Happy software developers solve problems better: psychological measurements in empirical software engineering

While software plays a crucial role in today's society, the industry and the academics have not been able to solve the so called *productivity gap*, where the need for software grows faster than our capability to deliver it. It has often been claimed that the most significant quality and productivity gains are achieved by focusing on people. This claim has rarely been verified in software engineering research, which is a young discipline that studies the aspects of the software construction process. Research in the field faces the challenge that the software construction process is difficult to be studied from a purely engineering viewpoint. Software development depends on human factors, and a multidisciplinary viewpoint should be adopted in empirical software engineering research. It has been established that software developers solve problems through cognitive processing abilities. Among the many skills required for software development, developers must possess high analytical problem-solving skills and creativity for the software construction process. According to the research done in the field of psychology, affective states—emotions, moods, and feelings—deeply influence the cognitive processing abilities and performance of workers, including creativity and analytical problem solving. Nonetheless, little research so far has investigated the correlation between affective states, creativity, and analytical problem-solving skills of programmers. There is a recent proposal to study human aspects in empirical software engineering with psychological measurements. This article echoes the call and reports a natural experiment with 42 participants to investigate the relationship between affective states, creativity, and analytical problem-solving skills of software developers. The results offer support for the claim that happy developers are indeed better problem solvers in terms of their analytical abilities. The contributions of this study are in (1) providing a better understanding of the impact of the affective states on the creativity and analytical problem-solving capacities of developers; in (2) introducing and validating psychological measurements, theories, and

concepts on affective states, creativity and analytical-problem-solving skills in empirical software engineering; and in (3) raising the need to study human aspects in software engineering by employing a multidisciplinary viewpoint.

Authors and affiliations

Daniel Graziotin, Free University of Bozen-Bolzano, Faculty of Computer Science, Piazza Domenicani 3, 39100 Bolzano, Italy

Xiaofeng Wang, Free University of Bozen-Bolzano, Faculty of Computer Science, Piazza Domenicani 3, 39100 Bolzano, Italy

Pekka Abrahamsson, Free University of Bozen-Bolzano, Faculty of Computer Science, Piazza Domenicani 3, 39100 Bolzano, Italy

Introduction

For several years now, the software industry has been faced with the problem of the *productivity gap* (Ecker, Müller, & Domer, 2009). The amount of software in electronic devices—such as personal computers, mobile phones, televisions, automobiles, and many others—is increasing at a rate that is faster than our capability to deliver software. While there have been several solutions proposed over time, the principal problem pertains even today. It has been claimed that an effective way to improve software developers' productivity and software quality is to focus on people, e.g., (Boehm & Papaccio, 1988). The software engineering field's latest trends concur with Boehm & Papaccio, (1988) early claim. In specific, the advocates of Agile software development methods have stressed the point that "If the people on the project are good enough, they can use almost any process and accomplish their assignment. If they are not good enough, no process will repair their inadequacy— 'people trump process' is one way to say this" (Cockburn & Highsmith, 2001). However, little evidence has been offered to verify this claim in empirical software engineering research.

Software engineering field faces an additional challenge, when compared to more traditional engineering fields. The challenge is that software development activities are rather different from industrial processes carried on by machines. Human aspects play a fascinating, complex role in the construction of software products, and the process itself is mainly intellectual (Darcy, 2005; Glass, Vessey, & Conger, 1992). Lately, the discipline of software engineering has begun to adopt a multidisciplinary view and is embracing theories from more established disciplines, such as psychology, organizational research, and human-computer interaction. A proposal to satisfy the need to study human aspects in empirical software engineering research was by "collecting psychometrics"[1] (Feldt, Angelis, & Samuelsson, 2008). Although this call has been mildly echoed, the present authors noted limited research on the role of emotions and moods of software developers as humans.

As individuals, we act based on emotions as we encounter the world through moods. Affects enable the *mattering* of things; they are the medium within which acting towards the world takes place (Ciborra, 2002, pp. 159–165). Numerous studies have shown that the happiness of people is an indicator of achievement in various life domains, including work achievements (Lyubomirsky, King, & Diener, 2005). Indeed, emotions play a role in the daily job; they pervade organizations, the relationships between workers, deadlines, work motivation, sense-making and human-resource processes (Barsade & Gibson, 2007). Organizational settings affect the worker's thoughts, feelings, and actions, and vice-versa (Brief & Weiss, 2002). Although emotions have been historically neglected in studies of industrial and organizational psychology (Muchinsky, 2000), an interest in the role of affect on job outcomes has accelerated over the past decades (Fisher & Ashkanasy, 2000). Of important interest for recent research has been the relationship between affect on the job and work-related achievements, including performance (Barsade & Gibson, 2007; Miner & Glomb, 2010; Shockley, Ispas, Rossi, & Levine, 2012), and problem-solving processes like creativity (T. Amabile, 1996; T. M. Amabile, Barsade, Mueller, & Staw, 2005).

Despite that the ability to sense the moods and emotions of software developers may be essential for the success of an Information Technology firm (Denning, 2012), software engineering research is lacking an understanding of how emotions have a role in the software

1 Two anonymous reviewers in the fields of psychology suggested that *psychometrics* actually means the field of study concerned with the implementation and validation of psychological measurements. The term has been mistakenly interpreted as "psycho"-metrics by the literature in Software Engineering, including this paper's authors. *Psychological measurements* is a correct term to be employed in this case, and it is present in this article.

construction process (Khan, Brinkman, & Hierons, 2010; Shaw, 2004). In software engineering research, the affective states of software developers have been rarely investigated, in spite of the fact that affective states have been a subject of other Computer Science disciplines such as human-computer interaction and computational intelligence (Lewis, Dontcheva, & Gerber, 2011; Tsonos, Ikospentaki, & Kouroupetrolgou, 2008). Thus, it is necessary to study the role of affective states—emotions, moods, and feelings—of software developers.

Software developers engage mostly in tasks where problem solving is required. They need to plan strategies in order to find a possible solution to a given problem, or to generate multiple creative and innovative ideas. Therefore, among the many skills required for software development, developers need to possess high analytical problem-solving skills and creativity. Both of them are cognitive processing abilities. Indeed, software development activities are not physical. Software development is complex and intellectual (Darcy, 2005; Glass et al., 1992), and it is accomplished through cognitive processing abilities (Fischer, 1987; Khan et al., 2010). Cognitive processing abilities have been shown to be deeply linked to the affective states of individuals (Ilies & Judge, 2002). Research in psychology and organizations has studied the correlation between the affective states of individuals, their work-related performance, and their cognitive processing abilities, including creativity and analytical problem-solving skills. No consensus has been reached on these linkages yet. However, there is a tendency to consider positive affective states as indicators of high job-related performance. The relationship between affective states and the creativity and analytical problem-solving skills of software developers in general has never been investigated.

This article reports a natural experiment with 42 participants that had the objective to analyze analytical and creative problem-solving skills of software developers, with respect to their affective states. The results of this study indicate that the happiest software developers are also those who possess the highest analytical problem-solving skills. Affective states may not influence the creativity of software developers, but the data inspire future research avenues.

This article offers several contributions. (1) It provides a better understanding of the impact of the affective states on the creativity and analytical problem-solving capacities of developers. (2) It introduces and validates psychological measurements, theories, and concepts on affective states, creativity and analytical problem-solving skills in empirical software engineering. (3) It raises the need to study human aspects in software engineering by employing a multidisciplinary viewpoint.

This article is organized as follows[2]: In the remaining part of this section, the background research and the related work of this study are reviewed. An overview is provided on the theories behind affective states and skills for software development in psychology and organizational research. The constructs, measurements, and the hypotheses are reported. The *Materials and Methods* section presents the design of the experiment, that is, the context, the participants' characteristics, the required instruments, and the procedures of the experiment are presented. The *Results* section describes the execution and the outcomes of the experiment, including the descriptive statistics and the hypothesis testing. The *Discussion* section provides the interpretation of the results, the limitations of the study, their implications of this study results, and the suggestions for future research. The last section concludes the paper.

2 It is an objective of this manuscript to bring concepts, theories, and measurements from psychology to the body of knowledge of software engineering. Therefore, some information provided in this article—especially in the Introduction—may appear redundant and obvious for a reader acquainted with psychology.

Affective states

The term *affective states* has been defined as “any type of emotional state [..] often used in situations where emotions dominate the person's awareness” (VandenBos, 2013). However, as it is shown in this section, it has been employed as a general term for emotions, moods, and feelings. Indeed, the terms emotions, moods, feelings, and happiness are freely interchanged in daily conversations. Thus, it is useful to define these terms first.

