



A Controlled Experiment on using Cognitive Work Analysis for System Engineering definition process

Daniel Prun, Camille Raymond

► To cite this version:

Daniel Prun, Camille Raymond. A Controlled Experiment on using Cognitive Work Analysis for System Engineering definition process. 2021 16th International Conference of System of Systems Engineering (SoSE), Jun 2021, Västerås, Sweden. 10.1109/SOSE52739.2021.9497498 . hal-03483593

HAL Id: hal-03483593

<https://enac.hal.science/hal-03483593>

Submitted on 16 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Controlled Experiment on using Cognitive Work Analysis for System Engineering definition process

Daniel Prun

ENAC, Université Fédérale de Toulouse

7 avenue Édouard-Belin 31055 Toulouse, Cedex 4

daniel.prun@enac.fr

Camille Raymond

Altran / ENAC

4, avenue Didier Daurat 31700 Blagnac

camille.raymond@gmail.com

Abstract— This paper describes an experiment about the relevance of using Cognitive Work Analysis (CWA) to improve the system definition process. CWA is a method for analyzing complex sociotechnical systems from a cognitive point of view. We submitted a set of exercises involving some specification tasks to a group of professional system engineers, some were asked to build on CWA data, while others to rely on some classical method results (Hierarchical Task Analysis). Beyond the results obtained from this specific experiment, this paper aims at demonstrating that controlled experiments are not only possible but also relevant in system engineering field.

Keywords — *Definition process, Controlled Experiment, Cognitive Work Analysis, Hierarchical Task Analysis*

I. INTRODUCTION

Several studies highlight the fact that human factors related elements are not adequately taken into consideration in complex systems definition and design (e.g. [1], [2], [3]). They list additional shortcomings that lead to human integration problems: human factors integration trends and standards are not captured; human performance metrics, targets, and limitations are not specified; human role design, job design, and organizational design are insufficiently captured, team activity and team requirements are insufficiently captured. Studies essentially identify a lack of shared methods, tools and formalisms precluding a meaningful communication and collaboration across system engineers, human factors practitioners and other personnel implied in System Engineering (SE). Furthermore [4] asserts that the integration of human factors within SE is not just a technical issue, but also cultural and organizational. The lack of accounting for cognitive factors during the upstream systems engineering processes contributes to incomplete system requirements, leading to design poorly usable systems. Without specific means, system design efforts will continue to be inconsistent, incomplete and redundant since these domains are not able to collaborate.

In this context, various solutions have been proposed. Hierarchical Task Analysis (HTA) [5] is the most popular of all human factors and ergonomics methods. It has been applied for more than 40 years in a lot of various domains and its popularity does not show signs of decrease within human factors and ergonomics communities. Based on the description of goals (“*what an operator is required to do or actually does, in terms of actions and/or cognitive processes*”), HTA belongs to normative or descriptive categories of work analysis modeling. The method aims at decomposing the overall main goal into sub-goals, operations and then plans. Plans specify the sequences and the conditions sub-goals have to be achieved by the operator in order to satisfy the upper-level goal. HTA found its place within a SE approach: the method is useful to identify operators, to describe operational scenarios involving humans and to

identify exchanged data between the system and the operators. Result of HTA is also very used in human error analysis techniques, functional allocation and human interface design. Moreover, the underlying formalism is very closed to major formalisms used by system engineer (e.g. System Modeling Language (SysML)) which contributes to its popularity. However, as stated by [6], the problem with HTA is that it remains centered on goals and so is highly dependent of situations which are related to actors and their finality. It addresses a list of enumerated situations which must be anticipated by the system engineer. If a specific situation has not been studied, then no related task and finally no related requirements are defined.

Cognitive Work Analysis (CWA) is an alternative human factors approach falling in the formative category ([7], [8]). It aims at identifying the purpose (“*the overarching intention that a work domain was designed to achieve*”) rather than goals. It focuses on the understanding of constraints forced upon the system, both physical and social. It is independent of situations (and so independent of actors and available physical resources) and can encompass a large variety of events, including non-predicted ones. The method aims at describing the system according to several viewpoints. The main used is Work Domain Analysis (WDA) which involves modelling the system based on its purposes and the constraints imposed by the environment [9]. Output of WDA is the Abstraction Hierarchy which contains 5 levels of description: Functional Purpose of the system is the reason why the system exists; Abstract Function is the criteria that can be used to judge whether the system is achieving its purposes; Generalized function is what functions are required to achieve the purpose of the work system; Physical function is the systems functional, capabilities and limitations; Physical object is the resources of the system. These elements are linked together through “means-ends” relationships. However, although some studies have shown how it can be useful for SE ([10], [11], [12]), mainly due to its capacity to help in identifying reasons why the system exists, mains functions, operational scenarios and physical components, CWA has rarely been used by system engineers and no studies have been carried out to assess CWA in a real situation ([13]).

