Dysfunctional User States in Interface Use and Their Dependency on Work Environment and Task Complexity

Alexandr V. Yakunin(✉) and Svetlana S. Bodrunova

St. Petersburg State University, St. Petersburg 199004, Russia
{a.yakunin,s.bodrunova}@spbu.ru

Abstract. Usability testing today comprises quantitative and qualitative approaches. Within the former, factors that shape test results include four major ones put together within the 'contextual fidelity' model. These factors include product features, task complexity, user traits (including cultural belonging and gender), and experiment settings. However, most usability tests only consider one or two, not four of them. Our earlier research [1] has shown that, when four factors are assessed in parallel, cumulative impact of all of them makes test results highly diverge. Our current study complements this research by showing how usability tests diverge when task complexity varies highly. Another gap in usability research is that they miss the point in their final target, as they measure relative efficiency of interfaces for various groups of users but do not study formation of dysfunctional psychological states that critically prevent efficient Internet use and task performance. This happens despite the growing evidence of mental harm brought by interfaces to, e.g., youngsters' health or office workers. By testing 60 assessors in either groups or individually on tasks that induce dysfunctional states, namely monotony and anxiety, and introducing tasks of varying complexity, we show that excessive cognitive load leads to rapid rise of user dysfunctionality. Our results suggest that anxiety reduction in real-world tasks may not be reached by reducing task complexity. We recommend group performance on monotonous tasks and individual performance on anxiety-inducing tasks. By that, we illustrate our doubt of the possibility to reach any 'objective' results of usability tests; the latter need to be treated as fundamentally conditioned by the contextual fidelity factors. This has two consequences: First, usability testing needs to be conducted in accordance with prospective goals of interface use; second, multi-functional interfaces of general use need to pass through multiple usability tests that would combine contextual fidelity factors in various ways.

Keywords: usability · usability testing · contextual fidelity model · user states · dysfunctional state · monotony · anxiety · interface · human-computer interaction

1 Introduction

Usability testing today comprises quantitative and qualitative approaches. Quantitative studies use objective measurements and aim at assessing both objective user performance and subjective user satisfaction, also measured by standardized instruments, unlike in

A. Coman and S. Vasilache (Eds.): HCII 2023, LNCS 14025, pp. 470–485, 2023.
https://doi.org/10.1007/978-3-031-35915-6_34

qualitative studies where assessment is often based on user narratives about human-computer interaction (HCI) experience.

Quantitative usability research has gradually accumulated knowledge on factors that critically affect (and, thus, may highly distort) the results of usability tests. Beside the qualities of the tested product itself (say, more or less harmonious interface design or more or less efficient navigation), task features, user traits, and experiment settings have been shown to highly influence test results. Several scholars have urged the research community to take these factors into account [1–5], as they were seen as key in the search for the optimal usability testing model. The four factors – product features (including web aesthetics), task features (especially task complexity), user traits (including socio-demographic features and cultural belonging), and experiment conditions (especially individual/group testing) – have been united in the so-called 'contextual fidelity' model [6] which we will below describe in more detail. However, this model, despite being formulated over a decade ago, has not gained proper popularity among the usability researchers. Most quantitative studies of usability, including those utilizing eye tracking and other instrumental facilities, have ignored the complex character of impact of the 'contextual fidelity' factors and habitually use research designs that employ only one or two factors and disregard their potential cumulative impact.

Another research gap that is extremely wide in usability testing in both academe and industry is the ultimate goal of testing. Most research aims at detecting some optimal features of interfaces that would allow for optimal, most efficient interface use (the best objective performance), as well as for the best levels of subjective user satisfaction. However, the other end of the quality spectrum is rarely taken into consideration. This other end is manifested via user states of critically low functional efficiency. Meanwhile, the theory of user states that comes from psychology of professions [7–11] has already formulated the concept of dysfunctional user states that critically prevent efficient work performance. Despite the growing evidence of negative impact of interface use upon mental states of younger audiences, office workers, or high-risk interface-dependent professionals (like military pilots), dysfunctional user states have not been the major focus of usability studies.

In 2022, we have contributed to today's usability research by uniting these two research gaps and showing that certain combinations of 'contextual fidelity' factors could lead, first, to significant differences in testing results and cumulative effects that arose from those combinations of influences and, second, to formation of dysfunctional user states [1]. However, in our previous research pipeline, tasks only varied one-dimensionally, as they induced two opposing dysfunctional states, namely monotony and anxiety. Today, we expand this research by making tasks vary multi-dimensionally. The second dimension will be task complexity, which will vary significantly, while the tasks will still aim at inducing dysfunctional states. To this, we add variations in experiment conditions, setting either individual or group test conditions for our assessors, as our previous research suggested that group conditions could compensate for or diminish the level of dysfunctionality that formed during the tests. Combining these two factors would better explain how task complexity relates to user dysfunctionality in various conditions of interface use.

To reach this goal, we test 60 assessors in either group or individual setting on tasks that induce dysfunctional states, namely monotony and anxiety. We provide the assessors with three tasks of varying complexity for each of the potential dysfunctional state.

The remainder of the paper is organized as follows. Section 2 provides an account on the 'contextual fidelity' model and on dysfunctional user states. Section 3 describes our methodology and the research pipeline. Section 4 demonstrates our results, and Sect. 5 discusses them. We conclude by posing the questions on future prospects of usability studies being conditioned by the 'contextual fidelity' factors.

2 User Dysfunctionality and Factors that Condition It in Usability Testing

2.1 The Concept of Dysfunctional User State and Its Application to Usability Research

Conceptualization of user states has mostly developed within Russian psychology of professions, but some works have also been published in English and have gained international recognition. We find them useful for detection of types of interfaces and tasks that critically affect user efficiency.