Emotions have been defined as the states of mind that are raised by external stimuli and are directed toward the stimulus in the environment by which they are raised (Plutchik & Kellerman, 1980). However, more than 90 definitions have been produced for this term (Kleinginna & Kleinginna, 1981), and no consensus in the literature has been reached. The term has been taken for granted and often defined with references to a list, e.g. anger, fear, joy, surprise (Cabanac, 2002).

Moods, on the other hand, have been defined as emotional states in which the individual feels good or bad, and either likes or dislikes what is happening around him/her (Parkinson, Briner, Reynolds S., & Totterdell, 1996).

Feelings have been defined as the conscious subjective experience of emotions (VandenBos, 2013). The term *happiness* has been defined as the emotional evaluation of life (Ed Diener, 1984; Dogan, Totan, & Sapmaz, 2013), and it can be measured as a sum of frequency of emotions in a timespan (E. Diener et al., 2009a; Dogan et al., 2013).

The difference between these terms is subtle for researchers outside of the psychology-fields and it is common to interchange them in everyday lives. Indeed, many authors over time consider mood, emotion, and feeling as interchangeable terms (Baas, De Dreu, & Nijstad, 2008; Schwarz & Clore, 1983; Schwarz, 1990; Wegge, Dick, Fisher, West, & Dawson, 2006), although it has been acknowledged that numerous attempts exist to differentiate them (Wegge et al., 2006; Weiss & Cropanzano, 1996). It has been suggested that a difference between moods and emotions lies in an absence of a causal factor in the phenomenal experience of the mood (Weiss & Cropanzano, 1996).

Emotions and moods are affective states (Fisher, 2000; Khan et al., 2010; Oswald, Proto, & Sgroi, 2008; Weiss & Cropanzano, 1996). It has also been argued that a distinction is not truly necessary for the sake of studying cognitive responses that are not strictly connected to the origin of the mood or emotion (Weiss & Cropanzano, 1996). For the purposes of this investigation, the authors have adopted the same stance and employed the noun *affective states* as an umbrella term for emotions, moods, and feelings.

There are two main theories to categorize affective states. One theory, namely the discrete approach, collects a set of basic affective states that can be distinguished uniquely (Plutchik & Kellerman, 1980), and that possess high cross-cultural agreement when evaluated by people in literate and preliterate cultures (Ekman, 1971; Matsumoto & Hwang, 2011). Ekman (1971) proposed a set of basic affects, which include anger, happiness, surprise, disgust, sadness, and fear. However, the list has received critique, leading to an extended version with other eleven elements (Ekman, 1992). They include amusement, embarrassment, relief, and shame. Plutchik & Kellerman (1980) proposed an alternative viewpoint, called the Wheel of Emotions. Eight primary, bipolar affective states were presented as coupled pairs: joy versus sadness, anger versus fear, trust versus disgust, and surprise versus anticipation. These eight basic emotions vary in intensity and can be combined with each other, to form secondary emotions. For example, joy has been set as the midpoint between serenity and ecstasy, whereas sadness has been shown to be the midpoint between pensiveness and grief. Developing a minimal list of basic affective states appears to be difficult with the discrete approach. Subsequent studies have come to the point where more than 100 basic emotions have been proposed (Shaver, Schwartz, Kirson, & O'Connor, 1987).

The other theory groups affective states in major dimensions that allow clear distinction among them (Lane, Chua, & Dolan, 1999; Russell, 1980). Affective states have been proposed to be characterized by their valence, arousal, and dominance. *Valence* (or pleasure) is the attractiveness (or adverseness) of an event, object, or situation (Lang, Greenwald, Bradley, & Hamm, 1993; Lewin, 1935). The term refers to the "direction of a behavioral activation associated toward (appetitive motivation) or away (aversive motivation) from a stimulus" (Lane et al., 1999, p. 990). *Arousal* represents the intensity of an emotional activation (Lane et al., 1999). It is the sensation of being mentally awake and reactive to stimuli, i.e., vigor and energy or fatigue and tiredness (Zajenkowski, Goryńska, & Winiewski, 2012). *Dominance* (or control, over-learning) represents a change in the sensation of the control of a situation (Bradley & Lang, 1994). It is the sensation by which an individual's skills are perceived to be higher than the challenge level for a task (Csikszentmihalyi, 1997a). The dimensional approach has been distinguished from the discrete approach in its lesser number of elements to be evaluated. Thus, it has been deemed useful in tasks where affective states must be evaluated quickly and preferably often. Indeed, it has been commonly adopted to assess affective states triggered by an immediate stimulus (Bradley & Lang, 1994; Ong, Cardé, Gross, & Manber, 2011), especially when repeated assessments of affective states are needed in very short timespans (Graziotin, Wang, & Abrahamsson, 2013).

This study opted for the discrete approach, as it has been adopted when studying the impact of affective states on a participant's characteristics without concern for the emotional reactions to external stimuli (Brand, Reimer, & Opwis, 2007; Forgeard, 2011).

Measuring affective states

Psychology studies have often grouped participants with respect to their affective states in terms of negative, neutral (less often), or positive. In the case of controlled experiments, the grouping has been based on the treatments to induce affective states. In the case of quasi-experiments and natural experiments, the grouping has been based on the values of affective state metrics, usually employed in questionnaires. Several techniques have been employed to induce affective states on participants, such as showing films, playing certain types of music, showing pictures and photographs, or letting participants remember happy and sad events in their lives (Lewis et al., 2011; Westermann & Spies, 1996). Recent studies have questioned the effects of mood-induction techniques, especially when studying pre-existing affective states of the participants (Forgeard, 2011).

One of the most notable measurement instruments for affective states is the Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988). It is a 20-item survey that represents positive affects (PA) and negative affects (NA). However, several shortcomings have been criticized for this instrument: PANAS reportedly omits core emotions such as *bad* and *joy* while including items that are not considered emotions, like *strong*, *alert*, and *determined* (E. Diener et al., 2009a; Li, Bai, & Wang, 2013). Another limitation has been reported in its non-consideration of the differences in desirability of emotions and feelings in various cultures (Li et al., 2013; Tsai, Knutson, & Fung, 2006). Furthermore, a considerable redundancy has been found in PANAS items (Crawford & Henry, 2004; Li et al., 2013; Thompson, 2007). PANAS has also been reported to capture only high-arousal feelings in general (E. Diener et al., 2009a).

Recent, modern scales have been proposed to reduce the number of the PANAS scale items and to overcome some of its shortcomings. Diener et al. (2009) developed the Scale of Positive and Negative Experience (SPANE). SPANE assesses a broad range of pleasant and unpleasant emotions by asking the participants to report them in terms of their frequency during the last four weeks. It is a 12-items scale, divided in two sub-scales. Six items assess positive affective states and form the SPANE-P scale. The other six assess negative affective states and form the SPANE-N scale. The answers to the items are given on a five-point scale ranging from 1 (*very rarely or never*) to 5 (*very often or always*). For example, a score of five for the *joyful* item means that the respondent experienced this affective state *very often* or *always* during the last four weeks. The SPANE-P and SPANE-N scores are the sum of the scores given to their respective six items. Therefore, they range from 6 to 30. The two scores can be further combined by subtracting SPANE-N from SPANE-P, resulting in the Affect Balance Score (SPANE-B). SPANE-B is an indicator of the pleasant and unpleasant affective states caused by how often positive and negative affective states have been felt by the participant. SPANE-B ranges from -24 (*completely negative*) to +24 (*completely positive*).

The SPANE measurement instrument has been reported to be capable of measuring positive and negative affective states regardless of their sources, arousal level or cultural context, and it captures feelings from the emotion circumplex (E. Diener et al., 2009a; Li et al., 2013). The timespan of four weeks was chosen in SPANE in order to provide a balance between the sampling adequacy of feelings and the accuracy of memory (Li et al., 2013), and to decrease the ambiguity of people's understanding of the scale itself (E. Diener et al., 2009a). The SPANE has been validated to substantially converge to other affective states measurement instruments, including PANAS (E. Diener et al., 2009a). The scale provided good psychometric properties in the introductory research (E. Diener et al., 2009a) and in numerous follow-ups, with up to twenty-one thousand participants in a single study (Dogan et al., 2013; Li et al., 2013; Silva & Caetano, 2011). Additionally, the scale proved consistency across full-time workers and students (Silva & Caetano, 2011).

Even if the SPANE-B score is a fuzzy indication of the affective states felt by individuals, it could be employed to split participants into groups using a median split. It is common to adopt the split technique on affective states measures (Berna et al., 2010; Forgeard, 2011; Hughes & Stoney, 2000). Regression analysis is also possible if the data are suitable for it.

Lastly, agreement measurements exist to provide validation of multiple SPANE affective states measurements. Cronbach (1951) developed the Cronbach's alpha as a coefficient of internal consistency and interrelatedness especially designed for psychological tests. It considers the variance specific to individual items. The value of Cronbach's alpha ranges from 0.00 to 1.00, where values near 1.00 mean excellent consistency (Cortina, 1993; Cronbach, 1951).