The first objective of this paper is to demonstrate the relevance of using CWA for the system definition processes. For this purpose, we have designed and carried out a scientific experiment to study the relevance of this method, in a professional context. This experiment is the second objective of this paper: to show how a scientific controlled experimentation can be successfully conducted in the field of SE. After a brief state-of-the-art on controlled experiments in SE field, we present the scientific experiment we conducted (organization and result) and we conclude.

II. CWA ANALYSIS IN SE PERSPECTIVE

A. Controlled experiments in SE field

Controlled experiments are a scientific research method based on procedures carried out to support, refute, or validate hypothesis. A controlled experiment is one in which everything is held constant except for one variable. As explained by [14], controlled experiments provide two major advantages. In mature areas of research, they are often the only way to isolate the effects of independent variables on dependent variables and demonstrate causality. In nascent areas, they can enable carefully instrumented observations of phenomena under a wider range of conditions than exist in a natural setting.

Even if it is very current to find experiments at the heart of research in several domains (medicine, social science, physical sciences), they remain rarely conducted in the SE field. Researchers prefer to use alternatives as unstructured or structured observations, historical archival researches, participant observations, ethnographic studies, survey researches, or case studies. Some explanations are identified in [14]: difficulty to find representative participants; difficulty to define a representative case; difficulty to identify and control all stable elements, and the variable ones; difficulty to collect and exploit data to dress definitive conclusions.

B. Analysis of HTA and CWA

During the 4 last decades for HTA and the 2 last decades for WDA, several studies have been performed to identify and analyze advantages provided by one or the other method. These studies have mainly adopted a same and unique deductive approach: from observation of properties of a method, some conclusions are forecasted and justified. Often, these studies have been completed by study-cases analysis with the objective to show examples where deducted conclusions are true.

In [15], the authors describe an experiment they conducted with the objective to compare HTA and CWA. The subject was a military mission planning system. Participants were two human factors researchers, each with significant experience in the application of both HTA and CWA. Provided information was a set of user documents (training manuals, standard operating instructions...) and an initial 2-days meeting was conducted in order to introduce participants to the mission planning process and to familiarize them with dedicated software tool. During the exercise, participants interviewed 4 experts of the domain and performed walkthroughs. The outputs of the experiment were an HTA and a WDA description of the system. Their analysis showed differences between them in term of content, granularity and level of abstraction. Even if this experiment is valuable and provided pertinent insights for the assessment of HTA and WDA, we would like to point out some relative weaknesses that limit its scope. Firstly, only 2 participants have been involved in the experiment. This is not enough to reach a statistical signification and so to draw definitive conclusion. Moreover, outputs were HTA and WDA models (tasks decomposition, hierarchical abstraction) and not SE artifacts (requirements, use cases, scenarios...). This experiment did not show how outputs of these methods are useful (or not) for SE activities. Finally, the analysis only concerned the quality of outputs and does not addressed usability of each method. Feedback from participant have not been collected nor analyzed. No conclusion has been drawn on this point.

III. EXPERIMENT

We report on a controlled experiment we carried out about the relevance of HTA and CWA in a system requirement definition process. We have followed the process described in [16] in which the authors propose a protocol to define and conduct a qualitative experiment-based research in the field of system engineering. The protocol contains 7 steps: choosing to employ qualitative experimental method, defining research questions, selecting case, conducting experiment, analyze data, compile the results of the experiment and infer and produce a theory. We describe the first 6 steps and address the last step in the conclusion. A full description is can be found in [17].

A. Description of the experiment

1) Step 1: Choosing to employ qualitative method

Qualitative methods are appropriated when the studied phenomenon is poorly understood and when it must be studied in an empirical context [16]. It is the case here.

2) Step 2: Defining research questions

We have defined several research questions through hypothesis (see TABLE I.). They are assertions that aim to be validated or not according to the results of the experiment. Q1 and Q2 addressed impact of the method relatively to temporal performances and quality of the productions while the others were centered on subjective feeling of participants.

