

Master's thesis presented to the Department of Psychology of the University of Basel for the
degree of Master of Science in Psychology

Measuring Trust – Quick and Dirty

Author: Denise Rieser, B.Sc.

Matriculation number: 09-911-959

Correspondence email: denise.rieser@stud.unibas.ch

Examiner: Prof. Dr. Klaus Opwis

Supervisor: Florian Brühlmann, M.Sc.

Center for Cognitive Psychology and Methodology, University of Basel

Submission date: 2.5.2017

Abstract

By now, the majority of existing instruments measuring trustworthiness in an online context are based on Likert scaling (e.g., Flavián, Guinalú, & Gurrea, 2006b; McKnight, Choudhury, & Kacmar, 2002). These however, are limited in applicability when using them in different research contexts. Declarative statements formed typically in Likert scales are most likely tailored to the website measured. Questionnaires using Likert-scaled format mostly require modifications to the declarative statements and may come along with loss in reliability and validity. The aim of this thesis is to develop and validate a trust scale using semantic differential to address the limitation discussed above. In study 1 20 out of 50 potential items haven been selected and reviewed by an expert panel. These items were tested for linguistic and psychological bipolarity. After a phase of item refinement in study 2 the final 14 items for the trust scale remained. All items load to their respective subscales well as shown in the EFA and CFA. The subscales of the trust questionnaire feature high reliability (Cronbach's α between .89 and .95). CFA resulted in a good model fit ($\chi^2(74) = 126.562$; $\chi^2/df = 1.710$, RMSEA = 0.034). Correlations between the trust questionnaire and other scales measuring trust, usability, and visual aesthetics suggest convergent and discriminant validity. Furthermore, results from the second study comparing low and high trust screenshots indicate a predictive validity for the trust scale. Overall, the main advantage of using a trust scale with semantic differential lies in its effortless and versatile utilization.

Measuring Trust – Quick and Dirty

Trust, in general is an important factor in many social interactions, involving uncertainty and dependency. Engagement in electronic commerce is not only characterized by uncertainty, but also by anonymity, lack of control and potential opportunism, making trust a crucial element of e-commerce. Further are online transactions associated with temporal and spatial separation between the participating parties. The lack of trust is one of the most cited reasons for not engaging in e-commerce, involving transactions of financial and personal information via the Internet (Hoffman, Novak, & Peralta, 1999; Lee & Turban, 2001). The importance of initiating, building and maintaining trust between consumers and merchants as key facilitators of successful e-commerce is increasingly being recognized in academic as well as in practitioner communities.

By now several questionnaires have been developed to measure trust (Bhattacharjee, 2002; Cho, 2006; Flavián et al., 2006b; Gefen, 2002; McKnight et al., 2002). However, those questionnaires all bare some limitations when it comes to applicability in research as well as in practice. First, most questionnaires incorporate Likert scales. Therefore the items are most likely tailored to measure trust on a specific website (e.g. legaladvice.com in McKnight et al., 2002) respectively on a website type. Using such a questionnaire on a different website may not be possible due to very specific items which can only be applied to this (type of) website. Further, rephrasing these questions may come along with a loss of reliability and validity. Second, translating longer statements into other languages can be a difficult and a long process which may further impact validity.

The aim of this thesis is to develop a quick and easy to use questionnaire for measuring trust in an online context. To reach this goal, the featured questionnaire makes use of semantic differential to meet the limitations discussed above.

Theoretical Background

In this section, various definition of trust taking different perspective into account are given. Subsequently, trust in online context will be discussed, since this is a relatively new research field. Further the web trust model by (McKnight et al., 2002) and the framework of trust inducing are introduced. Finally, advantages using semantic differential and its construction process are presented at the end of this section.

Trust

Trust is an essential factor, allowing people to act under uncertainty and with the risk of negative consequences (Flavián, Guinalú, & Gurrea, 2006a). The precise operationalization of trust is difficult and by now, there exists multiple definitions in the literature. Various approaches to the definition are provided to give a general overview of trust. The following examples are taken from the fields of psychology, management and marketing.

The focus in psychological literatures has been mainly on interpersonal trust. Psychologist generally agreed that trust is an important concept and vital to personality development (Erikson, 1963), cooperation institution (Deutsch, 1962), and social life (Rotter, 1967). Trust as defined by Erikson (1963) represents the first stage of his model of human development. He proposed that basic trust was a central ingredient in a healthy personality and had a major impact on individual traits. According to Deutsch (1962), “the initiation of cooperation requires trust whenever the individual, by his choice to cooperate, places his fate partly in the hands of others”. A regularly cited definition of interpersonal trust was given by Rotter (1967). He describes trust as “an expectancy held by individuals or groups that the word, promise, verbal, or written statement of another can be relied on” (Rotter, 1967).

In the discipline of management, trust has been studied intensively in organizational contexts. Organizational trust as defined by Driscoll (1979) is “the belief that the decision makers will produce outcomes favorable to the person’s interest without any influence by the person” (Driscoll, 1978). Mayer et al. (1995) identified trust as a control mechanism for enabling employees to work together more productively and effectively. Generally, management researchers believe that trust can enhance business performance (Grabner-Kräuter & Kaluscha, 2003).

In the discipline of marketing, research on trust has been conducted within the context of buyer-seller relationships. Within this context trust can be defined as, “a willingness to rely on an exchange partner in whom one has confidence” (Moorman, Deshpande, & Zaltman, 1993). Moorman et al. (1993) assumed an essential role in establishing and maintaining a long-term relationship between sellers and customers. Doney and Canon (1997) identified characteristics of salespersons, such as expertise, likeability, and similarity to customers, which played a significant role in building trust and strengthening the link between the customer and the supplier firm. Sideshmukh et al. (2002) analyzed consumer reports on behavior of front-line employees and practices and policies of management. They found that competence, benevolence, and problem solving, were related to reports of trust in the businesses. These characteristics described above relates to trust in online context too.

Trust in online context

There are four characteristics of trust, which are generally observed and accepted by researchers. First, there must exist two specific parties in a trusting relationship – a trusting party (trustor) and a party to be trusted (trustee). In an online context, the trustor is a consumer who is browsing an e-commerce website, and the trustee is the merchant that the website represents.

Second, trust involves vulnerability. Trust is only required when the environment is uncertain and risky. High complexity and anonymity is associated with e-commerce. Merchants may behave in an unpredictable manner on the Internet. Third, trust leads to actions, which are mostly comprised of risk-taking behaviors. Consumer trust in online merchants generates two specific actions from the consumer. First, making a purchase online from the merchant and second, providing personal and financial information. Finally, trust is a subjective matter. The level of trust considered sufficient to make transactions online is different for everyone. Additionally, people hold different attitudes toward machines and technology (Wang & Emurian, 2005).

Based on these characteristics of trust Mayer et al. (1995) define trust as “the willingness of a party [trustor] to be vulnerable to the actions of another party [trustee] based on the expectation that the other [trustee] will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party [trustee].”

Web Trust Model

To understand which role trust plays in an online context McKnight et al. (2002) proposed an integrative model (see Figure 1). The web trust model is integrated in the broad framework of the Theory of Reasoned Action (Ajzen & Fishbein, 1975). This theory posits that beliefs lead to attitudes, which lead to behavioral intentions, which then lead to the behavior itself. Applying the Theory of Reasoned Action to the web trust model, it proposes that trusting beliefs lead to trusting intentions, which in turn result in trust-related behaviors. Trust-related behaviors are actions that demonstrate a dependence on a Web vendor, that make one vulnerable to the vendor, or increase one’s risk.