Affective states and software developers

Limited research was found on the affective states of software developers. To the authors' knowledge, three publications exist that employed psychological tests to study the affective states of software developers.

Shaw (2004) observed that the role of emotions in the workplace has been the subject of study management research, but information systems research has focused on job outcomes such as stress, turnover, burnout, and satisfaction. However, little or no attention has been given to the emotions of information technology professionals. The study explored the emotions of information technology professionals and how emotions can help to explain their job outcomes. The paper employed Affective Events Theory (Weiss & Cropanzano, 1996) as a framework for studying the fluctuation of the affective states of 12 senior-level undergraduate students who were engaged in a semester-long implementation project for an information systems course. The participants were asked to rate their affective states during or right after their episodes of work on their project. At four intervals during the project they filled out a survey on stress, burnout, emotional labor, and identification with their teams. Shaw considered each student as a single case study, as statistical analysis was not considered suitable. The study showed that the affective states of a software developer may dramatically change during a period of 48 hours, and that Affective Events Theory proved its usefulness in studying the affective states of software developers while working. However, the research was a work in progress, and no continuation is currently known. The paper has called for research on the affective states of software developers.

This call was echoed by Khan et al. (2010). In the study, a correlation with cognitive processing abilities and software development was theoretically demonstrated. The authors constructed a theoretical two-dimensional mapping framework in two steps. In the first step, programming tasks were linked to cognitive tasks. For example, the process of constructing a program —e.g., modeling and implementation—were mapped to the cognitive tasks of memory, reasoning, and induction. In the second step, the same cognitive tasks were linked to affective states. The authors reported two empirical studies on affective states and software development. The studies were related to the impact of affective states to developers' debugging performance. In the first study, affective states were induced to software developers. Subsequently, the programmers completed a quiz on software debugging. In the second study, the participants were asked to write a trace on paper of the execution of algorithms implemented in Java. After 16 minutes of algorithm tracing, arousal was inducted to the participants. Subsequently, the participants continued in their debugging task. The overall study provided empirical evidence for a positive correlation with the affective states of software developers and their debugging performance.

Graziotin et al. (2013) conducted a correlational study on the affective states of developers and their self-assessed productivity while constructing software. The research employed the dimensional view of affective states and a pictorial survey to assess the affective states raised by the software development task. The study observed eight developers working on their individual software projects. Their affective states and their self-assessed productivity were measured in intervals of 10 minutes each. The analysis of the correlation employed a linear mixed-effects model. Evidence was found that valence and dominance towards a software development task are positively correlated with the self-assessed productivity of developers. The study called for future research on the topic.

Problem-solving skills and affective states

Among many theories, two ways of thinking for problem-solving have been distinguished in psychology which are strictly related to creativity and analytical problem-solving, namely divergent thinking and convergent thinking. Divergent thinking leads to no agreed upon solutions and involves the ability to generate a large quantity of not necessarily correlated ideas (Csikszentmihalyi, 1997b). Convergent thinking involves solving well-defined, rational problems that often have a unique, correct answer. It emphasizes speed and working out from what is already known. It leaves little room for creativity, as the answers are either right or wrong (Cropley, 2006; Csikszentmihalyi, 1997b). Although divergent and convergent thinking have not proven to be synonyms for creativity and analytical problem-solving capabilities—there are no clear definitions for these terms (T. M. Amabile, 1982)—these are the two dimensions that most studies analyze while still claiming to study creativity and analytical problem-solving (Csikszentmihalyi, 1997b).

A creative performance can be either defined in terms of creative outcome or by the processes that lead to the creation of the creative results (T. M. Amabile, 1982; Davis, 2009). In this paper, creativity was studied in terms of *creative outcome,* as it has been the most explored in the referenced disciplines.

According to a recent meta-analysis on the impact of affective states on creativity (i.e., creative outcomes) in terms of the quality of generated ideas, positive affective states lead to higher creativity than neutral affective states, but there are no significant differences between negative and neutral affective states or between positive and negative affective states (Baas et al., 2008). Another recent meta-analysis agreed that positive affective states have moderate effects on creativity with respect to neutral affective states. However, it showed that positive affective states also have small, non-zero effects on creativity with respect to negative affective states (Davis, 2009). Lewis et al. (2011) provided evidence for higher creativity under induced positive and negative affective states, with respect to non-induced affective states. Forgeard (2011) showed that participants low in depression possess higher creativity with induced negative affective states, and no benefits were found for individuals with induced positive affective states. Sowden & Dawson (2011) found that the quantity of generated creative ideas is boosted under positive affective states, but no difference in terms of quality was found in their study. On the other hand, there are studies empirically demonstrating that negative affective states increase creativity (George & Zhou, 2002; Kaufmann & Vosburg, 1997). Based on the literature review, the authors of this study have agreed with Fong (2006) that the nature of the relationship between affective states and creativity has not been completely understood and that more research is needed. No direction could be predicted on a difference between creativity and affective states of software developers.

First high-level hypothesis—a difference was expected in the creative work produced by software developers with respect to their affective states.

Few studies have assessed the relationship between affective states and the analytical problem-solving skills of individuals. The understanding of the relationship is still limited even in psychology studies. Analytical problem-solving skills—related to convergent thinking—are believed to be higher under the influence of negative affective states than positive affective states (Abele-Brehm, 1992; Melton, 1995). Somewhat contradictory to previous studies, Kaufmann & Vosburg (1997) reported no correlation between analytical problem-solving skills and the affective states of the participants. On the other hand, the processes of transferring, learning, and solving analytical problems have been reported to deteriorate with negative emotions (Brand et al., 2007). A consensus in research has yet to be consolidated. Again, no direction could be predicted on a difference between analytical problem-solving skills and affective states of software developers.

Second high-level hypothesis—a difference is expected in the analytical problem-solving skills of software developers with respect to their affective states.

Measuring problem-solving skills

Different tasks for measuring the creativity of the participants have been proposed. The most adopted task has been to generate creative ideas—e.g., the generation of captions for ambiguous photographs—or short solutions for uncommon and bizarre problems (Forgeard, 2011; Kaufman, Lee, Baer, & Lee, 2007; Lewis et al., 2011; Sowden & Dawson, 2011).

In order to measure the creativity, according to the Consensual Assessment Technique (T. M. Amabile, 1982), independent judges, expert in the field of creativity, score the captions using a Likert-item related to the creativity of the artifact to be evaluated. The judges have to use their own definition of creativity (T. M. Amabile, 1982; Kaufman et al., 2007). The Likert-item is represented by the sentence *This caption is creative*. The value associated to the item ranges from 1 (*I strongly disagree*) to 7 (*I strongly agree*).

The present study adopted measurements of quality and quantity for the assessment of creativity. The quality dimension of creativity was measured by two scores. The first quality score was the average of the scores assigned to all the generated ideas of a participant (*ACR*). The second quality score was the best score obtained by each participant (*BCR*), as suggested by Forgeard (2011), because creators are often judged by their best work rather than the average of all their works (Kaufman et al., 2007). The quantity dimension was represented by the number of generated ideas (*NCR*) as suggested by Sowden & Dawson (2011).

Measuring analytical problem-solving skills is less problematic than measuring creativity. There is only one solution to a given problem (Cropley, 2006). The common approach in research has been to assign points to the solution of analytical tasks (Abele-Brehm, 1992; Melton, 1995). This study employed the approach to combine measures of quality and quantity by assigning points to the achievements of analytical tasks and by measuring the time spent on planning the solution. The Tower of London game (a.k.a. *Shallice's test*) is a game aimed to determine impairments in planning and executing solutions to analytical problems (Shallice, 1982). It is similar to the more famous Tower of Hanoi game in its execution. Fig. 2 provides a screenshot of the game. The rationale for the employment of this task is straightforward.

The Tower of London game is a very high-level task that resembles an algorithm design and execution. This task reduced the limitations that would have been imposed by employing a particular programming language. Furthermore, such a level of abstraction permits a higher level of generalization, as the results are not bound to a particular programming language.

The Analytical Problem Solving (*APS*) score is defined as the ratio between the progress score achieved in each trial of the Tower of London Game (TOLSS), and the seconds needed to plan the solution to solve each trial (PTS). TOLSS ranges from 0 to 36 because there are 12 problems to be solved and each one can be solved in a maximum of 3 trials. PTS is the number of milliseconds that occur from the presentation of the problem to the first mouse click in the program. In order to have comparable results, a function to map the *APS* ratio to a range from 0.00 to 1.00 is employed.

