Supplements for

User-Centered Design Metrics and Motivational Aspects for Basic Study Design

Philipp Ziebell – University of Würzburg – Institute for Psychology

Psychological Intervention, Behavior Analysis, and Regulation of Behavior (Prof. Dr. Andrea Kübler)

Differential Psychology, Personality Psychology, and Psychological Diagnostics (Prof. Dr. Johannes Hewig)

User-Centered Design (UCD): Objective Measures

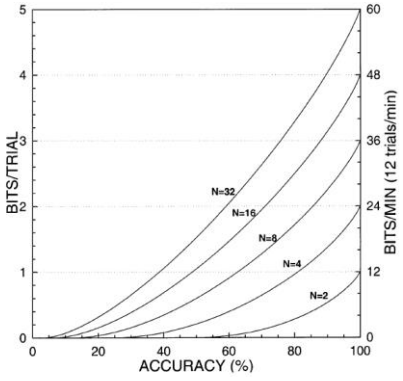
Effectiveness: Accuracy

A measure of how accurate and complete a user can accomplish a BCI-controlled application is how often the intended output can be achieved. Accuracy relates successful selections to the total number of attempted selections and can be expressed in percentage of correct responses, preferably online (vs. offline) [1, 2].

Efficiency: Information transfer rate (ITR) – utility metric

Efficiency relates the costs, i.e. effort and time, invested by the user to effectiveness. An objective measure of efficiency is the ITR and its modifications with regards to error probability, accuracy, and practicality. A common phenomenon in a BCI-controlled application is that high ITR can be achieved despite numerous miss-selections if the number of possible selections is high. However, such a BCI would be of no practical value as no meaningful communication would be possible. Thus, the utility metric was introduced and takes into account, that with an accuracy below 50%, no reliable communication can be achieved, i.e. ITR is 0 bits/minute for all accuracies below 50% [1, 2], see also following figure [3] and example [4].

Fig. 4. Information transfer rate in bits/trial (i.e. bits/selection) and in bits/min (for 12 trials/min) when the number of possible choices (i.e. N) is 2, 4, 8, 16, or 32. As derived from Pierce (1980) (and originally from Shannon and Weaver, 1964), if a trial has N possible choices, if each choice has the same probability of being the one that the user desires, if the probability (P) that the desired choice will actually be selected is always the same, and if each of the other (i.e. undesired) choices has the same probability of selection (i.e. $(1 - P)/(N - 1)$), then bit rate, or bits/trial (B), is: $B = \log_2 N + P \log_2 P + (1 - P)log_2[(1 - P)/(N - 1)]$. For each N, bit rate is shown only for accuracy $\geq 100/N$ (i.e. ≥chance) (from Wolpaw et al., 2000a).



Example

$$B = \log_2 N + P * \log_2 P + (1-P) * \log_2 \frac{1-P}{N-1} \quad (1)$$

With B standing for bits per selection, N standing for number of possible selection-targets and P standing for the estimated probability of a correct classification, based on the empirically found online-accuracy. To calculate the ITR, B was multiplied with the number of possible selections per minute (SPM), using the following formula (2), with S standing for the number of stimulus-selection-sequences that was chosen for each participant at each session and taking into account the duration of each stimulus-selection-sequence (3.75 s) as well as the post-stimulus-selection-sequence break (4 s):

$$SPM = \frac{60s}{S * 3.75s + 4s} \quad (2)$$

UCD: Subjective Measures

Efficiency: Workload

Workload constitutes a subjective measure of efficiency and is classically measured by the NASA-TLX. Workload in the NASA-TLX is defined as a ''hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance''. The NASA-TLX measures the overall workload experienced while operating a specific application and identifies main sources of workload, which is estimated across the dimensions mental, physical, and temporal demand, and performance, effort, and frustration. Subjective workload for each dimension has to be rated on twenty step bipolar scales with scores from 0 to 100. A weighting procedure combines the individual scores for each dimension into one total score. It has previously been used to assess workload of healthy subjects during BCI operation [1, 2].

