Author's Preprint of:
Wimmer, C., Biegler, S., Harms, J., Kappel, K., and Grechenig, T. (2018). SQT: A tool for the automated measurement of respondent behaviour and response quality in health-related gamified online surveys. In 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH), pages 1–8. IEEE.

# SQT: A Tool for the Automated Measurement of Respondent Behaviour and Response Quality in Health-related Gamified Online Surveys

Christoph Wimmer, Stefan Biegler, Johannes Harms, Karin Kappel, Thomas Grechenig
*Vienna University of Technology, Research Group for Industrial Software (INSO)*
*Wiedner Hauptstrasse 76/2/2, 1040 Vienna, Austria*
*{firstname}.{lastname}@inso.tuwien.ac.at*

***Abstract*—This paper addresses the necessity of taking detailed and accurate measures of respondent behaviour generally in health-related online surveys, and particularly in gamified variants of these surveys. Its specific contribution is the requirements analysis and implementation of an open-source survey quality tool (SQT) which automates the detection of negative respondent behaviour. The tool can be used in health and wellbeing research to evaluate the quality of responses in online surveys, as well as for researching the impact of survey gamification.**

## 1. Introduction

The Survey Quality Tool (SQT) presented in this paper allows researchers to empirically measure and detect negative respondent behaviour that affects data quality in online surveys. This is particularly important in the two areas of research that this paper focuses on: first, when using online surveys as a data gathering method, e.g., in health and well-being research, and second, when seeking to improve (possibly health-related) online surveys through gamification. The authors' personal motivation for developing the tool presented in this article stems from their own prior work in both of the above research areas, which revealed the need for a more detailed, data-driven understanding of how respondents behave in gamified health surveys.

Methodologically, we started by eliciting requirements for the survey quality tool. The tool was then implemented using web technology and is made publicly available along with this publication[1]. First practical experiences from using the tool in a case study are also published in this article, along with qualitative lessons learned.

The contributions of this paper are a systematic requirements analysis for automating the measurement of respondent behaviour in online surveys, as well as a first public release of the SQT tool implementing these requirements. The tool has been subject to a first, preliminary evaluation in a case study described in this paper. The contributions are directed at two target groups. First, the tool can be used by researchers who conduct online surveys as indicator of likely quality problems in response data. Second, the tool can also be used in research on survey design to compare the response quality in different design versions, e.g., when developing gamified survey designs.

The remainder of this paper is structured as follows. After describing related work and relevant foundations in the next section, Section 3 describes the requirements for and the implementation of the SQT. Section 4 then presents the health-related, gamified survey that was used as a case study for a first, preliminary evaluation. Corresponding results are presented as qualitative "lessons learned" in Section 4.3. Overall findings are then discussed in Section 5.

1. SQT: https://github.com/bigisoftStefan/SQT

## 2. Related Work

Surveys are a critical tool in healthcare. They are used to measure the health status, health behaviours and risk factors of the population and to assess the quality of received health care [1]. While the number of surveys in the healthcare domain is too numerous to provide a comprehensive overview, selected examples of health surveys on international, national and state levels can be found in [1]. With the increasing reach of the world wide web, online surveys (as opposed to paper- or telephone-based surveys) have increased in importance and popularity. Nonetheless, data quality of online surveys has been shown to be afflicted by negative respondent behaviour, as described in more detail in Section 2.2. According behaviour patterns introduce measurement error and lower data quality. The measurement and detection of such behaviour is an important prerequisite for proper response processing, which may involve the exclusion of low-quality responses.

### 2.1. Gamification of Online Surveys

A recent trend in online survey design is the use of gamification techniques. Gamification is defined as "the use of game design elements in non-game contexts" [2] to produce desired psychological and behavioural outcomes [3]. Hence, gamified survey designs employ game design elements in the context of online surveys [4].

One motivation for this area of research is that online surveys have been criticized for their dullness and lack of engagement, especially when there is a high respondent burden [5], such as questionnaires that are overly long or cover a topic of limited interest to the participant. This lack of engagement may result in negative respondent behaviour in order to finish the questionnaire more quickly (such as speeding or random responding) or it can lead to a participant's early termination of the survey before they are finished. This poses the problem of lower data quality for researchers, potentially skewing results.

The intended goals of online survey gamification are both psychological (e.g., making questionnaire filling a less boring and more enjoyable, engaging task) and behavioural benefits (e.g., to improve the response rate and data quality) [3], [4]. Both goals are also important in the context of online surveys about health and wellbeing.