According to Leonova [7, 10, 11], the impact of functional states on the emotional-volitional and cognitive sphere of a person is complex. Under the influence of certain activities, the most important mental processes and phenomena undergo significant changes. The sensitivity threshold of sensory systems and the tone of the sympathetic nervous system change [9], as well as stable changes are observed in the ability to memorize information, longer-term memory, reactions to stimuli, and the dynamics of intellectual-representative systems [12]. In particular, according to [8], the growth of mental fatigue or subjective anxiety causes changes in the whole spectrum of psychophysiological characteristics of a given user. Not only the sensitivity of analyzers changes, but also the ability to concentrate, to search in long-term memory, and to manipulate memorized objects.

According to the systemic approach in studying the functional systems of the human psyche, such changes in psychophysiological characteristics are part of a complex systemic reaction of an individual to factors that condition his/her activity during the work process [8: 1133]. Such an understanding of functional changes as reactions formed under external impact requires consideration of deviations from the state of operational rest as a response of functional systems on different levels to the factors/conditions of the activity performed by a worker (in our case, an interface user). As a rule, in studies based on theory of activity, the following factors are named that shape functional states: The content of the activity, conditions for its implementation, working person's functions, and its individual psychological traits [13, 14]. These factors, as we will see below, correspond well to the 'contextual fidelity' factors that have been studied in usability research, regretfully unrelated to wider psychology.

Due to the integral, systemic nature of functional states, the nature of the changes they cause in the psychophysiological state of users gains specific, recognizable, and stable shape. Via assessing psychophysiological indicators of the dominant user state in a particular moment *and* the factors that caused them, the researcher can accurately identify functional states, differentiate them both functionally and qualitatively, and describe them through combinations of indicators. For instance, as it follows from the results of several studies including our own, qualitative differences in the states of anxiety and monotony cause different effects on the user state. According to [9], the development of anxiety entails a change in the sense of time, transformation of the structure of intellectual operations, decrease in control over the performance quality, and a shift motivation [13]. The development of monotony is accompanied by subjective (boredom, apathy, drowsiness) and objective (decrease in the level of wakefulness and in the tone of the sympathetic nervous system) manifestations [15, 16].

Attempts to formulate conceptual foundations for detection of functional states in modern research have been made repeatedly. The first stream of studies is represented by the research dedicated to more general issues in the physiology of formation and development of functional states in human-machine interaction and methodological support for their diagnostics. In these studies, formation of (dys)functional states depends on both fundamental factors of their genesis [10] and external conditions of professional activity [14, 15]. Within the framework of such a universal perspective, approaches to classifying (dys)functional states are also presented, and their main types (operational rest, anxiety, monotony, fatigue, and psycho-emotional stress) are identified. Attempt were even made to characterize changes in psychophysiological traits specific to particular functional states [9]. However, these studies employ too general, universal indicators of functional states without taking into account the dynamics of user activity and without connection with task features. They do not pay much attention to shifts in functional states potentially brought by changes in intensity of influencing factors. The model they provide is static and, thus, less relevant to web interaction.

The second, smaller stream consists of studies that focus upon to the typology of affective and communicative states in human-computer interaction. This approach, based on detection and analysis of behavioral cues, attempts to identify persistent types of behavioral responses depending on user states [17]. In such works, the goal of the study is, as a rule, determining the behavioral patterns that accompany certain user states, e.g., the state of 'technological addiction' [18] or consumer behavior [19]. A special place in this group is occupied by studies on patterns of navigational behavior [20–22]. The design of these studies that aim at modeling user preferences in choosing a path through the website architecture best reflects the methodological flaw inherent in the entire group: The typology of user states and behavioral patterns is built upon machine processing of datasets of behavioral data, while external and internal factors in the formation of behavioral reactions and states are not considered.

Thus, as our review shows, neither of the research streams pays enough attention to.

the relationship between the intensity of exposure to factors that foster dysfunctionality and the intensity of experiencing dysfunctional states, which we see as a significant gap in word dysfunctionality studies. One of the promising directions in this regard is the study of the relationship between the degree of cognitive load and the intensity of the experienced dysfunctional state. The level of cognitive load is of decisive importance for the development of dysfunctions in one's cognitive-representative system [12].

2.2 The 'Contextual Fidelity' Model and Functional State Patterns

As we stated above, in usability studies, the concept of dysfunctional user states may be seen as consistent with the 'contextual fidelity' model, as the latter assesses similar variables, including user traits, product features, task complexity, and experimental settings. The model was proposed in 2011 and further developed in 2018 [16] and 2019 [2]. According to this model, various parameters related to product quality, user traits, and the test task may significantly affect the accuracy of testing. In this case, the user's functional state derives from the combined impact of the product, task, and testing environment, and is further conditioned by the user him-/herself.

Assessing the process of activity via the model, we get a flexible tool that allows for determining the external and internal factors affecting the workflow, at the intersection of which, a certain type of 'interaction experience' forms. Such combinations that link independent variables (e.g., the intensity of task plus the impact of the testing environment) and target variables (e.g., the level of user performance plus changes in the user's functional state) can be called *functional state patterns.* Such patterns describe stable linkages between combinations of 'contextual fidelity' factors, user performance, and user's functional state. The patterns are actually a sort of cumulative effects that persist and affect user performance. Finding them may be one of the goals of usability studies, as they show that interfaces need to be checked against formation of such patterns of inefficiency – that is, against particular combinations of 'contextual fidelity' factors.

