# Automatically tired?

-

# Evaluating the usability of a Sentence Completion Test to measure cognitive processes of fatigue

Bachelorthesis
Janis Sundermann (s1277243)

Faculty of behavioral sciences
Psychology
Department of Health Psychology

Primary Supervisor: Hanneke Kip
Secondary Supervisor: Christina Bode

Datum: 29.05.2015

UNIVERSITEIT TWENTE.

**Index**

## Abstract

**Objectives.** The main goal of this study was to scrutinize whether a Sentence Completion Test is suitable and useful to measure fatigue.

**Method.** An implicit measurement method, the Sentence Completion Test, was used to evaluate fatigue implicitly. To assess fatigue explicitly three explicit measurement methods, a Numeric Rating Scale (NRS), the Vitality/Fatigue subscale of the Research and Development 36-Item Health Survey (RAND-36), and the Checklist For Individual Strength (CIS-20-R) were additionally applied. In total, 50 participants completed the questionnaires. After completing the questionnaires the participant was asked to code his/her answers on the specific items of the implicit test as having regard to *fatigue* or *vitality* or none of them. In addition, the researcher coded the items independently. To evaluate the rater agreement between the three different raters a reliability analysis was conducted. At last, the Sentence Completion Test was correlated with each explicit test to evaluate possible relations.

**Results.** The data analysis showed weak to modest interrater-agreement between the three different raters. Further, the correlations between the implicit measurement methods and the explicit measurement methods showed no significant relation.

**Conclusion.** The results of the data analysis revealed only little indication that these implicit measurement methods are suitable and useful to measure fatigue. However, more methodologically accurate and well-investigated implicit measurement methods could be promising to measure fatigue more accurately and may serve as an useful alternative or addition next to existing standardized explicit measurements.

## Introduction

Fatigue is common complaint that was repeatedly reported by community and prime care studies. Studies from Kroenke & Price (1993) found high prevalent rates for fatigue of 23.6% in a US population. An epidemiological follow-up sample found that 20% of the population had a lifetime prevalence of inexplicable fatigue. Other studies indicated prevalence rates of 22% in a Norwegian sample (Loge, Ekegerg & Kaasa, 1998) and 18% in a UK sample (Pigeon, Sateia & Ferguson, 2003).

Fatigue is a complex and multidimensional phenomenon that involves a number of psychosocial and behavioral processes (Shen, Barbera & Shapiro, 2006). To scrutinize and explain the phenomenon fatigue and its sub-dimensions different approaches to define fatigue are needed. Firstly, the term fatigue is commonly used in everyday communication and according to English dictionaries defined as “extreme tiredness arising from mental or physical effort or illness” (Oxford Dictionaries, 2015). But, it is important to mention that dictionaries often do not distinguish between terms like sleepiness, tiredness and fatigue (Phillips, 2015). However, a discrimination between these terms is necessary because all of them bear reference to different phenomena or sub-dimensions of fatigue. In particular, sleepiness refers to a state of being sleepy, whereas tiredness describes a need for sleep or rest. At last, fatigue differs from the other terms in the way that it is attributed a specific cause, namely physical or mental effort (Phillips, 2015). Nevertheless, there is a huge overlap between these sub-dimensions which makes a clearly distinction often difficult. Therefore, an unclear usage of the definition of fatigue should be avoided and a definition with a multidimensional character could be helpful to establish a meaningful context. The different axes of the multidimensionality of fatigue bear reference to the subjectivity of experience of fatigue, psychosocial & physiological processes during fatigue and performance decrement at a physical and cognitive level due to fatigue (Bensing, Hulsman & Schreurs, 1999). Therefore

Phillips (2015) proposed a definition that incorporates the multidimensionality of the phenomenon fatigue:

> Fatigue is a suboptimal psychophysiological condition caused by exertion. The degree and dimensional character of the condition depends on the form, dynamics and context of exertion. The context of exertion is described by the value and meaning of performance to the individual; rest and sleep history; circadian effects; psychosocial factors spanning work and home life; individual traits; diet; health, fitness and other individual states; and environmental conditions. The fatigue condition results in changes in strategies or resource use such that original levels of mental processing or physical activity are maintained or reduced. (Phillips, 2015, p. 6)

Out of this, it becomes obvious that the concept fatigue has regard to several different sub-dimensions. Thus, the multidimensionality of fatigue leaves scope for different interpretations that all can have regard to the fatigue state but in different contexts. Moreover, next to divergent contexts or kinds of fatigue different time-dimensions of the fatigue state may play an significant role while measuring fatigue because fatigue may fluctuate in an individual (Yung & Wu, 2005). In particular, it needs to be questioned whether different time-dimensions, for example the state of fatigue during the past weeks, have impact on more recent states of fatigue because fatigue may fluctuate in an individual (Yung & Wu, 2005) and whether this is reflected in the measurement results. In conclusion, the fatigue state can be evaluated based on all these sub-dimensions which makes a precise definition and assessment of fatigue far more difficult. Nevertheless, there may be several sub-dimensions of fatigue but still they belong this to the same superordinate concept of fatigue in general.

Further, fatigue is a subjective experience, regardless of conceptualized as unitary or multidimensional phenomenon. Hence, to assess fatigue accurately, a measurement tool is needed that takes the multidimensionality, including sub-dimensions like different contexts or

time-dimensions, as well as the individual experience of fatigue into account. Explicit measurements were developed which meet these requirements (Vander Der Zee, Sanderman, Heyink, 1996; Vercoulen, Alberts & Bleijenberg, 1999, Hays & Morales, 2001) but their results are often distorted due to social desirability bias and the subjective evaluation of the respondent itself (Shen, Barbera & Shapiro, 2006).

In addition, fatigue is a non-specific symptom associated with significant level of physical and psychosocial morbidity. There are a number of acute and chronic illnesses that are consequently associated with fatigue like rheumatoid arthritis, cancer and multiple sclerosis. Fatigue can arise as primary symptom of these diseases or due to side effects of the required treatment (Wolfe &Michaud, 2004; Mollaoğlu & Üstün, 2009). Next to this, studies found that fatigue is often associated with psychological distress (Bültmann, Kant, Kasl et al., 2002) and depression (Chen, 1986). Additionally, fatigue is the primary symptom of chronic fatigue syndrome, an extreme and long lasting form of fatigue that endures about at least 6 months and is accompanied with several severe health complaints and life-limitations (Afari & Buchwald, 2014). Finally, fatigue is often one of the most important sources of disability and a prominent hindering and life-threatening condition (Shen, Barbera & Shapiro, 2006) which may affect the individual's performance and functioning in the occupational as well as in the home setting (Bültmann, Kant, Kasl et al., 2002). Nevertheless, there are only small proportions treated of those people who are affected by fatigue (Shen, Barbera & Shapiro, 2006). This fact implies that fatigue perhaps often is not diagnosed or recognized by physicians as physical or psychological disease. One of the contributing factors to this tendency is the intersection of different features of fatigue and sleepiness or tiredness (Shen, Barbera & Shapiro, 2006).

In addition to individual consequences, fatigue causes enduring substantial costs to the community and especially the health care system. A study from Buchwald, Pearlman, Umali et al. (1996) found that in Australia, where the health care system supports estimates of

financial costs, the chronic fatigue syndrome constitutes a loss to society of 59 million Australian Dollar per year. Further, at an individual level, financial costs of 9,429 $ per year arise for affected patients (Buchwald, Pearlman, Umali et al., 1996). Moreover, a study from Reynolds, Vernon, Bouchery & Reeves (2004) found that the chronic fatigue syndrome was responsible for an annual loss of productivity of 9.1 billion US-Dollar or about 20.000 US-Dollar per afflicted person.

As described above, fatigue is a highly prevalent phenomenon in general populations which also affects society. It is a common complaint that was repeatedly reported by community and prime care studies in working populations (Bültmann, Kant, Kasl et al., 2002). The high prevalence rates of fatigue (Kroenke & Price, 1993; Loge, Ekegerg & Kaasa, 1998) suggest that fatigue is a pervasive problem in society. Hence, this makes fatigue to an extreme financial and social burden to public health care and economy (Reynolds, Vernon, Bouchery & Reeves, 2004).

However, evaluating fatigue by measuring unconscious processes of fatigue could be useful alternative to existing explicit measurement methods. Therefore, it is also essential to briefly discuss the underlying biochemical processes that induce and proceed during fatigue. As a consequence, the procedure of assessing cognitive processes can be facilitated by evaluating these underlying processes in order to establish a context for an alternative measurement method. Firstly, the circadian rhythm alternates states of wakefulness and sleep of the human body. This 24-hour rhythm is mainly controlled by the suprachiasmatic nucleus (SCN) in the hypothalamus. The SCN regulates waking and sleeping by controlling the release of the hormone melatonin by the pineal gland. The secretion of melatonin influences circadian as well as circannual rhythms and makes the human body feeling sleepy (Kalat, 2015). Furthermore, diverse stimuli so-called ‘Zeitgeber’, for example light or noise, induce a reset of the circadian rhythm and prevent it from shifting (Rusak & Zucker, 1979). Secondly, fatigue can arise from other conditions like sleep deprivation, psychosocial stress,

physical/cognitive effort and the need for energy reservation/restoration or due to individual characteristics of a person (Irish, Kline, Gunn, Buysse & Hall, 2014). But, these conditions may not tell the whole story of the arising of fatigue. In particular, auditive cues (Ishii, Tanaka, Kim, Yamano & Watanabe, 2013), specific behavior before sleeping like lying on a bed or other stimuli in the environment of a person may serve as a trigger that induce fatigue due to a conditioned association between the trigger and fatigue. However, given that there could be other processes like implicit processes that induce fatigue, there is a need to fully understand the concept of fatigue. Unfortunately, there is rarely literature on which implicit and explicit cognitive processes proceed during fatigue.