Methodologically, gamification has been simplistically understood as addition of game elements into a non-game context [6]. One example of this approach is the addition of badges and achievements to a traditional survey design [7], [8]; further suitable game elements have been collected in related work [9], [10], [11]. Nonetheless, the simple addition of game elements has also faced criticism: Werbach and Hunter warn about the "lure of pointsification" [11], that is, the mindless addition of features least essential to games (such as points, thus "pointsification") in non-game contexts. Rather than this simplified approach, it has been argued that gamification should be understood as a complex design challenge that requires a holistic, creative and structured design process [4], [6].

Possible designs of gamified surveys have been explored in related work. Bailey et al. [12] distinguish between "hard gamification", where questions are embedded within a game, and "soft gamification", where more traditional web surveys are extended with game-like elements. A summary of experimental studies on web survey gamification by Keusch et al. [13] identified a broad spectrum of game mechanics that have been deployed in various studies, such as visuals, sound, avatars, scenarios and narrative framing, goals and quests, points, badges, progress indicators, levels, (time) challenges, feedback mechanisms and rewards. All of the studies deployed several (that is to say, more than one) of the above game mechanics, but none deployed all of them, highlighting the diversity of design approaches to gamified web surveys. Several studies have investigated the effect of different degrees of gamification: Downes-Le Guin et al. [14] compared four possible degrees ranging from text only, decoratively visual, functionally visual to fully gamified. Cechanowicz et al. [15] compared three designs, consisting of a plain survey, a partial game and a full game in their study. Mavletova [16] also evaluated three designs, in this case consisting of a text-only survey, a visual survey and a gamified survey.

Assessing the impact and effect of gamification on survey responses poses a challenge: previous studies have reported diverse (not always beneficial) psychological and behavioural effects regarding user experience, motivation, engagement, participation, satisfaction, enjoyment and the amount and quality of data. Hamari et al. [3] found that the effects of gamification were strongly influenced by users and context. Prior studies have reported beneficial psychological outcomes such as improved user experience [4], [14], [17] and increased motivation [15], as well as beneficial behavioural outcomes such as increased participation and engagement [15], [17], more elaborate feedback [18] and better data quality [17]. However, not all gamified surveys have produced significantly positive results [14].

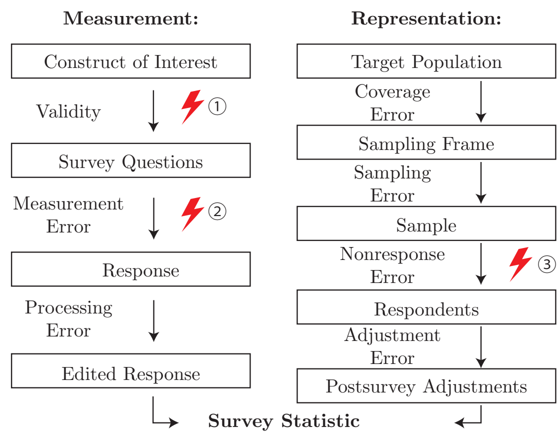


Figure 1. Groves' total error framework [22]; those components of statistical error that are potentially influenced (improved or worsened) by gamification are marked by red arrows. Figure from [4].

## 2.2. Negative Respondent behaviour

Negative respondent behaviour has also been described under the terms of careless responding [19], inattentive responding [19], and 'insufficient effort responding' [20]. The consequence of such behavior is the introduction of statistical error into survey results. This can produce bias and make spurious effects seem significant [19]. A common cause of negative behaviour patterns is when respondents are required to expend great effort for seemingly little or no reward. This can lead to a shift in response strategy called satisficing in order to reduce the cognitive effort necessary to finish the survey [21]. Satisficing, a portmanteau of satisfying and sufficing, describes an attitude and resulting behaviour where respondents expend a minimum or reduced amount of effort required in order to finish the survey, thus lowering data quality.

The overall statistical error that may be present in a survey's result can be understood using Groves' 'total error framework' [22] as consisting of several error components, as shown in Figure 1. The specific error components that may increase due to negative respondent behaviour are measurement error (e.g., due to biased answers) and non-response error (e.g., due to systematic non-response by a certain group of respondents). The benefits and potential pitfalls of gamified surveys can also be analyzed using Groves' framework [4]: in comparison to non-gamified surveys, the gamification of a survey design may reduce or increase any of three error components marked with red arrows in Figure 1. First, gamified questions influence construct validity if they correspond to a higher or lesser degree with the construct to be measured. Second, measurement error can be reduced if gamification successfully reduces negative respondent behaviour such as speeding, random responding, or lack of attention by motivating respondents to more deeply engage with the survey. On the other hand, game elements could bias the respondents' behaviour and motivation, or shift their attention away from their primary task of answering questions, thus increasing measurement error [13]. Third, non-response error is influenced if gamification leads to a different group of people responding (or not responding) to a gamified survey. For example, people who do not enjoy playing games might be less willing to participate in a gamified survey [14]. It is paramount for designers of gamified surveys to understand the impact of their design decisions on these three critical error components and to try to minimize them. The tool presented in this paper strives to support designers in this endeavor.