TABLE I. RESEARCH QUESTIONS

Q1	Temporal performance: WDA is a new method for system engineers who had just a short training before the experiment. We expect WDA participants to take more time than HTA ones to complete their tasks.
Q2	Quality of productions: because WDA brings more elements about human factors, we expect WDA participants productions to be better than HTA participants ones.
Q3	Cognitive effort: as WDA method is totally new and unknown to system engineers, and HTA is very close to their best used methods, we expect WDA participants cognitive effort to be higher than HTA.
Q4	Acceptation of the method: WDA is not as well described and formalized as classical SE methods, and HTA is very closed to best used SE methods, we expect WDA method to be less accepted (and less easy to use) than HTA one
Q5	Success/fail feeling: We expect HTA participants to have a better success feeling than WDA participants.
Q6	Experience impact: System engineers with a high experience and practice level are familiar to use their methods. We expect expert participants have more difficulties to accept WDA than junior ones.

3) Step 3: Selecting case

a) Participants

10 system engineers participated to the experiment (see TABLE II.). They worked as consultants in various domains (aeronautic, automotive, defense, healthcare...). Some of them, identified as junior, had less than 5 years of experience in SE while others (experts) had more. We checked through preliminary questionnaires that nobody had previously followed specific training in ergonomics. We verified also that no one had a knowledge about the finality of the study. Participants were divided in 2 groups of 5 (WDA and HTA) without any selective criterions.

TABLE II. GROUPS OF PARTICIPANTS

WDA group				HTA group			
<i>Id</i>	<i>Age</i>	<i>Sex</i>	<i>Expertise</i>	<i>Id</i>	<i>Age</i>	<i>Sex</i>	<i>Expertise</i>
WDA_P1	29	M	Expert (5)	HTA_P1	33	M	Expert (12)

WDA_P2	27	M	Junior (1.5)	HTA_P2	26	F	Junior (0.7)
WDA_P3	25	F	Expert (5)	HTA_P3	45	M	Expert (22)
WDA_P4	43	M	Expert (18)	HTA_P4	24	M	Junior (0.2)
WDA_P5	28	F	Junior (3)	HTA_P5	52	M	Expert (11)

b) Tasks

Each participant had to model a system which fulfils several on-ground services to travelers using air transport: flight reservation, boarding, security checking, information etc. Participants had access to a description of the initial high-level needs (see Fig. 1). Moreover, according to their group, participant had access to a task analysis (for HTA group participants – see Fig. 2) or an abstraction hierarchy (for WDA group participants – see Fig. 3) of the system previously built by ergonomists. On this basis, participants had to perform 4 exercises in less than 40 minutes: #1 write the set of requirements, #2 describe the set of use-cases, #3 write nominal scenarios and #4 write non-nominal scenarios. In order to avoid bias introduced by specific tools, models had to be designed using pen and paper.

To travel by plane, a traveler must:

- buy ticket through an online ticket provider, or using some automatic machines located to the airport or directly from agency desks;
- check-in thanks to an online access, or using automatic machine or directly from agency desk;
- possibly drop-off baggage directly to a dedicated desk or using an automatic drop-off system. Some airlines offer a service to take over baggage directly at the traveler residency.
- pass through airport security checks;
- attend the boarding gate to access to the plane.

All these tasks must respect rules related to security (prohibited articles, identification, etc.)

The airport is equipped with information boards and information desks for helping passengers to find their way in the airport.

....

