Chapter 29
Gamification of Survey Research: Empirical Results from Gamifying a Conjoint Experiment

Briana Brownell, Jared Cechanowicz, and Carl Gutwin

29.1 Introduction

29.1.1 Background and Motivation

One of the most important tools utilized by the marketing research industry is the consumer survey. This self-reported data is the foundation of many currently applied methodologies for measuring the success of marketing campaigns and strategies (Evans & Mathur 2005). As such, suppliers in the marketing research industry rely on the engagement and attentiveness of the individuals who participate in their research and respond to their surveys. Keeping these respondents engaged is important for reducing the drop-off rate (the rate at which respondents quit before completing a survey), increasing time spent on surveys (which is linked to the quality and quantity of responses), and improving respondents' subjective enjoyment (since a happy respondent is more likely to complete future surveys). There is evidence to suggest that engagement has an influence on data quality as well, since bored or inattentive respondents produce lower quality data (Cape, 2009). Keeping respondents engaged and willing to participate in research is critical both to industry providers and to clients who use the results of the research for their decision making.

Recently, gamification has become a popular way to increase engagement and participation in tasks that would otherwise be considered work (Deterding, Dixon, Nacke, O'Hara, & Sicart, 2011; Von Ahn & Dabbish, 2008). Researchers have

B. Brownell (✉)
Insightrix Research, Saskatoon, SK, Canada

University of Saskatchewan, Saskatoon, SK, Canada
e-mail: Briana.Brownell@insightrix.com; bri.brownell@gmail.com

J. Cechanowicz • C. Gutwin
University of Saskatchewan, Saskatoon, SK, Canada
e-mail: jcechanowicz@gmail.com; gutwin@cs.usask.ca

T. Reiners, L.C. Wood (eds.), *Gamification in Education and Business*,
DOI 10.1007/978-3-319-10208-5_29

successfully utilized gamification in a wide range of subject areas (Deterding et al., 2011; Dignan, 2011; Von Ahn & Dabbish, 2008), demonstrating that a gamified design can result in improved user motivation, subjective preference, and data quality. Because all of these improvements would be desirable in the area of marketing research, some research suppliers and academic researchers have already attempted to integrate gamification into surveys (Adamou, n.d.a, n.d.b; Guin, Baker, Mechling, & Ruylea, 2012; Puleston & Sleep, 2011). The majority of their results demonstrate an increase in engagement, subjective preference, or data quality. While these results are promising, research into gamified surveys is still relatively new and only a small number of survey types have been gamified.

To help address these gaps in the literature, we applied gamification to a survey containing a choice-based conjoint experiment. Specifically, we investigated the effect of scenario realism, points and feedback for correct responses, and level of gamification as compared to a benchmark survey. To our surprise, the game elements had either no effect or a negative effect on engagement. Our results provide evidence that improved engagement is not a certain outcome of gamification. Further, it seems likely that the success of a gamified design depends greatly on the numerous design decisions made during the process, up to and including the nature of the task being gamified. There also appears to be a strong expectation component to a user's experience: the perception of participating in surveys is that it is useful and helpful, while the perception of participating in games is that it is entertaining.

In this chapter, we first discuss the previous research related to our study. We then provide a detailed analysis of our results as they relate to user engagement and data quality. Finally, we discuss the implications of our findings to gameful designers. In particular, we highlight the design decisions, gamification choices, and experiential factors at play in an outcome that seems contrary to much of the research into gamification.

29.1.2 Application of Gamification

Gamification, "the use of game elements in non-gaming contexts" (Deterding et al., 2011), has never been more popular, as evidenced by the multitude of design guides (Dignan, 2011; Hunicke, LeBlanc, & Zubek, 2004; Lewis, Wardrip-Fruin, & Whitehead, 2012; Puleston & Sleep, 2011; Reeves & Read, 2009) that have been published on the topic. Although gamification is still a new area for researchers, many preliminary works have already been published; psychological research has shown that the motivational roots of gamification are already well understood (Ryan, Rigby, & Przybylski, 2006; Sweetser & Wyeth, 2005), and many successful implementations of gamified designs have been tested and studied (Chrons & Sundell, 2011; Flatla, Gutwin, Nacke, Bateman, & Mandryk, 2011; Korn, 2012; Puleston & Sleep, 2011; Von Ahn & Dabbish, 2008). While this research indicates that gamification can work, there is an element missing: the demonstration of an empirical link between the amount of gamification and the level of user motivation.

Researchers have only just begun to approach this problem (Guin et al., 2012; Jung, Schneider, & Valacich, 2010; Mekler, Brühlmann, Opwis, & Tuch, 2013) leaving much future work to be done.

29.1.3 Gamification of Work

The "non-gaming context" for gamification has often been work- and task-related; indeed, some of the earliest work in the area of gamification has been with systems designed to increase user motivation and performance in tasks that would otherwise be considered work (Deterding et al., 2011; Dignan, 2011; Reeves & Read, 2009; Shneiderman, 2004; Von Ahn & Dabbish, 2008). The process of gamification has made office and industrial assembly jobs more motivating (Korn, 2012; Nikkila, Byrne, Sundaram, Kelliher, & Linn, 2013), has made annoying computer tasks more fun (Chao, 2001; Flatla et al., 2011), and has enabled citizen-scientists to participate in conservational and bio-chemical research (http://www.fold.it; Mason, Michalakidis, & Krause, 2012).

But what game elements actually lead to improved motivation and performance? Some researchers have sought to answer this question by empirically comparing game elements. For instance, Mekler et al. (2013) demonstrated that giving participants points increases motivation in an image tagging application, as measured by the number of tags users gave to images. However, the quality of the tags they entered increased when the users were told that their tags would help scientific research into image recognition; this "meaningful framing" of the task, along with points, resulted in the best overall performance from users.