Specific patterns of negative respondent behaviour were identified in related work [19], [21], [23], and described along with metrics and thresholds for their detection based on user actions, as summarized in the following.

*Premature termination* describes a behaviour where respondents abandon a survey before they have successfully completed it. Premature termination directly impacts completion rate and leads to incomplete responses, which are commonly excluded from further analysis. Another possibility is to replace missing values with substitute values through weighting or imputation. Regardless of how incomplete responses are handled, they have a negative impact on data quality.

*Speeding* describes a behaviour where respondents answer a survey very quickly, which suggests that they are rushing through it as quickly as possible, without giving proper attention and care to the diligent answering of questions. Prior research has shown that speeding is linked to straightlining and that the two behaviours are likely to arise from a common satisficing tendency [23].

*Straightlining* is another behaviour indicative of satisficing and can occur when respondents are faced with several questions with the same response scales that are displayed in a table-like format, resulting in a grid structure of response options. Straightlining then describes a behaviour where respondents choose the same response option for every question in the grid, so that the selected responses form a straight, vertical line [21], [23].

*Don't know answers* can be another indicator of poor respondent engagement and motivation. Given the option, disengaged respondents might be inclined to claim they "don't know" an answer to save time and avoid further thought and consideration, even when they do know or could know the answer. A disproportionate amount of "don't know" responses thus indicates satisficing and careless responding [21], [24].

*Conflicting answers* are another sign of careless responding [19]. They can be observed by the use of bogus items or by examining conflicting answers, that is, the differences among responses to items that are highly similar in content or that are logically connected. Bogus items are items that could not possibly be true (e.g. "I was born on February 30th" [25]). An example of conflicting answers would be when respondents claim to be vegetarian in one question, and then contradict themselves by claiming to have eaten meat today in response to another question. While both answers could theoretically be true in the respondents' mind at the same time (depending on understanding and particular circumstances), it is still unlikely. Conflicting answers are indicative of low data integrity.

## 3. The SQT (Survey Quality Tool)

We analyzed requirements for creating a survey quality tool that automates the measurement and detection of negative respondent behaviour, and implemented one such tool according to the requirements.

### 3.1. Requirements

Requirements were first elicited from related work; these primarily concerned the behaviour patterns to be detected, as described in the previous section. Additional requirements were added based on the authors' experience in the field; these primarily concerned the technical implementation, but also compliance with data protection regulations to allow usage in a health-related context. The following requirements were defined for the survey quality tool:

R1: Overview on a survey's overall response quality
R2: Detailed measures of respondent behaviour, including custom thresholds (compare Section 3.2)
R3: Import/Export to support external data analysis
R4: Simple user interface, easy to understand visualizations
R5: Easy installation
R6: Flexibility to integrate with different survey platforms
R7: Special requirements for health-related surveys:
- In order for the tool to comply with information privacy regulations in the authors' country, the storage and processing of medical data must be separated from person-related data. Regarding the SQT, this requirement primarily concerns the logging of data used for measuring response quality and thus implies that logged data must not include medical or otherwise sensitive information.
- We reasoned that blind-folded (anonymized) studies implied no special requirements for the tool.

R8: Permissive, open-source license to ease dissemination and enable collaboration with the scientific community.

### 3.2. Measures of Respondent behaviour

The survey quality tool logs and analyzes the following five measures of respondent behaviour: premature termination, speeding, straightlining, "don't know" answers and conflicting answers (compare 2.2 for related work on these patterns of respondent behaviour).

**3.2.1. Premature Termination.** The tool detects whether a respondent has fully completed the survey or abandoned it prematurely. This measure is calculated as the percentage of incomplete responses in relation to the total number of responses to the survey.

**3.2.2. Speeding.** Speeding describes a behaviour where respondents rush through the survey without taking the time to carefully answer its questions. Speeding is measured by our tool as described in [26] and represented as an index. This index value ranges from 0 (speeding) to 1 (no speeding) and is calculated as follows: In a first step, the median of the duration of all respondents who completed the whole survey is calculated. To dismiss runaway values the top 5 percent quantile is excluded for calculating the median. For respondents who have a respond time between the median and the top 5 percent quantile an index value of 1 will be assigned. For respond times between one second and the median, the index value is calculated by dividing time (in seconds) through the median:

$$\text{Speeding Index: } I_s = T_{Respondent}/T_{Median}$$

Therefore, with decreasing response times (below the median) the index value will decrease. Based on the speeding index survey administrators and analysts can define a specific threshold for speeding detection.