Interestingly, a study from van der Linden, Frese & Meijman (2003) shows that respondents under fatigue that should resolve several cognitive tasks suffer more from compromised executive control than do not fatigued respondents. Also, tired respondents need prolonged planning time to solve the task and displayed more perseveration on one cognitive task. Unfortunately, other studies concentrate mainly on cognitive processes regarding the chronic fatigue syndrome and not on everyday fatigue of an individual. A study from Surawy, Hackmann, Hawton & Sharpe (1994) shows that patients suffering prolonged fatigue believe that their fatigue and the related symptoms are due to a disease with a physical etiology. The allocation of fatigue symptoms to a physical illness rather than to psychological factor leads to a misperception of their physical capabilities which ultimately provokes avoidance of activities (Knoop, Prins, Moss-Morris, Bleijenberg, 2010). Subsequently, it seems that there is a relation between cognitive processes and fatigue in the way that fatigue impairs cognitive processes and compromises executive control. However, it needs to be questioned whether implicit measurement methods can be used to measure fatigue and what kind of unconscious processes accompany during fatigue needs further research including this study. Assessing implicit processes that proceed during fatigue could be a meaningful and

helpful process to understand the complex phenomenon of fatigue, to scrutinize its causes and to help patients suffering fatigue with qualified and peculiarly adapted treatment.

Fatigue describes a condition as mainly subjective experience with a physiological and psychological character. Hence, the measurement of fatigue by means of questionnaires is also essentially subjective (Shen, Barbera & Shapiro, 2006). The measurement of fatigue mostly takes place by assessing the phenomenon via explicit measurement methods such as self-report questionnaires or physiological measurement methods (measuring heart rate, pupil dilating). Whereas physiological measurements of fatigue hold an objective evaluation of the fatigue state they deny the individual's perception of his/her fatigue (Phillips, 2015). Nevertheless, explicit measurements rely mostly on easily observed behaviors or on the subjective self-report evaluation of the respondent that are sensitive to bias. They evaluate explicit processing of events and are susceptible to impeding processes such as limitations of consciousness, distortion and social desirability (Stone, Bachrach, Jobe, Kurtmann et al., 2000). This needs to be taken into account while assessing the underlying processes during the fatigue state. Further, they are not suited to assess implicit and automatic processes within the respondent which may deliver meaningful insight in the underlying processes of fatigue. In contrast, implicit measurement methods are not open to introspection (Phillips, Hine & Thorsteinsson, 2008). Consequently, they could present a helpful and additional method to assess unconscious processes during fatigue next to explicit measurement methods because they can prevent distorted results of measuring fatigue due to bias.

Based on studies on dual process theory there are implicit processes that are relevant at similar concepts of fatigue like self-esteem or attitudes (Rooke et al., 2008; Rudman, Dohn & Fairchild, 2007; Rudman, Phelan & Heppen, 2008)). These phenomena are similar to fatigue because both are described by the subjective experiences of an individual. Consequently, implicit processes could play an essential role during fatigue. However, implicit processes are unconscious, automatic and not open to introspection. Whereas, explicit processes are

conscious, controlled, can be verbalized and usually are assessed by the means of questionnaires. The dual process theory implies that there are two distinct systems of information processing. At first, there is the experiential system that processes information effortless and rapid in a passive way and is intimately tied to the intuition and affect of the person. The second system is the rational-analytic system which operates intentional, logic-based and largely affect free (Phillips, Hine & Thorsteinsson, 2008). Based on the reason that explicit cognitions are bound to conscientiousness they are mostly evaluated by measures that need the respondent's thoughtful reflection and introspection. On the contrary, the automatic character of implicit cognitions, which are inaccessible for the respondent's introspection and thus not under his/her conscious control (e.g. reaction times, memory association), implies assessment under passive conditions (Phillips, Hine & Thorsteinsson, 2010, Gawronski, Lebel & Peters, 2007). These implicit processes may be a more accurate method to measure fatigue because they could show an undistorted and unbiased view of an individual's fatigue state. Consequently, it should be worth discussing the scrutiny of implicit cognition during fatigue.

To emphasize the relevance of implicit cognition during fatigue there is a need to consider other research regarding resembling phenomena to fatigue. Several studies show that a dual-process approach has been effective to implicit cognitions regarding attitudes and self-esteem (Rooke et al., 2008; Rudman, Dohn & Fairchild, 2007; Rudman, Phelan & Heppen, 2008). For example, a study from Rudman & Kilianski (2000) shows that implicit measurements of attitudes toward female authority assess similar constructs as do explicit measurements but can prevent the distortion of measurements due to bias. This encourages the assumption that implicit measurements can be adapted to measure implicit cognition regarding conceptions or states like fatigue because attitude or self-esteem are both phenomena described by subjective experiences of an individual and therefore delivering a related concept to fatigue. In contrast, results from a study from Greenwald & Farnham (2000), where the relation between explicit and implicit self-esteem was investigated by

means of the IAT (Implicit Association Test) and explicit measurements, show significant low correlations between these different methods of measurements. This result indicates that implicit methods of self-esteem measure a construct that is distinct, but correlated with, particularly the same construct of self-esteem assessed by explicit measurements (Greenwald & Farnham, 2000). However, these results support the suggestion that explicit measurements can be susceptible to limitations of awareness like during the fatigue state and social desirability and because of that show dissociations with implicit measurements. As a result, implicit measurements may often show dissociations in certain characteristics with explicit measurements but also associations because they are measuring the same construct. All together these results indicate that implicit measurement can give valuable insights in the phenomenon of fatigue and the related implicit cognitions by using free association to bypass conscious control of the individual in order to attain an unbiased concept of the fatigue state.

There are several methods to measure implicit cognitions. One of them is the sentence completion task. The sentence completion task refers rather to a group of tests that follow a similar format than to one individual test. It is a method to assess a respondent's perception, attitudes and implicit cognitions towards a certain construct in an indirect and semi-structured manner. Sentence completion tasks use incomplete sentences that the respondent needs to complete to evaluate a person's cognitions (Roger, Bishops & Lane, 2003). It is assumed that the indirectness of sentence completion simplifies the respondent's expression of attitudes, cognitions or feeling towards a certain construct or phenomenon. Further, the completion of the given sentence represents the cognitive association with a construct in question. At last, the respondent remains completely unaware of the test procedure (Roger, Bishops & Lane, 2003). Sentence completion tests were also commonly and successfully used to assess personality, attitudes or ego development of subjects (Rhode, 1946; Renner, Maher & Campbell, 1962; Lichtenstein, Pruski, Marshall, Blalock, Lee & Plaetke, 2003; Weiss, Zilberg & Genevro, 1989).

Hence, the question arises how to measure whether implicit and to what extent implicit processes are involved during fatigue. Therefore, within this study explicit measurements as well as implicit measurement are conducted to evaluate the relation between these processes in the light of fatigue.

In conclusion, this study, at first hand, aims to examine the multidimensionality of fatigue and its several underlying sub-dimensions that leave scope for different interpretations. Therefore, a interrater-reliability analysis is conducted to assess to what extent the different sub-dimensions correspond with each other dependent on evaluations of three different raters. Due to the fact that the sub-dimensions are still part of the main concept of fatigue, we expect a high agreement between the raters. Secondly, we want to scrutinize the ability of implicit measurement methods to measure fatigue in relation to explicit measurement methods. In particular, we want to examine whether different time-dimensions assessed by explicit measurements correlate positively with the results of the Sentence Completion Test because fatigue may fluctuate within in an individual and the fatigue state during the past weeks may have impact on a more recent state of fatigue. To specify the goals of this study we postulated 3 hypotheses:

(1) The interrater-reliability between the self-coding of the respondents and the coding of the researchers is high ($k > 0.60$)
(2) The results of explicit measurement regarding the momentary state of fatigue and the results of Sentence Completion Test correlate highly ($r \geq 0.7$)
(3) The results of explicit measurements regarding the state of fatigue during the past weeks and the results of the Sentence Completion Test correlate highly ($r \geq 0.7$)

# Methods

## Respondents

The sample of this study was established by using two different sampling methods: convenience sampling and snowball sampling. Hence, the students asked to participate were mainly friends and acquaintances of the researchers and thus were part of their direct social contacts. Additionally, those students as already existing study subjects, were asked to recruit further subjects from among their acquaintances. Inclusion criteria required that the respondent is over 18 years old, student and able to speak English at university level to fully understand the instructions and to completing the questionnaires. Hereby, the researchers build on the statements of the respondents regarding their ability to speak/understand the English language. In order to gather respondents, students of the University of Twente were asked face-to-face or via social media to participate in this study. Finally, for all students of the University of Twente it was possible to apply for the study via an online-subscribing research website of the University of Twente (utwente.sona-system.com) with their student accounts. By participating in the study the respondents have been able to gain reward points which are mandatory to complete their study requirements and are credited on their student account. Hereby, 3 students (6%) applied for the study via utwente.sona-system.com. In total, 50 respondents participated in the study, of which 56% were male and 44% were female. The age varied from 18 to 28, with a mean of 22.80 and a standard deviation of 1.88.