In 2021–2022, we explored the impact of web aesthetics upon the dynamics of user experience as two different adverse functional states developed, namely monotony and anxiety. We have shown that functional state patterns critically affect the results of usability testing; we have detected three cumulative effects that were playing both for and against higher quality of user performance [1, 23, 24]. In these works, however, the 'contextual fidelity' factors were fixed in a binary way – e.g., aesthetic/non-aesthetic design, individual/group performance, Western/Eastern provenance of assessors, and monotony-/anxiety-inducing tasks. Thus, the intensity of each parameter was not taken into account. Moreover, in creating the tasks, we oriented them to being inductive in terms of a particular dysfunctional state but did not deal with actual task complexity which is another task feature that is, by common sense, to additionally induce dysfunctionality as task complexity grows. At the same time, today's research (including our own earlier works) suggests that, of the whole four-factor 'contextual fidelity' model, the main factors that regulate the users' cognitive load are the task complexity (to a greater extent) and the impact of the testing environment (to a lesser extent). We will check user performance against the two 'contextual fidelity' factors, thus adding to discovery of dysfunctional state patterns.

2.3 The Research Questions and Hypotheses

Given all stated above, we have formulated the following simple research questions:

RQ1. How does task complexity affect formation of the dysfunctional states of monotony and anxiety in usability tests?

RQ2. How does individual/group regimes of testing affect user performance on the tasks of varying complexity? What are the dysfunctional state patterns that we can see under the combined impact of testing environment and task complexity?

In accordance with these questions and with our previous findings, we put forward a range of hypotheses, on the basis of which we developed the design of the experiment. The hypotheses on individual testing correspond to RQ1, while those on group testing open up the RQ2.

Individual testing:

H1a. In individual testing, intellectual lability drops the deeper, the higher task complexity for both types of tasks.

H1b. In individual testing, emotional stress rises in accordance with the growth of task complexity via the rise of indicators respective to each of the two dysfunctional states. Thus, fatigue and anxiety will steadily grow for the monotony- and anxiety-inducing tasks, respectively.

H1c. In individual testing, full-fledged dysfunctional states that combine low intellectual lability and high fatigue/anxiety (respective to the conditions of task completion) will develop for the tasks of middle-range and high complexity.

Group testing:

H2a. In group testing of monotony, in line with our previous results [1], intellectual lability does not drop significantly for easy tasks, as group-based distraction compensates for the impact of cognitive load. However, it drops for mid-range and complex tasks for similar percentages.

H2b. In group testing of anxiety, the group environment does not compensate for the drop of intellectual lability.

H2c. In group testing of emotional stress for both types of tasks, drops in the respective indicators will be smaller than in individual testing.

H2d. In group testing, full-fledged dysfunctional states only form for complex tasks.

3 The Research Method

In order to study in more detail the relationship between the intensity of the impact of 'contextual fidelity' factors (in particular, task complexity that induces varying cognitive load) and the intensity of experiencing dysfunctional states, we have developed an experiment that includes tasks of various levels of complexity fulfilled by assessors in two formats, namely the group one and the individual one.

3.1 The Experimental Design

According to a number of studies, the level of cognitive load that varies due to varying task complexity, is crucial for performance and productivity in solving interactive communication problems in human-computer interaction [25, 26]. At the same time, the correlation between the complexity of the task and the efficiency of its solution is inversely proportional: As the complexity of the task decreases, the accuracy and speed of its execution increase, and as the complexity grows, they fall [27].

In usability testing, the most common and relevant practice is to analyze the complexity of a task at three levels, namely structural, semantic, and cognitive.

The structural complexity defined by the information architecture of a web interface refers to the potential range of alternative paths to the target information in the hierarchical structure of a website. As studies in this area show, there is a relationship between the depth of target information, patterns of user behavior in the search process, and the subjectively perceived complexity of the search task [28]. The depth of the path to the target information is, thus, one of the key factors that affect the users' cognitive states.

The semantic complexity of the task is determined by the obviousness of the path to the target information, as stated in the task description. The formulation of the search goal in this case either indicates the optimal path to solving the problem or masks it making the task more difficult. At the same time, the relevance of the path is measured by the degree of semantic closeness between the description of the search goal and the wording of the headings of navigation links [29].

The cognitive complexity of a task is determined by the number of components involved simultaneously in the process of intellectual activity. Tasks like comparison or classification imply the achievement of goals based on several criteria and places higher demands on the user's cognitive resources. In experimental research, the Miller criterion is widely used as a metric for cognitive complexity [30], according to which the efficiency of intellectual operations depends upon the number of objects that the assessor is able to simultaneously manipulate in his/her short-term memory [31]. The threshold value here is 5 ± 2, and, beyond this value, the task becomes very difficult for most representatives of the online audience.

Thus, task complexity of the turns out to be a function of the interaction of the structural (depth), semantic (relevance), and cognitive aspects of the navigation path. Low task complexity is characterized by small depth, high relevance, and the number of objects of intellectual activity less than seven, while high complexity implies significant depth (long path to target information), low relevance of the goal description, and the number of objects of intellectual activity exceeding the threshold value.

In accordance with this approach, we have identified three types of complexity for our prospective tasks, with three corresponding levels of cognitive load on the user:

- 'Easy' tasks with the minimal and fairly comfortable level of cognitive load;
- 'Medium-level' tasks with a threshold level of cognitive load, close to the psychophysiological limit, beyond which the efficiency of decisions should drop sharply;
- 'Complex' tasks of a high level of complexity, implying mobilization of maximum user's cognitive resources in the process of task solving.

