Psychologie
Faculteit der Sociale Wetenschappen

The User Experience of Business Software:

Fulfilment of Psychological Needs as Predictor of Affect

Name: Youp van Veen

Supervisor: Pascal Haazebroek

Second reader: William Verschuur

Cognitive Psychology

Thesis MSc Applied Cognitive Psychology

Abstract

In recent years, many studies attempted to identify what actually causes differences in the User Experience (UX) of products and software. One factor that has been posed as an important influence, is the fulfilment of psychological needs. Previous studies found strong relationships between this fulfilment and UX-related affect. The current study investigated this relationship in the context of business software. In an online survey, 137 users of a specific business management software were asked to rate their typical day using it. Affect and the fulfilment of five psychological needs (autonomy, competence, relatedness, self-esteem, and influence) were measured with the Positive Affect Negative Affect Schedule (PANAS) and a – for this study translated – psychological needs questionnaire. Regression analyses and ANCOVA indicated the fulfilment of psychological needs were able to predict affect significantly. Age of the user and experience with the software were taken into account as moderators, but both did not provide strong results. Satisfactory validity and reliability of the translated psychological needs questionnaire were found. Two possible setups for future research are posed, to address the fundamental assumption of causality.

Keywords: HCI, User Experience, UX, business software, psychological needs, PANAS, positive affect, negative affect

The User Experience of Business Software:
Fulfilment of Psychological Needs as Predictor of Affect

The field of Human-Computer Interaction (HCI) has long been dominated by the notion of usability. It is a concept which probably emerged from the understanding of computers as a working-tool. In the early years of HCI – the 60s and 70s – few people interacted with computers. Yet, most of these interactions were work-related (Tractinsky, & Hassenzahl, 2005). When a broader audience started making more and more use of computers, it became necessary to develop a measure for its ease-of-use. Originating in a work-dominated computer culture, it developed into a task-oriented measure, focused on the product and preventing errors (Zimmermann, 2008). However, since the invention of the computer, there have been many technological developments. A computer is now more accessible and offers more functionalities than ever before, which enabled users to find new purposes for it. Consequently, the users' expectations have changed (Tractinsky, & Hassenzahl, 2005), which resulted in a need for new measurements to evaluate the interaction between humans and computers.

In recent years, research indeed found other aspects than usability to also influence the success of computers, software, interactive games, websites, and other tangible products like mobile phones. For example, aesthetics of websites was found to be a strong determinant of users' satisfaction and pleasure (Lavie, & Tractinsky, 2003). These findings asked for user-oriented measures, which are now subsumed under the umbrella of User Experience (UX). According to ISO 9241-210, UX is the "person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" (2010: p. 7). Compared to usability, UX is focused on the user, instead of the product used. It is considered to be more inclusive of high-level aspects of human functioning, also addressing human emotion and motivation (Roto, Law, Vermeeren, Hoonhout, 2011).

In spite of the ISO definition, under practitioners there is no strong consensus yet on what dimensions comprise UX, how to influence it, and how to research it (Lallemand, Gronier, & Koenig, 2015). As a multidisciplinary field, approached from, amongst other fields, psychology, design, and computer science, it is hard to find a middle ground. Now, for both UX design and research, three important goals remain. First, it is important to identify dimensions of UX. This will help in further communication and teaching about UX. Second, for design specifically, it is important to identify aspects with which these dimensions can be manipulated. This enables

designers to create better products. Third, for research it is important to gain knowledge on different methods and setups, and how to study UX. There is a strong division between qualitative and quantitative research, as well as what products, use situations, and contexts are studied. Although most research makes use of questionnaires, other data collection techniques include, but are not limited to: interviews (open and semi-structured), live user observation, diaries, & psychophysiological measures (Bargas-Avila, & Hornbæk, 2011). Working towards a research paradigm ensures growth of the field.

Previous Research

Dimensions of UX. Throughout the years, many concepts have been proposed as a dimension of UX. A review of 66 studies (Bargas-Avila, & Hornbæk, 2011) found more than 8 different aspects, ranging from frustration to fun. With 41% of the studies, the focus mostly was on "generic UX". These studies did not report what dimension specifically was investigated, and resorted to vague statements as "all-encompassing user experience" (Koca, Funk, Karapanos, Rozinat, & Van der Gaarden, 2008). The first most investigated dimension is affect, or an emotional response; an influential and well-supported notion, with 24% of the reviewed studies supporting this idea. For example, Mahlke and Thüring (2007) studied UX as an emotional response. They asked people to interact with two – computer simulated – versions of a portable audio player, which differed in usability and aesthetics. Participants then had to rate the different versions on a set of instrumental and non-instrumental qualities – aspects such as appeal and attractiveness – and indicate their emotions using a set of pictures. It was found that both categories of product qualities influenced the participants' emotional response, indicating there is a relationship between (the use of) products and emotions.

Although often used as synonyms in UX research, emotion and affect are not identical. As affect plays an important role in this study, it is important to clarify the difference. The term emotion refers to a mental state, which is high in intensity. It is usually a short-lived experience that can trigger the body into action, and is commonly caused by a specific stimulus or event (Scherer, 2005). Affect refers to a category of states, of which emotion is one – other states include, but are not limited to, preferences, attitudes, and moods (Scherer, 2005). According to most theories, these states consist of three different dimensions: valence, arousal, and motivational intensity (Harmon-Jones, Gable, & Price, 2013). The first is a positive to negative evaluation of the experienced state, whereas arousal displays the degree of activation. Motivational intensity

describes “the strength of urge to move toward/away from a stimulus” (Harmon-Jones, et al., 2013). Hence, affective states can be classified with these different dimensions. For example, moods are commonly seen as relatively low in arousal, whilst emotions are high in arousal (Scherer, 2005). It is not in this study’s scope to extensively discuss the differences between the different affective states. Yet, it is important to state explicitly that affect thus encompass emotions, but it also includes other aspects as preferences, moods, and attitudes. In the context of interaction with technology, affect is influenced by a great variety of aspects. Using a product, one might experience moments of surprise, delight, and irritation. The appearance of the product might not be appealing, or its function does not fulfil the users’ needs. These can all impact the affective state of the user, which underlines the importance of affect (Norman, 2004).

How to influence UX. With the different dimensions of UX come different ways to influence them. Already in the year 2000, pleasure was coined as an important aspect shaping the experience of a user. A book by Patrick Jordan (2000), *Designing Pleasurable Products*, pioneered in the change to broaden the scope of product design. It is a plea for the inclusion of pleasure as a criterion, drawing knowledge from anthropology theories on what is pleasurable in people’s lives. In the book, he set out how pleasure can help improving the experience with tangible products, and as a result increase the user satisfaction. Researchers in the field of HCI learned about this notion, and used it to improve the study of UX. For example, Van der Heijden (2004), who formulated pleasure as perceived enjoyment. The author considered UX to be user acceptance, which is the decision about when and how to use a product. The study focused on digital environments that are pleasure-related, like a movie website. Perceived enjoyment was found to be an important factor in explaining the degree to which users accepted the websites, indicating that pleasure might contribute to the decision to use a product.

Naturally, other concepts were also investigated for their influence on UX. One example is that of the fulfilment of psychological needs. The most widely accepted theory on psychological needs is the self-determination theory (SDT; Ryan, & Deci, 2000). This theory postulates autonomy (feeling that actions are self-chosen), competence (feeling effective in your actions), and relatedness (feeling close to others) as basic, universal needs. Fulfilling these needs is seen as an important determinant of well-being, whilst failing to fulfil them can cause severe psychological problems (Deci, & Ryan, 2000). Other theories also proposed the needs for self-esteem,

stimulation, physical thriving, self-actualization, security, influence, and luxury (Sheldon, Elliot, Kim, & Kasser, 2001).

Psychological needs are interesting aspects to investigate, as they are considered to be a user-oriented measure. Thus, this is more useful in measuring UX, than product-oriented measures traditionally used for usability research (Hassenzahl, Wiklund-Engblom, Bengs, Hägglund, & Diefenbach, 2015). Once identified, these "can be targeted to enhance personal thriving" (Sheldon, et al., 2001, *p.*1), and were found to be a source of affect in general life experiences (Sheldon, et al., 2001). In a series of experiments, participants were asked to describe the "single most personally satisfying event that [the participant] experienced during the last month". These experiences had then to be rated on a psychological needs questionnaire, and an affect questionnaire. Both the results of US and South Korean participants showed that the fulfilment of psychological needs was strongly associated with affect. With an increase of fulfilment of psychological needs, an increase of positive affective states, and decrease of negative affective states, was observed. Seeing the connection with affect – a dimension of UX considered to be important – this idea was inevitably soon investigated in the context of interaction with technology. Multiple studies showed the same relationship applies not only to life experiences in general, but also to experiences people have when using technology (e.g. Hassenzahl, et al., 2015; Partala, & Kallinen, 2012).

