Modeling mental workload via rule-based expert system: a comparison with NASA-TLX & Workload Profile

Lucas Rizzo
lucas.rizzo@mydit.ie

September, 2016

Research question

- *Can implementations of* **rule-based expert systems**, *compared to state-of-the-art* **MWL inference techniques**, *enhance the modelling of* **mental workload** *according to* **sensitivity** *and* **validity***?*

Human mental workload

- Basically it could be described as the amount of necessary cognitive work devoting to a certain task during a period of time.

Human mental workload

- Basically it could be described as the amount of necessary cognitive work devoting to a certain task during a period of time.
- The principal reason for measuring mental workload is to quantify the mental cost of performing a task in order to predict operator and system performance (Cain, 2007). A few application areas are: psychology, ergonomics, aviation, interface and web design.

Human mental workload

- Basically it could be described as the amount of necessary cognitive work devoting to a certain task during a period of time.
- The principal reason for measuring mental workload is to quantify the mental cost of performing a task in order to predict operator and system performance (Cain, 2007). A few application areas are: psychology, ergonomics, aviation, interface and web design.
- Mental workload (MWL) is a multi-faceted phenomenon with no clear and widely accepted definition

Human mental workload

- Basically it could be described as the amount of necessary cognitive work devoting to a certain task during a period of time.
- The principal reason for measuring mental workload is to quantify the mental cost of performing a task in order to predict operator and system performance (Cain, 2007). A few application areas are: psychology, ergonomics, aviation, interface and web design.
- Mental workload (MWL) is a multi-faceted phenomenon with no clear and widely accepted definition
- All studied measures of MWL do not take into consideration conflicts between pieces of evidence which might lead to contraction and loss of information.

State-of-the-art inference techniques

- **Nasa Task Load Index** (Hart and Staveland, 1988):

State-of-the-art inference techniques

- **Nasa Task Load Index** (Hart and Staveland, 1988):
  - Multidimensional scale, initially developed for the use in the aviation industry. It consists of six subscales: mental demand, physical demand, temporal demand, frustration, effort and performance

State-of-the-art inference techniques

- **Nasa Task Load Index** (Hart and Staveland, 1988):
  - Multidimensional scale, initially developed for the use in the aviation industry. It consists of six subscales: mental demand, physical demand, temporal demand, frustration, effort and performance
  - The computation of an overall workload index is made through a weighted average of these six subscales

State-of-the-art inference techniques

- **Nasa Task Load Index** (Hart and Staveland, 1988):
  - Multidimensional scale, initially developed for the use in the aviation industry. It consists of six subscales: mental demand, physical demand, temporal demand, frustration, effort and performance
  - The computation of an overall workload index is made through a weighted average of these six subscales
- **Workload Profile** (Tsang and Velazquez, 1996):

State-of-the-art inference techniques

- **Nasa Task Load Index** (Hart and Staveland, 1988):
  - Multidimensional scale, initially developed for the use in the aviation industry. It consists of six subscales: mental demand, physical demand, temporal demand, frustration, effort and performance
  - The computation of an overall workload index is made through a weighted average of these six subscales
- **Workload Profile** (Tsang and Velazquez, 1996):
  - Based on 8 dimensions: perceptual/central processing, response selection and execution, spatial processing, verbal processing, visual processing, auditory processing, manual output and speech output

State-of-the-art inference techniques

- **Nasa Task Load Index** (Hart and Staveland, 1988):
  - Multidimensional scale, initially developed for the use in the aviation industry. It consists of six subscales: mental demand, physical demand, temporal demand, frustration, effort and performance
  - The computation of an overall workload index is made through a weighted average of these six subscales
- **Workload Profile** (Tsang and Velazquez, 1996):
  - Based on 8 dimensions: perceptual/central processing, response selection and execution, spatial processing, verbal processing, visual processing, auditory processing, manual output and speech output
  - The operator is asked to provide the proportion of attentional resources, in the range 0 to 1, for each dimension. The overall workload is a sum of these scores

Expert Systems

- An expert system is a computer program designed to model the problem-solving ability of a human expert (Jack Durkin and Durkin, 1998)

Expert Systems

- An expert system is a computer program designed to model the problem-solving ability of a human expert (Jack Durkin and Durkin, 1998)
- The expert has to provide a **knowledge base** (Longo and Dondio, 2014), which will be translated into rules and applied into an **inference engine** capable of returning an numerical index for the overall workload.