**3.2.3. Straightlining.** Straightlining analysis detects a behavioural pattern where respondents select the same response option consecutively when presented with several questions and identical response scales. A series of consecutive response options on identical scales on a single page is required for straightlining detection. The LongString index is used for this purpose, defined as the longest string of the same response category on a single page. Since different cut off thresholds have been proposed [27] for this index, the SQT allows survey administrators and analysts to define a threshold of their own for straightlining detection. In addition to general straightlining, the presented tool detects and highlights two specific variants of straightlining: left edge straightlining and right edge straightlining, where a number of consecutive responses are detected in the left-most or right-most column in the response grid.

**3.2.4. "Don't know" answers.** A large amount of "don't know" or "no answer" responses are another indicator of satisficing [21], [24]. However, there are of course legitimate reasons for choosing these options when respondents truly don't know an answer or have no opinion on the subject matter. Therefore the SQT employs a LongString index, similar as for straightlining detection, in order to detect excessive, consecutive usage of "don't know" answers. Survey administrators and analysts can define their own threshold for "don't know" answer detection.

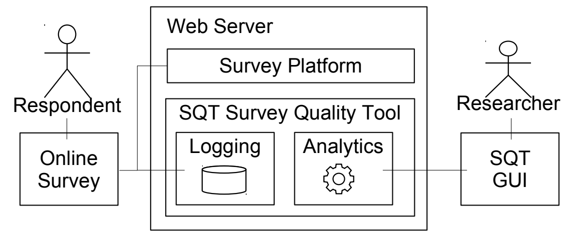


Figure 2. System Overview.

**3.2.5. Conflicting answers.** Conflicting answers are responses that are demonstrably false or that contradict themselves. Questions with potentially conflicting answers can be explicitly designed by the survey designers and defined as such in the survey platform of their choice. The SQT loads such definitions from the survey platform and analyzes the presence of conflicting answers in survey responses.

## 3.3. Architecture

The survey quality tool logs the respondents' behaviour while they fill out the survey, using a system architecture visualized in Figure 2. A web server hosts the survey platform as well as the survey quality tool, the latter of which is composed of logging and analytics components, and provides a web-based GUI (graphical user interface) for survey administrators and analysts.

*Survey Platform:* The online survey that is filled by respondents is provided by a survey platform, such as LimeSurvey [2] or SurveyMonkey [3]. Various products as well as custom survey implementations can be used for this purpose, provided they can load the JavaScript files needed for logging.

*Logging:* Respondent behavior is logged using the open source analytical tool Piwik. A client-side script includes platform-specific observers to detects meaningful events such as the answering of a question or the completion of the survey. The events are instantly sent to the Piwik server. An example log event is shown below.

```
{"actionDetails": {
  "eventCategory": "Answering",
  "eventAction": "Frage 1: How old are you?",
  "timeSpent": "15",
  "eventName": "A1;" }}
```

*Analytics:* The analytics component processes logged data in order to detect specific respondent behaviour patterns. Results from this analysis are shown in the web-based graphical user interface.

*Web GUI:* The Web GUI provides a user interface for survey administrators and analysts and presents detected negative respondent behaviour in an easy to understand way.

2. https://www.limesurvey.org/ (22.11.2017)
3. https://www.surveymonkey.com/ (22.11.2017)

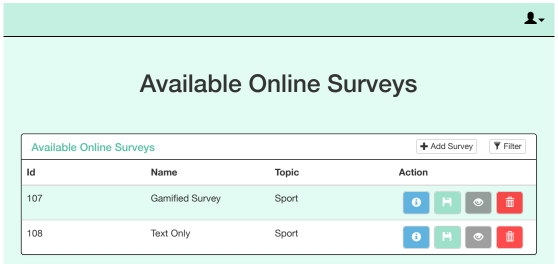


Figure 3. Screenshot of the SQT's List of Available Online Surveys.

### 3.4. Usage of the Survey Quality Tool

The tool provides a web-based administration interface for survey administrators and analysts. It consists of an overview of all the available online surveys, another overview for each individual online survey which highlights the presence of detected behaviour patterns, as well as detailed views of individual responses. This last view supports further analysis and the exclusion of individual responses on a case by case basis. Users can furthermore compare two online surveys side by side to get a quick overview of differences in response quality.

*User Administration:* The tool supports multiple user accounts for individual users. If enabled, analysts can register their own accounts in order to use the tool. Survey administrators can manage all registered users.

*List of Available Online Surveys:* After successful login, users are redirected to an overview of all the available online surveys that are presently stored in the database and accessible to the user (see Figure 3).

Survey administrators and analysts can add a new online survey for analysis on this screen. When adding a new survey, administrators and analysts can select which types of negative respondent behaviour (see section 2.2) should be monitored and detected by the tool. The survey data for analysis can be uploaded in JSON format. In addition, it is possible to filter the overview of available surveys. From this overview screen, administrators and analysts can select individual surveys to access a detailed view. In addition, they can manage their existing surveys and export survey data as a CSV file for later import into different statistics tools such as R[4] or IBM SPSS[5].