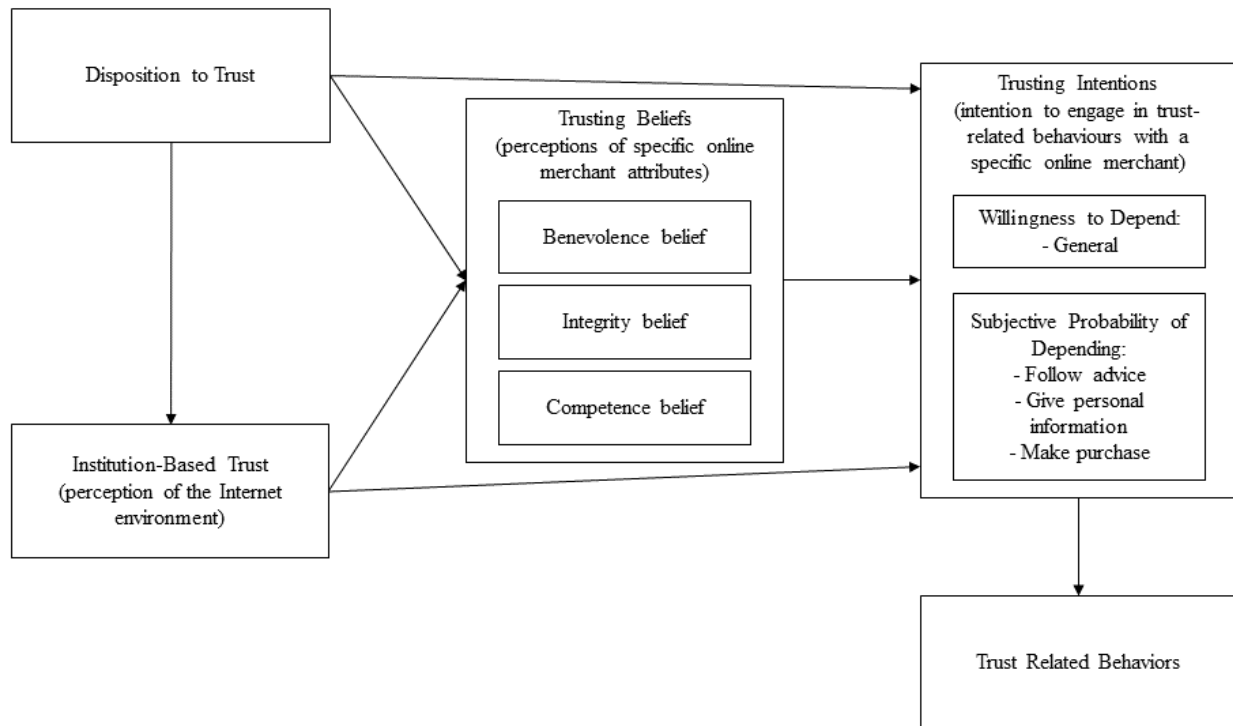


Figure 1. Web Trust Model proposed by McKnight et al. (2002)

Trust-related behavior are actions that increase one's risk or demonstrates dependence on a Web vendor, that make one vulnerable to the vendor. Such behavior in e-commerce include sharing personal information, making a purchase, or acting on information provided by a website. Trusting intentions means the trustor is securely willing to depend, or intends to depend, on the trustee. Trusting beliefs means the consumer's perception that the Web-based vendor has attributes that are beneficial to the consumer (McKnight et al., 2002).

Although there are multiple types of trusting beliefs found in the literature, three trusting beliefs are utilized most often (Bhattacharjee, 2002; Chen & Dhillon, 2003; Flavián et al., 2006b; Gefen, 2002; Mayer et al., 1995; McKnight et al., 2002): benevolence, integrity and competence. Benevolence is related to the user's belief that the other party is interested in his welfare, motivated by a search for a mutually beneficial relationship and without intention of

opportunistic behavior. In other words, the website owner is concerned with the present and future interests, desires and needs of his users and gives useful advice and recommendations. Integrity, sometimes referred as honesty, is the belief that the other party will keep his or her word, fulfill promises, and be sincere. This means, that there are no false statements and the information presented on the Website is sincere and honest. Competence on the other hand, means that the website owner has the resources and capabilities needed for the successful completion of the transaction and the continuance of the relationship (Casaló, Flavián, & Guinalú, 2007).

Trust in Web Design

Designing for trust in technology-mediated interaction is an important key element to build trust between consumers and merchants (Wang & Emurian, 2005). Online merchants depend mostly on their website to attract potential customers and to communicate with them. Therefore, applying trust-inducing features to the website of the online merchant is an important factor of building and enhancing trust. Wang et al. (2005) classified the trust inducing features of web design into four dimensions, namely graphic design, structure design, content design, and social-cue design. Graphic design refers to the graphical design of the website that usually makes up the first impression of the consumer. Structure design defines the overall organization and accessibility of the information displayed on the website. Whereas content design relates to the informational components – either textual or graphical – that can be included on the website. Social-cue design refers to embedding social cues, such as social presence and face-to-face interaction into the website via different communication media. Table 1 gives an overview over factors and characteristics (Seckler, Heinz, Forde, Tuch, & Opwis, 2015; Wang & Emurian, 2005) which can influence the perception of trustworthiness of the website.

Table 1

Overview of the Framework of trust-inducing features (Seckler et al., 2015; Wang & Emurian, 2005).

Dimension	Features
Graphic design	<p>Use of three-dimensional, dynamic, and half-screen size clipart</p> <p>Symmetric use of moderate pastel color of low brightness and cool tone</p> <p>Use of well-chosen, good-shot photographs</p> <p>Visual design (use of colors, site layout, layout complexity, photographs, overall appearance)</p>
Structure design	<p>Implementation of easy-to-use navigation (simplicity, consistency)</p> <p>Use of accessible information (e.g., no broken links and missing pictures)</p> <p>Use of navigation reinforcement (e.g., guides, tutorials, instructions)</p> <p>Application of page design techniques (e.g., white space and margin, strict grouping, visual density)</p> <p>Usability</p> <p>Pop-ups and ads</p> <p>Demands</p>
Content design	<p>Display of brand-prompting information (e.g., prominent company logo or slogan, main selling point)</p> <p>Up-front disclosure of all aspects of the customer relationship (e.g., company competence, security, privacy, financial, or legal concerns)</p> <p>Display of seals of approval or third-party certificate</p> <p>Use of comprehensive, correct, and current product information</p> <p>Use of a relevant domain name</p> <p>Privacy policy</p>
Social-cue design	<p>Inclusion of representative photograph or video clip</p> <p>Use of synchronous communication media (e.g., instant messaging, chat lines, video telephony)</p> <p>Customer service</p> <p>Users' social proof</p> <p>Friends' social proof</p>

Using Semantic Differential

Several questionnaires which measure trust have been developed by now (e.g., Bhattacharjee, 2002; Cho, 2006; Flavián et al., 2006b; Gefen, 2002; McKnight et al., 2002). However, neither of these questionnaires, although well operationalized and validated, are utilized as much as anticipated. This may be due to the fact that these questionnaires bare one common limitation on its applicability in other research studies and practice. All these questionnaires utilize longer sentenced statements in a Likert format, which most likely are tailored to a specific website (Flavián et al., 2006b; McKnight et al., 2002). Reusing these questionnaires without modifications into other research context may be difficult or even impossible. Further, modifications to the original questionnaire may also be accompanied with loss of reliability and validity.