To the knowledge of the authors of this article, there have not been studies in software engineering research that define software development tasks suitable for measuring the creativity and the analytical problem-solving skills of software developers. Although strict development tasks could be prepared, there would be several threats to validity. Participants with different backgrounds and skills are expected, and it is almost impossible to develop a software development task suitable and equally challenging for first year BSc students and second year MSc students. For example, asking a first year BSc students to implement a design pattern may result in a negative performance no matter their affective states. Additionally, Computer Science students typically know more than one programming language but no particular language can be expected to be shared across the participants of an experiment. For these reasons, the presented study remained at a higher level of abstraction. Creativity and analytical problem-solving skills were measured with validated tasks from psychology research.

Hypotheses

The literature review produced two high-level research hypotheses, on a difference in creativity and analytical problem-solving skills of software developers, with respect to their affective states. However, the high-level hypotheses can be refined with the acquired knowledge on the measurement of the constructs. This section presents the hypotheses of this study. As the literature review did not provide clear indications on the direction of such hypothesized difference, the alternative hypotheses were two-tailed.

From the first high-level hypothesis, a difference was expected in the creative work produced by developers with respect to their affective states. The hypotheses were developed. *Hypothesis H1* was on a difference in the best creativity score (*BCR*) of software developers with respect to their affective states. *Hypothesis H2* was on a difference in the average creativity score (*ACR*) of software developers with respect to their affective states. *Hypothesis H3* was on a difference in the number of creative solutions (*NCR*) generated by software developers with respect to their affective states.

From the second high-level hypothesis, a difference was expected in the analytical problem-solving skills of developers with respect to their affective states. *Hypothesis H4* was on a difference in the analytical problem-solving score (*APS*) between software developers with respect to their affective states.

Materials and methods

Participant Characteristics

Forty-two student participants were recruited from the Faculty of Computer Science from the Free University of Bozen-Bolzano. There were no restrictions in gender, age, nationality, or level of studies of the participants. The participation was voluntary and in exchange of credit points. The affective states of the participants were natural, i.e., random for the researchers. Of the 42 participants, 33 were male and 9 female. The participants had a mean age M = 21.50 years old (standard deviation (SD) = 3.01 years) and were diverse in provenience country: Italy 74%, Lithuania 10%, Germany 5%, and Ghana, Nigeria, Moldavia, Peru, U.S.A. all with a 2.2% frequency. The experience in terms of years of study has been recorded (M = 2.26 years, SD = 1.38).

An institutional review board approval for conducting empirical studies on human participants was not required by the institution. However, written consent was obtained from all subjects. Participants were advised, both informally and on the consent form, about the data retained and that anonymity was fully ensured. No sensitive data has been collected in this study. Participants were assigned a random participant code in order to link the gathered data. The code was in no way linked to any information that would reveal the participant's identity.

All the students participated to the affective states measurements sessions and the two experiment tasks. However, the working results of one participant from the creativity task and another from the analytical problem-solving task have been excluded; the two participants did not follow the instructions and submitted incomplete data. Therefore, the sample size for the two experiment tasks was N = 41. No participants reported previous experience with the tasks.

In addition to the experiment participants, the creativity task of this study required the recruitment of independent judges, who are experts in creativity-related fields (T. M. Amabile, 1982; Forgeard, 2011; Kaufman et al., 2007). Ten independent judges were contacted to rate the captions produced in the first task. Seven of them positively answered the request. Five judges completed the evaluation of the produced creative captions. Among the five recruited judges for evaluating the creative captions of the students, two are internationally recognized professors of Design & Arts. Two judges are local professors of humanistic studies; one of them is an expert in cinema and theatre arts. One judge is a professor in creative writing. The judges received the outcome of the creativity task. Therefore, they were involved after the execution of the experiment.

The participants and the judges were not aware of any experimental settings—e.g., existence of groups—nor of any purpose of the experiment. The judges were blind to the design and the scope of the experiment. That is, they received the six pictures with all the participants' captions grouped per picture. The judges were not aware of the presence of other judges and rated the captions independently.

Instruments

Six color photographs with ambiguous meanings were required for the creativity task. Fig. 1 is one of the six photographs. They are available from the authors upon request. The material required for the analytical problem-solving task was provided by the open-source PEBL software (Piper et al., 2011). PEBL instructed the participants, provided the task, and collected several metrics, including those employed to build the *APS* score. One computer per participant was required. The SPANE questionnaire instructions provided to the participants are available in the article by Diener et al. (2009), but are currently freely accessible on one author's academic website (E. Diener et al., 2009b).

Procedure

The experiment procedure was composed of four activities: (1) affective states measurement, (2) creativity task, (3) affective states measurement, (4) analytical problem-solving task. The second affective states measurement session happened to limit the threats to validity, as the first task may provoke a change in the affective states of the participants.

The participants arrived at the environment knowing only that they would be participating in an experiment. As soon as they sat at their workstation, they read a reference sheet, which is included in Supplemental Article S1. The sheet provided a summary of all the steps of the experiment. The researchers also assisted the participants during each stage of the experiment. The participants were not allowed to interact with each other.

For the two affective states measurement sessions, the participants completed the SPANE questionnaire through a Web form, which included the related instructions as well.

Creativity task

The participants received two random photographs from the set of the six available ones, one at a time. Participants imagined participating in the *Best Caption of the Year* contest and tried to win the contest by writing the best captions possible for the two photographs. They wrote as many captions as they wanted for the pictures. The creativity task instructions and the specific instructions for the judges are available as an appendix in the study by Forgeard (2011).

Analytical problem-solving task

The participants opened the PEBL software. The software was already set up to automatically display the Tower of London game, namely *Shallice test* ([1,2,3] pile heights, 3 disks, Shallice's 12 problems). The PEBL software displayed its instructions before the start of the task. The instructions stated how the game works and that the participants had to think about the solution before starting the task—i.e., doing the first mouse click. Fig. 2 provides a screenshot of the first level of the game. As PEBL is open-source software, the reader is advised to obtain it in order to read the instructions.

Although the participants did not have strict time restrictions to complete the tasks, they were advised on the time usually required to complete each task, and that the second task would begin only after every participant finished the first task. This apparent freedom was in reality based on the typical time limits known from previous studies.

Two supervisors were present during the experiment in order to check the progress of the participants and to answer to their questions. All the steps of the experiment were automatized with the use of a computer. An exception lay in the captions produced in the creativity task. The captions were manually transcribed in a spreadsheet file. For this reason, a third person double-checked the spreadsheet containing the transcribed captions. The data were then aggregated and analyzed using the open-source R software.

Results

The natural experiment was run in January 2012. The designed data collection process was fully followed. No deviations occurred. Both tasks required 30 minutes to be completed, and the participants completed the two surveys in 10 minutes each time. No participants dropped from the study.

Statistics and data analysis

The first SPANE questionnaire session happened as soon as the participants arrived at the laboratory. The SPANE-B value obtained from this measurement session let us estimate the SPANE-B population mean for software developers, $\mu_{\text{SPANE-B}_{\text{DEV}}} = 7.58$, 95% CI [5.29, 9.85]. The median value for SPANE-B was nine.

As the data could not permit insightful regression analysis on continuous values for the SPANE-B and the task scores, the data analysis was performed by forming two groups via a median split of the SPANE-B score. The two groups were called *N-POS* (as for *non-positive*) and *POS* (as for *positive*). Before the creativity task, 20 students were classified as *N-POS* while 21 students were part of the *POS* group.

The histograms related to the affective state distributions and the group compositions have been included as supplemental files of this article. They are not crucial for the purposes of this investigation. However, they have been attached to this article for the sake of completeness. The same holds for the graphs representing non-significant data.

Table 1 summarizes the task scores of the two groups for the two tasks. The two creativity scores *ACR* and *BCR* showed many commonalities. The scatterplots of the *ACR* (Supplemental Fig. S6) and *BCR* (Supplemental Fig. S7) scores versus the SPANE-B score suggested a weak trend of higher creativity when the SPANE-B value tended to its extreme values (-24 and +24). The median for the number of generated captions (*NCR*) was four for the *N-POS* group and six for the *POS* group. However, the lower quartiles of the two groups were almost the same and there was a tiny difference between the two upper quartiles.

Research hypotheses H1 and H2 were tested with unpaired, two-tailed t-tests. Regarding the *BRC* score, there was no support for a difference between the *N-POS* ($M = 4.02$, $SD = 0.76$) and the *POS* ($M = 3.98$, $SD = 0.76$) groups, $t(39) = 0.20$, $p > .05$, 95% CI [-0.43, 0.53]. In terms of the *ACR* score, there was no support for a difference between *N-POS* ($M = 3.13$, $SD = 0.45$) and *POS* ($M = 3.08$, $SD = 0.58$), $t(39) = 0.31$, $p > .05$, 95% CI [-0.28, 0.38].