Details and example see following figures [5, 6].

RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	*Low/High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	*Low/High*	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	*Good/Poor*	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

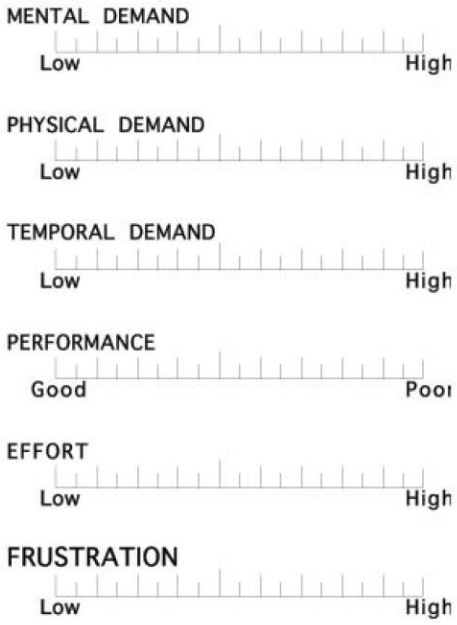


Table 1

Number of Approximate Results on Google Scholar™ for Common Workload Measures as of June 2015

Search Term	Results
"NASA TLX"	10,300
SWAT workload	3,620
"Cooper Harper" workload	2,260
"Subjective Workload Dominance"	229
"Bedford workload"	177

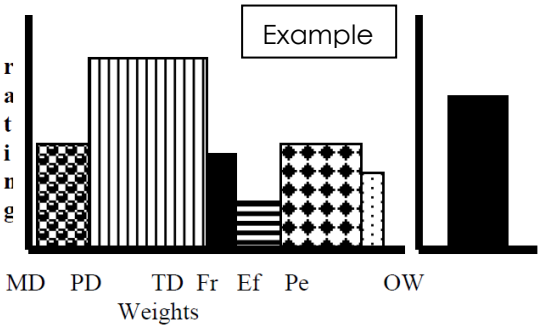


Fig 1: Graphic Representation of weighted subscale ratings and an overall workload value

Satisfaction: Overall satisfaction – BCI-related aspects – interview

User satisfaction refers to the perceived comfort and acceptability while using the product. We suggest several measures to assess device satisfaction. The Extended QUEST 2.0 (see following page) is not suitable to be applied after every BCI session as it requires time and it is unlikely that basic aspects contributing to satisfaction change substantially across sessions with the same BCI-controlled application. However, we consider it valuable to obtain a coarse rating of overall satisfaction at the end of each BCI session. Visual analogue scales (VAS) can provide such a measure. Users can be easily asked after each session to indicate their overall satisfaction on a VAS ranging from 0 (not at all satisfied) to 10 (maximally satisfied). Such a rating does not provide any in depth information about sources of satisfaction/dissatisfaction, but it allows for easy monitoring specifically in long-term studies [1, 2].

Concrete examples from a relatively basic study design see subsequently [4].

Example

Dear participant,

Please complete the following questions:

Please use the line below to rate how difficult it has been for you to control the BCI. The number 0 means that it was not difficult at all and the number 10 means that it was extremely difficult. As a marker, please put a cross in the area that you think best represents your experience.

0 = not difficult at all 10 = extremely difficult

Please use the line below to rate how satisfied you are with the BCI system.

0 = not satisfied at all 10 = extremely satisfied

Could you imagine using a BCI for communication? What problems would need to be addressed? What would be major improvements? Did you notice anything else?

