HALO EFFECTS IN CONSUMER SURVEYS

Master Thesis
Erasmus University Rotterdam
Erasmus School of Economics

Author: Lidia V. Lüttin

Student Number: 352879

E-mail address: 352879ll@eur.nl

Supervisor: Prof. Dr. Martijn G. de Jong

Study Program: Business Economics

Specialization: Marketing

Date: 02.08.2012

Abstract

The halo effect, a systematic response error, is often neglected when marketing constructs are measured with multi-item scales. However, it can distort the results obtained by consumer surveys and result in wrong conclusions and strategic decisions. This thesis provides an extensive compilation of the present knowledge of the halo effect, such as definitions, methods to measure and detect halo effects, and statistical, as well as design-oriented approaches to reduce halo effects in surveys. Additionally, the findings of a conducted experiment to examine the effect of five design-oriented approaches on halo effect are reported. The results indicate a halo-reducing effect for survey length, intermixing scale items, screen-by-screen design, and the combination of both.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AB	Acquiescence Bias
ANOVA	Analysis of Variance
BARS	Behaviorally Anchored Rating Scales
CP	Components of Involvement
ERS	Extreme Response Style
GE	General Evaluation
I.I.D.	Independent and Identically Distributed
LB	Leniency Bias
M	Mean
MA	Multi-Attribute
MDA	Multiple Discriminant Analysis
MI	Multi-Item
MR	Midpoint Responding
OLS	Ordinary Least Squares
SD	Standard Deviation
SDB	Social Desirability Bias
SERVQUAL	Service Quality
TSLS	Two-Stage Least Squares
UI	Use Innovativeness
ZC	Zone Counting

1. Introduction

Consumer surveys play an important role in marketing. They can help businesses track customer satisfaction, measure brand equity and brand awareness, improve customer retention, pinpoint areas for improvement, and many more. Surveys are popular because they are quick, easy and cheap to administer and can help businesses increase their long-term profitability. To better manage critical factors for success such as customer satisfaction, brand image and brand equity, companies invest millions of dollars (Wirtz 2001).

However, in many cases the quality of the results of consumer surveys suffers from response effects, which can result in discrepancies between the obtained measurements and the respondent's true value assessment. In practice, researchers often ignore response bias, although it has been shown that such effects exist, and that they can affect the validity of research findings (Baumgartner & Steenkamp 2006). One important effect is the so-called "halo-error", first labeled by Thorndike (1920) in the psychological context of personal evaluations. His study revealed that supervisors were unable to evaluate subordinates independently on different characteristics which in consequence led to high correlations of their ratings on different characteristics with their overall impression (Thorndike 1920).

In the context of marketing and consumer marketing research halo effects play a role especially when multi-item scales are used to measure beliefs and attitudes. These data are used in consumer surveys to measure several personality-, behavior-related-, and attitudinal-marketing constructs, such as preferences, satisfaction, awareness and involvement (Parasuraman, Grewal & Krishnan 2007; Leuthesser et al.1995). In this regard the halo effect can be described as an effect influencing the respondent's rating to items of a scale by prior items of this scale. The particular response strategy adapted by the respondent depends on the type of multi-item measure that is applied. It can be distinguished between general multi-item scales, and multi-attribute-attitude scales.

If no attention is paid to this effect when analyzing the outcomes of consumer surveys, it can have negative consequences for the marketing strategy and finally for the long term success of a company.

1.1. Problem Statement and Research Questions

In marketing research halo effects occur in the application of two types of multi-item measures in consumer surveys to measure beliefs and attitudes: Multi-attribute scales, where the measurement of a construct (e.g. satisfaction) is directed to several objects such as brands, and general multi-item scales where often only a construct is measured without concentration on a specific object. A major problem exists in the underlying assumption of these measures, that several items together explain each respondent's score on the construct and in this manner, measure beliefs directly.

To assure the interpretability of multi-item scales in marketing research, the scales should deliver internally valid and reliable results (King & Bruner 2000). However, the halo effect causes confusion whether the obtained results represent true content and thus unbiased beliefs of the respondents, or are simply another measure of prior rated items (Beckwith & Lehmann 1975). Whereas in general multi-item scales the ratings are influenced by preceding items, for multi-attribute scales the overall attitude towards a product or brand shapes the ratings.

In the context of marketing construct measurement, such as satisfaction, where summated item ratings serve the identification of the drivers of satisfaction, results can be distorted, if respondents are unable to assess the items individually from their memory and experience (Bueschken et al. 2010; Wirtz 2003).

In practice, this may result in higher spurious correlations between items and consistently higher/lower ratings on these items and therefore less variability in the data, leading to inflated reliability and lower predictive validity. Unless it is known whether ratings are influenced by the halo effect, the interpretation of data obtained by multi-item measures may be ambiguous (Wirtz & Bateson 1995; Bradlow & Fitzsimons 2001).

The halo effect is assumed to affect the quality of ratings negatively and to lower the usefulness of the results for several purposes (Murphy et al. 1993). Data distorted by halo effects limit the interpretability of marketing metrics, such as customer satisfaction and is hardly informative about the individual drivers of overall satisfaction (Bueschken et al. 2010; Wirtz 2003). Furthermore, Wirtz and Bateson (1995) and Wirtz (2000) demonstrate that halo contaminated data can obscure the identification of product strengths and weaknesses and make attribute-based comparisons among brands unreliable (Wirtz 2003). Finally, this may result in wrong strategic decisions such as investments for the improvement of weaknesses, or misleading conclusions about competitive positioning (Wirtz 2003; Leuthesser et al. 1995).

The objective of this master thesis is to increase the understanding of the effect of halo regarding the results of consumer surveys in the marketing field and to work out implications for consumer marketing research. The major research question is:

How do halo effects affect consumer surveys?

In order to give a comprehensive answer to this research question, the following sub-questions will be analyzed in detail:

What is the halo effect in consumer surveys?

Which methods can be used to detect halo effects?

How can halo effects be reduced post-hoc?

How can halo effects be reduced ex-ante?

1.2. Academic and Managerial Relevance

This research aims to increase our understanding of the halo effect in consumer surveys. Until now systematic response effects in surveys and evaluations in general, such as social desirability (e.g. Krosnick 1999; Mick 1996), leniency (e.g. Podsakoff et al. 2003; Schriesheim et al. 1979), acquiescence (e.g. Baumgartner & Steenkamp 2001; Greenleaf 1992), positive and negative affectivity (e.g. Bagozzi 1994; Baumgartner & Steenkamp 2001), extreme response style (e.g. Greenleaf 1992; O'Donovan 1965), as well as the effects of consistency and illusory correlations, which can be seen as similar to halo effects (e.g. McGuire 1966; Salancik & Pfeffer 1977; Berman & Kenny 1976) have been examined in a wide range in literature (Baumgartner & Steenkamp 2001; Podsakoff et al. 2003; Baumgartner & Steenkamp 2006).

Until now the halo effect has been studied broadly in the context of evaluations from products and brands, retail stores, cities (e.g. Wilkie et al. 1973; Beckwith & Kubilius 1978; Wu & Petroshius 1987), people, such as performance appraisal, personnel recruitment and interpersonal judgment (e.g. Fisicaro & Lance 1990; Murphy et al. 1993), as well as in pre-choice evaluations (e.g. Beckwith et al. 1978). In the field of consumer surveys, and in this regard in the application of multi-item scales, halo effects have been investigated in satisfaction & image measurement, brand evaluations and preferences (e.g. Wirtz 2000; Leuthesser et al.1995).

Table 1 illustrates an overview of studies, examining the halo effect in a marketing context. Several of the presented studies are reviewed in the literature review. Furthermore, halo research in the field of psychology and organizational behavior provides potentially useful insights which still have to be examined in a marketing context. For example, there is not much research yet on how the interpretability of data obtained by multi-item measures is improved by design-oriented halo reducing methods such as alternative design of rating scales.

Furthermore, it can be observed, that many studies in academic journals rely on data which are affected by the halo effect. The results, based on questionable data may possibly alter the outcomes of some studies (Rosenzweig 2007).

This master thesis builds on existing literature and studies on how the halo effect and similar response effects influence ratings to multi-item scales in consumer surveys. Relevant existing theories of the psychological and behavioral economics research are applied to this context, and indistinct findings of the related marketing literature are further investigated.

For managers seeking to make decisions based on data obtained from consumer surveys, the halo effect is a potential source of risk in regard to faulty decisions (Leuthesser et al.1995). Therefore, in the managerial context, this master thesis can help marketers and market researchers to improve their consumer surveys by extending their knowledge about the effect of halos. Finally, this will support them in improving the quality and reliability of conducted consumer research, and will therefore help marketers in making more accurate and evidence-based decisions.

1.3. Structure of the Thesis

The first chapter illustrated the research objectives and research questions, as well as their academic and managerial relevance. The second chapter is dedicated to the theory, on which this master thesis builds upon, therefore existing knowledge on the halo effect and relevant concepts and models existing in the literature will be reviewed. Furthermore, in this chapter the conceptual model and the main hypotheses are developed. In the third chapter the research methology and design, applied to test the formulated hypotheses, are described. In chapter four, the data analysis is conducted and hypotheses are tested. Finally, results in regard to the research questions are discussed, and limitations, as well as future research possibilities are pointed out in chapter five.

Table 1 Overview of Studies of Halo Effect in Marketing Research

Authors	Object	Dependent Variable	Independent Variables/Technique	Approach	Findings
Wilkie & Mc Cann (1972)	Toothpaste	Preferences	Instructions Brand intermixing	Design	Halo is reduced by brand intermixing and providing warm-up instructions
Wilkie; McCann & Reib-stein (1973)	Toothpaste	Brand Performance	Instructions & brand-intermixing	Design	Halo is reduced for brand-intermixing
Beckwith & Lehmann (1975)	TV-Shows	Brand Preference	Develop a simultaneous equation model to estimate halo	Statistical	Find strong halo effects & stronger halos for less im-portant, vague, ambiguous attributes
Beckwith & Kubilius (1977)	Retail Store	Image	Develop regression model for estimating true locations for judged objects corrected for halo-like effects (e.g. familiarity)	Statistical	Find halo effects
Moore & James (1978)	Automobiles	Product Perfor-mance	Apply regression model as in Beckwith & Lehmann 1975	Statistical	Find halo to be unimportant in multi-attribute model
James & Carter (1978)	Cities	Preferences	IV: object preference, familiarity, attributes with physical corre-lates	Statistical	Halo is less for objects with high preference, attributes with clearly defined physical correlates
Bemmaor & Huber (1978)	Cities	Preferences	Test Beckwith & Lehmann's (1975) single equation model for specification errors (single vs. simultaneous equation)	Statistical	Find that the specification of the model affects halo estimates
Holbrook & Huber (1979)	Piano Recordings	Preferences	Combine regression, factor- and discriminant analysis to correct for halo	Statistical	Remove halo effects
Holbrook (1983)	Piano Recordings	Preferences	Develop a structural model of halo to assess perceptual distor-tion due to affective overtones	Statistical	Only find weak halo effects
Dillon; Mulani & Frederick (1984)	Jazz Recordings	Preferences	Apply double centring-technique to partialling- out the halo	Statistical	Remove halo effects
Wu & Petroshius (1987)	Retail Store	Image	Gender, brand-intermixing, familiarity, attribute importance	Design	Halo is reduced for familiarity, attribute importance & females
Wirtz & Bateson (1995)	Online Banking	Satisfaction	Induce halo by manipulating an attribute in an experiment	Design	Find halo effects & show that halo effects can lead to wrong conclusions in satisfaction measures
Leuthesser; Kohli & Harich (1996)	Household Products	Product Perfor-mance Brand Equity	Apply double centring-technique to partialling- out the halo	Statistical	Find the level of halo varying over different brands
Wirtz (2000)	Travel Agency	Satisfaction	Attribute importance Halo additive function of number of halo-causing attributes	Design	Find halo effects Halo is stronger for important attributes
Wirtz (2001)	Service of front-line staff	Satisfaction	Number of attributes; Relative rating scales Time delay between consumption and rating	Design	Halo is reduced by relative scales and direct rating after consumption
Wirtz (2003)	Fast Food Restaurant	Satisfaction	Number of attributes Involvement Purpose of evaluation	Design	Halo is reduced for developmental purpose, more attributes & for high involvement
Gilbride; Yang & Allenby (2005)	Digital Cameras	Purchase Intention/ Brand Performance	Develop a Bayenesian Mixture Model to model simultaneity and brand halos	Statistical	Find halo effects
Van Doorn (2008)	B2B Service	Satisfaction	Develops a two-level asymmetric model to estimate dynamic effects on both the level of attribute and overall evaluation	Statistical	Find only weak halo effects
Büschken; Otter & Allenby (2011)	Hospitals & student evaluation of instructors	Satisfaction	Develop a Bayenesian Mixture Model that separates-out halo	Statistical	Remove halos and find improved fit to the data, stronger driver effects, more reasonable inferences

2. Literature Review

This part concentrates on the presentation of the most relevant issues surrounding halo effects in consumer surveys. The chapter is organized as follows: Firstly, an overview of the scales applied in consumer surveys and cognitive response processes is provided. This is followed by the conceptual definition of the halo effect and statistical techniques to detect and correct the halo effect post hoc. Thirdly, causes of halo effects in consumer surveys and design-oriented approaches as suggested in literature are reviewed and discussed. This is followed by the development of hypotheses and the conceptual model underlying this study.

2.1. Systematic Measurement Error in Consumer Surveys

In marketing research nearly 30% of the empirical studies published in the Journal of Marketing and Journal of Marketing Research, during 1996 and 2005, apply surveys as research method (Rindfleisch et al. 2008). The study of halo effects in consumer surveys necessitates a closer look at cognitive response processes and the scales applied in these surveys to measure constructs of interest and the types of systematic biases that can occur and distort result of consumer surveys.

2.1.1. Cognitive Response Process in Surveys

The cognitive processes which take place unconsciously during the response process are crucial to understand how and why respondents halo.

Often cited in literature is the **belief sampling model** of Tourangeau et al. (2000). The model divides the response process into four stages: comprehension, retrieval, judgment, and response. First, a question is read and interpreted, followed by the retrieval of information from memory, which is then assembled to form a judgment about the particular issue of the question, and eventually the judgment is assigned to one of the offered response categories of the scale. The authors point out that these four components are part of a cognitive tool set, which respondents use to compose their answer to a survey question. For an optimal response, free from response error and generating high quality data, the respondent should carry out the complete process separately for each item of a multi-item scale. Therefore, the quality of the response depends on how exactly these steps are carried out by the respondent (Tourangeau et al. 2000). It is likely that not every respondent carries out this response process thorough and therefore each stage provides a basis for the respondent to bias his/her responses. For instance, respondents might be unable to assign their judgments to an appropriate response category, or unwilling to retrieve information from memory, and rather use information from more accessible sources such as previous answered items (Tourangeau et al. 2000; Tourangeau & Rasinski 1988).

Krosnick and Alwin (1987) and Krosnick (1991) call this phenomenon **satisficing**. The authors point out that respondents satisfice to reduce the cognitive effort related to the response process. As a result, questionnaire items are not processed with the required depth to give an optimal answer (Krosnick 1991). Krosnick (1991) describes several reasons that may lead the respondent to sacrifice. Firstly, the respondent might be unable to carry out all four stages of cognitive processes completely due to the lack of item-relevant knowledge or familiarity of the topic addressed by the item. The respondent might therefore not be able to retrieve information from memory and instead has to rely on other cues to respond to the item. Secondly, another reason leading the respondent to sacrifice is a lack of sufficient motivation. In consumer surveys, incentives such as monetary reward to trigger extrinsic motivation are mostly not given. Respondents, who like to engage in cognitive thinking, have a so-called need for cognition, and therefore have an intrinsic motivation to optimize their responses. However, respondents who dislike cognitive effort might be less motivated and therefore satisfice in their survey responses. The author points out that motivation decreases with response time, therefore satisficing behavior can be expected to be stronger at the end of a questionnaire (Krosnick 1991).

Jobe and Herrmann (1996) review seven cognitive response models in their study, and state that little research has been conducted on how accurate these models portray the response process (Jobe & Herrmann 1996). However, although the cognitive processes that take place when responding to questionnaires are not fully understood yet, these models help to better understand the processes that lead to halo effects.

2.1.2. Multi-Item Scales in Consumer Surveys

So-called multi-item scales or summated rating scales are commonly applied in consumer surveys are to measure the variable of interest. As distinct from single-item scales, where a construct is obtained by only one single attribute, multi-item scales contain several items which are summed up or averaged in order to measure a construct (Spector 1992). Such scales have been developed for a huge number of constructs in marketing (see Bearden &

Netemeyer 1999); a popular example is SERVQUAL – a multi-item scale to measure service quality (Parasuraman et al. 1986).

Multi-item measures have the reputation to deliver more reliable results compared to single-item measures. Furthermore they provide more detailed information than can be obtained by a single-item measure by capturing more facets of a construct (Baumgartner & Homburg 1996). Moreover, taken all the items together, they provide a more discriminating response scale, by offering all over more response categories, which allows making more exact distinctions among respondents (Churchill 1979; Bergkvist & Rossiter 2007)

First developed by Likert (1932) to assess attitudes, the underlying rationale of multi-item scales stems from classical test theory. Basically, it defines the relationship between observed score (e.g. measured satisfaction level of respondent) and a true score (e.g. actual satisfaction level of respondent) of a respondent on the construct. As the true score is unobservable, it has to be estimated by assessing observed scores. Within a multi-item scale the several items together are designed to be an observation of the measured construct. The observed score (O) is assumed to consist of the true score (T) and a random measurement error component (E): $O=T+E$. When combining the multiple items to obtain an estimate of the true score, errors are assumed to average approximately to zero, resulting in a reliable measure of the construct (Spector 1992).

A special type of multi-item scales, which is often effected by halo, are multi-attribute-attitude scales. Multi-attribute models to measure attitudes were first developed by Rosenberg (1956) and Fishbein (1967). These models are based on the assumption that beliefs about attributes of objects and the importance attached to these attributes together compose a respondent's attitude towards an object (e.g. brand) (Bass et al. 1972). The attitude towards an object is the sum of the weighted beliefs for the attributes: $A_{jk} = \sum_{i=1}^{n} I_{ik} * B_{ijk}$; where i=attributes, j=brand, k=respondent, A_{jk}=respondent k's attitude score for brand j, I_{ik}=importance weight given attribute i by respondent k, B_{ijk}=respondent k's belief on attribute i of brand j (Lehmann 1971).

In marketing these models are commonly applied to measure choice or purchase intent (Wilkie & Pessemier 1973) and was transferred to the assessment of satisfaction levels (e.g. Churchill & Surpenant 1982).It is the most common model used in satisfaction studies (Wirtz 2003, 2000). The model serves as an analytical tool to assess brand strength and weaknesses and brand performance, as well as to compare among brands (Wilkie & Pessemier 1973;

Wirtz 2000). Throughout the literature several model modifications have been advanced, and many studies find high correlations between overall attitude and the summed weighted beliefs on the attributes and report high predictability of attitude towards objects (e.g. Bass et al. 1972; Bass & Talarzyk 1972) (Beckwith & Lehmann 1975).

Multi-attribute scales are distinct from general multi-item scales since they direct the measurement of a construct (e.g. satisfaction) to several objects such as brands, whereas with general multi-item scales often only a construct is measured without concentration on a specific object (e.g. Component of Involvement scale (Lastovicka & Gardner 1979)) or the focused lies only on one object (e.g. SERVQUAL (Parasuraman et al. 1986)).

2.1.3. Response Bias in Multi-Item Ratings

Classical test theory, describing a simple relationship between observable and actual scores, ignores the possibility of systematic influences which in contrast to random error can alter the relationship among the individual items of the construct, leading to deviations between true and observed score (Spector 1992; Podsakoff et al. 2003). One of these systematic influences is common method bias, which is one of the main sources of measurement errors in multi-item scales. Method bias is a serious issue because it provides an additional source for the observed relations among measures of the different items, leading to inflated or deflated correlations among them (Williams et al. 1989; Podsakoff et al. 2003).

Measurement error can stem from context-, method-, and respondent-related sources. If the error can be attributed to the respondent, caused by a tendency to respond to scale items in a systematic way other than the items were designed for, it can be referred to as response bias, response set or response style (Baumgartner & Steenkamp 2001; Cronbach 1949; Paulhus 1991). However, this definition cannot be seen as exclusive, since the respondent can be influenced by the method used or the context s/he is confronted with (Baumgartner & Steenkamp 2006). Response biases are especially important, because they are common in market research, and although they are known to have a confounding influence on the results of consumer surveys which leads to wrong conclusions, researchers often do not pay attention to them (Baumgartner & Steenkamp 2006; Podsakoff et al. 2003).

An overview of common response biases that are, besides the halo effect, often present in multi-item ratings, and which have been discussed in literature, namely social desirability bias, acquiescence bias, extreme response style, leniency bias, midpoint responding, and consistency bias, is given in Appendix 1.

2.2. The Halo Effect in Consumer Surveys

This part of the thesis focuses on the halo effect in multi-item scales. First an overview of several conceptual definitions of halo effect is given, followed by a description of the form of occurrence of the halo in ratings, and a discussion of the effect of halo on multi-item measures.

2.2.1. Definition of the Halo Effect

In the literature, several conceptual definitions of the "halo effect" can be found, differing in their underlying assumed causal nature of the halo effect. There is wide spread consensus in regard to the operational definition of the halo effect, being an systematic response bias causing an inflation of correlation among scale items above natural levels and leading to a reduced variance in the data (Murphy et al. 1993; Murphy & Jako 1989; Beckwith et al. 1978).

Fisicaro and Lance (1990) categorized the several conceptual definitions of halo, which circulate in literature, into three causal models. They distinguish among general impression-, salient dimension-, and inadequate discrimination- halo effect, as distinct concepts (Fisicaro & Lance 1990). Wirtz (2003) reviewed these models, extended them with the associonist model of halo effects, and applied it to the context of consumer satisfaction (Wirtz 2003).

A. *General Impression Halo Effect:* Thorndike (1920), who was the first to label this kind of systematic response bias as "halo error", discovered an inability of supervisors to rate subordinates independently on different dimensions, leading to inflated correlations with their global impressions. He described this phenomenon as the "tendency to think of the person in general as rather good or inferior and to color the judgments by this general feeling" (Thorndike 1920, p.25). Nisbett and Wilson (1977) defined the halo effect in the same sense as the "influence of a global evaluation on evaluations of individual attributes of a person" (Nisbett &Wilson 1977, p.250). Balzer and Sulsky (1992) referred to this type of halo as "general impression halo", defining it as a bias of general impression that leads to the rating of performance dimensions being consistent with one's general impression (Balzer and Sulsky 1992). Figure 1 shows, that the general impression of the respondent (G) has a casual effect on the rated attributes R1 and R2, leading to increased partial correlations between these two attributes and the general impression (Wirtz 2003; Fisicaro & Lance 1990).
B. *Salient Dimension Halo Effect:* Fisicaro and Lance (1990) categorized the second type of halo effects as salient dimension halo effect which can be defined as the influence of the

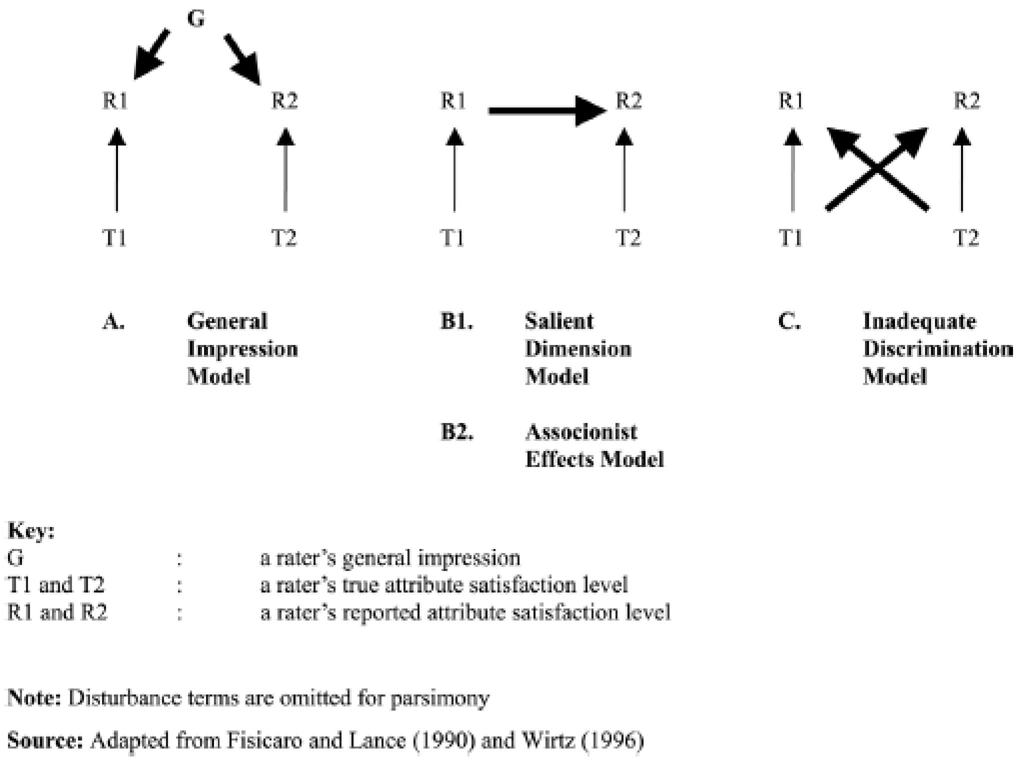
rating of less salient items by the evaluation of one or more dominant items (Nisbett & Wilson 1977, Kozlowski et al. 1986; Cooper 1981, Fisicaro & Lance 1990). As shown in Figure 1, the salient item R1 directly influences item R2, resulting in an increased correlation among items (Wirtz 2003; Fisicaro & Lance 1990). A special type of the salient dimension halo effect is the *associonist halo effect*, defined by Wirtz (1996). This type of halo effect influences the ratings in the same way as the salient dimension halo effect does. However the sequence in which items are rated plays a role here. The rater might get primed by an item, which in subsequence affects the evaluation of other items (Wirtz 1996, 2003).