Fig. 1. Extract of initial high-level user needs as provided to all participants

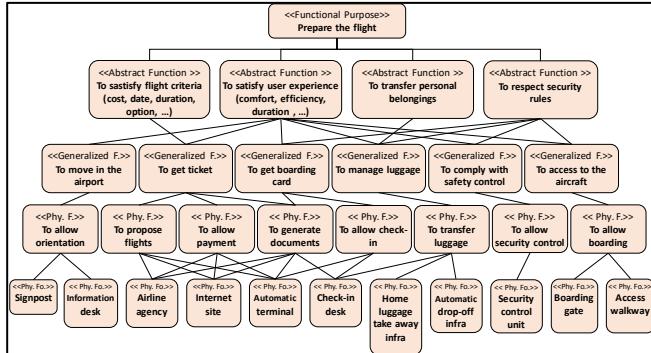


Fig. 2. Abstraction Hierarchy as provided to CWA group

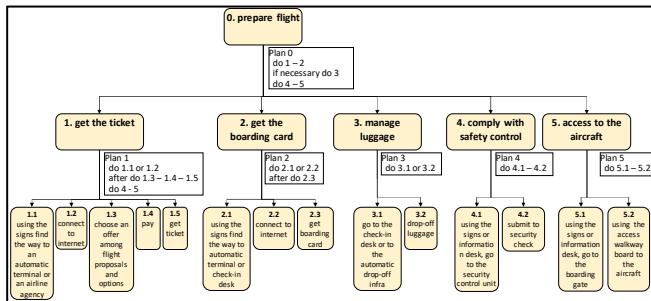


Fig. 3. Task decomposition as provided to HTA group

4) Step 4: Experimental protocol

The experimentation consisted in 4 steps.

Step 4.1: Collect information about participants profile (age, sex, experience etc.);

Step 4.2: Train participants to the WDA based method (for WDA group) and to the HTA base method (for HTA group). At the end of this step, a multiple-choice question has been submitted to participants to check their understanding of the methods. Participants had to answer to 9 questions (for HTA trainees) and 13 questions (for WDA trainees). To be able to compare results, marks have been scaled up to 20;

Step 4.3: Perform the 4 required exercises. During this step, completion times have been measured;

Step 4.4: Collect participants feedback. 5 questionnaires have been proposed to collect participant feedback: pass/fail, modeling, NASA-TLX, TAM and SUS. **Pass/fail** is a set of questions we have specifically designed to collect the feedback of participants relative to their feeling about the success (or unsuccess) in performing the exercises. **Modeling** is a set of questions we have specifically designed to encourage participants to describe how they used input materials to produce the models. Collected information is useful to better understand processes involved in the exploitation of WDA and HTA based methods. **NASA-TLX** (Task Load Index) [18] is a widely used subjective method to evaluate workload perceived to perform a task. It contains 6 questions (see TABLE III. assessed using 20 points rating scale (from easy to difficult). Answers are combined to derive a sensitive and reliable estimate of global workload. **TAM** (Technology Acceptance Model) [19] is a method for assessing user acceptance (utility and ease-of-use) of a computer-based system. It proposes 16 questions (see TABLE IV. to the performer, with 7 points rating scale from “strongly agree” (level 1) to “strongly disagree” (level 7). **SUS** (System Usability Scale) [20] is used to evaluate how people perceive the usability of a system or a method. It is dedicated to measure performer efficiency and satisfaction. It contains 10 questions (see TABLE V. with a 5 levels rating scale. The result is computed from the collected answers is a mark between 0 and 100. Usability is “poor” below 35, “ok” and “good” between 35 and 86 and “excellent” above.

TABLE III. NASA-TLX QUESTIONS

Q1	Mental demand: How much mental and perceptual activity was required?
Q2	Physical demand: How much physical activity was required?
Q3	Temporal demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred?
Q4	Overall performance: How successful were you in performing the task?
Q5	Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?
Q6	Frustration level: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

TABLE IV. TAM QUESTIONS

TAM questions related to utility	
Q1	(HTA or WDA) would improve quality of my design task
Q2	(HTA or WDA) would improve quality of my productions
Q3	(HTA or WDA) would improve my productivity
Q4	(HTA or WDA) would address main aspects of my task
Q5	(HTA or WDA) would address unusual aspects of my task
Q6	(HTA or WDA) would improve my efficiency in doing my task
Q7	(HTA or WDA) would make easier my task
Q8	(HTA or WDA) is useful in doing my task
TAM questions related to easy-of-use	

Q9	(HTA or WDA) is heavy to use
Q10	(HTA or WDA) is easy to learn
Q11	(HTA or WDA) is heavy to use in a SE process
Q12	(HTA or WDA) is clear and understandable
Q13	(HTA or WDA) is agile
Q14	(HTA or WDA) is easy to remember in order to use it
Q15	(HTA or WDA) requires an important mental workload
Q16	(HTA or WDA) requires an important workload to be able to use it adequately.

TABLE V. SUS QUESTIONS

Q1	I think that I would like to use (HTA or WDA) frequently.
Q2	I found (HTA or WDA) unnecessarily complex.
Q3	I think (HTA or WDA) is easy to use.
Q4	I think that I would need the support of a technical person to be able to use (HTA or WDA).
Q5	I found the various functions in (HTA or WDA) were well understood.
Q6	I thought there was too much inconsistency in (HTA or WDA).
Q7	I would imagine that most people would learn to use (HTA or WDA) very quickly.
Q8	I found (HTA or WDA) very cumbersome to use.
Q9	I felt very confident using (HTA or WDA).
Q10	I needed to learn a lot of things before I could get going with (HTA or WDA).