How to study UX. An important aspect all UX research is faced with, is the timespan of the experience. This refers to the point in time that the experience of interest takes place (Roto, et al., 2011). It has been proposed that it is best to focus on an experience at an earlier point in time, than to measure it at the moment the experience is happening (Norman, 2009). By measuring UX in a retrospective way using the participants' memories, the measurements are intentionally influenced by the imperfect aspects of memory. This increases the ecological reliability of the results, and gives more insight into how a user will decide on using a product on the long-term (Schachter, Addis, & Buckner, 2007; Oishi, & Sullivan, 2005). Also, previous research has shown measuring experience in a retrospective way is an effective method (for examples, see: Hassenzahl, Diefenbach, & Göritz, 2010; Hassenzahl et al., 2015; Partala, & Kallinen, 2012).

In a UX whitepaper, four different time spans of UX were distinguished, of which two are retrospective: episodic and cumulative (Roto, et al., 2011). For episodic UX, a user reflects on one specific experience, which is a well-defined frame in time. To the contrary, cumulative UX relates to multiple experiences divided over a longer period of time. It is the sum of the views on a system collected after having used a product for various times. For a schematic overview of the different timespans, see Figure 1.

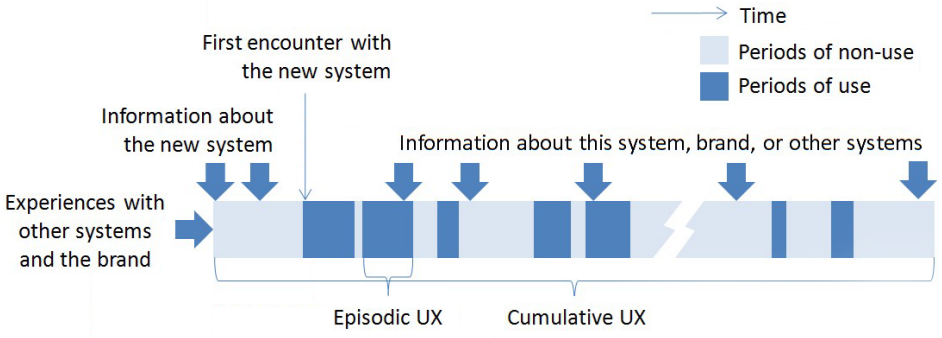

Figure 1: A schematic overview of retrospective time spans of UX, adapted from Roto et al. (2011)

Important limitation of UX research. Albeit many studies showed promising results with different UX dimensions and influencers, most of them only consider a context of leisure, personal use. For example, people reflecting on their experiences with their mobile phones, playing video games, or interactive technology in general (e.g. Partala, & Kallinen, 2012). As UX is heavily subject to the situation, and people's goals and expectations differ per context, it is necessary to study UX in different contexts (Roto, et al., 2011). One context that is not yet studied often, is the business context (Bargas-Avila, & Hornbæk, 2011).

This Study

The current study was, therefore, intended to study UX in the business context, with a focus on the dimension of affect. This is the most supported notion of UX (Bargas-Avila, & Hornbæk, 2011), and has proven to provide fruitful results in other contexts (e.g. Müller, Mekler, & Opwis, 2016; Partala, & Kallinen, 2012). Only the valence of affect was considered, which is the positive to negative evaluation. This allowed for a relatively simple measurement, and yet provided valuable results. In the used questionnaire, positive and negative affect are seen as two relatively independent states, and measured with their own respective subscale. Furthermore, the influencing element inspected was the fulfilment of psychological needs. This has not only been regarded as a user-oriented construct, it has also been tested in relation to affect with good results (e.g.

Hassenzahl, et al., 2010; 2015). Therefore, the approach used in these previous experiments formed the basis for this study.

More precisely, users of a particular business management software tool were asked to rate their past experiences with the software. A set of questionnaires designed to measure affect and the fulfilment of psychological needs were used to provide insight into how (interacting with) the software is related to both aspects. Unlike previous research, which considered the episodic UX, this study focused on the cumulative UX. Although research into the episodic UX has proven to give sufficient results, this study assumed it is not yet a proper approach in the context of business software. Due to the typically prolonged use of this type of software, it presumably does not provide clear, well-defined satisfying or unsatisfying experiences. This would make it difficult for people to rate a specific event. Hence, the focus was on a more general evaluation of their experiences. Specifically, this study asked: what psychological needs best predict UX-related affect on a typical day?

For this study, the needs for autonomy, competence, and relatedness were deemed to be important, as they are widely accepted as universal, basic needs, and play a central role in SDT, which is an important theory in occupational psychology (Gagné, 2014). Therefore, their predictive value of positive and negative affect was examined. Based on this theory, the first hypothesis was stated as: (1) the fulfilment of the needs for autonomy, competence, and relatedness, is positively related to positive affect, and negatively to negative affect. Included in this study was also the need for self-esteem, because this has shown promising results (Partala, & Kallinen, 2012; Sheldon, et al., 2001), and influence, as this was assumed to be an important aspect of a job environment. An overall degree of need fulfilment has been found to be related to positive and negative affect in general life experiences (Sheldon, et al., 2001) and this has been replicated in the context of human-computer interaction (Hassenzahl, et al., 2010; 2015; Partala, & Kallinen, 2012). As a result, the second hypothesis was formulated as: (2) the overall need fulfilment is significantly correlated positively with positive affect, and negatively with negative affect. Additionally, users with more experience were expected to be more competent with the software due to training (Langdon, Lewis, & Clarkson, 2007). Thus, it was expected that this need is more important for users with low experience. Based on this, the third hypothesis was: (3) the correlation between fulfilment of competence, and positive and negative affect respectively, is moderated negatively and positively by amount of experience with the product. Further, moderation of the

age of the users was explored, to see if it regulated the influence of need fulfilment on affect. There was no specific hypothesis formulated for this, yet the focus was on the SDT needs. Furthermore, previous research results have shown autonomy, competence, relatedness, self-esteem, and influence to be important predictors of affect (e.g. Hassenzahl, et al., 2010; 2015; Partala, & Kallinen, 2012). Therefore, it is interesting to find out which need was the best predictor of affect. There was no specific hypothesis formulated for this. The same question was also investigated for negative affect. This study was conducted in cooperation with a Dutch supplier of business software, who provided participants for this research.

Methods

Design

This study was of correlational nature. The participants were not selected beforehand on specific conditions. Rather, all invitees were accepted to participate, and were not assigned to experimental conditions.

Of each participant, through an online survey, eight main variables were measured: positive affect, negative affect, overall need fulfilment and the fulfilment of autonomy, competence, relatedness, self-esteem, and influence. Positive and negative affect refer to, more specifically, experience-related affect. These were measured with 10 statements each, which had to be rated on a scale from 1 to 5 each. The sum of these ratings depicts the total score, effectively ranging from 10 to 50. These are interval level. The fulfilment scores were gained by rating 3 statements for each need, on a scale from 1 to 5 each. The average of the respective scores corresponded to the fulfilment score. This, thus, ranges also from 1 to 5, and was treated as ratio data. The overall need fulfilment is an average of all the need scores, and was treated likewise.

Next to these main variables, the age group of the participant was asked, as well as the experience with the product in months. The different age groups were as follows: <26 years, 26-35, 36-45, 46-55, 56-65, >65 years old. As the participants came from a working population, these groups were considered to be meaningful. This is ordinal data. Experience with the product, as asked in months, is ratio level.

In an attempt to have the participants fully focused on their experiences with the product, 5 prime questions were developed. These asked for specific parts of the product, and were carefully formulated in order to be relevant for all participants. This was supposed to help them retrieve

their own memories of their encounters with the product. These questions were not intended to measure anything, and thus are discussed further under *Apparatus*.

Positive and negative affect functioned as outcome variables, whereas the fulfilment scores were seen as predictor measures. The age group and experience with the product were, for the relevant hypotheses, covariates. As a result, this study did not have a strong experimental character. Yet, this is in line with earlier research, and the results were thus interpreted similarly.

Participants

To retrieve participants for this study, a Dutch supplier of business software cooperated. Their software is divided into different categories, depicting different fields of work. To have a clear focus, and to eliminate some error caused by different work fields, only the Dutch user group of one specific product group was of interest. This was the largest group of users and thus provided the most data to work with. Of this group, a randomly chosen 2000 people were invited via email to participate in this research. The survey was online for 2 weeks, and after the first week a reminder was sent. The only inclusion criterion was to have worked with the software at least once. There were no restrictions on age or experience with the product. In total 222 people started the survey, of which 137 people (n = 87 males, n = 50 females) fully completed the survey. Of the participants who did not fully complete the survey, there is no data on why or after how long. The chosen survey system did not allow for this data to be stored. The distribution over the age groups can be found below in Table 1.

Table 1. *Distribution of Participants over the Age Groups*

Age group	<26 years	26 – 35 years	36 – 45 years	46 – 55 years	56 – 65 years	>65 years
n	4	17	27	50	34	5

For partaking in this survey, the participants were not compensated. The ethical aspects of this study were reviewed by the Ethical Committee Psychology of Leiden University and approval was granted. Also, the study was conducted in accordance with the applicable laws and guidelines.