_____

_____

_____

Satisfaction: General aspects of assistive technology

The Quebec User Evaluation of Satisfaction with assistive technology (QUEST 2.0), allows for quantifying satisfaction with general aspects of a product. The questionnaire consists of items that cover 12 aspects. The QUEST 2.0 is considered invalid if scores for more than six (of 12) satisfaction items are missing; thus, it is important to check whether the items are adequate to assess a specific application. We considered the items ''durability, service delivery, repairs/servicing, and follow-up services'' inadequate for the evaluation of a BCI-controlled application during development and removed those from the questionnaire. ''Durability'' was removed also because EEG amplifiers have already demonstrated their long-term functionality, electrodes have to be replaced depending on the frequency of use, and our evaluation procedure did not span a time frame of years such that durability could become an issue. Items are rated on a Likert-type scale from 1 to 5. Whenever users are not ''very satisfied'' they are invited to comment. The arithmetic mean across all items provides the total satisfaction score. Demers and colleagues explicitly invite researchers to add few items to render the questionnaire more suitable for a specific piece of technology. Concrete example see below [1, 2, 7, 8, https://www.midss.org/sites/default/files/questeng.scoring_sheetpdf_0.pdf].

Example

Table 3

Satisfaction ratings of N = 4 end-users after the free spelling, email and internet task with regard to (1) different aspects of the BCI QW-device (extended QUEST 2.0) and (2) overall device satisfaction (VAS)

(1) Extended QUEST 2.0 (ratings from 1-5)													Average (M)* over users and tasks
	End-user A			End-user B			End-user C			End-user D			
TASKS	**FS**	**E**	**I**	**FS**	**E**	**I**	**FS**	**E**	**I**	**FS**	**E**	**I**	
Items													
Dimensions	5	5	5	2	2	2	4	5	5	3	2	2	3.50
Weight	5	5	5	4	2	3	5	5	5	4	3	3	4.08
Adjustment	4	4	4	3	3	3	3	3	4	2[a]	2	2[a]	3.08
Safety	5	5	5	5	4	4	5	5	5	5	4	4	4.67
Comfort	4	4	4	3[a]	3[a]	3[a]	3	4	4	3[a]	4[a]	3[a]	3.50
Ease of use	5[a]	4[a]	4[a]	4	3	4	4	4	4	2	3	2	3.58
Effectiveness	4[a]	5[a]	4[a]	4[a]	3[a]	3[a]	4[a]	4[a]	4[a]	3	3[a]	2	3.58
Prof. services (information/ instructions)	5	5	5	4	4	4	5	5	5	4	4	4	4.50
QUEST total score	**4.75**	**4.63**	**4.50**	**3.63**	**3.00**	**3.38**	**4.13**	**4.38**	**4.50**	**3.25**	**3.13**	**2.75**	
Reliability	5[a]	4[a]	5[a]	4	4[a]	3[a]	5[a]	4[a]	5[a]	3	3	4	4.08
Speed	5	5	4	4	2	4	3[a]	3[a]	3[a]	2[a]	2[a]	2[a]	3.25
Learnability	5	5	5	3[a]	4	4	5	5	5	4	4	4	4.42
Aesthetic design	4	4	4	2	2	2	4	4	4	2	2	3	3.08
Added items total score	**4.75**	**4.50**	**4.50**	**3.25**	**3.00**	**3.25**	**4.25**	**4.00**	**4.25**	**2.75**	**2.75**	**3.25**	
(2) VAS overall device satisfaction (0-10)													
	End-user A			End-user B			End-user C			End-user D			
TASKS	**FS**	**E**	**I**	**Fs**	**E**	**I**	**Fs**	**E**	**I**	**Fs**	**E**	**I**	
	6.3	6.0	6.8	5.7	7.6	8.6	7.4	9.0	7.5	5.7	4.6	4.2	

Note. FS=Free spelling task, E=Email task, I=Internet task; (1) QUEST 2.0: 1=not satisfied at all, 2=not very satisfied, 3=more or less satisfied, 4=quite satisfied, 5=very satisfied; M=mean; a=three most important aspects out of the 12 items. (2) VAS: 0=not satisfied at all, 10=absolutely satisfied. *Means were calculated only to give an idea of overall performance and were not interpreted, because the number of subjects did not permit parametric statistics.