*Survey Quality Overview:* The survey quality overview screen provides an overview of the various negative respondent behaviours that are tracked and that were detected by the tool. Analysts can see the percentage of incomplete survey responses, the average speeding index of the survey, and whether any other types of negative respondent behaviour (straightlining, "don't know" answers or conflicting answers) were detected by the tool. Custom thresholds can be set in dialogues that are accessible from this screen. To further investigate the various measures on a case by case basis, analysts can select each individual measure to access a detailed view of all survey responses that exhibit the corresponding negative respondent behaviour. Furthermore it is possible to compare two different online surveys side by side, in order to get a better overview of similarities in respondent behaviour as well as quality differences between them.

4. https://www.r-project.org (22.11.2017)
5. https://www.ibm.com/analytics/at/de/technology/spss/ (22.11.2017)

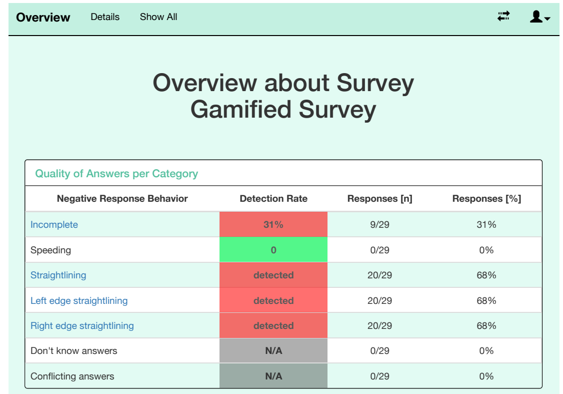


Figure 4. Screenshot of one Selected Online Survey in SQT.

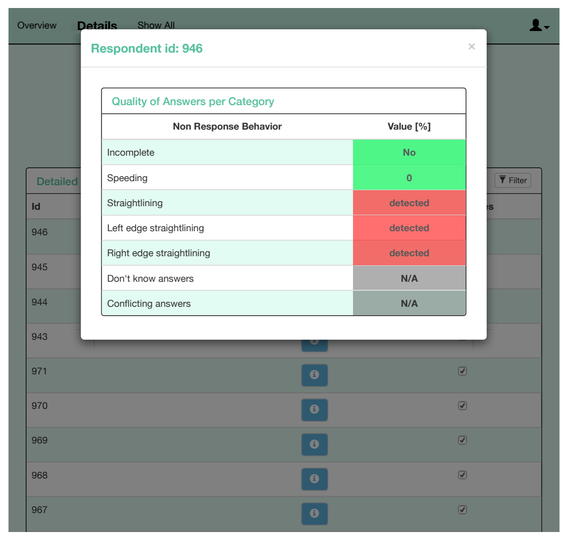


Figure 5. Screenshot of the SQT's Details View of an Online Survey.

*Survey Detail View:* Analysts can access a detailed view of the behaviour of the respondent in the currently selected online survey by switching to the survey detail view (see Figure 5). The view lists all available responses for the current survey. Analysts can filter this list based on the presence of detected behaviour patterns (e.g. speeding, etc) or can exclude individual cases from further analysis if the response quality is deemed insufficient. By selecting a response from the list, analysts can access a detailed view showing detected behaviour patterns in order to assess each individual response on a case by case basis.


Figure 6. Gamified survey design with achievement badges

## 4. Case Study: HealthSurvey

The survey quality tool was employed in a case study about sports and health-related behaviour amongst teenagers and young adults. The goal was to assess the validity of the requirements postulated in Section 3.1 and to experience how well the tool fulfilled these requirements.

### 4.1. Gamified and Conventional Survey Designs

The survey chosen for the case study is a publicly available online survey[6] about sports and health-related behaviour among teenagers and young adults. This survey was chosen due to its following, beneficial characteristics: The survey's questions are easy to understand and answer without requiring domain-specific expert knowledge from participants. Furthermore, the survey addresses teenagers and young adults, a target population known to respond well to gamification [16].

Two different survey gamification approaches were explored and evaluated in prior experiments:

One gamified design (Figure 6) explored the effects of a simple, low-effort and low-cost gamification process [28]: A set of 10 achievements, designed to encourage positive behaviour without introducing a bias to participants' responses were added to the survey. Each achievement consisted of a badge designed to fit the survey's sports theme. The collection of badges was displayed in the topmost part of the screen and served both as a visualization of past achievements, as well as further challenges in the shape of badges yet to be achieved. Aside from these game elements, the rest of the survey employed a traditional design.