Apparatus

As this study was an online survey, all questionnaires were provided digitally to the participants. They were free to answer the questionnaires with a device, and in a place, of their choice. To setup the questions and gather the data, software by Verint was used. This software was

in license of the cooperating company. It does not allow insight in the time it took participants to fill in the separate questionnaires. Further analysis of the data was performed with SPSS. The survey consisted of 4 parts: prime questions, the Positive Affect Negative Affect Schedule (PANAS), a psychological needs questionnaire, and 3 demographic questions; for gender, age group, and amount of experience. In total, 43 questions had to be answered. These parts are discussed in the order they were presented to the participants.

Prime questions. The first part consisted of 5 prime questions which were intended to activate the memory of the participants' experiences with the software. These were formulated for this study specifically, in collaboration with product experts from the cooperating company. The questions consider a broad set of product aspects, to ensure the participants did not focus too much on one experience. Also, the covered aspects are essential for using the software, and are thus known to be used by every user. An example of one prime question can be found in Figure 2. Participants were asked to rate their emotion with these aspects using pictures of 5 emotions, ranging from sad to neutral to happy. These pictures were familiar to most participants, because they are also used in the feedback form of the software. These questions were not intended to measure anything; they were merely supposed to activate the memory of the participants. An emotional response was considered to be most appropriate, to fall in line with the measurement of affect. The questions were stated in a way to reflect the appropriate timeframe. Refer to Appendix A for an overview of all questions.

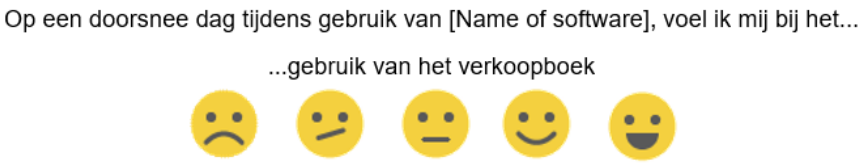

Figure 2: An example of a prime question

PANAS. The second part was aimed to measure UX-related affect. For this, the PANAS was used. Specifically, it is the validated Dutch translation of the original English questionnaire (Engelen, De Peuiter, Victoir, Van Diest, & Van den Bergh, 2006). It was mainly developed for the clinical context, but previous research has used it in the context of UX research before – with success (e.g. Partala, & Kallinen, 2012). The questionnaire aims to measure two dominant affective dimensions, namely positive affect and negative affect. Positive affect "reflects the extent to which a person feels enthusiastic, active, and alert" (Watson, Clark, & Tellegen, 1988, p. 1063).

On the other hand, negative affect reflects "aversive mood states, including anger, contempt, disgust, guilt, fear, and nervousness" (Watson, et al., 1988, p. 1063). These are, counterintuitively, relatively independent, meaning a high positive affect can occur together with high negative affect (Watson, et al., 1988).

The questionnaire consisted of 20 words describing emotions and feelings – 10 for each subscale – which the participants had to rate on a scale from 1 (=very little) to 5 (=very much). The score for each subscale was the summation of the 10 relevant ratings, effectively giving a range of 10 - 50. A low positive affect score indicated a low level of experienced positive emotions, whilst a high score showed a high level of experienced positive emotions. The same applied to the negative affect scale. Important to note is that the subscales were treated as independent. As a consequence, for experience-related affect there were two scores, which were used autonomously in the analyses. The timeframe of the question stem was changed to "on a typical day using [Name of the software] I feel…" to be in line with the research question, and the psychological needs questionnaire. This would provide affect that is related to the participants' experiences with the product. The PANAS allows explicitly for modulation of the timeframe (Watson et al., 1988), so this should not have given problems. The complete questionnaire can be found in Appendix B.

Psychological needs. To measure the fulfilment of the separate psychological needs and overall need fulfilment, the questionnaire as used by Sheldon and colleagues (2001) was adapted. This questionnaire has given reliable results in previous research in the context of UX (e.g. Hassenzahl, et al., 2010; 2015; Partala, & Kallinen, 2012). Only 15 of the original 30 statements were used, as only 5 out of the 10 needs were in the scope of this study. The individual needs were regarded as subscales, and the research of Sheldon and colleagues (2001), as well as other research which adapted the method (e.g. Hassenzahl, et al., 2010, 2015;) has shown sufficient reliability and validity. The original statements follow the stem "During this event I felt...", but since the evaluation of a typical day is of interest, the stem was changed accordingly to "On a typical day using [Name of the software], I feel..." Participants were asked to rate these statements on a scale from 1 (=very little) to 5 (=very much). The score for every need was calculated by averaging the ratings of the 3 relevant statements. As a consequence, the scores for each need ranged from 1 - 5. A low score indicated low fulfilment of this particular need, whilst a high score indicated high satisfaction. To come to the total need fulfilment, the average of all the needs was calculated. Again, this score ranged from 1 - 5. Interpretation of the scores was similar to that of the subscales.

The documentation team of the software company, which is concerned with the specific task of translating communication and documents into different languages, translated the questionnaire from English to Dutch. Reliability and validity analyses are conducted in this study to evaluate the questionnaire. See Appendix C for a complete overview of all the statements.

Procedure

The cooperating company send out emails to all invitees, asking for help in an online study. The study was online for 2 weeks, and after the first week a reminder was sent to all people who did not enter the survey yet. As there were no different conditions, instructions and tasks for all participants were the same. The email stated the survey was intended to improve the company's products and services, and that it would not take more than 10 minutes. Clicking a button saying "participate" redirected the person to the website with the survey. Here, participants were explained the duration of the survey, their right to stop at any time they want, and the confidentiality of their answers. Further, they were told the study is about the experience of the product, and that their answers could help in improving the services of the company. The participants could fill in the survey when and where they wanted, as long as they completed it in one session. There was no possibility to save the answers and continue at a later time. It was indicated that by clicking the ''continue'' button the participant would agree to the set conditions, and start with the survey.

Prime questions. Having started the survey, the participants were given instructions for the first part, the prime questions. It was mentioned the participant would be presented with 5 statements about their daily experiences with the company's product, and they were asked to indicate how they felt with each statement. For the exact instructions, see Appendix A. Having clicked ''continue'', the 5 statements were presented with the smiley pictures to rate. To indicate the emotion, the desired picture had to be clicked. After all statements were rated, participants could continue by clicking a button at the bottom of the page.

PANAS. This button brought them to the second instruction page – for the PANAS. The instructions specified the next questionnaire consists of a series of words depicting feelings and emotions. Furthermore, the participants were asked here to give a rating for each word, to indicate to what degree they experience this feeling on a typical day using the company's product. See Appendix B for the full instructions. To start the actual questionnaire, the "continue" button had to be clicked. At this page, the question stem "on a typical day using [Name of the software] I

feel…" was presented at the top. Below, the 20 words were given. By clicking the desired radio button, participants could then indicate their rating.

Psychological needs. On the next page, the instructions for the psychological needs questionnaire were presented. These said that, again, statements would be displayed, which had to be rated on a 1 to 5 scale. The question stem was presented, together with an example of a statement. For full instructions, see Appendix C. After this page, the statements were given in random order, with the question stem on top. Clicking the desired radio button indicated the degree to which the participant experienced the feeling.

At the next and final page, participants were able to enter any comments they had, and were asked for their gender, age group, and months of experience with the software. Gender could be indicated with a female, male, and neutral option. Age group had to be selected from the predefined age groups, whereas the months of experience with the software could be entered in a free-form text field. The debriefing was displayed to finish the survey. A thank you message and an explanation of the study's goal were shown. On average, it took 8 minutes and 55 seconds to fully complete the survey.

Analysis

After collecting all the data, the desired variables were computed from the relevant scores. Specific information on these calculations can be found under *Apparatus.* Next, the data was inspected for any abnormalities. Hereafter, reliability and validity checks of the measurements were performed, and regression analyses and ANCOVA's were executed for the relevant hypotheses. A specific, regression based, macro was used to test the moderation as set out in the third hypothesis. For these tests, the necessary assumptions were checked, which are mentioned in the *Results* section. Conclusions were drawn based on the results of all analyses.

While checking the data for any abnormalities, outliers were identified and the distribution of the measurements was inspected. Univariate outliers were detected via boxplots. Any score marked as such by SPSS was excluded for further analysis. In ambiguous cases, the standardized scores were used to label outliers. Any standardized score lower than -2.68 or higher than 2.68 were also excluded from further analysis. During the statistical tests the histograms of standardized residuals were visually checked for possible multivariate outliers. Missing data was not relevant, as only fully completed surveys were taken into account.

The reliability and validity of the psychological needs questionnaire and the PANAS was administered with reliability analyses and confirmatory factor analyses. For each subscale, reliability is expressed in Cronbach's alpha, and exclusion of individual items is based on the alpha if item is deleted. Validity is inspected per subscale with the Kaiser-Meyer-Olkin (KMO) measure, with factor loadings for each item. Factor loadings equal to or higher than .50 were considered to indicate important items. In each factor analysis, the explicit extraction of factors was set to 1. Further, intercorrelations between the psychological needs subscales is reported, as well as those between the two PANAS subscales.