Hypothesis H3 required a Mann-Whitney U test because the assumptions of normality were not met (Shapiro-wilk test for normality, $W = 0.89$, $p = 0.02$ for *N-POS* and $W = 0.87$, $p = 0.01$ for *POS*). In terms of the *NCR* score, there was no support for a difference between *N-POS* ($M=4.70$, $SD=2.34$) and *POS* ($M=5.90$, $SD=3.46$), $W = 167.50$, $p > .05$, 95% CI [-2.00, 1.00].

The second SPANE questionnaire session was performed right after the participants finished the creativity task. The average value of SPANE-B was $M = 8.70$ ($SD = 6.68$), with a median value of 10. Therefore, a slight change in the group composition happened, with 19 students composing the *N-POS* group and 22 students belonging to the *POS* group.

The boxplots for the *APS* score in Fig. 3[3] showed a difference between the two groups. The distribution of the *N-POS* group appeared as negatively skewed, but the *POS* group was likely positively skewed in its distribution. Although the two medians of the groups were not far away from each other, the two distributions seemed to overlap only between the upper quartile of the *N-POS* group and the lower quartile of the *POS* group.

The scatterplot of the *APS* score in Fig. 4 was appealing: the *APS* points for the *N-POS* group looked linearly, negatively correlated with the SPANE-B score. The *APS* scores for the *POS* did not present an evident correlation with the SPANE-B score. It seemed that excellent *APS* score were achieved only in the *POS* group.

H4 was tested with an unpaired, two tailed t-test with Welch's correction, because a significant difference in the variances of the two groups was found (F-test for differences in variances, $F(21, 18) = 3.32$, $p = 0.01$, 95% CI [1.30, 8.17]). With respect to the *APS* score, there was support for a difference in analytical problem-solving skills of *N-POS* (M = 0.14, SD = 0.04) and *POS* (M = 0.20, SD = 0.08), $t(33.45) = -2.82$, $p = 0.008$, 95% CI [-0.11 , -0.02]. A two-sample permutation test confirmed the results, $t(168)$, $p = 0.01$, CI [-13.19 -1.91].

In the second SPANE measurement session, there was a significant increase in the SPANE-B value: 1.02, $t(39) = 3.00$, $p < 0.01$, 95% CI [0.34, 1.71]. The Cronbach's alpha reliability measurement for the two SPANE questionnaire sessions was $\alpha = 0.97$, 95% CI [0.96, 0.98].

Discussion

The empirical data did not support a difference in the creativity with respect to the affective states of software developers in terms of average creativity, best creativity, and the number of generated creative ideas. The results of this study agree with those of Sowden & Dawson (2011), who did not find a difference in the creativity of the generated ideas with respect to the affective states of the participants. The results of this study are in line with those of Sowden & Dawson (2011) regarding the number of produced ideas, but not significantly. Instead, the results of this study deviate from those in the study by Forgeard (2011), where non-depressed participants provided more creative captions under negative affective states. Nevertheless, it must be noted that the depression factor has not been controlled in this study.

It should be noted that, as a result of this study, it can be hypothesized fur future studies that a tendency of higher creativity (*ACR* and *BCR* scores) is obtained when the SPANE-B approaches extreme values such as -24 or +24 (see the supplemental Fig. S6 and Fig. S7). An interpretation, concordant with the related work, could be that higher creativity is achieved when individuals feel extreme affective states, either in the positive or negative directions. The tendency could be even stronger for the *POS* group. However, any specific relationships cannot be claimed at this stage.

The empirical data supported a difference in analytical problem-solving skills of software developers with respect to their affective states. More specifically, the results suggest that happy software developers are better analytical problem solvers. The results are in contradiction to those of Abele-Brehm (1992) and Melton (1995), where higher analytical problem-solving skills were found in participants feeling negative affective states. The results of this study are also in

---

3 The color scheme for the graphs of this study have been generated by following guidelines for producing Colorblind-friendly graphics (Okabe & Ito, 2008).

contradiction to those of Kaufmann & Vosburg (1997), where no support was found for a difference in analytical problem-solving skills with respect of the affective states of participants.

It should also be noted that a curved, V-shaped relationship of the analytical problem-solving skills and the SPANE-B may exist (as Fig. 4 suggests), having a vertex in proximity of the value 5 for the SPANE-B score. This should be addressed in future studies with a larger amount of participants.

The average value of the SPANE-B score for the sample as soon as the participants arrived at the laboratory was 7.58 (SD=7.06). The 95% CI [5.29, 9.85], estimated for the true mean score, suggests that software developers are moderately happy on average.

Additionally, it has been observed how the participants enjoyed the creativity task and happily committed to it. This observation was mirrored by the data: the participants generated 220 captions, averaging 5.24 captions per participant. This enjoyment of the first task was reflected by the second SPANE measurement session, as the SPANE-B value significantly increased by 1.02, $t(39) = 3.00$, $p < 0.01$, 95% CI [0.34, 1.71]. This validates the capabilities of the adopted measurement instrument for the affective state measurements, and shows that even simple and short activities may impact the affective states of software developers. The Cronbach's alpha value of 0.97 of the two SPANE measurement sessions present evidence that the participants provided stable and consistent data.

Limitations

The primary limitation of this study lies in the sample: the participants were all Computer Science students. Even though there is diversity in nationality and experience in years of study, they have limited software development experience with respect to the professionals. However, Kitchenham et al. (2002) & Tichy (2000) argued that students are the next generation of software professionals. Thus, they are remarkably close to the interested population and may even be more updated on the new technologies (Kitchenham et al., 2002; Tichy, 2000). Höst, Regnell, & Wohlin (2000) have been along the same line as they found non-significant differences in the performance of professional software developers and students on the factors affecting the lead-time of projects. There is awareness that not all the universities offer the same curricula and teaching methods, and that students may have different levels of knowledge and skills (Berander, 2004). Still, given the high level of abstraction provided by the tasks in this study, a hypothetical difference between this study's participants and software professionals would likely be in the magnitude, not in the direction of the results (Tichy, 2000). Lastly, the employed affective states measurement instrument, SPANE, provided consistent data across full-time workers and students (Silva & Caetano, 2011).

Another limitation is that a full coverage of the SPANE-B range in the negative direction could not be obtained. Even though 42 participants were recruited, the SPANE-B score didn't fall below the value of -9, and its average value was always greater than +7 on a scale of [-24,+24]. Before the run of the experiment, a more homogeneous distribution of the participants with respect to the SPANE-B score was expected. On the other hand, there is actually no evidence that the distribution of SPANE-B for the population of software developers should cover the full range of [-24, +24]. Additionally, studies estimating the SPANE-B mean for any population are not known. For this reason, an estimation of the affective states population mean for software developers was offered by this study: $\mu_{\text{SPANE-B}_{\text{DEV}}} = 7.58$, 95% CI [5.29, 9.85]. Thus, it may be that the population's true mean for SPANE-B is above +7 and significantly different from the central value of the measurement instrument.

A third limitation lies in the employment of a median split to compose the groups. Employing median splits removed the precision that would have been available in a continuous measure of the SPANE-B[4]. Despite that, this was necessary as no known regression technique could yield valid results; median splits on affective state measurements are not uncommon in similar research (Berna et al., 2010; Forgeard, 2011; Hughes & Stoney, 2000).

Implications and future research

The theoretical implications of this study are that positive affective states of software developers are indicators of higher analytical problem-solving skills. Although the same is not shown for creativity, the data trends offer inspiration to continue also this study avenue. An implication for research in software engineering is that the study of affective states of the various stakeholders, who are involved in the process of software construction, should be taken into account and should become an essence of the research in the field.

Future research should provide additional details to the claims reported in this article. The weak yet promising trends in the creativity data inspire a prosecution of research in this area. A replication of this experiment with a larger order of magnitude may provide significant data, and could even enable regression analyses to verify how the intensity of affective states may matter in the creativity of software developers. It is necessary to study the affective states of software developers from a process-oriented view, to observe a possible correlation with work-related achievements and productivity while developing software. There is the need to focus on practical implications as well. Qualitative research should answer how the creativity of software developers influences design artifacts and the source-code of a software system. Research can be done on how mood induction effects may affect the quality of a software system and the productivity of a developer.

Conclusions

While software plays an important role in today's society and the software productivity problem is imminent, no satisfactory solutions have been proposed by the research in software engineering. For decades, it has been claimed without strong empirical evidence that a way to improve software developers' productivity and software quality is to focus on people. A proposal to effectively study human aspects in empirical software engineering research has been to adopt psychological measurements.

By observing the reference fields—primarily psychology and organizational research—it has been understood that software developers solve problems in creative and analytical ways, through cognitive processing abilities. Cognitive processing abilities are deeply linked with the affective states of individuals, i.e., emotions, moods and feelings.