The relationship between collected data and research question is highlighted in TABLE VI.

TABLE VI. LINK RESEARCH QUESTIONS - COLLECTED DATA

Research questions	Collected data						
	Pass/fail	Modeling	Nasa-TLX	TAM	SUS	Completion time	Models quality
Q1						x	
Q2							x
Q3		x	x				
Q4		x		x	x		
Q5	x						
Q6				x			

During questionnaires, participants were asked to explain and justify their responses "in their own words". Data collected during this **verbalization** process have been useful during the data analysis step, in order to identify qualitative data and supporting the findings.

5) Step 5: Data analysis

Quantitative collected data have been evaluated according to the mean value (noted u_{hta} for HTA group and u_{wda} for WDA group) and standard deviation (noted σ_{hta} for HTA group and σ_{wda} for WDA group). It was then possible to have a rough assessment of the global trend and representativity of the data.

Qualitative data (responses to the exercises) have been analyzed using peer reviews, carried out by SE experts. As we provided participants with little detailed input materials, several acceptable solutions were possible and only an assessment performed by expert was relevant. Some quantitative indicators have been produced from these reviews: **requirements** (production of exercise #1) have been checked according to 3 suitability criteria: correct, inferred (requirements not related to the initial need nor to HTA and WDA models) and over-specified (requirement related to an initial need or an HTA or WDA element but adding extra properties); **use-cases** (production of exercise #2) have been

evaluated according to the identified actors, the relevance of the use case itself (5 major uses cases have previously been identified) and its suitability (correct, inferred or over-specified); **scenarios** (production of exercise #3 and #4) have been analyzed according to their relevance.

To compare results between WDA and HDA groups, we selected the Mann-Whitney statistic test [21]. This test is used to verify whether 2 distribution laws are similar. It is useful on a low number of participants and when the distribution of variables does not necessarily follow the normal law, which is the case here. Equation (1) defines this test for HTA group. According to Mann-Whitney pre-computed table, considering 5 participants per group and a level of significance $a=0.05$, distributions are correlated if $\min(U_{hta}, U_{wda}) > 2$.

$$U_{hta} = n_{hta}n_{wda} + \frac{n_{hta}(n_{hta}+1)}{2} - R_{hta} \quad (1)$$

where n_{hta} is the size of HTA group, u_{wda} the size of WDA group and R_{hta} the sum of ranks for HTA group.

B. Results of the experiment

1) Assessment of participants after training sessions

Mean of marks obtained by participants were $u_{hta} = 15.55$ and $u_{wda} = 18.40$. These results are very good, and the standard variation is low: $\sigma_{hta} = 2.1$ and $\sigma_{wda} = 1.34$. We conclude that participants understood very well the basis of both methods and they were ready to perform the exercises.

2) Completion time of exercises

Measures performed during step 3 (see Fig. 4) showed that the time required by HTA group to complete exercises 1, 2 and 4 is smaller than WDA group (an average of 7'30" vs 10"). Conversely exercise 3 took less time for WDA group (9'35" vs 12'48"). However, standard variation is high reflecting a great variability between participants. So, a definitive conclusion is hazardous.

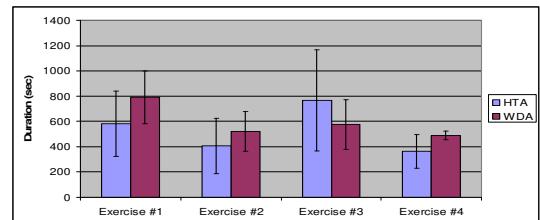


Fig. 4. Mean and standard deviation of completion time per exercise

3) Pass/fail analysis

Collected feedback showed that feeling about the success was very high for all participants.

4) Analyze of the productions

Requirements: HTA group has proposed slightly less correct requirements than WDA group ($u_{hta} = 7.8, \sigma_{hta} = 4.15, u_{wda} = 10, \sigma_{wda} = 2.65$). HTA group over specified ($u_{hta} = 2.4, \sigma_{hta} = 1.67, u_{wda} = 0.8, \sigma_{wda} = 0.83$) and very few requirements are inferred.

