

8-1-2010

# Crowdsourcing with Semantic Differentials: A Game to Investigate the Meaning of Form

Roland M. Müller

*University of Twente*, r.m.mueller@utwente.nl

Katja Thoring

*Anhalt University of Applied Sciences*, k.thoring@design.hs-anhalt.de

Ruben Oostinga

*University of Twente*, r.j.s.oostinga@student.utwente.nl

Follow this and additional works at: <http://aisel.aisnet.org/amcis2010>

## Recommended Citation

Müller, Roland M.; Thoring, Katja; and Oostinga, Ruben, "Crowdsourcing with Semantic Differentials: A Game to Investigate the Meaning of Form" (2010). *AMCIS 2010 Proceedings*. Paper 342.

<http://aisel.aisnet.org/amcis2010/342>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# Crowdsourcing with Semantic Differentials: A Game to Investigate the Meaning of Form

**Roland M. Müller**

University of Twente, The Netherlands  
r.m.mueller@utwente.nl

**Katja Thoring**

Anhalt University of Applied Sciences, Germany  
k.thoring@design.hs-anhalt.de

**Ruben Oostinga**

University of Twente, The Netherlands  
r.j.s.oostinga@student.utwente.nl

## ABSTRACT

This paper presents a tool to collect empirical data about the collaborative meaning of form. We developed an online crowdsourcing game, in which two users rate randomly assigned three-dimensional shapes. The more similar the ratings are, the more points both players get. This crowdsourcing method allows identifying what certain shapes mean to people. This paper is a contribution on two levels: First, the game presents a particular research method—an experimental survey using semantic differentials—which adds a motivational benefit for the participants: It is fun to play. Also, it involves a quality control mechanism through the pairing of two participants who rate the same image and therefore act as verification. Second, the semantic collection of forms might help designers to better control the connotative meanings embedded in their designs. This paper is focused on introducing the game; the analysis of the data will be covered in further research.

## Keywords

Crowdsourcing, Collaborative Intelligence, Research Methods, Semantic Differential, Semantics

## INTRODUCTION

The concept of semantics is a sub-category of semiotics, the theory of signs, which is originally a part of linguistics. While semantics describe the meaning of certain signs, syntactics describe the grammar and layout of the signs without touching the meaning, and pragmatics deal with the influence of context and usage. This paper is focused on the aspect of semantics. Not only linguistics, but also in different areas of design the possibility of communicating through signs (colors, materials, and forms) is important and helps to support a certain message to a user or observer. In our research we focus on the semantics of form—the analysis of color and materials is not covered in this article.

The idea to transfer the linguistic theory of signs to the area of product design is not new. In the nineteen-fifties and sixties the “Hochschule für Gestaltung Ulm” in Germany has developed a ‘semiotic approach’ to design, which mainly covered the syntactic aspect of semiotics—they investigated the fundamental formal aspects as means of designing (Bürdek, 1994, pp. 136). The attempt to establish a science-based education for design was revolutionary for that time, but critics also called it a ‘scientification’ of design.

In the nineteen-seventies, the concept of product semantics was wide-spread around Europe and the US. Its source can be found in Germany, specifically at the ‘Hochschule für Gestaltung Offenbach’ in Germany, where the ‘theory of product language’ was developed, which was based mainly on semiotics (Bürdek, 1994, pp. 12). Steffen (2000) summarizes the ‘Offenbach approach’ in her book “design as product language”.

The term ‘product semantics’ was coined by Krippendorff and Butter (1984). They are in-line with Wittgenstein’s (1953) definition of meaning as use, culminating in the axiom that “humans do not see and act on the physical qualities of things, but on what they mean to them” (Krippendorff (2006), pp. 47). According to Wittgenstein, a person knows the meaning of a statement, if they can react in an intelligent way; they can participate in the “language game”.

Semantics is a collaborative concept. As long as a person is the only person in the world who understands one specific word,

the meaning of this word is practically meaningless—communicating this word will not be possible. On the other hand, the concept of semantic meaning is also based on the intuitive associations of the observer, which makes it even more difficult to determine ‘one common meaning’. Both, intuitive associations and collaborative conventions might differ according to context and cultural background of the observer. Of course, there already exist some general understandings—you could also say ‘clichés’—e.g. that round shapes look ‘more feminine’, or that slanted shapes look ‘more dynamic and sporty’, but what is missing is an empirical analysis of such collaborative understandings of forms; as well as a structured database of such semantic shapes.

The goal of our work is to develop a research method to collect empirical data about a common meaning of forms. We want to discuss the following question: How can the design of such a research tool motivate a lot of people to participate in the survey, and how can we ensure high quality of the collected data? The aspired result is some kind of database of semantic forms.