## Measurement Instruments

The questionnaire was a paper and pencil questionnaire. In total, the questionnaire (see Appendix B) consists of four different parts. Firstly, the respondent was asked to provide information over his/her age and gender. Secondly, two different implicit association tests, one word association test and one sentences completion test were taken to evaluate subconscious processes with regard to fatigue. In this study only the sentence completion test

was analysed because the Word Association Test is part of another similar bachelor thesis which made use of the same data gathered in the collection process. Hence, the Word Association Test will not further be described. To avoid bias, the order of the two implicit tests was randomized. Consequently, some respondents received at first the sentence completion test and then the word association test whereas other respondents received the tests in the reversed order. Thirdly, three different explicit measurements, the RAND-36, CIS-20-R and the Numeric Rating Scale, were used to assess fatigue via self-report. Lastly, the participant needed to classify his/her answers on the implicit tests, if applicable, on the basis of three different categories.

**Sentence Completion Test.**

A sentence completion test is a method to assess a respondent's implicit cognitions towards a certain construct in an indirect and semi-structured manner by using indirect sentences that respondent needs to complete (Roger, Bishops & Lane, 2003). In order to construct the Sentence Completion Test the researchers made use of several studies that used an Sentence Completion Test to assess personality, attitudes or ego development (Rhode, 1946; Renner, Maher & Campbell, 1962; Lichtenstein, Pruski, Marshall, Blalock, Lee & Plaetke, 2003; Weiss, Zilberg & Genevro, 1989). In total, the sentence completion test in use consisted of twenty different items (incomplete sentence stems), from which ten items were ambiguous sentence stems that may elicit associations related to fatigue, and ten control items. To avoid bias, the order of the twenty items was randomized. The generating process of the ambiguous items mostly took place by considering everyday situations and contexts that had regard to fatigue which were then reflected in incomplete sentence beginnings that could be interpreted in several different ways (as having regard to fatigue/vitality or another subject). The developing process of the items was supported by an expert for implicit test construction. Further, a pilot-test was conducted to evaluate the created items and to assess whether the items are appropriate to measure fatigue. Because the pilot test proceeded without

complications no further changes were applied. Further, the ambiguity of the items should guarantee that only tired respondents classify their answers as having regard to fatigue in contrast to not tired respondents. The ambiguous items had a distinct relation to fatigue and referred mainly to activities that take place prior to sleeping, activities that are related to physical/mental exhaustion and activities that may evoke feelings of fatigue (see table 1 below for some examples).

Tabel 1

*Examples of incomplete sentence beginnings (ambiguous items) of the Sentence Completion Test*

| **prior to sleeping** | **physical/mental exhaustion** | **evoking feelings of fatigue** |
|---|---|---|
| "When I lay down…" | "After a busy day…" | "After dinner…" |
| "On the sofa…" | "After a lecture…" | "In the evening…" |
| "When I close my eyes…" | "Learning for my exams…" | "Staying out really late…" |

Additionally, ten control items without any relation to fatigue were used in order to get possible comparisons between the ambiguous and control items. Hence, the participants were asked to complete the incomplete sentence beginning immediately with whatever comes to their mind in order to get a full sentence.

**Word Association Test.**

A word association test is compiled to assess participant's attitudes and cognitions towards fatigue and how he/she reacts to ambiguous words with regard to fatigue. The Word Association Test was similar to the Sentence Completion Test but used only single words as items instead of incomplete sentence beginnings.

**Numeric Rating Scale.**

The first explicit test that was used to evaluate fatigue is a numeric rating scale. In order to assess the momentary state of fatigue, the participant was asked to set a cross on a given line

indicating his/her subjective estimation of the construct (Hartrick, Kovan & Shapiro, 2003). The Numeric Rating Scale ranged from the extreme on the left "not tired at all" (0) to the extreme on the right "very tired"(10). Studies have shown that the Numeric Rating Scale is a valid measurement that participants find easier to use and are more responsive to a NRS than to comparable scales, like visual analogue scales, or verbal rating scales (Ferreira-Valente, Pais-Ribeiro & Jensen, 2011; Hjermstad, Fayers, Haugen, Caraceni, Hanks, Loge, et al., 2011).

**Research and Development 36-Item Health Survey (RAND-36; The Finnish Medical Society Duodecim, 2001).**

Secondly, the Energy & Fatigue subscale of the RAND-36 was used, which evaluated the participant's fatigue during the last four weeks (Hays & Morales, 2001). It consists of four items, which are supposed to be answered on a six-point Likert scale. The participant is asked to indicate how often during the last four weeks he/she had feelings of fatigue or vitality, for example "Did you feel full of pep?" for vitality or "Did you feel tired?" for fatigue. The Energy & Vitality subscale of the RAND-36 was used in the study because it showed a high reliability, with $\alpha = 0{,}82$ in a study from Vander Der Zee, Sanderman & Heyink (1996). Furthermore, the RAND-36 is a valid instrument with high sensitivity, indicating good psychometric qualities (Vander Der Zee et al., 1996).

**Checklist Individual Strength (CIS-20-R).**

To assess fatigue during the last two weeks with another instrument, participants were at last asked to complete the CIS-20-R. The CIS-20-R is a 20-item self-report questionnaire and evaluates the level of fatigue during the last two weeks (Vercoulen, Alberts & Bleijenberg, 1999). It is composed of four subscales, namely, *Subjective Feeling of Fatigue, Concentration, Motivation,* and *Physical Activity*. Participants were asked to indicate how often the given statements apply to how they have felt during the past two weeks on a seven-point Likert scale. In a study from Beurskens, Bültmann, Kant, Vercoulen, Bleijenberg &

Swaen (2000) the CIS-20-R showed a high reliability with a Cronbach's $\alpha = 0.90$ and holds a sufficient discriminant, as well as convergent validity. Due to a good psychometric quality (Beurskens et al.,2000), this scale was used.

**Self-Coding.**

At the end of the questionnaire, the participant received the instruction to classify his/her answers on the implicit tests. On the basis of his/her own interpretation, the participant needed to categorize every given answer on every item of the implicit tests into one of three categorizations, namely "*has regard to fatigue*", "*has regard to vitality*", or "*does not have any regard to one of these terms*". Indication took place by setting a "+" for "*has regard to fatigue*", a "-" for "has regard to vitality", and a "0" for "*does not have any regard to one of these terms*" beside each given association.

*Procedure*

**Test Conditions.**

The study took place in a quiet room in the library of the University of Twente. During the testing, only the two researchers and the respondent were present to avoid interruptions and to guarantee the best possible test conditions. Respondents applied for the study via utwente.sona-system.com and can choose between different time slots. The duration of the whole test procedure was scheduled for 30 minutes. Completing the whole questionnaire took approximately 15 minutes; the remaining 15 minutes were used for the debriefing and possible questions of the respondent. For respondents that did not apply via utwente.sona-system.com a meeting was scheduled.

**Test Process.**

Before test start, the respondent received an informed consent printed on paper (see appendix A) regarding anonymity and privacy of his/her personal data. With agreeing on the informed consent by signing it, the respondent received the questionnaire. On the first page of the questionnaire the participant could find instructions with regard to test procedure and order of

the questionnaires. Following the order of the questionnaires was mandatory. The respondent received sufficient time to read the instructions and, if questions arise, to ask the researchers. Additionally, the participant was told that during the testing it is allowed to ask questions regarding test procedure, but not regarding the content of the questionnaires. If all potential questions were answered, the respondent started with the questionnaire.

**Debriefing.**

During the debriefing the participant received information about the main aim of the study whether implicit measuring methods are appropriate to measure fatigue. Furthermore, the respondents received the information that the two implicit tests contained control items, as well as ambiguous items, which were designed to evoke unconscious associations that might have regard to fatigue and vitality. The aim of self-coding was to determine whether the associations of the respondent had regard to fatigue or vitality or had no regard to the two terms at all, according to the respondent. This process served as indicator whether fatigue and vitality are indeed distinct sub-dimensions of fatigue or are related to each other.

**Data Analysis**

All statistical analysis used in this study was executed via IBM's software package SPSS (Statistical Package for Social Sciences), version 21. All relevant variables were tested for normality using the Saphiro-Wilk Test (Razali & Wah, 2011). For all scales descriptive statistics (mean & standard deviation) were calculated. For all variables that were not normal distributed, the median was calculated instead of the mean, because it happened to be a more robust measure for not normally distributed data (Pappas & DePuy, 2004). See table 2 for the distributions of normality of the single variables. If the p-value was significant ($p<0.05$) the considered variable was not normally-distributed, but if the p-value was not significant ($p>0.05$) the variable was normally distributed. Referring to table 2 only the variable CIS-20-R was normally distributed, the remaining variables were not normally distributed.