Hypotheses 1 to 3. For the first hypothesis, a regression analysis with the scores for autonomy, competence, and relatedness as predictor variables was executed. In this instance, the desired method was enter. In two separate analyses, respectively positive and negative affect were the outcome variable. A regression analysis was repeated for the second hypothesis, but with the overall needs score as predictor variable. To test the moderation of experience, on the relationship between competence and affect, the PROCESS macro by Hayes (2013) was used. It is a regression based method to check for mediation, moderation, and conditional processes. In the macro, either positive or negative affect was set as outcome variable, competence as predictor, and experience as proposed moderator. The test was set to model number 1, as this specifies the moderation model.

Explorative analyses. In contrast to the previous tests, the following analyses were of explorative nature. The moderation of the age group on the relationship between autonomy, competence, and relatedness on the one hand, and affect on the other hand, was tested with ANCOVA tests. The PROCESS macro was not suitable for this, because the age group is a categorical variable. The age groups were the fixed factor and affect the outcome variable, with the need fulfilments as a covariate. For each need a separate analysis was performed. Distribution of the participants over the age groups is heavily asymmetrical. It ranges from 4 scores for the <26 years old group, to 50 scores for the 45 - 54 years old group. To increase robustness to non-normality, the two age groups with fewer scores than 15 were excluded from the analysis; these are the <26 years old (n = 4) and >65 years old (n = 5) age groups. This still means the size of the largest group, the 46 - 55 years group (n = 50), was more than 1.5 times larger than the smallest group, the 26 - 35 years old (n = 17). As a result, the F-test was relatively not robust to unequal group variances, and thus Levene's Test Equality of Error Variances is reported in case of significant results.

Finally, for the last explorative question – which need best predicts affect? – two approaches were considered to be meaningful. First, the question is answered on the condition that all needs are in the model. To achieve this, two regression analyses with, respectively, positive and negative affect as outcome variable, and the five need scores as predictor variable, were executed. This test showed which need explains the most variance uniquely. With the second approach, it was tested how much this specific need can explain individually. The need that was found to explain the most variance uniquely – given that all the needs are in the model – was now the only predictor variable in a regression analysis. Again, positive and negative affect were the outcome variables.

Results

To test all aforementioned hypotheses, regression analyses and ANCOVA's were performed. Reliability and validity of the measurements was tested. Further, assumptions for the analyses were checked, and the specific hypotheses were tested with the relevant variables. First, however, the data was inspected for any abnormalities.

In total, 137 participants (n = 87 males, n = 50 females) fully completed the survey. Partially filled in surveys were not used for the analyses, eliminating the chance of missing data. This was supported by visual inspection of the datafile. Median age group is 46 – 55 years. Based on visual inspection of the boxplots for all the variables, three outliers on negative affect were detected. An additional score was suspected of being an outlier. Inspection of the standardized scores showed this score was lower than -2.68, and thus was marked as an outlier. All four outliers were excluded from further analyses with negative affect. The independence between positive and negative affect allowed for all positive affect scores to be included in the analyses. All in all, for the positive affect analyses 137 scores were taken into account; whilst for the negative affect analyses this was 133.

Despite these exclusions, the distribution of the negative affect scores (M = 14.5, SD = 6.6) was still heavily, positively skewed, with skewness of 1.76 (SE = 0.21) and kurtosis of 2.27 (SE = 0.42). However, for completeness of the analysis, the negative affect scale was included in its current state. Transformation of the scale was not considered appropriate, as it would make the interpretation of the results fuzzy. Implications of these problems are elaborated on in the discussion section. The distributions of all other scales showed no serious abnormalities. See Table 2 for an overview of all descriptive statistics.

The validity and reliability of the psychological needs questionnaire were assessed with confirmatory factor analyses and reliability analyses for all subscales individually. Each analysis explicitly extracted 1 factor, relating to the investigated subscale. All KMO values are relatively low, with a range from .62 for autonomy to .91 for overall fulfilment. The factor loadings of each item on its respective subscale are high, with .77 as lowest for the first item of the competence scale. The reliability analyses showed high reliability of all subscales, with competence (3 items; Cronbach's α = .73) being the lowest. The exclusion of no items was found to increase the reliability. See Table 2 for an overview of the most important scores and Appendix D for all results.

In identical manner, the validity and reliability of positive affect and negative affect were investigated. The factor analysis of positive affect returned an adequate KMO value of .88. Factor loadings were medium to high, with a range from .46 for item *uitgelaten* (excited) to .81 for item *sterk* (strong). The same analysis with negative affect returned comparable results, with a KMO value of .92, and factor loadings ranging from .77 for item *nerveus* (nervous) to .92 for item *angstig* (scared). Reliability of positive affect was found to be sufficient (10 items; α = .897). Deletion of items *aandachtig* (attentive) and *actief* (active) would result, respectively, in an increase of α to .903 and .90. The complete use of the scale was considered to be more important than this small increase, and therefore all items were included in the analysis. Reliability of negative affect was also found to be sufficient (10 items; α = .953), and no items were excluded from the analysis. See Table 2 for an overview of the most important scores, and Appendix E for an overview of all validity and reliability scores of positive affect and negative affect.

Table 2. *Descriptive Statistics of All Measurements*

Scale	*N*	No. of items	*M* (*SD*)	Skewness (*SE*)	Kurtosis (*SE*)	α	*KMO*
Positive affect	137	10	31.4 (7.90)	-0.62 (0.21)	0.10 (0.41)	0.90	0.89
Negative affect	134	10	14.5 (6.60)	1.76 (0.21)	2.27 (0.42)	0.95	0.92
Autonomy	137	3	3.5 (0.89)	-0.89 (0.21)	0.67 (0.41)	0.76	0.62
Competence	137	3	3.2 (0.98)	-0.42 (0.21)	-0.43 (0.41)	0.73	0.65
Relatedness	137	3	3.07 (1.00)	-0.55 (0.21)	-0.32 (0.41)	0.81	0.71
Self-esteem	137	3	3.12 (0.98)	-0.51 (0.21)	-0.19 (0.41)	0.79	0.70
Influence	137	3	3.2 (0.89)	-0.73 (0.21)	0.19 (0.41)	0.81	0.71
Overall fulfilment	137	15	3.2 (0.90)	-0.64 (0.21)	0.09 (0.41)	0.95	0.91

The subscales of the psychological needs have high intercorrelations. The lowest is between competence and influence $r(137) = .734$, $p < .001$, and the highest between relatedness and self-esteem $r(137) = .881$, $p < .001$. This makes it difficult to decipher the individual predictive values of variables. Implications of this will be further discussed in the discussion section. A full overview of the intercorrelations can be found below in Table 3. Similar to the needs subscales, positive affect and negative affect showed intercorrelation. It was, however, a weaker, negative relationship: $r(134) = -.224$, $p = .005$.

Table 3. *Intercorrelations of all Psychological Need Subscales*

	Need				
Need	Autonomy	Competence	Relatedness	Self-esteem	Influence
Autonomy	1				
Competence	.819**	1			
Relatedness	.815**	.854**	1		
Self-esteem	.799**	.834**	.881**	1	
Influence	.818**	.734**	.814**	.857**	1

**$p < .01$

Hypothesis 1

The first hypothesis was concerned with the predictive value of autonomy, competence, and relatedness for, respectively, positive and negative affect. To begin with, a regression analysis was performed with the three needs as predictor variables, and positive affect as outcome variable. The method for this analysis was enter. A significant model was found, $F(3, 133) = 49.80$, $p < .001$, with an adjusted R^2 of .518. Of the variance of positive affect, autonomy uniquely explained 2.0%, competence uniquely explained 4.7%, and relatedness 0.03%. Collinearity statistics were sufficient, with tolerance between .23 and .28. Participants' predicted positive affect was equal to 10.25 + 2.44*(autonomy) + 3.68*(competence) + 0.29*(relatedness). Only relatedness was not significant ($p = .766$). The regression was repeated with negative affect as outcome variable. This produced a significant model, $F(3, 133) = 4.44$, $p = .005$, with an adjusted R^2 of .072. Of the variance of negative affect, autonomy uniquely explained 2.6%, competence uniquely explained 3.0%, and relatedness 7.6%. Collinearity statistics were sufficient, with tolerance ranging from .26 to .31. Participants' predicted negative affect was equal to 18.7 – 2.29*(autonomy) - 2.41*(competence) + 3.72*(relatedness).

Hypothesis 2

The second hypothesis stated the relationship between the overall need fulfilment to be correlated positively with positive affect, and negatively with negative affect. First, a regression analysis was performed with the overall needs score as predictor, and positive affect as outcome variable. A significant regression equation was found, $F(1, 135) = 150.59$, $p < .001$, with an adjusted R^2 of .524. Participants' predicted positive affect was equal to 10.711 + 6.421*(overall needs). However, one possible multivariate outlier was detected in the standardized residual plot. With exclusion of this outlier, a new significant model was found, $F(1, 134) = 178.28$, $p < .001$, with an adjusted R^2 of .568. The regression coefficients differed slightly. Participants' predicted positive affect was equal to 9.86 + 6.73*(overall needs). To investigate the relationship between overall need fulfilment and negative affect, a regression analysis was performed with negative affect as outcome variable. This produced a non-significant model ($p = .770$).