The aim of this thesis is to provide an easy to use questionnaire which addresses the situation above. By using semantic differential as opposed to Likert scale format, it is aimed to have the flexibility to use this questionnaire in different research settings without modifications. The semantic differential makes use of a set of bipolar scales, whereas Likert-scaling demands from respondents to indicate the extent to which they disagree or agree with declarative statements. Each of the bipolar item that make up a semantic differential consists of an antonym pair, which are usually two adjectives. Literature suggests that this format is well suited to measure complex and multidimensional constructs (Verhagen, van Den Hooff, & Meents, 2015). The adoption of semantic differential contains several advantages over other measurement techniques.

First, semantic differentials can reduce the acquiescence bias sometimes provoked by Likert scales. The acquiescence bias refers to a category of response bias in which respondents to

a survey have a tendency to agree with all items or to indicate a positive connotation (Friborg, Martinussen, & Rosenvinge, 2006). Second, it has been demonstrated that semantic differential outperforms Likert-based scaling on robustness (Hawkins, Albaum, & Best, 1974), reliability (Wirtz & Lee, 2003), and validity (Van Auken & Barry, 1995). Similar questionnaires in the field of human-computer-interactions have adopted semantic differential successfully and achieve high reliability and validity (Hassenzahl, Burmester, & Koller, 2003; Laugwitz, Held, & Schrepp, 2008; Verhagen et al., 2015). Third, it has been shown that semantic differential functions effectively as a short-form scale format to reduce survey completion time (Chin, Johnson, & Schwarz, 2008). Finally, the semantic differential offers respondents the opportunity to express their opinion about a concept more fully, that is, ranging from the negative polar to the positive polar. Whereas, Likert-type scale respondents can only indicate the extent to which they disagree or agree. For example, a strongly disagree on such a scale does not automatically imply that a respondent sees the concept as the opposite (Chin et al., 2008).

Framework for Developing and Applying Semantic Differentials

Verhagen et al. (2015) introduced a framework for developing and applying semantic differentials. The construction of the trust questionnaire presented in this thesis follows the suggested framework. By using the framework, it is intended to ensure that the development process of the questionnaire results in a well operationalized and validated trust scale.

Verhagen et al. (2015) divides the construction process into six stages. The first step of the construction process of a scale using semantic differential is to establish a sample of relevant bipolar scales. To achieve the first step, it is recommended to collect existing bipolar scales, convert scales using other formats into bipolar scales, and generate new scales based on literature study, observation, or expert interviews. The next step involves ensuring the linguistic and

psychological bipolarity. Linguistic bipolarity implies that the scale anchors of each semantic differential reflects a contrasting relationship, meaning that, the scale anchors function as grammatical antonyms. Psychological bipolarity assumes that the selected scale anchors are bipolar in relation to the concept to be measured. It is suggested to make use of an expert panel to test and establish linguistic matching of each of the bipolar scales in terms of the concept under the study. In the following step a pilot test of the draft questionnaire is recommended. A pretest helps to ensure clarity and understandability of the bipolar scales, their introductions, and the involved concept(s). The fourth step involves the assessment and establishment of the dimensionality of the semantic differential scale. This can be achieved by applying exploratory factor analysis and confirmatory analysis. Followed by an analysis of reliability, convergent validity, and discriminant validity. The following step includes an analysis of the sensitivity of the semantic differential to anchoring effects between the dimensions and items. For this purpose, a measurement invariance test is suggested. Finally, it is recommended to apply psychometric tests to establish whether the scales do indeed show required psychometric qualities in the final data collection.

Study 1: Item Scale Construction

The goal of study 1 is to find potential candidates of items to validate in a subsequent study. To build up an initial item pool relevant existing trust questionnaires were reviewed. Afterwards, an item-sort task and a test for linguistic and psychological bipolarity was performed by an expert panel in an online survey.

Method

Materials. The first step toward constructing a scale using semantic differential is establishing a sample of relevant bipolar scales. For the initial item pool relevant trust questionnaires (Bart, Shankar, Sultan, & Urban, 2005; Bhattacharjee, 2002; Cho, 2006; Corbitt, Thanasankit, & Yi, 2003; Flavián et al., 2006b; Gefen, 2002; Jian, Bisantz, & Drury, 2000; Koufaris & Hampton-Sosa, 2004; Lu, Wang, & Hayes, 2012; Mayer et al., 1995; McCroskey & Teven, 1999; McKnight et al., 2002; Pavlou & Gefen, 2004; Rieser & Bernhard, 2016) were reviewed. Likert-scaled items taken from the reviewed questionnaires were then converted into bipolar scales. Positive anchored adjectives were complemented with antonyms consulting online thesauri (www.thesaurus.com; www.merriam-webster.com). To enlarge the item pool further synonyms of positive and negative anchored adjectives have been included to the item pool. This process resulted in 28 potential positive anchored adjectives and 47 potential negative anchored adjectives giving a total of 50 potential items taking different adjective combinations into account.

Procedure. The next step in the construction process of the trust questionnaire is to ensure that the items are associated with their respective subscales - namely benevolence, integrity, and competence - and testing the bipolarity of the selected scale anchors. The online survey consisted of three parts. In the first part, the participants were given a brief introduction

and the definition of the three subscales with relevant examples. Then, they were asked to allocate the given positive anchored adjectives to the three subscales. Additionally, to the three subscales they had an option to indicate not to know where to assign the adjective.

In the second part, the task was to rate the bipolarity of the semantic differential. The participants were given for each positive anchored adjective a selection of negative anchored adjectives to choose from. They had also the chance to enter an alternative adjective, if they thought that their suggestion fits better as an antonym. Further, they were given the opportunity to indicate that no adjective fits as antonym. Finally, the proficiency level of the English language as well as the knowledge level in trust related research were measured.

Participants. A total of 18 participants completed the online survey. Recruitment for the online survey took place at the Department of Psychology at the University of Basel. Students and employees from the research group of human computer interaction were either asked personally to participate or invited via e-mail. Participants had to fulfil two requirements to participate in the online survey. First, they need to have at least a basic knowledge in field of human computer interaction and trust research. Second, they should master the English language at least on a conversational level. The participation was fully voluntary and no reimbursement was given.

Results

From the initial item pool 30 items were excluded. The exclusion followed the criteria by Howard et al. (2016), in which they proposed a new statistical significance formula for item-sort tasks. According to Howard et al. (2016) is the critical value for conformity at 13 or 72% with a sample size of 18. Each positive anchored adjective was paired with the negative anchored adjective which achieved the highest agreement. A total of 20 items remained for the subsequent

validation study. Table 2 features an overview of the remained items used in the subsequent study.

Table 2

Items of the trust questionnaire resulted from study 1.