Today the technological structures make it easy to access a lot of people through e.g. online surveys. The problem with such surveys, however, is to motivate people to participate, and to prevent cheating or carelessness of the participants. The concept of ‘crowdsourcing’ addresses these problems. In this paper we compare different applications that use crowdsourcing techniques to gather data from a crowd of people, such as Amazon’s Mechanical Turk (MTurk) (Amazon, 2005), Facestat (O’Connor and Biewald, 2009), Google Image Labeler (Google, 2006), Galaxy Zoo (Lintott, Schawinski, Slosar, Land, Bamford, Thomas, Raddick, Nichol, Szalay, Andreescu, Murray, and Vandenberg, 2008), and Peakaboom (von Ahn, Liu, and Blum, 2006). We compare the different approaches of motivating participants, as well as the mechanisms to prevent cheating, and then present our rating game, which combines different aspects of the aforementioned applications. The result of our research is a working prototype, which can be used to collect statistical data about what certain shapes mean to people.

To the best of our knowledge, such an empirical analysis of the meaning of forms has not been conducted, so far. Design practitioners would benefit from such a database, because they could use it to determine certain feelings and associations that the majority of people perceive when seeing a specific shape. This could then be incorporated into the design of objects, to enhance an intended message or mood. Researchers, on the other hand, could benefit from our developed prototype of the rating game, since this could also be adapted for other research questions and it is an example of the use of crowdsourcing in the innovation process.

## METHODOLOGY

This paper is a design science contribution. We follow the guidelines for design science research of Hevner, March, Park and Ram (2004). Design science is creating a viable artifact, either in the form of a construct, a model, a method, or an instantiation. The paper presents an instantiation of a crowdsourcing application. Also, the solution space for crowdsourcing applications will be analyzed and we will discuss why specific design choices were made.

## RELATED WORK

### Crowdsourcing

The word crowdsourcing was coined by Howe (2006). He describes a model for problem solving or production using a crowd of people. The problem or assignment is broadcasted to a group of people. Some of the people within the crowd submit a solution or participate in the assignment. In some cases this labor is well compensated, either monetarily, with prizes, or with recognition. In other cases the only rewards may be reputation or intellectual satisfaction.

Also the quality of a crowd can be remarkable good. However, for achieving this kind of “wisdom of the crowds” (Surowiecki, 2004) four requirements have to be fulfilled:

1. **Diversity.** The crowd includes people with different backgrounds and perspectives.
2. **Independence.** Each participant chooses their decision relatively independent of the others.
3. **Decentralization.** The decisions are based on local and specific knowledge of the individuals rather than of an all-knowing central planner.
4. **Aggregation.** There is some aggregation function that turns individual judgments into a collective decision.

If these requirements are met, a group can be remarkably intelligent; often smarter than the smartest people in them (Surowiecki, 2004).

## Collective Intelligence

Collective Intelligence describes the emergent capability of a complex system (e.g. group of people) for a kind of shared or group intelligence based on collaboration or competition of many individuals in this group (Kapetanios, 2008). Human-computer systems could facilitate this group intelligence in which computers collect large amounts of human-generated information and enable emergent knowledge through analyzing and inferring over these information (Kapetanios, 2008).

## Games

A game play is the formal interaction, in which the designed rules and structures players follow result in an experience (Salen and Zimmerman, 2004). Raybourn (2007) defines serious games as interactive digital technologies for training and education. This definition would exclude games with other purposes. Crowdsourcing games or “games with a purpose” are games people play and as a side effect of playing, they perform tasks computers are unable to do (von Ahn and Dabbish, 2008).

## SOLUTION SPACE OF CROWDSOURCING APPLICATIONS

This section discusses the solution space for crowdsourcing applications, including the pros and cons of each solution. A model of the solution space has been made using a morphological chart; see Table 1.

Element	Choice				
	Money		Altruism	Usefulness	Fun
Motivation					
Number of Players	1		2		>2
Concurrency	Yes			No	
Input	Image		Text		Other
Output	Text/Label	Rating	Binary	Multiple Choice	Pointing/tracing
Winning Condition	Output Agreement		Input Agreement	Inversion-problem	Other / not applicable
Points	Output amount	Agreement	Similarity	Close to average	None
User Accounts	Yes			No	
High Scores	Yes			No	
Timer	Yes			No	

**Table 1: Morphological Chart for the Solution Space of Crowdsourcing Applications**

## Motivation

The biggest problem with this project is how to motivate people to use crowdsourcing applications. We identified four motives for users to participate in crowdsourcing applications: money, altruism, usefulness, and fun.