Hypothesis 3

The third hypothesis considered the predictive value of competence for affect to be moderated by experience with the product. To test this, the PROCESS macro by Hayes (2013) was used. In the macro, positive affect was set as outcome variable, competence as predictor, and experience as proposed moderator. Model 1 was chosen, as this specifies the moderator model. The macro returned a significant regression, $F(3, 133) = 44.82$, $p < .001$, with an R^2 of .503. However, the interaction effect was not significant ($p = .83$). The same analysis with negative affect as outcome variable, returned a non-significant result ($p = .07$).

Exploration of age moderation

To explore the moderation of age on the predictive value of the needs, six ANCOVA tests were performed. For autonomy, competence, and relatedness, the interaction with the age groups was analysed. In the ANCOVA's, the age group is the fixed factor, where the need score is the covariate. The analyses are executed with both positive and negative affect as outcome variable. Only the analyses with significant interaction effects are stated in-text. For a full overview of all results, see Appendix F.

The ANCOVA with autonomy, age, and negative affect, showed no main effect of autonomy, $F(1, 117) = .001$, $p = .971$, $\eta 2 = .00001$. A main effect of age was observed, $F(3, 117) = 3.14$, $p = .028$, $\eta 2 = .07$, and an interaction of autonomy and age, $F(3, 117) = .287$, $p = .040$, $\eta 2 = .07$. Levene's test is not significant, with $p = .518$, indicating equality of error variances. The

effect size is relatively low. A scatterplot, seen in Figure 3a, shows the direction of the interaction effect. For the groups of 26 – 35 years old, and 46 – 55 years old, a negative relationship is seen. To the contrary, for the groups of 36 – 45 years old, and 56 – 65 years old, a positive relationship is detected. The same analysis, but for relatedness, showed similar results: no main effect of relatedness $F(1, 117) = 3.12$, $p = .080$, $\eta 2 = .02$, an effect of age $F(3, 117) = 2.77$, $p = .044$, $\eta 2 = .06$, and an interaction effect $F(3, 117) = 2.76$, $p = .045$, $\eta 2 = .06$. Levene's test is not significant, with $p = .518$. The scatterplot of relatedness and negative affect shows a positive relationship for the 36 – 45 and 56 – 65 years old, and a negative relationship for the 46 – 55 years old. This plot is shown in Figure 3b. The Mahalanobis Distance showed no multivariate outliers in both analyses.

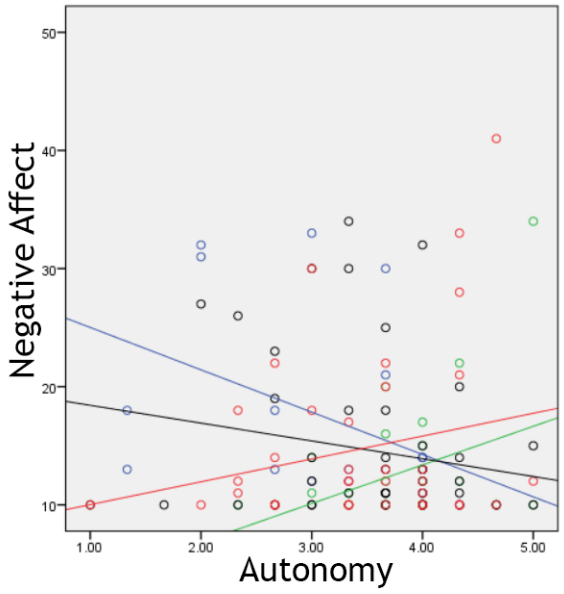

Figure 3a: Scatterplot of autonomy and negative affect for the different age groups

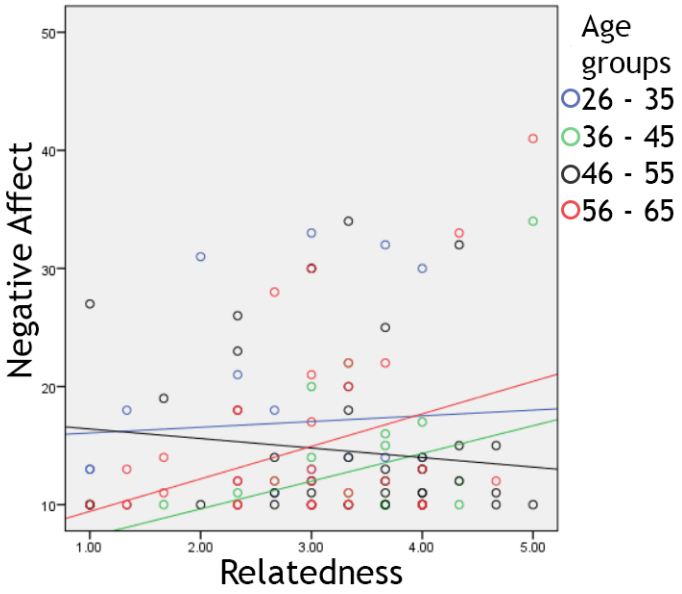

Figure 4b: Scatterplot of relatedness and negative affect for the different age groups

Best predicting needs

To explore which needs best predict affect, there were two steps. First, interest lied in the need with the most uniquely explained variance, given that all the needs are in the model. A regression analysis with all separate needs as predictor and positive affect as outcome variable, gave a significant regression, $F(5, 131) = 32.89$, $p < .001$, with an adjusted R^2 of .540. The variable with most uniquely explained variance was competence, with 2.9%. This same test was executed with negative affect as outcome variable, and gave a significant result, $F(5, 127) = 3.93$, $p = .002$, with an adjusted R^2 of .100. Here, the need with most uniquely explained variance was autonomy, with 5.3%.

The second step was to check how much these needs, competence and autonomy, could explain independently. Therefore, two additional regression analyses were performed. To begin with, a regression analysis was executed with competence as predictor variable, and positive affect as outcome variable. This gave a significant regression, $F(1, 135) = 136.20$, $p < .001$, with an adjusted R^2 of .499. In other words, it explained 49.9% of the variance of positive affect. In Figure 4 a scatterplot is shown, depicting the scores of competence and positive affect. For negative affect, autonomy was found to be the predictor explaining uniquely most variance. A regression analysis with negative affect as outcome variable and autonomy as predictor variable, however, returned a non-significant model ($p = .147$).

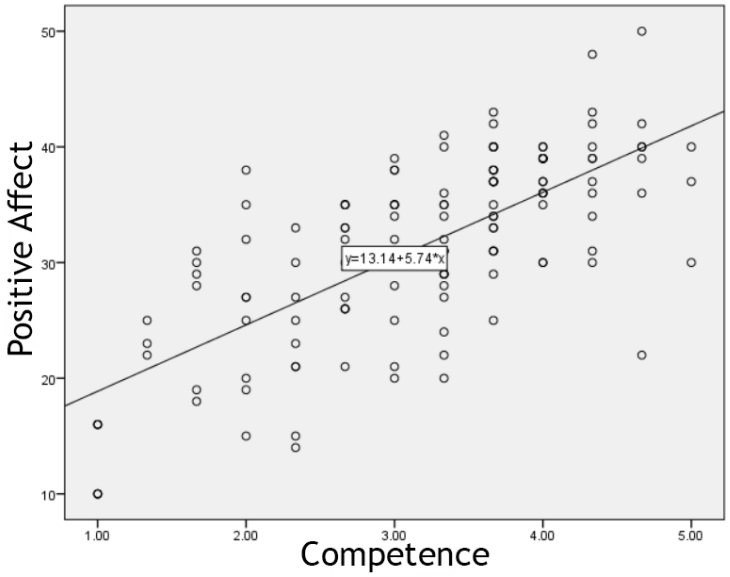

Figure 4: Scatterplot of competence and positive affect

Discussion

All in all, this study aimed to identify the impact of fulfilling psychological needs on the user's experience of business software. Of the experience, it was the affective dimension that was of interest. Participants were asked to rate their typical day working with a particular software tool, on a set of questionnaires used to measure affect and the fulfilment of psychological needs. With the examination of the relationship between these two aspects in the context of business software, an attempt has been made to establish their importance in a context not studied often before. Additionally, the moderation of experience with the software and age of the participants was investigated.

Findings

The results presented in this paper are largely in line with previous research. The idea that the psychological needs have an impact on the affective state, in the context of interacting with software, seems to be supported.

Need fulfilment as predictor of affect. For the first and second hypotheses, regression analyses were performed to examine the predictive value of need fulfilment for affect. Hypothesis 1 was concerned with autonomy, competence, and relatedness – the three needs from the SDT – whereas hypothesis 2 considered only the overall need fulfilment. For autonomy, competence, and relatedness as predictors, a significant, positive relationship with positive affect, and a significant, negative relationship with negative affect was found. The first hypothesis is hereby fully accepted. In a separate analysis with the overall need fulfilment, a significant, positive relationship was found with positive affect. Contrarily, no significant relationship was found with negative affect. The second hypothesis is, therefore, partially accepted.

The analyses show support for the notion of psychological needs as a source of affect when interacting with business software, despite the second hypothesis being partially rejected. The results are mostly in line with research findings in other contexts of software use. Previous research identified a negative relationship – although weak – between overall need fulfilment and negative affect (Hassenzahl, et al., 2010), but there it was regarded as an unwanted side-effect. That study only considered positive experiences, and deemed negative affect as an unintentional outcome. Research into general life experiences did find a strong and consistent relationship (Sheldon, et al., 2001), and similar results have been replicated in the context of interaction with computers (e.g. Partala, & Kallinen, 2012). It is, thus, interesting to see these findings did not show here. A reason for this can be found in the considered timespan of UX, which will be discussed later under *Limitations*.