ID	Item
Benevolence	
BEN1	ignoring – caring
BEN2	malicious – benevolent
BEN3	rude – cordial
BEN4	insensitive – sensitive
BEN5	inconsiderate – empathic
Integrity	
INT1	dishonest – honest
INT2	insincere – sincere
INT3	dishonorable – honorable
INT4	unbelievable – believable
INT5	untruthful – truthful
INT6	fraudulent – credible
Competence	
COM1	clueless – knowledgeable
COM2	incompetent – competent
COM3	unskilled – skillful
COM4	unqualified – proficient
COM5	incapable – capable
COM6	uninformed – informed
COM7	inexperienced – experienced
COM8	ineffective – effective
COM9	resourceful - inept

Note. Items shown in bold were retrieved from Rieser & Bernhard (2016)

Study 2: Item Refinement and Scale Validation

Study 2 is aimed towards multiple goals. The next step towards constructing a trust questionnaire was to assess and establish the dimensionality of the concept. Since the trust questionnaire is founded on a theoretical basis, it is intended to replicate the three dimensions - benevolence, integrity, and competence. Further, an analysis of the psychometric properties of the items was performed to reduce THE item number. Finally, an analysis was conducted to assess the general validity of the trust questionnaire.

Method

Participants. A total of 714 participants have finished the online survey successfully. Responses were excluded from the final data set when one of the three following criteria applied. First, if the response time of the participant was under 150 seconds. Second, if a repeated response pattern (e.g. crossing only the middle response option for a specific questionnaire) was detected. Third, if the participant themselves stated not to use the data for the final data analysis. After data exclusion, responses from 601 participants (42% women, 58% men, $M_{age} = 38$ years, age range: 18 - 84) remained. Recruitment took place on Amazon Mechanical Turk. For participation, the participants were reimbursed with 0.60\$. Only workers from Amazon Mechanical Turk living in the United States were eligible to participate in the survey.

Procedure. The online survey is divided into two main parts. Subsequently to the brief overview of the study the participants were asked to give their consent to collect their data for research purposes. In the first part of the survey the participants were given the task to rate a mock online shop based on a screenshot provided. The rating was performed with the trust questionnaire constructed in study 1. The participants were randomly assigned into two groups. The first group were presented a screenshot with a high trustworthy online shop and the second

group on the other hand received a screenshot with a low trustworthy online shop. Both screenshots were presented for at least four seconds. After four seconds the participants were given the option to continue to the next page where they were asked to rate the online shop owner in terms of trustworthiness via the trust questionnaire. The data collected for this part is used to assess if the trust questionnaire can differentiate between high and low trustworthiness.

In the second part of the survey the participants were asked to perform two tasks on a real existing website. Participants were randomly assigned into two groups. The first group received the link to an online shop (www.crazysales.com.au), where they were asked to find a product of their liking and to inform their self about the return policy of the company. The second group were given the link to an website (sunshineloans.com.au), which is specialised in giving small loans. On this website, the participants were asked to inform their self about the loan costs and if security is required to apply for a loan. Afterwards, they were asked to rate the website in terms of trust, usability, and visual aesthetics. Data collection for this part is utilized to assess the dimensionality of the questionnaires as well as analysing the validity of the questionnaire. Finally, general demographics and the propensity to trust were measured.

Materials. For the mock online shop used in the first part of the study two screens were designed using the framework of trust-inducing features (Wang & Emurian, 2005) The first screen represents a high trustworthy online shop (see Figure 2) whereas the second screen presents a low trustworthy online shop (see Figure 3). For each underlying dimension of the framework it was aimed to manipulate at least one aspect. Table 3 gives a detailed overview of the aspects which were manipulated.

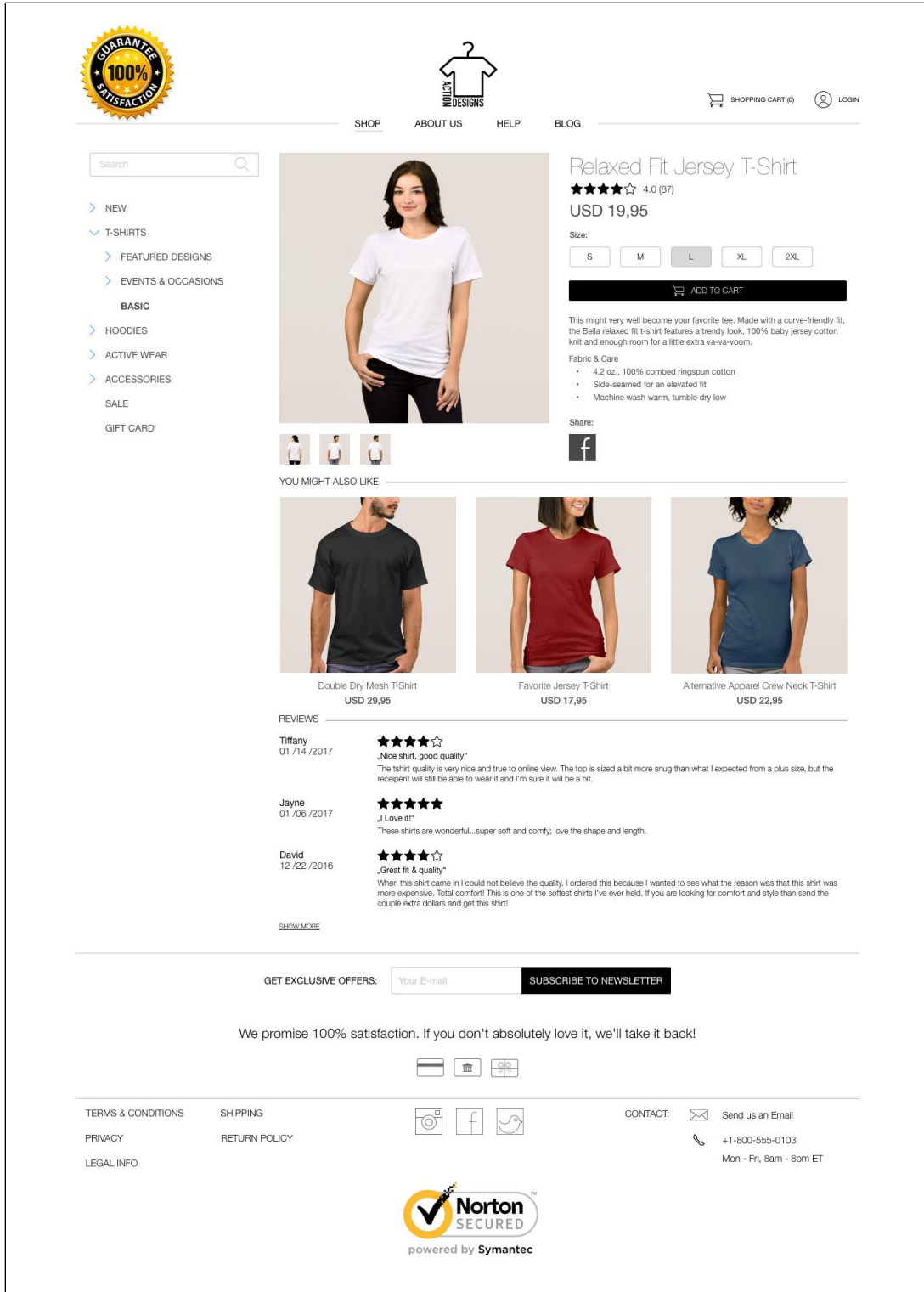


Figure 2. Mock online shop with high trustworthiness.