Table 2

*Shapiro-Wilk test for distributions of normality of the single variables*

| | Shapiro-Wilk | | |
|---|---|---|---|
| | Statistic | df | Sig. |
| Numeric Rating Scale | .937 | 50 | .010 |
| SCTFatigue (Respondent) | .959 | 50 | .080 |
| SCTVitality (Respondent) | .927 | 50 | .004 |
| SCTTotal (Respondent) | .898 | 50 | .000 |
| SCTFatigue (Researcher) | .944 | 50 | .020 |
| SCTVitality (Researcher) | .836 | 50 | .000 |
| RAND-36 | .934 | 50 | .008 |
| CIS-20-R | .965 | 50 | .150 |

**Implicit Tests.**

The first step of the data analysis of the implicit tests had regard to the coding of the answers on every particular item. The answers were classified independently by the respondent and both researchers into three different categories: “has regard to fatigue”, “has regard to vitality” and “does not have any regard to one of these terms”. For both implicit tests, sum scores of the frequency of the two classifications of associations were calculated. Additionally, a total score of the sum scores of the two classifications, fatigue and vitality was compiled. Afterwards, an interrater-reliability analysis using Cohen’s Kappa (manually) was executed in order to assess the correspondence of the coding between the raters.

**Correlation Explicit & Implicit Tests.**

In total, 9 correlations were calculated between the two implicit and the three explicit measures to find out possible relationships. For every correlation, the coding of the respondents was used, because it might show the highest degree of validity. Firstly, the

fatigue respectively vitality scores, as well as the sum score of both classifications of the sentence completion test were correlated with the scores of each particular explicit scale (SCT vs. NRS, SCT vs. RAND-36, SCT vs. CIS-20-R). For variables with a normal distribution, Pearson´s correlation coefficient was used; for variables that did not have a normal distribution, Spearman's rank correlation coefficient (rho) was applied. Pearson's correlation coefficient was calculated for the correlation between the variable SCTFatigue (Sentence Completion Test, Respondent, Fatigue scale) and the CIS-20-R. Spearman's rank correlation coefficient (rho) was calculated for all remaining correlations between the SCT and the explicit measurements (SCTFatigue vs. NRS, SCTVitality vs. NRS; SCTFatigue vs. RAND-36, SCTVitality vs. RAND-36; SCTVitality vs. CIS-20-R). Lastly, three correlations between the explicit measurements were calculated to evaluate the validity and reliability of the NRS, RAND-36 and CIS-20-R using Spearman's correlation coefficient rho.

**Qualitative/Quantitative study.**

The completion of the incomplete sentence beginnings on the Sentence Completion Test indicated a qualitative data analysis. However, due to the self-coding process by the respondent of the given answers, as having regard to vitality or fatigue or none of them, a quantitative data set was compiled by attributing a defined value ("+" for fatigue, "-" for vitality or "0" for none of them) to the given answer. As a consequence, a quantitative data analysis was conducted.

# Results

## Descriptive Statistics

The descriptive statistics of all questionnaires are presented in table 3 below. In comparison with other studies which assessed fatigue explicitly by means of the RAND-36 and the CIS-20-R we found slightly lower scores for the vitality scale of the RAND-36 for the general population but considerably higher scores of fatigue on the CIS-20-R in a student population. In particular, considering this study the median for the vitality scale of the RAND-36 was 59.7 with a standard deviation of 17.77, whereas a study from Van der Zee & Sandermann (2012) found a mean score of 67.4 with a standard deviation of 19.9 in the general population. Moreover, we found a mean score of 96.96 with a standard deviation of 14.74 on the CIS-20-R, in contrast, a study from Bültmann, Kant, Kasl, Beurskens & van de Brandt (2002) found a mean score of 57.2 with a standard deviation of 23.7 in the general population. Hence, the results of this study indicate slightly lower state of vitality and considerably higher state of fatigue among students in comparison with the general population.

Table 3

*Descriptive Statistics of all questionnaires (Sentence Completion Test, Numeric Rating Scale, RAND-36, CIS-20-R), N=50*

| | Mean | Median | Std. Deviation | Min. | Max. |
|---|---|---|---|---|---|
| SCT Fatigue | 5.5200 | 6.0000 | 1.80972 | 2.00 | 9.00 |
| SCTVitality | 2.0600 | 2.0000 | 1.55721 | .00 | 6.00 |
| SCTTotal | 7.5800 | 8.0000 | 1.88539 | 3.00 | 10.00 |
| NRS | 4.48 | 4.00 | 2.314 | 0.00 | 10.00 |
| RAND-36 | 59.7000 | 65.0000 | 17.76864 | 20.00 | 85.00 |
| CIS-20-R | 96.9600 | 98.0000 | 14.73017 | 62.00 | 121.00 |

***Interrater-reliability***

In order to assess the correspondence of the coding on the sentence completion test (SCT) between the respondent and the researchers themselves an interrater reliability analysis using Cohen's Kappa was manually executed, apart for the *fatigue* and *vitality* items and *combined*. For the *fatigue* items moderate correlations were found between the coding of the respondent, researcher 1 and researcher 2 indicating modest agreement (see figure 1 below for further details). Moreover, low correlations were found regarding the *vitality* items and thus indicating little agreement between the three raters (see figure 1 below for further details). At last, for the agreement regarding the *combined* scale of *fatigue* and *vitality* low to moderate agreement was indicated by the correlations between the three raters (see figure 2 on p. 24 for further details).

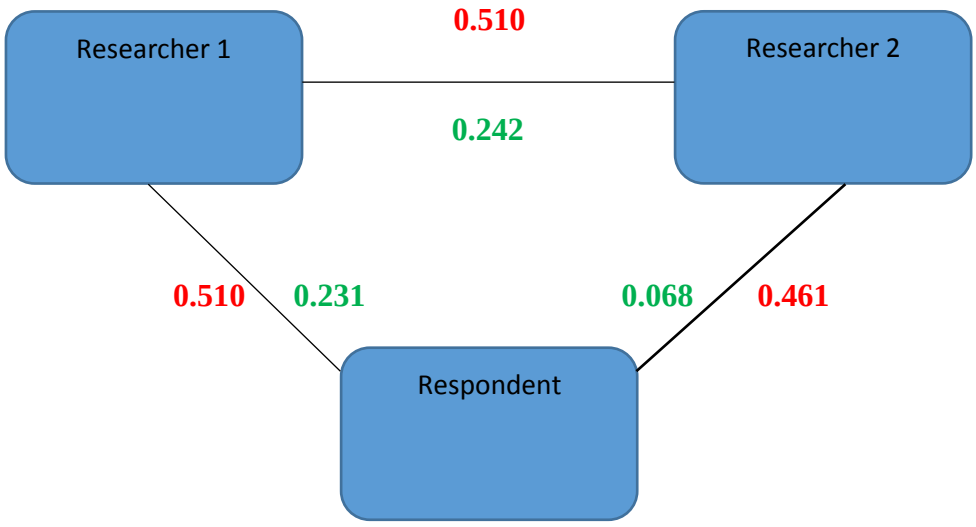



*Figure 1:* Interrater-reliability regarding fatigue (in red) and vitality (in green), measurements are given in Cohen's Kappa κ

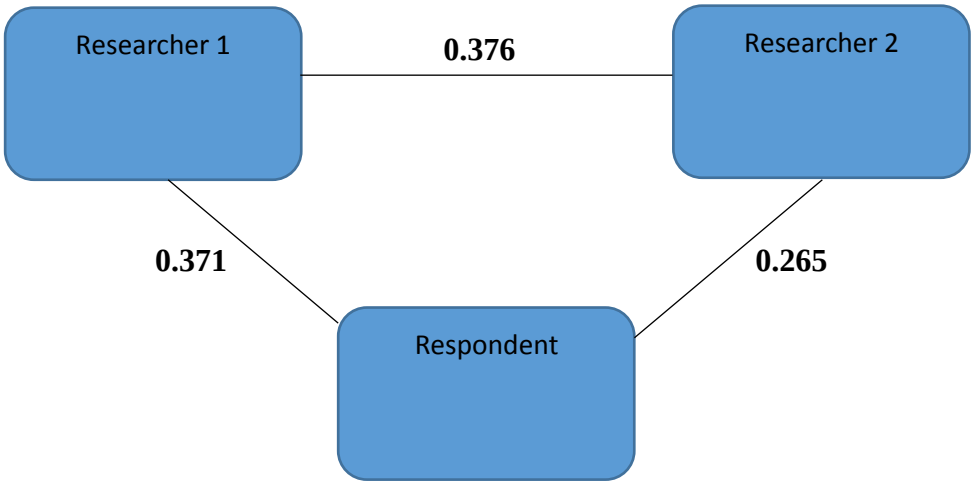



*Figure 2:* Interrater-reliability regarding both scales, fatigue and vitality, measurements are given in Cohen's Kappa κ

**Correlation, Numeric Rating Scale vs. Sentence Completion Test**

To scrutinize a possible relation between the NRS and the SCT the sum scores of both tests were correlated using Spearman's rho. The following analysis revealed no significant correlations at all between the NRS and the SCT considering the *fatigue* and *vitality* items as well both scales combined (see table 4 for further details).