For more accurate monitoring of changes in the user states, we have chosen the method of sequential complication of the task based on increasing the values of a discrete parameter with a minimum step according to the formula $(n + 1)$ as the basis for measuring structural complexity. In particular, the number of levels of website architecture that needs to be passed to the target content acts as a metric. As a criterion of semantic complexity, the ratio of the wording of the search goal and menu headings was chosen, and, for cognitive complexity, the number of criteria in intellectual operations that had to be operated simultaneously was fixed.

As a result, three tasks with different expected cognitive load were compiled, the complexity criteria for which are presented in detail in Table 1.

In essence, the task itself for inducing monotony and anxiety is of the same type; it aims at information search. However, an essential part of the task is how it is to be

Table 1. Tank complexity as part of the research design

Complexity level	Level of cognitive load	Structural complexity (the number of website layers)	Cognitive complexity (the number of criteria in smart operations)	Semantic complexity (goal statement vs. the menu headings)
easy	minimum	2	1	complete match
medium	threshold	3	2	non-complete match
complex	excessive	4	3	

Table 2. Assessor groups based on the research design, 5 assessors per group

Test format	Dysfunctional state	Task complexity level		
		Easy	Medium	Complex
Individual test	Monotony	sub-group1	sub-group5	sub-group9
	Anxiety	sub-group2	sub-group6	sub-group10
Group test	Monotony	sub-group3	sub-group7	sub-group11
	Anxiety	sub-group4	sub-group8	sub-group12

performed. Via additionally conditioning the tasks, we made them be more monotony-inducing or anxiety-inducing (see below).

As a task, we offered the assessors to evaluate and sort Master programs offered by Syracuse University, USA according to a range of criteria. The university website was chosen due to its recent redesign and high usability qualities. The target information bits that had to be found and evaluated by the assessors were the following:

1. Presence/absence of GRE (Graduate Record Examination) results in the admission requirements.
2. Presence/absence of GPA/TOEFL exam results in the admission requirements.
3. The minimum eligible amount of credit hours.
4. Presence/absence of final essay in the requirements for completion of the study program.
5. A requirement to present an official transcript of achievement as undergraduate.
6. The number of training trajectories.
7. The amount of financial support available via the scholarship program (in %).

In accordance with what is stated above, the differences between the tasks were determined by the website 'layer' on which the entry point for navigation situated, as well as by the number of sorting criteria. The 'easy' task meant searching for curricula according to the first three criteria to the depth of no more than two levels of architecture, while the criteria were clearly defined by the task statement. The task of medium complexity implied comparison of programs within the architecture of three levels of immersion, based on five criteria. In the high-complexity task, the search was carried out four levels deep and across the entire list of criteria. In the ‘medium’ and ‘complex’ tasks, the formulations of the criteria were identical and did not correspond to particular page headings.

Just as in our previous studies, we aimed at checking the formation of two dysfunctional states, namely monotony and anxiety. The former is to appear in the conditions of repetitive completion of a large number of typical operations with an insignificant creative component. The main factors that induce monotony, in our task, were the unawareness of the fixed time for task completion and a significant number of repetitive operations of comparing educational programs in order to sort them. The dysfunctional state of anxiety is formed in conditions of lack of time and information when solving a problem. Also, the development of this state may be associated with unforeseen changes in operating modes: for example, with a sharp complication of the task or a sudden change in its conditions combined with an unforeseen reduction in the time to solve. The induction of the anxiety mode was fostered by an unexpected reduction in timing and a change in the goal three minutes after the start of the test.

3.2 Measurements and Instruments

To measure the quality of user experience, we have chosen the following indicators for tracking the dynamics of assessors’ (dys) functional states.

Intellectual Lability. This metric characterizes the ability to switch attention, that is, the ability to quickly move from solving one problem to solving another one without making mistakes. According to the theory of functional states, high levels of anxiety provoke impulsivity and an excessively high rate of reactions; in the case of monotony, the rate decreases down to apathy. We measure this indicator using the so-called Gorbov – Schulte table [32], a short test of attention switching before and after the experiment.

Emotional Condition. Here, we use four metrics that, taken together, make up the integral indicator of a user’s emotional state. In accordance with the method of self-assessment of emotional states by Wessman and Ricks [33], we intend to measure the level of anxiety (calmness/anxiety scale), fatigue (energy/ fatigue scale), arousal (excitement/depression scale), and confidence (self-confidence/helplessness scale).

Experimental Design. To test the hypotheses, we developed the 4x3 study design. Test tasks were performed in twelve subgroups of assessors, 5 assessors per group. Each group worked with a task of a certain level of complexity in one of the two testing formats (either group or individual) and in one of two cognitive modes of ac-tivity, that is, in conditions that fostered either anxiety or monotony. Thus, our exper-iment included 60 assessors and had the following design (see Table 2).

The experiment had the following steps.

1. Input testing for the cognitive and emotional state, used to fix the initial state of cognitive efficiency and emotional stress in ‘operational comfort’ of each assessor:
   - assessment of the current level of emotional stress (self-assessment of emotional conditions by Wessman and Ricks);

- test on intellectual lability (via the Gorbov – Schulte table).

2. Performing tasks of one of three levels of complexity that form dysfunctional states (either anxiety and monotony) in a group/individual test.
3. Output testing for the cognitive and emotional state, used to fix changes in the indicators of cognitive efficiency and emotional stress of each user.