Testing another pair of game elements in a distributed "idea generation" system, Jung et al. (2010) found that feedback on the ideas was far more important than providing set goals. Providing only feedback resulted in increased quality and quantity of responses, but providing only set goals actually resulted in a small decrease in the quality and quantity of responses. It was the combination of set goals and feedback that resulted in the highest level of user response quality and quantity.

29.1.4 Gamification of Surveys

The marketing research industry has empirically studied many of the factors that affect engagement and data quality in traditional and online surveys, including the effect of survey length (Brown, 2003), the effect of answering device (slider, textbox, etc.) (Guin et al., 2012; Malinoff, 2010), and the effect of social motivation (Thomas, Bremer, Terhanian, & Couper, 2007). However, there has been little research into the effects of gamification in the context of surveys.

In a study comparing four different versions of a survey, Guin et al. (2012) included one gamified version. This version included a narrative "quest" as a

motivating element, a customizable in-game avatar, and collectable items that were required to progress through the survey. The empirical results were inconclusive, but respondents seemed to prefer the gamified version over the others.

Much of the publicly released research into gamified survey design has been performed by Puleston and Sleep (Puleston & Sleep, 2011), and Adamou (Adamou, n.d.a, n.d.b) respectively. Over a number of studies they have tested gamified survey designs that include 3D environments (Puleston & Sleep, 2011; http://www.researchthroughgaming.com/), in-game story and narratives (Korn, 2012; http://www.youtube.com/watch?v=2s63gLuO-W0), response feedback such as point scoring systems (Guin et al., 2012; Puleston & Sleep, 2011), game timers (Puleston & Sleep, 2011), and in-game avatars (Adamou, n.d.a; Guin et al., 2012). In most cases, their results show that the addition of these game elements increases the length and quantity of responses, and respondents typically prefer the gamified version to the standard survey version. However, their research does not compare the effectiveness of game elements in gamified surveys. They have also found that that some gamified survey designs can lead to compromised respondent data (Puleston & Sleep, 2011).

As we have identified, there is a significant gap in the literature with regards to the effectiveness of using games as a way to gather marketing research data these in two key areas. First, it is unknown whether gamified data-gathering tools really work to provide greater engagement or participation among research participants. Some of the case studies discussed above have suggested value in the approach, but there is still little empirical evidence that directly compares gamified and non-gamified results. There is also little understanding about what aspects of gamification may be the source of any effect. Second, it is unknown whether the data gathered from gamified marketing research instruments is statistically reliable and reconcilable with traditional research methods.

29.1.5 Respondent Engagement in Survey Research

Presently, the most common method of data collection in the marketing research industry is through online surveys (Greenbook Industry Trends, 2013). Researchers conducting data collection by way of an online survey often rely on a participant pool recruited from the population of interest. Within the past decade, response rates have declined considerably (Cape, 2009), presenting an issue for researchers whose main objective is to collect accurate data. Interest in the subject matter of the survey, short survey length and relevance are commonly cited as motivating factors that encourage respondents to complete surveys (Ray & Tabor, 2003). However, individuals may not be passionate about subjects on which researchers would like to collect data.

Various strategies for increasing response rates and respondent engagement have been considered in survey research, such as incenting respondents to complete by providing a small monetary honorarium or entry into a prize draw. However, the effectiveness of these methods has been debated (Porter & Whitcomb, 2003). It has

been suggested that barriers due to disinterest cannot necessarily be overcome by monetary incentives: "Response will be improved if the survey is short, relevant and of interest to the respondent. Failure to meet these criteria cannot be compensated by incentives" (Ray & Tabor, 2003, pp. 35).

Comparisons in the respondent data to short and long surveys have shown that individuals are willing to participate in fairly long questionnaires; however, it has been noted that maintaining the respondents' engagement is a key factor in ensuring the data quality does not suffer: "It appears that even a surprisingly long questionnaire can be administered without large-scale and pervasive deterioration of the quality of the data, particularly if efforts are made to maintain respondent motivation." (Herzog & Bachman, 1981, pp. 559).

29.1.6 Conjoint Experiments

Conjoint experiments are commonly used methods for gathering information about the relative value of different product attributes. The method, developed by Luce and Tukey (1964), has been in use since the 1960s and has produced a wide body of literature.

The principal advantage to the methodology is that it allows researchers to collect information that is very difficult to gather accurately in other ways. Choice-based-conjoint (CBC) style experiments are the most common type of conjoint method used in marketing research (Carson et al., 1994), as it allows the direct calculation of preference share and most directly mirrors the real choice process (Carson et al., 1994; Orme, 2010). In this type of conjoint experiment, the respondent is asked to choose his or her most preferred product from several alternatives rather than to rate the products. Although easier for respondents, the choice-based approach requires each respondent to complete more choice sets in order to achieve the same level of accuracy (Orme, 2010). Therefore, utilizing this methodology means that respondents must complete a longer survey.

Researchers have focused on experimental designs that reduce the required number of choices respondents must make (Chrzan & Terry, 1995) rather than improving the overall experience of completing the task. Survey fatigue may negatively affect responses, and some researchers have found that respondents may be using artificial simplification strategies in their later choice sets (Johnson & Orme, 1996). At present, there is no universally accepted measure for how many choice sets are too many. It is difficult to estimate the number of choice sets that respondents would typically be asked to perform, as products vary widely in their complexity; however, Johnson and Orme's (1996) analysis of 21 datasets to assess the reliability of CBC data contained between eight and 20 choice sets per respondent.

Because of its usefulness for researchers but its difficulty for respondents, a conjoint experiment was chosen as a survey style for which respondents are in great need of increased motivation, and thus used in the empirical research presented herein.

29.2 Methodology

29.2.1 Conjoint Experiment Design

Conjoint experiments can handle a wide variety of product trade-offs. We selected a product for which the respondents would be familiar: a wireless cell phone package. Products of this type have many directly comparable features which have a definite ranking order. For instance, a cell phone package which includes 200 calling minutes is at least as good as the package which includes 100 min for the same price. For this reason, this type of product is ideal for use in a conjoint analysis. The only attribute for which there was no clear ranking order between attributes was the wireless provider. Table 29.1 shows the attributes and levels that were used in the conjoint experiment.