Expert Systems

- An expert system is a computer program designed to model the problem-solving ability of a human expert (Jack Durkin and Durkin, 1998)
- The expert has to provide a **knowledge base** (Longo and Dondio, 2014), which will be translated into rules and applied into an **inference engine** capable of returning an numerical index for the overall workload.
- The knowledge base was developed for the inference of MWL in the field of human computer interaction. It comprises 21 different attributes all quantified, through a subjective question, in the range $[0, 100] \in \mathbb{R}$.

Expert Systems

- Rules are mainly represented in the form IF ... THEN

Expert Systems

- Rules are mainly represented in the form IF ... THEN

- *Given a low degree of performance there is a reason to believe that the mental workload exerted by a user on task T is high*

Expert Systems

- Rules are mainly represented in the form IF ... THEN
- *Given a low degree of performance there is a reason to believe that the mental workload exerted by a user on task T is high*
  - IF low performance THEN overload workload

Expert Systems

- Rules are mainly represented in the form IF … THEN
- *Given a low degree of performance there is a reason to believe that the mental workload exerted by a user on task T is high*
  - IF low performance THEN overload workload
  - IF medium performance THEN optimal workload

Expert Systems

- Rules are mainly represented in the form IF ... THEN

- *Given a low degree of performance there is a reason to believe that the mental workload exerted by a user on task T is high*
  - IF low performance THEN overload workload
  - IF medium performance THEN optimal workload
  - IF high performance THEN underload workload

Inference Engine

- Four mental workload levels are defined: *underload*, *fitting*$^{-}$, *fitting*$^{+}$ and *overload*.

Inference Engine

- Four mental workload levels are defined: *underload*, *fitting*$^-$, *fitting*$^+$ and *overload*.
- Rules have to be processed in order to infer a workload level (*forecast rules*)

Inference Engine

- Four mental workload levels are defined: *underload*, *fitting*$^-$, *fitting*$^+$ and *overload*.
- Rules have to be processed in order to infer a workload level (*forecast rules*)
- Rules might deny other rules (*undercutting and rebutting rules*)

Inference Engine

- Four mental workload levels are defined: *underload*, *fitting*$^-$, *fitting*$^+$ and *overload*.
- Rules have to be processed in order to infer a workload level (*forecast rules*)
- Rules might deny other rules (*undercutting and rebutting rules*)
  - (1) low effort $\rightarrow$ underload workload (*forecast rule*)

Inference Engine

- Four mental workload levels are defined: *underload*, *fitting*$^{-}$, *fitting*$^{+}$ and *overload*.
- Rules have to be processed in order to infer a workload level (*forecast rules*)
- Rules might deny other rules (*undercutting and rebutting rules*)
  - (1) low effort $\rightarrow$ underload workload (*forecast rule*)
  - (2) high motivation $\rightarrow$ not low effort (*undercutting rule*)

Inference Engine

- Four mental workload levels are defined: *underload*, *fitting*$^-$, *fitting*$^+$ and *overload*.
- Rules have to be processed in order to infer a workload level (*forecast rules*)
- Rules might deny other rules (*undercutting and rebutting rules*)
  - (1) low effort $\rightarrow$ underload workload (*forecast rule*)
  - (2) high motivation $\rightarrow$ not low effort (*undercutting rule*)
  - (2) denies (1)

Inference Engine

- Four mental workload levels are defined: *underload*, *fitting*$^-$, *fitting*$^+$ and *overload*.
- Rules have to be processed in order to infer a workload level (*forecast rules*)
- Rules might deny other rules (*undercutting and rebutting rules*)
  - (1) low effort $\rightarrow$ underload workload (*forecast rule*)
  - (2) high motivation $\rightarrow$ not low effort (*undercutting rule*)
  - (2) denies (1)
- At the end of the process, from the set of forecast rules, there will be a resulting set of **surviving** rules

Inference Engine

- A MWL index has to be calculated according to the set of surviving rules.

Inference Engine

- A MWL index has to be calculated according to the set of surviving rules.
- Surviving rules are evaluated by a function according to the input value of their attribute.