The first option is to give the users monetary incentives or other extrinsic incentives (Leimeister, Huber, Bretschneider, and Krcmar, 2009) for completing the tasks. Amazon's Mechanical Turk (MTurk) is a marketplace that enables coordination of the use of human intelligence to perform micro-tasks in return of a small payment (Amazon, 2005). The advantage of paying users is that the application does not have to be well-known to generate a lot of data. The major disadvantage is that users might be more interested in completing as much sessions as possible, in order to make more money, than to actually give accurate information. A way to overcome this is to check workers by regularly including control-tasks, where the correct answers are already known (Snow, O'Connor, Jurafsky, and Ng, 2008), or by letting workers check the tasks of other users.

The second choice is to hope for the altruism of the users. The enjoyment of helping others is an intrinsic benefit that is rooted in the concept of Altruism. Altruism is defined as an “individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system” (Organ, 1997). Galaxy Zoo (Lintott, et al., 2008) is an example of a crowdsourcing application that relies on the altruism of the users to manually classify images of galaxies. Because of the motivational crowding-out effect, extrinsic incentives can crowd out intrinsic motives (Osterloh and Frey, 2000). Therefore it is problematic to combine monetary and altruistic motivations.

Third, individuals could use an application because it solves some individual problem of the user (usefulness). As a by-product of this usage the user creates shared data. An example are social bookmarking sites like Delicious.com, that allow users to save interesting bookmarks on the site and add tags (keywords) for finding them later again. As a by-product the user adds this user-generated content to the common repository of tagged sites. The sites do not require altruistic users. The users tag the bookmarks because it is for them individually useful.

Fourth, users could use an application because it is fun to do so. The ESP game (von Ahn and Dabbish, 2004) and Google Image Labeler (Google, 2006) are examples in which people label random images just for fun. However, hedonic information systems have different usage acceptance criteria than productivity-oriented systems (Van der Heijden, 2004). For explaining why a game is fun (Chen, 2007), the concept of *flow* from Csikszentmihályi (1990) is helpful. Flow is a focused mental state in which a person is immersed within an activity. For creating flow, the activity must be challenging but at a difficulty level that is in line with the skill level of the user. A too easy task will be boring. A too difficult task will be frustrating. The activity should have clear goals and give direct feedback. The person needs to feel a sense of control of the activity. While performing flow activities, participants lose the awareness of time and self (Agarwal and Karahanna, 2000).

### **Number of Players**

Most *games with a purpose*, like the ESP game (von Ahn and Dabbish, 2004), are two-player games. However, also single-player and multi-player games are possible. The advantage of making a single player game is that it is easier to implement. However, this does not take advantage of the fact that users like to play against other players. Therefore making a two player game could greatly increase the user's interest and participation. However, the question is what happens after one of the players leaves the game. In the ESP game (von Ahn and Dabbish, 2004), this problem has been solved by using a prerecorded session to imitate a user. Another possibility is to have more than 2 players in a game at one time. This has not been done before. There might be a small advantage in that it is more fun to play in groups. However, this greatly increases the complexity of the system.

### **Concurrency**

With two-player or multi-player games, the question arises if the users play concurrently together or not. Concurrent playing requires many players at the site so that the user does not have to wait for another player for a long time. Also the technical complexity of a concurrent game is higher.

### **Input**

A crowdsourcing application could offer different input types for the users. The ESP game (von Ahn and Dabbish, 2004) displays images to the player. The game Verbosity (von Ahn and Dabbish, 2008) shows text. Other input types like video or complex data structure are also possible.

### **Output**

Based on the input, the users of the crowdsourcing application will create an output. There are different output types possible. In the ESP game users have to enter text (labels) for images. In the Facestat game (O'Connor and Biewald, 2009), users provide text and ratings for faces. In TagATune (Law, von Ahn, Dannenberg and Crawford, 2007) both players create text descriptions about an image that are visible for both. However, the players don't know, if the other player has the same image. The players have to judge through the labels of the other if both see the same image (binary output). In the Squigl game (Law and von Ahn, 2009), both players see the same image and word. Both players can trace the object in the image that is described by the word (pointing/tracing output).

### **Winning Condition**

Von Ahn and Dabbish (2008) describe three prototypical game designs. In output-agreement games, players have to create the same output for an input. The goal of the ESP game (von Ahn and Dabbish, 2004) is to find the same keyword as the other player. In input-agreement games, the players should agree that they have the same input or not. In TagATune (Law et al., 2007), both players have to decide if both see the same image based on the descriptions of the other player. In inversion-problem games, one player is the 'describer' and the other player is the 'guesser'. The describer produces an output based on some input. This output is sent to the guesser who tries to produce the original input. An example is Peekaboom (von Ahn, et al., 2006).