Usage: Match between product and user – overall usability – use in daily life

For a detailed description with a focus on hearing the needs of clinical users see [2].

Motivational Aspects

See overview in following figure [9] and concrete examples on following page [4].

Table 1 | Summary of desirable properties of a good instructional design with corresponding suggestions to improve human training protocols for BCI.

Level	Properties of a good instructional design	Corresponding suggestions for BCI training protocols
Feedback	- Non-evaluative and supportive feedback (Hattie and Timperley, 2007; Shute, 2008) - Feedback that conducts to a feeling of competence (Ryan and Deci, 2000)	Provide positive feedback (feedback only indicating when the user did right) only for beginners, and disconfirmatory feedback for advanced users
	- Clear and meaningful feedback (Hattie and Timperley, 2007)	Start with a subject-independent classifier for users with poor initial performances
	- Explanatory and specific feedback (Hattie and Timperley, 2007; Shute, 2008) (Moreno and Mayer, 2007) - Feedback that signals a gap between current and desired performances (Hattie and Timperley, 2007; Shute, 2008)	Provide more information about what was right or wrong about the EEG patterns produced by the user: - Provide as feedback the value of a few (less than seven) relevant EEG features - Provide as feedback some measure of quality of the mental imagery
	- Multimodal feedback (Ainsworth, 2006) (Merrill, 2007)	Provide a multimodal feedback (e.g., visual + haptic), with the same granularity and specificity for each modality, with some redundancy between them
	- Engaging feedback and environment (Ryan and Deci, 2000)	Represent the feedback as an interaction with a game element (e.g., a 3D car)
Instructions	- Goals should be clearly defined (Hattie and Timperley, 2007; Shute, 2008)	Expose the real goal of BCI training, i.e., to produce clear, specific and stable EEG patterns
	- The meaning of the feedback should be explained (Ainsworth, 2006)	Explain what the BCI feedback means, particularly for non-intuitive feedback such as the classifier output.
	- Prior knowledge should be activated (Merrill, 2007; Moreno and Mayer, 2007) - The skill to be learned should be demonstrated (Merrill, 2007)	- Instruct the users to remember situations in which they used the task they will imagine - Demonstrate successful BCI use and BCI feedback during correct task performance
Tasks	- Progressive and adaptative tasks (Ainsworth, 2006; Merrill, 2007) - Tasks that are challenging but still achievable (Hattie and Timperley, 2007; Shute, 2008)	Use adaptive BCI training protocols with increasing difficulty (e.g., progressively increasing the number of mental tasks to be mastered)
	- Need for autonomy and work at the user's own pace (Ryan and Deci, 2000; Shute, 2008) (Moreno and Mayer, 2007)	Include more training sessions with free and/or self-paced BCI use
	- Motivation and positive emotions promote learning (Ryan and Deci, 2000; Um et al., 2012)	Using positive emotion-inducing training tasks e.g., including gaming mechanisms
	- Need for variability over tasks and problems (Sweller et al., 1998; Ainsworth, 2006)	Include variety in the mental tasks to be performed, e.g., change in speed or duration of the mental imagery
	- Adapt the training procedure to the student (Hattie and Timperley, 2007; Shute, 2008)	Matching BCI training protocols to users' characteristics

It should be noted that such suggestions are only based on theory, and will need to be formally validated.