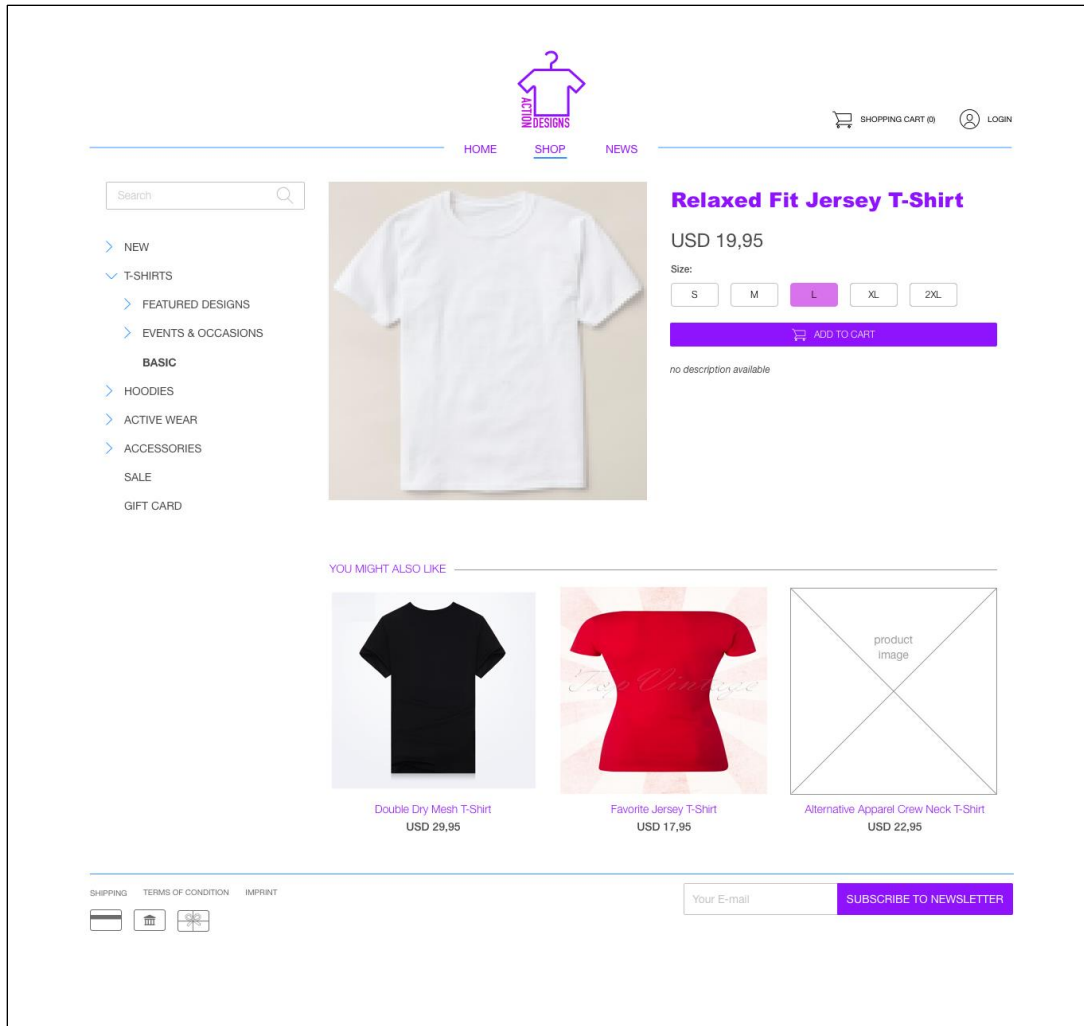


Figure 3. Mock online shop with low trustworthiness.

Table 3

Overview of the dimensions and respective features manipulated in the mock online shop.

Dimension	Features
Graphic design	Symmetric use of moderate pastel color of low brightness and cool tone Use of well-chosen, good-shot photographs Visual design (use of colors, site layout, layout complexity, photographs, overall appearance)
Structure design	Use of accessible information (e.g., no broken links and missing pictures)
Content design	Up-front disclosure of all aspects of the customer relationship (e.g., company competence, security, privacy, financial, or legal concerns) Display of seals of approval or third-party certificate Use of comprehensive, correct, and current product information Use of a relevant domain name Privacy policy
Social-cue design	Customer service Users' social proof

For the second part of the study two different website types were chosen, to assess the measurement invariance of the trust questionnaire. Both websites (www.crazysales.com.au, sunshineloans.com.au; see Figure 4 and Figure XX) were selected considering the website traffic and the website ranking from the United States, since the target audience of the survey came from the United States. Both metrics were taken from www.alexa.com and www.similarweb.com. It was aimed to select a relatively unknown website to prevent any bias from previous experience with the website.

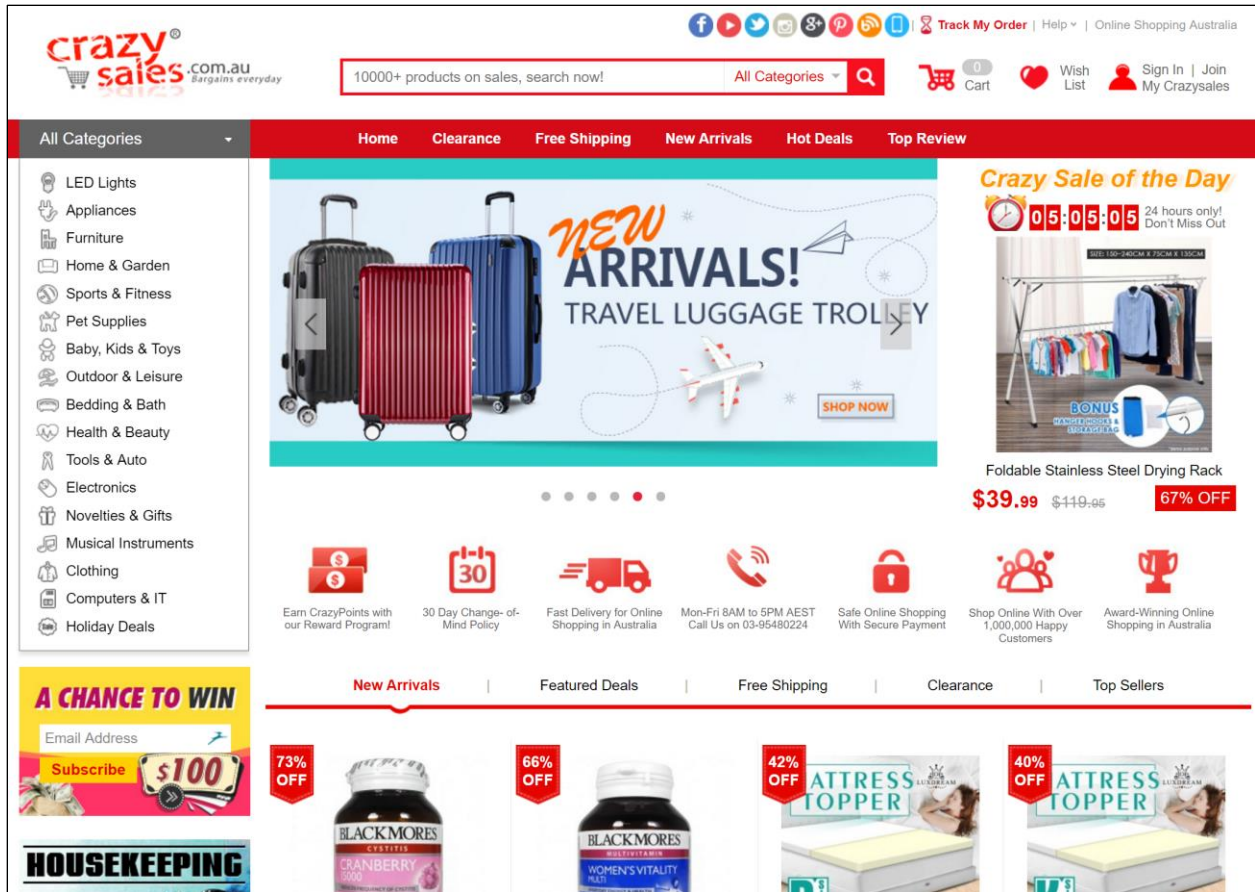


Figure 4. Screenshot from www.crazysales.com.au.