Inference Engine

- A MWL index has to be calculated according to the set of surviving rules.
- Surviving rules are evaluated by a function according to the input value of their attribute.
- Finally, surviving rules values are aggregated based on the MWL level inferred by each of them. Four heuristics are constructed in order to capture the following ideas:

Inference Engine

- A MWL index has to be calculated according to the set of surviving rules.
- Surviving rules are evaluated by a function according to the input value of their attribute.
- Finally, surviving rules values are aggregated based on the MWL level inferred by each of them. Four heuristics are constructed in order to capture the following ideas:
  - $\boldsymbol{h_1}$: to give more importance to the largest point of view.

Inference Engine

- A MWL index has to be calculated according to the set of surviving rules.
- Surviving rules are evaluated by a function according to the input value of their attribute.
- Finally, surviving rules values are aggregated based on the MWL level inferred by each of them. Four heuristics are constructed in order to capture the following ideas:
  - $\boldsymbol{h_1}$: to give more importance to the largest point of view.
  - $\boldsymbol{h_2}$: to consider the pessimistic point of view.

Inference Engine

- A MWL index has to be calculated according to the set of surviving rules.
- Surviving rules are evaluated by a function according to the input value of their attribute.
- Finally, surviving rules values are aggregated based on the MWL level inferred by each of them. Four heuristics are constructed in order to capture the following ideas:
  - $\boldsymbol{h_1}$: to give more importance to the largest point of view.
  - $\boldsymbol{h_2}$: to consider the pessimistic point of view.
  - $\boldsymbol{h_3}$: to give equal importance to all surving rules.

Inference Engine

- A MWL index has to be calculated according to the set of surviving rules.
- Surviving rules are evaluated by a function according to the input value of their attribute.
- Finally, surviving rules values are aggregated based on the MWL level inferred by each of them. Four heuristics are constructed in order to capture the following ideas:
  - $\boldsymbol{h_1}$: to give more importance to the largest point of view.
  - $\boldsymbol{h_2}$: to consider the pessimistic point of view.
  - $\boldsymbol{h_3}$: to give equal importance to all surving rules.
  - $\boldsymbol{h_4}$: to give equal importance to all point of views.

Comparative criteria

- Since there is no ground truth indicating the quantity of mental workload imposed, several criteria have been proposed for the selection and development of measurement techniques (Eggemeier, 1988; Tsang and Velazquez, 1996)

Comparative criteria

- Since there is no ground truth indicating the quantity of mental workload imposed, several criteria have been proposed for the selection and development of measurement techniques (Eggemeier, 1988; Tsang and Velazquez, 1996)
- Two criteria were employed:

Comparative criteria

- Since there is no ground truth indicating the quantity of mental workload imposed, several criteria have been proposed for the selection and development of measurement techniques (Eggemeier, 1988; Tsang and Velazquez, 1996)
- Two criteria were employed:
    - **Validity**: determine whether the MWL measure is measuring MWL. Technique: *Correlation*. Two types of validity:

Comparative criteria

- Since there is no ground truth indicating the quantity of mental workload imposed, several criteria have been proposed for the selection and development of measurement techniques (Eggemeier, 1988; Tsang and Velazquez, 1996)
- Two criteria were employed:
  - **Validity**: determine whether the MWL measure is measuring MWL. Technique: *Correlation*. Two types of validity:
    - **Convergent validity**: determine to what extent a model correlate with other model of MWL

Comparative criteria

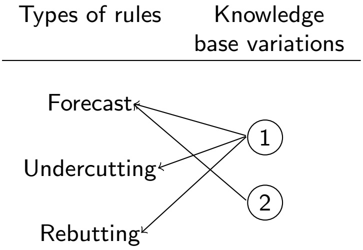
- Since there is no ground truth indicating the quantity of mental workload imposed, several criteria have been proposed for the selection and development of measurement techniques (Eggemeier, 1988; Tsang and Velazquez, 1996)
- Two criteria were employed:
  - **Validity**: determine whether the MWL measure is measuring MWL. Technique: *Correlation*. Two types of validity:
    - **Convergent validity**: determine to what extent a model correlate with other model of MWL
    - **Concurrent validity**: determine the extent to which a model correlate with task completion time