Because of the nature of the game (described below), we elected to use a CBC method rather than a traditional conjoint method. This approach corresponded nicely with the game scenario which asked respondents to choose the best of several options.

A randomized design was used to determine the choice sets. This method allowed us to use all of the data collected from respondents and also allows the estimation of the relevant parameters in an efficient way (Orme, 2010), despite the necessity to have each individual complete more choice sets as compared with a block design. Respondents could select one out of three different options, or they could select "none". This design was used to add to the realism of the choice task (Carson et al., 1994).

29.2.2 Measurement of Engagement

Two common measures of engagement in surveys are drop-off rates and self-reported enjoyment. Both have been used as the primary measure in a number of research studies involving gamification as well as other survey attributes such as length and complexity (Deutskens, de Ruyter, Wetzels, & Oosterveld, 2004; Guin

Table 29.1 Attributes and levels used in the conjoint experiment

Attribute	Number of levels	Levels
Brand	3	SaskTel/Telus/Rogers
Price	4	$30/$40/$50/$60
Minutes	3	100/200/400
Data	4	None/250 MB/1GB/2GB
Long distance	2	25¢ per minute/unlimited
Calling features	2	None/Call waiting, call transfer, three-way calling

Table 29.2 Self-reported engagement questions

Question	Scale
How enjoyable was this survey game?	(1—not enjoyable) to (7—highly enjoyable)
How motivated were you to complete the survey game?	(1—not motivated) to (7—highly motivated)
How interested were you in the subject matter of the questions?	(1—not interested) to (7—highly interested)

et al., 2012). Presumably, respondents who become bored are more likely to drop off part-way through the survey. However, some researchers have found a remarkable resilience to survey length and, despite declining response rates to online surveys, have found little effect on participation rates (Cape, 2009). For this reason, we did not wish to rely on participation and completion rates as our only measure of engagement. Time spent has also been used to measure engagement (O'Brien & Toms, 2010) in gamification research. We chose to include time spent as well in an attempt to quantify the degree of engagement of participants.

Because our main goal was to determine whether gamification would increase the quantity of data we were able to collect in a conjoint experiment, we directly used the number of choice tasks completed as the ultimate measure of the success the gamification of the research. The quantity of data collected has been used as a measure of success in gamification (Jung et al., 2010; Mekler et al., 2013) but to our knowledge, it has not been assessed in the context of survey research.

Following (Guin et al., 2012), we also included survey questions so that the respondents could self-report their engagement levels. A seven-point scale was used for three semantic-anchor questions. Table 29.2 lists the questions used.

We were interested to see whether an increased enjoyment of the survey games would also lead to a higher measured engagement in the subject matter. In other words, we were interested to know whether the way in which the subject was presented, within a game or within a survey, affects the subjective evaluation of how interesting it was perceived to be.

Finally, we collected qualitative data on the participants' subjective experience by allowing them the option to provide comments about the games in which they participated.

29.2.3 Applicability of Gamification to Conjoint Analysis

The gamification of the survey needed to ensure that the core tasks of the process remained the same. A CBC experiment consists of a series of choice tasks in which the respondent must evaluate and choose the most preferred alternative from several choices (Sawtooth Software Inc.,1993), and so we created a game where the goal mimicked this decision: a salesperson scenario. Respondents were situated in a

store and customers came in wanting to purchase a product. They were asked to serve the customer by choosing the best product from several alternatives (the choice task) for the customer.

Several elements of a gamification framework were included:

- **Fantastical scenario and immersive experience**: Two versions of the game were developed. The first was a "realistic" scenario where customers were located in a store. The second was a completely fantastical scenario where the customers were monsters from an alien planet.
- **Autonomy of the participant**: Instead of going through a survey where the questions are displayed automatically, respondents were able to "play" by choosing which customer they wished to serve in the game.
- **Novelty**: To add an element of novelty, the instruction statements which explained the task to the respondent were different for each customer they clicked on (although the sentiment stayed the same). In the most gamified version, we added humourous elements and sound effects.
- **Feedback and scoring system**: Participants were given points for successfully serving customers. In two of the versions tested, the instruction statement from the customer had a correct and incorrect answer, and the respondent was given feedback as to whether or not the customer was successfully served.
- **Progress**: The game was divided into rounds of six questions each. Although there was no increase in difficulty as the rounds progressed, participants would receive higher point rewards in the higher rounds of the game.
- **Time pressure**: We included a time element in the more fully gamified versions to determine whether adding a time pressure would have any effect on the engagement in the task.

Several gamification elements were considered for inclusion but ultimately were not included because of their potentially negative effects on the data quality:

- **Increasing challenge**: We did not include any change to the level of challenge throughout the game as the choice task requires the data to be collected in the same format.
- **Reputation or ranking**: No out-of-game implications such as a leader board was included as we felt that respondents may cease to pay attention to the task at hand in an attempt to make it to the top of the list, resulting in compromised data.

29.2.4 Experimental Design

In our research, we sought to answer four key questions in two areas. First, we wanted to determine the effect of gamification on the accuracy of the data collected, as determined by the similarity of the results between a survey version and a gamified version. Second, we sought to explore the effect of gamification on participant engagement. Our experimental design strived to answer the following questions:

- Does the act of completing a conjoint experiment in a game setting affect the results? Further, does the realism of the game scenario affect the results?
- Does providing responses in a game setting allow us to collect more data from each participant? Further, does the possibility of correct and incorrect answers improve respondents' engagement with the series of choice tasks?