This paper reported a natural experiment—relying on observations of the variables under study—on the importance of affective states on important software development skills and capacities, namely analytical problem-solving skills (convergent thinking) and creativity (divergent thinking). It has been shown that happy software developers are significantly better analytical problem solvers. Although the same could not be shown with respect to creativity, the data trends inspire future research on this matter.

The understanding provided by this study should be part of basic science—i.e., essence—in software engineering, rather than leading to direct, applicable results. This work (1) provides a better understanding of the impact of the affective states on the creativity and analytical

4 The authors are thankful to an anonymous reviewer for pointing out this issue.

problem-solving capacities of developers; (2) introduces and validates psychological measurements, theories, and concepts on affective states, creativity and analytical-problem-solving skills in empirical software engineering; (3) raises the need to study human aspects in software engineering by employing a multidisciplinary viewpoint.

Although the claim *people trump process* is far from being empirically validated yet, this study provides tools, evidence, and an attitude towards its validation. This study calls for further research on the affective states of software developers.

Software developers are unique human beings. By embracing a multidisciplinary view, human aspects in software engineering can be effectively studied. By inspecting how cognitive activities influence the performance of software engineers, research will open up a completely new angle and a better understanding of the creative activity of the software construction process.

Acknowledgments

The authors would like to thank the students participating in the experiment. The authors would also like to acknowledge Elena Borgogno, Cristiano Cumer, Federica Cumer, Kyriaki Kalimeri, Paolo Massa, Matteo Moretti, Maurizio Napolitano, Nattakarn Phaphoom, and Juha Rikkilä for their kind help during this study. Last but not least, the authors are grateful for the insightful comments and understanding offered by the assigned academic editor Shane Mueller and two anonymous reviewers.

References

Abele-Brehm A. 1992. Positive versus negative mood influences on problem solving: A review. *Polish Psychological Bulletin 23*:187–182.

Amabile T. 1996. Creativity and innovation in organizations. *Harvard Business School Background* Note 396–239. Retrieved from http://hbr.org/search/396239

Amabile TM. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology 43*:997–1013. doi:10.1037//0022-3514.43.5.997

Amabile TM, Barsade SG, Mueller JS, Staw BM. 2005. Affect and Creativity at Work. *Administrative Science Quarterly 50*:367–403. doi:10.2189/asqu.2005.50.3.367

Baas M, De Dreu CKW, Nijstad BA. 2008. A meta-analysis of 25 years of mood-creativity research: hedonic tone, activation, or regulatory focus? *Psychological bulletin 134*:779–806. doi:10.1037/a0012815

Barsade SG, Gibson DE. 2007. Why Does Affect Matter in Organizations? *Academy of Management Perspectives 21*:36–59. doi:10.5465/AMP.2007.24286163

Berander P. 2004. Using students as subjects in requirements prioritization. In *Proceedings 2004 International Symposium on Empirical Software Engineering 2004 ISESE 04* (pp. 167–176). Ieee. doi:10.1109/ISESE.2004.1334904

Berna C, Leknes S, Holmes E a, Edwards RR, Goodwin GM, Tracey I. 2010. Induction of depressed mood disrupts emotion regulation neurocircuitry and enhances pain unpleasantness. *Biological psychiatry 67*:1083–90. doi:10.1016/j.biopsych.2010.01.014

Boehm BW, Papaccio PN. 1988. Understanding and controlling software costs. *IEEE Transactions on Software Engineering 14*:1462–1477. doi:10.1109/32.6191

Bradley MM, Lang PJ. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry 25*:49–59. doi:10.1016/0005-7916(94)90063-9

Brand S, Reimer T, Opwis K. 2007. How do we learn in a negative mood? Effects of a negative mood on transfer and learning. *Learning and Instruction 17*:1–16. doi:10.1016/j.learninstruc.2006.11.002

Brief AP, Weiss HM. 2002. Organizational behavior: affect in the workplace. *Annual review of psychology 53*:279–307. doi:10.1146/annurev.psych.53.100901.135156

Cabanac M. 2002. What is emotion? *Behavioural Processes 60*:69–83. doi:10.1016/S0376-6357(02)00078-5

Ciborra C. 2002. *The Labyrinths of Information: Challenging the Wisdom of Systems* (1st ed.). New York, New York, USA: Oxford University Press, USA.

Cockburn A, Highsmith J. 2001. Agile software development, the people factor. *Computer, IEEE 34*:131–133. doi:10.1109/2.963450

Cortina JM. 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology 78*:98–104. doi:10.1037/0021-9010.78.1.98

Crawford JR, Henry JD. 2004. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *The British journal of clinical psychology / the British Psychological Society 43*:245–65. doi:10.1348/0144665031752934

Cronbach L. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika 16*:297–334. doi:10.1007/BF02310555

Cropley A. 2006. In Praise of Convergent Thinking. *Creativity Research Journal 18*:391–404. doi:10.1207/s15326934crj1803_13

Csikszentmihalyi M. 1997a. *Finding flow: The psychology of engagement with everyday life*. *Psychology Today* (1st ed., Vol. 30, p. 46). New York, New York, USA: Basic Books.

Csikszentmihalyi M. 1997b. *Creativity: Flow and the Psychology of Discovery and Invention*. *Personnel* (p. 464). Harper Perennial.

Darcy DP. 2005. Exploring Individual Characteristics and Programming Performance: Implications for Programmer Selection. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (Vol. 14, p. 314a–314a). IEEE. doi:10.1109/HICSS.2005.261

Davis M. 2009. Understanding the relationship between mood and creativity: A meta-analysis. *Organizational Behavior and Human Decision Processes 108*:25–38. doi:10.1016/j.obhdp.2008.04.001

Denning PJ. 2012. Moods. *Communications of the ACM 55*:33. doi:10.1145/2380656.2380668

Diener E., Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, Biswas-Diener R. 2009a. New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings. *Social Indicators Research 97*:143–156. doi:10.1007/s11205-009-9493-y

Diener E., Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, Biswas-Diener R. 2009b. Scale of Positive and Negative Experience (SPANE). *Ed Diener personal Website*. Retrieved March 19, 2013, from http://internal.psychology.illinois.edu/~ediener/SPANE.html

Diener Ed. 1984. Subjective well-being. *Psychological Bulletin 95*:542–575. doi:10.1037/0033-2909.95.3.542

Dogan T, Totan T, Sapmaz F. 2013. The Role Of Self-esteem, Psychological Well-being, Emotional Self-efficacy, And Affect Balance On Happiness: A Path Model. *European Scientific Journal 9*:31–42. Retrieved from http://eujournal.org/index.php/esj/article/view/1559

Ecker W, Müller W, Domer R. 2009. Hardware-dependent Software: Introduction and Overview. In *Hardware-dependent Software: Principles and Practice* (1st ed., pp. 1–13). Berlin, Heidelberg: Springer Verlag. doi:10.1007/978-1-4020-9436-1_1

Ekman P. 1971. Universals and cultural differences in facial expressions of emotion. In *Nebraska Symposium on Motivation* (pp. 207–283).

Ekman P. 1992. An argument for basic emotions. *Cognition & Emotion 6*:169–200. doi:10.1080/02699939208411068

Feldt R, Angelis L, Samuelsson M. 2008. Towards individualized software engineering: empirical studies should collect psychometrics. In *International Workshop on Cooperative and human aspects of software engineering* (pp. 49–52). ACM. doi:10.1145/1370114.1370127

Fischer G. 1987. Cognitive View of Reuse and Redesign. *IEEE Software 4*:60–72. doi:10.1109/MS.1987.231065

Fisher CD. 2000. Mood and emotions while working: missing pieces of job satisfaction? *Journal of Organizational Behavior 21*:185–202. doi:10.1002/(SICI)1099-1379(200003)21:2<185::AID-JOB34>3.0.CO;2-M

Fisher CD, Ashkanasy NM. 2000. The emerging role of emotions in work life: an introduction. *Journal of Organizational Behavior 21*:123–129. doi:10.1002/(SICI)1099-1379(200003)21:2<123::AID-JOB33>3.0.CO;2-8

Fong CT. 2006. The Effects Of Emotional Ambivalence On Creativity. *Academy of Management Journal 49*:1016–1030. doi:10.5465/AMJ.2006.22798182

Forgeard MJ. 2011. Happy people thrive on adversity: Pre-existing mood moderates the effect of emotion inductions on creative thinking. *Personality and Individual Differences 51*:904–909. doi:10.1016/j.paid.2011.07.015

George JM, Zhou J. 2002. Understanding when bad moods foster creativity and good ones don't: The role of context and clarity of feelings. *Journal of Applied Psychology 87*:687–697. doi:10.1037//0021-9010.87.4.687

Glass RL, Vessey I, Conger SA. 1992. Software tasks: Intellectual or clerical? *Information & Management 23*:183–191. doi:10.1016/0378-7206(92)90043-F