Example

Properties of a good instructional design	Corresponding design-aspects in the current study
Feedback	
- Non-evaluative and supportive feedback, that conducts to a feeling of competence	- Positive feedback after correct selection (not given after wrong selection)
- Engaging feedback and environment	- "Star Wars"-specific sound (R2D2)
Instruction	
- Goals should be clearly defined	- Briefing and Debriefing (explanation of paradigm and research purpose, suggesting strategies for successful use)
- The meaning of the feedback should be explained	- Explanation of the P300 and ways of subjectively influencing it (Johnson, 1986, 1993), showing Calibration results to users
Task	
- Progressive and adaptive tasks	- New Calibration in each Training-Session, allowing users to reach online-accuracies > 70% without guaranteeing near 100%
- Need for autonomy and work at user's own pace	- Adjustable breaks during Training-Sessions, with time to drink or eat little refreshments
- Motivation and positive emotions promote learning	- "Star Wars"-Theme (Absolving "Missions" that were inspired by movie-scenes from "Star Wars"), creation of a positive atmosphere (little refreshments)
- Need for variability over tasks and problems	- Varying "Missions", each one to a new movie-scene from "Star Wars"
- Adapt the training procedure to the student	- Chance to try the alternative version of the Streaming Paradigm in the Transfer-Session after the three Training-Sessions

In addition to considering the listed motivational aspects, motivation can be assessed via relatively global measures (e.g. VAS) or via more specific measures (e.g. QCM-BCI) [4, 10, 11].

Literature

[1] Kübler, A., Holz, E. M., Riccio, A., Zickler, C., Kaufmann, T., Kleih, S. C., …, & Mattia, D. (2014). The user-centered design as novel perspective for evaluating the usability of BCI-controlled applications. *PLoS ONE, 9(12),* doi: 10.1371/journal.pone.0112392.

[2] Kübler, A., Nijboer, F., & Kleih, S. (2020). Hearing the needs of clinical users. In Ramsey, N. F. & Millán, J. R. (Eds.), *Handbook of Clinical Neurology*. Amsterdam: Elsevier.

[3] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, *113*(6), 767-791.

[4] Ziebell, P., Stümpfig, J., Eidel, M., Kleih, S. C., Kübler, A., Latoschik, M. E., & Halder, S. (2020). Stimulus modality influences session-to-session transfer of training effects in auditory and tactile streaming-based P300 brain–computer interfaces. *Scientific Reports, 10(1),* doi: 10.1038/s41598-020-67887-6.

[5] Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9)*, 904-908.

[6] Grier, R. A. (2015). How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59(1)*, 1727-1731.

[7] Demers, L., Weiss-Lambrou, R., & Ska, B. (2002). The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): an overview and recent progress. *Technology and Disability*, *14*(3), 101-105.

[8] Zickler, C., Riccio, A., Leotta, F., Hillian-Tress, S., Halder, S., Holz, E., … & Kübler, A. (2011). A brain-computer interface as input channel for a standard assistive technology software. *Clinical EEG and Neuroscience, 42(4),* 236-244.

[9] Lotte, F., Larrue, F., & Mühl, C. (2013). Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. *Frontiers in Human Neuroscience, 7,* doi: 10.3389/fnhum.2013.00568.

[10] Nijboer, F., Furdea, A., Gunst, I., Mellinger, J., McFarland, D. J., Birbaumer, N., & Kübler, A. (2008). An auditory brain–computer interface (BCI). *Journal of Neuroscience Methods, 167(1),* 43-50.

[11] Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). QCM: A questionnaire to assess current motivation in learning situations. *Diagnostica*, *47*(2), 57-66.

Examples for Further Reading

Chavarriaga, R., Fried-Oken, M., Kleih, S., Lotte, F., & Scherer, R. (2017). Heading for new shores! Overcoming pitfalls in BCI design. *Brain-Computer Interfaces, 4,* doi: 10.1080/2326263X.2016.1263916.

Choi, I., Rhiu, I., Lee, Y., Yun, M. H. & Nam, C. S. (2017). A systematic review of hybrid brain-computer interfaces: Taxonomy and usability perspectives. *PLoS ONE, 12(4),* doi: 10.1371/journal.pone.0176674.

Mladenović, J. (2021). Standardization of protocol design for user training in EEG-based brain–computer interface. *Journal of Neural Engineering*, *18*(1), 011003.