The other gamified design (Figure 7) was created using a more elaborate gamification design process as outlined in [7]: The goal was to produce a design which elicits a rich visual sensation, includes micro-games for providing answers and allows participants to freely explore and discover various survey areas. The resulting gamified design shares little similarities with a traditional survey design, but features a highly game-like appearance instead.

6. https://www.jugendportal.at/befragung/bewegung-und-sport (22.11.2017)

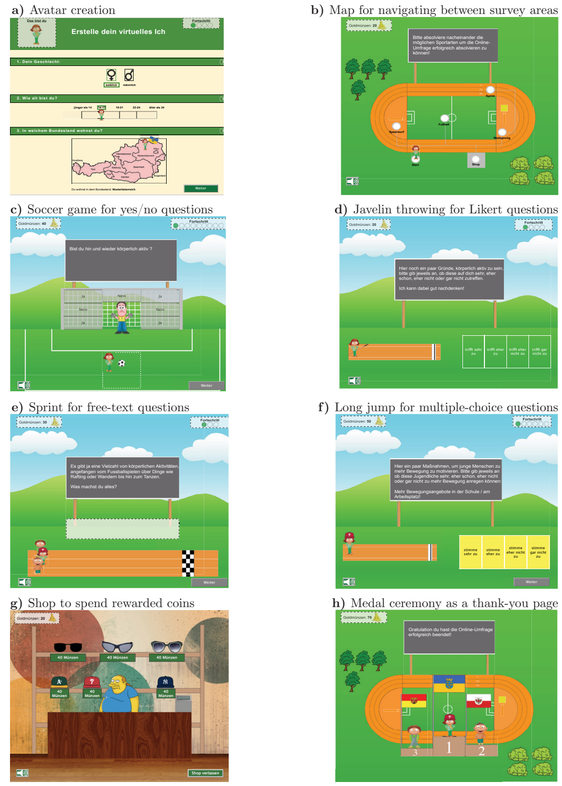


Figure 7. Gamified survey design with sports-themed micro-games

### 4.2. Methodology

By using the survey quality tool (SQT) in the case study, we subjected it to a first, preliminary evaluation. The goal was to gain experience and lessons learned from using the tool. Since detailed log data of respondent behaviour was readily available from both the gamified and non-gamified survey versions, we retrospectively imported and analyzed this pre-existing data. The authors then reflected on their experience with the tool to formulate the following results.

### 4.3. Results

Results obtained from the case study are lessons learned about the survey quality tool. More specifically, these lessons learned concern the validity of requirements formulated in Section 3.1, as well as how well the requirements were satisfied by the tool, as detailed in Table 1. In summary, the survey quality tool proved to be a valuable aid in assessing the overall quality of responses. It successfully demonstrated the presence (or absence) of the various kinds of negative respondent behaviour that are currently detected by our tool. In addition, the tool provided valuable insight about the response behaviour of individual respondents in order to decide whether to include these responses in further analysis on a case by case basis. In contrast to other methods of respondent behaviour analysis, which often favour a quantitative approach to quality assessment, the tool proved especially valuable as a diagnostic aid in order to facilitate a detailed assessment of individual response quality.

All of the requirements proved to be valid, i.e., we found that they referred to valuable and necessary functionality. We furthermore found that all requirements were satisfied at least in a basic manner and documented possible improvements, see Table 1. We discovered one additional requirement, namely support for ad-hoc hypothesis testing: While in its current form the tool allows to compare two surveys and to export data for further statistical analysis, statistical tests for comparing survey versions would be a useful improvement.

## 5. Discussion

This work set out to create a tool for the automated measurement and detection of negative respondent behaviour. Towards this goal, we presented a list of requirements, described the implementation of an according tool, and presented results from a first, preliminary evaluation in a case study. Results indicated all requirements to be valid and satisfied by the tool. Possible improvements and one additional requirement were identified and documented. Further evaluation and application of the tool is ongoing.

Besides functional requirements, the bigger challenge for future work is to formally validate the tool's output. The validity of implemented measures of respondent behavior currently relies on findings and definitions from related work (as described in Section 3.2), but has not been empirically and formally evaluated. The challenge with conducting such an evaluation is that the ground truth (whether or not a response was influenced by certain negative respondent behaviour) is unknown for many behaviour patterns such as speeding and straightlining. Validation of the tool's output may nevertheless be achieved by comparing it with simulated data. Such data could either be produced in a computer simulation or by instructing respondents to enact certain behaviour patterns.

In addition, a large number of different methods and thresholds for the detection of negative respondent behaviour has been proposed in related work and prior studies. Even though the tool currently affords analysts some flexibility in their analysis by defining their own thresholds and cut-off values for various types of detected behaviours, the tool could benefit from the inclusion of additional methods of negative respondent behaviour detection, including the use of machine learning.