**Correlation, RAND-36 vs. Sentence Completion Test**

Again, to assess a possible relation between the RAND-36 and the SCT Spearman's rho was calculated, apart for the *fatigue* and *vitality* items. The results of the analysis revealed no significant correlations at all between the RAND-36 and the SCT considering the fatigue and vitality items as well as both scales combined (see table 4 for further details).

**Correlation, Checklist Individual Strength vs. Sentence Completion Test**

Due to the normal-distribution of the sum scores of the CIS-20-R and of the SCT regarding *fatigue* Pearson's correlation coefficient was used to assess a possible connection between these scores. The following analysis revealed no significant correlations between the CIS-20-

R and the SCT considering the fatigue items (see table 5 for further details. In addition, considering the *vitality* items and both scales combined Spearman's rho was used because these scores showed no normal-distribution. Hereby, no significant correlations were found between the CIS-20-R and the SCT considering the vitality items as well as both scales combined (see table 4 for further details).

**Correlations between the explicit measurements (NRS, RAND-36, CIS-20-R)**

To evaluate the validity and reliability of the three explicit measurements, Numeric Rating Scale, RAND-36 and CIS-20-R, correlations using Spearman's rho were calculated. Firstly, there was no significant correlation between the NRS and the RAND-36 (see table 6 for further details. Secondly, there was a significant negative correlation between the NRS and the CIS-20-R with $r_s$ = -,427, p = .002. Thirdly, there was a significant correlation between the CIS-20-R and the RAND-36 with $r_s$ = .659, p = .000.

For an overview of the correlations between the SCT and the explicit measurements; see table 4 for the non-parametric correlations and table 5 for the parametric correlation. For an overview of the correlations between the explicit measurements see table 6.

Tabel 4

*Correlations Sentence Completion Test (both scales FAT/VIT) vs. explicit tests (NRS, RAND-36 and CIS-20-R)*

| | | | SCTFatigue | SCTVitality | SCTTotal | NRS | RAND-36 | CIS-20-R |
|---|---|---|---|---|---|---|---|---|
| Spearman's rho | SCTFatigue | Correlation Coefficient | 1.000 | -.389** | .580** | .219 | -.084 | -.177 |
| | | Sig. (2-tailed) | . | .005 | .000 | .127 | .562 | .219 |
| | SCTVitality | Correlation Coefficient | -.389** | 1.000 | .456** | -.169 | -.014 | -.027 |
| | | Sig. (2-tailed) | .005 | . | .001 | .241 | .921 | .853 |
| | SCTTotal | Correlation Coefficient | .580** | .456** | 1.000 | .039 | -.129 | -.168 |
| | | Sig. (2-tailed) | .000 | .001 | . | .788 | .373 | .244 |

**. Correlation is significant at the 0.01 level (2-tailed).

Tabel 5

*Correlation Sentence Completion Test (FAT) vs. CIS-20-R, N=50*

| | | SCTFatigue | CIS-20-R |
|---|---|---|---|
| SCTFatigue | Pearson Correlation | 1 | -.165 |
| | Sig. (2-tailed) | | .253 |

Tabel 6

*Correlations between explicit measurements (Numeric Rating Scale, RAND-36 and CIS-20-R)*

| | | | NRS | RAND-36 | CIS-20-R |
|---|---|---|---|---|---|
| Spearman's rho | NRS | Correlation Coefficient | 1.000 | -.133 | -.427** |
| | | Sig. (2-tailed) | . | .359 | .002 |
| | RAND-36 | Correlation Coefficient | -.133 | 1.000 | .659** |
| | | Sig. (2-tailed) | .359 | . | .000 |
| | CIS | Correlation Coefficient | -.427** | .659** | 1.000 |
| | | Sig. (2-tailed) | .002 | .000 | . |

**. Correlation is significant at the 0.01 level (2-tailed).

## Discussion

The main goal of this study was to scrutinize whether the sentence completion test as implicit measurement is appropriate and useful to measure the phenomenon fatigue.

The usefulness of the Sentence Completion Test to measure fatigue was evaluated by assessing a possible relation between the results of three standardized and validated explicit measurements and the newly constructed Sentence Completion Test. Additionally, to evaluate the validity and reliability of the Sentence Completion Test and the three explicit measurements an interrater-reliability analysis and correlations between the three explicit measurements were conducted.

The main findings of this study were (i) that there was a moderate agreement between the three raters on the coding of the items regarding fatigue; with regard to the coding of the vitality items only a low agreement between the three raters was found; at last, combining both scales, fatigue and vitality low to moderate agreement between the three raters was found. Consequently, the 1st hypothesis cannot be fully confirmed. (ii) that there was no significant correlation at all between the results of the explicit measurements and the results of the implicit measurement regarding fatigue, independent of time period of measurement, thus, disagreeing with the 2nd and 3rd hypothesis;

### Evaluation

The results revealed that there is no significant correlation between the explicit measurements and the implicit measurement, the Sentence Completion Test, regarding the momentary state of fatigue as well as regarding the state of fatigue during the past weeks. This result indicates that the Sentence Completion Test either did not measure fatigue as intended because the designed items were not applicable to measure fatigue or the Sentence Completion Test measured a different dimension of fatigue than the explicit measurements. These problems are discussed in the following.

The concept of fatigue is a highly complex phenomenon that unifies several different axes of dimensions, regarding the physical and mental state, psychosocial stress and the health status of an individual (Phillips, 2015). Therefore, this multidimensionality of fatigue facilitates a large scope for interpretation when it comes to the coding of particular items as having regard to fatigue or vitality or none of them. For example, considering the meaning of the item "Staying out really late…" could be interpreted in two ways. At first, the respondent might interpret the item as being awake really late and therefore being tired which would be considered as having regard to fatigue. In contrast, the respondent might understand the item in the context of social activities like outgoing in the evening with other people which would be considered as having regard to vitality. Consequently, the item intended to measure fatigue had have regard to two different concepts which may be related to each other, but were nevertheless interpreted in different ways by the raters resulting in a low agreement. Further, the explicit and implicit measurements may differed in the measured dimension of fatigue. Due to the fact, that the coding of the respondent represents an essential part in the evaluation process a large scope for interpretation on one item could result in insignificant outcomes because each individual differs from the other regarding the interpretation of one item. Hence, interpretations of the same item by three different coders may be contrary to each other or have no relation at all, resulting in a diverging concept of fatigue with several different dimensions and therefore in a poor interrater-reliability because the raters held different concepts of fatigue. This fact holds true for the coding process of the respondent and the researchers as well as for designing process of the items by the researchers. The designing process of the items could be impeded due to the fact that the researchers themselves may hold different interpretations of the concept of fatigue. Therefore, designing items intended to measure fatigue and that are additionally ambiguous to disguise the intended construct seemed to be a difficult endeavor. At last, the designed items did not seem to bear reference to the multidimensionality of the concept fatigue, rather they were developed to measure fatigue

in general.

In addition, explicit measurements are often susceptible to bias and are constrained to limitations of awareness (Stone, Bachrach, Jobe, Kurtmann et al., 2000). Thus, the results of the explicit measurement may be distorted because the subjective notions of an individual about his/her fatigue state could have impact on his/her answers on the questionnaires resulting in a subjective concept of fatigue. Whereas, implicit measurements can bypass consciousness, identify the true or "other" concept of fatigue and are not sensitive to subjectivity (Phillips, Hine & Thorsteinsson, 2008). Hence, the results of the implicit measurement may represent a raw concept of fatigue not influenced by subjective notions of the respondent. Therefore, the results between the explicit and the implicit measurements may show no significant correlation at all. This means not necessarily that both measurements methods are measuring different phenomena but they are rather measuring different sub-dimensions of fatigue. A similar result was found in a study from Greenwald & Farnham (2000). In this study explicit and implicit measurement methods were used to assess self-esteem and only significant low correlations between the different kinds of measurement regarding self-esteem were found. Hence, this results indicates that implicit measurement of self-esteem measure something different from what is measured by the explicit measurements of self-esteem. In particular, implicit measurements could measure constructs that distinct, although correlated with, nominally the same constructs as measured by explicit measurements (Greenwald & Farnham, 2000). Further, the low correlations between the results of explicit and implicit measurements may be emerged from the different measurement qualities of both methods regarding limitations of awareness. This becomes relevant while measuring the phenomenon fatigue, a condition that refers to a limited state of awareness of an individual. The mere fact that fatigue limits the awareness of the individual is not taken into account by explicit measurements, but implicit measurements may compensate this explanatory gap by bypassing consciousness representing an unbiased concept of fatigue, not

influenced by the fatigue state of the individual. In contrast, a study from Rudman & Kilianski (2000) that assessed attitudes towards female authority, found that implicit measurements evaluate similar constructs as do explicit measurements. Unfortunately, the present study could not fully confirm this hypothesis regarding the measurement of fatigue. Hence, the difference in results of this study might be evolved from measuring different sub-dimensions of the main phenomenon fatigue. Assessing the complex phenomenon of fatigue by implicit measurements requires that the measurement method bears reference to all sub-dimensions of fatigue. Nevertheless, the results of the studies of Greenwald & Farnham (2000) and Rudman & Kilianski (2000) enhance the notion that implicit measurement methods can be useful to assess constructs that are related to subjective experience as do fatigue or attitudes. Finally, the results of this study indicate that different concepts of fatigue were measured resulting in insignificant correlations between the explicit measurements and the Sentence Completion Test.

