A Study on User Preferential Choices about Rating Scales

Federica Cena, University of Torino, Torino, Italy
Fabiana Vernero, University of Torino, Torino, Italy

ABSTRACT
Websites usually offer the same rating scale for all users and all tasks, but users can have very different preferences. In this paper, the authors study rating scales from the point of view of preferential choices, investigating i) if user preferences for rating scales depend on the object to evaluate, and ii) if user preferences change after they have rated an object repeatedly, gaining a high level of experience with the evaluated object. The authors first defined a model of rating scales, identifying generic classes based on features like granularity and visual metaphor. Then, the authors had users choose between three scales, one for each class, for rating two objects with opposite features, first in a condition where users had a low level of experience, and then in a condition where their level of experience was high. Results showed that user choices depend on the evaluated objects, while their level of experience influences their overall preferences, but not their choices when they have to rate a specific object. The authors conclude with some insights and guidelines for designers of interactive systems.

Keywords: Interactive Systems, Rating Scales, User Choices, User Interface, User Studies

1. INTRODUCTION
Rating scales are visual widgets that are characterized by specific features (e.g. granularity, numbering, presence of a neutral position, etc.) which allow users to provide quantitative input to a system. Each system uses its own different rating scale, with different features such as granularity and visual presentation. Examples of rating scales are stars in Amazon (Amazon), Anobii (Anobi) and Barnes & Noble (Barnes & Noble), thumbs in Facebook (Facebook) and YouTube (Youtube), circles in Tripadvisor (Tripadvisor), squares in LateRooms (LateRooms), bare numbers in Criticker (Criticker). In recommender systems (Adomavicius and Tuzhilin, 2005), users rate items to receive personalized suggestions about other items (similar to the previous ones or liked by similar users).

Understanding how users perceive rating scales, and why they might prefer one to another, is very important for interface designers in order to create more effective and pleasant web sites. This problem can be framed in terms of preferential choices (Jameson, 2012), i.e., when two or more options are available, none of which can be defined as “incorrect”, but one of which can be preferred for some reason (e.g., tasks, user skills, usage context, habits, etc. (Jameson et al., 2011)).

Rating scales are widely studied in the literature, especially in survey design (Garland, 1991; Colman et al., 1997; Amoo & Friedman, 2001; Dawes, 2008) and Human-Computer

DOI: 10.4018/ijthi.2015010103
Interaction (HCI) (Cosey et al., 2003; van Barneveld & van Setten, 2004; Nobarany et al., 2012; Herlocker et al., 2004), but not in terms of preferential choices, i.e. not focusing on users’ decision making process. In the survey design field, scales are compared according to their psychometric properties, i.e., their ability to detect “real” user opinions. In the HCI field, scales are mainly studied from a usability point of view. In a sense, the question that all previous works aimed to answer was: “What is the scale that measures the best?”. We instead aim to answer a different question: “What is the scale that users would choose?”, assuming that there may be other criteria beside precision and usability. With this paper, we investigate how users choose rating scales when they are offered this opportunity. In particular, we study whether users prefer different scales for evaluating different objects, and whether user’s choices change after they have rated a certain object repeatedly, gaining a higher level of experience with the evaluated object.

To answer these questions, we first analysed existing rating scales in order to define an abstract model, which allowed us to identify three generic “classes” of rating scales. Then, we carried out a user study to investigate user choices with respect to three scales chosen as representatives of each class. According to our findings, user choices are influenced by the evaluated objects and overall preferences for rating scales can change after their repeated use. Based on our results, we formulate some guidelines for systems designers.

The main contributions of this paper are:

- A general model of rating scales;
- The results of a user study on preferential choices about rating scales in a website;
- New insights which can help system designers to include the most appropriate rating scales.

The paper is structured as follows. Section 2 presents a preliminary study to devise a model for describing rating scales. Section 3 clarifies how we chose the scales, the objects to evaluate and the use case for our user study, while Sections 4 and 5 describe the study and its results. Section 6 provides the theoretical background for our research and analyzes related works, offering a systematic comparison with the results obtained by the most relevant ones. Section 7 concludes the paper with some guidelines derived from our results and with some possible directions for future work.

2. A MODEL OF RATING SCALES

Aiming at defining a general model of rating scales, we examined thirteen rating scales described by van Barneveld & van Setten (2004), Gena et al. (2011) and Nobarany et al. (2012): 3-, 5- and 10-point stars, bare numbers, smileys, sliders ranging -10/+10, -1/+1 and 0/10, likert-like scales ranging -10/+10 and 1/5 and 1-, 2-, and 3-point thumbs (see Figure 1). First, we identified a list of features which could be used to describe them, based on the literature and our insights, and we organized it in a model (Section 2.1). Then, observing how these features tend to combine in the examined scales, we identified three general classes of rating scales (Section 2.2).

2.1. Analysis of Rating Scales

van Barneveld & van Setten (2004), Gena et al. (2011) and Nobarany et al. (2012) described rating scales through different features, as reported in Table 1.

If we consider the thirteen rating scales we selected, however, some of these features do not seem useful to distinguish among them: continuity (all scales allow to input only discrete values), use of colour (colour is not used in a semantic way, i.e., to convey meaning), measurement scale (all scales allow only absolute input) and recall support (no information is provided about previously recorder opinions). Conversely, in our opinion, features such as “step”, “icon”, “point mutability”, “positive/negative scale” and “diffusion” should be taken into account, and a feature such as “numbering”
The Practical Accomplishment of Location-Based Game-Play: Design and Analysis of Mobile Collaborative Gaming
[www.igi-global.com/article/the-practical-accomplishment-of-location-based-game-play/116484?camid=4v1a](www.igi-global.com/article/the-practical-accomplishment-of-location-based-game-play/116484?camid=4v1a)

An Empirical Exploration of Mental Representations in the Individual Saving Decision Process
[www.igi-global.com/article/an-empirical-exploration-of-mental-representations-in-the-individual-saving-decision-process/116789?camid=4v1a](www.igi-global.com/article/an-empirical-exploration-of-mental-representations-in-the-individual-saving-decision-process/116789?camid=4v1a)