C. *Inadequate Discrimination Halo Effect:* In the literature the halo effect is also often defined as the inability or unwillingness of a rater to distinguish among the evaluated attributes, leading to cross effects among these attributes, and resulting in increased correlations. Assumed relations among the items serve as basis of the response process rather than the actual information retrieved from long-term memory (Banks & Murphy 1985; Balzer & Sulsky 1992; Cooper 1981). Figure 1 shows these cross effects; whereas a rater's true satisfaction level T1 influences the rating on the item R2 and vice versa (Wirtz 2003; Fisicaro & Lance 1990).

Although the categorization of Fisicaro and Lance (1990) and Wirtz (2006) seems to be reasonable; empirically, it is difficult to differentiate among them, because of their identical outcome, namely an inflated correlation among attributes (Cooper 1981). Furthermore, the usefulness of the distinction for the analysis of halo effect in consumer surveys is questionable. From a different perspective the above described causal models cannot be assumed to be truly distinct, since they can occur at the same time and can influence each other. Especially the concept of inadequate discrimination seems to be more a prerequisite of salient dimension and general impression halo effect, since also there, the inability and/or unwillingness of a rater to differentiate or to recognize that items/attributes should be rated independently, motivates the rater to rely on global impression or salient items to rate items. Furthermore, the general impression halo can only occur in surveys where multi-attribute scales are applied and that actually contain a question in regard to global impression towards an object of the respondent. In contrast, salient dimension halo can occur in all multi-item measures.

For this thesis the following definition is chosen: The halo effect is the *unwillingness or inability of a respondent to differentiate among several items of a scale*. This inability or unwillingness leads the respondent to halo. This response strategy implies that the responses to attributes/items of a scale are influenced either by the general impression of an object for multi-attribute scales, or by salient items (such as preceding items) of a scale for general multi-item measures.

Figure 1 Conceptual Models of Halo Effect



Source: Wirtz 2003 (adapted from Fisicaro and Lance (1990) and Wirtz (1996))

2.2.2. Occurrence of Halo Effect in Consumer Surveys

Halo effects and effects similar to halo have been researched by a variety of researchers, using different notations, such as priming (Salancik & Pfeffer 1977), proximity effect (Weijters et al. 2008, Schwarz et al. 1991), state-dependence (De Jong et al. 2012), consistency (Salancik & Pfeffer 1977), salience effects (Schuman & Presser 1981), non-differentiation (Krosnick 1991), and logical error (Balzer & Sulsky 1992), as well as context effects (Podsakoff et al. 2003, Bickart 1993) and content carry over effects (Bickart 1993, Tourangeau et al. 1989, Tourangeau & Rasinski 1988).

Non-differentiation among the items of a scale is the origin of the halo effect. The respondent is unable or unwilling to differentiate among scale items because s/he invests too less cognitive effort due to a lack of motivation, boredom or fatigue. Since too less cognitive effort is spent, the response process as described by Tourangeau et al. (2000) is not carried out thoroughly, and all stages can be affected by prior items of a scale (Tourangeau & Rasinski 1988). The particular response strategy that respondents adapt when haloing, can be twofold: Non-differentiation can occur either as response carry-over or content carry-over from prior items to the current rated item.

Content carry-over refers to the transfer of beliefs from prior to subsequent items. For instance, in the retrieval stage, information retrieved for prior items is made highly accessible and can be retrieved easily for later questions from short-term memory to save cognitive effort. Therefore preceding items prime the respondent through making particular beliefs which are associated with prior items more salient. Content carry-over seems to be related especially to general impression halo effect, which occurs in multi-attribute scales. For instance, respondents can get primed if the global evaluation is located at the beginning of the questionnaire, so that they transfer retrieved beliefs and information to subsequent attribute ratings. Furthermore, it could evoke the desire to stay consistent and logic with the overall evaluation in the particular attributes ratings, although they should be rated individually. Moreover, general impression halo effect also seems to be related to response substitution, which is the tendency of respondents to express attitudes and beliefs that they were not asked for in the question. For instance, if a respondent has a highly positive attitude towards a brand in general, s/he might want to express this attitude as well in the attribute ratings, although they should be rated independently (Gal & Rucker 2011).

In contrary, when non-differentiation occurs in form of *response carry-over*, respondents do not necessarily assume a similarity of content among the items, but simply do not invest enough cognitive effort to assess the item individually, and therefore transfer their response from the preceding item to the current item (Rindfleisch et al. 2008; De Jong et al. 2012). Here, for instance a desire to stay consistent and logic with the previous response plays a role. Preceding items can define a scope for the interpretation of upcoming items, and induce a feeling of redundancy with earlier responses.

Both applied response strategies result in the repeated choice of identical or nearly identical response categories for several items.

2.2.3. Effect of Halo on Multi-Item Ratings

Many studies report high reliabilities and predictive validity of multi-item scales in consumer surveys. However, several researchers criticize the optimistic results obtained by multi-item scales, and point out that these results can be partly attributed to halo effects, and their role in the relationship among beliefs and attitudes (e.g. Beckwith & Lehmann 1975; Holbrook 1983) (Wirtz 2001).

For instance, Wilkie and Pessemier (1973) stress the relevance of the halo effect in multi-attribute models as a central concern surrounding their use. Halo effects are a problem since they affect the dispersion of belief ratings across attributes for an object, which is the basis of interferences on strength and weaknesses of brands on attributes (Wilkie et al. 1973). Particularly, the halo effect is problematic in regard to the basic assumptions of the model, as it suggests a dual causality in a way that not only beliefs form attitudes, but also attitudes influence beliefs (Beckwith & Lehmann 1975).

Although past research on halo effects in consumer surveys mainly focuses on multi-attribute scales, also general multi-item scales are affected by halo, residing in the underlying assumption of these models that several items/attributes are taken together to measure a certain construct such as satisfaction. Here the halo effect also causes higher spurious correlations between items and less variability in the data, leading to a spurious inflated reliability or lower predictive validity (Bradlow & Fitzsimons 2001).

2.3. Methods to Detect Halo Effects

In the literature several statistical methods to detect the halo effect and its magnitude are discussed. However, to measure the halo effect in ratings one has to take a closer look at its actual dimensions. Cooper (1981) states that the observed halo effect in ratings is the sum of a true halo level, representing natural correlations among items, and an illusory halo level, representing the respondent’s inability/unwillingness to discriminate among items (Cooper 1981; Pulakos et al. 1986). In this thesis, the halo effect is used synonymously with illusory halo. It is obvious that this relationship has to be taken into account when measuring the halo effect, since it is likely that some degree of true halo in ratings occur in a field setting. One can conclude that the observed halo effect is only partly an error, and therefore it has to be isolated from the true level in order to be properly measured (Pulakos et al. 1986). Taking this into account, true halo is set as a baseline and the excess correlation over this level can be assumed to be caused by halo effect (Cooper 1981; Pulakos et al. 1986). However, it might be possible to obtain true halo scores in a laboratory setting, whereas in field setting it is almost impossible.

In the following part methods to detect halo effects that are discussed in literature are presented. The majority of these techniques was developed to measure halo levels in the field of psychology where for instance in performance appraisals rater-ratee-dimension relations are measured. In marketing research these measures were therefore mainly applied to obtain halo levels in multi-attribute models, focusing on a rater-object-attribute relation. Eight measures for halo effect in multi-item measures are presented. They will be reviewed in short and criticized in regard to their appropriateness as a measure for halo effects. Furthermore, in Table 2 an overview is given, which measures can be applied to which type of multi-item scale.

2.3.1. Correlation Measures

One way to measure halo is to compute **correlation coefficients** ($r_{d,d'}$) between each pair of items (d, d'): ($r_{1,2} = {}^{\sigma_{12}}/_{\sigma_1 \sigma_2}$), resulting in a DxD correlation matrix, and for multi-attribute scales additionally between the overall rating and each of the attributes (Balzer & Sulsky 1992). These correlations can then be averaged across items by either simply averaging r scores (e.g. Guildford & Fructer 1973, p. 310-320) or by first transforming r to z scores using Fisher's r-to-z transformation (for a comparison see Strube (1988)) to obtain an average absolute measure of the inter-item correlations (Pulakos et al. 1986). A halo effect is assumed to be present if the correlations are higher than the level of true correlation among the items (Cooper 1981) (as discussed later in this thesis). The higher the correlations are, the less able the respondent was to discriminate among different items of the construct, and thus the stronger is the halo effect (Saal et al. 1980).

Another technique to detect halo effects is to conduct a **principal component or factor analysis** of the DxD correlation matrix to identify the inter-item factor structure. The fewer factors emerge relative to the number of items, accounting for most of the variance, the stronger is the effect of halo in the data (Saal et al. 1980; Jacobs & Kozlowski 1985). The identification of only one single factor indicates a maximum level of halo (Saal et al. 1980). The theory behind this is that such a data structure does not represent multidimensionality and is an indication for the inability of respondents to discriminate among items (Cooper 1981). For multi-attribute measures the first or common extracted factor is assumed to be a measure of the overall attitude, reflecting halo effect (James & Carter 1978; Leuthesser et al.1995).

Partial Correlation Measure: Partialling-out is a statistical technique to remove halo effects from data obtained by multi-attribute scales, however, the procedure can also be used to measure indirectly whether a halo effect is present or not. Halo is estimated by statistically removing the effect of an overall impression (g) rating from the ratings of the attributes (d) by estimating partial correlation coefficients $\left(r_{d,d',g}\right)$ for each pair of attributes (d,d'): $r_{d,d',g} =$

$r_{d,d'} - (r_{d,g} * r_{d',g})/(1 - r_{d,g}^2)(1 - r_{d',g}^2)$. Subsequently the corrected data is compared with the raw data, applying one of the above mentioned halo measurement techniques (Leuthesser et al. 1995). A halo effect is detected, if the processed ratings show lower inter-correlations among attributes, respectively if more factors emerge (Landy et al. 1980; Balzer and Sulsky 1992; Leuthesser et al. 1995).

2.3.2. Dispersion Measures:

Another approach is to obtain a measure of halo by computing the **variance** i $({\sigma^2}_i)$ or **standard deviation** (σ_i) of the ratings for each respondent, averaged across items (Pulakos et al. 1986; Cooper 1981). For multi-attribute scales the variance/standard deviation is additionally averaged across objects (e.g. brands) to obtain an absolute halo measure (Pulakos et al. 1986). Less dispersion among the attribute ratings, indicated by a small standard deviation or variance, is assumed to indicate halo effects (Jacobs & Kozlowski 1985; Saal et al. 1980).

For multi-attribute models a Rater x Object x Attribute **Analysis of Variance (ANOVA)** can serve as another measure for halo effects, first applied by Guilford (1954). A significant Rater x Object interaction $(MS\ (Raters\ x\ Objects)/MS\ (Raters\ x\ Objects\ x\ Attributes))$, that optimally explains a huge proportion of the variance, is seen as evidence for halo effects (Saal et al. 1980; Cooper 1981).

2.3.3. True Halo Measure

For all abovementioned measures, to assess the actual degree of halo in the ratings, the level of true halo has to be known. The halo effect is therefore estimated by comparing the difference between observed item inter-correlations and actual inter-correlations among the items (Fisicaro 1988; Cooper 1981), using true halo as a baseline. Obviously the main drawback of this method is how to assess true halo levels. One attempt employed in behavioral research is to use expert ratings to obtain true halo levels (Balzer and Sulsky 1992; Murphy & Jako 1989). In a consumer survey context, true halo scores can be estimated by measuring perceived similarity among items of a scale of the respondents. Another method which was applied for multi-attribute scales is to obtain true halo levels in an experiment. This involves the comparison of correlations between manipulated attributes in a treatment survey (halo effect is artificially induced) and non-manipulated attributes in a control survey. The correlation between the attributes in the control group is assumed to reflect true halo levels, whereas the difference in correlations to the treatment group represents the halo effect (Wirtz 2001, 2003).

2.3.4. Alternative Measures

An alternative approach to variance- and correlational-based measures in multi-attribute models is the method from Beckwith and Lehmann (1975), who estimate halo as a **regression coefficient** in a simultaneous equation model**.** In the first level of the model for each individual *(k)* overall attitude (A_{ik}) towards an object (i) is modeled as a function of the average global evaluation of all respondents (A_i^*) for each object and the respondent's individual belief (B_{ijk}) toward the attribute (j) for an object: $A_{ik} = \sum_{j=1}^{n} \omega_j B_{ijk} + \gamma A_i^* + u_0$. The estimated coefficients are measures for the relative importance of attributes for the individual respondent (ω_j), and the social impact of the attitudes of other individuals (γ). In the second level of the model, the beliefs on attributes towards each object are estimated for each individual as a function of overall attitude and average beliefs (${B_{ij}}^*$): $B_{ijk} = \beta_j A_{ik} + \gamma_j B_{ij}^* + u_{ijk}$ (Beckwith & Lehmann 1975). The estimated coefficients (β_j) are taken as measures of halo effect: the larger relative size and significance, the stronger is the influence of global evaluations on the respondent's belief and therefore the stronger is the halo effect (Holbrook & Huber 1979). The main advantage of this method is that it enables the measurement of the degree of halo effect that originates from individual attributes for each respondent (James & Carter 1978). This method has been applied in several subsequent studies (e.g. Bemmaor & Huber 1978; Moore & James 1978). Holbrook and Huber (1979) criticize the method in regard to its exclusive concentration on the individual level, and argue that the possible existence of common perceptual distortions (halo is conducted by all respondents) is ignored (Holbrook & Huber 1979).

A measure concentrating halo originating from the respondent's unwillingness or inability to differentiate among items of a scale is to simply **count identical or nearly identical ratings** to each item. Holbrook et al. (2003), for instance, use a count measure to obtain non-differentiation in ratings. This measure is also successfully applied a study of Tourangeau et al. (2004) A high number of similar or identical ratings is assumed to indicate halo effect.

Table 2 Overview of Halo Measures and Their Applicability Depending on Scale Type

Measure	General Multi-Item Scales	Multi-Attribute Scales
Inter-Item Correlations	✓	✓
Factor Analysis	✓	✓
Partial Correlations	×	✓
Standard Deviation/Variance	✓	✓
True Halo Measure	✓	✓
Beckwith & Lehmann (1975) Regression Model	×	✓
Counting Measure	✓	✓

2.3.5. Limitations of Methods to Detect Halo Effects

The measures presented share several methodological and conceptual limitations, affecting their usefulness to measure halo. Firstly, one critical point is that all variance-based and correlation-based measures compute halo across multiple respondents. However Balzer and Sulsky (1992) state that the unit of analysis should be at an individual level, since the halo effect is conceptually a response bias and therefore committed by the individual respondent. In this way the measure can be influenced by differences among respondents that are independent from halo effects (Balzer & Sulsky 1992). In contrast, other researchers claimed that the computation of halo effect across respondents delivers same results as individual level estimates (Solomonson & Lance 1997).

Secondly, most of these measures do not take into account the nature and actual appearance of the halo effect in ratings, which is the inability or unwillingness of respondents to distinguish among several items of a scale resulting in same or similar ratings across items of a scale. Therefore these measures themselves do not allow identifying the kind of response bias present in the ratings. For instance, high correlations or low variance in the data can be caused by several response biases such as state dependence or extreme response style.

A further problem for the correlational and ANOVA measures stems from the dependence of the measures on the amount of variance in the ratings. Little variance in the ratings has the effect that the magnitude of halo cannot be calculated accurately. Furthermore, in multi-attribute models a low variance among the objects can be seen as a necessary but not sufficient evidence of halo effects, since this measure does not assess a direct relation between overall and the individual attribute ratings. In an ANOVA a lack of variance influences negatively the emergence of a significant rater x object interaction. In the most severe case of no dispersion among the ratings, these measures cannot be applied at all (Balzer and Sulsky 1992).

Fourthly, researchers discuss logical problems with the ANOVA approach which assumes that the attributes are not conceptually independent, and only represent different levels of the same aspect. However, for instance multi-attribute models which are applied in consumer surveys assume the attributes to be independent factors which together explain the attitude of a respondent (Saal et al. 1980). Moreover, it is criticized that an ANOVA does not allow for the separation of object effects from effects of rating error, resulting in a possible overestimation of the halo effect to the extent of other factors distorting the attribute ratings (James & Carter 1978; Saal et al. 1980). Furthermore, the ANOVA measure can only be applied to data where a complete rater x object x attribute/item design is available. However in multi-item scales without a specific object, only a rater x item design is applied. In this case it is not possible to identify any halo effect, which is indicated by a significant rater x object interaction (Saal et al. 1980).

Furthermore, as already mentioned above, the level of true halo serves as a baseline to measure illusory halo. Most studies that apply correlation- or variance-based measures ignore possible natural correlations between individual attributes and overall evaluation (true halo). Since the true amount of correlation among attributes is unknown, it is always a somewhat subjective assessment of the researcher, when correlations are too high and to which degree a systematic response bias is present and therefore a halo effect can be reported. Leuthesser et al. (1995) state that one can assume as a general "rule of thumb" that a halo effect is indicated by average inter-correlations of 0.6-0.7 or higher (Leuthesser et al. 1995). Although there are indeed attempts to obtain true halo levels (e.g. Bormann 1977, 1978, Nisbett &Wilson 1977, Kenny & Berman 1980), there is no general agreement on a criterion to ensure that expert ratings are adequate. It is obvious that in this case the measurement of halo effect depends on the estimated of true halo, and can therefore only be as good as them, and literature suggests that they are not very good (Balzer & Sulsky 1992).

Moreover the above presented measures assume that the halo effect affects ratings in only one direction (inflation), however there are studies that find negative halo effects that actually deflate correlations (e.g. Murphy & Jako 1989; Kozlowski & Kirsch 1987; Fisicaro 1988).

Balzer and Sulsky (1992) examined the frequency of employment of different methods from 1980 to 1990 in studies of halo effects, and showed that the inter-correlational measure was applied in the most studies (37%), followed by the factor-analysis measure (21%). The less used technique is the partialling measure (Balzer & Sulsky 1992). Jacobs and Kozlowski (1985) point out that different measures of halo are not comparable because of their incongruence. Saal et al. (1980) state that the applied halo measure may cause a rating format to appear superior, whereas another measure would lead to another result.

2.4. Methods to Reduce Halo Effects

Several methods were developed in literature with the aim to reduce the occurrence of halo effects in consumer surveys. A distinction can be made between design-oriented approaches, that attempt to ex-ante prevent halo effects by targeting the causes of halo effects, and statistical-oriented approaches that aim to reduce halo effects post-hoc by applying statistical methods to the halo-contaminated data.

2.4.1. Statistical-Oriented Methods to Reduce Halo Effects (Post-Hoc)

For multi-attribute scales one procedure to control for halo effects is to **remove the first or common factor** that emerges of a principal component analysis or factor analysis. This factor is assumed to be a measure of the overall attitude and to reflect halo effect (Dillon et al. 1984).

The technique applied by far the most in literature for multi-attribute scales is **partialling-out** the halo effect (e.g. Holzbach 1978; Landy et al. 1980). This method involves the computation of partial correlation coefficients for each pair of attributes to remove the effect of overall evaluation (see chapter 2.3.1) (Leuthesser et al.1995; Cooper 1981).

The third technique is the **chaining strategy** as proposed by Holbrook and Huber (1979) which builds on the partialling-out technique and combines it with a principal component- and multiple-discriminant- analysis to remove halo effect. The researchers take up the critic points on Beckwith and Lehmann (1975)'s regression model to measure halo effects. The model neglects "common" halo effect, conducted by a group of respondents with shared preferences. The authors state that focusing only on individual haloing can result in biased belief ratings. The first step of the method involves the partialling-out of common halo effect. The halo-corrected correlation coefficients serve then as input for a principal components analysis to reduce the number of attributes for a subsequent multiple discriminant analysis (MDA). The resulting factors are used as independent variables in a MDA, whereas the rated objects are the dependent variables to be predicted. The resulting discriminant functions build a multidimensional space with data points for the ratings of each object by each respondent. Halo effect is removed by using centroid-based distance scores of the resulting discriminant functions (Holbrook & Huber 1979; Dillon et al. 1984).

Both the partialling-out and chaining strategies have the underlying assumption that the natural correlations among attributes ratings and overall ratings are zero. As some degree of true halo can be expected to be present, these techniques remove not only illusory halo effect but also true halo (Cooper 1981; Leuthesser et al.1995). Landy et al. (1980) therefore suggest that the residual scores should only be used as supplements to raw scores but not replace them (Landy et al. 1980). The advantage of the chaining approach over the partialling-out method is that it takes both common and individual haloing into account (Dillon et al. 1984). All three described techniques have only been applied to remove halo from multi-attribute scales.

A third approach is a **double-centering transformation** of a raw score matrix as performed by Dillon et al. (1984). This technique involves two steps, first the columns (corresponding to attributes) of the raw data matrix are standardized to attach meaning to rank ordering of respondents scores, and second, rows (corresponding to respondents) are transformed into ipsative data (rows sum to the same value) to eliminate irrelevant mean differences (halo) among respondents. In this way the centroid of both the respondents and the attributes is moved to the same point, and the influence of overall evaluation is removed, with respect to keeping response profiles across attributes intact (Dillon et al. 1984; Leuthesser et al.1995). Leuthesser et al. (1995) employ the double-centering method to assess halo effects as indicator of brand equity in their study. When comparing the centered data with the raw data they find halo effects varying in magnitude over the brands (Leuthesser et al.1995). The advantage of this technique is that it avoids the removal of true halo effects, since double centering a raw score matrix only eliminates elevation of correlations (Dillon et al. 1984). However the disadvantage of this technique lies in the associated statistical problems related to the analysis of ipsative data, residing in its mathematical interdependence. For instance, software for standard statistical estimation methods such as maximum likelihood cannot be routinely applied without adjustment (Cheung & Chan 2002). Moreover, the approach has only been applied to remove halo effect from multi-attribute models.

Another approach to remove halo effects is the application of a **mixture model.** Bueschken et al. (2010) develop a Bayesian mixture model that separates-out the haloed ratings and controls for heterogeneous scale use. The theory behind this approach is that survey responses consist of a mixture of formed and haloed responses. Respondents who halo can be identified by the low covariance of their ratings and are grouped in order to be separated from the respondents who form their ratings by assessing each attribute individually. The model integrates the factor of heterogeneous scale use, which is associated with the isolation of haloed responses, obscuring the model that generated the responses. Bueschken et al. (2010) apply their model to two customer satisfaction datasets, and are able to show improved fit compared to a regression that does not discriminate among the two groups (Bueschken et al. 2010). A disadvantage can be seen in the uncertainty associated to the identification of haloed responses. In the model, halo effect is identified by low covariance in the ratings, however this indices does not assure whether the identified response bias is halo effect or any other response bias causing little dispersion in ratings.

2.4.2. Design-Oriented Methods to Reduce Halo Effects (Ex Ante)

Several factors that favor the occurrence of halo effects in surveys are discussed in literature. Although the halo effect is considered to be a response effect, it can also be caused by contextual factors or by the method used, since the respondent's behavior is influenced by these factors.

The studies in marketing and psychological literature have investigated both internal and external factors which are suspected to cause halo effects. With regard to the internal factors, studies examined the tendency to maintain cognitive consistency (e.g. Cohen & Ahtola 1971*;* Bass & Talarzyk 1972*;* Beckwith and Lehmann 1975*)*, the positive cognitive reevaluation of important attributes of preferred brands (e.g. Cohen & Houston 1972), and the influence of demographics such as gender and education (e.g. Beckwith et al. 1978; Wu & Petroshius 1987).

External factors, causing halo effects are associated with the situational context of surveys, the rating procedure, the features of the rating task, and the rating scale applied. Most design-oriented attempts to reduce halo effects focus on external causes of halo effects, these will be reviewed as follows, as well as suggested solutions are discussed.

Context as Cause of Halo Effects

One cause of halo effects, associated with the context of consumer surveys is the *time span* between consumption of a product or service, and survey. Many studies found that a delay between rating and consumption can lead to halo effects by forcing the respondent to rely on global impressions or salient items instead of rating attributes individually. The theory behind is, that the rater has to recall the information from memory; however details are forgotten over time leading to the loss and addition of information (Cooper 1981; Murphy & Reynolds 1988). A straight-forward approach would be to reduce the time span between surveying the respondent and consumption to a minimum, with best expected results direct after the consumption. Wirtz (2001) examined this in a marketing context, and found halo effects to be reduced in a satisfaction survey of service conducted directly after consumption (Wirtz 2001).

Another contextual factor discussed in literature is the *survey purpose* as perceived by the respondent. Research has shown that raters seem to be reluctant to rate items independently if the rating is perceived as being part of a performance evaluation, as consumer don't want to feel responsible for negative consequences for employees as a result of a poor evaluation (Wirtz 2003). This especially plays a role in the evaluation of services, studies of Wirtz (2006, 2003) have shown that customers become more involved if the purpose is of a developmental nature (e.g. improvement of service). The theory behind is that customers are willing to invest more cognitive effort if they assume that their evaluation improves services or products (Wirtz 2003).