### **Points**

For the way in which points are rewarded there exist multiple options. One option is to give points for completed answers. Another option is to give points if the players agree. Also the game could reward points based on the similarity of the output of both players or the similarity to the average output. The problem with only giving points for completing the questions is that it does not persuade the user to think about their answers and actually rewards filling in random answers. In some crowdsourcing applications no score is kept.

**User Accounts**

Either users have to register and create a user account or they are just identified by cookies or their IP-address.

**High Scores**

High score lists could be shown to motivate the users.

**Timer**

The game could either end after a specified time or after a specified event like the number of answered question.

**CLASSIFICATION OF KNOWN CROWDSOURCING APPLICATIONS**

According to the developed solution space, we classify the following crowdsourcing applications: Amazon Mechanical Turk, Galaxy Zoo, Google Image Labeler, and Peekaboom.

Amazon Mechanical Turk (Amazon, 2005) is a marketplace for micro tasks. Users participate because they get monetary compensation for completing tasks.

Element	Choice				
Motivation	Money		Altruism	Usefulness	Fun
Number of Players	1		2	>2	
Concurrency	Yes			No	
Input	Image		Text		Other
Output	Text/Label	Rating	Binary	Multiple Choice	Pointing/tracing
Winning Condition	Output Agreement		Input Agreement		Inversion-problem Other / not applicable
Points	Output amount	Agreement	Similarity	Close to average	None
User Accounts	Yes			No	
High Scores	Yes			No	
Timer	Yes			No	

**Table 2: Classification of Amazon Mechanical Turk**

Galaxy Zoo (Lintott, et al., 2008) allows users to classify images of galaxies according to a classification schema. Users participate mainly because of altruism.

Element	Choice				
Motivation	Money		Altruism	Usefulness	Fun
Number of Players	1		2	>2	
Concurrency	Yes			No	
Input	Image		Text		Other
Output	Text/Label	Rating	Binary	Multiple Choice	Pointing/tracing
Winning Condition	Output Agreement		Input Agreement		Inversion-problem Other / not applicable
Points	Output amount	Agreement	Similarity	Close to average	None
User Accounts	Yes			No	
High Scores	Yes			No	
Timer	Yes			No	

**Table 3: Classification of Galaxy Zoo**

Google Image Labeler (Google, 2006) is the licensed version of the ESP game developed by Luis von Ahn (von Ahn and Dabbish, 2004). This means that the two applications are essentially the same. Two users, or players, connect to the website and start the game. The players are paired up randomly from everyone playing. The goal is to get as much points as possible by giving the same label or tag to an image. At the start of the game both players are presented with the same image. The

players submit their labels without being able to see the other player's labels. When both players guess the same label points are rewarded and the players move on to the next image. The labels don't have to be guessed at the same time, therefore the server maintains lists of all the labels guessed by the players. Commonly guessed labels for images become taboo (ESP game). This makes sure new labels will be provided for the images. After completing the round of the game the players can see the labels of their partner. When the application isn't used very often it might be a problem that there are not enough users to form a pair. The ESP game solves this by making it able to play against previously recorded sessions when no one is available (von Ahn and Dabbish, 2004).

Element	Choice				
Motivation	Money		Altruism	Usefulness	Fun
Number of Players	1		2	>2	
Concurrency	Yes			No	
Input	Image		Text		Other
Output	Text/Label	Rating	Binary	Multiple Choice	Pointing/tracing
Winning Condition	Output Agreement		Input Agreement	Inversion-problem	Other / not applicable
Points	Output amount	Agreement	Similarity	Close to average	None
User Accounts	Yes			No	
High Scores	Yes			No	
Timer	Yes			No	

**Table 4: Classification of Google Image Labeler**

Facestat (O'Connor and Biewald, 2009) is a crowdsourcing application, which is used to determine how photos of people are perceived by the crowd. Users can upload their own photos and have the photo judged by the crowd. Also they can look at and judge photos of other users. The judgments are based on various pre-selected questions. Facestat asks multiple questions like: "How old do you think this person is?" and "Describe this person in one word." Rating the faces in Facestat is—besides the regular users—also done by workers of Amazon's Mechanical Turk (MTurk).