29.2.5 Effect of the Game Setting on Engagement

We hoped that adding a gamification element to the conjoint experiment would encourage research participants to be more engaged in the research process. The five measures of engagement used were

- drop-off rates
- time spent participating in the research
- self-reported enjoyment, motivation and interest
- the proportion of minimum efforts
- the number of choice tasks done by each respondent

Drop-off rates were calculated by determining the number of respondents who dropped out of the survey part-way through and dividing it by the total number of respondents who started the survey, excluding those who were disqualified from participation due to demographic quotas.

When the respondent clicked into the survey, the time was recorded. The time was again recorded when they submitted their completed responses. Because the respondents could have opened the link in their browser and then left it open to complete later, the accuracy of this measure is somewhat ambiguous. Attempts were made to exclude outliers from the results: any time greater than 5 h were excluded from calculations.

After participating in the game, respondents were routed to an online survey system and asked to rate their enjoyment, motivation and interest on a seven-point semantic-anchor question. Since a conjoint experiment is tedious for the respondent, we expected low levels of engagement for the survey overall.

The proportion of minimum efforts and the total number of rounds completed were used to measure the extent to which respondents were willing to provide more data to researchers. Participants were told that they only needed to complete a single round (six questions) in order to receive their honorarium and they were able to continue playing as long as they wished. Therefore, the honorarium provided no additional incentive to continue participation. We measured the proportion of respondents who completed only the minimum number of choice tasks to receive their incentive.

We also directly used the number of rounds that the participants completed as a measure of their ongoing engagement in the various versions of the games. This allowed us to see whether those incented by the gamified versions had increased motivation to *continue* providing data.

Table 29.3 In-game statements used in the salesperson scenario game

In-game statement from the customer	Correct response
I need your help! Which is the best value?	No correct response
I'm not sure… Which should I pick?	No correct response
Can you do me a favour? Choose for me!!	No correct response
I've narrowed it down to these… Which do you think is best?	No correct response
"I don't want to spend much! I want the cheapest price!"	Least expensive
"I make a lot of work calls to cities across the country."	Must contain long distance
"I don't know what I'd do without internet on the bus!"	Must contain data
"I don't know how much I talk on the phone - but it's a LOT!"	Large amount of calling minutes

29.2.6 Additional Motivating Factors

As designed, a conjoint experiment does not have correct or incorrect answers since the purpose is to solicit an individual's subjective opinion of which is the best option. However, in a game setting, feedback is an important motivating mechanism. All of the gamified versions featured points that respondents would receive upon answering each question.

Incorporating the possibility of getting incorrect answers had the potential to motivate respondents by providing more challenging gameplay (Mekler et al., 2013; Reeves & Read, 2009). From a data collection standpoint, the effect on engagement would have to be substantial in order for it to be worthwhile to include these planted choice sets since they do not provide the researcher with data that is usable for the conjoint experiment results.

To measure the effect of correct and incorrect answers on engagement, we added planted choice sets. The respondent was asked to choose the best plan for the customer. The questions were phrased so that there were clear correct answers among the choices. Several examples of in-game statements used are summarized in Table 29.3.

The appearance of the planted choice sets was random and the ratio of planted choice sets to real choice sets was set to 50/50. We hypothesized that providing feedback to respondents would provide a more immersive experience and encourage them to continue playing.

29.3 Description of Game Versions

29.3.1 Version 1.0: Simple Choice Sets

The first version was as close as possible to a basic survey version. The respondent would click a "Next Question" button to reveal each conjoint question in the round. Respondents were instructed to select the best plan from the options presented. A screenshot of the survey-style version is contained in Fig. 29.1.

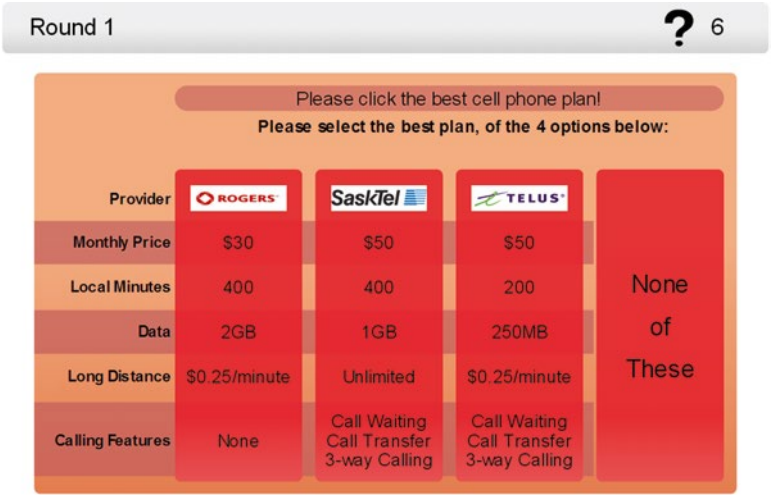

Fig. 29.1 Version 1.0 Survey-style Choice Sets

29.3.2 Version 2.1: Salesperson: No Correct Answers

The next, more gamified version of the game included animated human characters that represented each new question. Respondents were asked to pretend that they were the cell phone shop owner, and each conjoint question was revealed by clicking on the customer characters that walked onto the screen. The game story had basic sounds and graphics, and a simple story where the respondent was asked to help customers in the virtual store decide on their cell phone purchase. Figure 29.2 illustrates version 2.1 of the game.

29.3.3 Version 2.2: Salesperson: With Correct Answers

As in version 2.1, the respondents who played version 2.2 were asked to act as the cell phone shop owner and to recommend cell phone plans to customers. This version differed from 2.1 in several ways. Version 2.2, illustrated in Fig. 29.3, included planted cards with correct and incorrect responses, a point scoring system, and a game timer.