Data Analysis and Interpretation. For each assessor, the results of input and output testing were fixed. Then, in each group, mean values and standard deviations for them were calculated for each indicator. After that, the difference between the indicator values before and after task completion (delta, or Δ) was calculated, and its mean and its standard deviation were, too. The deltas are the secondary indicator that allows for comparing individual and group testing, as well as the very assessment of the user cognitive and emotional states. As the emotional indicators were measured from 10 (the best state) to 1 (the worst state), drops in values indicate the growth of dysfunctionality. The same is true for intellectual lability, though it is measured in points that combine time and precision of finding the targets on the Gorbov – Schulte table.

4 Results

The results of our experiment are shown in Table 3 (for the individual testing environment) and Table 4 (for the group one).

RQ1. As the results show, the changes in assessors' functional states vary depending on the level of task complexity. This is especially noticeable in reactions to the cognitive load from attention and intellect via the intellectual lability indicator. However, the results for indicators of emotional stress point to multi-directional effects caused by combinations of testing settings and task complexity.

Individual task completion. In line with expectations, for both monotony and anxiety, intellectual lability decreases least intensely when the 'easy' task is solved. In particular, intellectual lability drops by ~ 15% (187.15 to 171.13, $\Delta = -16.02$) for the anxiety-inducing task and by ~ 8.5% (from 187.05 to 182.93, $\Delta = -4.12$) for the monotony-inducing one. It is worth noting that the anxiety-inducing task has caused twice as big a drop in lability than the monotonous task. For mid-range task complexity, expectedly, lability drops more significantly but nearly equally for both task types – for ~ 23% for anxiety and for ~ 27% for monotony. This pattern continues for the complex tasks: Lability drops for ~ 45% anxiety and for ~ 44% for monotony, thus making users lose nearly a half of their initial multi-tasking capacity. This confirms **H1a** and hints that, for more complex tasks, complexity plays a bigger role in defining the user states than the nature of the task; with growth of complexity, anxiety- and monotony-inducing task solving becomes equally dominated by cognitive load, rather than by testing conditions.

H1b, though, cannot be confirmed, as the results for indicators of emotional stress differ from those expected. Thus, there is no pattern of steady growth of fatigue/anxiety with the growth of task complexity; in neither case (including the group results) emotions 'responsible' for a given dysfunctional state grow in relation to task complexity. For monotony, fatigue, indeed, grows on average, but, for 'easy' and mid-range tasks, the results are within the standard deviation interval, which tells that the assessors' results

Table 3. The results of individual testing

User state	Indicator		Task complexity level								
			Easy			Medium			Complex		
			before task	after task	*delta (Δ)*	before task	after task	*delta (Δ)*	before task	after task	*delta (Δ)*
Monotony	Intellectual lability		187,05 (1,057)	182,93 (1,228)	***-4,12 (1,504)***	188 (1,00)	132,2 (1,300)	***-50,8 (2,167)***	188,54 (0,756)	105,6 (1,148)	***-82,94 (1,357)***
	Emotional stress	Calmness/ anxiety	7 (0,471)	7 (0,471)	*0 (0,666)*	6,8 (0,447)	7 (0,707)	*0,2 (0,836)*	7 (0)	7 (0,707)	*0 (0,707)*
		Energy/ fatigue	6,9 (1,197)	5,5 (0,527)	*-1,4 (1,505)*	6,6 (1,513)	5,4 (0,547)	*-1,2 (1,923)*	7 (0,707)	5,2 (0,447)	***-1,8 (0,836)***
		Excitement/ depression	6,9 (0,567)	6,9 (316)	*0 (0,816)*	7,2 (0,447)	6,8 (0,447)	*-0,4 (0,894)*	8 (0)	7 (0,707)	***-1 (0,707)***
		Self-conf./ helplessness	6.8 (1,475)	7,1 (1,100)	*0,3 (1,766)*	6,8 (1,095)	6,8 (1,303)	*0 (0,707)*	6,6 (1,140)	6,6 (1,140)	*0 (0,707)*
Anxiety	Intellectual lability		187,15 (1,285)	171,13 (3,067)	***-16,02 (2,791)***	194,2 (3,193)	149,4 (2,073)	***-44,8 (2,049)***	191 (2,549)	105,05 (2,132)	***-85,94 (3,112)***
	Emotional stress	Calmness/ anxiety	7,1 (0,875)	6,1 (0,875)	***-1 (0,471)***	7,2 (1,095)	6,2 (1,095)	***-1 (0)***	7,2 (1,095)	6,2 (1,095)	***-1 (0)***
		Energy/ fatigue	6 (0,666)	6 (0,666)	*0 (0,666)*	5,6 (0,547)	5,8 (0,836)	*0,2 (0,447)*	5,7 (0,447)	5,7 (0,670)	*0 (0,353)*
		Excitement/ depression	6,9 (0,567)	7 (0,471)	*0,1 (0,567)*	7 (0,707)	7,2 (0,447)	*0,2 (0,447)*	6,4 (0,547)	7,4 (0,547)	***1 (0,707)***
		Self-conf./ helplessness	7,1 (0,370)	7,3 (0,251)	*0,2 (0,788)*	6,8 (1,095)	6,8 (1,303)	*0 (0,707)*	6,8 (1,095)	6,8 (1,303)	*0 (0,707)*

were too divergent to form a stable pattern. Fatigue significantly grows only for the ‘complex’ task. In case of anxiety, calmness drops equally for all the levels of complexity; this pattern also repeats for the group performance, however, with nearly two times higher rise of anxiety. Such stable drops of calmness could be induced by the very fact of task completion; however, this does not seem to be true, as the figures for group work are stably bigger than those for individual task solving, and are non-significant for monotonous tasks. More probably, the rises of anxiety are induced by the task conditions, and, in this case, it is anxiety that dominates over complexity in shaping the users’ state. Due to worse performance in groups, as judged by the level of anxiety, we recommend that the tasks potentially leading to high anxiety are performed in the individual regime.