Element	Choice				
Motivation	Money		Altruism	Usefulness	Fun
Number of Players	1		2	>2	
Concurrency	Yes			No	
Input	Image		Text		Other
Output	Text/Label	Rating	Binary	Multiple Choice	Pointing/tracing
Winning Condition	Output Agreement		Input Agreement	Inversion-problem	Other / not applicable
Points	Output amount	Agreement	Similarity	Close to average	None
User Accounts	Yes			No	
High Scores	Yes			No	
Timer	Yes			No	

**Table 5: Classification of Facestat**

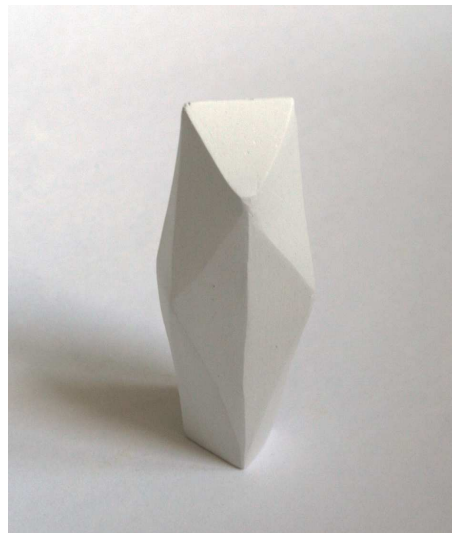
Peekaboom (von Ahn, et al., 2006) is a two-player game, with two roles: Peek and Boom. Boom gets an image and a word and must reveal parts of the image for Peek to guess the correct word. Peek can enter multiple guesses which are visible to Boom.

Element	Choice				
Motivation	Money	Altruism	Usefulness	Fun	
Number of Players	1	2	>2		
Concurrency	Yes			No	
Input	Image		Text		Other
Output	Text/Label	Rating	Binary	Multiple Choice	Pointing/tracing
Winning Condition	Output Agreement		Input Agreement		Inversion-problem Other / not applicable
Points	Output amount	Agreement	Similarity	Close to average	None
User Accounts	Yes			No	
High Scores	Yes			No	
Timer	Yes			No	

**Table 6: Classification of Peekaboom**

**THE RATING GAME PROTOTYPE**

The first step in this project was to develop around 80 different shapes to use as a foundation for the survey application. These shapes were developed in a class for product design fundamentals. Each student had to design a plaster shape according to a predetermined semantic phrase (such as “elegant”, “cheap”, or “aggressive”.) The starting point for each shape was the same cuboid with fixed dimensions, in order to keep the comparability of the shapes. This cuboid should then be transformed according to the semantic phrase. The decision, which shape matches best to the required phrase, was left to the student. This resulted in quite intuitive and arbitrary shapes, but it served the purpose to develop lots of different, yet comparable shapes. An example of such a shape can be found in Figure 1.



**Figure 1: Example of a Shape**

The second step was the design of the application to use as a crowdsourcing application. The application should meet the following requirements: The developed shapes should be stored in a database. People should be able to access this database online in order to rate and comment the shapes. The goal of this application is to receive many ratings and comments for the shapes from a crowd of people, in order to detect either obvious impressions as well as connotative or ambiguous meanings of the shapes.

**Our Solution: The Rating Game**

Our application is a combination of Google Image Labeler (Google, 2006) or the ESP game (von Ahn and Dabbish, 2004), and Facestat (O’Connor and Biewald, 2009). Because the user’s task is relatively easy, making it fun to play is the best way to motivate people to participate for free. The goal is to make the application as attractive to users as possible. This is kept in mind while making every design choice. Paraphrasing Wittgenstein, we define the meaning of a form as a game people play.

If only labels are used as an output, this would just describe the shapes, which is less useful and harder to analyze to



determine the actual meaning of the shapes. The rating game will generate data that is easier to analyze. Since it involves the rating of shapes instead of people, it is unlikely that people find it interesting enough to generate enough data without a game design. Because of these problems both parts of the ESP game and Facestat are combined into a single game. There are multiple ways to combine both parts. A possibility is to ask the user to rate the image at the end of each round of the rating game. Points can be awarded when users give similar ratings. Another possibility is to have just one round per image where the user has to answer multiple types of questions. A question which asks the user to label the image could be included. The second approach has been chosen because it generates more rating data that is easier to analyze statistically.

As already mentioned returning users are especially useful. To increase the chance a user will return there must be something to achieve by returning. This is done just as in Google Image Labeler and the ESP game by keeping the high scores. We identify a player by a cookie and the IP-address and assign them a unique number. This number is shown while playing a game.

Another choice that has to be made is whether or not to use a synchronized game with two players, or just make a single player game. Research showed (von Ahlen, 2004) that users like to play against other players. Therefore it has been chosen to make a two player game. We use prerecorded sessions to imitate a user if no matching partner is available, similar as in the ESP game.

For this application, questions concerning the 3d-shapes will have to be established. These can be simple multiple-choice questions, but can also be ratings on a scale or even open questions. Multiple-choice questions and ratings are easier to analyze than open questions because the user's answers are limited, however, open question can give more creative and original results. In order to give an idea of how the questions will look, here is an example for each type of question:

Multiple-choice question: "Which of the following words describes the shape most accurate?" The answers can be picked between different labels.