Graziotin D, Wang X, Abrahamsson P. 2013. Are Happy Developers More Productive? In *14th International Conference on Product-Focused Software Process Improvement (PROFES 2013)* (Vol. 7983, pp. 50–64). Paphos, Cyprus: Springer Berlin Heidelberg. doi:10.1007/978-3-642-39259-7_7

Höst M, Regnell B, Wohlin C. 2000. Using Students as Subjects — A Comparative Study of Students and Professionals in Lead-Time Impact Assessment. *Empirical Software Engineering 5*:201–214. doi:10.1023/A:1026586415054

Hughes JW, Stoney CM. 2000. Depressed mood is related to high-frequency heart rate variability during stressors. *Psychosomatic medicine 62*:796–803. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11138999

Ilies R, Judge T. 2002. Understanding the dynamic relationships among personality, mood, and job satisfaction: A field experience sampling study. *Organizational Behavior and Human Decision Processes 89*:1119–1139. doi:10.1016/S0749-5978(02)00018-3

Kaufman JC, Lee J, Baer J, Lee S. 2007. Captions, consistency, creativity, and the consensual assessment technique: New evidence of reliability. *Thinking Skills and Creativity 2*:96–106. doi:10.1016/j.tsc.2007.04.002

Kaufmann G, Vosburg J. 1997. “Paradoxical” Mood Effects on Creative Problem-solving. *Cognition & Emotion 11*:151–170. doi:10.1080/026999397379971

Khan IA, Brinkman W, Hierons RM. 2010. Do moods affect programmers’ debug performance? *Cognition, Technology & Work 13*:245–258. doi:10.1007/s10111-010-0164-1

Kitchenham BA, Pfleeger SL, Pickard LM, Jones PW, Hoaglin DC, Emam K El, Rosenberg J. 2002. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on Software Engineering 28*:721–734. doi:10.1109/TSE.2002.1027796

Kleinginna PR, Kleinginna AM. 1981. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion 5*:345–379. doi:10.1007/BF00992553

Lane RD, Chua PM-L, Dolan RJ. 1999. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia 37*:989–997. doi:10.1016/S0028-3932(99)00017-2

Lang PJ, Greenwald MK, Bradley MM, Hamm AO. 1993. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology 30*:261–273. doi:10.1111/j.1469-8986.1993.tb03352.x

Lewin K. 1935. *A dynamic theory of personality*. *Trans by Adams F and Zener K New York McGrawHill* (Vol. 2008, p. 77). New York, New York, USA: McGraw-Hill. doi:10.1109/IEMBS.2008.4649489

Lewis S, Dontcheva M, Gerber E. 2011. Affective computational priming and creativity. *Proceedings of the 2011 annual conference on Human factors in computing systems* 735–744. doi:10.1145/1978942.1979048

Li F, Bai X, Wang Y. 2013. The Scale of Positive and Negative Experience (SPANE): psychometric properties and normative data in a large Chinese sample. *PloS one 8*:e61137. doi:10.1371/journal.pone.0061137

Lyubomirsky S, King L, Diener E. 2005. The benefits of frequent positive affect: does happiness lead to success? *Psychological bulletin 131*:803–55. doi:10.1037/0033-2909.131.6.803

Matsumoto D, Hwang HS. 2011. Reading facial expressions of emotion. *Psychological Science Agenda, American Psychological Association*. Retrieved August 30, 2013, from http://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx

Melton RJ. 1995. The Role of Positive Affect in Syllogism Performance. *Personality and Social Psychology Bulletin 21*:788–794. doi:10.1177/0146167295218001

Miner AG, Glomb TM. 2010. State mood, task performance, and behavior at work: A within-persons approach. *Organizational Behavior and Human Decision Processes 112*:43–57. doi:10.1016/j.obhdp.2009.11.009

Muchinsky PM. 2000. Emotions in the workplace: the neglect of organizational behavior. *Journal of Organizational Behavior 21*:801–805. doi:10.1002/1099-1379(200011)21:7<801::AID-JOB999>3.0.CO;2-A

Okabe M, Ito K. 2008. Color Universal Design (CUD) / Colorblind Barrier Free - http://jfly.iam.u-tokyo.ac.jp/color/index.html.

Ong JC, Cardé NB, Gross JJ, Manber R. 2011. A two-dimensional approach to assessing affective states in good and poor sleepers. *Journal of Sleep Research 20*:606–10. doi:10.1111/j.1365-2869.2011.00907.x

Oswald AJ, Proto E, Sgroi D. 2008. Happiness and productivity. *The Warwick Economics Research Paper Series TWERPS no 882 1*:44 pages. Retrieved from http://wrap.warwick.ac.uk/1335/

Parkinson B, Briner R, Reynolds S., Totterdell P. 1996. *Changing moods: The psychology of mood and mood regulation*. *Longman* (1st ed., p. 4). London: Addison-Wesley Longman.

Piper BJ, Li V, Eiwaz MA, Kobel Y V., Benice TS, Chu AM, … Mueller ST. 2011. Executive function on the Psychology Experiment Building Language tests. *Behavior Research Methods*. doi:10.3758/s13428-011-0096-6

Plutchik R, Kellerman H. 1980. *Emotion, theory, research, and experience*. *Theories of Emotion* (Vol. 1). London: Academic Press.

Russell JA. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology 39*:1161–1178. doi:10.1037/h0077714

Schwarz N. 1990. Feelings as information: informational and motivational functions of affective states. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of Motivation and Cognition: Foundations of Social Behavior* (pp. 527–561). New York, NY, US: Guilford Press. Retrieved from http://psycnet.apa.org/psycinfo/1990-98254-015

Schwarz N, Clore GL. 1983. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology 45*:513–523. doi:10.1037/0022-3514.45.3.513

Shallice T. 1982. Specific Impairments of Planning. *Phil Trans R Soc B 298*:199–209. doi:10.1098/rstb.1982.0082

Shaver P, Schwartz J, Kirson D, O’Connor C. 1987. Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology 52*:1061–1086. doi:10.1037/0022-3514.52.6.1061

Shaw T. 2004. The emotions of systems developers. In *Proceedings of the 2004 conference on Computer personnel research Careers, culture, and ethics in a networked environment - SIGMIS CPR '04* (p. 124). New York, New York, USA: ACM Press. doi:10.1145/982372.982403

Shockley KM, Ispas D, Rossi ME, Levine EL. 2012. A Meta-Analytic Investigation of the Relationship Between State Affect, Discrete Emotions, and Job Performance. *Human Performance 25*:377–411. doi:10.1080/08959285.2012.721832

Silva AJ, Caetano A. 2011. Validation of the Flourishing Scale and Scale of Positive and Negative Experience in Portugal. *Social Indicators Research 110*:469–478. doi:10.1007/s11205-011-9938-y

Sowden PT, Dawson L. 2011. Creative feelings. In *Proceedings of the 8th ACM conference on Creativity and cognition - C&C '11* (p. 393). New York, New York, USA: ACM Press. doi:10.1145/2069618.2069712

Thompson ER. 2007. Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology 38*:227–242. doi:10.1177/0022022106297301

Tichy W. 2000. Hints for reviewing empirical work in software engineering. *Empirical Software Engineering 5*:309–312. doi:10.1023/A:1009844119158

Tsai JL, Knutson B, Fung HH. 2006. Cultural variation in affect valuation. *Journal of personality and social psychology 90*:288–307. doi:10.1037/0022-3514.90.2.288

Tsonos D, Ikospentaki K, Kouroupetrolgou G. 2008. Towards modeling of Readers' Emotional State response for the automated annotation of documents. *2008 IEEE International Joint Conference on Neural Networks IEEE World Congress on Computational Intelligence* 3253–3260. doi:10.1109/IJCNN.2008.4634260

VandenBos GR. 2013. *APA dictionary of clinical psychology*. (G. R. VandenBos, Ed.)*APA dictionary of clinical psychology*. American Psychological Association. doi:10.1037/13945-000

Watson D, Clark LA, Tellegen A. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology 54*:1063–1070. doi:10.1037/0022-3514.54.6.1063

Wegge J, Dick R Van, Fisher GK, West M a., Dawson JF. 2006. A Test of Basic Assumptions of Affective Events Theory (AET) in Call Centre Work1. *British Journal of Management 17*:237–254. doi:10.1111/j.1467-8551.2006.00489.x

Weiss H, Cropanzano R. 1996. Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior 18*:1–74. Retrieved from http://psycnet.apa.org/psycinfo/1996-98665-001

Westermann R, Spies K. 1996. Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology 26*:557–580. doi:0.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4

Zajenkowski M, Goryńska E, Winiewski M. 2012. Variability of the relationship between personality and mood. *Personality and Individual Differences 52*:858–861. doi:10.1016/j.paid.2012.01.007

Figure 1

A photograph for the creativity task.