29.3.4 Version 3: Monsters: With Correct Answers

The final game version was intended to push the boundaries of our gamified design; we selected a fantastical game story and setting as a contrast to the more grounded and realistic game story from the other versions tested. The human characters were

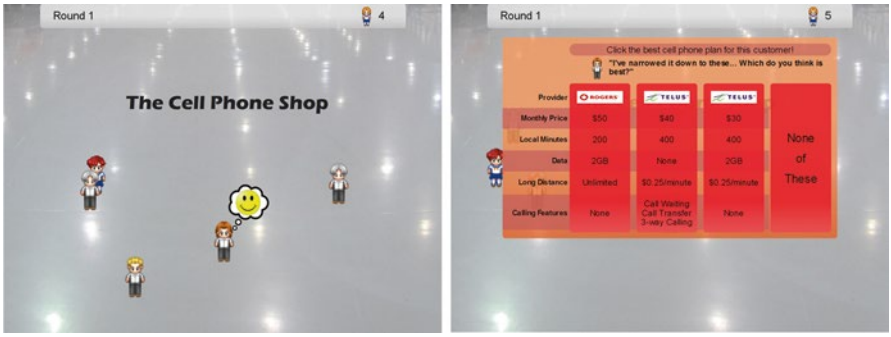

Fig. 29.2 Salesperson Scenario Game with No Correct Answers

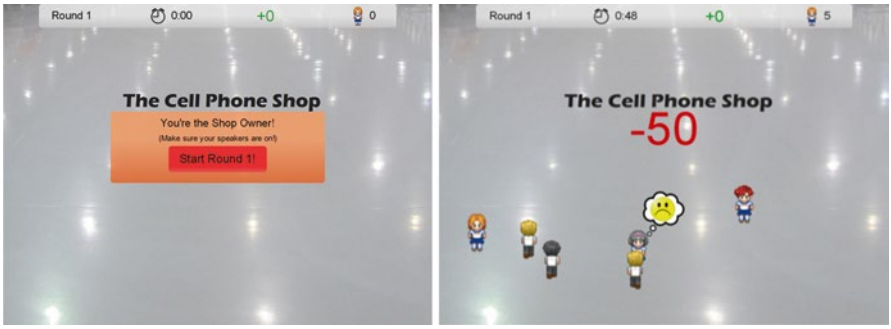

Fig. 29.3 Salesperson Scenario Game with Correct and Incorrect Answers

Fig. 29.4 Monster Scenario Game

Table 29.4 In-game statements used in the monster scenario game

In-game statement from the customer	Correct response
Monster stubbed toe—Ouch! Best phone plan will fix!	No correct response
Eee a rowr! (Translation: Help me pick out the best one.)	No correct response
ROAR!!!… Sorry… Habit. You help me pick?	No correct response
ARHR! Monster hate shopping!!! (Maybe just pick the best one for him)	No correct response
Me like stay up to date on news! Monster on Twitter! #ROOOOOAR	Must have data plan
Monster want browse internet on bus! No internet?? Monster EAT BUS!!	Must have data plan
What's WhoTube? My face-look? Monster don't need that! Monster just want PHONE!	No data plan
Monster like spend least possible! ME WANT FREE!	Lowest cost

replaced with monsters and the store environment background image was replaced with an alien planet to reinforce the fantasy setting. The task descriptions for each conjoint question were intended to be humorous, and we used additional audio and video elements (such as random talk and speech bubbles) to increase interest. Figure 29.4 illustrates the Monster Scenario Game used.

Table 29.4 shows a selection of the in-game statements used for the categorization of correct and incorrect answers.

29.4 Experimental Results

29.4.1 The Research Setting

Respondents who had previously consented to participate in research through Insightrix Research's proprietary online panel were invited to participate in the experiment. The sample was drawn from approximately 12,000 individuals and was balanced by age and gender. Each potential participant was sent an email inviting him or her to participate in the research and then randomly assigned to one of the four versions tested. Quotas were set to ensure that an appropriate demographic mixture of respondents participated in the research, as interest in games varies by these variables. Once a quota group was full, the respondent was disqualified from the research pool and was not able to participate.

Each version was programmed to show six questions per round. There was a maximum of 192 questions per game session, which represented the respondent seeing and evaluating every possible combination of levels and attributes.

Approximately 100 respondents participated in each of the various versions of the game. The sample sizes and the demographic breakdown are given in Table 29.5.

Table 29.5 Demographic breakdown of responses

		Version 1—simple choice sets	Version 2.1—salesperson—no incorrect answers	Version 2.2—salesperson—with incorrect answers	Version 3—monsters—with incorrect answers
Age	18–34	26	20	31	31
	35–54	38	35	36	35
	55+	38	35	35	35
	Total	102	90	102	101
Gender	Male	46	43	39	48
	Female	58	47	63	53
	Total	104	90	102	101

29.4.2 Drop-Off Rate

Approximately 40 % of those who were qualified to complete the survey dropped off before completion in each of the gamified versions. Because of the greater complexity of the conjoint experiment compared to a typical survey, it was expected that the drop-off rate would be higher than for a typical survey. This was indeed the case: other researchers have found a fairly consistent drop-off rate of 25 % in online surveys (Cape, 2009). Table 29.6 details the drop-off rates of the four different games.

Gamification did not significantly reduce the survey drop-off rate at the 95 % confidence level, although the sample size may not have been powerful enough to show a difference. Respondents who participated in the regular survey-style conjoint experiment were just as likely to discontinue participating in the survey as those who played the gamified versions.

No effect was found on drop-off with regards to the inclusion of correct and incorrect answers: including questions that respondents could get wrong did not affect the drop-off rates of the survey.

Additionally, the scenario realism also had no significant effect on the drop-off rates. Respondents were just as likely to drop out of the realistic version of the game as the fantastical version.

29.4.3 Time Taken

The total time take to complete the survey was measured for each respondent to determine whether he or she would spend more time on the task in the gamified version. The time taken includes the total time participating in the research: answering the beginning demographic questions, playing the games or participating in the survey, and completing the final few survey questions where they rated their enjoyment and provided comments.