Rating Procedure as Cause of Halo Effects

Research related to rating procedures states that the *order of items* can cause halo effects, which is in line with the salient dimension/associationist model of halo effects, proposing that the rating of an item or attribute primes a respondent by activating information, affecting ratings of subsequent items (Murphy et al. 1993). For multi-attribute scales, solutions involve a modification of the rating procedure so that all objects on one attribute are rated before rating the next attribute (Symonds 1925; Cooper 1981; Wilkie et al. 1973). For instance, in a survey where different brands are to be rated, the respondent would be asked to rate all attributes at a time for one brand instead of all brands on one attribute at a time. If rating is conducted in this way, the ratings may be less affected by halo effects because the respondent focuses more on the attribute-brand relation and less on the relation among the attributes, since respondents tend to rely on their general attitude toward the brand when rating attributes at the same time (Wu & Petroshius 1987). Moreover, due to the same reasons the positioning of the global rating dimension and salient attributes in the questionnaire is important, and is assumed to reduce halo when it is positioned at the end (Cooper 1981). It is also suggested to randomize the order of items in the questionnaire (Wirtz 2001). Researchers do not agree on the effectiveness of this method, since results are not conclusive (Wu & Petroshius 1987). For general multi-item scales, instead of dispersing brands/products, the items of two or more scales are intermixed (e.g. Bradlow & Fitzsimons (2001); Feldman & Lynch 1988; Weijters et al. 2009). De Jong et al. (2012) study the effect of dispersing items of scales on the degree of state dependence and find that items, each dispersed through the inclusion of two unrelated "filler" questions, lead to a reduced number of identical ratings compared to ratings of items that were grouped together (De Jong et al. 2012).

Another cause of halo effects is the loss of motivation of respondents due a *lack of information* on the progress made in the questionnaire. When respondents are left in uncertainty about the length of a survey and their actual position, own expectations are made. During the rating process, when fatigue and boredom increase, this can result in a loss of motivation and cause haloing. A proposed solution in the literature is to provide information about the questionnaire length with a status bar, which indicates progress made either by question or screen (e.g. Greinöcker 2009; Couper et al. 2001). Couper et al. (2001) find an increased average response time when a progress bar was present in the survey. The authors conclude that this could reside in an increased cognitive effort of respondents when rating (Couper et al. 2001). However, the approach has only been applied in the research of non-response error in consumer surveys, but not in the specific context of halo effects.

As a further cause of halo that can be attributed to the rating procedure, is the *grouping of items on a screen* in web-based surveys. Often all items of one scale are presented clustered on a single screen. This leads to an increased time spend on one screen, which in turn can increase boredom and fatigue, since no feeling of progress is perceived. Furthermore, the respondent can see all items of a scale at once, and assumes a relation between, which leads to less differentiation, and thus halo effects. Solutions proposed in the literature focus on splitting scale items across several screens (e.g. Couper et al. 2001; Tourangeau, et al. 2004; Bradlow & Fitzsimons 2001). It could be shown that spreading items of a scale across several screens instead of presenting them grouped, does reduce inter-item correlations (Tourangeau et al. 2004; Couper et al. 2001). These studies show that the spread of items leads to a displacement of context among items of a scale due to the lack of visual presence, and therefore can motivate the respondent to retrieve new information to rate each item independently from previous ones, resulting in lower correlations among the items of a scale. This approach has not yet been applied to the specific context of halo effects.

Another factor widely discussed in literature is the *lack of knowledge* of the rater on the rating task, which involves the individual assessment of each item. Psychological research counteracts this problem by training the rater (e.g. Borman 1979, Cooper 1981). In the context of consumer surveys the respondents can be “trained” by providing *instructions* at the outset of the questionnaire, pointing out the importance of an individual assessment of the items, and in this way derogating the abilities of consumer to rate items independently (Wilkie et al. 1973). McCann and Wilkie (1972), and Wilkie et al. (1973) could report a positive effect on halo effect by including instructions in their surveys.

Rating Features as Causes of Halo Effects

Features of the object (e.g. product) can favor the occurrence of halo effects in consumer surveys. In this regard *familiarity* and *knowledge* of the respondent with the surveyed construct plays a role. It is assumed that less familiarity is connected with higher levels of halo effects. For multi-attribute scales it was found that raters who are more experienced with the brand/product being rated tend to be more objective in their ratings, having more knowledge and therefore are not forced to rely on global impressions or on salient items for the evaluation of individual items (Koltuv 1962; Kozlowski et al. 1986). The less experience the raters have with the product, the less differentiated are their ratings which can lead to halo effects (Murphy et al. 1993). Most studies report the above described negative relationship (e.g. McCann & Wilkie 1972; Wilkie et al. 1973; Koltuv 1962; Wu & Petroshius 1987; Wilkie et al. 1973; Wirtz 2003). Ex-ante methods focus on selecting only participants with sufficient experience and knowledge for the survey (Beckwith et al. 1978). This can be done by screening respondents for their suitability at the beginning of the questionnaire (Beckwith et al. 1978).

Furthermore, the product/brand itself or the *product category* can influence the degree of haloing. Research in psychology and marketing has proposed that the willingness of a respondent to discriminate among items of a scale might be determined by the level of involvement with the product (Banks & Murphy 1985; Wirtz 2003; Cooper 1981). Several studies report differences in the magnitude for halo effects in different product categories. For example, the halo effect has been found to be less for products that involve more consideration, such as cars (Moore & James 1978), and has been found to be stronger for household products such as toothpaste (Wilkie et al. 1973). Therefore, products that are purchased on a routine basis seem to be more affected by the halo effect, since the consumer might not be strongly involved, and considers all items independently and therefore might be influenced by salient important attributes or general impression when evaluating (Moore & James 1978; Wu & Petroshius 1987; Beckwith et al. 1978). Moreover, the halo effect can be stronger if perceived differences between products or brands are smaller (Moore & James 1978).

Rating Scale as Cause of Halo Effects

Halo effects can also be provoked by the rating scale and its items. There are many references in relation to multi-attribute scales that particular types of attributes may be more susceptible to halo effects, such as *important attributes* (Moore & James 1978). In general it would be expected that respondents rely less on general impressions or other salient attributes to form evaluations, the more important attributes are perceived because they are willing to put more effort in their ratings (Wu & Petroshius 1987). Whereas this view is supported by the result of some studies (e.g. Beckwith & Lehmann 1975; Wu & Petroshius 1987), there are also studies that report contradicting results, where ratings for important attributes were more haloed (e.g. Cohen & Houston 1972; James & Carter 1978*)*. Furthermore, also transferable to general multi-item scales, several more types of items/attributes are suspected to be prone to halo effects, such as items without physical correlates (Hubert & James 1976, James & Carter 1978), vague and ambiguous items (Beckwith & Lehmann 1975) and affective and subjective items such as style and appearance (Moore & James 1978; James & Carter 1978). One solution for this involves engaging the respondent in the rating scale construction, so that only attributes that are important by the individual respondents are rated (Cooper 1981). In psychology research, it was found that raters haloed less when the raters were asked to take part in scale construction before rating (Friedman and Cornelius 1976). However this method has not yet been tested in the context of consumer surveys, although one approach might be adaptive choice based conjoint where the scale is composited dependent on preferences of the respondent.

Another factor is the *lack of concreteness* of rating scales. If items are perceived as overlapping, ambiguous and redundant, the rater relies on prior items or general impressions due to the lack of distinctiveness, which in turn leads to higher halo effects (Leuthesser et al.1995; Cooper 1981). Design-oriented approaches suggest applying scales with only empirically derived, concrete and descriptive items (Cooper 1981). To some extent this can be influenced with questionnaire wording, which was found to influence the magnitude of halo effects (Wilkie et al. 1973; McCann & Wilkie 1972).

However, to cover all causes of halo effects that derive from rating scales, specific scaling methods have to be developed, such as relative rating scales (Cooper 1981). For instance, Bartlett (1983) states that scales that involve respondents to compare attributes or rated objects can decrease halo effects (Bartlett 1983). Popular scales, developed in psychology research, are the so-called forced choice scales (Bartlett 1983), and behaviorally anchored rating scales (BARS) (Smith & Kendall 1963), that describe the construct in terms of behavior (Cooper 1981). Kavanagh & Duffy (1978) used this approach in the rating attributes for a TV-show. The BARS scaling method has found to be effective in the reduction of halo effects (Cooper 1981). However there are no applications of this scaling method to consumer surveys to examine halo effects.

Another factor discussed in literature is the *number of attribute/items* (also often referred to as *under-sampling)* included in the scale. In studies of halo effect in multi-attribute scales, Murphy et al. (1993) and Wirtz (2003) reported that halo effects are stronger if there are only a few highly relevant attributes, whereas halo effects are less likely to occur if many attributes are rated, which are perceived as relatively unrelated to overall performance ((Murphy et al. 1993; Wirtz 2001). The theory behind, which also applied to general multi-item scales, is that the more items are required to evaluate; the more cognitive effort is needed in the rating process and therefore respondents are less likely rely on prior items or global impressions to form their evaluations. However, the number of items/attributes is limited, since fatigue and boredom, resulting from too long surveys, lead to less cognitive effort and can increase the likelihood of halo effects (Wirtz 2003, 2001; Murphy et al. 1993). Design-oriented approaches concentrate on the increase of the number attributes/items to be rated in surveys. However, there is no consent yet on the limits and definition of “few” and “many” attributes. Wirtz 2003, for example, used 5 and 10 attributes in his study, and suggested the relationship between number of attributes and halo effects to be u-formed (Wirtz 2003). It is assumed that if irrelevant items are explicitly rated and the rater is forced to fall back on prior items or general impressions, the relevant dimensions will be rated less haloed. However, the challenge is to find items or attributes that are related to the measured construct but less relevant (Cooper 1981).

Furthermore, the more items are included in the rating scale, the more difficult it gets for the respondent to distinguish among them, resulting in an increased degree of haloing (Murphy et al. 1993). In literature this is especially discussed regarding negative halo effects (deflation of correlations). Similar attributes/items are assumed to be naturally correlated (true halo). Some researchers state that the level of true halo may influence the level of illusory halo, in a way that if true halo is high, illusory halo might decrease natural correlations (Murphy & Jako 1989; Murphy & Reynolds 1988; Murphy et al. 1993; Cooper 1981). However, there are neither studies examining this in the context of consumer surveys, nor any design-oriented approaches to prevent this.

2.5. Hypotheses and Conceptual Models

In this study the halo effect in consumer surveys is examined. The focus hereby lies on how to prevent or reduce the occurrence of halo effects in a questionnaire by ex-ante design-oriented approaches. As noted in the literature review, general multi-item scales and multi-attribute models are often applied in surveys to measure marketing constructs. It has been shown many times that the ratings of these measures are distorted by halo effects. Therefore, in this study both kinds of such scales will be used to investigate the effect of halo.

Several studies in survey research report that survey design features such as scale presentation and context affects systematically influence the respondent's ratings. Four modifications of questionnaire design will be examined.

Consumer surveys nowadays are mainly administered web-based. It seems plausible that the specific conditions of this instrument such as design, visual presentation and the absence of an interviewer influence the ratings of respondents (Evans & Mathur 2005). One of these design features, which is only applied in web-based surveys, is the progress or status bar, indicating the remaining questions or the percentage of already answered questions, and signaling the actual position within the survey. The main purpose of a progress bar is to keep the respondent motivated during the rating process and to prevent a cancellation before the end by providing information on the progress made (Greinöcker 2009; Couper et al. 2001). Research mainly concentrates on the effect of status bars on non-response bias such as drop-out rates and question skipping (e.g. Couper et al. 2001; Heerwegh & Loosveldt 2006) and delivers mixed results. Couper et al. (2001) for instance find no significant differences in drop-out rates in an experiment, examining the effect of displaying a status bar. However, they could observe that the average response time was significantly higher when a progress bar was present. The authors conclude that this could, besides increased download times due to the progress bar, also partly reside in an increased effort of respondents when rating (Couper et al. 2001).

As the progress bar could be shown to prevent tiring and loss of motivation throughout the survey, a positive influence in regard to the reduction on the degree of haloing can be expected. If the respondent's motivation can be increased by the provision of information on surveying progress, then the respondent might put more effort into the ratings and therefore be more willing and able to distinguish among items of a scale. Since there is no research yet that examines the effect of status bars on response bias in surveys, this will be done in this study:

H1a: The presence of a status bar reduces halo effect in multi-attribute measures.

H1b: The presence of a status bar reduces halo effect in general multi-item measures.

The effect of proximity of items of a particular scale in a questionnaire on intra-construct correlations, perceived similarity, and response bias is the focus of several studies in marketing research. Schwarz et al. (1991) find that respondents perceive items of a scale as more related to each other if they are grouped (Schwarz et al. 1991). If respondents cannot distinguish among items of a scale they tend to use the same beliefs for all ratings because they perceive the question as redundant and therefore reduce their cognitive effort to assess new information. Furthermore, they transfer beliefs of previous answered items to the current item because it is still easily accessible in short term memory (Weijters et al. 2009; Tourangeau et al. 2000; Feldman & Lynch 1988). Feldman and Lynch (1988) find that the transfer of beliefs to a subsequent item increases as a function of proximity among the items (Feldman & Lynch 1988; Weijters et al. 2009). De Jong et al. (2012) study the effect of dispersing items of scales on the degree of state dependence and find that items, each dispersed through the inclusion of two unrelated "filler" questions, lead to a reduced number of identical ratings compared to ratings of items that were grouped together (De Jong et al. 2012). Moreover, Podsakoff et al. (2003) discuss the possible effect of intermixing items of different scales in a questionnaire. The authors argue that this could lead to reduced intra-construct correlations at the cost of higher inter-construct correlations. This relationship is confirmed in a study of Bradlow and Fitzsimons (2001) who examine the effect of dispersing versus clustering together the items of sub-scales in a multi-item scale. They find that if items are dispersed, then items of the same subscales are mainly uncorrelated. However, respondents then also tend to rely more on previous items in their rating, regardless the related construct (Bradlow & Fitzsimons 2001).

In regard to halo effects in multi-item scales, the dispersing of items can be expected to have a reducing effect since the overall construct is obscured and thus might reduce the reliance on general impression/prior rated attributes (De Jong et al. 2012). For multi-attribute models this already has been shown, where the halo effect could be reduced by modifying of the rating procedure so that all brands are rated on one attribute before rating the next attribute (e.g. Wilkie et al. 1973). The effect of dispersing items of general multi-items scales has been investigated in regard to response effects similar to halo, such as state dependence (e.g. De Jong et al. 2012; Bradlow & Fitzsimons 2001). However it has not yet been researched in the specific context of halo effect, which will therefore be done in this study:

H2a: The intermixing of brands in multi-attribute scales reduces halo effect.

H2b: The intermixing of the items of two general multi-item scales reduces halo effect.

Another design issue in web-administered surveys is whether the items of a scale are presented grouped on the same screen or each on a single screen. The effect of item presentation spread on several screens versus single page presentation in regard to context effects and correlations among items has been researched by several researchers (e.g. Couper et al. 2001; Tourangeau, et al. 2004; Bradlow & Fitzsimons 2001). Screen-by-screen design leads to a loss of context and therefore supports the respondent in focusing only on the current item to be rated. In contrast, the presentation of items grouped on one page increases perceived similarity (Toepoel et al. 2009). Tourangeau et al. (2004) find consistently lower correlations among separated items than among items presented grouped on a screen. They conclude that respondents perceive items as more related to each other in the single page design. Furthermore they recognize a higher tendency of respondents to choose the same answer categories for all items and reason that respondents might attach meaning to the grouping of items at the expense of carefully processing the question and therefore are less able to differentiate among items of a scale (Tourangeau et al. 2004). Couper et al. (2001) find small (& insignificant), but consistently higher correlations among items that are presented grouped on a single-screen (Couper et al. 2001).

These studies show that the spread of items leads to a displacement of context among items of a scale due to the lack of visual presence, and therefore can motivate the respondent to retrieve new information to rate each item independently from previous ones, resulting in lower correlations among the items of a scale. The effect of screen-by-screen designs versus single-page designs has not yet been studied in the context of response bias, and will therefore be examined in this study:

H3a: The presentation of attributes of a multi-attribute scale on separate pages reduces halo effect.

H3b: The presentation of items of a general multi-item scale on separate pages reduces halo effect.

Both the intermixing of items of two scales and the spreading of items over several pages is expected to influence the willingness of the respondent to distinguish among several items of a scale. Therefore it can be expected that the halo-reducing effect of combining both approaches is stronger compared to the single application:

H3c: A significant interaction exists between the intermixing of brands of multi attribute scales in combination with the spreading of the attributes on several pages that reduces halo effect.

H3d: A significant interaction exists between the intermixing of items of two general multi-item scales in combination with the spreading of the items on several pages that reduces halo effect.

There are some studies in response bias research focusing on the effect of the order of attributes and global dimension on the degree of halo in multi-attribute measures. Salancik and Pfeffer (1977) discuss order effects as a methodological problem in their need-satisfaction model. They describe priming and consistency effects, induced by salient attributes, as cause of order effects, leading to inflated correlations among attributes. They argue that these two factors shape the ratings because of the respondent's awareness to his precedent response (priming) and its logical inference to the next response (consistency). Due to the same reasons in regard to halo effects the positioning of the global dimension and salient attributes in a questionnaire seems to be important. Gal and Rucker (2011) discuss this phenomenon from a different point of view. They describe it as the tendency of respondents to convey attitudes and beliefs that were not asked for through their ratings to items, and refer to it as "response substitution" (Gal & Rucker 2011). This behavior seems similar to global dimension halo effect, where the rating of several dimensions is influenced by the global evaluation. The issue is whether the positioning of the global dimensions in the questionnaire plays a role for the respondent's desire to express his/her overall opinion. Whereas order effects of salient attributes can be controlled for by randomizing attributes so that the order is different for each respondent (Wirtz 2001), the positioning of the global dimension at the beginning of the scale can prime the respondent and induce consistent ratings. This could be controlled by placing this dimension at the end of the questionnaire (Cooper 1982).

Yet, there is no agreement on the efficiency of the positioning of the global dimension in consumer surveys (Wu & Petroshius 1987). Therefore this will be analyzed in this study:

H4: Halo effect is reduced in multi-attribute scales when the global evaluation is positioned at the end of the scale.

Furthermore, in multi-attribute models there are several factors apart from scale features that can increase halo effect by influencing the ability and willingness of a respondent to discriminate among the items of a scale, and that therefore need to be controlled for. Firstly, familiarity with the product category or brand has an impact on halo effects in the direction that higher levels of halo can be expected when a respondent is less familiar with the object rated (e.g. Wirtz 2003; Wilkie et al. 1973). Furthermore it has to be controlled for attribute importance, since attributes that are perceived as important by the respondent have been found to be either more prone to halo effects (e.g. Cohen & Houston 1972) or to be less susceptible to halo effects (e.g. Beckwith & Lehmann 1975). Moreover, researchers point out that the respondent's involvement with a product class influences the willingness and ability to discriminate among items of a scale and therefore is related to the degree of haloing (e.g. Wirtz 2003).

Figure 2 Conceptual Model: Halo Effect in General Multi-Item Scale

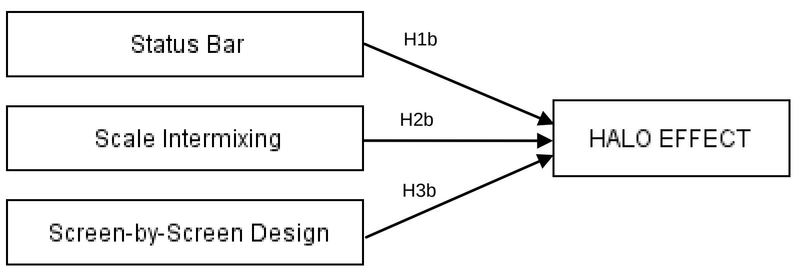
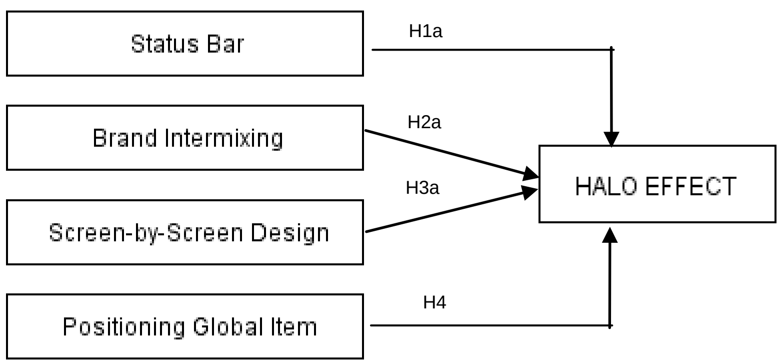


Figure 3 Conceptual Model: Halo Effect in Multi-Attribute Scale



3. Method

After setting-up the theoretical and conceptual framework and the development of hypotheses, this section presents the method and research design applied to the study of halo effect in consumer surveys and to test the hypotheses. First, the research design is discussed, followed by a description of the measures and manipulations, where the measures such as independent, dependent and control variables are described. Moreover, details about the applied questionnaire versions and about sampling and procedure are provided. Finally, the method of analysis in order to measure the variables of interest is presented.

3.1. Research Design

To explore the effect of halos in consumer surveys, an experiment is the most appropriate method, since it is possible to identify and isolate causal effects of manipulated independent variables on the dependent variable. In respect to the issues raised by the measurement of halo effects, such as the difficulties of obtaining true halo levels, it needs a true experimental approach (Murphy et al. 1993).

An issue surrounding experiments is their external and internal validity. In order to guarantee external validity, the ability of the results to be generalized to other settings besides the specific experimental scenario, the conditions should orientate on “populations, settings, times of the real setting” and include additional influential variables from the real world (Malhotra & Birks 2007). In marketing research external validity often goes at the cost of internal validity, which is the degree to which the observed outcomes can be drawn back to the treatments. If internal validity is not ensured results can be distorted by confounding effects such as extraneous variables, and therefore the measured causal reference is misleading. Therefore, often internal validity is also seen as pre-conditional to external validity. Internal validity can be best achieved in a laboratory environment, where the researcher is able to control for extraneous variables. An experimental design that is internally as well as externally valid is preferable (Malhotra & Birks 2007). In this study, to guarantee external validity, the environment of the experiment should be similar to the real scenario of consumer surveys in marketing practice, which are mostly distributed online. Therefore the experiment for this study is conducted in the field in the form of online surveys. Internal validity is maintained by the control for extraneous variables. Confounding effects can be minimized by selecting a sample size that is large enough, minimizing the probability that one of the groups is exposed to a third factor in isolation. Other researchers such as De Jong et al. (2012) and Rucker et al. (2011) successfully conducted experiments to examine response effects in a field setting.

As a true experiment is aimed to be conducted in this study, a precondition is that subjects are randomized into treatments to prevent selection bias and to have one or more control groups (Malhotra & Birks 2007). The most appropriate design for the study at hand is a between-subjects 2x2x2x2 fractional factorial between-subjects design. As it would go beyond the scope of this study to measure all of the 11 possible interaction effects, only a fractional design is applied. Therefore, the main effect of four independent variables and only one interaction effect between two variables on the degree of halo are measured.

3.2. Manipulations and Measures

To measure the effect of the treatments on the degree of halo effect, the ratings to the questionnaire versions with manipulated scales are compared to the ratings to a control version. To test the hypotheses developed in Chapter 2 the experiment consists of five treatments with each two levels and a control condition. Additionally, in a seventh condition it is controlled for the effect of questionnaire length.

3.2.1. Independent Variables & Manipulations

1. *Status Bar:* A status bar, indicating progress per screen, is displayed along the whole survey. The effect of the presence of a status-bar on the degree of halo effect in ratings of two general multi-item scales, and in a multi-attribute scale for smartphones is measured.
2. *Scale Intermixing:* In a multi-attribute scale brands are intermixed, so that each attribute is rated for all brands at a time. In order to disperse the items of a general multi-item scale, the items of a “focal” scale are intermixed with the items of a “filler” scale, in a way that two items of the “filler” scale follow on one item of the “focal” scale. The effect of intermixing items on the degree of halo effect is measured in the multi-attribute measure as well as in both multi-item scales.
3. *Screen-by-Screen:* For the multi-attribute measure, the items for each brand are shown on a single screen; whereas in the other conditions the items are shown in groups of three brands per screen. Furthermore, the items of the two general multi-item scales are displayed in groups of five per screen; whereas in the other conditions the items are presented in groups of ten per screen. The effect of displaying items screen-by-screen on the halo effect in the ratings in the multi-attribute scale and the general multi-item scales is measured.

4. *Combining Scale Intermixing and Screen-by-Screen:* The treatments as described under point 2 and 3 are combined.
5. *Positioning of the Global Evaluation at the End:* The global evaluation item for each brand in the multi-attribute scale is positioned at the end of the rating of the according bran; whereas in the other conditions the global evaluation item is placed before the brand rating.
6. *Control Version:* None of the treatments, described in points 1-7, is applied.
7. *Questionnaire Length:* 105 items of general multi-item scales are positioned additionally at the beginning of the questionnaire.

3.2.2. Measurement of the Dependent Variable

In this study the dependent variable is the degree of halo effect in the ratings of the respondents. As reviewed in Chapter 2 there are several problems in regard to the measurement of halo effects, such as unknown levels of true halo and unknown baselines of correlational and variance measures. Furthermore, all existing measures have several advantages and disadvantages. For instance, although correlational and variance measures are often used in halo research, a researcher cannot be sure whether higher values are due to a halo effect, or if they are caused by another response bias.

To test the hypotheses the ratings in different questionnaire versions are compared to the control version. In this way, the halo-reducing effect can be obtained without dependence to the availability of true halo levels (Wirtz 1996). To take into account the form of appearance of halo effects (the assignment of identical or nearly identical ratings to distinct scale items) in the ratings, a count measure is applied. More precisely, a zone counting technique is applied by making use of a point system. For identical ratings to subsequent items two points are assigned; whereas for nearly identical ratings (+/- one category on the response scale) one point is assigned. To obtain a score per respondent, the points are summed for each scale. Whereas for general multi-item scales similarity to the preceding item was counted, for multi-attribute scales similarity to the overall attitude rating was measured. If average of nearly identical and identical ratings is significantly less in the treatments than in the control version it can be assumed that halo effect was reduced.