Comparative criteria

- Since there is no ground truth indicating the quantity of mental workload imposed, several criteria have been proposed for the selection and development of measurement techniques (Eggemeier, 1988; Tsang and Velazquez, 1996)
- Two criteria were employed:
  - **Validity**: determine whether the MWL measure is measuring MWL. Technique: *Correlation*. Two types of validity:
    - **Convergent validity**: determine to what extent a model correlate with other model of MWL
    - **Concurrent validity**: determine the extent to which a model correlate with task completion time
  - **Sensitivity**: investigate how many pairs of tasks each measure is capable of differentiating. Technique: *One way analysis of variance + post hoc*

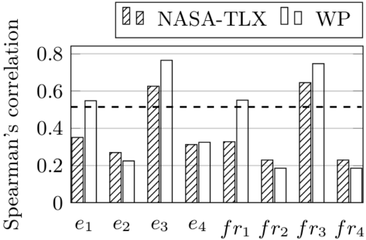
Design and methodology

Table: Experiments set up: types of rules employed by two variations of the same knowledge base (left) and name of each model, variation used, heuristic adopted (right).

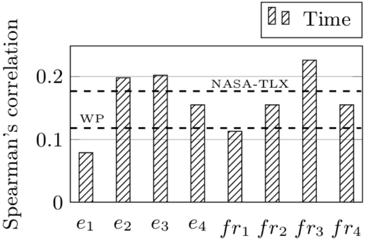


Model	KB variation		Heuristics			
	1	2	h_1	h_2	h_3	h_4
e_1	✓		✓			
e_2	✓			✓		
e_3	✓				✓	
e_4	✓					✓
fr_1		✓	✓			
fr_2		✓		✓		
fr_3		✓			✓	
fr_4		✓				✓

Results: validity

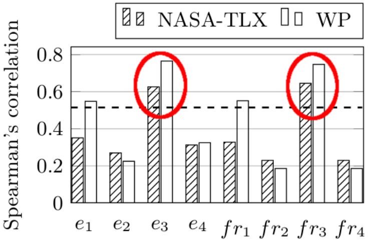


Convergent validity: $p < 0.05$.

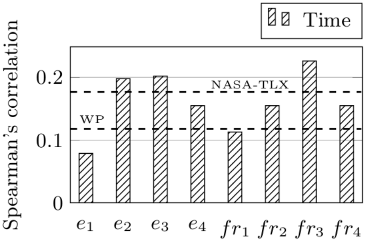


Concurrent validity: $p < 0.05$.

Results: validity

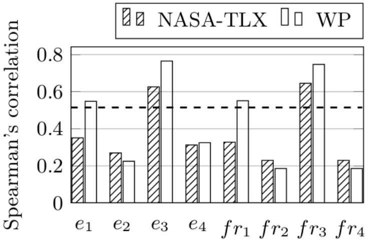


Convergent validity: $p < 0.05$.

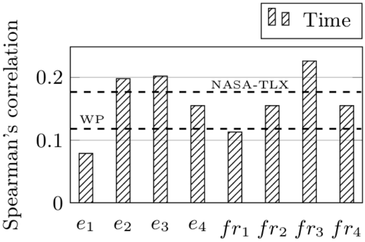


Concurrent validity: $p < 0.05$.

Results: validity

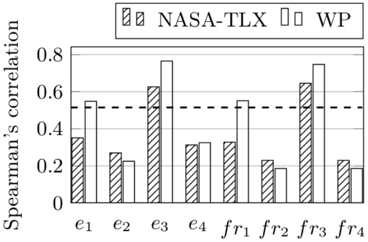


Convergent validity: $p < 0.05$.

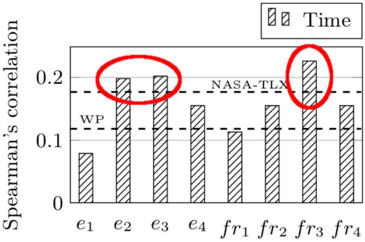


Concurrent validity: $p < 0.05$.

Results: validity



Convergent validity: $p < 0.05$.



Concurrent validity: $p < 0.05$.

Results: sensitivity

Table: Sensitivity of MWL models with Games-Howell post hoc analysis. The maximum pairwise comparisons of 9 tasks is $\binom{9}{2} = 36$).