Users of the survey quality tool should carefully interpret its output and, if in doubt, manually review questionable responses, e.g., using rich additional data such as online screen recordings. Despite the issues left for future work, the tool was successfully employed in the case study for its intended purpose, and it is ready to be employed by

TABLE 1. LESSONS LEARNED FROM THE CASE STUDY

| Requirements: Validity | Requirements: Fulfilment | Lessons Learned from the case study (Legend: ✓ ok, ☞ potential improvement for future work) |
|---|---|---|
| **R1: Overview on a survey's response quality** | | |
| ✓ | ✓ | Regarding this requirement, we found that the tool provided necessary overview by aggregating response quality data on several levels: It allows to drill down from a list of multiple surveys to an overview page for one specific survey, to detailed views for individual responses. |
| **R2: Detailed measures of respondent behavior** | | |
| ✓ | ✓ ☞ | The tool takes five measures regarding premature termination, speeding, straightlining, "don't know" answers, and conflicting answers and shows the results for these measures per survey and per response. ☞ Also, for some of the measures, such as speeding it would be interesting to compare different measurement approaches that have been described in related work. |
| **R3: Import/Export to support data analysis with external tools** | | |
| ✓ | ✓ | We used the tool's JSON import functionality to evaluate data that had previously been acquired, as described in the case study. The CSV export functionality provides the additional advantage of being able to perform further analysis of respondent behaviour using third-party statistical software, but was not used extensively in our case study. |
| **R4: Simple user interface and visualizations that are easy to understand** | | |
| ✓ | ✓ ☞ | This requirement proved to be very important in the case study. We detected misunderstandings, conducted informal "hallway" usability tests, and incorporated usability improvements. ☞ Future work: The tool would likely benefit from further design iterations. |
| **R5: Easy installation** | | |
| ✓ | ✓ ☞ | The SQT was installed as web application on a standard Apache + PHP + MySQL stack. ☞ Future work: Installation could further be eased by packaging the tool as a PHP 'composer' package. |
| **R6: Flexibility to integrate with different survey platforms** | | |
| ✓ | ✓ ☞ | The tool can be integrated with different survey platforms. It is currently integrated with LimeSurvey. New integrations require the implementation of additional connectors. ☞ In the future, integrations could further be eased by providing out-of-the-box connectors for more survey platforms. |
| **R7: Special requirements for health-related surveys** | | |
| ✓ | ✓ | The SQT logs only the user behaviour, but not the actual responses. |
| **R8: Permissive license** | | |
| ✓ | ✓ | Published under an MIT license; the source code is publicly available free of charge. |
| *Additional Requirements Uncovered during the Case Study* | | |
| **Hypothesis testing** | | |
| new | ☞ | Future work: The tool does not currently support ad-hoc hypothesis testing. This would be a useful addition that could be added in future work. |

others. It can particularly prove beneficial in areas similar to the case study, i.e., when using online surveys as data gathering method in health-related and other domains, and when evaluating the effects of innovative, gamified survey designs.

## 6. Conclusion

Online surveys are an important instrument for data gathering in health-related research and many other domains where data quality is important. But survey results have been shown to be afflicted by various patterns of negative respondent behaviour. The measurement and detection of such behaviour is also important for any research aiming to influence and improve respondent behaviour, for example by means of gamification. In light of the above, this work presents a tool that automates the measurement and detection of negative behaviour patterns. The specific contributions are a requirements analysis and the implementation of an according survey quality tool, which is made publicly available along with this publication. The tool was employed in a case study. This allowed to present lessons learned and discuss potential future improvements. In summary, results from the case study showed the requirements to be valid and satisfied by the tool. The tool can now be used in health and wellbeing research as an indicator of likely problems in response quality, as well as for researching the impact of survey gamification on respondent behaviour.

## References


[1] L. A. Aday and L. J. Cornelius, *Designing and conducting health surveys: a comprehensive guide*. John Wiley & Sons, 2006.

[2] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, “From game design elements to gamefulness: defining gamification,” in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. ACM, 2011, pp. 9–15.

[3] J. Hamari, J. Koivisto, and H. Sarsa, “Does gamification work?–a literature review of empirical studies on gamification,” in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*. IEEE, 2014, pp. 3025–3034.

[4] J. Harms, C. Wimmer, K. Kappel, and T. Grechenig, “Gamification of online surveys: conceptual foundations and a design process based on the mda framework,” in *Proceedings of the 8th Nordic conference on human-computer interaction: Fun, fast, foundational*. ACM, 2014, pp. 565–568.

[5] N. Bradburn, “Respondent burden,” in *Proceedings of the Survey Research Methods Section of the American Statistical Association*, vol. 35, 1978.