3.2.3. Questionnaire and Applied Multi-Item Scales

The questionnaire is programmed with the online survey provider "Surveygizmo.com". Each experimental design is applied in an own questionnaire, resulting in six versions (see Table 3).

Each of the questionnaire versions (see Appendix 2) starts with introducing the respondents to a smart phone survey conducted in the framework of a master thesis with the aim to find out more about customers of smartphones. This is followed by a block of questions concentrating on demographics such as age and education.

In the next part respondents state their beliefs towards the four attributes of smartphones (performance, ease of operation, features and physical design) for five brands (iPhone, Blackberry, HTC, Nokia, Samsung) on 5-point scales from “Not at all favorably” to “Very favorably”. Additionally they state their overall attitude toward the brands (global evaluation). In all versions, except version 5, the global evaluation item is rated before the attribute items. In version 5 the global evaluation is placed after the attribute ratings. In all conditions, besides 2 and 4, the brands are rated separately for all attributes, whereas in the intermixing conditions the scales are dispersed by the brands, so that all brands are rated simultaneously for each of the attributes.

The next part consists of the Components of Involvement (CP) Scale from Lastovicka and Gardner (1979) (focal scale), containing 22 items to measure a person’s involvement with a product. The items are rated on 5-point Likert-type scales. The scale consists of three subscales (familiarity, commitment, and normative importance). To obtain a measure for each factor the items can be summed within each factor (Bearden et al. 2011, pp. 237). The scale was chosen because it can be easily applied to several product classes and it is subdivided into three subscales. To shorten the scale in the applied questionnaire only a subset of ten items was administered. Furthermore, the items’ wordings were slightly modified by exchanging general expressions such as “the product” with “the smartphone” to ensure that the respondent understands what product to rate when the scale is intermixed in version 2 and 4. In all versions the scale items are displayed in groups of ten items on the screen, with exception to version 3 and 4, where each item will be displayed in a group of only five items per screen.

This scale is followed by the Use Innovativeness (UI) Scale from Price & Ridgway (1983) (filler scale), measuring variety seeking in the use of products. It consists of 44 items and five subscales (Creativity/Curiosity, Risk Preferences, Voluntary Simplicity, Creative Reuse, and Multiple Use Potential). The items are measured with 5-point Likert-type scales. To obtain the overall use innovativeness measure items ratings are summed (Bearden et al. 2011, pp. 237). This scale was chosen because it is different to the CP scale and low correlations can be expected. Therefore the respondent can be expected to notice the difference among the constructs since the questions do focus more on personality than on a product. 20 items of the scale were chosen randomly.

In version 2 and 4 the CP and UI Scale are intermixed in a way that each item of CP-Scale is followed by two items of the UI Scale, resulting in a mixed scale of 30 items. In version 4 the items are shown in groups of five per screen, whereas in version 2 items are grouped in 10. In version 1 a status bar, indicating the respondent's progress, is shown throughout the whole survey. Version 6 is the control questionnaire, where none of the treatments is applied.

Additionally, a 7th questionnaire version is applied to control for the effect of questionnaire length on the halo effect. In this version, 105 items of several general multi-item scales are placed prior to the multi-attribute scales. Whereas the average time to complete questionnaire versions 1-6 is 7-10 minutes, it is 15-20 minutes for version 7.

Furthermore, questions to measure attribute importance (How important are the following smartphone attributes for you?) are employed in all versions.

Table 3 Questionnaire Versions

Version	Status Bar	Scale Intermixing	Screen-by-Screen	Global Evaluation	Survey Length	Hypothesis
1	yes	no	no	no	short	H1
2	no	yes	no	no	short	H2
3	no	no	yes	no	short	H3a
4	no	yes	yes	no	short	H3b
5	no	no	no	yes	short	H4
6	no	no	no	no	short	
7	no	no	no	no	long	

3.3. Sampling and Procedure

To be representative the sample should represent the typical target respondents of a consumer survey on smartphones. Since smartphone users heavily participate in social networking, respondents are recruited via online channels such as Linkedin.com, Xing.de, Facebook.com and Twitter.com. The participants are randomly assigned to one of the six questionnaire versions. To prevent over- or under-sampling of the several treatments a quota of 40 samples is set for each version.

3.4. Method of Analysis

The objective of the study is to reduce halo effects by ex-ante design approaches. As a first step it has to be assured that a halo effect is present in the data. As there is no general measure for the halo effect several analyses will be applied, such as principal component analysis, average inter-correlations of items of scales, as well as a linear regression. A halo effect can be assumed to be present one applied measure indicates halo effect. The next step of analysis is the testing of the hypotheses. Hereby, the count measure is applied and each version will be compared separately to the control version (6). To ensure that differences are statistically significant the independent sample t-test will be applied. To control the results a robustness check is conducted.

4. Data Analysis and Results

In this part, details to the conducted analysis are presented. Firstly, a brief summary of the main results of the smartphone study is given, followed by details to the conducted preliminary analysis, such as sample characteristics and drop-out rates. Further, the hypotheses are tested. The softwares SPSS Statistics 19 and StataSE 11 are applied to conduct the statistical analyses.

4.1. Consumer Survey on Smartphones

The content of the survey which served as the framework for the experiments was “Smartphones”. In this chapter the main results of the survey on smartphones are reported.

Sample

As shown in Table 3 the total sample size is 280 respondents, while male participation predominates (62%). Furthermore, most respondents have a higher education or an university degree (95%). Almost half of the respondents own an iPhone (42%), followed by Samsung (18%) (Figure 4), which can also be interpreted as market shares.

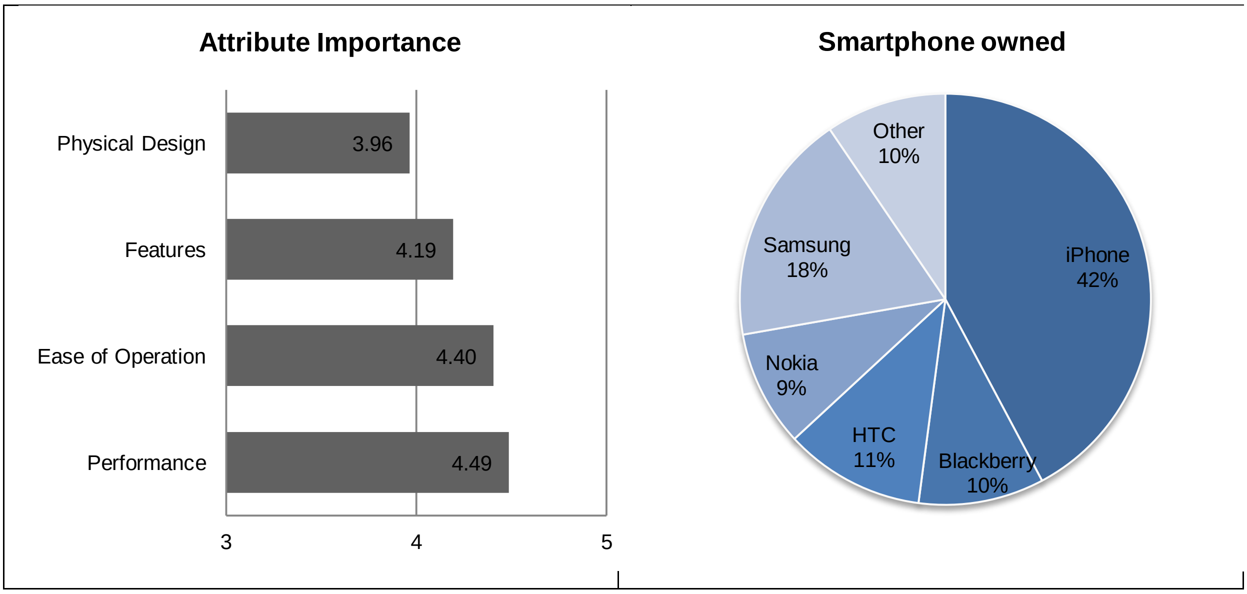
Attribute Importance

The respondents were asked to state their importance for four smartphones attributes: performance, ease of use, features and physical design. The results (Figure 4) show that “performance” is the most important attribute (M 4.49; SD 0.73), closely followed by “ease of operation” (M 4.4, SD 0.81). The least important attribute is “physical design” (M 3.96, SD 0.9). For “performance” and “features” no significant differences between owners of specific smartphone brands and perceived importance can be observed, implying that these attributes are equally important for users of all brands. However, significant differences can be observed for the attributes “ease of operation” and “physical design”. For instance, iPhone users do attach more importance to “ease of operation” (M 4.64, SD 0.63) than do respondents who own a Samsung (M 4.23, SD 0.66) or a Blackberry (M 4.08, SD 0.94). In regard to the attribute “physical design” it can be noticed that iPhone users (M 4.26, SD 0.75) value this attribute more than Samsung users (M 3.75, SD 0.84). For age groups significant differences exist only for the attribute “features”, in the sense that older people (>34) do attach more importance to this attribute (M 4.38, SD 0.68) than younger people (18-34) (M 4.13, SD 0.86). In regard to sex, differences exist for the attribute “physical design”, which women (M 4.09, SD 0.97) value more than men (M 3.88, SD 0.85).

For manufacturers of smartphones these findings could imply that they should focus on “performance”, as it is the most important smartphone attribute for all participants. Furthermore, the results seem to be consistent to the reported findings of other smartphone surveys. For instance, users of the iPhone value the attributes “ease of operation” and “physical design” more than users of other brands. The iPhone distinguishes itself by its operating system (iOS) which is easy to use, and its elegant design, whereas other brands, such as Android-based phones, set the focus more on functionality. Furthermore, the findings suggest that smartphone manufacturers who develop models that target women should besides the other attributes, especially focus on a nice physical design (e.g. color), which is especially important to women. Moreover, if the target group is the age group over 45, the focus should lie on features (e.g. navigation).

When interpreting attribute importance ratings, it should be noticed that attribute importance was self-reported by the respondents on a five-point scale. Attribute importance obtained in this way, does not necessarily measure the actual weight of attributes in purchase decisions. The results can be distorted, since respondents can confuse importance with salience for instance. Thus, more reliable results could be obtained by a conjoint study, where respondents make choices based on trade-offs between products with different attribute levels. In this way attribute importance can be measured indirectly (Heeler et al. 1979).

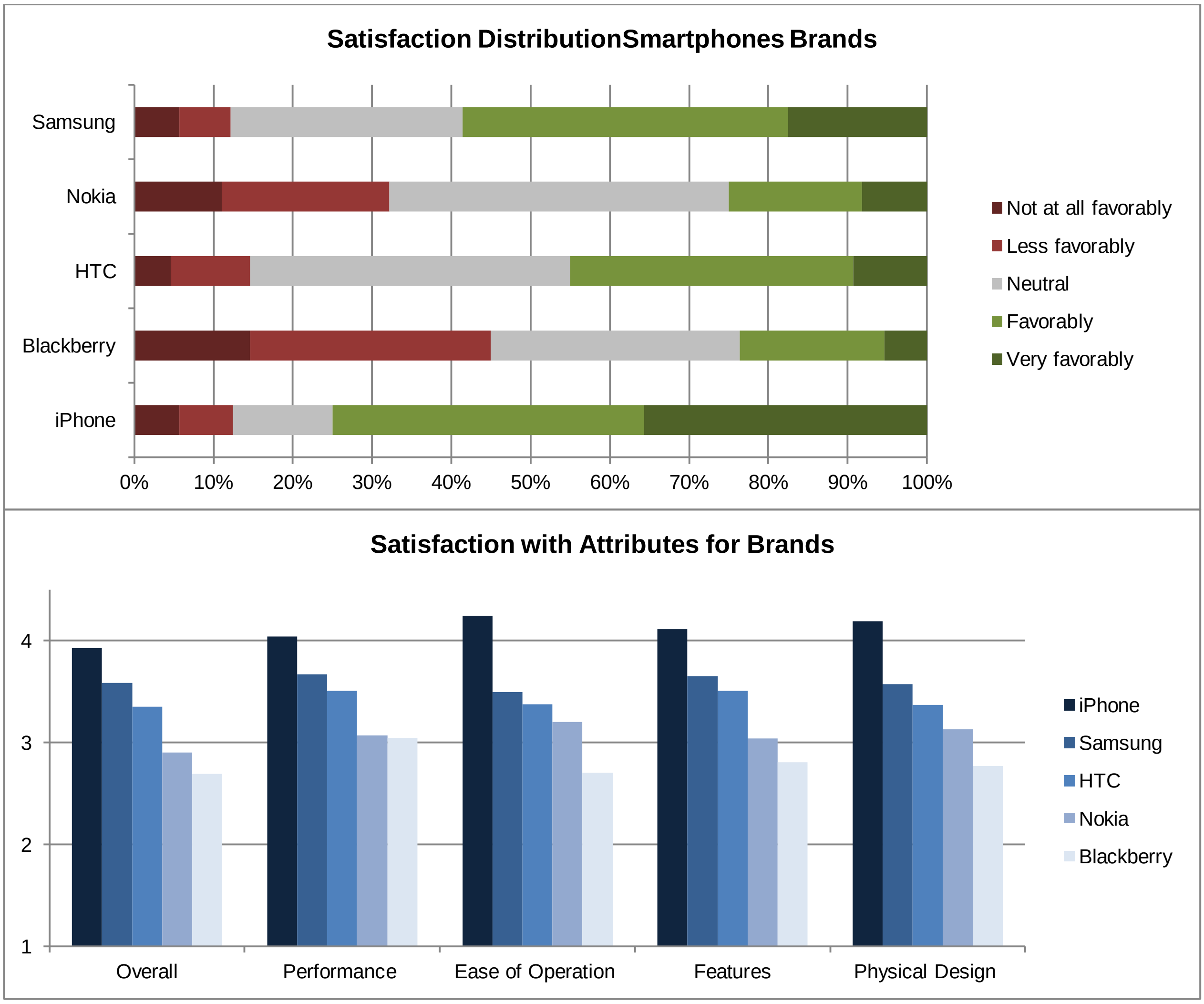
Figure 4 Attribute Importance and Smartphone owned



Overall & Attribute Satisfaction

Figure 5 displays the proportions of respondents who assigned their satisfaction level to each of the categories of the 5-point scale from "Not at all favorably" to "Very favorably". For instance, the iPhone is rated by more than 80% of the respondents as "favorably" or "very favorably", followed by Samsung with 60%. Last positioned in this ranking is Blackberry with 45% of the respondents reporting negative satisfaction ("not at all favorably" or "less favorably"). Taking a closer look at the satisfaction in regard to the attributes of the brands, it strikes that the iPhone is in the lead for all attributes and Blackberry and Nokia are rated the worst on all attributes.

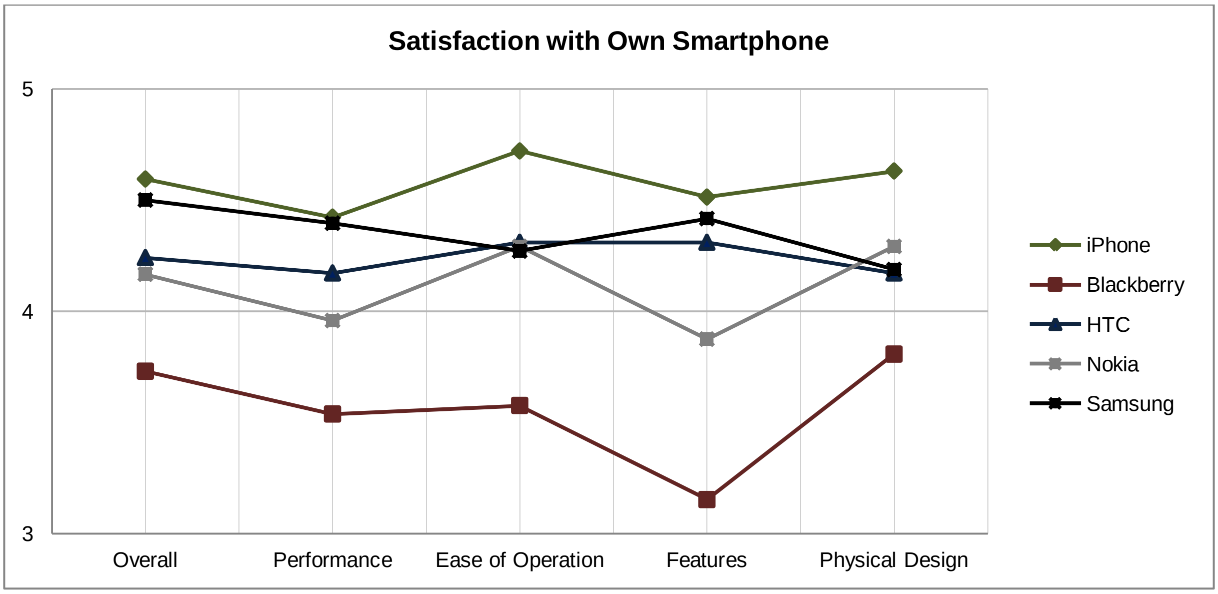
Figure 5 Satisfaction Distribution & Satisfaction with Attributes for Smartphone Brands



Satisfaction with the Own Smartphone

A similar pattern can be observed, when considering satisfaction ratings for the brand the respondents own themselves (see Figure 6). IPhone users report the highest satisfaction levels on all attributes, whereas Blackberry users indicate the lowest satisfaction levels throughout. Striking are the low ratings for Nokia and Blackberry on the attribute “features”. This might be explained by the operating systems (Blackberry OS and Symbian OS/WindowsPhone OS for Nokia), which are less popular and do not offer as many features (e.g. apps) as iOS- and or Android-based smartphones.

Figure 6 Satisfaction with Own Smartphone



Familiarity

Familiarity with smartphones was measured with a subscale of the Components of Involvement (CP) Scale from Lastovicka and Gardner (1979). The ratings were summed to obtain a score, indicating the degree of familiarity with smartphones. The results show that significant differences exist only for respondents who do not own a smartphone. Non-smartphone owners score the lowest (M 19.3 SD 4.016); whereas the highest scores are obtained by iPhone (M 24.63, SD 3.52) and HTC users (M 25.03, SD 3.51).

Creativity

Creativity was measured with a subscale of the Use Innovativeness Scale from Price and Ridgway (1983). No significant differences could be observed among owners of different brands.

4.2. Preliminary Analysis

The preliminary analysis provides details to the characteristics of the sample such as demographics, drop-out rates for the different questionnaire versions, and the degree of halo effect that is present in the underlying data.

4.2.1. Sample Characteristics

Overall 404 respondents filled in the seven questionnaires. After the removal of partial responses a sample of 280 respondents remained, equally dispersed over the seven questionnaire versions (40 per version), which was achieved by quotas. As shown in Table 4, across the seven versions 61% of the participants are male and 39% are female. The average age is 29.8 years, and 83% of the respondents went to university, 13% have a higher education (10-12 years). When comparing demographics of the several versions, it can be noticed that in each version more men than women participated. Furthermore, the average age lies between 27 and 34 years, and in all versions most of the respondents (>75%) have a university degree.

Table 4 Sample Characteristics

		Total	1	2	3	4	5	6	7
Sample Size		280	40	40	40	40	40	40	40
Gender	Male	61.7%	57.5%	72.5%	57.5%	70.0%	60.0%	52.5%	57.50%
	Female	38.3%	42.5%	27.5%	42.5%	30.0%	40.0%	47.5%	42.50%
Age		29.1	28.7	29.5	30.5	28.5	27.2	30.2	34.3
Education	No education	0.4%	0.0%	2.5%	0.0%	0.0%	0.0%	0.0%	2.5%
	1-4 years	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	4-9 years	4.2%	7.5%	0.0%	0.0%	7.5%	5.0%	5.0%	2.5%
	10-12 years	13.3%	7.5%	10.0%	15.0%	17.5%	17.5%	12.5%	10.0%
	University	82.1%	85.0%	87.5%	85.0%	75.0%	77.5%	82.5%	85.0%

4.2.2. Drop-Out Rates

Although per se a “non-response error”, the degree of “drop-out” in the questionnaires can provide additional information in regard to the effect of the treatments. Lack of motivation, as well as boredom and fatigue can lead to either response error, characterized by a reduced willingness to spend cognitive effort and resulting in non-differentiation (halo effect), or to non-response error which is the termination of the questionnaire before its end, resulting in an incomplete response.

The respondents were randomly assigned to one of the questionnaire versions, whereas for each version a response quota of 40 was set. Table5 shows the drop-out rates across versions, as well as the results of an independent samples t-test to compare the ratio of drop-out across versions. It can be observed that the longer version (7) has, as expected, the most drop-outs (68). This can mainly be attributed to the fact that 15-20 minutes were required to complete this version, in contrary to the shorter versions with a duration of only 7-10 minutes.

Interestingly, version 5 has the least number of drop-outs. It is surprising that the positioning of the global evaluation element should have an influence on drop-out. For this reason, the actual points of drop-out of each partial response was studied more precisely. When comparing across the different versions it can be observed that on average one third of the respondents terminate in the beginning of the questionnaire, during the multi-attribute scale part. Unfortunately this does not allow drawing a conclusion in regard to the number of drop-outs in version 5. Whether this effect is attributed to the treatment or to other factors in the questionnaire could be examined in future research.

The highest ratio of drop-outs in the multi-attribute-part takes place in version 1, where the effect of a status bar was tested. This as well is surprising, since often status bars are implemented in questionnaires to reduce the drop-out rate. The theory behind is that the respondent is informed about the length of the survey and can observe his/her progress, which should motivate to proceed. However, this process does not seem to be linear. A status bar can also lead to the contrary effect, an increase of drop-out especially at the beginning of a questionnaire, attributed to the fact that the respondent is able to build expectations on the length of the survey. This phenomenon can be observed here.

Significant differences in drop-out rates to version 6 exist for version 3 (screen-by-screen) and 4 (screen-by-screen & intermixing). For version 3 and 4 this might be attributed to the frequent screen changes, motivating the respondent because s/he perceives to proceed faster through the questionnaire. The difference between version 3 and 4 is not significant (p=0.152), therefore the additional intermixing of scales does not add to the reduction of drop-outs.

Table 5 Comparison of Drop Outs across Versions

	1	2	3	4	5	7	6
No. Drop-outs	13	17	12	7	3	63	16
Mean	0.75	0.7	0.77	0.85	0.93	0.39	0.71
Std. Deviation	0.43	0.46	0.43	0.36	0.26	0.49	0.46
Mean Difference to v6	0.04	-0.01	0.05	0.14	0.22	-0.33	
p-value (one-tailed)	0.318	0.442	0.26	0.046	0.002	0.00	

4.2.3. Measurement of the Halo Effect

Additionally, it is interesting to examine the applicability of different halo measures to this study, that were presented in the literature review. Thus, several halo indices such as factor analysis, correlational analysis, and linear regression are applied to measure halo effect in the control questionnaire version (6). Since no baselines for these measures are available that would allow to derive a reliable conclusion in regard to the degree of halo present in the data, halo effect is additionally compared to a longer questionnaire version (7), which included 105 additional items of general multi-item scales. The halo effect can be assumed to increase with survey length; therefore a comparison to version 6 allows judging the degree of halo present.

Factor Analysis

A technique often applied to detect halo effects is to conduct a factor analysis to identify the inter-item factor structure in the data. Based on the number of emerged factors and the amount of variance accounted for by the first factor a conclusion can be drawn whether a halo effect is present. The fewer factors emerge relative to the number of items, the stronger is the halo effect in the data (Saal et al 1980; Jacobs & Kozlowski 1985). For multi-attribute scales, the first or common extracted factor is assumed to be a measure of overall attitude (James & Carter 1978; Leuthesser et al.1995). The theory behind this is that such a data structure does not represent multidimensionality and is an indication for the inability of respondents to discriminate among items (Cooper 1981).

A principal component analysis was conducted for version 6 separately for the multi-attribute scales and the general multi-item scales. The findings show that for the MA-scales 6 components emerge (Eigenvalues >1). Relative to the number of items (25) the ratio is 0.24. The first factor accounts for 22% of the variance (Table 6). For the MI-scales this ratio is 0.30 (9 components, 30 items). For the longer questionnaire version (7) the ratio of emerged components to the number of items is 0.36 for the MA-scales (9 factors, 25 items) and 0.17 for the MI-scales (5 components, 30 items).

When comparing version 6 to 7 it can be noticed that for the multi-attribute scales more factors emerged in version 7 than in version 6, and less variance explained by the first factor. This would indicate that in version 7 the halo effect is less. Assuming the difference is statistically significant, this could be attributed to the fact that the MA-scales are perceived as a “structural break”, compared to the prior 105 items from MI-scales. This variation could motivate the respondent to invest more cognitive effort and therefore to reduce non-differentiation. For the MI-scales the findings are reversed, the ratio of emerged components is less for version 6. This could be explained by the same argument. Since the 105 items at the beginning of the questionnaire are of the same type as the two MI-scales in the second part of the questionnaire, the respondents could perceive them as repetitive and therefore redundant, resulting in a loss of motivation and an increase in halo effect.

On a first view, according to the factor measure one could conclude that halo is dependent on the length of the questionnaire especially for the MI-Scales, where the differences are higher. However, unfortunately testing the ratio of emerged factors between two samples is not straightforward. Therefore the conclusions drawn are only based on the assumption that the differences are statistically significant. The difference could be tested by applying a bootstrapping approach, where both samples are "resampled" several thousands of times. The difference of emerged factor structure could be tested by comparing the bootstrap distributions of both groups. However this goes beyond the scope of this study.