Figure 2

The first level of the Tower of London game.

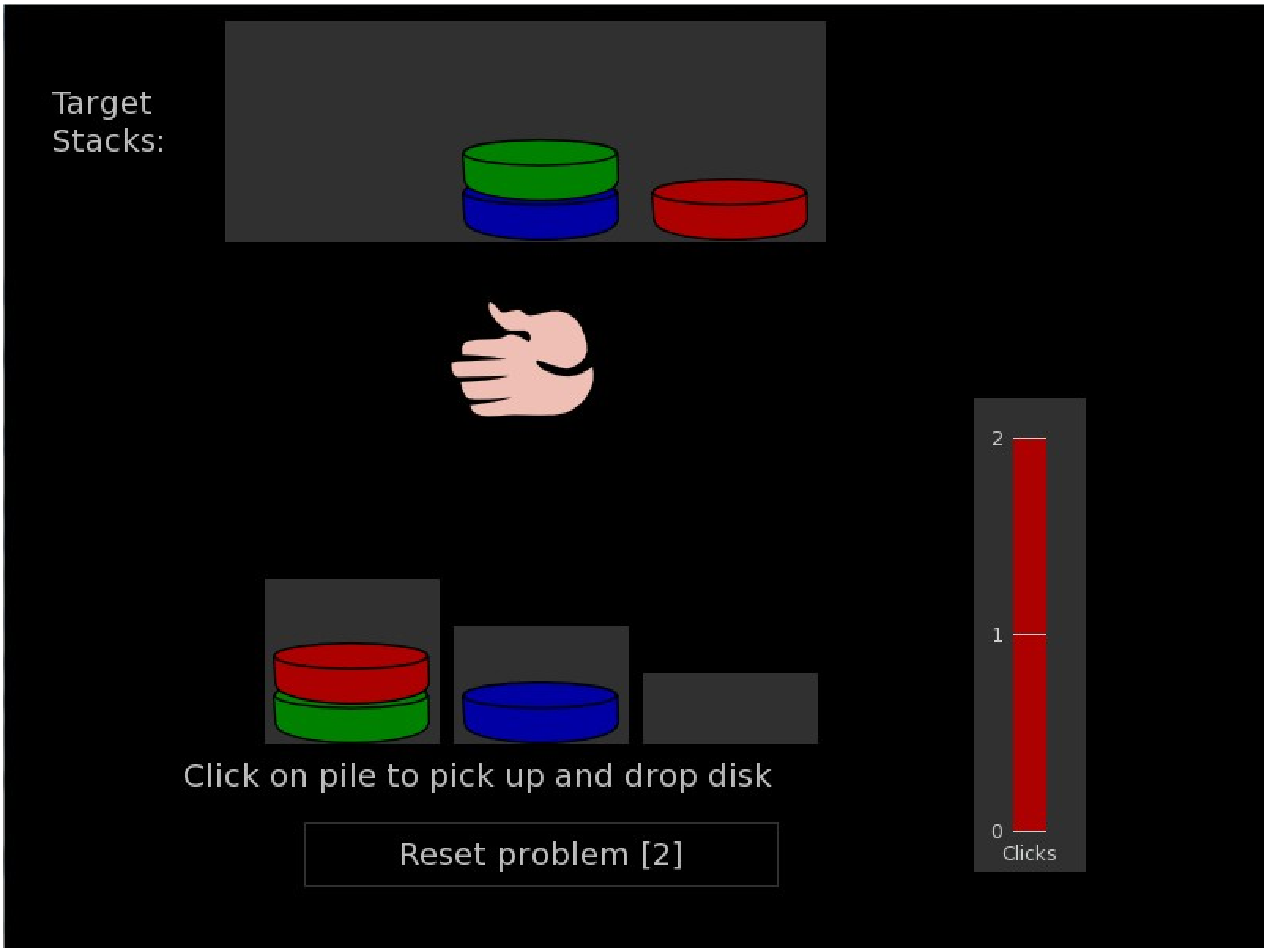

Table 1 (on next page)

Mean and Standard Deviation of the Task Scores divided by the groups.

	N-POS		POS	
Variable	M (SD)	95% CI	M (SD)	95% CI
ACR	3.13 (0.45)	[2.92, 3.35]	3.08 (0.58)	[2.81, 3.35]
BCR	4.02 (0.76)	[3.67, 4.38]	3.98 (0.76)	[3.63, 4.32]
NCR	4.70 (2.34)	[3.60, 5.50]	5.90 (3.46)	[4.00, 7.50]
APS	0.14 (0.04)	[0.12, 0.17]	0.20 (0.08)	[0.17 0.25]

Figure 3

Boxplots for the analytical problem-solving (APS) of the N-POS and POS groups.

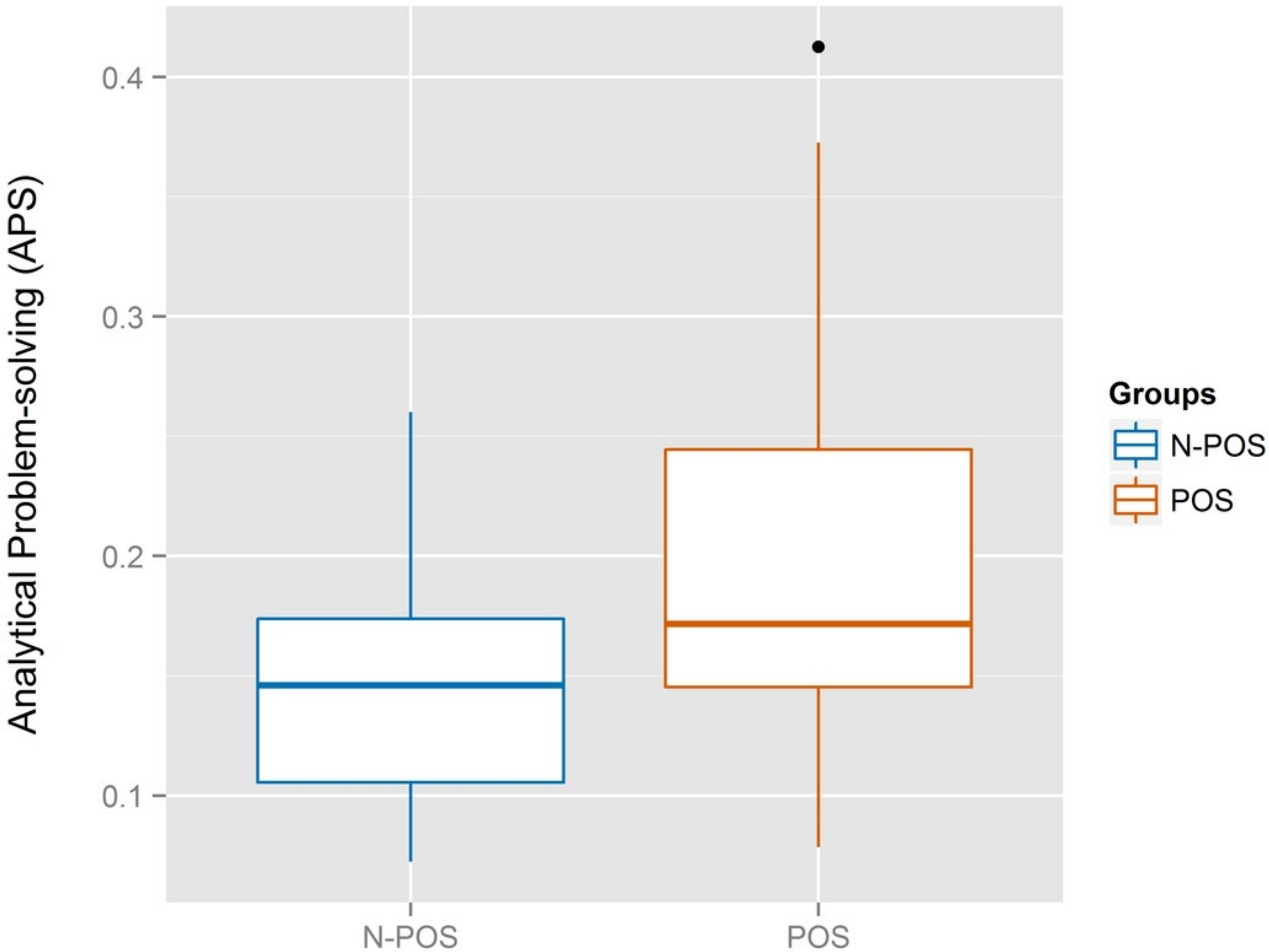

Figure 4

Scatterplot for the analytical problem-solving (APS) vs. the affect balance (SPANE-B) between the N-POS and POS groups.