Moderation by competence and age. The moderation analyses showed to be less fruitful. For the third hypothesis, the amount of experience with the software was expected to moderate the relationship between affect and competence. The idea was that a more experienced user is more competent due to training effects, which decreases the importance of a feeling of competence. Nonetheless, neither the analyses with positive affect and negative affect returned significant results. This could indicate that the need for a feeling of competence is always important,

independent of the experience – and consequently training – with a product. It could imply for experienced users it is still important to be challenged, in order to create competence experiences.

The exploratory analyses into the moderation of the participants' age on the relationship between the fulfilments of autonomy, competence, and relatedness, only showed significant results for autonomy and relatedness. In both cases, it was in relationship with negative affect. The scatterplots for the different age groups show an interaction effect. For the groups with 26 – 35 years old and 46 – 55 years old, a low autonomy score is paired with a lower negative affect score, when compared to the scores of the groups of 36 – 45 years old and 56 – 65 years old. On the other hand, the former groups show a higher negative affect score for a higher autonomy score compared to the latter groups. This same interaction effect is found with the need of relatedness, except for the group of 26 – 35 years old. The scores for this group remain around the same negative affect score, independent of the degree of relatedness fulfilment.

These results could indicate the effect of fulfilling needs is different for people of different ages. At first sight, it seems the 36 – 45 years old and 56 – 65 years old react negatively on the fulfilment of autonomy and relatedness. However, despite the significant moderation found, the effect size is relatively low and the regression lines show heavy under- and overestimation of the scores. The effect that is observed is, thus, not strong and of low predictive value for individual cases.

Best predicting needs. In order to find which needs best predicted affect, regression analyses were performed. For positive affect, it appeared competence explained the most variance uniquely, whilst for negative affect autonomy is the best predictor. Nonetheless, in both cases the explained variance is low. The analysis with competence as an independent predictor of positive affect shows an increased explained variance to almost half, but its scatterplot shows many under- and overestimates. Furthermore, relatedness independently does not predict negative affect. As with the moderation analyses, the high intercorrelation between the needs presumably causes the effects of individual needs to become hazy. It makes it hard to distinguish the importance of distinct needs, which results in the low uniquely explained variance.

For most of the previous studies, a different set of needs was found to be most important. For example, in a study with Finnish students, autonomy was found to be the most important need (Partala, & Kallinen, 2012). Another study, with a general German population, found relatedness to be the best predicting need (Hassenzahl, et al., 2010). Although for general life experiences,

autonomy, competence, relatedness, and self-esteem were consistently found to be the most important (Sheldon, et al., 2001), it is no surprise that the relative importance varies for different participant populations. With specific user groups, come specific experiences with technology. A study which asked participants to rate an interaction with a product, found different types of interaction to fulfil different psychological needs (Hassenzahl, et al., 2015). In experiences in which a participant was spectator of someone playing a game with technology, relatedness was most fulfilled. On the other hand, when one was playing themselves, the fulfilment of competence increased. This is not necessarily merely a function of the experience itself; different experiences can be sought after and form a means to achieve satisfaction of a specific need (Sheldon, & Gunz, 2009). The population of the current study – a working class – might endeavour in using technology with other motivations than a group of students would. The business context implies the subjected interactions with the software came from a result-driven environment. It is, therefore, plausible that the participants experienced more competence satisfying experiences, because they were motivated to do so. Fulfilment of the other needs would have less an impact on the affective state, as the context mainly desires competence.

Questionnaire validation. Additionally, sufficient results have been found for the validity and reliability of the psychological needs questionnaire. Although some of the KMO values are of medium level, they are satisfactory to the standard for new questionnaires. Added to that, the reliability scores are high. This could indicate the translation is of good quality, and the results are in line with other studies using the original questionnaire (e.g. Sheldon, et al., 2001). For further use, it is advised to optimize the Dutch translation to improve the KMO values.

Limitations

Naturally, this study also faced problems – some of which not encountered by other studies. These can be roughly divided into statistical issues and methodological limitations, but cannot be seen separately; the former might be a result of the latter.

Negative affect floor effect. Firstly, negative affect shows a strongly, positively skewed distribution. A strong floor effect is observed, which decreases the ability to distinguish between responses on the lowest scores. It makes it troublesome for the psychological needs to predict the scores. The floor effect could indicate the statements – as asked for in the negative affect scale – were too negative for most of the participants. Accordingly, questions can be asked about the use of this scale in this specific context. As these results show, the statements might not be appropriate

in the context that is tested here. For this study, it is important to recognize such shortcomings for the negative affect analyses. It decreases the strength of the findings and could cause non-significant results. For future research, alternatives should be considered to increase the ability to distinguish respondents with low negative affect. The use and limitations of PANAS can, for example, be weighed against the use of the Self-Assessment Manikin (SAM; Bradley, & Lang, 1994). It is also used frequently in UX research with good results (Bargas-Avila, & Hornbæk, 2011).

Intercorrelation of the psychological needs. Another problem observed is the high intercorrelation between the psychological needs scores. It causes the effects of the psychological needs to be hard to segregate. This is mainly expressed in the degree to which variance of affect is uniquely explained by a specific psychological need. In the analyses these were found to be low, albeit the total variance explained was high.

Interestingly, earlier research has found low to moderate intercorrelations (e.g. Hassenzahl, et al., 2010). Hence, there can be mainly two reasons to have caused the problems: the translation of the psychological needs questionnaire, and differences in the approach. A specialised documentation team of the cooperating company performed the translation, but it was not tested in a pre-study. Furthermore, one difference with other studies, is the context of the user's experience (business versus personal). However, the context of use is thought to influence the relative importance of the needs only (Hassenzahl et al., 2010; 2015; Partala, & Kallinen, 2012; Sheldon, et al., 2001). There is no reason to think other contexts would increase the intercorrelation between the needs.

Timespan of UX. In contrast to this, a second difference with other studies can have important implications for the interpretation of all results. Instead of one specific experience (episodic UX), participants were asked to rate their experiences with the product on a typical day. This was chosen to represent the cumulative UX timespan. Reason for it was that if a person has to decide on using particular products on the long-term, information about past experiences retrieved from memory provides an important foundation for that decision (Schachter, et al., 2007; Oishi, & Sullivan, 2005). Human memory is not perfect, though, and is influenced by different processes. For example, for specific experiences, the emotional peak-value is the most positive or negative moment, whilst the emotional end-value is the emotional value at the end of the experience. It was found that these two values are important predictors for how an experience is

recalled (Do, Rupert, & Wolford, 2008). Although this research has not been conducted in the context of UX, the peak-end rule is assumed to affect experiences with interactive products as well (Norman, 2009).

The episodic and cumulative timespans do come with their own considerations. Both timespans might rely on different memory systems, depending on the point of time the remembered experience took place. In two experiments (Geng, Chen, Lam, & Zheng, 2013), the length of retention interval was found to influence the type of information used to reconstruct retrospective evaluations of events. For events that took place maximum 3 to 7 weeks ago, participants used episodic information, and were influenced by the peak-end rule. On the other hand, construction of older experiences was based on both episodic and semantic information, and not influenced by the peak-end rule. Episodic information refers to event-specific information, that is closely tied to a specific occurrence (e.g. "my experience yesterday with product X was pleasant"). Semantic information is not linked to a specific event, and consists of more abstract generalisations (e.g. "my experiences with product X are always pleasant"; Tulving, 1984). Effectively, it means the episodic and cumulative UX rely on different types of information, and can thus show different results.

The exact retention interval of the two different timespans is not clear, and it is thus dependent on the participants' experiences whether the reconstructed evaluations rely only on episodic information. Most other studies asked participants to evaluate an experience from the past 6 months, which means either memory type could be used. However, the formulation of a single experience might seduce participants to rely mostly on episodic memory. Contrarily, the formulation as used in this study ("a typical day"), might push towards a more general evaluation based on semantic memory. Unfortunately, the retention interval is not considered in this experiment. This could provide valuable information for future research.

Other contextual influences. Added to that, this study also did not consider other contextual influences on both affect and psychological need fulfilment. This poses serious threat for the interpretation of the results to be UX-related. There is a possibility the findings might be more job-related, being influenced by factors such as job satisfaction. As the software is used during the participants' workday, it is plausible to expect an impact of these factors on the measurements. The formulation of the questions' stem (On a typical day using [Name of the software], I feel…), might not zoom in enough on the use of the product. Although the notion of

UX includes the effect of the context (Roto, et al., 2011), it is important that the effects cannot be ascribed to the context alone.