Thus, it is for ‘complex’ tasks only that full-fledged dysfunctional states may be detected; this is why they were not discovered in our previous studies which did not offer complex enough tasks to the assessors. Moreover, in each case, the dysfunctional states not only combine low lability and high stress by the corresponding indicator (either fatigue or anxiety) but is also complemented by changes in one other emotional indicator. For monotony, depression also rises in both individual and group testing, which is quite understandable intuitively. For anxiety, though, a counter-intuitive rise in excitement is observed (which is also noticeable for ‘easy’ and mid-range tasks but is not significant), and this pattern does not repeat for group testing (see below). Thus, **H1c** is confirmed for high-complexity tasks but not for borderline cognitive load.

Table 4. The results of group testing

User state		**Indicator**	**Task complexity level**								
			Easy			**Medium**			**Complex**		
			before task	after task	*delta (Δ)*	before task	after task	*delta (Δ)*	before task	after task	*delta (Δ)*
Monotony		Intellectual lability	190,2 (0,836)	188,2 (6,906)	*-2 (6,442)*	185 (4,527)	127,5 (1,581)	***-57,5 (3,640)***	183,4 (3,189)	86,2 (2,196)	***-97,2 (3,701)***
	Emotional stress	Calmness/ anxiety	6 (0)	6 (0,707)	*0 (0,707)*	6 (0,547)	6 (0,707)	*0 (0,547)*	6 (0,707)	6 (0,707)	*0 (0,707)*
		Energy/ fatigue	6,4 (0,547)	6,2 (0,447)	*-0,2 (0,836)*	7 (0,707)	6 (0,707)	*-1 (1,224)*	7,6 (0,547)	6 (0,707)	***-1,6 (0,894)***
		Excitement/ depression	7,2 (0,836)	7 (0)	*-0,2 (0,836)*	7,2 (0,836)	7 (0)	*-0,2 (0,836)*	7,8 (0,447)	6,6 (0,547)	***-1,2 (0,447)***
		Self-conf./ helplessness	6,6 (0,894)	7,4 (0,547)	*0,8 (0,836)*	7 (0)	7 (0)	*0 (0)*	7,4 (0,547)	7,4 (0,547)	*0 (0,707)*
Anxiety		Intellectual lability	188,64 (1,499)	159,1 (3,398)	***-29,54 (4,450)***	189,2 (1,303)	132,4 (5,594)	***-56 (6,140)***	193 (2,915)	96,5 (2,179)	***-96,5 (3,240)***
	Emotional stress	Calmness/ anxiety	7 (1,870)	5,2 (1,788)	***-1,8 (0,836)***	7 (1,870)	5,2 (1,788)	***-1,8 (0,836)***	6,8 (1,303)	5,2 (1,483)	***-1,6 (1,140)***
		Energy/ fatigue	6,4 (0,547)	7,2 (0,447)	*0,8 (0,836)*	7 (0)	7 (0)	*0 (0)*	6,4 (0,894)	6,4 (1,140)	*0 (1,870)*
		Excitement/ depression	7 (0,707)	7,6 (0,547)	*0,6 (0,894)*	7 (0,707)	8 (0,707)	*1 (1,224)*	6,6 (0,894)	7,6 (0,547)	*1 (1,224)*
		Self-conf./ helplessness	6,8 (0,447)	7,4 (0,547)	***0,6 (0,547)***	6,8 (0,447)	7,8 (0,836)	***1 (0,707)***	6,8 (0,447)	5,6 (0,894)	***-1,2 (0,836)***

RQ2. The group testing has, in general, demonstrated similar dysfunctional state patterns. In both individual and group testing, emotional stress indicators mostly responded in an expected way: Thus, for monotony, there was growth of fatigue and no changes in anxiety; for anxiety, there was growth of anxiety and no changes in fatigue.

Group Testing Environment. Here, we juxtapose the results of individual and group testing. First, intellectual lability for the 'easy' task solved in groups is the only case of insignificant lability change, in line with our previous research [1]. If for individual testing, intellectual lability dropped by ~ 15% for anxiety and by ~ 8.5% for monotony, for the group test, the decrease in intellectual lability was very similar (~15.6%) for anxiety but insignificant for monotony (~1%; $\Delta = -2$, SD = 6.442). For the 'easy' task, we see that group testing compensates for pressures created via insignificant cognitive load; however, this is only true for monotony-, not for anxiety-inducing performance procedures. For mid-range and 'complex' tasks, lability drops sharply and similarly for both monotony and anxiety (~31%/ ~ 30% and ~ 53%/50%, respectively). This fully confirms **H2a.** This is why we here repeat our recommendation for group-based completion of monotonous tasks, as group-based distraction compensates for the impact of cognitive load. This, however, may not work for more complex tasks. One more consequence of our study is the following: We could test any given task for complexity and identify whether this or that task is easy enough if group testing compensates for intellectual lability drops, and lability remains unchanged.

For anxiety, as we have just stated, no compensation happens, and lability drops in a notable way even for 'easy' tasks (~15,6%), and this drop intensifies with the growth of complexity. Thus, **H2b** is also confirmed.

As to the emotional stress, we had expected that group testing would compensate for negative emotions and allow for lower indicator values. However, **H2c** has to be rejected. We could only assess the emotional stress indicators for high-complexity tasks where they have shown significant drops; the changes of emotional states are multi-directional. There are only two stable patterns that we could detect. First, as already stated above, for the anxiety-inducing task, group values on anxiety are nearly two times worse than those for the monotonous task, contrary to expectations. Second, in both individual and group testing, excitement slightly (and insignificantly) rises for the anxiety task, while self-confidence behaves highly unpredictably. Thus, for individual testing, it remains completely unchanged, and in group testing it first rises significantly (allegedly, group support allows for better self-confidence when the task is solved), but then sharply drops. This result is not counter-intuitive, as successful solution of 'easy' and 'mid-range' tasks may foster a rise of self-confidence, while the excessive task complexity may leave people irritated and dissatisfied.