Further, the correlations between the explicit measurements themselves suggest that the measured-time dimensions of the three different questionnaires have influence on the correspondence between Numeric Rating Scale, the RAND-36 and the CIS-20-R. The low negative correlation between the NRS and CIS-20-R points out that the momentary state of fatigue is independent from the fatigue state in the past weeks. This result contradicts a study from Yung & Wu (2005) which found that fatigue can fluctuate within an individual. Further, the time references of the explicit measurements, momentary state of fatigue and state of fatigue in the past weeks, may have regard to two different time dimensions, similar to the concepts of states and traits (McAdams, 1990). States refer to temporary conditions or behaviors that depend on situations and motives at a particular time (McAdams, 1990), similar to the momentary state of fatigue of the individual that refers to the condition of fatigue at the exact moment of measurement assessed by the Numeric Rating Scale. Whereas traits have regard to characteristic dispositions that are consistent and long lasting (McAdams,

1990), similar to a prolonged fatigue state over several weeks measured by RAND-36 and the CIS-20-R. This comparison suggests that the momentary state of fatigue is highly dependent on the situation of the individual. For example, before completing the Numeric Rating Scale the individual might have exercised or learned for an exam resulting in physical or mental fatigue that is reflected in the results of the Numeric Rating Scale. In contrast, the fatigue state of the past weeks, evaluated by the RAND-36 and the CIS-20-R, reflects a more consistent and long lasting state of fatigue that might have evolved from illness or psychological stress over a long time (Wolfe &Michaud, 2004; Mollaoğlu & Üstün, 2009; Philipps, 2015). Indeed, the high correlations between the RAND-36 and CIS-20-R indicate that both measurement methods measure the same construct of fatigue in relation to a prolonged state of fatigue. In conclusion, the Numeric Rating Scale respectively the RAND-36 and the CIS-20-R measure indeed the phenomenon fatigue (Hays & Morales, 2001; Vercoulen, Alberts & Bleijenberg, 1999) but dependent on different time dimensions resulting in low negative correlations between each other.

In addition, the items of the Sentence Completion did not refer to a specific time dimension of the fatigue as do the Numeric Rating Scale (momentary state of fatigue), or the RAND-36 or CIS-20-R (state of fatigue during the past weeks) resulting in a disproportion of results. Rather, the items of the SCT had regard to specific situations, behaviors, thoughts or attitudes that may represent habits or thoughts of the respondents independent of a specified time dimension.

At last, regarding the used implicit measurement method it needs to be mentioned that the construction of the Sentence Completion Test was rarely based on a theoretical foundation because no literature existed to adapt correspondingly to the given aim of the study. In addition, no item pool existed to gather suitable ambiguous items measuring fatigue. Therefore, the researcher made them up by themselves for designing the Sentence Completion Test. The insufficient theoretical foundation and the no existing item pool may have resulted

in a low validity of the Sentence Completion test. Hence, the Sentence Completion Test may have not measured the construct of fatigue as indicated by the low correlations between the explicit measurements and the Sentence Completion Test.

**Limitations**

This study had several limitations with respect to sample selection, measurement instruments and theoretical foundation. At first, the sample consisted only of students of the University Twente. Other participants with a different occupation or a population with a different age distribution were not included in the sample of this study. Because of this no representative sample of the general population existed. As a consequence the results of this study have only regard to the student population. A study from Lee, Chien & Chen (2007) found high prevalence rates of fatigue among graduated students. Risk factors included lack of physical activity (Lee, Chien & Chen, 2007), irregular sleep cycles and insufficient sleep (Kang & Chen, 2009). The general population might differ from the student population in extent and kinds (sub-dimensions) of fatigue. Therefore, considering a representative sample of the general population is recommended. At last, the study made among others use of snowball-sampling to gather participants. Hence, most participants were Psychology students at the same university like the researchers making a measurement bias due to scientific knowledge about implicit tests possible. This becomes relevant if the gathered respondents recognize while completing the Sentence Completion Test the measured construct. Consequently, the Sentence Completion Test could not bypass consciousness and is biased to the subjective attitude of the individual regarding his/her fatigue state. Hence, the results of the implicit measurement would be distorted and do not represent the concept of fatigue as intended.

As described above the construction process of the Sentence Completion Test lacked sufficient theoretical foundation and an appropriate item pool for items that measure fatigue. To compensate the missing theoretical foundation and the item pool an expert for test construction was consulted and several studies using a Sentence Completion Test to measure

similar concepts of fatigue like self-esteem or attitude (Greenwald & Farnham, 2000; Rudman & Kilianski, 2000) were studied to develop appropriate items to measure fatigue.

Finally, in this study it became obvious that both concepts, fatigue and vitality, show similarities as well as differences in features dependent on the interpretation of both concepts by the raters. Interpretations of fatigue by three different raters holding different interpretations of the concept fatigue and vitality themselves or by referring to different sub-dimensions of the complex phenomenon of fatigue resulted in a poor interrater-reliability making a reliable conclusion about the usefulness of a sentence completion test to measure fatigue hardly possible. In relation to the coding process proceeded by the two researchers and the respondent it is important to mention that the test instructions how to code the items appropriately was sometimes unclear and confusing for the respondent because they did not understand how exactly to code their answers based on three categories, or to be exact they did not grasp the semantic meaning of the two concepts, fatigue and vitality or the difference between them. The researchers received several questions regarding the coding of the items, unable to answer them adequately because the participants would be biased if they receive further information about the test. Hence, several participants coded the items not as intended by the researcher. For example, some participants tried to code as many items as possible as fatigue or vitality. As a consequence several items were coded as fatigue or vitality but had only little relation to these terms. Therefore, it was not possible to measure fatigue by means of these items appropriately. Rather, the items should only be coded as having regard to fatigue or vitality if there is a distinct relation between the given answer and one of the constructs, otherwise the items should be coded as not having regard to any of these terms. For a future study these test instructions should be made explicit and clear to the participant to avoid misinterpretation and misperformance by instructing the respondents only to code their answers as having regard to fatigue or vitality if there is a clear relation to one of these terms. Furthermore, survey fraud or social desirability are most threatening factors to distort

measurement results of self-report questionnaires (Stone, Bachrach, Jobe, Kurtmann et al., 2000). This becomes relevant if, for example, participants felt ashamed because the researchers could see afterwards what they answered on particular items regarding their of fatigue or, especially for the Sentence Completion Test, what they have thought while completing the test. Hereby, participants could felt inhibited and, consequently, tended to give social desirable answers that did not reflect the true answers of the participants like if there was no researcher present or anonymity was completely guaranteed. In the present study, this problem might have resulted in answers that did not represent the unconscious association which the respondent truly had while completing the test. As a consequence, the given answers were not necessarily related to the fatigue state of the respondent as it was intended.

**Future Research**

One of the main obstacles and the essential part of the evaluation process of the Sentence Completion Test represents the coding of the respondent as well as the coding of the researchers. The diverging concept of fatigue and due to that the large scope for interpretation of fatigue terms resulted in a poor interrater-reliability. Therefore, more research is needed to improve the coding process itself and thereby the interrater-reliability. For example, studies from Torstrick, McDermut, Gokberk, Bivona, & Walton (2015) and Feher, Vandecreek, & Teglasi (1983) using an alternative version of a Sentence Completion Test, the Rotter Incomplete Sentence Blank, found a near-to-perfect interrater-agreement between the investigator and the interviewer after they received a special coding training. Based on this results a coding training for the researcher could improve the interrater reliability. Furthermore, after receiving an appropriate training the researcher would be able to give concise instructions to the respondent avoiding misunderstandings or misperformance. At last, there is more research needed to design suitable items that are intended to measure fatigue while disguising the construct intended to measure. In the present study, ambiguous items were used that had regard to fatigue as well as vitality at the same time, for example “Staying

out really late", resulting in an intersection of both concepts. As a consequence, there was no clearly distinction between both concepts possible and the respondents did not know in which way to code the item. Rather, the items should be ambiguous in order to distinguish between fatigue and a not-fatigue-related concept respectively between vitality and a not-vitality-related concept. Further, these ambiguous items require scientific validation. For this purpose it could be useful to interview physicians or psychologists working in sleep medicine to attain a theoretical basis in order to construct an item pool for the concept fatigue and to classify behaviors, thoughts or attitudes representing a specific state of fatigue that can be used in the developing process of the ambiguous items. Additionally, the items need to represent situations or conditions that have a direct relation to fatigue as well as a relation to a situation or condition that has no relation at all to fatigue to guarantee ambiguity. Otherwise, the Sentence Completion Test cannot circumvent consciousness and reveals the construct intended to measure.