Figure 5. Screenshot from www.sunshineloans.com.au.

To assess convergent validity the trust questionnaire developed by Flavián et al. (2006b) was included to the survey. Slight modifications to the declarative statements of the items were carried out to fit the measured website better. To assess the discriminant validity visual aesthetics using VisAWI (Moshagen & Thielsch, 2010) and usability using a short scale version of the SUS (Brooke, 1996), namely UMUX (Borsci, Federici, Bacci, Gnaldi, & Bartolucci, 2015) were measured. Finally, a short scale measuring the individual propensity to trust (McKnight et al., 2002) was added to the survey. All items from the above-mentioned questionnaires were presented in a random order within their subsection of the survey.

Results

Item refinement. The data set ($N = 601$) of the second study was considered for the analysis. A two-sample Kolmogorov-Smirnov test was performed, to confirm that the data sets from each website don't differ significantly ($D = 0.090$, $p = .169$) from each other in terms of their distribution. The item refinement process took place in three steps. First, the distribution values for each item were analysed (see Table 4). For the competence subscale three items (COM1, COM7, COM8) show a slight negative skew, suggesting an anchoring effect. For this reason, they were excluded from further analysis.

Table 4

Descriptive statistics of the items in the trust questionnaire.

	<i>M (SD)</i>	skew	kurtosis
Benevolence			
BEN1	4.49 (1.24)	-0.04	-0.51
BEN2	4.49 (1.24)	-0.05	-0.28
BEN3	5.08 (1.16)	-0.27	-0.24
BEN4	4.32 (1.20)	-0.03	0.18
BEN5	4.52 (1.22)	-0.11	-0.03
Integrity			
INT1	4.82 (1.36)	-0.46	-0.13
INT2	4.75 (1.30)	-0.45	0.07
INT3	4.62 (1.33)	-0.22	-0.34
INT4	5.08 (1.38)	-0.71	0.18
INT5	4.93 (1.36)	-0.40	-0.36
INT6	5.06 (1.43)	-0.58	-0.26
Competence			
COM1	5.56 (1.17)	-0.91	1.04
COM2	5.51 (1.21)	-0.75	0.33
COM3	5.39 (1.18)	-0.59	0.15
COM4	5.55 (1.19)	-0.70	0.45
COM5	5.48 (1.20)	-0.78	0.58
COM6	5.60 (1.20)	-0.65	0.26
COM7	5.51 (1.22)	-0.89	0.62
COM8	5.51 (1.24)	-0.88	0.66
COM9	5.43 (1.23)	-0.78	0.53

Note. $N = 601$.

Second, an explorative factor analysis was conducted on the 17 remaining items with oblique rotation, since factors were expected to be correlated. The Kaiser-Meyer -Olkin measure verified the sampling adequacy for the analysis, $KMO = .97$ ('marvellous' according to Hutcheson & Sofroniou, 1999), and all KMO values for individual items were greater than .95, which is well above acceptable limit of .5 (Field, 2013). The Bartlett's Test of sphericity, which tests the overall significance of all correlations within the correlation matrix, was significant ($\chi^2(136) = 9633.92, p < .001$), suggesting that using an exploratory factor analysis is appropriate. In an initial analysis of the eigenvalues only two factors had eigenvalues over Kaiser's criterion of 1. However, the parallel analysis suggested three factors which in combination explained 73% of the variance. The exploratory factor analysis was performed with three factors, since this solution is in line with the theoretical model. After the first exploratory analysis, a total of three items (BEN3, INT2, and INT3) were eliminated because they did not contribute to the factor structure and failed to meet the minimum criteria of having a primary factor loading of .4 or above, and no cross-loading of .3 or above. Table 5 shows the factor loadings after the rotation.

Table 5

Results of the second exploratory factor analysis.

Item	Factor loadings with oblique rotation			h^2
	Benevolence	Integrity	Competence	
BEN1: ignoring – caring	.78	.07	.07	.77
BEN2: malicious – benevolent	.62	.18	.05	.63
BEN3: rude – cordial	.44	.08	.31	.53
BEN4: insensitive – sensitive	.85	-.02	-.01	.69
BEN5: inconsiderate – empathic	.86	.00	.02	.76
INT1: dishonest – honest	.13	.86	-.08	.83
INT2: insincere – sincere	.39	.51	.01	.74
INT3: dishonorable – honorable	.46	.44	.04	.76
INT4: unbelievable – believable	.06	.72	.07	.67
INT5: untruthful – truthful	-.03	.75	.15	.71
INT6: fraudulent – credible	-.04	.75	.20	.76
COM2: incompetent – competent	-.04	.12	.83	.79
COM3: unskilled – skillful	.09	-.07	.84	.70
COM4: unqualified – proficient	.00	.05	.83	.76
COM5: incapable – capable	-.06	.11	.84	.79
COM6: uninformed – informed	-.02	.06	.84	.76
COM9: inept - resourceful	.12	-.14	.86	.69
Eigenvalues	3.73	3.74	4.89	
% of variance	22	22	29	

Note. Factor loadings above .3 are marked in bold. $N = 601$.

A second exploratory factor analysis of the remaining 14 items with oblique rotation was conducted, with the three factors explaining 74% of the variance. All items had primary loadings above .5 and load to their corresponding factor. The factor loadings of the final solution are presented in Table 6 and the correlations between the factors are presented in Table 7.

Finally, the reliability of each subscale was analysed. Benevolence ($\alpha = .89$), integrity ($\alpha = .95$), and competence ($\alpha = .93$) achieved a high internal consistency. No substantial increase in Cronbach's alpha for any of the scales could have been achieved by eliminating more items.

Table 6

Results of the second exploratory factor analysis.

Item	Factor loadings with oblique rotation			h^2
	Benevolence	Integrity	Competence	
BEN1: ignoring – caring	.78	.08	.06	.78
BEN2: malicious – benevolent	.60	.18	.05	.61
BEN4: insensitive – sensitive	.83	.00	-.01	.68
BEN5: inconsiderate – empathic	.90	-.02	.01	.79
INT1: dishonest – honest	.14	.88	-.11	.83
INT4: unbelievable – believable	.06	.73	.05	.66
INT5: untruthful – truthful	-.01	.77	.11	.72
INT6: fraudulent – credible	-.02	.76	.18	.77
COM2: incompetent – competent	-.04	.12	.83	.79
COM3: unskilled – skillful	.08	-.06	.83	.70
COM4: unqualified – proficient	.00	.05	.83	.76
COM5: incapable – capable	-.06	.10	.85	.80
COM6: uninformed – informed	-.01	.04	.85	.77
COM9: inept - resourceful	.12	-.14	.87	.69
Eigenvalues	2.80	2.96	4.59	
% of variance	20	21	33	
α	.89	.95	.93	

Note. Factor loadings above .3 are marked in bold. $N = 601$.

Table 7

Correlations between the factors.

Factor	Benevolence	Integrity	Competence
Benevolence	–		
Integrity	.76	–	
Competence	.52	.72	–

Note. $N = 601$.