Just as in individual testing, in group testing, full-fledged dysfunctional states only form under the impact of tasks of maximum complexity. This shows that group testing does not fully compensate for drops in performance efficiency, and complex enough tasks may eventually destroy functionality of web users. **H2d** is confirmed.

5 Discussion and Conclusion

After our research, one may say that stable patterns of emotional-cognitive reactions are observed that are specific for tasks of different levels of complexity. However, accumulation of dysfunctionality is non-linear, and certain dysfunctional state patterns deserve special attention.

Monotonous tasks cause relatively less dysfunctionality on low and middle levels of complexity. When solving an 'easy' task in both group and individual conditions, cognitive functions are close to operational rest in both individual and group regimes of testing. On medium levels of complexity, only intellectual capacities diminish in a notable way (approximately one fourth of initial capacity), while emotional stress grows insubstantially. But as soon as the level of complexity grows beyond a certain level, the dysfunctional state forms, with a nearly half loss of brain lability and growth of fatigue and depression. Here, the sharp break in performance beyond the threshold load points an area for future research in the field of managing the formation of dysfunctions.

Anxiety-inducing tasks cause somewhat bigger drops of intellectual lability, which also stably increases with the growth of task complexity. What differs the anxiety-inducing tasks from monotonous ones is immediate rise of emotional stress exactly via loss of calmness; anxiety remains on the same level in both individual and group environments regardless of task complexity, which may mean that anxiety reduction in real-world HCI cannot be reached by reducing task complexity and introducing simpler, less complex tasks. This deserves immediate attention by education professionals and interface developers. However, as anxiety in groups is nearly two times higher on average than that registered for individual performance, it is not that assessors' anxiety reaches its maximum from the very beginning; individual format allows for lesser growth of anxiety, as well as prevents the feeling of helplessness. This is a counter-intuitive finding, as, usually, group task solution is characterized by lesser helplessness due to potential mutual help of participants and lesser expression of helplessness in front of fellow assessors. However, here, we see that group performance enhances helplessness even for low- and mid-range task complexity.

For anxiety-inducing tasks, the pattern of sharp formation of dysfunctionality is less evident for individual testing, where brain lability gradually diminishes with the growth of complexity but anxiety remains stably higher than in rest. However, for group performance, the rise and then sharp fall of self-confidence reveals that sharp drops of functionality may also happen for anxiety tasks.

Thus, the formation of dysfunctional states is shaped by a combination of inner conditions of task solving and task complexity, as, both types of dysfunctionality sharply rise in any external conditions of task solving. Along with that, group performance may compensate for the drops of intellectual lability for monotonous tasks, while enhances helplessness for anxiety-inducing tasks. Thus, our recommendations that aim at harm reduction call for group performance on monotonous tasks and, vice versa, individual performance on anxiety-inducing tasks.

To conclude, we need to state one thought that derives from our studies of the 'contextual fidelity' model. As 'contextual fidelity' factors are multiple and their combinations affect testing results, this poses a simple question: What do we see as the 'objective' result of quantitative usability testing? Is it the result of individual or group testing? Perceiving usability test results as objective might be misleading. What is needed instead is conduction of target-oriented usability research, where the usability test model would correspond to prospective uses of a given interface. E.g., if an interface implies office work or use in school class, group tests need to be compulsory, while interfaces for personal use need to pass testing in individual regimes. This also implies that multi-function interfaces, for which one cannot easily predict preferential modes of use need to pass multiple testing with various combinations of 'contextual fidelity' factors, in order to detect the threshold cognitive loads and find out how and when dysfunctionality arises.

Acknowledgements. . This research has been supported in full by the project 'Center for International Media Research' of St. Petersburg State University, year 3, #94033584.

References

1. Yakunin, A.V., Bodrunova, S.S.: Cumulative impact of testing factors in usability tests for human-centered web design. Future Internet **14**(12), 359 (2022)
2. Sauer, J., Sonderegger, A., Heyden, K, Biller, J., Klotz, J., Uebelbacher, A.: Extra-laboratorial usability tests: An empirical comparison of remote and classical field testing with lab testing. Appl. Ergon. **74**, 85–96 (2019)