Table 29.6 Drop-off rates

	Version 1—simple choice sets	Version 2.1—salesperson—no incorrect answers	Version 2.2—salesperson—with incorrect answers	Version 3—monsters—with incorrect answers
Total completes	104	90	102	101
Total drop-offs	128	58	68	63
Total qualified to complete	232	148	170	164
Percentage of drop-offs over the total who started	55 %	39 %	40 %	38 %

Table 29.7 Time spent

	Version 1—simple choice sets	Version 2.1—salesperson—no incorrect answers	Version 2.2—salesperson—with incorrect answers	Version 3—monsters—with incorrect answers
Average time spent (minutes)	13.42[a]	10.69	11.83	12.67
Standard Deviation	*12.87*	*11.40*	*13.02*	*16.83*
Standard error	*1.27*	*1.20*	*1.29*	*1.68*

[a]One outlier with a time of more than 5 h—corresponding to a respondent who may have left his or her browser window open—was excluded

We expected respondents to spend more time, on average, taking part in the research process when they were participating in the gamified versions. However, respondents spent approximately the same amount of time or longer completing the simple choice sets as they did playing the games. The average time taken for each version of the game is given in Table 29.7.

A univariate ANOVA test of average completion times did not find a significant difference between the four versions ($F_{3,390}=0.517$, $p=0.67$). The realism of the scenario did not significantly affect the time spent participating in the research; respondents spent a similar amount of time playing the realistic game as the fantastical game. There was also no significant effect on the time spent when correct and incorrect responses were included.

29.4.4 Self-Reported Engagement

After respondents decided to stop completing choice sets, they were asked to rate the enjoyableness of the games that they played. They were also able to provide comments about their experience.

Table 29.8 Self-reported engagement

	Version 1: simple choice sets	Version 2.1: salesperson—no incorrect answers	Version 2.2: salesperson—with incorrect answers	Version 3: monsters—with incorrect answers
Enjoyment (How enjoyable was this survey game?)	2.9	3.0	3.4	3.1
Standard deviation	*1.79*	*1.69*	*1.90*	*1.83*
Standard error	*0.18*	*0.19*	*0.18*	*0.18*
Motivation (How motivated were you to complete the survey game?)	3.0	3.3	3.5	3.1
Standard deviation	*1.70*	*1.78*	*1.88*	*1.82*
Standard error	*0.18*	*0.19*	*0.18*	*0.18*
Interest (How interested were you in the subject matter of the questions?)	3.5	3.7	3.6	3.0
Standard deviation	*1.76*	*1.73*	*1.86*	*1.72*
Standard error	*0.18*	*0.19*	*0.18*	*0.18*

Table 29.8 shows the average ratings for self-reported enjoyment, motivation and interest in the subject matter. Ratings were given on a seven-point semantic-anchor scale.

Univariate ANOVA tests showed that gamification had no significant effect on respondent ratings of how enjoyable the game was ($F_{3,390} = 1.12$, $p = 0.34$) or how motivated they were to complete the survey ($F_{3,390} = 1.95$, $p = 0.12$). We did find a significant difference in respondent interest in the subject matter of the questions ($F_{3,390} = 2.84$, $p = 0.038$).

As there were no significant differences found between versions 2.1 and 2.2, the addition of correct and incorrect answers in the salesperson scenario does not seem to have had an effect on enjoyment, motivation or interest. However, it seems likely that some element of the design of version 3 had an effect on respondents' interest in the subject matter: the fantastical setting or the different tone and writing style of the in-game customer statements *decreased* interest in the subject, opposite to our expectations.

The comments from the four versions we deployed showed that there was a marked difference between the non-gamified and gamified versions. The general tone and types of comments we received regarding the non-gamified version were typical for any survey; generally minor complaints, with no major positives or negatives cited. Here are a few examples:

- "*This was too repetitious*."
- "*A bit too long*."
- "*Did not know when the rounds would end—got tired/bored*."

However, the tone of the comments from the gamified versions tended to two extremes. Some of the comments we received were strikingly positive, such as the following:

- "*Do more surveys in the form of games, it's much more fun*!"
- "*Fun*!"
- "*I love games. I was quite excited to do this one*."

On the other extreme, some respondents were clearly not pleased with various aspects of the gamified design. Here are a few of these types of comments:

- "*I am not motivated by highschool-looking characters*."
- "*The game was stupid…*"
- "*What was the point of the survey. Waste of time*."

Our gamified designs were quite divisive; it seems clear that respondent expectations and preferences must be strongly considered when designing and deploying gamified surveys. Some respondents will be more motivated by games and game elements than others, and what seems like a well-selected game mechanic may not appeal to everyone. Moreover, the positive and negative qualitative comments were spread across demographics and gaming experience. Clearly some respondents may prefer the standard survey version over even a perfectly designed gamified survey, and it would be prudent to retain this option for such respondents.

29.4.5 Number of Choice Tasks Completed

To our surprise, the game elements did not have a positive effect on the number of choice sets that respondents completed. Although a univariate ANOVA test did not show a significant difference ($F_{3,390} = 1.17$, $p = 0.322$), respondents who received the survey-like version completed an average of 10 *more* choice sets than in any of the game versions, completely contrary to expectations.

As there is no significant difference between the number of choice sets completed between the three gamified versions, it seems that the inclusion of correct and incorrect answers (for versions 2.2 and 3) did not have a notable effect on the average number of choice tasks completed. Table 29.9 summarizes these results.

It is also worth noting that the average number of choice sets completed by respondents was substantially higher than the eight to 20 choice task range in the data sets analysed by Johnson and Orme (O'Brien & Toms, 2010). This suggests that respondents may be willing to complete many more choice sets than are commonly employed in CBC research studies.

