v36)	answer provided	1.0000 ***	1.0000 ***	perfect
		quantity of the answer	.5302***	.2667***	moderate
		accuracy of the answer	.7887***	.3119**	substantial
		fit with vignette	.8417***	.5005***	substantial
Vignette 4	First vignette (v16)	answer provided	1.0000***	1.0000 ***	perfect
		quantity of the answer	.6097***	.5835***	moderate
		accuracy of the answer	.7829***	.6836***	substantial
		fit with vignette	.9360***	.6932***	perfect
Vignette 4	First vignette (v22)	answer provided	1.0000***	1.0000***	perfect
		quantity of the answer	.5520***	.5479***	moderate
		accuracy of the answer	.7015***	.5903***	moderate
		fit with vignette	.9149***	.6533***	perfect
Vignette 5	Second vignette v18	answer provided	.9830***	.7921***	perfect
		quantity of the answer	.6968***	.5625***	moderate
		accuracy of the answer	.7793***	.4512***	substantial
		fit with vignette	.8637***	.5912***	substantial

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

<i>Vignette 5</i>	Third vignette v26	answer provided	.9727***	.7570***	perfect
		quantity of the answer	.6613***	.4585***	moderate
		accuracy of the answer	.8328***	.3934**	substantial
		fit with vignette	.8888***	.6908***	perfect
<i>Vignette 6</i>	Third vignette v20	answer provided	1.0000***	1.0000***	perfect
		quantity of the answer	.6357***	.5394***	moderate
		accuracy of the answer	.5394***	.5487***	substantial
		fit with vignette	.8653***	.6624***	substantial
<i>Vignette 6</i>	Second vignette v24	answer provided	1.0000***	1.0000***	perfect
		quantity of the answer	.6873***	.5738***	moderate
		accuracy of the answer	.7614***	.4624***	substantial
		fit with vignette	.9132***	.6010***	perfect

\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

## Appendix D

### Selected open answers on knowledge-sharing behavior (KSB)

Note: Own translation of German answers

#### Core public administration

##### KSB-Index = 1 (does not fit the question)

- Everybody benefits from it.
- In my job, information and collegiality are the foundations of our conduct. It wouldn't work any other way.
- It pushes the team forward.

##### KSB-Index = 2

- In a one-on-one conversation
- Very precise
- 1: jour fixe 2: mailing list

##### KSB-Index = 3

- New information is transferred personally.
- A written report or e-mail
- Via e-mail or during coffee break time

##### KSB-Index = 4

- A group e-mail addressed to the department to inform about the outcome of the issue.
- Orally in conversation, and in some circumstances by looking at the file.
- I announce this experience within the team, so everyone can decide for himself or herself whether they can or want to use this information.

##### KSB-Index = 5

- If it is something very interesting in my point of view, I would forward an e-mail with the necessary documents to my colleagues. If it is an issue under the category "...just so you have heard about it...", I would seek direct talks and elaborate on the case.
- An electronic submission such as an appendix via e-mail, or a written submission such as copies distributed or an offer to approach me if required, depending on relevance.
- I prepare to address the issue in the next meeting and to share the new information with my co-workers.

#### Health sector

##### KSB-Index = 1 (does not fit the question)

- Leads to shorter working hours
- Because the simplification of work is good
- Improvement within the team

##### KSB-Index = 2

- In a team meeting
- In conversations
- By demonstration

##### KSB-Index = 3

- Not sure, it depends on the co-workers and their mood.
- I show them my approach.
- I address it briefly within the team and wait to see if somebody is interested.

##### KSB-Index = 4

- In a conversation/ team meeting I make suggestions to restructure work processes and ask for opinions of others, after which the team has to decide.
- I would post it in our company network or speak about it in a team meeting.
- I gather co-workers, who work in a similar task field, and tell them about my experiences.

##### KSB-Index = 5

- I will share this experience in personal conversation and will passionately talk about my findings. Hopefully, they will keep this knowledge in mind when they are doing their job.
- I would share this with my co-workers in the first situation possible and not wait for the next team meeting. I always get excited when I discover something new that simplifies my work and I want to share this joy with others instantly.
- I inform my manager and ask for permission to change something actively. When it comes to random things, which don't require official procedure instructions, I would train the new co-workers directly, which is the better way of doing things.