Use cases: a total of 24 correct use cases have been identified by participants. Analysis showed that HDA group modelled 2 times less correct use cases than WDA. It is also important to note that WDA group have a very low variability contrary to HDA ($u_{hta} = 6.2, \sigma_{hta} = 5.59, u_{wda} = 12, \sigma_{wda} = 1.58$). Moreover, all WDA group participants found all 5 main use cases contrary to HTA group participants who found only 60%.

Scenarios: because of the lack of formalism imposed to express scenarios, quantitative analyze was hard to run and we mainly analyzed qualitative aspects collected through Modeling questionnaire and verbalization. HTA group participants expressed that task model provided as input was very close to models they traditionally used. They proceeded by reading the task model and produced a new scenario each time a choice appeared to the actor. For WDA group, they analyzed the abstraction hierarchy and mainly started from generalized and physical functions. Concerning non nominal scenarios, none HTA group participant used the task model (mainly because this model does not contain information related to fault). On the contrary, some WDA group participant directly relies on the analysis of the abstraction hierarchy: they studied physical function description layer by analyzing possible failure of each physical element, deducing non nominal scenarios from this analysis. Globally, analysis of the productions shows a global positive trend for WDA group: they found more human factors related requirements and they had a relevant use of the abstraction hierarchy to find non-nominal scenarios.

5) Workload analysis

Results from NASA-TLX test (see Fig. 5) showed that the three last factors covering interactions between performer and task (performance, effort and frustration) were rather neutral (around 45) and very closed between HTA and WDA groups. Verbalization showed that the exercises were classical for participants, but time pressure and novelty of the method (for WDA group) tended to raise the effort. Frustration is low but largely scattered ($u_{hta} = 38, \sigma_{hta} = 23.61$, $u_{wda} = 45, \sigma_{wda} = 23.18$). HTA_P2 and HTA_P3 had a high feeling of frustration (60 and 65 respectively) due to the temporal stress (HTA_P2) and to the frustration to not complete 2 exercises (HTA_P3). WDA_P3 expressed a great frustration because of the novelty of the method.

Among the first three factors related to constraints from the task, Mental demand stayed moderate but was higher for WDA. WDA_P1 expressed a high mental workload because of the novel nature of the method. Physical demand was low because it was only asked to use paper and pencil for the exercises. Hight Temporal demand showed that participants were under pressure during the experiment. However, this pressure did not prevent participants to complete their exercises within the allocated time (exception with HTA_P3).

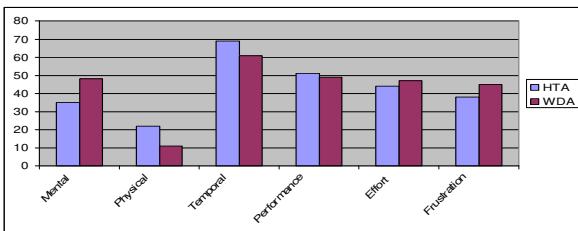


Fig. 5. NASA-TLX results

Statistical Mann-Whitney test showed that for each factor of NASA-TLX, the distributions of collected data are identical. We conclude that there are no significative differences between both methods.

6) Usability and easiness analysis

Results from TAM test (see Fig. 6) show there are few differences between HTA and WDA groups. On average, HTA is slightly better. The first 8 questions related to utility showed a high level of acceptation for both methods. In the

detail, HTA got a better assessment regarding suitability for the task (question 4), usability (question 9) and clarity and understandability (question 12). WDA had been assessed better regarding workload required to implement it (question 16). The main arguments put forward relate to the fact that WDA was new to participants while HTA was closed to data they frequently used; WDA abstraction hierarchy can turn to be a “plate of spaghetti” and become unusable; HTA requires a specific workload to understand the hierarchical task description which is less true for the abstraction hierarchy.

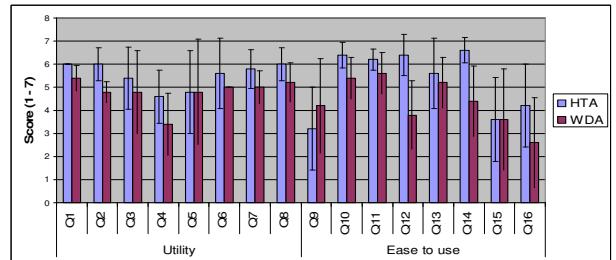


Fig. 6. Mean and standard deviation of TAM

The analysis of TAM results showed that both methods have similar impacts. Mann-Whitney test showed that the distributions of data are identical.