Table 6 Factor Analysis Comparison Version 6 & 7

	Version	6	7
MA	Ratio of emerged components	0.24	0.36
	Variance explained by 1st component	22%	17%
MI	Ratio of emerged components	0.30	0.17
	Variance explained by 1st component	22%	35%

Inter-Item Correlations

The degree of halo in the data can also be measured by calculating correlation coefficients between each pair of items of a scale. To obtain a single measure, the coefficients are averaged per scale. Since Pearson's product-moment correlation coefficients (r) are not additive, a transformation to z scores by applying Fisher's r-to-z transformation is necessary. For the multi-attribute scales the coefficients were first averaged per brand, and then again averaged across brands to obtain a single coefficient for the multi-attribute scales. Similarly, for the general multi-item scales the average coefficients were first calculated separately for focal and filler scale, and then averaged to obtain a single measure. Only significant correlation coefficients were included in the average calculation.

To obtain whether a halo effect is present in the data, the correlations should be compared to true halo levels, which are not available. A halo can be assumed to be present in the data if the correlations are high. For instance, Leuthesser et al. (1995), define as a rule of thumb that inter-item correlations of 0.6 indicate a halo effect (Leuthesser et al. 1995). Cohen (1988) defines a correlation coefficient to be "large" if it is >0.5 and to be "medium" if it is >0.3 (Cohen 1988, pp.79-81). Table 7 shows the average correlation coefficients for both MI- and MA-scales for questionnaire versions 6 and 7. It can be seen that for control version 6, the average correlation for the MA-scales is 0.57, and can be classified as "large" after Cohen (1988) and is close to Leuthesser et al. (1995)'s rule of thumb. For the MI-scales the average correlation is 0.44 which can be classified after Cohen (1988) as "medium". For version 7 the average correlation coefficient for the MA-scales is 0.76 ("large"), and for the MI-scales 0.39 ("medium"). To compare the coefficients for version 7 and 6 a z-test was conducted. For the MA scales the difference between version 6 and 7 is marginally significant at the 10%-level, whereas for the MI-scales no significant differences exist.

It can be concluded that the average inter-item correlation for the MA-scales can be classified as "large", indicating halo effect. For the MI-scales only medium average inter-item correlation are present. Based on the correlational measure and the given baselines a halo effect can be assumed to be present for the MA-scales. When controlling for questionnaire length, the findings show that for the MA-scales correlations are marginally significant higher than in the shorter version (6). Therefore, questionnaire length has an influence on the degree of halo. For the MI-scales the length of the survey does not influence the degree of halo.

Table 7 Correlational Analysis Comparison Version 6 & 7

Version	6	7	p-value (one-tailed)
Average Correlation MA	0.57	0.76	0.068
Average Correlation MI	0.44	0.39	0.386

Halo as a Regression Coefficient

The degree of halo effect can also be estimated in the form of a regression coefficient. Thus in this study the Beckwith and Lehmann (1975) approach is applied. As their approach is only directly applicable to multi-attribute scales, a similar modified approach was transferred and applied to general multi-item scales.

Multi-Attribute Scales

Beckwith and Lehmann (1975) measure the halo effect in multi-attribute scales on an individual level with a regression coefficient. In the first level of the model for each individual *(k)* overall attitude (A_{ik}) towards an object (i) is modeled as a function of the average overall attitude of all respondents (A_i^*) for each object and the respondent's individual belief (B_{ijk}) toward the attribute (j) for an object: $A_{ik} = \sum_{j=1}^{n} \omega_j B_{ijk} + \gamma A_i^* + u_0$. The estimated coefficients are measures for influence of beliefs on overall attitude for each attribute (ω_j), and the social impact of the attitudes of other individuals (e.g. conformity seeking) (γ). In the second level of the model, the beliefs on attributes towards each object are estimated for each individual as a function of overall attitude and average beliefs (${B_{ij}}^*$): $B_{ijk} = \beta_j A_{ik} + \gamma_j B_{ij}^* + u_{ijk}$ (Beckwith & Lehmann 1975). The actual measure of halo is β_j, measuring the influence of overall attitude on beliefs. The larger relative size and significance, the stronger is the impact of overall attitude on the respondents attribute ratings, and therefore the stronger is the halo effect (Holbrook & Huber 1979). The influence of average beliefs (γ_j) is a measure of the true position of all respondents on the attributes. The functions are both estimated separately via ordinary least squared (OLS), and simultaneous via two stage least squares (TSLS).

To measure halo effect the Beckwith & Lehmann (1975) approach was applied to this study. To assure that the coefficients do not reflect individual weighting of important attributes, the variables were likewise standardized to have a mean of zero. The regressions were estimated via OLS. Contrary to the Beckwith & Lehmann (1975) version, the model was not estimated for the individual, but pooled across respondents. The main reason for this is that due to instability of individual estimates, comparison across the different questionnaire versions is compounded (e.g. different sample sizes). Additionally, critics state that the approach would focus too intensively on the individual level, neglecting halo effect that is conducted by the sample as a whole ("common perceptual distortions") (Holbrook & Huber 1979). Furthermore, pooling causes errors to be correlated since each respondent rated all brands, and therefore violates one of the assumptions of linear regression. This has can cause standard errors to be biased. To correct for serial correlation, robust standard errors were estimated by clustering respondents and allowing errors to be correlated within clusters but not between (see Roger 1993, Williams 2000).

Table 8 shows that for the control version (6) both all the halo coefficients (β_j) of overall attitude, as well as all the coefficients for average position on attributes (γ_j) are significant.

Comparing the sum of the halo coefficients (2.16) to the sum of the specific average belief coefficients (1.95) (as Beckwith and Lehmann (1975)), it can be seen that halo is a stronger contributor to the beliefs than true average position. In Table 9 the findings of the attitude regression are presented. The results show that overall attitude is formed by beliefs. However, whether this represents actual consumer behavior or whether the findings are distorted by the halo effect cannot be derived. To examine this relationship the functions need to be estimated simultaneously (Beckwith and Lehmann 1975). Furthermore, the insignificance of the average overall attitude coefficient indicates that conformity seeking of the respondent does not have influence on the overall attitude towards a brand.

For the longer questionnaire version (7), after summing the significant coefficients, it can be seen that the influence of halo effect seems to be much stronger (2.64) than average beliefs (0.91). To test whether a version-specific effect exists and to compare the coefficients of version 6 and 7, version-interactions with the dependent variables are included in the regression (Gujarati 1970). A group-specific effect for beliefs is only present for the attributes performance and physical design. Thus, the halo effect is stronger in version 7 only for the attribute performance, whereas it is less for the physical design. For the influence of beliefs on overall attitude no significant differences between the versions could be found.

Table 8 Coefficients of the Belief Equations Versions 6 & 7

Dependent Variable Belief B_j	β_j coefficient γ_j coefficient	6	7	p-value (one tailed)
Performance	Overall Attitude	0.535*	0.720*	0.005
	Average Performance	0.512*	-0.143	0.026
Ease of Operation	Overall Attitude	0.444*	0.694*	0.150
	Average Ease of Use	0.627*	0.324**	0.282
Features	Overall Attitude	0.556*	0.667*	0.140
	Average Features	0.489*	0.357***	0.428
Physical Design	Overall Attitude	0.625*	0.555*	0.000
	Average Physical Design	0.320***	0.226**	0.000

Annotation: 1) Beliefs on attributes (j) towards each brand (i) were estimated pooled across individuals (k) and brands (i) as a function of overall attitude (A_{ik}) and average beliefs (B_{ij}^*): $B_{ijk} = \beta_j A_{ik} + \gamma_j B_{ij}^* + u_j$.
2) * significant at 1%-level, ** significant at 5%-level, *** marginally significant at 10%-level.
3) p-value: significance-level of version interaction

Table 9 Coefficients for the Attitude Equations Version 6 & 7

ω_j **coefficient** γ **coefficient**	**6**	**7**	**p-value (one tailed)**
Performance	0.243*	0.295*	0.240
Ease of Operation	0.186**	0.289*	0.089
Features	0 .225**	0.168***	0.052
Physical Design	0.326 *	0.131	0.192
Average Overall Attitude	0.054	0.367***	0.306

Annotation: 1) Overall Attitude (A_{ik}) towards each object (i) were estimated pooled across individuals (k) and brands (i) as a function of beliefs (B_{ijk}) and average overall attitude (A^*_{ij}): $A_{ik} = \sum \omega_j B_{ijk} + \gamma A^*_{ij} + u_0$
2) * significant at 1%-level, ** significant at 5%-level, *** marginally significant at 10% level.
3) p-value: significance-level of version interaction

To allow a better comparison between the versions, additionally the belief function was estimated pooled across attributes. In this case it was assumed that the halo effect is not attribute-specific, thus overall attitude influences belief ratings for all attributes in the same way. Table 10 presents that in the pooled estimation a significant main effect for both halo effect and average beliefs can be found. The halo coefficient is higher, implying that halo effect contributes stronger to belief ratings. Compared to version 7, halo effect is significantly higher in version 7 than in version 6 (p=0.037). To conclude, the regression technique indicates that beliefs are both formed and haloed. Moreover, the halo effect is stronger in the longer survey version.

Table 10 Pooled MA Regression Coefficients Comparison Version 6 & 7

β_j **coefficient** γ_j **coefficient**	**6**	**7**	**p-value (one tailed)**
Overall Attitude	0.561*	0.730*	0.037
Average Attribute Rating	0.450*	0.278*	0.130

General Multi-Item Scales

To apply a similar measure to the general multi-item scales, the Beckwith and Lehmann (1975) approach was adapted. The difference to multi-attribute scales is that non-differentiation does not relate to the overall attitude, but to preceding items. For this reason the following function was estimated via OLS: $R_{s,i,k} = \beta * R_{s,i-1,k} + \gamma * R^*_{s,i} + \varepsilon$; whereas the rating of respondent k to item i of scale s: $R_{s,i,k}$ is modeled as a function of $R_{i-1,k}$, each respondent's rating to the previous item, and $R^*_{s,i}$; the average rating of all respondents for item i of scale s. Halo effect is measured through β, the coefficient of the previous rated item, reflecting non-differentiation through the influence of the rating to the previous item. The coefficient γ of the average item rating of all respondents measures the true content by averaging ratings across all respondents.

The coefficients have to be interpreted with caution. The halo coefficient β cannot be assumed to be a pure measure of halo, since it also reflects true content. For the comparison among versions the assumption is therefore that there is no difference in content across versions, and that differences in scores can be attributed to a change in the degree of halo effect.

In Table 11 it can be seen that both the rating to the preceding item, as well as the average item rating has an influence on the ratings, whereas the average item rating has the stronger influence. This implies that there is a halo effect present, but true content seems to be the main contributor to item ratings. Comparing to version 7 a marginally significant difference (p=0.098) can be found for the halo effect. This implies that the halo effect is stronger in the longer questionnaire version.

Table 11 Pooled MI Regression Coefficients Comparison Version 6 & 7

	6	7	p-value (one tailed)
Rating to Previous Item (β)	0.182*	0.242*	0.098
Average Item Rating (γ)	0.822*	0.761*	0.207

1) p-value: significance-level of version interaction

Conclusion on the Degree of Halo in the Data

To draw a conclusion in regard to the degree of halo in the data the three applied measures are compared. The *factor analysis* indicated that halo effect is stronger for the MA-scales; however no baselines for obtained factor structure are available. Questionnaire length seems to have an influence, indicating that the halo effect is reduced for the MA-scales and increased for the MI-scales, assuming statistical significance, which was not tested. The *inter-item correlations* showed that based on the frameworks given by Cohen (1988) and Leuthesser et al. (1995) for the MA-scales a halo effect is present, since correlations can be classified as "high". For the MI-scales halo effect seems to be less with "medium" inter-item correlations. Survey length has an increasing effect on correlations for the MA-scales. Moreover, the applied *regression measure* indicates a significant influence of halo effect on both scales, with a stronger impact in MA-scales. For both scales a halo increasing effect of questionnaire length could be found. Regarding survey length, a significant difference could only be found for the MI-scales indicating that length has a stronger influence on halo effect in MI-scales, which might be attributed to both the positioning of these scales in the survey and their similarity to the additional rated items.

To conclude, all three applied halo measures indicate that a halo effect is present in both MI- and MA-scales. However, non-differentiation seems to be stronger for the latter. Regarding questionnaire length the results are more ambiguous. For the MA-scales only inter-item correlations and regression measures indicate a stronger halo effect, whereas the factor analysis indicates an opposite result. For the MI-scales also only two measures show higher halo effects, whereas the correlational measure does not show an effect.

4.3. Hypothesis Testing

To test the hypotheses the number of (nearly) identical ratings (count measure) of each version is compared to the control version 6, which serves as the baseline for the level of halo. If the average score obtained by zone counting technique in the treatment version is less than in version 6, and this difference is statistically significant, it is assumed that the halo effect was reduced and the corresponding hypothesis is supported. To compare the different versions with the control version, an independent sample t-test is applied. This test allows comparing the mean of one variable of two independent groups. The assumptions are the independence of the two compared groups, and the equality of variances. The latter is controlled for with the Levene's for Equality of Variances.

4.3.1. Hypothesis 1 a) and b)

Hypothesis 1 was stated as follows: a) *The presence of a status bar reduces halo effect in multi-attribute measures* and b) *The presence of a status bar reduces halo effect in general multi-item measures.*

In a direct comparison (Table 12) of the means of the count measure, it can be noticed that for the multi-attribute scale (ZC MA GE) the mean for version 1 (M 30.40, SD 5.48) is lower than in control version 6 (M 31.53, SD 5.12). However the independent sample t-test indicates that the difference is not statistically significant ($p=0.173$). For the general multi-item scales (ZC MI) the mean of the count measure in version 1 is almost equal (M 33.43, SD 6.54) to version 6 (M 33.45, SD 5.45).

No difference exits for both scales. The Hypotheses 1 a) and b) therefore are rejected. The presence of a status bar does not reduce the level of halo effect.

Table 12 Count Measure T-Test Comparison Version 6 & 1

	Version	Mean	Std. Deviation	Std. Error Mean	Mean Difference	p-value . (one-tailed)
ZC MA GE	6	31.53	5.12	5.12		
	1	30.40	5.48	5.48	1.13	0.173
ZC MI	6	33.45	5.45	0.86		
	1	33.43	6.54	1.03	1.35	0.493

4.3.2. Hypothesis 2 a) and b)

Hypothesis 2 was formulated as follows a) *The intermixing of brands in multi-attribute scales reduces halo effect* and b) *The intermixing of the items of two general multi-item scales reduces halo effect.*

A direct comparison of the means of the count measure for the multi-attribute scales shows that the mean of the count measure is lower for version 2 (M 29.18, SD 5.42) compared to the control version 6 (M 31.53, SD 5.12). The independent sample t-test indicates that the difference is statistically significant (p=0.025) (see Table 13).

The comparison of means for the general multi-item scales shows that the mean is higher for the control version 6 (M 33.45, SD 5.45) compared to version 2 (M 31.35, SD 6.43). The t-test indicates that the difference is marginally significant at the 10%-level (p=0.060).

Hypotheses 2a) and b) are supported; the intermixing of brands in multi-attribute scales and the intermixing of two general multi-item scales reduces halo effect.

Table 13 Count Measure T-Test Comparison Version 6 & 2

	Version	Mean	Std. Deviation	Std. Error Mean	Mean Difference	p-value (one-tailed)
ZC MA GE	6	31.53	5.12	5.12		
	2	29.18	5.42	5.42	2.35	0.025
ZC MI	6	33.45	5.45	0.86		
	2	31.35	6.43	1.02	1.33	0.060

4.3.3. Hypothesis 3 a) and b)

Hypothesis 3 was stated as follows: a) *The presentation of attributes of a multi-attribute scale on separate pages reduces halo effect* and b) *The presentation of items of a general multi-item scale on separate pages reduces halo effect.*

Table 14 shows that for the multi-attribute scales the mean of the count measure is lower for version 3 (M 29.55, SD 5.07) than for the control version 6 (M 31.53, SD 5.12). The t-test shows that the differences are statistically significant (p=0.044). For the general multi-item scales the mean of the count measure for treatment condition 3 is lower (M 32.05, SD 4.30) than in the control version 6 (M 33.45, SD 5.45), however the difference is not statistically significant (p=0.103).

Hypothesis 3a) is supported; the presentation of multi-attribute scales in a screen-by-screen format does reduce halo effect. Hypothesis 3b) is rejected; the presentation of general multi-item scales in a screen-by-screen format does not reduce halo effect.

Table 14 Count Measure T-Test Comparison Version 6 & 3

	Version	Mean	Std. Deviation	Std. Error Mean	Mean Difference	p-value (one-tailed)
ZC MA GE	6	31.53	5.12	5.12		
	3	29.55	5.07	5.07	1.98	0.044
ZC MI	6	33.45	5.45	0.86		
	3	32.05	4.30	0.68	1.10	0.103

4.3.4. Hypotheses 3 c) and d)

Furthermore, the interaction between the intermixing of scales and the presentation per screen was stated in Hypothesis 3 c) *A significant interaction exists between the intermixing of brands of multi attribute scales in combination with the spreading of the attributes on several pages that reduces halo effect* and d) *A significant interaction exists between the intermixing of items of two general multi-item scales in combination with the spreading of the items on several pages that reduces halo effect.*

In Table 15 it can be seen that for the multi-attribute scales a significant difference (p=0.011) in mean for the count measure between treatment condition 4 (M 28.65, SD 5.79) and the control condition 6 (M 31.53, SD 5.12) exists.

For the general multi-item scales also a difference between control version 6 (M 33.45, SD 5.45) and treatment version 4 (M 31.30, SD 4.45) is present. This difference is statistically significant (p=0.029).

Hypotheses 3c) and d) are supported; the combination of intermixing of the brands of a multi-attribute scale, as well as the items of two general multi-item scales in combination with a screen wise presentation does reduce halo effect.

Table 15 Count Measure T-Test Comparison Version 6 & 4

	Version	Mean	Std. Deviation	Std. Error Mean	Mean Difference	p-value (one-tailed)
ZC MA GE	6	31.53	5.12	5.12		
	4	28.65	5.79	5.79	2.88	0.011
ZC MI	6	33.45	5.45	0.86		
	4	31.30	4.45	0.70	1.11	0.029

4.3.5. Hypothesis 4

Hypothesis 4 was formulated as follows: *Halo effect is reduced in multi-attribute scales when the global evaluation is positioned at the end of the scale.*

Table 16 shows that the mean of the count measure is higher (M 32.10, SD 4.87) in treatment version 5 compared to control version 6 (M 31.53, SD 5.12). The difference is not statistically significant (p=0.304).

Hypothesis 4 is rejected. The positioning of the global evaluation item at the end of the multi-attribute scale does not reduce halo effect.

Table 16 Count Measure T-Test Comparison Version 6 & 5

	Version	Mean	Std. Deviation	Std. Error Mean	Mean Difference	p-value (one-tailed)
ZC MA GE	6	31.53	5.12	5.12		
	5	32.10	4.87	4.87	-0.58	0.304
ZC MI	6	33.45	5.45	0.86		
	5	34.05	5.97	0.94	1.28	0.320

4.3.6. Conclusion Hypothesis Testing

An overview of the results of the hypotheses testing is presented in Table 17. To summarize, no significant differences between treatment and control version could be found when the effect of a status bar on multi-attribute and general multi-item scales was tested. Both hypotheses H1a and H1b therefore were rejected. Secondly, the testing of the effect of intermixing of brands in MA-scales, and of items of two MI-scales delivered positive results. Both hypothesis H2a and H2b were supported. Further, a positive impact of screen-by-screen design on the degree of halo effect in MA-scales could be proven, thus Hypothesis H3a was supported. Moreover, also the combination of screen-by-screen design with brand intermixing in MA-scales proved to be effective, thus H3c was supported. Whereas for the MI-Scales the displaying of items screen-by-screen did not have any effect on the degree of halo which led to the rejection of H3b, the combination with scale intermixing showed a reducing effect on halo effect in MI-scales. H3d was therefore supported. Lastly, the positioning of the global evaluation item at the end of MA-scales did not prove to be effective, thus H4 was rejected.

Table 15 Overview Results Hypotheses Testing

Hypothesis	Treatment	Supported
H1a	Statusbar in MA-Scales	×
H1b	Statusbar in MI-Scales	×
H2a	Brand-Intermixing in MA-Scales	✓
H2b	Scale-Intermixing in MI-Scales	✓
H3a	Screen-by-Screen Design for MA-Scales	✓
H3b	Screen-by-Screen Design for MI-Scales	×
H3c	Combination Brand Intermixing & Screen-by-Screen Design for MA-Scales	✓
H3d	Combination Scales Intermixing & Screen-by-Screen Design for MI-Scales	✓
H4	Positioning of Overall Attitude Item at the End for MA-Scales	×

4.4. Robustness Check

To substantiate the results obtained through the hypothesis testing, a robustness check of the count measure is conducted. Furthermore, an additional robustness check is conduced, using correlational analysis and regression measure as control measures.

4.4.1. Robustness Check of the Count Measure

To assure that the conclusions derived from hypothesis testing are not attributed to the applied counting technique, a robustness check of this measure is necessary. Mainly there are two counting methods that can be applied to measure halo effect, a) a single-counting measure, which involves simply counting the number of cases a respondent assigns the same scale category to two subsequent items of a scale, and b) a zone-counting measure, which also takes into account nearly identical ratings to subsequent items by making use of a point system. For identical ratings to subsequent items two points are assigned; whereas for nearly identical ratings (+/- one category on the response scale) one point is assigned. To obtain a score per respondent, the points are summed for each scale.

Furthermore, for multi-attribute scales it is possible to either count the number of (nearly) identical ratings across the whole scale (for all brands and attributes), or to count the number of ratings that are (nearly) identical to the general evaluation item per brand. Moreover, in regard to the general multi-item scales, identical ratings can be either counted across both scales or separately for each scale.

Table 18 shows the results of comparing the number of (nearly) identical ratings of each questionnaire version to the control version (6). A "*" indicates significant differences at 1%-level, "**" at the 5%-level, "***" indicates marginal significant differences at the 10%-level, and "NS" no significant differences, tested with the independent samples t-test.

For the multi-attribute scales (MA) both the single count measure (a) and the zone count measure (f) indicate similar results (significant/marginally significant differences for version 4). Furthermore, counting the number of ratings that are identical to the ratings of the general evaluation item (MA GE), for both single- (b) and zone-count (g), significant differences exist for version 2, 3, and 4, but the zone-count indicates a higher significance for version 3. Moreover, comparing MA GE to MA, it can be seen that for both, single and zone-counting, more significant results are achieved with the MA GE counting method.

For the general multi-item scales, it can be seen that if the count measure is applied across focal and filler scale (MI), for both zone count (j) and single count (e) significant differences exist for version 2 and 4. For the focal scale, both counting techniques do not indicate any significant differences. For the filler scale, both counting techniques indicate significant differences for version 2 and 4. Additionally when the zone count (i) is applied, marginally significant differences exist for version 3. To have one measure for the general multi-item scales the MI counting technique was applied to measure halo effect.

Version 7 is considered separately as it is not related to hypothesis testing, but to control for the dependence of halo effect on the length of a survey. Both zone and single counting technique indicate that significant differences exist. However, results for the single count show stronger significance and do not indicate differences for the multi-attribute scales, whereas the zone count does.

Table 16 Significant Differences To Version 6 Based on the Counting Method Applied

Treatment/ Version Counting Method	Status Bar (1)	Intermixing (2)	Screen-by-Screen (3)	Combination 2&3 (4)	Global Evaluation (5)	Survey Length (7)
a) Single Count MA	NS	NS	NS	**	NS	NS
b) Single Count MA-GE	NS	**	***	*	NS	NS
c) Single Count Focal	NS	NS	NS	NS	-	**
d) Single Count Filler	NS	**	NS	**	-	**
e) Single Count MI	NS	**	NS	***	-	NS
f) Zone Count MA	NS	NS	NS	***	NS	***
g) Zone Count MA-GE	NS	**	**	**	NS	NS
h) Zone Count Focal	NS	NS	NS	NS	-	NS
i) Zone Count Filler	NS	**	***	**	NS	***
j) Zone Count MI	NS	***	NS	**	NS	**

Annotation: NS denotes difference is not significant, * difference is statistically significant at 1%-level, ** difference is significant at 5%-level, *** difference is significant at 10%-level

4.4.2. Robustness Check Based on Other Halo Measures

Status Bar (Hypothesis 1 a & b)

Inter-Item Correlations

The analysis of inter-item correlations shows higher average correlations for MA-scales, whereas for MI-scales average correlations are lower for version 1 (Table 19). However, the differences for both scales are not statistically significant.

Table 17 Correlational Analysis Comparison Version 6 & 1

Version	1	6	p-value (one-tailed)
Average Correlation MA	0.68	0.57	0.219
Average Correlation MI	0.36	0.44	0.343

Regression Coefficient

For the MA-scales, the halo coefficient is higher when a status bar is present (version 1), than in the control version (Table 20). In contrary, for the MI-scales the halo effect seems to be stronger for the control version (6). However, differences are not statistically significant for both scales.

Table 18 Regression Coefficients Comparison Version 6 & 1

	Coefficients	1	6	p-value (one-tailed)
MA	Overall Attitude	0.677*	0.561*	0.107
MI	Rating to Previous Item	0.153*	0.182*	0.435

Conclusion

Both the correlational and the regression measure indicate effects in the same direction. For the MA-scales halo effect seems to be increased by the presence of a status bar, whereas it seems to be decreased for MI-scales. As discussed before in the context of drop-out rates, this indicates a non-linear effect of the status bar depending on the position of the scales in the survey. Halo effects are higher for MA-scales which are positioned at the beginning of the questionnaire, and less for MI-scales which are positioned at the end. However, in both measures the observed differences are not statistically significant. Thus, the robustness check confirms the results obtained by the count measure in hypotheses testing (Hypothesis 1 a & b).