Model	$p < 0.05$	$p < 0.01$
NASA-TLX	18	12
WP	9	4
e_1	2	1
e_2	5	3
e_3	13	10
e_4	0	0

Model	$p < 0.05$	$p < 0.01$
NASA-TLX	18	12
WP	9	4
fr_1	2	0
fr_2	4	1
fr_3	17	10
fr_4	4	1

Results: sensitivity

Table: Sensitivity of MWL models with Games-Howell post hoc analysis. The maximum pairwise comparisons of 9 tasks is $\binom{9}{2} = 36$).

Model	$p < 0.05$	$p < 0.01$
NASA-TLX	18	12
WP	9	4
e_1	2	1
e_2	5	3
e_3	13	10
e_4	0	0

Model	$p < 0.05$	$p < 0.01$
NASA-TLX	18	12
WP	9	4
fr_1	2	0
fr_2	4	1
fr_3	17	10
fr_4	4	1

Findings

- Successful mental workload modelling and assessment according to validity and sensitivity quantification using rule-based expert systems.

Findings

- Successful mental workload modelling and assessment according to validity and sensitivity quantification using rule-based expert systems.
- Rule-based expert systems can be seen as more appealing and dynamic:

Findings

- Successful mental workload modelling and assessment according to validity and sensitivity quantification using rule-based expert systems.
- Rule-based expert systems can be seen as more appealing and dynamic:
  - Rules are built with terms that are closer to the way humans reason.

Findings

- Successful mental workload modelling and assessment according to validity and sensitivity quantification using rule-based expert systems.
- Rule-based expert systems can be seen as more appealing and dynamic:
  - Rules are built with terms that are closer to the way humans reason.
  - It imitates experts problem-solving ability.

Findings

- Successful mental workload modelling and assessment according to validity and sensitivity quantification using rule-based expert systems.
- Rule-based expert systems can be seen as more appealing and dynamic:
  - Rules are built with terms that are closer to the way humans reason.
  - It imitates experts problem-solving ability.
  - Heuristics for aggregating rules with a better capacity of handling uncertainty and conflicting pieces of information compared to fixed formulas.

Findings

- Successful mental workload modelling and assessment according to validity and sensitivity quantification using rule-based expert systems.
- Rule-based expert systems can be seen as more appealing and dynamic:
  - Rules are built with terms that are closer to the way humans reason.
  - It imitates experts problem-solving ability.
  - Heuristics for aggregating rules with a better capacity of handling uncertainty and conflicting pieces of information compared to fixed formulas.
  - Easier comparison of knowledge-bases and beliefs of different MWL designers.

Conclusions and future work

- The researched presented a new way of modelling and assessing the construct of Mental Workload by means of rule-based expert systems.

Conclusions and future work

- The researched presented a new way of modelling and assessing the construct of Mental Workload by means of rule-based expert systems.
- Different heuristics were designed aimed at inferring MWL as a numerical index.

Conclusions and future work

- The researched presented a new way of modelling and assessing the construct of Mental Workload by means of rule-based expert systems.
- Different heuristics were designed aimed at inferring MWL as a numerical index.
- A comparison with two state-of-the-art MWL assessment techniques was presented.

Conclusions and future work

- The researched presented a new way of modelling and assessing the construct of Mental Workload by means of rule-based expert systems.
- Different heuristics were designed aimed at inferring MWL as a numerical index.
- A comparison with two state-of-the-art MWL assessment techniques was presented.
- Rule-based expert systems demonstrated to be a flexible approach for translating knowledge-bases and thus enhancing the understanding of the construct of mental workload itself.

Conclusions and future work

- The researched presented a new way of modelling and assessing the construct of Mental Workload by means of rule-based expert systems.
- Different heuristics were designed aimed at inferring MWL as a numerical index.
- A comparison with two state-of-the-art MWL assessment techniques was presented.
- Rule-based expert systems demonstrated to be a flexible approach for translating knowledge-bases and thus enhancing the understanding of the construct of mental workload itself.
- Future works include the use of different knowledge-bases, incorporation of fuzzy representation of rules and incorporation of acceptability semantics from argumentation theory.

Thank you!

Questions?

Bibliography I

Amgoud, L. & Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, *34*(1-3), 197–215.

Baroni, P., Guida, G., & Mussi, S. (1997). Full non-monotonicity: a new perspective in defeasible reasoning. In *European symposium on intelligent techniques* (pp. 58–62).