[6] M. Jacobs, “Gamification: Moving from ‘addition’to ‘creation’,” in *Proceedings of the ACM CHI 2013 Workshop on Designing Gamification: Creating Gameful and Playful Experiences*, 2013.

[7] J. Harms, S. Biegler, C. Wimmer, K. Kappel, and T. Grechenig, “Gamification of online surveys: Design process, case study, and evaluation,” in *Human-Computer Interaction*. Springer, 2015, pp. 219–236.

[8] S. Schacht, F. Keusch, N. Bergmann, and S. Morana, “Web survey gamification–increasing data quality in web surveys by using game design elements,” 2017.

[9] C. Lewis, N. Wardrip-Fruin, and J. Whitehead, “Motivational game design patterns of’ville games,” in *Proceedings of the International Conference on the Foundations of Digital Games*. ACM, 2012, pp. 172–179.

[10] B. Reeves and J. L. Read, *Total engagement: How games and virtual worlds are changing the way people work and businesses compete*. Harvard Business Press, 2009.

[11] K. Werbach and D. Hunter, *For the win: How game thinking can revolutionize your business*. Wharton Digital Press, 2012.

[12] P. Bailey, G. Pritchard, and H. Kernohan, “Gamification in market research Increasing enjoyment, participant engagement and richness of data, but what of data validity?” *International Journal of Market Research*, vol. 57, no. 1, pp. 17–28, 2015.

[13] F. Keusch and C. Zhang, “A review of issues in gamified surveys,” *Social Science Computer Review*, vol. 35, no. 2, pp. 147–166, 2017.

[14] T. Downes-Le Guin, R. Baker, J. Mechling, and E. Ruyle, “Myths and realities of respondent engagement in online surveys,” *International Journal of Market Research*, vol. 54, no. 5, pp. 1–21, 2012.

[15] J. Cechanowicz, C. Gutwin, B. Brownell, and L. Goodfellow, “Effects of gamification on participation and data quality in a real-world market research domain,” in *Proceedings of the First International Conference on Gameful Design, Research, and Applications*. ACM, 2013, pp. 58–65.

[16] A. Mavletova, “Web surveys among children and adolescents: is there a gamification effect?” *Social Science Computer Review*, vol. 33, no. 3, pp. 372–398, 2015.

[17] S. Dolnicar, B. Grün, and V. Yanamandram, “Dynamic, interactive survey questions can increase survey data quality,” *Journal of Travel & Tourism Marketing*, vol. 30, no. 7, pp. 690–699, 2013.

[18] J. Puleston, “Online research–game on!: A look at how gaming techniques can transform your online research,” in *Shifting the Boundaries of Research. Proceedings of the 6th ASC (Association for Survey Computing) International Conference*, 2011, pp. 20–50.

[19] A. W. Meade and S. B. Craig, “Identifying careless responses in survey data.” *Psychological methods*, vol. 17, no. 3, p. 437, 2012.

[20] J. L. Huang, M. Liu, and N. A. Bowling, “Insufficient effort responding: Examining an insidious confound in survey data.” *Journal of Applied Psychology*, vol. 100, no. 3, p. 828, 2015.

[21] J. A. Krosnick, “Response strategies for coping with the cognitive demands of attitude measures in surveys,” *Applied cognitive psychology*, vol. 5, no. 3, pp. 213–236, 1991.

[22] R. M. Groves, F. J. Fowler Jr, M. P. Couper, J. Lepkowski, E. Singer, and R. Tourangeau, “Survey methodology (2nd Edition),” *Hoboken: John Wiley and Sons*, pp. 97–98, 2009.

[23] C. Zhang and F. Conrad, “Speeding in web surveys: The tendency to answer very fast and its association with straightlining,” in *Survey Research Methods*, vol. 8, no. 2, 2014, pp. 127–135.

[24] S. Fricker, M. Galesic, R. Tourangeau, and T. Yan, “An experimental comparison of web and telephone surveys,” *Public Opinion Quarterly*, vol. 69, no. 3, pp. 370–392, 2005.

[25] D. A. Beach, “Identifying the random responder,” *The Journal of psychology*, vol. 123, no. 1, pp. 101–103, 1989.

[26] M. Steinbrecher, J. Roßmann, and M. Bergmann, “The short-term campaign panel of the german longitudinal election study 2009: Design, implementation, data preparation, and archiving; version 5.0. 0,” 2015.

[27] J. A. Johnson, “Ascertaining the validity of individual protocols from web-based personality inventories,” *Journal of research in personality*, vol. 39, no. 1, pp. 103–129, 2005.

[28] J. Harms, D. Seitz, C. Wimmer, K. Kappel, and T. Grechenig, “Low-cost gamification of online surveys: Improving the user experience through achievement badges,” in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, 2015, pp. 109–113.