Brand/Scale Intermixing (Hypothesis 2 a & b)

Inter-Item Correlations

The correlational analysis indicates a stronger halo effect for the MA-scales when brands are intermixed (version 2). For the MI-scales, average inter-item correlations are slightly lower in version 2, than in control version 6. The differences for both scales are not statistically significant (Table 21).

Table 21 Correlational Analysis Comparison Version 6 & 2

Version	2	6	p-value (one-tailed)
Average Correlation MA	0.70	0.57	0.175
Average Correlation MI	0.41	0.44	0.428

Regression Coefficient

Also, when applying the regression measure the halo effect seems to be slightly stronger when brands are intermixed in the MA-scales (Table 22). However, the difference is not statistically significant. For MI-scales the halo effect is marginally significant lower in version 2 than in version 6 ($p=0.095$).

Table 19 Regression Coefficients Comparison Version 6 & 2

Scale	Coefficients	2	6	p-value (one tailed)
MA	Overall Attitude	0.577*	0.561*	0.437
MI	Rating to Previous Item	0.096*	0.182*	0.095

Conclusion

Both the regression and the correlational measure indicate effects in the same direction. However, a marginally significant difference could only be obtained with the regression measure for MI-scales. Whereas with the count measure a significant halo-reducing effect could be found for both scales, the robustness check only can confirm the reducing effect of intermixing two MI-scales (Hypothesis 2b).

Screen-by-Screen Design(Hypothesis 3 a & b)

Inter-Item Correlations

The correlational analysis indicates that halo effect is stronger for the MA-scales and less for the MI-scales, when scales are presented on single pages. However, both differences are not statistically significant (Table 23).

Table 20 Correlational Analysis Comparison Version 6 & 3

Version	3	6	p-value (one-tailed)
Average Correlation MA	0.66	0.57	0.249
Average Correlation MI	0.38	0.44	0.374

Regression Coefficient

Table 24 shows that halo is stronger for version 3, compared to the control version 6 for MA-scales. For MI-scales the halo effect seems to be reduced when items are presented on single pages. However, the differences between versions are for both scales statistically insignificant.

Table 21 Regression Coefficients Comparison Version 6 & 3

	Coefficient	3	6	p-value (one-tailed)
MA	Overall Attitude	0.612*	0.561*	0.283
MI	Rating to Previous Item	0.136*	0.182*	0.290

Conclusion

Both measures signal effects in the same directions, but the differences to the control version are not statistically significant. This finding confirms the result obtained by the count measure in regard to the MI-scales (Hypothesis 3b). In contrast, for the MA-scales a significant difference to version 3 could be found with the count measure. The robustness check cannot confirm this finding.

Combination of Scale Intermixing and Screen-by-Screen Design (Hypothesis 3 c & d)

Inter-Item Correlations

According to the correlational analysis there are no differences in correlations for MA-scales. A lower halo effect can be observed for MI-scales in version 4, however the difference is not statistically significant (Table 25)

Table 22 Correlational Analysis Comparison Version 6 & 4

Version	4	6	p-value (one-tailed)
Average Correlation MA	0.57	0.57	0.482
Average Correlation MI	0.37	0.44	0.362

Regression Coefficient

A reduced halo effect can be observed for both types of scales when screen-by-screen design is combined with intermixing of brands/scales (version 4) (Table 26). However, the difference is only statistically significant for MI-scales (p=0.031).

Table 23 Regression Coefficients Comparison Version 6 & 4

		4	6	p-value (one-tailed)
MA	Overall Attitude	0.529*	0.561*	0.381
MI	Rating to Previous Item	0.066**	0.182*	0.031

Conclusion

Whereas the correlational measure does not indicate any significant differences, the regression measure shows a reduced halo effect for MI-scales when screen-by-screen design is combined with intermixing of scales. In contrary, with the count measure a significant halo-reducing effect for both scales could be measured. The findings, obtained with the count measure, can therefore only partly be confirmed with the robustness check for the MI-scales (Hypothesis 3d).

Global Evaluation (Hypotheses 4)

Inter-Item Correlations

Table 27shows that halo effect seems to be higher for MA-scales when the global evaluation item is positioned at the end of the scale. However, the difference is not statistically significant.

Table 24 Correlational Analysis Comparison Version 6 & 5

Version	5	6	p-value (one-tailed)
Average Correlation MA	0.69	0.57	0.192

Regression Coefficient

Applying the regression measure, halo is found to be is significantly higher for version 5 compared to the control version 6 (p=0.039) (Table 28).

Table 25 Regression Coefficients Comparison Version 6 & 5

		5	6	p-value (one-tailed)
MA	Overall Attitude	0.713*	0.561*	0.039

Conclusion

The results obtained by the regression measure in contrast to the results obtained with the regression measure and the count measure, which do not indicate an effect of the position of the global evaluation item. The results can therefore only partly be confirmed with the robustness check (Hypothesis 4).

4.4.3. Conclusions on the Robustness Check

To conclude for the count measure, both counting techniques mainly deliver the same results, only differing in the strength of significance, and therefore support the results obtained in the hypotheses testing. In general, which counting method to apply therefore depends how "non-differentiation" is interpreted. Whereas the single-counting technique, by only taking into account identically assigned response categories, is based on a stricter definition, the zone count is a broader measure that also considers nearly identical ratings. As "non-differentiation" not necessarily means to continually stick to the same response category, for the study at hand the zone counting technique was applied. Furthermore, for the multi-attribute scales the general evaluation measure (MA GE) delivers more significant results. This might be attributed to the fact that the scales per brand are quite short as respondents only rate four attributes per brand. It is therefore difficult to recognize a tendency of general non-differentiation to the prior attribute for scales of only four items. However, the MA GE count measure focuses on a special form of halo effect, namely the general impression halo effect, where non-differentiation is not attributed to the preceding attribute, but to overall attitude which is stated before the attributes for each brand. If the ratings to all four attributes are (nearly) identical to the rating to this item for several brands, a tendency can be recognized. Therefore, for the multi-attribute scales the number of nearly identical ratings to the general evaluation item was applied.

The robustness check conducted based on regression and correlational measure delivers less clear findings. The results obtained with the count measure in hypotheses testing for Hypotheses 1 a) & b), 2 b), 3 b) & d), and 4 could be confirmed (at least one measure derived the same result as the count measure). That results are more ambiguous, might reside in the incongruence the correlational measure and the regression measure. Jacobs and Kozlowski (1985) for instance point out that different measures of halo are not comparable. Moreover, Saal et al. (1980) state that the applied halo measure may cause a rating format to appear superior, whereas another measure would lead to another result. It strikes that the correlational measure never indicates any statistically significant differences. This can be mainly attributed to the fact that, as previously criticized, measures that are based on inter-correlations do not measure halo exclusively. In that sense, one cannot be sure whether the correlations are caused by halo effect or other response effects that increase correlations between items. In this case, therefore halo could have been reduced by the treatments, but the correlational measure can be still influenced by other response effects that cause inflated correlations. The regression in contrast delivers some significant differences, and partly confirms results obtained with the count measure. However, the measure can also be criticized in regard to its assumptions, which imply that the content in the versions is the same, and that differences in coefficients can be alone attributed to a change in the degree of halo. If this assumption is not met, the comparison can be biased.

5. Conclusions

In this chapter answers are given to the research questions formulated in Chapter 1. The questions will be answered based on the review of literature and the results of the conducted experiments. Further, managerial and academic contribution, as well as limitations and areas for future research are discussed.

5.1. General Discussion & Research Questions

The problem statement in Chapter 1 pointed out that the main issue surrounding halo effects in consumer surveys where general multi-item- and multi-attribute measures are applied, is the distorting effect of haloing on the results. A consequence is that results can be misleading and wrong strategic decisions can be made (Wirtz 2003; Leuthesser et al. 1995). The objective of this master thesis therefore was to improve the understanding of theory and processes surrounding the halo effect in consumer surveys and to work out implications for marketing research.

The main research question examined was: *How do halo effects affect consumer surveys?* To be able to give a complete answer, it was subdivided into several questions: *What is the halo effect in consumer surveys? Which methods can be used to detect halo effects? How can halo effects be reduced post hoc? How can halo effects be reduced ex ante?*

Most of these questions were discussed in the literature review. The empirical focus of this study was on examining methods to reduce halo effect by altering the survey design. For this purpose an experiment with seven conditions was conducted.

5.1.1. What Is the Halo Effect in Consumer Surveys?

In the literature there is little agreement in regard to the conceptual definition of halo effects. For instance, Fisciaro and Lance (1990) distinguish among three conceptual definitions of halo effect: General Impression Halo Effect, Inadequate Discrimination Halo Effect, and Salient Dimension Halo Effect. Wirtz (1996) adds the Associonist Halo Effect as a variant of the latter. A comparison showed that these definitions are partly overlapping. The definition chosen for this thesis was that the halo effect is the "unwillingness or inability of a respondent to differentiate among several items of a scale", which is in line with the inadequate discrimination halo effect as categorized by Fisciaro and Lance (1990). How this inability or unwillingness actually shapes the respondent's rating strategy depends on the type of scale which is applied.

In consumer surveys constructs are mainly measured with multi-item scales. A distinction was made between general multi-item scales and multi-attribute-attitude scales as a special type of multi-item measures. For multi-attribute scales where beliefs on several independent attributes together explain a respondent's overall attitude towards an object (e.g. brand), the halo effect relates to a non-differentiation between the overall evaluation of an object and attributes of that object. The respondents fail to recognize that attributes should be rated independently and transfer the overall attitude to the attribute ratings. This definition is in line with Fisciaro and Lance (1990)'s general impression halo effect. For general multi-item scales, where a construct is measured as the sum or average of ratings to items, halo effect relates to the non-differentiation between items of a scale. Whereas each item should be rated independently from the preceding, respondents assume dependence among the items and are influenced by their ratings to prior items in the scale.

Moreover the thesis examined cognitive processing that takes place during the response process, which is responsible for the respondent's unwillingness/inability to discriminate among items. Central here is the belief sampling model from Tourangeau et al. (2000), which describes the response process as a complex cognitive process consisting of several stages. If the respondent fails to carry out this process throughout, too less cognitive effort is invested, resulting in haloed ratings. This phenomenon is labeled "satisficing" by Krosnick and Alwin (1987). The reasons for satisficing are for instance lack of knowledge and familiarity on the construct of the scale, or lack of motivation due to a lack of incentives, non-interest in the topic of the survey, or simply boredom and fatigue initiated through too long questionnaires.

In a statistical sense, non-differentiation leads to a reduced variance in the ratings and an excess correlation above natural levels between items. These natural correlations among scale items are referred to as "true halo".

5.1.2. How Can Halo Effects Be Reduced Post-Hoc?

To reduce halo effects post-hoc from the data statistical methods can be applied. Several methods discussed in literature were reviewed. A technique that is based on increased correlations between attributes is to remove the first or common emerged factor of a factor analysis which is assumed to represent overall attitude. The most popular method is to partial-out the halo effect by calculating partial correlation coefficients and to remove halo effect by correcting the raw data. Another technique is to combine partialling-out with principal component and multiple-discriminant analysis (chaining technique). Halo effect is removed by using centroid based distance scores of the resulting discriminant functions. Furthermore, halo effect can also be removed from the data by double centering which implies to standardize raw data and transform it into ipsative data. Another suggested technique is to identify haloing respondents and to separate them with a mixture model. None of the techniques presented was generally accepted yet in literature, since all presented solutions have their disadvantages. For instance, techniques based on partialling-out ignore natural correlations and thus also remove true halo effect, or the double centering technique comes along with statistical problems of ipsative data.

5.1.3. Which Methods Can Be Used to Detect Halo Effects?

Several indices for halo effects were presented in the literature review, as well as applied to the study at hand. A review of literature showed that measures to detect halo effects can mainly be categorized into three classes depending on their foundation. The first type of measures is based on inter-item correlations of a scale. These correlational measures imply calculating average correlation coefficients for the items of a scale, conducting a factor analysis, or partialling-out halo effect. The second type is dispersion measures which are based on the variance in the ratings, and imply calculating average standard deviations or conducting ANOVA to identify Rater x Object interactions. However, the big disadvantage of these measures is the non-availability of true halo levels which serve as a baseline. Therefore to compute the degree of halo effect in the data, these true halo levels have to be known. Literature proposes to assess true levels with expert ratings, perceived similarity scales or by experiments. Unless true halo levels are unknown, the degree of halo present in the data is always a subjective assessment of the researcher. The third type is alternative measures which aim to measure halo through as a regression approach or to count nearly identical and identical ratings to items in a scale.

Inter-item correlations, factor analysis, and the regression measure were applied in this study to demonstrate the presence of a halo effect in the data. As true halo levels were not available, the different measures were compared and a conclusion was drawn. From all three applied measures the same results could be derived. A halo effect was present in the data. It was stronger for the multi-attribute scales than for the general multi-item scales. Furthermore, the dependence of the degree of halo effect of the length of the survey was examined. For this purpose the measures were applied to two questionnaire versions and compared. The results were ambiguous. For the multi-attribute scales only inter-item correlations and regression measures indicated a stronger halo effect, whereas the factor analysis indicated an opposite result (which was not tested for significance). For the MI-scales, two measures showed a higher halo effect, except the correlational measure which did not indicate any effect of questionnaire length. The conformity of the applied measures in regard degree of halo effect shows that traditional measures and alternative measures derive the same results. To apply different measures and to compare their results can be an alternative to assess the level of halo effect, when true halo levels are not available.

However, inter-item correlations and regression measure were also applied to control the results obtained from hypothesis testing with the count measure. Here, the findings were more ambiguous and thus indicate an incongruence of the halo measures. It could be found that both correlational measure and regression measure are inappropriate for hypothesis testing. Whereas the correlational measure does not measure halo effect exclusively and also reflects other response bias that cause inflated correlations, the regression measure has the underlying assumption that content between two compared versions is the same, and that differences between the coefficients can be attributed to halo effects alone. However, other response effects such as context and/or content-carry over effects which occur apart from halo effect, can weaken the assumption.

Points of criticism regarding the correlational and the regression measure support the use of a count measure to measure halo effect, since it is the only measure that takes into account the occurrence of halo effect in the ratings (assignment of (nearly) identical ratings to subsequent items of a scale). Furthermore, the count measure has proven as robust towards different counting techniques, and can therefore be assumed to indicate halo effect reliable.

5.1.4. How Can Halo Effects Be Reduced Ex-Ante?

Existing ex-ante approaches to prevent/reduce halo effect that focus on questionnaire design were reviewed. The solutions proposed in literature could mainly be categorized into four groups: approaches that focus on context, rating procedure, rating features and rating scale.

Context-oriented methods concentrate on time span between consumption and surveying, and survey purpose. Thereafter, the halo effect seems to be stronger with an increasing time lag and if the survey purpose is perceived a performance evaluation. Solutions focus on reducing the time lag and to declare the purpose as developmental. Approaches that focus on the rating procedure, concentrate on the order and specific positioning of items in a scale. The underlying theory is that items can prime respondents by increasing the accessibility of specific beliefs, and in this way initiate non-differentiation. Solutions propose randomizing scale items or intermixing brands items of a scale, or to position a global evaluation item at the end of a multi-attribute scale. Another proposal is to provide instructions at the beginning of the questionnaire to train respondents to rate items independently. Solutions in the group of rating features focus on lack of familiarity, knowledge, interest or involvement associated with the construct being rated. Halo effect was observed to be higher for less familiar, involved and interested respondents. It is suggested to screen respondents upfront and only survey those who are less susceptible to halo. Last, approaches that focus on halo effects that can be attributed to the rating scale concentrate on the specific items of a scale. For instance, halo effect was observed to be stronger for less important and vague attributes and attributes without physical correlates. Solutions propose to involve the respondents in scale construction. Furthermore, it was suggested to increase the number of attributes to be rated to reduce halo effects.

The focus of this study was on how to prevent or reduce the occurrence of halo effects in a questionnaire by ex-ante design-oriented approaches. Hereby the effect of a status bar, scale intermixing, screen-by-screen presentation and positioning of the global evaluation item on halo effect was examined.

Status Bar

In the experiment no effect of a status bar could be proven for both types of scales. Two things might be the cause: First, the halo-reducing effect of a status bar could only unfold at the end of a questionnaire where boredom and tiring are the strongest, and for this effect to occur the questionnaire was not long enough with a length of 7-10 minutes. Or second, the presence of a status bar only has an effect on non-response bias, such as drop-out rates. Comparing the drop-out rate to the control version the number of drop-outs is less, but not significant. However it strikes that compared to all other versions, a much higher percentage of respondents terminated during the multi-attribute part of the questionnaire. This implies that a status bar indeed already has an effect at the beginning of the questionnaire, but not on halo effect.

Intermixing

An effect of scale intermixing for both general multi-item scales and multi-attribute scales could be found. For the multi-attribute scales the intermixing of brands seems to be more effective. A reason for this could be that rating all brands for each attribute at a time initiates a direct comparison among brands on an attribute basis, which reduces the influence of overall attitudes. Furthermore it is more demanding for the respondent to compare brands, which leads to the investment of more cognitive effort and thus less non-differentiation. The findings confirm results of other researchers such as Wilkie et al. (1973). For the general multi-item scales also a halo reducing effect could be proven. The intermixing of the items of two distinct scales could motivate the respondent though variation in constructs and lead him/her to invest more cognitive effort, which in turn results in less non-differentiation. The findings of other researchers (e.g. De Jong et al. 2012, Bradlow & Fitzsimons 2001) could be confirmed.

Screen-by-Screen

For multi-attribute scales, where the overall evaluation of a brand was separated from its attribute ratings through a page break, a halo reducing effect was proven. This might be attributed to the fact that the accessibility of overall attitude is less when separated through page break, since respondents are not able to refer back to the item. This could have the effect that attributes are more independently and are more compared to each other instead of matched to overall evaluation for a given brand. For the multi-item scales, presenting items in groups of five instead of ten did not prove effective. An insignificant effect could be observed, however probably the size of item clusters was still too high to reduce non-differentiation.

Intermixing & Screen-by-Screen

For both general multi-item scales and multi-attribute scales the combination of scale intermixing and screen-by-screen presentation proved to be effective in reducing the halo effect. For the multi-attribute scales, the halo reducing effect seems stronger compared to the main effects. First stating overall attitude for all brands at once and then rating all brands on attributes on single pages, combines the effect of directly comparing brands on attributes and not being able to directly referring back to the overall evaluation. In general multi-item scales the combination also seems to reduce halo effect. However, the effect is only slightly (insignificant) stronger than scale intermixing alone. So no added effect of displaying the intermixed items in groups of five could be shown.

Positioning of the Global Evaluation Item

No significant effects could be found for the positioning of the global evaluation item in multi-attribute scales. A lack of effect could be due to the fact that the overall attitude was positioned after the attribute ratings, but still on the same page, which prevents an independent assessment of the attributes. This finding contributes to the disagreement in literature where no clear effect of the positioning of the global evaluation item could yet be proven (e.g. Wu & Petroshius 1987).

5.2. Academic Contribution

In the past little attention was paid to halo effects in marketing research and consumer surveys, and it was neglected compared to other systematic response effects such as social desirability. Mostly the halo effect has been studied in the psychological and organizational field, and the least in the marketing field. Furthermore most studies examine the occurrence of halo effects for constructs which are measured on an attribute-level, such as product and brand evaluation or satisfaction measurement. However, research on halo effects in general multi-item scales, which differ in regard to the occurrence of halo effects from multi-attribute scales, is rare. The aim of this study was therefore to increase the understanding of halo effect in consumer surveys for both types of multi-item scales.

For this purpose this thesis provides an extensive compilation of types of response effects, cognitive processes which explain halo effects, a review of definitions of halo effect, an overview of different measurement, and statistic- and design-oriented techniques to reduce halo effects in consumer surveys.

Furthermore, different halo measures were applied and compared to each other, and the results showed that from all measures the same conclusions could be derived. These results shed a positive light on different halo measures which were criticized because of their incomparability in other studies (e.g. Jacobs and Kozlowski 1985; Saal et al. 1980). Moreover, it was found that multi-attribute scales are stronger affected by halo effects than general multi-item scales. The strength of halo effect therefore also depends on type of scale applied.

The results obtained from examining four different approaches to reduce halo effect by questionnaire design confirm both findings of other research and provide new insight on how halo effect can be reduced in consumer surveys.

5.3. Managerial Implications

For managers seeking to make decisions based on data obtained from consumer surveys, the halo effect is a potential source of risk in regard to faulty decisions (Leuthesser et al.1995). Therefore, in the managerial context, this master thesis aimed to help marketers and market researchers to improve their consumer surveys by extending their knowledge about the effect of halos. This thesis points to the importance of halo effect by demonstrating the consequences of ignoring halo effects in consumer marketing research. Furthermore, a work of reference for companies that conduct consumer research is provided, by presenting an extensive overview of literature surrounding halo effect. Moreover, it can help businesses to improve the design of their studies, by providing instructions and references on how to detect and measure halo effects, and on how to prevent and remove them.

Based on the study of halo effect in this thesis the following recommendations can be given to managers of market research firms or market intelligence departments who conduct surveys to gather information about consumers:

Firstly, it could be shown in this study that halo has a distorting effect on data obtained by both general multi-item scales and multi-attribute scales. Managers therefore should not ignore halo effects when analyzing data obtained by consumer surveys, but apply proposed techniques to detect halo effects, prevent them by adapting survey design, or remove them by applying statistical techniques. In regard to halo detection, as baselines (true halo levels) are mostly not available, managers can face difficulties when applying these measures to their surveys. Based on this study, the count measure can be recommended, since it is easy and fast to apply, and it takes into account the specific occurrence of halo effect in the ratings. To get a feel for the amount of halo that is present in the data, the ratio of nearly identical and identical ratings to the total numbers of items of a scale should be calculated. Although the interpretation of this ratio is always subjective, high ratios (e.g. 70% similar ratings) give an indication of the amount of halo that is present. However, it is important to backup the so obtained results with a control measure, for instance with a correlational measure. If also high inter-item correlations (>0.6) can be found, it can be concluded that a halo effect is present. It should be noticed that correlational measures cannot be recommended to be applied in isolation, since high correlations can also be attributed to other response effects, such as extreme response style.

Secondly, managers should already pay attention to possible halo effects when designing questionnaires, and apply techniques that were tested as successful in reducing halo effect in this and other studies. For instance, when multi-attribute scales are applied to compare between different products or brands, the brand should be intermixed in a way that all brands are rated for one attribute at a time. The dispersing of brands has shown to reduce halo effect. The same applies for general multi-item scales, when individual specific constructs such as involvement are measured. Here, items of the focal scale should be dispersed with items of an unrelated filler scale. An intermixing ratio of 1: 2 has been proven effective in this study.

Thirdly, although an effect of a status bar on halo could not be measured in this study, managers should not conclude that a progress bar should not be implemented in a survey. It can be assumed that the effect of a status bar on halo effect only unfolds in longer surveys, and also has an influence on the degree of other response effects which were not measured in this study.

Fourthly, halo effect could be shown as being dependent from the length of a questionnaire. In this study a questionnaire of 7-10 minutes response time, was compared to a version with a length of 15-20 minutes. The degree of halo seems to increase with the length of the survey. Thus, surveys should be kept as short as possible to prevent a loss of motivation and in this way to reduce halo effects.

5.4. Limitations and Directions for Future Research

The limitations of this study can be seen as relating to the limitations of the applied research method, the applied medium and the sample.

The research method chosen for this study was an experiment. Results of different versions were compared and based on differences conclusions were drawn. A main limitation of this study might therefore lie in the basic assumption that there are no differences in content between the different questionnaire versions, and that found differences can only be attributed to differences in halo effect. Without this assumption a comparison across different versions would not have been possible. However other response effects, such as content carryover, could have biased the results. Future research could develop methods that separate true content from halo effect and in this way increase comparability of different treatments.

A second source for limitations is the applied medium for the experiments which was a survey. Therefore the results obtained might be dependent from specific characteristics of the applied questionnaire. For instance, the findings could be influenced through the position of the scales in the questionnaire. No significant effects were found for the focal scale (10 items) at all when measured alone. A reason could be that the halo effect was not strong enough in the middle part of the questionnaire, and the effect of switching over to another type of scale from the multi-attribute scales was still present. Furthermore, probably the number of items is too small to recognize a tendency of non-differentiation. This is supported by the fact that for the filler scale (20 items) significant effects were found. Additionally this scale was located at the end of the questionnaire, where halo effects can be assumed to be the strongest. In future research the positioning of different types of scales could therefore be randomized. Also, the dependence of halo effect from questionnaire length provides potential for further research. It could be shown that halo effects are partly stronger in a longer questionnaire version, however results are ambiguous. In the applied longer questionnaire version a main limitation can be found in the fact that the additional items were from general multi-item scale. This could explain a stronger effect of questionnaire length on the general multi-item scales. In future research the effect of questionnaire length could be tested by the extending the questionnaire with additional items of both types of scales.

Survey length also probably plays a role in regard to the ineffectiveness of a status bar on the degree of halo effect in this study. The questionnaire might have been too short for a status bar to show a reducing effect. Thus, testing the dependence of the effect of a status bar on the degree of halo effect from the length of a questionnaire and scale position could be a topic of future research. Furthermore, a reason for the finding of ineffectiveness of screen-by-screen design for general-multi item scales might reside in the fact that the difference between presenting items in groups of five instead of ten is too less. Probably the effect would have been stronger if less than five items would have been shown on a screen. Future research could examine the optimal number of items that are grouped on one screen to be effective in regard to the degree of halo effect. Research could even explore whether the extreme case of presenting each item of a scale on a single screen has a decreasing or increasing effect on halo.

Moreover, also the positioning of the global evaluation item subsequent to the attribute ratings has proven to be ineffective. This could mainly be attributed to the fact that both attributes and overall evaluation items were presented on the same page. As the respondents were able to see them at the same time attributes were probably not rated independently. Future research could investigate the effect of combining positioning of the global evaluation at the end with screen-by-screen design and/or brand intermixing.