3. Doi, T., Doi, S., Yamaoka, T.: The peak-end rule in evaluating product user experience: the chronological evaluation of past impressive episodes on overall satisfaction. Hum. Fact. Ergon. Manufact. Serv. Indust. **32**(3), 256–267 (2022)
4. Cha, S.S., Lee, S.H.: The effects of user experience factors on satisfaction and repurchase intention at online food market. J. Indust. Distribut. Bus. **12**(4), 7–13 (2021)
5. Holden, R.J., et al.: Human factors engineering and human-computer interaction: supporting user performance and experience. Clin. Inform. Study Guide: Text Rev., 119–132 (2022)
6. Sauer, J., Sonderegger, A.: The influence of product aesthetics and user state in usability testing. Behav. Inf. Technol. **30**(6), 787–796 (2011)
7. Leonova, A.B.: Functional status and regulatory processes in stress management. In: Hockey, G. R.J., Gaillard, A.W.K., Burov, O., (Eds.) Operator functional state: the assessment and prediction of human performance degradation in complex tasks, pp. 36–52. IOS Press, Amsterdam (2003)
8. Ushakov, I.B., Bogomolov, A.V., Kukushkin, Y.A.: Psychophysiological mechanisms of formation and development of functional states. Russian J. Physiol. (Formerly I. M. Sechenov Physiol. J.) **100**(10), 1130–1137 (2014)
9. Ushakov, I.B., Bogomolov, A.V., Kukushkin, Y.: A: Patterns of operator functional states [Patterny funktsional'nyh sostoyaniy operatora]. Nauka, Moscow (2010)
10. Leonova, A.B., Medvedev, V.I.: Functional states of a person in labor activity: A textbook [Funktsional'nye sostoyaniya cheloveka v trudovoy deyatel'nosti]. Lomonosov Moscow State University Publishing House, Moscow (1981)
11. Leonova, A.B.: The concept of human functional state in Russian applied psychology. Psychol, Russia: State of the Art **2**, 517–538 (2009)
12. Chuprikova, N.I.: Human reaction time: Physiological mechanisms, verbal-semantic regulation, connection with intelligence and properties of the nervous system [Vremya reaktsiy cheloveka: Fiziologicheskie mehanizmy, verbal'no-smyslovaya regulyatsiya, svyaz' s intellektom i svoystvami nervnoy systemy]. Yask Publishing House, Moscow (2019)
13. Suvorova, G.A.: Psychology of activity [Psihologiya deyatelnosti]. PER SE Publishing House, Moscow (2003)
14. Kaptelinin, V., Nardi, B.: Activity Theory in HCI: Fundamentals and Reflections. Morgan & Claypool Publishers, San Rafael (2012)
15. Bodrov, V.A.: Psychological foundations of professional activity [Psihologicheskie osnovy professional'noy deyatel'nosti]. PER SE / Logos, Moscow (2007)
16. Ilyin, E.P.: Psychophysiology of human states [Psihpfiziologiya sostoyaniy cheloveka]. Piter, St. Petersburg (2005)
17. Scherer, S., et al.: A generic framework for the inference of user states in human computer interaction. J. Multimodal User Interf. **6**, 117–141 (2012)
18. Laor, T., Galily, Y..: In WAZE we trust? GPS-based navigation application users' behavior and patterns of dependency. PLoS ONE **17**(11), e0276449 (2022)
19. Li, X., Zhao, X., Pu, W.: Measuring ease of use of mobile applications in e-commerce retailing from the perspective of consumer online shopping behaviour patterns. J. Retail. Consum. Serv. **55**, 102093 (2020)
20. Diwandari, S., Hidayat, A.T.: Comparison of classification performance based on dynamic mining of user interest navigation pattern in e-commerce websites. J. Phys. Conf. Ser. **1844**(1), 012025 (2021)
21. He, Q., Borgonovi, F., Suárez-Álvarez, J.: Clustering sequential navigation patterns in multiple-source reading tasks with dynamic time warping method. J. Comput. Assist. Learn. (2022). https://doi.org/10.1111/jcal.12748
22. Gao, Y., Cui, Y., Bulut, O., Zhai, X., Chen, F.: Examining adults' web navigation patterns in multi-layered hypertext environments. Comput. Hum. Behav. **129**, 107142 (2022)

23. Yakunin, A.V., Bodrunova, S.S.: Website aesthetics and functional user states as factors of web usability. In: Ahram, T., Taiar, R. (eds.) IHIET 2021. LNNS, vol. 319, pp. 394–401. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-85540-6_51
24. Yakunin, A.V., Bodrunova, S.S.: Cumulative distortions in usability testing: combined impact of web design, experiment conditions, and type of task and upon user states during internet use // design, user experience, and usability: ux research, design, and assessment. In: 11th International Conference, DUXU 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Proceedings, pp. 526–535. Springer Nature, Cham (2022)
25. Campbell, D. J.: Task complexity: A review and analysis. Academy of Management Review, 13(1), 40–52 (19880
26. Hong, W., Thong, J.Y., Tam, K.Y.: The effects of information format and shopping task on consumers' online shopping behavior: A cognitive fit perspective. J. Manag. Inf. Syst. **21**(3), 149–184 (2004)
27. Byström, K.: Information and information sources in tasks of varying complexity. J. Am. Soc. Inform. Sci. Technol. **53**(7), 581–591 (2002)
28. Gwizdka, J., Spence, I.: What can searching behavior tell us about the difficulty of information tasks? A study of web navigation. In: Proceedings of the 69th Annual Meeting of the American Society for Information Science and Technology (ASIS&T), vol 43. Information Today, Inc, Medford (2006), asistdl.onlinelibrary.wiley.com/doi/pdf/https://doi.org/10.1002/meet.14504301167, (Assessed 23 Feb 2023)
29. Van Oostendorp, H., Ignacio Madridb, R., Carmen Puerta Melguizo, M.: The effect of menu type and task complexity on information retrieval performance. Ergonom. Open J. **2**(1), 64–71 (2009)
30. Jiang, Z., Benbasat, I.: The effects of presentation formats and task complexity on online consumers' product understanding. MIS Q. **31**(3), 475–500 (2007)
31. Michailidou, E., Harper, S., Bechhofer, S.: Visual complexity and aesthetic perception of web pages. In: Proceedings of the 26th Annual ACM International Conference on Design of communication, pp. 215–224. ACM (2008)
32. Mashin, V.A.: Analysis of heart rate variability based on the graph method. Hum. Physiol. **28**, 437–447 (2002)
33. Wessman, A.E., Ricks, D.F.: Mood and personality. Holt, Rinehart and Winston, New York (1966)