Rating question: "How smooth is this shape?" The rating questions are actually a type of semantic differential (Osgood, Suci and Tannenbaum, 1957). In semantic differentials a respondent is asked to choose where his or her position lies, on a scale between two bipolar adjectives (for example: "Adequate-Inadequate", "Good-Evil" or "Valuable-Worthless") (Osgood et al. 1957). In the rating game, a semantic differential of 5 points is used, with the values -2, -1, 0, +1, and +2, where 0 means 'neutral.'

Open question: "Describe this shape in one word." This question is also used in Facestat to describe a photo.

The score is calculated using the similarity of the answers of the players. Rating questions can be answered with a value between -2 and +2. If the answers are the same, both players get 100 points. If there is a difference of 1, they get 50 points, and 25 points if there is a difference of 2. Multiple choice answers only get 100 points if both answers are the same. For open questions both players get 200 points, if they enter exactly the same phrase.

Although implementing a timer enhances the fun of the game, we decided not to use a timer, but to limit one round by the amount of shapes being rated. Thereby people can take more time thinking about their answers, which will probably result in higher quality of the results. In one round, 10 shapes have to be rated. After finishing one round, both players can see the results of their partner.

Table 7 shows an overview of the classification of the rating game, and Figure 2 shows a screenshot of the application.

Element	Choice				
	Money	Altruism	Usefulness	Fun	
Number of Players	1		2		>2
Concurrency	Yes			No	
Input	Image		Text		Other
Output	Text/Label	Rating	Binary	Multiple Choice	Pointing/tracing
Winning Condition	Output Agreement		Input Agreement		Other
Points	Output amount	Agreement	Similarity	Close to average	None
User Accounts	Yes			No	
High Scores	Yes			No	
Timer	Yes			No	

Table 7: Classification of the Rating Game Prototype

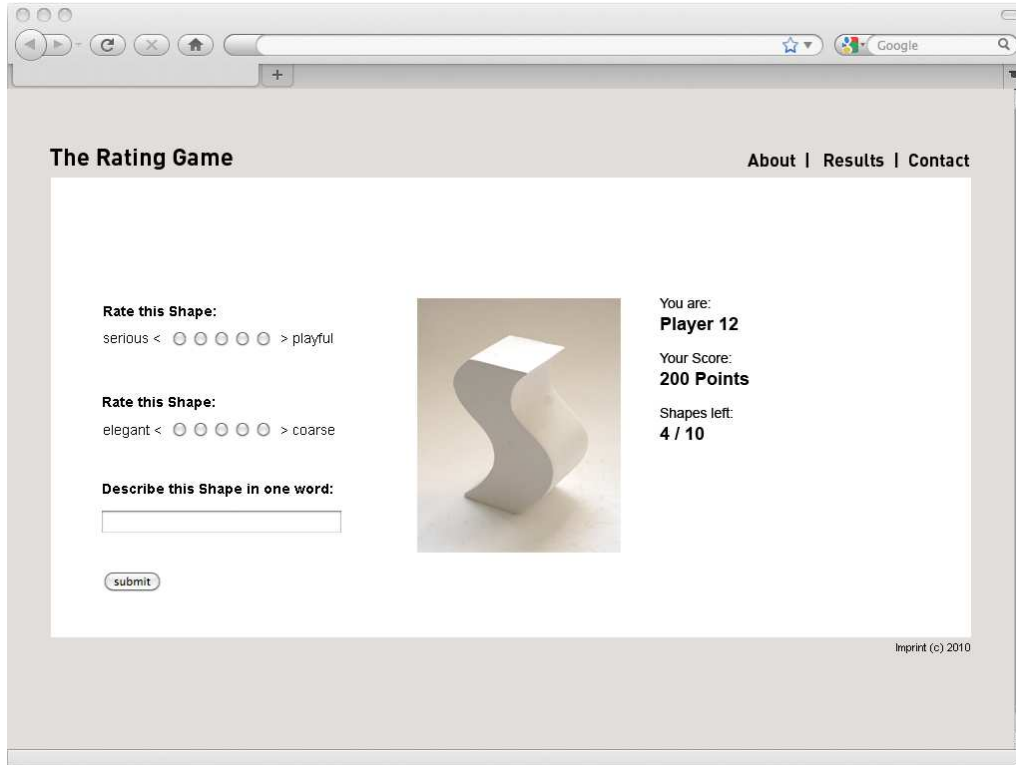
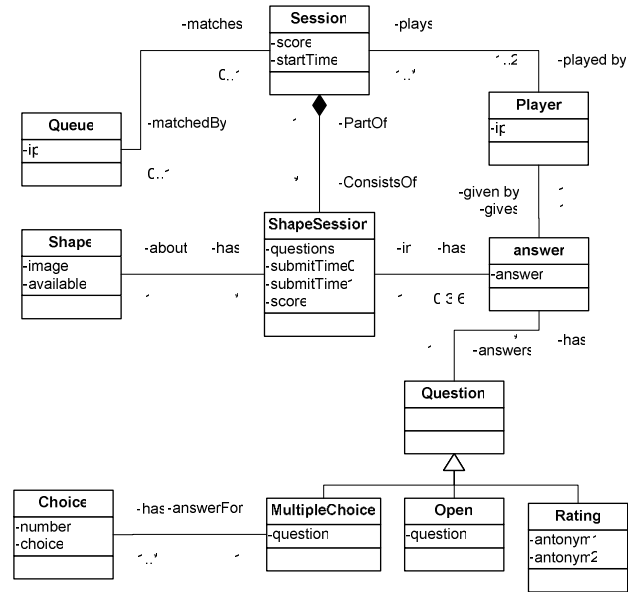