Another limitation could reside in the sample, which was multi-cultural, but not equally spread across countries/cultures. Haloing could be different for respondents of different origin. Therefore the dependence of halo effect from country/culture of origin of the respondent could be researched in future.

A further interesting topic of future research could be whether the degree of halo effect depends from the applied product category. For this study smartphones were chosen, which is a product where consumers make well-informed purchase decisions and presumably not have very high emotions. However, research already showed that products/brands that evoke strong emotions can lead to stronger halo effects (=affective overtones). This could maybe also be the case for the iPhone, a brand for which many consumers feel strong feelings of passion and even identification. Whether there are differences in the degree of haloing between iPhone users and users of other smartphones could be examined.

Moreover, a surprising finding of this study is the low number of drop-outs in the questionnaire version where the positioning of the global evaluation item in multi-attribute scales was tested. An explanation could not be given based on a review of point of drop-out or halo measures. In future research it could be examined whether this effect occurs regularly, and if - why it does occur.

Last but not least, of interest for future research is also to examine variations of the intermixing technique of items in multi-item scales. In this study, a mixing ratio of 2:1 was chosen, however halo effect could be further reduced if items of the focal scale were stronger dispersed.

Appendix

Appendix 1 Overview of Response Effects Occurring in Multi-Item-Scales

Social Desirability Bias (SDB) is the tendency of participants to respond to items in a socially desirable way to present themselves favorably according to cultural norms *(*Mick 1996; Crowne & Marlowe 1964; Krosnick 1999). In consumer surveys this is reflected by an over-reporting of favorable attitudes and an under-reporting of unfavorable attitudes (Baumgartner & Steenkamp 2001, 2006*)*. SDB is known to especially occur in questions on sensitive topics such as income or education (Baumgartner & Steenkamp 2006) and can obscure the relationship among different variables by increasing/decreasing correlations (Podsakoff et al. 2003). Most prominent measures are the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe 1960) and the Balanced Inventory of Desirability Responding (Paulhus 1991). Design-oriented attempts (ex ante) to control for SDB include for example the assurance of anonym data processing, the placement of sensitive items at the end of the questionnaire to prevent carry-over effects to subsequent questions, and the change of the wording of sensitive question to a more generalized way that involves surveying respondents about their general opinion on the topic. Statistically-oriented attempts (post-hoc) to control for SDB involve for instance the incorporation of an SDB-scale in the survey, which allows regressing individual responses on the focal measured construct, and provides information on the individuals generalized tendency to engage in social desirable responding (Baumgartner & Steenkamp 2006).

Acquiescence Bias (AB), also referred to as Yea-Saying or Nay-Saying, is the tendency for people to respond consistently positive or negative to survey items irrespective of the question content (Spector 1992; Krosnick 1999; Baumgartner & Steenkamp 2001; Greenleaf 1992). In literature AB is attributed to factors such as lack of involvement, distraction, time pressure, and vague scale items and leads to deflated or inflated correlations (Viswanathan 2005; Spector 1992; Baumgartner & Steenkamp 2005). AB can be measured by assessing the ratio of (dis-)agreement with heterogeneous items from multiple scales or the ratio of agreement to the items of a balanced scale, where equal number of items is worded negatively and positively. Advantage of the application of balanced scales is that they have a build-in control for AB, since the bias cancels out (Baumgartner & Steenkamp 2001, 2006). Statistic-oriented attempts include the partialling-out of variance caused by AB, on the basis of a measure, such as a scale for the tendency of acquiescence responding or a score based on the participants (dis-) agreement to heterogeneous items (Baumgartner & Steenkamp 2006).

Extreme Response Style (ERS) can be defined as the tendency to respond to items with only the most extreme response categories (positive and negative) regardless of the item content (Greenleaf 1992).

Leniency Bias (LB) is the tendency of respondents to rate objects such as brands which they are familiar with higher than warranted (Baumgartner & Steenkamp 2006; Podsakoff et al. 2003; Guilford 1954). In consumer surveys this is a problem when comparing scores of different brands, since correlations of the well-known brand will be artificially increased (Podsakoff et al. 2003).

Midpoint Responding (MR) is the tendency to use the middle scale category irrespective of the content of the item and can be caused by factors such as indifference or indecision. MR causes correlational systematic error if the sample mean is different from the middle category, and is measured by assessing the ratio or number of items that are affected by MR (Viswanathan 2005; Baumgartner & Steenkamp 2001, 2006). Design-oriented attempts to avoid MR include the replacement of the middle category with a “don’t know” category. Statistic-oriented attempts focus on partialling-out the caused variance with an above mentioned measure of MR (Baumgartner & Steenkamp 2006).

State Dependence (SD) is the tendency of respondents to stay consistent with a rating given to previous items and to transfer them to subsequent items, independent from item content. De Jong et al (2012) attribute SD to a reduced cognitive effort of respondents resulting in the repeated choice of the same response categories for several items (Rindfleisch et al. 2008; De Jong et al. 2012). To obtain the degree of SD, researchers can count the number of same responses in a scale. Design-oriented attempts concentrate on the intermixing of the focal multi-item scale with unrelated filler items (De Jong et al. 2012).

Appendix 2 Questionnaire

Questionnaire

I. Introduction ***(all versions)***

Thank you for taking the time to complete this survey about smartphones. This study is conducted within the framework of a master thesis at Erasmus University Rotterdam and aims to find out more about the customers of smartphones.

II: Demographics ***(all versions)***

Q1. What is your gender?

□ Male □ Female

Q2. What is your age? *(Open Question)*

Q3. What is the highest level of education you have completed?

□ No formal education

□ Elementary: 1-4 years

□ High school: 4-9 years

□ Higher Education: 10-13 years

□ University (Bachelor/Master/PhD)

Q4. Do you use a smart-phone as your primary mobile phone?

□ Yes □ No

Page Break

III: Brand Attitude/Beliefs

In this part of the survey we would like to know your opinion towards several attributes of smartphones.

Q5. How important are the following smartphone attributes for you?

Performance	□ Not at all important □ Not that important □ Neutral □ Quite important □ Very Important
Ease of operation	□ Not at all important □ Not that important □ Neutral □ Quite important □ Very Important
Features	□ Not at all important □ Not that important □ Neutral □ Quite important □ Very Important
Physical Design	□ Not at all important □ Not that important □ Neutral □ Quite important □ Very Important

Q6. What brand of smartphone do you use? (select your primary mobile phone if you have more than one)

□ iPhone □ Blackberry □ HTC □ Nokia □ Samsung □ Other □ None

Page Break

VERSION 1: Statusbar ***(a status bar is displayed along the whole survey).***

Please rate the following smartphone brand on the displayed attributes:

Q7. How do you rate the iPhone in general?

iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q8. How do you rate the iPhone in regard to the following attributes?

iPhone	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q9. How do you rate the Blackberry smartphones in general?

Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q10. How do you rate Blackberry smartphones in regard to the following attributes?

Blackberry	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q11. How do you rate HTC smartphones in general?

HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q12. How do you rate HTC smartphones in regard to the following attributes?

HTC	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q13. How do you rate Nokia smartphones in general?

Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q14. How do you rate Nokia smartphones in regard to the following attributes?

Nokia	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q15. How do you rate Samsung smartphones in general?

Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q16. How do you rate Samsung smartphones in regard to the following attributes?

Samsung	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

IV. Involvement (Lastovicka and Gardner 1979)

Please state whether you agree or disagree with the following statements:

Q17. A smartphone is a product that I could talk about for a long time.

□Strongly Disagree □ □ □ □ Strongly Agree

Q18. I understand the features of smartphones well enough to evaluate the brands.

Q19. A smartphone is a product that interests me.

Q20. I have a preference for one or more brands in the smartphone product class.

Q21. A smartphone is a product for which I have no need whatsoever.

Q22. I am not at all familiar with smartphones.

Q23. If I had made a brand choice in the smartphone product class before actually making the purchase, I might easily change my intended choice upon receiving discrepant information.

Q24. My use of a smartphone allows others to see me as I would ideally like them to see me.

Q25. A smartphone helps me attain the type of life I strive for.

Q26. I can make many connections or associations between experiences in my life and the smartphone.

Page Break

V. Use Innovativeness (Price & Ridgway 1983)

Q27. I never throw something away that I might use late.

□Strongly Disagree □ □ □ □ Strongly Agree

Q28. I save broke appliances because I might be able to use parts from them.

Q29. I am less interested in the appearance of an item than in what makes it tick.

Q30. As a child, I really enjoyed taking things apart and putting them back together again.

Q31. I often make gifts instead of buying them.

Q32. Curiosity is one of the permanent and certain characteristics of a vigorous intellect.

Q33. I am very curious about how thing work.

Q34. If I can't figure out how something works, I would rather tinker with it than ask for help.

Q35. I like to build things for my home.

Q36. I never take anything apart because I know I'll never be able to put it back together again.

Page Break

Q37. I like to fix things around the house.

Q38. I have gotten instruction in self-reliance skills (eg. carpentry, car tune-up, etc.).

Q39. I would rather fix something myself than take it to someone to fix.

Q40. When I try to do projects on my own, I'm afraid I will make a worse mess of them than if I had just left them alone.

Q41. Even if I don't have the right tool for the job, I can usually improvise.

Q42. When I try to do projects on my own, without extra directions, they usually work out really well.

Q43. I enjoy thinking of new ways to use old things around the house.

Q44. I often buy a food item for a particular recipe but end up using it for something else.

Q45. I'm uncomfortable working on projects different from types I'm accustomed to.

Q46. I always follow manufacturer's warnings regarding how to use a product.

Page Break

VERSION 2: Scale Intermixing

Q7. How do you rate these smartphone brands in general?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q8. How do you rate these smartphone brands in regard to their Performance?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q9. How do you rate these smartphone brands in regard to their Ease of Operation?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q10. How do you rate these smartphone brands in regard to their Features?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q11. How do you rate these smartphone brands in regard to their Physical Design?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Intermixing: Innovativeness/Involvement

Please state whether you agree or disagree with the following statements:

Q12. A smartphone is a product that I could talk about for a long time.

□Strongly Disagree □ □ □ □ Strongly Agree

Q13. I never throw something away that I might use late.

Q14. I save broke appliances because I might be able to use parts from them.

Q15. I understand the features of smartphones well enough to evaluate the brands.

Q16. I am less interested in the appearance of an item than in what makes it tick.

Q17. 4. As a child, I really enjoyed taking things apart and putting them back together again.

Q18. A smartphone is a product that interests me.

Q19. I often make gifts instead of buying them.

Q20. Curiosity is one of the permanent and certain characteristics of a vigorous intellect.

Q21. I have a preference for one or more brands in the smartphone product class.

Page Break

Q22. I am very curious about how thing work.

Q23. If I can't figure out how something works, I would rather tinker with it than ask for help.

Q24. A smartphone is a product for which I have no need whatsoever.

Q25. I like to build things for my home.

Q26. I never take anything apart because I know I'll never be able to put it back together again.

Q27. I am not at all familiar with smartphones.

Q28. I like to fix things around the house.

Q29. I have gotten instruction in self-reliance skills (eg. carpentry, car tune-up, etc.).

Q30. If I had made a brand choice in the smartphone product class before actually making the purchase, I might easily change my intended choice upon receiving discrepant information.

Q31. I would rather fix something myself than take it to someone to fix.

Page Break

Q32. When I try to do projects on my own, I'm afraid I will make a worse mess of them than if I had just left them alone.

Q33. My use of a smartphone allows others to see me as I would ideally like them to see me.

Q34. Even if I don't have the right tool for the job, I can usually improvise.

Q35. When I try to do projects on my own, without extra directions, they usually work out really well.

Q36. A smartphone helps me attain the type of life I strive for.

Q37. I enjoy thinking of new ways to use old things around the house.

Q38. I often buy a food item for a particular recipe but end up using it for something else.

Q39. I can make many connections or associations between experiences in my life and the smartphone.

Q40. I'm uncomfortable working on projects different from types I'm accustomed to.

Q41. I always follow manufacturer's warnings regarding how to use a product.

Page Break

VERSION 3: Screen-by-Screen

Please rate the following smartphone brand on the displayed attributes:

Q7. How do you rate the iPhone in general?

iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q8. How do you rate the iPhone in regard to the following attributes?

iPhone	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q9. How do you rate the Blackberry smartphones in general?

Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q10. How do you rate Blackberry smartphones in regard to the following attributes?

Blackberry	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q11. How do you rate HTC smartphones in general?

HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q12. How do you rate HTC smartphones in regard to the following attributes?

HTC	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q13. How do you rate Nokia smartphones in general?

Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q14. How do you rate Nokia smartphones in regard to the following attributes?

Nokia	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q15. How do you rate Samsung smartphones in general?

Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q16. How do you rate Samsung smartphones in regard to the following attributes?

Samsung	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

IV. Involvement (Lastovicka and Gardner 1979)

Please state whether you agree or disagree with the following statements:

Q17. A smartphone is a product that I could talk about for a long time.

□Strongly Disagree □ □ □ □ Strongly Agree

Q18. I understand the features of smartphones well enough to evaluate the brands.

Q19. A smartphone is a product that interests me.

Q20. I have a preference for one or more brands in the smartphone product class.

Q21. A smartphone is a product for which I have no need whatsoever.

Page Break

Q22. I am not at all familiar with smartphones.

Q23. If I had made a brand choice in the smartphone product class before actually making the purchase, I might easily change my intended choice upon receiving discrepant information.

Q24. My use of a smartphone allows others to see me as I would ideally like them to see me.

Q25. A smartphone helps me attain the type of life I strive for.

Q26. I can make many connections or associations between experiences in my life and the smartphone.

Page Break

V. Use Innovativeness (Price & Ridgway 1983)

Q27. I never throw something away that I might use late.

□Strongly Disagree □ □ □ □ Strongly Agree

Q28. I save broke appliances because I might be able to use parts from them.

Q29. I am less interested in the appearance of an item than in what makes it tick.

Q30. As a child, I really enjoyed taking things apart and putting them back together again.

Q31. I often make gifts instead of buying them.

Page Break

Q32. Curiosity is one of the permanent and certain characteristics of a vigorous intellect.

Q33. I am very curious about how thing work.

Q34. If I can't figure out how something works, I would rather tinker with it than ask for help.

Q35. I like to build things for my home.

Q36. I never take anything apart because I know I'll never be able to put it back together again.

Page Break

Q37. I like to fix things around the house.

Q38. I have gotten instruction in self-reliance skills (eg. carpentry, car tune-up, etc.).

Q39. I would rather fix something myself than take it to someone to fix.

Q40. When I try to do projects on my own, I'm afraid I will make a worse mess of them than if I had just left them alone.

Q41. Even if I don't have the right tool for the job, I can usually improvise.

Page Break

Q42. When I try to do projects on my own, without extra directions, they usually work out really well.

Q43. I enjoy thinking of new ways to use old things around the house.

Q44. I often buy a food item for a particular recipe but end up using it for something else.

Q45. I'm uncomfortable working on projects different from types I'm accustomed to.

Q46. I always follow manufacturer's warnings regarding how to use a product.

Page Break

VERSION 4: Scale Intermixing & Screen-by-Screen

Q7. How do you rate these smartphone brands **in general**?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q8. How do you rate these smartphone brands in regard to their **Performance**?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q9. How do you rate these smartphone brands in regard to their **Ease of Operation**?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q10. How do you rate these smartphone brands in regard to their **Features**?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Q11. How do you rate these smartphone brands in regard to their **Physical Design**?

Brand	Rating
iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

Please state whether you agree or disagree with the following statements:

Q12. A smartphone is a product that I could talk about for a long time.

□Strongly Disagree □ □ □ □ Strongly Agree

Q13. I never throw something away that I might use later.

Q14. I save broke appliances because I might be able to use parts from them.

Q15. I understand the features of smartphones well enough to evaluate the brands.

Q16. I am less interested in the appearance of an item than in what makes it tick.

Page Break

Q17. 4. As a child, I really enjoyed taking things apart and putting them back together again.

Q18. A smartphone is a product that interests me.

Q19. I often make gifts instead of buying them.

Q20. Curiosity is one of the permanent and certain characteristics of a vigorous intellect.

Q21. I have a preference for one or more brands in the smartphone product class.

Page Break

Q22. I am very curious about how thing work.

Q23. If I can't figure out how something works, I would rather tinker with it than ask for help.

Q24. A smartphone is a product for which I have no need whatsoever.

Q25. I like to build things for my home.

Q26. I never take anything apart because I know I'll never be able to put it back together again.

Page Break

Q27. I am not at all familiar with smartphones.

Q28. I like to fix things around the house.

Q29. I have gotten instruction in self-reliance skills (eg. carpentry, car tune-up, etc.).

Q30. If I had made a brand choice in the smartphone product class before actually making the purchase, I might easily change my intended choice upon receiving discrepant information.

Q31. I would rather fix something myself than take it to someone to fix.

Page Break

Q32. When I try to do projects on my own, I'm afraid I will make a worse mess of them than if I had just left them alone.

Q33. My use of a smartphone allows others to see me as I would ideally like them to see me.

Q34. Even if I don't have the right tool for the job, I can usually improvise.

Q35. When I try to do projects on my own, without extra directions, they usually work out really well.

Q36. A smartphone helps me attain the type of life I strive for.

Page Break

Q37. I enjoy thinking of new ways to use old things around the house.

Q38. I often buy a food item for a particular recipe but end up using it for something else.

Q39. I can make many connections or associations between experiences in my life and the smartphone.

Q40. I'm uncomfortable working on projects different from types I'm accustomed to.

Q41. I always follow manufacturer's warnings regarding how to use a product.

VERSION 5: Global Evaluation

Please rate the following smartphone brand on the displayed attributes:

Q7. How do you rate the iPhone in regard to the following attributes?

iPhone	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q8. How do you rate the iPhone in general?

iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q9. How do you rate Blackberry smartphones in regard to the following attributes?

Blackberry	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q10. How do you rate the Blackberry smartphones in general?

Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q11. How do you rate HTC smartphones in regard to the following attributes?

HTC	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q12. How do you rate HTC smartphones in general?

HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q13. How do you rate Nokia smartphones in regard to the following attributes?

Nokia	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q14. How do you rate Nokia smartphones in general?

Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q15. How do you rate Samsung smartphones in regard to the following attributes?

Samsung	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q16. How do you rate Samsung smartphones in general?

Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

IV. Involvement (Lastovicka and Gardner 1979)

Please state whether you agree or disagree with the following statements:

Q17. A smartphone is a product that I could talk about for a long time.

□Strongly Disagree □ □ □ □ Strongly Agree

Q18. I understand the features of smartphones well enough to evaluate the brands.

Q19. A smartphone is a product that interests me.

Q20. I have a preference for one or more brands in the smartphone product class.

Q21. A smartphone is a product for which I have no need whatsoever.

Q22. I am not at all familiar with smartphones.

Q23. If I had made a brand choice in the smartphone product class before actually making the purchase, I might easily change my intended choice upon receiving discrepant information.

Q24. My use of a smartphone allows others to see me as I would ideally like them to see me.

Q25. A smartphone helps me attain the type of life I strive for.

Q26. I can make many connections or associations between experiences in my life and the smartphone.

Page Break

V. Use Innovativeness (Price & Ridgway 1983)

Q27. I never throw something away that I might use late.

□Strongly Disagree □ □ □ □ Strongly Agree

Q28. I save broke appliances because I might be able to use parts from them.

Q29. I am less interested in the appearance of an item than in what makes it tick.

Q30. As a child, I really enjoyed taking things apart and putting them back together again.

Q31. I often make gifts instead of buying them.

Q32. Curiosity is one of the permanent and certain characteristics of a vigorous intellect.

Q33. I am very curious about how thing work.

Q34. If I can't figure out how something works, I would rather tinker with it than ask for help.

Q35. I like to build things for my home.

Q36. I never take anything apart because I know I'll never be able to put it back together again.

Page Break

Q37. I like to fix things around the house.

Q38. I have gotten instruction in self-reliance skills (eg. carpentry, car tune-up, etc.).

Q39. I would rather fix something myself than take it to someone to fix.

Q40. When I try to do projects on my own, I'm afraid I will make a worse mess of them than if I had just left them alone.

Q41. Even if I don't have the right tool for the job, I can usually improvise.

Q42. When I try to do projects on my own, without extra directions, they usually work out really well.

Q43. I enjoy thinking of new ways to use old things around the house.

Q44. I often buy a food item for a particular recipe but end up using it for something else.

Q45. I'm uncomfortable working on projects different from types I'm accustomed to.

Q46. I always follow manufacturer's warnings regarding how to use a product.

VERSION 6: Control

Please rate the following smartphone brand on the displayed attributes:

Q7. How do you rate the **iPhone in general?**

iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q8. How do you rate the iPhone in regard to the following attributes?

iPhone	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q9. How do you rate **Blackberry smartphones in general?**

Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q10. How do you rate Blackberry smartphones in regard to the following attributes?

Blackberry	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q11. How do you rate **HTC smartphones in general?**

HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q12. How do you rate HTC smartphones in regard to the following attributes?

HTC	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q13. How do you rate **Nokia smartphones in general?**

Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q14. How do you rate Nokia smartphones in regard to the following attributes?

Nokia	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q15. How do you rate **Samsung smartphones in general?**

Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q16. How do you rate Samsung smartphones in regard to the following attributes?

Samsung	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

IV. Involvement (Lastovicka and Gardner 1979)

Please state whether you agree or disagree with the following statements:

Q17. A smartphone is a product that I could talk about for a long time.

□Strongly Disagree □ □ □ □ Strongly Agree

Q18. I understand the features of smartphones well enough to evaluate the brands.

Q19. A smartphone is a product that interests me.

Q20. I have a preference for one or more brands in the smartphone product class.

Q21. A smartphone is a product for which I have no need whatsoever.

Q22. I am not at all familiar with smartphones.

Q23. If I had made a brand choice in the smartphone product class before actually making the purchase, I might easily change my intended choice upon receiving discrepant information.

Q24. My use of a smartphone allows others to see me as I would ideally like them to see me.

Q25. A smartphone helps me attain the type of life I strive for.

Q26. I can make many connections or associations between experiences in my life and the smartphone.

Page Break

V. Use Innovativeness (Price & Ridgway 1983)

Q27. I never throw something away that I might use later.

□Strongly Disagree □ □ □ □ Strongly Agree

Q28. I save broken appliances because I might be able to use parts from them.

Q29. I am less interested in the appearance of an item than in what makes it tick.

Q30. As a child, I really enjoyed taking things apart and putting them back together again.

Q31. I often make gifts instead of buying them.

Q32. Curiosity is one of the permanent and certain characteristics of a vigorous intellect.

Q33. I am very curious about how thing work.

Q34. If I can't figure out how something works, I would rather tinker with it than ask for help.

Q35. I like to build things for my home.

Q36. I never take anything apart because I know I'll never be able to put it back together again.

Page Break

Q37. I like to fix things around the house.

Q38. I have gotten instruction in self-reliance skills (eg. carpentry, car tune-up, etc.).

Q39. I would rather fix something myself than take it to someone to fix.

Q40. When I try to do projects on my own, I'm afraid I will make a worse mess of them than if I had just left them alone.

Q41. Even if I don't have the right tool for the job, I can usually improvise.

Q42. When I try to do projects on my own, without extra directions, they usually work out really well.

Q43. I enjoy thinking of new ways to use old things around the house.

Q44. I often buy a food item for a particular recipe but end up using it for something else.

Q45. I'm uncomfortable working on projects different from types I'm accustomed to.

Q46. I always follow manufacturer's warnings regarding how to use a product.

Page Break

VERSION 7: Questionnaire Length

Please state whether you agree or disagree with the following statements:

Q7. I reflect about my health a lot.

□Strongly Disagree □ □ □ □ Strongly Agree

Q8. I notice how I feel physically as I go through the day.
Q9. I am constantly examining my health.
Q10. I am generally attentive to my inner feelings about my health.
Q11. I am very involved with my health.
Q12. If a product is on sale, that can be a reason for me to buy it.
Q13. When I buy a brand that's on sale, I feel that I am getting a good deal.
Q14. I have favorite brands, but most of the time I buy the brand that's on sale.
Q15. Compared to most people, I am more likely to buy brands that are on special.
Q16. I am more likely to buy brands on sale.

Page Break

Q17. Some of the most important achievements in life include acquiring material possessions.
Q18. The things I own say a lot about how well I am doing in life.
Q19. I don't place much emphasis on the amount of material objects people own as a sign of success.
Q20. I don't pay much attention to the material objects other people own.
Q21. I like to own things that impress people.
Q22. I admire people who own expensive homes, cars, and clothes.
Q23. The Apple brand reflects who I am.
Q24. I can identify with the Apple brand.
Q25. I feel a personal connection to Apple.
Q26. I use the Apple brand to communicate who I am to other people.
Q27. I think the Apple brand helps me become the type of person I want to be.
Q28. The Apple brand suits me well.
Q29. I see myself as someone who has an active imagination.
Q30. I see myself as someone who gets nervous easily.
Q31. Things used to be better in the good old days
Q32. I am probably more likely to return an unsatisfactory product than most people I know
Q33. I see myself as someone who is talkative.
Q34. I would be willing to stop buying products from companies guilty of polluting the environment, even though it might be inconvenient for me.
Q35. I often discuss environmental issues with my friends.
Q36. I become incensed when I think about the harm being done to plant and animal life by pollution.
Q37. When I think of the ways industries are polluting the environment, I get frustrated and angry.
Q38. I seldom read articles or watch TV-programs on environmental issues.