Reductionist approach. Furthermore, this study took into account a reductionist approach to UX. This is a view that allows for describing experiences in single constructs – like affect. To the contrary, the holistic approach dictates that experiences should be studied in its whole, and not be reduced to (a selection of) measures. The idea is that any experience cannot be described in constructs, simply because of experiences' inherent complexity. An advantage of the reductionist approach, and motivation for this study to choose it, is that it allows for empirical testing and increases generalizability (Law, Van Schaik, & Roto, 2014). However, interpretations of the results might differ based on the approach to UX taken into account. If seen from a holistic approach, this study's results might merely relate to a small, if any, part of what can be called an experience.

Non-causality. Lastly, it is important to stress that all implied relationships here are of correlational nature. Causality was not addressed in this study. In new experiments, this is an important aspect to consider. This could not only provide theoretical support for the model, but can also bring about new implications for practitioners of UX design. If indeed a causal relationship is found, psychological needs could provide a point of intervention to establish an experience through the use of a product. This is valuable information for any designer focused on providing positive experiences with their designs.

Significance

Having conducted this experiment in a working context – as opposed to a personal, leisure context – an attempt has been made to further validate the notion of psychological needs as a source of UX-related affect. As UX is seen as highly contextual (Roto, et al., 2011), it is an important step to consider different situations of product use. The findings showed the fulfilment of psychological needs is strongly related with affect in the context of business software, and therewith helped in providing support for the framework. This does not only improve the field on a theoretical level, but also shows it is a usable, and practical, foundation for UX research in this particular context. As a result, the field is a step closer to achieving two of its main goals: establishing UX dimensions, and identifying ways to influence these.

Furthermore, this study set a step in the focus on cumulative UX; a time span not often reported yet (Bargas-Avila, & Hornbæk, 2011). With positive results having used this approach,

it might motivate future research to also consider cumulative UX. As stated before, there are substantial, theoretical reasons arguing for the focus on this particular time span. As mentioned, knowledge about how users remember their experiences, might provide insight into how they will decide to use a product on the long-term (Schachter, et al., 2007; Oishi, & Sullivan, 2005). Using the participant's memory to retrieve the experiences can help in gaining the knowledge needed to achieve this. By showing cumulative UX is a fruitful approach, this study hopefully enlightened relatively new methodologies.

Moreover, this study made a first step into providing a Dutch translation for the psychological needs questionnaire, as used in other studies (e.g. Hassenzahl, et al., 2015). The validity and reliability results indicate that it has potential to be used in a Dutch population, and thus enable future research to apply it. Although further improvement of the statements is advised, in its current form it seems to retain satisfactory levels of validity and reliability.

Future Research

Yet, the most important challenge for future research is to show causality of the relationships. This is an important assumption in the experiments up to this date, and needs to be investigated to continue improving UX research and design. Considerable data is gathered to support the notion of psychological needs as a source of affect, but causality has not yet been shown with experimental setups. Two ideas to achieve this are discussed.

The first is a study design which makes use of a priming effect in four different conditions, which participants are assigned to randomly. These different conditions reflect four different primes: autonomy, competence, relatedness, and control. To begin the experiment, a participant is shown a set of scrambled words. They have to unscramble them to form a sentence, which is related to the psychological need of the condition they are in. For example, a participant in the competence prime group receives the words: good, I, am, did, what, I, in. Unscrambled, the words form the sentence "I am good in what I did", and possibly primes the need for competence. Then, the participant has to think of an experience with technology, and rate it with the PANAS and the psychological needs questionnaire. The scores are compared to a control group, which is not primed. In a specific prime group, it is expected that experiences with high fulfilment of the corresponding need are reported more than in the other groups. The difference on affect between the primed group and the control group, could indicate the need fulfilment as source of affect. The score on the needs questionnaire could function as a manipulation check for the prime.

Unfortunately, this setup would not indicate the original strength of the relationship between the psychological need fulfilment and affect. Yet, it is an easy setup that can be executed without complex methods or tools.

The second setup would be to use a chatbot to fulfil different needs. In an experiment, participants are asked to chat with a bot, of which four different versions exist. The first three should fulfil the needs for, correspondingly, autonomy, competence, relatedness. A fourth version should be simple with a low to none fulfilment of needs. For example, in the autonomy condition, participants are asked many questions on things they want to do. The bot provides elaborate customisation features, which are commanded by the user via chat messages. A set of features, to satisfy specific needs, can be determined in a pre-study. The participant is then asked to rate the experience with the chatbot on the PANAS and psychological needs questionnaire. Similar to the previous idea, the affect in the need fulfilment conditions will be compared to the control group. Differences in these scores could indicate a causal relationship. The challenge in this approach lies mainly in the development of the different versions of the chatbot. It might not be easy to fulfil specific needs with specific functions or other aspects, so this requires elaborate work. Essentially, the bot could be approached as a new questionnaire which has to be developed. Albeit it does not measure something, an iteration process based on validity and reliability analyses could result in distinctive versions. Furthermore, the psychological needs questionnaire could function as a manipulation check after the interaction with one of the chatbots. Compared to the first proposed setup, the development of a bot is more complex – but hopefully yields better results.

Conclusion

All in all, this study aimed to identify the fulfilment of psychological needs – through the use of a product – as a source of affect. The focus was specifically on the context of business software, as this is a context that is not often studied yet. The results showed support for the notion of psychological needs as a source of affect, but was not able to specify for the effect of individual needs. Additionally, the translated psychological needs questionnaire was proven to retain sufficient validity and reliability properties. Having found these results in the context of business software, pleas for the use of broader criterions for assessing HCI in this specific context. As the traditional notion of usability is founded in the idea of computers as mainly work-related tools, actually obtaining these findings in a work-related context might indicate that it is time to shift the focus to UX measures.

Acknowledgments

Special thanks to both my supervisors, Anette Hiltunen and Pascal Haazebroek, for their supervision and support throughout the project. Without their knowledge and patience, this thesis could not have come to a successful completion.

References

Bargas-Avila, J., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, USA, 2689-2698.

Bradley M., & Lang P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49-59.

Deci, E.L., & Ryan, R.M., 2000. The ‘‘what’’ and ‘‘why’’ of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry 11*(4), 227–268.

Do, A. M., Rupert, A. V., & Wolford, G. (2008). Evaluations of pleasurable experiences: The peak-end rule. *Psychonomic Bulletin & Review, 15*(1), 96-98.

Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., & Van den Bergh, O. (2006). Verdere validering van de Positive and Negative Affect Schedule (PANAS) en vergelijking van twee Nederlandstalige versies. *Gedrag & Gezondheid*, *34*(2), 89-102.

Gagné, M. (2014). *The Oxford handbook of work engagement, motivation, and self-determination theory* (Oxford Library of Psychology). Oxford: Oxford University Press.

Geng, X., Chen, Z., Lam, W., & Zheng, Q. (2013). Hedonic evaluation over short and long retention intervals: The mechanism of the peak-end rule. *Journal of Behavioral Decision Making, 26*(3), 225-236.

Harmon-Jones, E., Gable, P. A., & Price, T. F. (2013). Does negative affect always narrow and positive affect always broaden the mind? Considering the influence of motivational intensity on cognitive scope. *Current Directions in Psychological Science, 22*(4), 301-307.

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect and interactive products – Facets of user experience. *Interacting with Computers, 22*, 353-362.

Hassenzahl, M., Wiklund-Engblom, A., Bengs, A., Hägglund, S., & Diefenbach, S. (2015) Experience-oriented and product-oriented evaluation: Psychological need fulfillment, positive affect, and product perception. *International Journal of Human-Computer Interaction, 31*(8), 530-544.

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Methodology in the Social Sciences). New York: The Guilford Press.

ISO. (2010). *ISO 9241-210. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407).* Switzerland: International Organization for Standardization (ISO).

Jordan, P. W. (2000). *Designing pleasurable products: An introduction to the new human factors.* London, Taylor & Francis.

Koca, A., Funk, M., Karapanos, E., Rozinat, A., & Van der Gaarden, N. (2008). Grasping product pragmatics: A case with internet on TV. In *Proceedings of the 1st international conference on Designing interactive user experiences for TV and video,* Silicon Valley, California, USA, 193-202.

Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior, 43*(1), 35-48.

Langdon, P., Lewis, T., & Clarkson, J. (2007). The effects of prior experience on the use of consumer products. *Universal Access in the Information Society, 6*(2), 179-191.

Law, E., Van Schaik, P, & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies, 72*(6), 526-541.

Lavie, T., & Tractinsky, N. (2003). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies, 60*(3), 269-298.

Mahlke, S., & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In *Proceedings of the CHI Conference on Emotion & Empathy (CHI 2007),* San Jose, USA, 915-918.

Müller, L. J., Mekler, E. D., & Opwis, K. (2016). Hedonic enjoyment and personal expressiveness in positive user experiences. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI 2016),* San Jose, California, USA, 3166-3172.

Norman, D. (2004). *Emotional design: Why we love (or hate) everyday things.* New York: Basic Books.

Norman, D. A. (2009). Memory is more important than actuality. *Interactions, 16*(2), 24-26.

Oishi, S., & Sullivan, H. W. (2005). The predictive value of daily vs. retrospective well-being judgments in relationship stability. *Journal of Experimental Social Psychology, 42*(4), 460-470.

Partala, T., & Kallinen, A. (2012). Understanding the most satisfying and unsatisfying user experiences: Emotions, psychological needs, and context. *Interacting with Computers, 24*, 24-34.

Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (2011). *User experience white paper: Bringing clarity to the concept of user experience.* Retrieved from http://www.allaboutux.org/files/UX-WhitePaper.pdf

Ryan, R. M, & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Schachter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews, 8,* 657-661.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information 44*(4), 695-729.

Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology, 80*(2), 325-339.

Sheldon, K. M., & Gunz, A. (2009). Psyhological needs as basic motives, not just experiential requirements. *Journal of personality, 77*(5), 1467-1492.

Tractinsky, N., & Hassenzahl, M. (2005). Arguing for aesthetics in human-computer interaction. *I-com, 4*(3), 66-68.

Tulving, E. (1984). Précis of elements of episodic memory. *The Behavioral and Brain Sciences, 7*(2), 223-238.

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly 28*(4), 695–704.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Zimmermann, P. G. (2008). *Beyond usability – Measuring aspects of user experience* (Doctoral dissertation). Retrieved from Networked Digital Library of Theses and Dissertations. (DISS. ETH NO. 17901)

Appendix A (Prime questions)

Instructies

Allereerst volgen 5 vragen over uw dagelijkse ervaringen met [Naam van software]. Geef per activiteit aan hoe u zich voelt.

Op een doorsnee dag tijdens gebruik van [Naam van software], voel ik mij bij het...

...gebruik van het verkoopboek

...gebruik van het inkoopboek

...verkrijgen van een financieel overzicht

...afletteren

...verwerken van uitstaande posten

Appendix B (Positive Affect Negative Affect Scale; PANAS)

Instructies

Deze vragenlijst bestaat uit een aantal woorden die verschillende gevoelens en emoties beschrijven. Geef voor elk woord aan in welke mate u zich op **een doorsnee dag tijdens gebruik van [Naam van software] zo voelt**. Lees elk woord en omcirkel daarnaast uw antwoord. Vul in op een schaal van 1 (= heel weinig) tot 5 (= heel veel).

		1 Heel weinig	2 Een beetje	3 Matig	4 Veel	5 Heel veel
1.	Geïnteresseerd	O	O	O	O	O
2.	Overstuur	O	O	O	O	O
3.	Uitgelaten	O	O	O	O	O
4.	Van streek	O	O	O	O	O
5.	Sterk	O	O	O	O	O
6.	Schuldig	O	O	O	O	O
7.	Angstig	O	O	O	O	O
8.	Vijandig	O	O	O	O	O
9.	Enthousiast	O	O	O	O	O
10.	Trots	O	O	O	O	O
11.	Prikkelbaar	O	O	O	O	O
12.	Alert	O	O	O	O	O
13.	Beschaamd	O	O	O	O	O
14.	Geïnspireerd	O	O	O	O	O
15.	Nerveus	O	O	O	O	O
16.	Vastberaden	O	O	O	O	O
17.	Aandachtig	O	O	O	O	O
18.	Rusteloos	O	O	O	O	O
19.	Actief	O	O	O	O	O
20.	Bang	O	O	O	O	O

Appendix C (Psychological needs questionnaire)

Instructies

Lees deze uitleg alstublieft zorgvuldig.

Op de volgende pagina's treft u stellingen aan die u helpen bij de evaluatie. Elke stelling geeft een gevoel weer. Deze gevoelens kunnen helpen bij het beschrijven van de kwaliteiten. Alle stellingen volgen op de basiszin: “Op een doorsnee dag gebruik van [Naam van software] voel ik...” Geef voor elke stelling aan in welke mate deze van toepassing is. Vul in op een schaal van 1 (= heel weinig) tot 5 (=heel veel).

Een voorbeeld

Op een doorsnee dag tijdens gebruik van [Naam van software] voel ik...

...dat ik iemand ben wiens advies wordt gevraagd en opgevolgd.

O	◐	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

Deze evaluatie geeft aan dat de stelling een beetje van toepassing is.

Besteed geen tijd aan nadenken over de stellingen. Probeer een spontaan antwoord te geven. Het kan zijn dat u vindt dat bepaalde stellingen het product niet goed beschrijven. Geef ook in zo'n geval toch antwoord.

Onthoud dat er geen goede of foute antwoorden zijn. Uw persoonlijke mening is wat telt!

Op een doorsnee dag tijdens gebruik van [Naam van software] voel ik...

...dat mijn keuzes gebaseerd zijn op mijn interesses en waarden.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat ik vrij ben om dingen te doen op mijn eigen manier.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat mijn keuzes mijn “ware ik” reflecteren.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat ik met succes moeilijke taken en projecten afrond.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat ik moeilijke uitdagingen aanga en overwin.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...mij bekwaam in wat ik doe.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...verbondenheid met mensen die om mij geven, en om wie ik geef.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...mij dichtbij en verbonden met mensen die belangrijk voor mij zijn.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...vertrouwelijkheid met de mensen waar ik tijd mee doorbreng.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat ik veel positieve kwaliteiten bezit.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...mij tevreden met wie ik ben.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...een sterk gevoel van zelfrespect.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat ik iemand ben wiens advies wordt gevraagd en opgevolgd.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat ik een sterke invloed heb op opvattingen en gedrag van anderen.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

...dat ik een sterke impact heb op wat anderen doen.

O	O	O	O	O
1	2	3	4	5
Heel weinig	Een beetje	Matig	Veel	Heel veel

Appendix D (Validity and reliability measures of all psychological need subscales)

Scale/Item	KMO	Factor Loading	Cronbach's α	α if item deleted
Autonomy	0.62		0.76	
… dat mijn keuzes gebaseerd zijn op mijn interesses en waarden.		0.78		0.75
...dat ik vrij ben om dingen te doen op mijn eigen manier.		0.90		0.53
...dat mijn keuzes mijn "ware ik" reflecteren.		0.79		0.74
Competence	0.65		0.73	
...dat ik met succes moeilijke taken en projecten afrond.		0.77		0.69
...dat ik moeilijke uitdagingen aanga en overwin.		0.86		0.53
...mij bekwaam in wat ik doe.		0.79		0.67
Relatedness	0.71		0.81	
...verbondenheid met mensen die om mij geven, en om wie ik geef.		0.88		0.70
...mij dichtbij en verbonden met mensen die belangrijk voor mij zijn.		0.85		0.74
...vertrouwelijkheid met de mensen waar ik tijd mee doorbreng.		0.83		0.77
Self-esteem	0.70		0.79	
...dat ik veel positieve kwaliteiten bezit.		0.81		0.77
...mij tevreden met wie ik ben.		0.84		0.73
...een sterk gevoel van zelfrespect.		0.87		0.67
Influence	0.71		0.81	
...dat ik iemand ben wiens advies wordt gevraagd en opgevolgd.		0.82		0.79
...dat ik een sterke invloed heb op opvattingen en gedrag van anderen.		0.87		0.72
...dat ik een sterke impact heb op wat anderen doen.		0.87		0.71

Note. KMO = Kaiser-Meyer-Olkin

Appendix E (Validity and reliability measures of positive affect and negative affect)

Scale/Item	KMO	Factor Loading	Cronbach's α	α if item deleted
Positive Affect	0.89		0.90	
…geïnteresseerd		0.67		0.89
…sterk		0.81		0.88
…enthousiast		0.80		0.89
…trots		0.74		0.90
…alert		0.53		0.88
…geïnspireerd		0.81		0.88
…vastberaden		0.80		0.88
…aandachtig		0.77		0.88
…actief		0.81		0.88
…uitgelaten		0.46		0.903
Negative Affect	0.92		0.953	
…overstuur		0.91		0.94
…van streek		0.85		0.95
…schuldig		0.79		0.95
…angstig		0.92		0.94
…vijandig		0.81		0.95
…prikkelbaar		0.78		0.953
…beschaamd		0.85		0.95
…nerveus		0.77		0.95
…rusteloos		0.85		0.95
…bang		0.91		0.95

Note. KMO = Kaiser-Meyer-Olkin

Appendix E (ANCOVA analyses of age moderation on autonomy, competence, and relatedness)

Predictor variable(s)	Outcome variable	d.f.	*F*	*P*	$\eta 2$
Autonomy	Positive affect	1, 120	93.1	<.001	.38
Age		3, 120	.92	.436	.01
Autonomy*Age		3, 120	1.31	.272	.02
Autonomy	Negative affect	1, 117	.001	.971	.00001
Age		3, 117	3.14	.028	.07
Autonomy*Age		3, 117	2.87	.040	.07
Competence	Positive affect	1, 120	98.6	<.001	.40
Age		3, 120	1.00	.392	.01
Competence*Age		3, 120	1.40	.235	.01
Competence	Negative affect	1, 117	.11	.740	.00009
Age		3, 117	2.44	.068	.06
Competence*Age		3, 117	2.31	.080	.05
Relatedness	Positive affect	1, 120	87.0	<.001	.37
Age		3, 120	1.59	.196	.02
Relatedness*Age		3, 120	1.97	.122	.03
Relatedness	Negative affect	1, 117	3.12	.080	.02
Age		3, 117	2.77	.044	.06
Relatedness*Age		3, 117	2.76	.045	.06