Figure 2: Screenshot of the application

### Technical Implementation

To make the application widely available and easy to start it is web based. For the interface there are multiple options available. One option is to use browser plugins such as Sun's Java or Macromedia's Flash. Another option is to use html pages as the interface. Because plugins have to be installed it has been chosen to make the application available without them. To make the application more responsive, JavaScript and XML are used in an AJAX approach (Paulson, 2005). This makes it possible to run the game without having to refresh the page. Communication with the server after the loading of the initial page is done by the JavaScript.

At the start of the implementation a model of the database has been made (see Figure 3). For each game an entry in a session instance is created. Each session consists of multiple ShapeSessions which are made for each shape. Two players play a session of the game. A ShapeSession is linked to a Shape. A player has to answer 3 questions for each shape. There are two types of questions—open questions and rating. The score is stored for each Session and ShapeSession.



**Figure 3: Class Diagram of the Rating Game**

At the beginning of each game, two clients have to be matched. This involves communication between PHP sessions, and therefore the first step is to put the client's IP-address and the PHP session-id in the table queue. In order to ensure mutual exclusivity a lock on the table queue is acquired first. After the client is added to the queue, the queue is checked for another client with a different IP-address. If a client is available, the IP-addresses of both clients are added to the players' tables. From now on clients are called players. A session is made with a reference to both player entries. If no client is available the lock is given up and the thread waits for half a second and then retries to match. This retrying is done for approximately 5 seconds.

When there is no partner available or a partner leaves during a game, a player has to be simulated using pre-recorded sessions. Of course it is possible that the question has not been answered for this shape yet. When that's the case a different shape is used. If that is also not available a random answer will be generated. However, the answers of the pre-recorded sessions are not stored, while the answers of the player are. This way user data can still be collected when there is only one player at a time.

## CONCLUSION

In this paper we present a working prototype of an online rating game, which can be used to gather empirical data about the perception of users when seeing certain shapes. The article focuses on the design and implementation of the game, which combines elements from different crowdsourcing applications. Our application can be viewed as a novel research method with the goal to collect statistical data from a group of people, while offering a) a motivational game to stimulate participation, and b) implying some kind of quality control mechanism to prevent cheating of the participants. As a result, this kind of method for gathering statistical data has several advantages compared to standard questionnaires and surveys: We will probably collect more data in a shorter period of time, since people enjoy participating. Moreover, we are able to collect multi-dimensional data through the use of a 5-point rating Likert scale. From the analyzed crowdsourcing applications, only Facestat offers this kind of multidimensional rating system. To the best of our knowledge, the use of crowdsourcing games to enhance online surveys with semantic differentials has not been developed, so far.

As a result, the application can be used to support the design process by gathering subjective meanings of form. We believe, that this paper contributes to the design community, who might be interested in collecting data about the semantics of forms, as well as to the research community in general, who could adapt the concept of this prototype for other research questions.

The first step of our future work will be to collect data about the semantic shapes. For this purpose, the prototype will be promoted within the design community. After a critical amount of data has been gathered, we will implement an analysis tool that automatically generates diagrams from the gathered data. We consider also the implementation of a different layout of the game—with only one player—in order to compare the participation rate for both research games.

## ACKNOWLEDGMENTS

We thank the students of the Anhalt University of Applied Sciences, Department of Design in Dessau, Germany, for contributing the plaster shapes, which were developed during the class ‘two- and three-dimensional design basics’ between 2008 and 2010.