Furthermore, in this study two different constructs, fatigue and vitality, were assessed by the means of the Energy and Vitality subscale of the RAND-36 (Hays & Morales, 2001) and the coding process of items of the Sentence Completion Test. Both concepts show distinct relations as well as intersections that make the evaluation process of these constructs extremely difficult. Hence, in future research studies should concentrate on one construct and exclude constructs that show similarities and dissimilarities at the same time. However, vitality does not need to be a completely different concept than fatigue. Rather, vitality could represent one concept of the multidimensional phenomenon of fatigue, but these sub-dimension needs to be considered in the developing process of the Sentence Completion Test in order to gather items that measure different concepts of fatigue by constructing ambiguous items that are related to every sub-dimensions of fatigue.

At last, strong points of this study include the easy implementation of the questionnaires and the possible measurement of a true and unbiased concept of fatigue not

influenced by the subjective attitude of the individual regarding his/her fatigue state by revealing unconscious processes during fatigue. Further, results of this study indicate that it is indeed possible to measure fatigue implicitly. But, careful investigation of the complex phenomenon of fatigue and appropriate items that measure fatigue and that bear reference to the single sub-dimensions of fatigue are needed.

**Conclusion**

The data analysis gave only little indication that implicit measurements method are a suitable and useful instrument to measure the construct fatigue. But, the low correlations might be caused by measuring different sub-dimension of fatigue. Therefore, intersections and dissimilarities of distinct sub-dimensions of fatigue made a precise measurement of the intended construct extremely difficult. However, more methodologically accurate and well-investigated implicit tests could be promising to measure fatigue more accurately and may serve as an alternative or addition to the existing standardized explicit measurements if they bear reference to the multidimensionality of the complex phenomenon fatigue. Additionally, it might be possible to assess only one sub-dimension of fatigue by the implicit measurement preventing an intersection of different dimensions. However, a special coding-training for the interviewer is advised to improve the interrater-reliability to guarantee concise instructions for the respondent and to constrain the scope for interpretation during the coding process. At last, there is more research necessary based on a theoretical foundation to construct an implicit test with scientifically validated items to measure fatigue.

## References

Afari, N., & Buchwald, D. (2014). Chronic fatigue syndrome: a review.

Bensing, J. M., Hulsman, R. L., & Schreurs, K. M. (1999). Gender differences in fatigue: biopsychosocial factors relating to fatigue in men and women.*Medical care*, *37*(10), 1078-1083.

Buchwald, D., Pearlman, T., Umali, J., Schmaling, K., & Katon, W. (1996). Functional status in patients with chronic fatigue syndrome, other fatiguing illnesses, and healthy individuals. *The American journal of medicine*, *101*(4), 364-370.

Bültmann, U., Kant, I., Kasl, S. V., Beurskens, A. J., & van den Brandt, P. A. (2002). Fatigue and psychological distress in the working population: psychometrics, prevalence, and correlates. *Journal of psychosomatic research*,*52*(6), 445-452.

Bültmann, U., Kant, I. J., Van Den Brandt, P. A., & Kasl, S. V. (2002). Psychosocial work characteristics as risk factors for the onset of fatigue and psychological distress: prospective results from the Maastricht Cohort Study.*Psychological medicine*, *32*(02), 333-345.

Bültmann, U., Kant, I., Kasl, S. V., Beurskens, A. J., & van den Brandt, P. A. (2002). Fatigue and psychological distress in the working population: psychometrics, prevalence, and correlates. *Journal of psychosomatic research*,*52*(6), 445-452.

Chen, M. K. (1986). The epidemiology of self-perceived fatigue among adults. *Preventive medicine*, *15*(1), 74-81.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241.

Feher, E., Vandecreek, L., & Teglasi, H. (1983). The problem of art quality in the use of human figure drawing tests. *Journal of clinical psychology*, *39*(2), 268-275.

Ferreira-Valente, M. A., Pais-Ribeiro, J. L., & Jensen, M. P. (2011). Validity of four pain intensity rating scales. *PAIN®*, *152*(10), 2399-2404.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us?: scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*(2), 181-193.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of personality and social psychology*, *79*(6), 1022.

Hartrick, C. T., Kovan, J. P., & Shapiro, S. (2003). The numeric rating scale for clinical pain measurement: a ratio measure?. *Pain Practice*, *3*(4), 310-316.

Hays, R. D., & Morales, L. S. (2001). The RAND-36 measure of health-related quality of life. *Annals of medicine*, *33*(5), 350-357.

Hjermstad, M. J., Fayers, P. M., Haugen, D. F., Caraceni, A., Hanks, G. W., Loge, J. H., ... & European Palliative Care Research Collaborative (EPCRC. (2011). Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *Journal of pain and symptom management*, *41*(6), 1073-1093.

Irish, L. A., Kline, C. E., Gunn, H. E., Buysse, D. J., & Hall, M. H. (2014). The role of sleep hygiene in promoting public health: A review of empirical evidence.*Sleep medicine reviews*.

Ishii, A., Tanaka, M., Iwamae, M., Kim, C., Yamano, E., & Watanabe, Y. (2013). Fatigue sensation induced by the sounds associated with mental fatigue and its related neural activities: revealed by magnetoencephalography.*Behav Brain Funct, 9*, 24.

Kalat, J. (2015). *Biological psychology*. Cengage Learning.

Kang, J. H., & Chen, S. C. (2009). Effects of an irregular bedtime schedule on sleep quality, daytime sleepiness, and fatigue among university students in Taiwan. *BMC Public Health*, *9*(1), 248.

Knoop, H., Prins, J. B., Moss-Morris, R., & Bleijenberg, G. (2010). The central role of cognitive processes in the perpetuation of chronic fatigue syndrome.*Journal of psychosomatic research*, *68*(5), 489-494.

Kroenke, K., & Price, R. K. (1993). Symptoms in the community: prevalence, classification, and psychiatric comorbidity. *Archives of Internal Medicine*,*153*(21), 2474-2480.

Lee, Y. C., Chien, K. L., & Chen, H. H. (2007). Lifestyle risk factors associated with fatigue in graduate students. *Journal of the Formosan Medical Association*, *106*(7), 565-572.

Lichtenstein, M. J., Pruski, L. A., Marshall, C. E., Blalock, C. L., Lee, S., & Plaetke, R. (2003). Sentence completion to assess children's views about aging. *The Gerontologist*, *43*(6), 839-848.

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45-65.

Loge, J. H., Ekeberg, Ø., & Kaasa, S. (1998). Fatigue in the general Norwegian population: normative data and associations. *Journal of psychosomatic research*, *45*(1), 53-65.

McAdams, D. P. (1990). *The person: An introduction to personality psychology*. Harcourt Brace Jovanovich.

Mollaoğlu, M., & Üstün, E. (2009). Fatigue in multiple sclerosis patients. *Journal of clinical nursing*, *18*(9), 1231-1238.

Oxford Dictionaries (2015). Definition of fatigue in English. Oxford Dictionaries Online. Retrieved 24.02.2015.

Phillips, R. O. (2015). A review of definitions of fatigue–And a step towards a whole definition. *Transportation Research Part F: Traffic Psychology and Behaviour*, *29*, 48-56.

Phillips, W. J., Hine, D. W., & Thorsteinsson, E. B. (2010). Implicit cognition and depression: A meta-analysis. *Clinical Psychology Review*, *30*(6), 691-709.

Pigeon, W. R., Sateia, M. J., & Ferguson, R. J. (2003). Distinguishing between excessive daytime sleepiness and fatigue: toward improved detection and treatment. *Journal of psychosomatic research*, *54*(1), 61-69.

Ramsay, A. M. (1988). *Myalgic encephalomyelitis and postviral fatigue states: the saga of Royal Free disease*. Gower Medical for the Myalgic Encephalomyelitis Association.

Renner, K. E., Maher, B. A., & Campbell, D. T. (1962). The validity of a method for scoring sentence-completion responses for anxiety, dependency, and hostility. *Journal of Applied Psychology*, *46*(4), 285.

Rohde, A. R. (1946). Explorations in personality by the sentence completion method. *Journal of Applied Psychology*, *30*(2), 169.

Rogers, K. E., Bishop, J., & Lane, R. C. (2003). Considerations for the use of sentence completion tests. *Journal of Contemporary Psychotherapy*, *33*(3), 235-242.

Rooke, S. E., Hine, D. W., & Thorsteinsson, E. B. (2008). Implicit cognition and substance use: a meta-analysis. *Addictive Behaviors*, *33*(10), 1314-1328.

Rudman, L. A., Dohn, M. C., & Fairchild, K. (2007). Implicit self-esteem compensation: Automatic threat defense. *Journal of Personality and Social Psychology*, *93*(5), 798.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and social psychology bulletin*, *26*(11), 1315-1328.

Rudman, L. A., Phelan, J. E., & Heppen, J. B. (2007). Developmental sources of implicit attitudes. *Personality and Social Psychology Bulletin*, *33*(12), 1700-1713.

Rusak, B. & Zucker, I. (1979). Neural regulation of circadian rhythms. *Physiological reviews, 59*(3), 449-526.

Shen, J., Barbera, J., & Shapiro, C. M. (2006). Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep medicine reviews, 10*(1), 63-76.

Soames-Job, R. F., & Dalziel, J. (2000). Defining fatigue as a condition of the organism and distinguishing it from habituation, adaptation and boredom. In P. A. Hancock & P. A. Desmond (Eds.), Stress, workload and fatigue. NY: CRC Press.