Bench-Capon, T. J. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, *13*(3), 429–448.

Bench-Capon, T. J. & Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial intelligence*, *171*(10-15), 619–641.

Besnard, P. & Doutre, S. (2004). Checking the acceptability of a set of arguments. In *Nmr* (Vol. 4, pp. 59–64).

Cain, B. (2007). *A review of the mental workload literature*. DTIC Document.

Chang, C. F., Ghose, A., & Miller, A. (2009). Mixed-initiative argumentation: a framework for justification management in clinical group decision support. In *Aaai fall symposium: the uses of computational argumentation*.

Craven, R., Toni, F., Cadar, C., Hadad, A., & Williams, M. (2012). Efficient argumentation for medical decision-making. In *Kr*. Citeseer.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, *77*(2), 321–357.

Durkin, J. [Jack] & Durkin, J. [John]. (1998). *Expert systems: design and development*. Prentice Hall PTR.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. *Advances in Psychology*, *52*, 41–62.

Fox, J., Black, L., Glasspool, D., Modgil, S., Oettinger, A., Patkar, V., & Williams, M. (2006). Towards a general model for argumentation services. In *Aaai spring symposium: argumentation for consumers of healthcare* (pp. 52–57).

Glasspool, D., Fox, J., Oettinger, A., & Smith-Spark, J. (2006). Argumentation in decision support for medical care planning for patients and clinicians. In *Aaai spring symposium: argumentation for consumers of healthcare* (pp. 58–63).

Hart, S. G. & Staveland, L. E. (1988). Development of nasa-tlx (task load index): results of empirical and theoretical research. *Advances in psychology*, *52*, 139–183.

Henkind, S. J. & Harrison, M. C. (1988). An analysis of four uncertainty calculi. *Systems, Man and Cybernetics, IEEE Transactions on*, *18*(5), 700–714.

Hunter, A. & Williams, M. (2010). Argumentation for aggregating clinical evidence. In *Tools with artificial intelligence (ictai), 2010 22nd ieee international conference on* (Vol. 1, pp. 361–368). IEEE.

Jackson, P. (1986). *Introduction to expert systems*. Addison-Wesley Pub. Co., Reading, MA.

Bibliography II

Longo, L. (2012). Formalising human mental workload as non-monotonic concept for adaptive and personalised web-design. In *User modeling, adaptation, and personalization* (pp. 369–373). Springer.

Longo, L. & Dondio, P. (2014). Defeasible reasoning and argument-based systems in medical fields: an informal overview. *In proceeding of: The 27th International Symposium of Content-based Medical Systems.*

Longo, L. & Hederman, L. (2013). Argumentation theory for decision support in health-care: a comparison with machine learning. In *Brain and health informatics* (pp. 168–180). Springer.

Longo, L., Kane, B., & Hederman, L. (2012). Argumentation theory in health care. In *Computer-based medical systems (cbms), 2012 25th international symposium on* (pp. 1–6). IEEE.

Longo, L., Rusconi, F., Noce, L., & Barrett, S. (2012). The importance of human mental workload in web design. In *Webist* (pp. 403–409).

Meshkati, N. & Hancock, P. A. (2011). *Human mental workload*. Advances in Psychology. Elsevier Science.

Reid, G. B. & Nygren, T. E. (1988). The subjective workload assessment technique: a scaling procedure for measuring mental workload. *Advances in Psychology*, *52*, 185–218.

Schvaneveldt, R. W., Reid, G. B., Gomez, R. L., & Rice, S. (1997). *Modeling mental workload*. DTIC Document.

Toni, F. (2010). Argumentative agents. In *Computer science and information technology (imcsit), proceedings of the 2010 international multiconference on* (pp. 223–229). IEEE.

Tsang, P. S. & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, *39*(3), 358–381.

Young, M. S. & Stanton, N. A. (2001). Mental workload: theory, measurement, and application. *International encyclopedia of ergonomics and human factors*, *1*, 507–509.

NASA-TLX

Table: Questionnaire - Part 1

Dimension	Question
Mental demand	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	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	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?

NASA-TLX

Table: Questionnaire - Part 2

Dimension	**Question**
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Performance	How successful do you think you were in accomplishing the goals, of the task set by the experimenter (or yourself)?
Frustration	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?