Table 29.9 Number of choice tasks completed

	Version 1—simple choice sets	Version 2.1—— salesperson—no incorrect answers	Version 2.2— salesperson—with incorrect answers	Version 3—monsters— with incorrect answers
Average number of choice sets completed	47	37	36	37
Standard deviation	52.9	48.2	39.6	45.8
Standard error	5.08	4.73	4.32	4.21

Table 29.10 Number of choice tasks completed—percentages

	Version 1—simple choice sets	Version 2.1— salesperson—no incorrect answers	Version 2.2— salesperson—with incorrect answers	Version 3—monsters— with incorrect answers
Percentage who did at least 36 choice tasks	41 %	24 %	33 %	30 %
Percentage who completed all 192 choice tasks	7 %	3 %	0 %	3 %
Percentage who did only the minimum (6 choice tasks)	21 %	29 %	29 %	29 %

29.4.6 Minimum Efforts

The percentage who completed the minimum number of choice tasks to receive the incentive varied slightly across the versions; however, the difference was not statistically significant at the 95 % level. Directionally, the non-gamified version again proved to be the best.

Some respondents exhibited a remarkable willingness to complete the choice tasks and completed all (192) choice tasks possible in the entire experiment, which was much greater than expected.

The proportion who did the minimum was lowest among those who completed the simple survey. Table 29.10 summarizes the minimum efforts across all four versions.

29.4.7 Reliability of Results

The results from the five games showed a remarkable amount of consistency in the calculation of the part worth utilities from the conjoint dataset Table 29.11 summarizes these results. This result suggests that there was no negative effect on the

Table 29.11 Reliability of results

Price	Version 1—simple choice sets	Version 2.1—sales-person—no incorrect answers	Version 2.2—sales-person—with incorrect answers	Version 3—mon-sters—with incorrect answers
$30	33 %	36 %	36 %	31 %
$40	30 %	27 %	24 %	28 %
$50	22 %	22 %	23 %	20 %
$60	14 %	15 %	17 %	21 %
Provider				
Brand 1	44 %	48 %	45 %	47 %
Brand 2	28 %	24 %	27 %	24 %
Brand 3	28 %	28 %	28 %	30 %
Minutes				
100	25 %	26 %	29 %	26 %
200	36 %	35 %	33 %	32 %
400	39 %	39 %	38 %	43 %
Data				
None	15 %	14 %	17 %	15 %
250 MB	20 %	23 %	25 %	24 %
1GB	29 %	29 %	27 %	29 %
2GB	36 %	34 %	31 %	32 %
Long distance				
$0.25/min	40 %	41 %	42 %	44 %
Unlimited	60 %	59 %	58 %	56 %
Calling features				
No calling features	45 %	45 %	41 %	44 %
Call waiting, call transfer, 3-way calling	55 %	55 %	59 %	56 %

accuracy of the results despite the drastic changes in the realism of the different games. These results suggest a high degree of engagement already in the respondent population, which did not appear to be improved by gamifying the research process.

29.5 Summary and Conclusions

Our results contrast some of the results in the literature where gamification has been shown to assist in engagement in tedious tasks. None of our measures of engagement were improved by gamifying the conjoint survey, and one, interest in the subject matter, was significantly reduced. The addition of the gamification elements also had, if anything, a detrimental effect on the *amount* of data that respondents

were willing to provide. This result contrasts those of Mekler et al. (2013) and Jung et al. (2010) who both found an increase in the quantity of data provided by respondents.

It is important to note that, while gamification reduced the number of questions completed, the data collected did not appear to be negatively affected. Our results on drop-off rates are also consistent with Guin et al. (2012) who also found no difference in the engagement measures of survey drop-off and completion rates in their gamified version. Guin et al. also found that gamification increased self-reported engagement. However, our results did not show a strong effect on the enjoyability of the task or motivation to complete the task, and interest in the subject matter was lower in the very fantastical scenario. As well, including questions with correct and incorrect responses did not appear to influence our engagement results significantly between versions.

Our results could have been caused by several factors. Previous researchers have highlighted the importance of carefully pairing work tasks and game elements when developing a gamified design (Von Ahn & Dabbish, 2008). It is possible that the "salesperson" game metaphor that we selected was not as appropriate for a conjoint survey as we initially thought. In a typical conjoint survey respondents are asked to select the best option with the implication that they are selecting the best option *for themselves*. In our game versions, respondents are asked to recommend the best options *for someone else* as a result of the design of the game. This changes the respondent experience in two ways. First, a respondent is being asked to pretend that the character on screen (who is "asking" them for help) is real enough to warrant giving their genuine opinion: the task might be more successful if respondents believed they were giving a recommendation to a real person instead. Second, asking the respondent to make a recommendation for someone else may actually make the task less personal and interesting: some respondents may prefer discussing and making choices for themselves rather than for others.

It is also possible that respondents were not given enough instruction and training regarding the game mechanics to make them feel comfortable with playing the game. Although the conjoint questions were administered in exactly the same manner in all four versions, the game setting may have been enough to confuse respondents, make them second guess their task, and reduce their confidence and comfort with the game version of the conjoint; some respondent comments seem to support this as a possible conclusion. Further research where some participants receive information about the game mechanic and others do not would be helpful in determining the true effect of additional instruction on engagement in a gamified survey task.

Another possible explanation for our findings could be in the framing of the task: Mekler et al. (2013) found that simply informing participants that their results would be used to improve scientific knowledge increased their motivation and willingness to complete more tasks. In the case of a survey, the very nature of the survey instrument makes it clear to the respondent that the data is being collected for a particular purpose. However, respondents may have been less clear about the purpose of the questions they were answering in the game version and how their participation in the game versions would be as helpful as in a typical survey.

Compared to the method tested by Mekler, gamifying surveys and occluding the survey instrument could actually prove to be problematic. If the obviousness of the data collection method proves to be a major motivating factor for survey participation, gamification in survey research may in fact be more of a detriment than a positive. Experiments where participants are explicitly informed of the value of the game to researchers could assist in determining whether this is indeed a factor. However, making the usefulness of the survey game explicit could also have a negative effect. Since part of the reason for using gamified data collection methods is to capture *naturalistic* data from participants, informing them prior to their participation in the survey game could have other negative effects. This is an important area for future research.