## REFERENCES

1. Agarwal, R. and Karahanna, E. (2000) Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage, *MIS Quarterly*, 24, 4, 665-694.
2. Amazon (2005). Amazon Mechanical Turk, URL (accessed 20.2.2010): <https://www.mturk.com/mturk/welcome>, first online November 2005
3. Bürdek, B. E. (1994) Design, Geschichte, Theorie und Praxis der Produktgestaltung, DuMont, Köln, 2. Edition
4. Chen, J. (2007) Flow in Games (and everything else), *Communications of the ACM*, 50, 4, 4.
5. Csikszentmihályi, M. (1990) Flow: The Psychology of Optimal Experience, Harper and Row, New York.
6. Google (2006) Google image labeler. URL (accessed 20.2.2010): <http://images.google.com/imagelabeler/>, (accessed 20.2.2010), first online August 2006
7. Hevner, A.R., S.T. March, J. Park, and S. Ram (2004) Design science in information systems research, *MIS Quarterly*, 28, 1, 75-105.
8. Howe, J. (2006) The rise of Crowdsourcing, *Wired*, 14, 6 URL (accessed 20.2.2010): <http://www.wired.com/wired/archive/14.06/crowds.htm>
9. Kapetanos, E. (2008) Quo Vadis computer science: From Turing to personal computer, personal content and collective intelligence, *Data & Knowledge Engineering*, 67, 2, 286-292.
10. Krippendorff (2006) The Semantic Turn; A New Foundation for Design. Taylor&Francis, Boca Raton.
11. Krippendorff, K. and Butter, R. (1984). Exploring the Symbolic Qualities of Form. *Innovatios*, 3, 2, 4-9.
12. Law, E. and von Ahn, L. (2009) Input-agreement: a new mechanism for collecting data using human computation games, in *Proceedings of the 27th international conference on Human factors in computing systems (CHI'09)*, Boston, MA, USA, 1197-1206
13. Law, E.L.M., von Ahn, L., Dannenberg, R.B., and Crawford, M. (2007) TagATune: A game for music and sound annotation, in *Proceedings of the Eighth International Conference on Music Information Retrieval*, Vienna, Austria, Sept. 23–30, Austrian Computer Society, 361–364.
14. Leimeister, J. M., Huber, M., Bretschneider, U. and Krcmar, H. (2009) Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition, *Journal of Management Information Systems*, 26, 1, 197-224.
15. Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R.; Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. (2008). Galaxy Zoo : Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, *Monthly Notices of the Royal Astronomical Society*, 389, 3, 1179-118
16. O'Connor, B. and Biewald, L. (2009) Superficial Data Analysis: Exploring Millions of Social Stereotypes, In: Toby Segaran and Jeff Hammerbacher, *Beautiful Data*, O'Reilly, Sebastopol, CA, 279-301
17. Organ, D.W. (1997) Organizational citizenship behavior: It's construct clean-up time, *Human Performance*, 10, 2, 85-97.
18. Osgood, C. E., Suci, G. J. and Tannenbaum, P. H. (1957). The measurement of meaning. University of Illinois Press, Urbana, USA.
19. Osterloh, M., and Frey, B.S. (2000) Motivation, Knowledge Transfer, and Organizational Forms, *Organization Science*, 11, 5, 538-550.
20. Paulson, L. (2005) Building rich web applications with AJAX. *Computer*, 38, 10, 14–17.
21. Raybourn, E. (2007) Applying simulation experience design methods to creating serious game-based adaptive training systems, *Interacting with Computers*, 19 ,2, 206-214.
22. Salen, K. and Zimmerman, E. (2004) Rules of Play, The MIT Press, Cambridge.
23. Snow, R., B. O'Connor, D. Jurafsky, and A. Y Ng. (2008) Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language*

*Processing*, 254–263.

24. Steffen, D. (2000) Design als Produktsprache – Der „Offenbacher Ansatz“ in Theorie und Praxis. Verlag form, Frankfurt/Main
25. Surowiecki, J. (2004) The wisdom of the crowds: why the many are smarter than the few, Abacus, London.
26. Van der Heijden, H. (2004) User Acceptance of Hedonic Information Systems, *MIS Quarterly*, 28, 4, 695-704.
27. von Ahn, L. and Dabbish, L. (2004) Labeling Images with a Computer Game. *Proceedings of the 24th international conference on Human factors in computing systems (CHI'04)*, 319-326.
28. von Ahn, L., Liu, R., and Blum, M. (2006) Peekaboom: A Game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montreal, Apr. 22–27, ACM Press, New York, 55–64.
29. Wittgenstein, L. (1953) *Philosophical Investigations*, Blackwell, Oxford.