Page Break

Q39. Technological change will insure a brighter future.
Q40. It is important that I do my job better than others.
Q41. I like to know what brands and products make good impressions on others.
Q42. When buying products, I generally purchase those brands that I think others will approve of.
Q43. I see myself as someone who can be tense.
Q44. I'd rather depend on myself than on others.
Q45. I rely on myself most of the time; I rarely rely on others.
Q46. I often do "my own thing".
Q47. My personal identity, independent of others, is very important to me.
Q48. It is important that others like the products and brands I buy.
Q49. It is my duty to take care of my family, even when I have to sacrifice what I want.
Q50. I see myself as someone who is primarily striving to become the self I "ought" to be – fulfill my duties, responsibilities, and obligations.
Q51. I see myself as someone who is inventive.
Q52. American people should always buy American-made products instead of imports.
Q53. If a co-worker gets a prize, I would feel proud.
Q54. The well-being of my co-workers is important to me
Q55. To me, pleasure is spending time with others.
Q56. I feel good when I cooperate with others.

Page Break

Q57. Competition is the law of nature.
Q58. I felt others would be horrified if they knew of my spending habits.
Q59. I see myself as someone who is emotionally stable, not easily upset.
Q60. If I have any money left at the end of the pay period, I just have to spend it.
Q61. Foreigners should not be allowed to put their products on our markets.
Q62. I would attempt to notify store management if I thought service in a story was particularly bad.
Q63. I see myself as someone who is original, comes up with new ideas.
Q64. I see myself as someone who worries a lot.
Q65. I see myself as someone who generates a lot of enthusiasm.
Q66. I see myself as someone who tends to be quiet.
Q67. I sometimes tell lies if I have to.
Q68. I never cover up my mistakes.
Q69. I always obey laws, even if I am unlikely to get caught.
Q70. I have said something bad about a friend behind his or her back.
Q71. When I hear people talking privately, I avoid listening.
Q72. I have received too much change from a salesperson without telling him or her.
Q73. When I was young I sometimes stole things.
Q74. I have done things that I don't tell other people about.
Q75. I never take things that don't belong to me.

Page Break

Q76. I'd rather do almost anything than return a product to the store.
Q77. It may cost me in the long run but I prefer to support American products.
Q78. It is not right to purchase foreign products.
Q79. Parents and children must stay together as much as possible.
Q80. I frequently imagine how I will achieve my hopes and aspirations.
Q81. It is important to me that I respect the decisions made by my groups.
Q82. If a defective product is inexpensive, I usually keep it rather than put up a fuss or complain.
Q83. I rarely purchase the latest fashion styles until I am sure my friends approve of them.
Q84. I see myself as someone who is reserved.

Page Break

Q85. I see myself as someone who is outgoing, sociable.
Q86. I see myself as someone who does a thorough job.
Q87. I see myself as someone who can be somewhat careless.
Q88. I see myself as someone who tends to be disorganized.
Q89. I see myself as someone who tends to be lazy.
Q90. I see myself as someone who does things efficiently.
Q91. I see myself as someone who makes plans and follows through with them.
Q92. I feel like I have made progress toward being successful in life
Q93. I see myself as someone who prefers work that is routine.
Q94. Products are getting shoddier and shoddier.
Q95. I usually obeyed rules and regulations that were established by my parents.
Q96. I see myself as someone who remains calm in tense situations.
Q97. I see myself as someone who is helpful and unselfish with others.
Q98. I see myself as someone who starts quarrels with others.
Q99.I see myself as someone who can be cold and aloof.
Q100. I see myself as someone who is considerate and kind to almost everyone.
Q101. I see myself as someone who likes to cooperate with others.
Q102. I see myself as someone who is sometimes rude to others.
Q103. When another person does better than I do, I get tense and aroused.
Q104. Only those products that are unavailable in the US should be imported
Q105.We are experiencing a decline in the quality of life.
Q106. I see myself as someone who is relaxed, handles stress well.
Q107. Winning is everything.
Q108. In most ways my life is close to my ideal.
Q109. The conditions of my life are excellent.
Q110. I am satisfied with my life.

Page Break

Please rate the following smartphone brand on the displayed attributes:

Q111. How do you rate the iPhone in general?

iPhone	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q112. How do you rate the iPhone in regard to the following attributes?

iPhone	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q113. How do you rate Blackberry smartphones in general?

Blackberry	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q114. How do you rate Blackberry smartphones in regard to the following attributes?

Blackberry	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q115. How do you rate HTC smartphones in general?

HTC	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q116. How do you rate HTC smartphones in regard to the following attributes?

HTC	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q117. How do you rate Nokia smartphones in general?

Nokia	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q118. How do you rate Nokia smartphones in regard to the following attributes?

Nokia	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q119. How do you rate Samsung smartphones in general?

Samsung	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Q120. How do you rate Samsung smartphones in regard to the following attributes?

Samsung	Rating
Performance	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Ease of Operation	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Features	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably
Physical Design	□ Not at all favorably □ Less favorably □ Neutral □ Favorably □ Very Favorably

Page Break

IV. Involvement (Laurent and Kapferer ... 1985)

Please state whether you agree or disagree with the following statements:

Q121. A smartphone is a product that I could talk about for a long time.

□Strongly Disagree □ □ □ □ Strongly Agree

Q122. I understand the features of smartphones well enough to evaluate the brands.

Q123. A smartphone is a product that interests me.

Q124. I have a preference for one or more brands in the smartphone product class.

Q125. A smartphone is a product for which I have no need whatsoever.

Q126. I am not at all familiar with smartphones.

Q127. If I had made a brand choice in the smartphone product class before actually making the purchase, I might easily change my intended choice upon receiving discrepant information.

Q128. My use of a smartphone allows others to see me as I would ideally like them to see me.

Q129. A smartphone helps me attain the type of life I strive for.

Q130. I can make many connections or associations between experiences in my life and the smartphone.

Page Break

V. Use Innovativeness (Price & Ridgway 1983)

Q131. I never throw something away that I might use later.

□Strongly Disagree □ □ □ □ Strongly Agree

Q132. I save broken appliances because I might be able to use parts from them.

Q133. I am less interested in the appearance of an item than in what makes it tick.

Q134. As a child, I really enjoyed taking things apart and putting them back together again.

Q135. I often make gifts instead of buying them.

Q136. Curiosity is one of the permanent and certain characteristics of a vigorous intellect.

Q137. I am very curious about how thing work.

Q138. If I can't figure out how something works, I would rather tinker with it than ask for help.

Q139. I like to build things for my home.

Q140. I never take anything apart because I know I'll never be able to put it back together again.

Page Break

Q141. I like to fix things around the house.

Q142. I have gotten instruction in self-reliance skills (eg. carpentry, car tune-up, etc.).

Q143. I would rather fix something myself than take it to someone to fix.

Q144. When I try to do projects on my own, I'm afraid I will make a worse mess of them than if I had just left them alone.

Q145. Even if I don't have the right tool for the job, I can usually improvise.

Q146. When I try to do projects on my own, without extra directions, they usually work out really well.

Q147. I enjoy thinking of new ways to use old things around the house.

Q148. I often buy a food item for a particular recipe but end up using it for something else.

Q149. I'm uncomfortable working on projects different from types I'm accustomed to.

Q150. I always follow manufacturer's warnings regarding how to use a product.

Page Break

Reference List

Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology*, 77(6), 975-985.

Banks, C. G. & Murphy, K. R. (1985). Toward narrowing the research-Practice gap in performance appraisal. *Personnel Psychology,* 38(2), 335 – 345.

Bartlett, C. J. (1983). What's the difference between valid and invalid halo? Forced-choice measurement without forcing a choice. *Journal of Applied Psychology*, 68(2), 218-226.

Bass, F. M. & Talarzyk W. W. (1972). An attitude model for the study of brand preference. *Journal of Marketing Research,* 9, 42-96.

Bass, F. M., Pessemier, E. A. & Lehmann, D. R. (1972). An experimental study of relationships between attitudes, brand preference, and choice. *Behavioral Science*, 17(6), 532–541.

Bass F. M. & Wilkie W.L. (1973). A Comparative Analysis of Attitudinal Predictions of Brand Preference. *Journal of Marketing Research*, 10(3), 262-269.

Baumgartner, H. & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13(2), 139-161.

Baumgartner, H. & Steenkamp, J. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, 38(2), 143-156.

Baumgartner, H. & Steenkamp, J. (2006). Response Biases in Marketing Research. In: Grover, R. & Vriens, M. (2006). The Handbook of Marketing Research: Uses, Misuses, and Future Advances. Thousand Oaks: SAGE Publications, Inc.

Bearden, W.O & Netemeyer, R. G. (1999). Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research (Association for Consumer Research Series). SAGE Publications Inc.

Beckwith, N. E. & Lehmann, D. R. (1975). The Importance of Halo Effects in Multi-Attribute Attitude Models, *Journal of Marketing Research* , 12(3), 265-275.

Beckwith, N. E., & Lehmann, D. R. (1976). Halo Effects in Multiattribute Attitude Models: An Appraisal of Some Unresolved Issues. *Journal of Marketing Research*, 12, 418-421.

Beckwith, N. E. & Kubilius, V. U. (1978). Empirical Evidence of halo effects in store image research by estimating true locations. *Advances in Consumer Research*, 5, 485-493.

Beckwith, N. E.; Kassarjian, H. H. & Lehmann, D. R. (1978), Halo Effects in Marketing Research: Review and Prognosis. *Advances in Consumer Research*, 5, 465-467.

Bemmaor, A. C. & Huber, J. C. (1978), Econometric Estimation of Halo Effect: Single vs. Simultaneous Equation Models, *Advances in Consumer Research*, 05, 477-480.

Bergkvist, L., & Rossiter, J. R. (2007). The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs. *Journal of Marketing Research*, 44(2), 175-184.

Berman, J. S., & Kenny, D. A. (1976). Correlational bias in observer ratings. *Journal of Personality and Social Psychology*, 34(2), 263-273.

Bickart, B. A. (1993). Carryover and Backfire Effects in Marketing Research. *Journal of Marketing Research*, 30(1), 52-62.

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, 64(4), 410-421.

Bradlow, E. T. & Fitzsimons, G. J (2001). Subscale Distance and Item Clustering Effects in Self-Administered Surveys: A New Metric. *Journal of Marketing Research* , 38(2), 254-261.

Bueschken, J. , Otter, T. & Allenby, G. M. (2010). Do We Halo or Form? A Bayesian Mixture Model for Customer Satisfaction Data. Fisher College of Business Working Paper No. 1620863.

Churchill, G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, 16(1), 64-73.

Churchill, G. A. Jr. & Surprenant, C. (1982), An Investigation into the Determinants of Customer Satisfaction. *Journal of Marketing Research*, 19, 491 -504.

Cheung, M., W.-L. & Chan W. (2002). Reducing Uniform Response Bias With Ipsative Measurement in Multiple-Group Confirmatory Factor Analysis, Structural Equation Modeling. *A Multidisciplinary Journal*, 9(1), 55-77.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Couper, M. P.; Traugott, M. W. & Lamias, M. J. (2001). Web Survey Design and Administration. *The Public Opinion Quarterly*, 65(2), 230-253.

Cronbach, L. J. (1949). Response Sets and Test Validity. *Educational and Psychological Measurement*, 6(4), 475-494.

Crosby, L., A. & Stephens, N. (1987). Effects of Relationship Marketing on Satisfaction, Retention, and Prices in the Life Insurance Industry. *Journal of Marketing Research*, 24(4), 404-411.

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349-354.

Crowne, D. P. & Marlowe, D. (1964). The approval motive: Studies in evaluative dependence. New York: Wiley.

Cox, E. P. (1980).The Optimal Number of Response Alternatives for a Scale: A Review. *Journal of Marketing Research*, 17(4), 407-422.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90(2), 218-244.

Cohen, J. B. & Ahtola O. T (1971). An expectancy X value analysis oft he relationship between consumer attitudes and behavior. Proceedings of the Second Annual Conference of the Association for Consumer Research. Association for Consumer Research, 344-364

Cohen, J. B., & Houston, M. J. (1972). Cognitive consequences of brand loyalty. *Journal of Marketing Research*, 9(1), 97-97.

Dawes, J. (2008). Do data characteristics change according to the number of points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50, 61–77.

Dillon, W.R; Mulani, N. & Frederick; D. G. (1984). Removing Perceptual Distortions in Product Space Analysis. *Journal of Marketing Research*, 21(2), 184-193.

Erickson, G., M., Johansson, J. K. & Chao, P. (1984). Image Variables in Multi-Attribute Product Evaluations: Country-of-Origin Effects. *Journal of Consumer Research*, 11(2), 694-699.

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

Feldman, J. M.; Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, 73(3), 421-435.

Fishbein, M. (1967). A Behavior Theory Approach to the Relations between Beliefs about an Object and the Attitude toward the Object," in M. Fishbein, ed., Readings in Attitude Theory and Measurement. New York: Wiley, 1967, 389-99.

Fisicaro, S. A. (1988). A reexamination of the relation between halo error and accuracy. *Journal of Applied Psychology*, 73, 239-244.

Fisicaro, S. A. & Lance, C. E. (1990). Implications of Three Causal Models for the Measurement of Halo Error. *Applied Psychological Measurement*, 14(4), 419-429.

Friedman, B. A. & Cornelius, E. T. (1976). Effect of rater participation in scale contruction on the psychometric characteristics of two rating scale formats. *Journal of Applied Psychology*, 61(2), 210-216.

Gal, D. & Rucker, D. D. (2011). Answering the Unasked Question: Response Substitution in Consumer Surveys. *Journal of Marketing Research*, 48(1), 185-195.

Gilbride, T.; Yang, S. & Allenby, G. (2005). Modeling Simultaneity in Survey Data. *Quantitative Marketing and Economics*, 3(4), 311-335.

Greenleaf, E. A. (1992). Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research*, *29*(2), 176-188.

Greenleaf, E. A. (1992). Measuring Extreme Response Style. *Public Opinion Quarterly*, 56(3), 328-351.

Greinöcker A. (2009), Visual Design Effects on Respondents' Behavior in Web-Surveys. An Experimental Comparison of Different Input Controls for Scalar Questions and Visual Analog Scales, and a Software Solution. Dissertation, Universität Innsbruck.

Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York, NY, US: McGraw-Hill.

Guilford, J. P. & Fruchter, B. (1973). Fundamental statistics in psychology and education. McGraw-Hill, New York, 5th ed.

Gujarati, D. (1970). Use of Dummy Variables in Testing for Equality between Sets of Coefficients in Two Linear Regressions: A Note. *The American Statistician*, 24(1), 50-52.

Heeler, R., M., Okechuku, C. & Reid, S. (1979). Attribute Importance: Contrasting Measurements. *Journal of Marketing Research*, 16(1), 60-63.

Heerwegh, D. & Loosveldt, G. (2006). An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys. *Journal of Official Statistics*, 22(2), 191-210.

Holbrook, M., B. (1983). Using a Structural Model of Halo Effect to Assess Perceptual Distortion Due to Affective Overtones. *Journal of Consumer Research*, 10 (2), 247-252.

Holbrook, M. B. & Hubert, J. M. (1975). Multi-Attribute Attitude Models: A Comparative Analysis. *Advances In Consumer Research*, 2(1), 375.

Holbrook, M. B. & Huber, J. (1979). Separating perceptual dimensions from affective overtones: An application to consumer aesthetics. *Journal of Consumer Research*, 5(4).

Holbrook, A. L.; Green, M. C. & Krosnick, J. A. (2003). Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias
Public Opinion Quarterly, 67(1), 79-125.

Holzbach, R. L (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology*, 63(5), 579-588.

Huber, J. & James, B. (1976). The Monetary Worth of Physical Attributes: A Dollarmetric Approach, paper presented at the Seventh Annual Attitude Research Conference, Hilton Head, South Carolina.

Huber, J. & McCann, J. (1982). The Impact of Inferential Beliefs on Product Evaluations. *Journal of Marketing Research*, 19(3), 324-333.

Jacobs, R. & Kozlowski, S. W. (1985). A Closer Look at Halo Error in Performance Ratings, , *The Academy of Management Journal*, 28(1), 201-212.

James, W. L. & Carter, F. S. (1978). Halo effects and location preferences. *Advances in* Consumer *Research*, 5, 474-476.

Jobe, J. B., & Herrmann, D. J. (1996). Implications o f Models o f Survey Cognition for Memory Theory. In D. J. Herrmann, C. McEvoy, C. Herzog, P. Hertel, & M. K. Johnson (Eds.), Basic and Applied Memory Research: Practical Applications. Vol. 2, p.193-205). Mahwah, New Jersey: Erlbaum.

Jong de, M. G.; Lehman, D. R. & Netzer, O. (2012). State-dependence Effects in Surveys. Forthcoming.

Jong de, M. G.; Steenkamp, J-B. E.; Fox, J-P. & Baumgartner, H. (2008). Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation. *Journal of Marketing Research* 45(1), 104-115.

Kavanagh, M. J. & Duffy, J. F. (1978). An extension and field test of the retranslation method for developing rating scales. *Personnel Psychology*, 31(3), 461-470.

King, M., F.& Bruner, G., C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing*, Special Issue: Social Desirability Bias, 17(2), 79–103.

Koltuv, B. (1962). Some Characteristics of Intrajudge Trait Intercorrelations. *Psychological Monographs*, 76.

Kozlowski, S. W. J.; Kirsch, M. P. & Chao, G. T. (1986). Job Knowledge, Ratee Familiarity, Conceptual Similarity and Halo Error: An Exploration. *Journal of Applied Psychology*, 71(1).

Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures In Surveys. *Applied Cognitive Psychology*, 5, 213-236.

Krosnick, J. A. (1999). Survey Research. *Annual Review of Psychology*, 50(1), 537-567.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly*, 51, 201-219.

Lance, C. E. & Woehr D. J. (1986). Statistical control of halo: Clarification from two cognitive models of the performance appraisal process. *Journal of Applied Psychology*. 71(1):45-49.

Landy, F. J., Barnes-Farrell, J. L., Vance, R. J., & Steele, J. W. (1980). Statistical Control of Halo Error in Performance Ratings. *Journal of Applied Psychology*, 65(5), 501-506.

Lastovicka, J. L. & David M. Gardner (1979). Components of Involvement. Attitude Research Plays for High Stakes, Eds. J. C. Maloney and B. Silverman, Chicago: AMA, 53-73.

Lehmann D. R. (1971). Television Show Preference: Application of a Choice Model *Journal of Marketing Research*, 8(1), 47-55.

Leuthesser, L., Kohli, C. S. & Harich, K. R. (1995). Brand equity: the halo effect measure. *European Journal of Marketing*, 29(4), 57-66.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140).

Malhotra, N. & Birks, D. (2007). Marketing Research: An Applied Approach, Financial Times Press, 3eds.

McCann, J.M & Wilkie, W.L (1972). The halo effect and related issues in multi-antribute attitude models: an experiment. Institute for Research in the Behavioral, Economic and Management Sciences (Lafayette); Purdue University. Herman C. Krannert Graduate School of Industrial Administration Paper / Purdue University, 377, 19.

McGuire, W. J. (1966). Attitudes and opinions. *Annual Review of Psychology*, *17*, 475-514.

Mick, D. G. (1996). Are Studies of Dark Side Variables Confounded by Socially Desirable Responding? The Case of Materialism. *Journal of Consumer Research*, 20(2), 106-119.

Moore, W. L. & James, W. L. (1978). A study of the amount of halo in the perceptions of automobiles. *Advances In Consumer Research*, 5(1), 481-484.

Murphy, K.R. (1982). Difficulties in the statistical control of halo. *Journal of Applied Psychology*, 67(2), 161-164.

Murphy, K. R. & Jako, R. (1989). Under what conditions are observed intercorrelations greater or smaller than true intercorrelations? *Journal of Applied Psychology*, 74(5), 827-830.

Murphy, K.R, Reynolds, D.H (1988). Does true halo affect observed halo? *Journal of Applied Psychology*, 73, 235-238.

Murphy, K. R., Jako, R. A. & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology*, 78(2), 218-225.

Nisbett, R. E. & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality & Social Psychology*, 35(4), 250-256.

O'Donovan, D. (1965). Rating extremity: Pathology or meaningfulness? *Psychological Review*, 72(5), 358-372.

Parasuraman, A., Grewal D., & Krishnan, R. (2007). Marketing Research. Boston: Houghton Mifflin.

Paulhus, D. L. (1991). Measures of personality and social psychological attitudes. Measures of personality and social psychological attitudes. Measures of social psychological attitudes, 1, 17-59. San Diego, CA, US: Academic Press Robinson, John P. (Ed); Shaver, Phillip R. (Ed); Wrightsman, Lawrence S. (Ed), (1991). xiv, 753 pp.

Peterson, R. & Wilson, W. (1992). Measuring customer satisfaction: Fact and artifact. *Journal of the Academy of Marketing Science*, 20(1), 61-71.

Podsakoff, P. M., MacKenzie, S.B., Lee, J-Y & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.

Pulakos, E. D.; Schmitt, N. & Ostroff, C. (1986). A Warning About the Use of a Standard Deviation Across Dimensions Within Ratees to Measure Halo
Journal of Applied Psychology, 71(1), 29–32.

Rindfleisch, A., Malter, A., Ganesan, S., & Moorman, C. (2008). Cross-Sectional Versus Longitudinal Survey Research: Concepts, Findings, and Guidelines. *Journal Of Marketing Research*, 45(3), 261-279.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*, 13.19–23.

Rosenberg, M. J. (1956). Cognitive Structure and Attitudinal Affect. *Journal of Abnormal and Social Psychology*, 53, 367-72.

Rosenzweig, P. (2007). The halo effect, and other managerial delusions. *McKinsey Quarterly*, 1(1), 77-85.

Price, L. L. & Ridgway, N. M. (1983). Development of a Scale to Measure Use Innovativeness. *Advances in Consumer Research*, 10, 679-684.

Saal, F. E.; Downey, R. G. & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88(2), 413-442.

Salancik, G. R. & Pfeffer, J. (1977). An Examination of Need-Satisfaction Models of Job Attitudes. *Administrative Science Quarterly*, 22(3), 427-456.

Schuman, J., & Presser, S. (1981). Questions and answers in attitude surveys: Experiments in question form, wording, and context. New York: Academic Press

Schriesheim, C. A., Kinicki, A.J. & Schriesheim, J.F. (1979). The effect of leniency on leader behavior descriptions. *Organizational Behavior and Human Performance*, 23(1), 1-29.

Schwarz, N.; Strack, F. & Mai, H-P. (1991). Assimilation and Contrast Effects in Part-Whole Question Sequences: A Conversational Logic Analysis. *The Public Opinion Quarterly*, 55(1), 3-23.

Smith, P. C. & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47(2), 149-155.

Solomonson, A. L. & Lance, C. E. (1997). Examination of the Relationship Between True Halo and Halo Error in Performance Ratings. *Journal of Applied Psychology*. 82(5):665-674.

Spector, P. E. (1992). Summated Rating Scale Construction: An Introduction. Sage University Papers Series. Quantitative Applications in the Social Sciences; 7(82), Sage Publications, Inc.

Strube, M. (1988). Averaging Correlation Coefficients: Influence of Heterogeneity and Set Size. *Journal of Applied Psychology*. 73(3), 559-568.

Thorndike, E.L. (1920). A constant error on psychological rating. *Journal of Applied Psychology*, IV, 25-29.

Toepoel, V., Das, M., & Van Soest, A. (2009). Design of Web Questionnaires: The Effects of the Number of Items per Screen. *Field Methods*, 21, 200-213.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103(3), 299-314. business.illinois.edu.

Tourangeau, R., Rasinski, K. A., Bradburn, N. & D'Andrade, R. (1989). Carryover Effects in Attitude Surveys. *Public Opinion Quarterly*, 53(4), 495-524.

Tourangeau, R.; Rips, L. J. & Rasinski, K. (2000). The Psychology of Survey Response. Cambridge University Press.

Tourangeau, R.; Couper, M. P. & Conrad, F. (2004). Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly*, 68(3): 368-393.

Van Doorn, J. (2008). Is There a Halo Effect in Satisfaction Formation in Business-to-Business Services? *Journal of Service Research*, 11(2), 124-141.

Viswanathan, M. (2005). Measurement Error and Research Design. Thousand Oaks: Sage Publications.

Weijters, B.; Geuens, M. & Schillewaert, & N. (2008).The Proximity Effect: The Role of Interitem Distance on Reverse-Item Bias. Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 08/546.

Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*, 56, 645–646.

Williams, L. J.; Cote, J. A. & Buckley, M. R (1989). Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology*, 74(3), 462-468.

Willingham, W. W. (1958). Interdependence of successive absolute judgments. *Journal of Applied Psychology*, 42(6), 416-418.

Wilkie W., McCann J. & Reibstein D. (1973). Halo Effects in Brand Belief Measurement: Implications for Attitude Model Development, Proceedings, Fourth Annual Conference, Association for Consumer Research, November, 1973.

Wilkie W. L. & Pessemier E. A. (1973). Issues in Marketing's Use of Multi-Attribute Attitude Models. *Journal of Marketing Research*, 10(4), 428-441.

Wirtz, J (1996). Controlling halo in attribute-specific customer satisfaction measures – towards a conceptual framework. *Asian Journal of Marketing*, 5(1), 41-58.

Wirtz, J. (2000). An examination of the presence, magnitude and impact of halo on consumer satisfaction measures. *Journal of Retailing and Consumer Services*, *7*(2), 89-99.

Wirtz, J. (2001). Improving the measurement of customer satisfaction: A test of three methods to reduce halo. *Managing Service Quality*, 11(2), 99-112.

Wirtz, J. (2003). Halo in customer satisfaction measures: The role of purpose of rating, number of attributes and customer involvement. *Journal of Service Management*, 14(1), 96-119.

Wirtz, J. & Bateson, J. E. G. (1995). An experimental investigation of halo effects in satisfaction measures of service attributes. *Journal of Service Management, 6*(3), 84-84.

Wu B. T. & Petroshius A. M. (1987). The Halo Effect in Store Image Measurement. *Journal of the Academy of Marketing Science*, 15(3), 44-51.