Validation of the trust scale. A confirmatory factor analysis was performed to assess the validity of the scale. The CFA was conducted on the same sample ($N = 601$) taken from the second part of study 2. To test for multivariate normality Mardia's test was performed. This test is based on multivariate extensions of skewness and kurtosis. The results show that the distribution of the data differs significantly (Mardia's skewness = 24.52, $p < .001$, Mardia's kurtosis = 338.23, $p < .001$) from multivariate normal distribution. For this reason, a maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic was used for the confirmatory factor analysis. For the CFA, the respective items were modelled as reflective indicators of their latent factors (benevolence, integrity, and competence). The results are presented in Figure 6. Goodness-of-fit metrics for CFA typically include χ^2 . However, the χ^2 statistic is sensitive to sample size, meaning that, the probability of model rejection increases with increasing sample size (Bhattacharjee, 2002). Therefore, to assess Goodness-of-fit for the model following metrics were considered. For larger sample sizes it is recommended to consider χ^2/df also referred to as relative χ^2 , for which the value should not exceed three for models with good fit (Kline, 2011). Further, RMSEA together with NNFI, and CFI which are robust against large sample size, are included to the report. RMSEA should be below .05 and NNFI as well as CFI should exceed .97 for a good fit of the model (Kline, 2011). The model shows overall a good

fit ($\chi^2(74) = 126.562$; $\chi^2/df = 1.710$, RMSEA = 0.034, NNFI = 0.99, CFI = 0.99). All paths are significant and all factor loadings exceed 0.70. However, results from the measurement invariance test, for which online shop were compared to the loan website, couldn't confirm measurement invariance ($\chi^2(11) = 25.976$, $p < 0.01$).

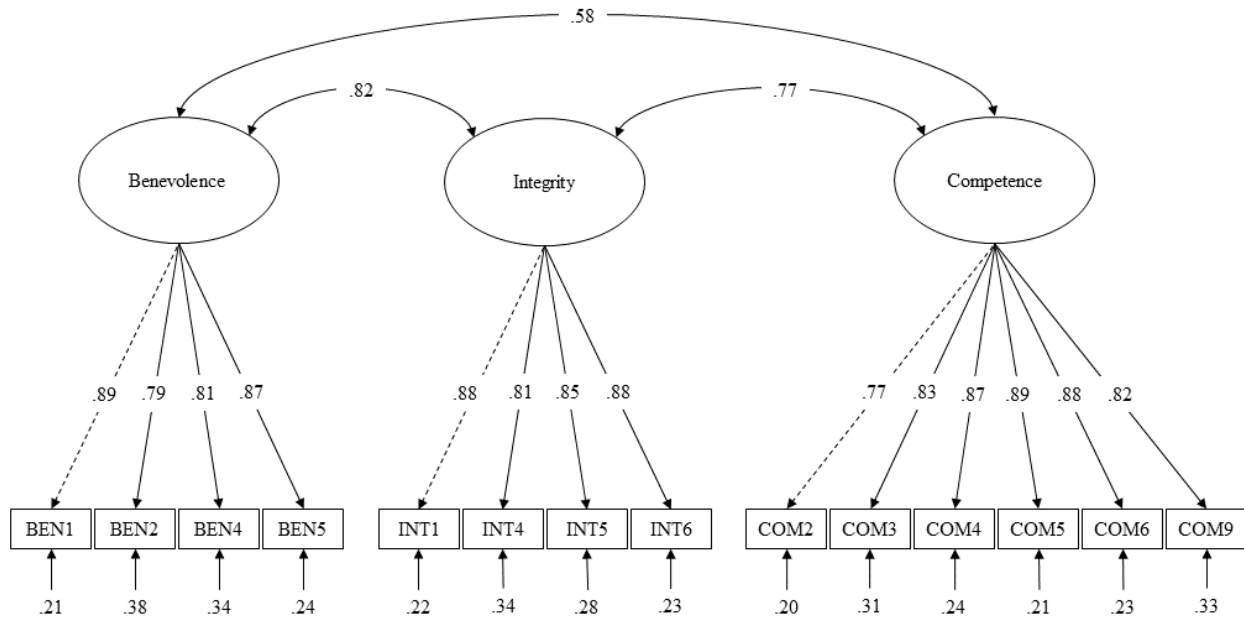


Figure 6. Standardized solution of the CFA. $N = 601$. All paths are significant ($p < .001$). Dotted arrows indicate the fixed paths of the CFA. ($\chi^2(74) = 126.562$; $\chi^2/df = 1.710$, RMSEA = 0.034, NNFI = 0.99, CFI = 0.99).

To assess convergent and discriminant validity correlations between the different scales were compared (see Table 8). The correlation between the trust scale developed in this thesis and the trust scale developed by Flavián (2006b) show a higher correlation compared to the correlation between trust and visual aesthetics as well as usability.

Table 8

Correlations between trust, visual aesthetics, and usability.

Questionnaires	Trust	Trust (Flavian)	VisAWI	UMUX
Trust	–			
Trust (Flavian)	.68	–		
VisAWI	.47	.51	–	
UMUX	.52	.54	.75	–

Note. All correlations are significant ($p < .001$). $N = 601$.

To evaluate the predictive validity the data from the first part of the study were analysed. All three subscales as well as the whole trust scale show a significant difference. See detailed results in Table 9. On average, participants given the high trustworthy screenshot of an online shop ($M = 5.41$, $SD = 0.8$) rated the trustworthiness of the online merchant significantly ($t(592) = -12.722$, $p < 0.001$) higher than those given the low trustworthy screenshot of an online shop ($M = 4.46$, $SD = 1.02$). The results indicate that the trust questionnaire can differentiate between low and high trustworthiness of an online merchant.

Table 9

Difference in means between low and high trustworthy screenshot for trust and the subscales.

	low		high		$t(df)$
	M	SD	M	SD	
Trust	4.46	1.02	5.41	0.80	-12.72(592)***
Benevolence	4.27	0.91	4.91	0.85	-8.95(596)***
Integrity	4.81	1.11	5.62	0.96	-9.66(599)***
Competence	4.36	1.31	5.60	0.94	-13.50(576)***

Note. *** $p < .001$. $N_{low} = 320$. $N_{high} = 281$.

General Discussion

The purpose of this thesis was to develop and validate a questionnaire for measuring individual trust incorporating semantic differential, for potential use in future empirical studies in online contexts. Scale construction is an important step in confirmatory research because the quality of a measurement scale determines the extent to which empirical results are meaningful and accurate (Bhattacharjee, 2002). As discussed earlier, large proportion of existing trust questionnaire make use of Likert-scale format. Likert scaled items are mostly tailored to the specific website measured, which makes it difficult to use these questionnaires in other research contexts. Based on the three dimensions of trust (benevolence, integrity, and competence) a 14-item trust scale using semantic differential meeting the limitation of Likert-scaled format was constructed. These items were tested by an expert panel for appropriate linguistic bipolarity as well as psychological bipolarity. All items load to their respective factors well as shown in the exploratory factor analysis and confirmatory factor analysis. Furthermore, the reliability for the three subscales are very promising. The correlations with other scales have indicated a convergent as well as discriminant validity. Moreover, the results support the predictive validity, meaning that, it can assess the difference between high and low trustworthiness.