Stone, A. A., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.). (1999). *The science of self-report: Implications for research and practice*. Psychology Press.

Torstrick, A., McDermut, W., Gokberk, A., Bivona, T., & Walton, K. E. (2015). Associations Between the Rotter Incomplete Sentences Blank and Measures of Personality and Psychopathology. *Journal of personality assessment*, (ahead-of-print), 1-12.

Van der Zee, K., & Sanderman, R. (1993). RAND-36. *Groningen: Northern Centre for Health Care Research, University of Groningen, the Netherlands, 28*, 2e druk (2012).

Van der Zee, K. I., Sanderman, R., Heyink, J. W., & de Haes, H. (1996). Psychometric qualities of the RAND 36-Item Health Survey 1.0: a multidimensional measure of general health status. *International journal of behavioral medicine*, *3*(2), 104-122.

Vercoulen, J. H. M., Alberts, M., & Bleijenberg, G. (1999). De checklist individuele spankracht (CIS). *Gedragstherapie, 32*, 131-136.

Weiss, D. S., Zilberg, N. J., & Genevro, J. L. (1989). Psychometric properties of Loevinger's Sentence Completion Test in an adult psychiatric outpatient sample. *Journal of personality assessment, 53*(3), 478-486.

Wolfe, F., & Michaud, K. (2004). Fatigue, rheumatoid arthritis, and anti-tumor necrosis factor therapy: an investigation in 24,831 patients. *The Journal of rheumatology*, *31*(11), 2115-2120.

Yang, C. M., & Wu, C. H. (2005). The Situational Fatigue Scale: A different approach to measuring fatigue. *Quality of Life Research, 14*(5), 1357-1362.

# Appendix

## Appendix A, Informed Consent

UNIVERSITY OF TWENTE.

### Study Information Sheet

| | | | |
|---|---|---|---|
| **Title of Project:** | Measuring psychometric quality | **Ethics Approval Number:** | 15155 |
| **Investigator(s):** | Dion Schlesiger<br>Janis Sundermann | **Researcher Email:** | d.h.schlesiger@student.utwente.nl<br>j.h.sundermann@student.utwente.nl |

Aims of the Study:
Measuring of psychometric quality of a questionnaire

Eligibility Requirements:
Students over 18, able to speak English

What you will need to do and time commitment:
Completion of a questionnaire, including self-report measurements. The whole test procedure will endure approximately 30 minutes.

Confidentiality of your data:
All data will be treated anonymously and will not be published

Details of any payments/credits (*must be approved by ethics committee)*
0,5 SONA credits

Remember that participation in this research study is completely voluntary. Even after you agree to participate and begin the study, you are still free to withdraw at any time and for any reason.

If you would like a copy of this consent form to keep, please ask the researcher. If you have any complaints or concerns about this research, you can direct these, in writing, to the secretary of the Ethics Committee Faculty Behavioral Sciences of the University of Twente, J. Rademaker (phone: 053-4894591, e-mail: j.rademaker@utwente.nl, Postbus 217, 7500AE Enschede)

UNIVERSITY OF TWENTE.

## RESEARCH INFORMED CONSENT FORM

| | | | |
|---|---|---|---|
| **Title of Project:** | Measuring psychometric quality | **Ethics Approval Number:** | |
| **Investigator(s):** | Dion Schlesiger<br>Janis Sundermann | **Researcher Email:** | d.h.schlesiger@student.utwente.nl<br>j.h.sundermann@student.utwente.nl |

Please read the following statements and, if you agree, initial the corresponding box to confirm agreement:

| | Initials |
|---|---|
| I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. | |
| I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. | |
| I understand that my data will be treated confidentially and any publication resulting from this work will report only data that does **not** identify me. | |
| I freely agree to participate in this study. | |

Signature:

| | | |
|---|---|---|
| Name of participant (block capitals) | Date | Signature |
| Dion Schlesiger<br>Researcher 1 (block capitals) | Date | Signature |
| Janis Sundermann<br>Researcher 2 (block capitals) | Date | Signature |

If you would like a copy of this consent form to keep, please ask the researcher. If you have any complaints or concerns about this research, you can direct these, in writing, to the secretary of the Ethics Committee Faculty Behavioral Sciences of the University of Twente, J. Rademaker (phone: 053-4894591, e-mail: j.rademaker@utwente.nl, Postbus 217, 7500AE Enschede)

# Appendix B, Questionnaire

**Research Questionnaire – Bachelor thesis**

Thank you for participating in this study. This questionnaire consists of four different parts including personal information, a word association test, a sentence completion test, and three self-report questionnaires. The test procedure is going to endure approximately 30 minutes. All collected data is treated anonymously and will not be published. At the beginning of every part, instructions will be presented. Please follow these instructions and maintain the chronological question order.

In case of questions or further interest in the study, please contact us:

Dion Schlesiger: d.h.schlesiger@student.utwente.nl
Janis Sundermann: j.h.sundermann@student.utwente.nl

## 1. Demographic variables

Age: _____ years

Gender: male / female

Study: _____

**2a)**

Write the first word you think of next to each word. For example, if the word "doctor" is presented, you might write "nurse". Work as quickly as possible; write the first thing that comes to your mind.

| | |
|---|---|
| Blanket | |
| Reading | |
| Family | |
| Workout | |
| Lamp | |
| Music | |
| House | |
| Alarm clock | |
| Battery | |
| Handkerchief | |
| Shopping | |
| Water | |
| Energy | |
| Fruit | |
| Table | |
| Television | |
| Tree | |
| Exam | |
| School | |
| Morning | |

**2b)**

Please complete each of the following incomplete sentences. Like before, write the first thing down that comes to your mind.

a) In the evening _____

b) After a lecture _____

c) A healthy diet _____

d) When I close my eyes _____

e) Drinking alcohol in the weekend _____

f) Homophobia makes me feel _____

g) Working in a group with other students _____

h) I've never been to _____

i) When I lay down _____

j) After dinner _____

k) On the sofa _____

l) When I meet my friends _____

m) Staying out really late _____

n) After a busy day _____

o) Vegetarians are _____

p) I often go _____

q) I like _____

r) My university is _____

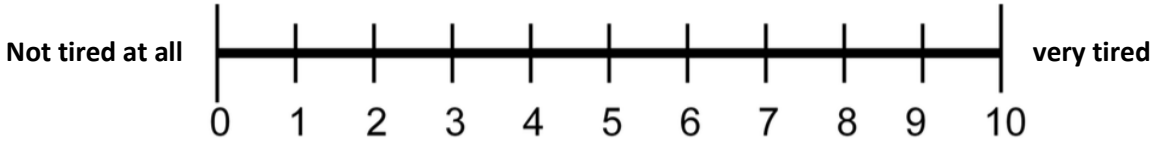
s) Learning for my exams _____

t) One of my favorite sports is _____

**3. Self-report questionnaire**

**a) Numeric rating scale**

How tired are you right now? Please indicate the intensity of your current state of fatigue by setting a cross on one of the numbers on the presented scale.




**b) RAND-36**

These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the past 4 weeks…

| | All of the time | Most of the time | A good bit of the time | Some of the time | A little of the time | None of the time |
|---|---|---|---|---|---|---|
| Did you feel full of pep? | | | | | | |
| Did you have a lot of energy? | | | | | | |
| Did you feel worn out? | | | | | | |
| Did you feel tired? | | | | | | |

**c) Checklist Individual Strength (CIS20R)**

Instruction: In the following you find 20 statements. With these statements we wish to get an impression of how you have felt during the past two weeks. Do not skip any statement and place only one cross for each statement in one of the boxes.

| **FROM** 1 = "No, not at all" **TO** 7 = "Yes, totally" | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| I feel very tired | | | | | | | |
| I feel very active | | | | | | | |
| Thinking requires effort | | | | | | | |
| Physically I feel exhausted | | | | | | | |
| I feel like doing all kinds of nice things | | | | | | | |
| I feel fit | | | | | | | |
| I do quite a lot within a day | | | | | | | |
| When I am doing something, I can concentrate quite well | | | | | | | |
| I feel weak | | | | | | | |
| I don't do much during the day | | | | | | | |
| I can concentrate well | | | | | | | |
| I feel rested | | | | | | | |
| I have trouble concentrating | | | | | | | |
| Physically I feel I am in a bad condition | | | | | | | |
| I am full of plans | | | | | | | |
| I get tired very quickly | | | | | | | |
| I have a low output | | | | | | | |
| I feel no desire to do anything | | | | | | | |
| My thoughts easily wander | | | | | | | |
| Physically I feel in a good shape | | | | | | | |

**4.**

Please classify your answers on part 2a) and part 2b).

1. If your answer has regard to “fatigue”, please indicate this by setting a “**+**” beside your answer.

**OR**

2. If your answer has regard to “vitality” please indicate this by setting a “**-**” beside your answer.

**OR**

3. If your answer does not have any relation with these two terms, please indicate this by setting a “**0**” beside your answer.

Classify your answers on the basis of your **own** subjective interpretation (“what did I mean with my answer?”).

Thank you very much for participating!
In the following you will get further information about the context of this study.