It is reassuring that there were no negative data quality implications of gamifying the conjoint experiment. This suggests that despite the relatively fantastic nature of the game terrain itself, individuals' preferences are fairly fixed: what they believe is better does not appear to strongly depend on the context in which they are asked. This remains true for a fairly realistic game scenario – a store front – as well as an extremely fantastical scenario – monsters on an alien planet.

Our use of gamification as a motivating component did not have a positive effect on any of the engagement measures that we used, suggesting that gamifying surveys is not a guaranteed way to increase respondent engagement in the survey process. The results suggest that in the context of survey research, the core task of providing opinions to a researcher is a very important motivator in itself. Therefore, meaningful framing must be a crucial consideration for researchers who wish to employ gamification as part of a data collection process and should guide future gamification strategies used in survey research.

References

Adamou, B. (n.d.a). *Using gamification in youth surveys*. Retrieved from http://rwconnect.esomar.org/2012/07/17/using-gamification-in-surveys-for-children/

Adamou, B. (n.d.b). Giving research the NOS effect. *Proceedings of 2012 Net Gain 6.0 MRIA Conference.*

Brown, J. (2003). Survey metrics ward off problems. *Marketing News, 17*, 17–20.

Cape, P. (2009). Questionnaire length, fatigue effects and response quality revisited. *SSI white paper.*

Carson, R., Louviere, J., Anderson, D., Arabie, P., Bunch, D., Hensher, D., et al. (1994). Experimental analysis of choice. *Marketing Letters, 5*(4), 351–368.

Center for Game Science at University of Washington: Foldit. Retrieved from http://www.fold.it

Chao, D. (2001). Doom as an interface for process management. *Proceedings of CHI 2001* (pp. 152–157).

Chrons, O., & Sundell, S. (2011) Digitalkoot. Making old archives accessible using crowdsourcing. *Proceedings of HCOMP.*

Chrzan, K., & Terry E. (1995). Partial profile choice experiments: A choice based approach for handling large numbers of attributes. *1995 Advanced Research Techniques Conference Proceedings*. Chicago, IL: American Marketing Association.

Deterding, S., Dixon, D., Nacke, L., O'Hara, K., & Sicart, M. (2011). Gamification: Using game design elements in non-gaming contexts. *CHI 2011 Ext. Abstracts*.

Deutskens, E., de Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters, 15*(1), 21–36.

Dignan, A. (2011). *Game frame: Using games as a strategy for success*. New York: Free Press.

Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research, 15*(2), 195–219.

Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., & Mandryk, R. L. (2011). Calibration games: Making calibration tasks enjoyable by adding motivating game elements. *Proceedings of UIST 2011* (pp. 403–412).

Greenbook Industry Trends Report (2013, Winter).

Guin, T. D., Baker, R., Mechling, J., & Ruylea, E. (2012). Myths and realities of respondent engagement in online surveys. *International Journal of Market Research, 54*(5), 613–633.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly, 45*, 549–559.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. *Proceedings of AAAI04 WS on Challenges in Game AI* (pp. 1–5).

Johnson, R., & Orme, B. (1996). How many questions should you ask in choice-based conjoint studies? *Sawtooth software research paper series*

Jung, J., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science, 56*(4), 724–742.

Korn, O. (2012). Industrial playgrounds: How gamification helps to enrich work for elderly or impaired persons in production. *Proceedings of EICS 2012* (pp. 313–316).

Lewis, C., Wardrip-Fruin, N., & Whitehead, J. (2012). Motivational game design patterns of 'ville games. *Proceedings of FDG 2012* (pp. 172–179).

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new scale type of fundamental measurement. *Journal of Mathematical Psychology, 1*(1), 1–27. doi:10.1016/0022-2496(64)90015-X.

Malinoff, B. (2010). Sexy questions, dangerous answers. *Proceedings of CASRO 2010 Technology Conference*.

Mason, A., Michalakidis, G., & Krause, P. (2012). Tiger nation: Empowering citizen scientists. *Proceedings of IEEE DEST*.

Mekler, D., Brühlmann, F., Opwis, K., & Tuch, N. (2013). Disassembling gamification: The effects of points and meaning on user motivation and performance. *CHI 2013*.

Nikkila, S., Byrne, D., Sundaram, H., Kelliher, A., & Linn, S. (2013). Taskville: Visualizing tasks and raising awareness in the workplace. *CHI 2013 Ext. Abstracts.*

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology, 61*(1), 50–69.

Orme, B. (2010). *Getting started with conjoint analysis: Strategies for product design and pricing research* (2nd ed.). Glendale, CA: Research Publishers LLC.

Porter, S., & Whitcomb, M. (2003). The impact of contact type on web survey response rates. *Public Opinion Quarterly, 67*(4), 579–588.

Puleston, J., & Sleep, D. (2011). The game experiments: Researching how gaming techniques can be used to improve the quality of feedback from online research. *Proceedings of ESOMAR Congress*.

Ray, N. M., & Tabor, S. W. (2003). Cyber surveys come of age. *Marketing Research, 15*, 32–37.

Reeves, B., & Read, J. (2009). *Total engagement: Using games and virtual worlds to change the way people work and businesses compete*. Boston: Harvard Business Press.

Research through gaming: Pimple crisis. Retrieved from http://www.youtube.com/watch?v=2s63gLuO-W0

Research through gaming: The Playspondent House. Retrieved from http://www.researchthrough-gaming.com/

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 347–363.

Sawtooth Software Inc. (1993). The CBC system for choice-based conjoint analysis—Version 8. *Sawtooth software technical paper series*

Shneiderman, B. (2004). Designing for fun: How can we design user interfaces to be more fun? *Interactions, 11*(5), 48–50.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment, 3*(3), 1–24.

Thomas, R., Bremer, J., Terhanian, G., & Couper, M. P. (2007). Truth in measurement: Comparing web-based interview techniques. *Proceedings of ESOMAR Congress*.

Von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM, 51*(8), 58–67.