The proposed trust questionnaire, can not only be applied in consumer trust but also in business to business e-commerce. From a practitioner standpoint, the trust scale presented here provides a very convenient way, meaning that it can be applied without modifications to assess their customers' level of trust in their firm. Firms whose revenue structure depends on frequent and continuous user transactions may lose a lot if they fail to assess consumer trust in their services. An early identification of users with low trust levels may help to ensure their retention by targeting them specifically with specialized interventions.

The presented findings are subject to some limitations. First, although the presented trust questionnaire show some promising results in the confirmatory factor analysis the measurement invariance could not be confirmed. Verhagen (2015) recommends to assess initial measurement invariance using split-half method. Therefore, a large sample should rate only one specific website and then randomly be assigned to either a group to determine the measurement invariance. Second, the trust questionnaire is based on the prototype trust scale by Rieser & Bernhard (2016). Items which have shown promising results in Rieser & Bernhard (2016) have now been eliminated due to various reasons discussed above. Further research should inspect these differences thoroughly to determine the reason for such a difference in psychometrical values. Third, relatively high correlations between the three subscales were observed. This indicates that the theoretical model may not be appropriate to reflect the construct. Fourth, convergent and discriminant validity were assessed only on a descriptive level. Further research could assess convergent and discriminant validity using structural equation modelling by defining a measurement model. Further research should also consider testing the trust questionnaires on more than two websites.

Overall, using the trust questionnaire based on semantic differential format can be an easy to use alternative to existing Likert scaled format questionnaires. The trust questionnaire was carefully operationalized and therefore show promising results on its validation. The main advantage of this approach lies in its effortless and versatile utilization

References

- Ajzen, I., & Fishbein, M. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study. *Journal of Marketing*, 69(4), 133–152.
- Bhattacharjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of Management Information Systems*, 19(1), 211–241.
- Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing User Satisfaction in the Era of User Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience. *International Journal of Human-Computer Interaction*, 31(8), 484–495. <https://doi.org/10.1080/10447318.2015.1064648>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194), 4–7.
- Casaló, L. V., Flavián, C., & Guinalú, M. (2007). The role of security, privacy, usability and reputation in the development of online banking. *Online Information Review*, 31(5), 583–603. <https://doi.org/10.1108/14684520710832315>
- Chen, S. C., & Dhillon, G. S. (2003). Interpreting dimensions of consumer trust in e-commerce. *Information Technology and Management*, 4(2–3), 303–318.
- Chin, W. W., Johnson, N., & Schwarz, A. (2008). A fast form approach to measuring technology acceptance and other constructs. *MIS Quarterly*, 687–703.
- Cho, J. (2006). The mechanism of trust and distrust formation and their relational outcomes. *Journal of Retailing*, 82(1), 25–35. <https://doi.org/10.1016/j.jretai.2005.11.002>

- Corbitt, B. J., Thanasankit, T., & Yi, H. (2003). Trust and e-commerce: a study of consumer perceptions. *Electronic Commerce Research and Applications*, 2(3), 203–215.
- Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. *Nebraska Symposium on Motivation*, (10), 275–318.
- Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *The Journal of Marketing*, 35–51.
- Driscoll, J. W. (1978). Trust and participation in organizational decision making as predictors of satisfaction. *Academy of Management Journal*, 21(1), 44–56.
- Erikson, E. H. (1963). *Childhood and Society*. 2d ed., rev. and enl. New York, Norton.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll* (4th edition). Los Angeles: Sage.
- Flavián, C., Guinalú, M., & Gurrea, R. (2006a). The influence of familiarity and usability on loyalty to online journalistic services: The role of user experience. *Journal of Retailing and Consumer Services*, 13(5), 363–375. <https://doi.org/10.1016/j.jretconser.2005.11.003>
- Flavián, C., Guinalú, M., & Gurrea, R. (2006b). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & Management*, 43(1), 1–14.
- Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences*, 40(5), 873–884. <https://doi.org/10.1016/j.paid.2005.08.015>
- Gefen, D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM Sigmis Database*, 33(3), 38–53.

- Grabner-Kräuter, S., & Kaluscha, E. A. (2003). Empirical research in on-line trust: a review and critical assessment. *International Journal of Human-Computer Studies*, 58(6), 783–812.
[https://doi.org/10.1016/S1071-5819\(03\)00043-0](https://doi.org/10.1016/S1071-5819(03)00043-0)
- Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Eds.), *Mensch & Computer 2003* (Vol. 57, pp. 187–196). Wiesbaden: Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-80058-9_19
- Hawkins, D. I., Albaum, G., & Best, R. (1974). Stapel scale or semantic differential in marketing research? *Journal of Marketing Research*, 11(3), 318–322.
- Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM*, 42(4), 80–85.
- Howard, M. C., & Melloy, R. C. (2016). Evaluating item-sort task methods: The presentation of a new statistical significance formula and methodological best practices. *Journal of Business and Psychology*, 31(1), 173–186.
- Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage.
- Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53–71.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). New York: Guilford Press.

- Koufaris, M., & Hampton-Sosa, W. (2004). The development of initial trust in an online company by new customers. *Information & Management*, 41(3), 377–397.
<https://doi.org/10.1016/j.im.2003.08.004>
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group* (pp. 63–76). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-89350-9_6
- Lee, M. K., & Turban, E. (2001). A trust model for consumer internet shopping. *International Journal of Electronic Commerce*, 6(1), 75–91.
- Lu, J., Wang, L., & Hayes, L. A. (2012). How do technology readiness, platform functionality and trust influence C2C user satisfaction? *Journal of Electronic Commerce Research*, 13(1), 50.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709. <https://doi.org/10.2307/258792>
- McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communications Monographs*, 66(1), 90–103.
- McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, 13(3), 334–359.
- Moorman, C., Deshpande, R., & Zaltman, G. (1993). Factors affecting trust in market research relationships. *The Journal of Marketing*, 81–101.
- Moshagen, M., & Thielsch, M. T. (2010). Facets of visual aesthetics. *International Journal of Human-Computer Studies*, 68(10), 689–709. <https://doi.org/10.1016/j.ijhcs.2010.05.006>

- Pavlou, P. A., & Gefen, D. (2004). Building Effective Online Marketplaces with Institution-Based Trust. *Information Systems Research*, *15*(1), 37–59.
<https://doi.org/10.1287/isre.1040.0015>
- Rieser, D. C., & Bernhard, O. (2016). Measuring Trust: The Simpler the Better? (pp. 2940–2946). ACM Press. <https://doi.org/10.1145/2851581.2892468>
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665.
- Seckler, M., Heinz, S., Forde, S., Tuch, A. N., & Opwis, K. (2015). Trust and distrust on the web: User experiences and website characteristics. *Computers in Human Behavior*, *45*, 39–50. <https://doi.org/10.1016/j.chb.2014.11.064>
- Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*, *66*(1), 15–37.
- Van Auken, S., & Barry, T. E. (1995). An assessment of the trait validity of cognitive age measures. *Journal of Consumer Psychology*, *4*(2), 107–132.
- Verhagen, T., van Den Hooff, B., & Meents, S. (2015). Toward a better use of the semantic differential in IS research: an integrative framework of suggested action. *Journal of the Association for Information Systems*, *16*(2), 108.
- Wang, Y. D., & Emurian, H. H. (2005). An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior*, *21*(1), 105–125.
<https://doi.org/10.1016/j.chb.2003.11.008>
- Wirtz, J., & Lee, M. C. (2003). An examination of the quality and context-specific applicability of commonly used customer satisfaction measures. *Journal of Service Research*, *5*(4), 345–355.