To bring an answer to research question Q6 (Experience level impact), we correlated answers collected to TAM questionnaires with the experience level of participants. Results showed that junior participants had a better perception of utility and use of HTA than experts. Concerning WDA, feelings are similar (see Fig. 7).

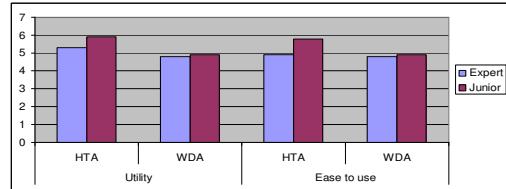


Fig. 7. TAM result according to experience

Results from SUS test shows 2 main trends (see Fig. 8): HTA got an excellent global score and WDA got very variables feedbacks. This is explained by the facility participants use HTA (this method is very close to their usual practices) and by the novelty of the WDA method. Verbalizations showed difficulties participants have to use a new method. Some of them are suspicious and wish to learn more before adopting a definitive position. On the contrary, others welcome the method, already giving it some merits.

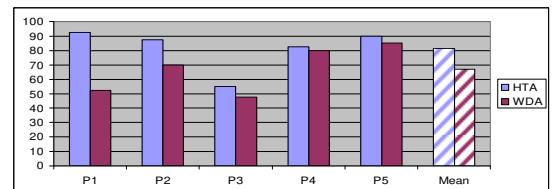


Fig. 8. SUS score for each participant and global mean values

Mann-Whitney test showed that the distributions of SUS data are identical.

C. Summary

Results allows us to draw up answers to research questions defined at step 2. They are summarized in TABLE VII.

TABLE VII. ANSWERS TO RESEARCH QUESTIONS

Q1	Temporal performance Completion time showed no significative differences between HTA and WDA groups. Mann-Whitney statistic test showed distributions are identical. Therefore, it is not possible to give a definitive answer.
Q2	Quality of productions Results are slightly better for WDA for number and quality of human factors requirements, use cases and non-nominal scenarios. Therefore, the hypothesis is validated
Q3	Cognitive effort Mental demand assessed with NASA-TLX test has been shown to be higher for WDA group. Therefore, the hypothesis is validated.
Q4	Acceptation of the method SUS test and the first part of TAM test showed that WDA is harder to use than HTA. Therefore, the hypothesis is validated.
Q5	Success/fail feeling Results showed success feeling is very high for both methods. Therefore, the hypothesis is rejected.
Q6	Experience impact TAM test showed no difference between experienced and novice system engineers regarding acceptation of WDA. Therefore, the hypothesis is rejected.

These results confirm that WDA is relevant in a SE definition process because it allows to find more pertinent data. It requires a higher cognitive effort and is less accepted by participants. This is due to the novelty of the method and the lack of training and practice. Even if some participants remain skeptical about this method, a majority is enthusiastic.

IV. CONCLUSION & FUTURE WORK

In this paper, we related a controlled experiment in SE field. We selected a representative panel of participants and we set up an experimental framework according to recommended state-of-the-art experimental protocol. We collected qualitative and quantitative data about the quality of the outputs and also about the usability of the methods. The use of statistical tests allowed us to draw conclusions with a defined level of confidence.

We would like to advocate for a wider use of such experimental approaches in SE. Indeed, the characteristics of this domain (where practices are empirical, not always formalized, and involve humans' activities) justify using experiments to get valuable results about validation of assumptions made by researchers. We showed how this approach can be deployed and controlled and we hope this it will be largely borrowed by future researches.

On the content, results demonstrated how valuable WDA is for system definition activities. Due to its formative approach, WDA allows system engineer to identify more human factors related requirements, and to have a better covering of non-nominal situations. However, WDA is more distant of system engineers practices than HTA is. Therefore, it requires higher workload from them, and usability gain need to be enhanced.

Finally, we would argue for the complementarity of the methods. CWA models the intentional, functional and topological structure of the work domain while task analysis allows navigation within this structure to find tasks to be performed to reach objectives. An obstacle to overcome remains the lack of a formalized method: contrary to HTA

which has a well-versed, step-by-step methodology CWA is more a framework than a methodology and does not restrict system engineer to specific methodologies for each phase.

ACKNOWLEDGMENT

We would like to thank C. Bouvier, E. Fournac, J. Françoise, L. Giraldo and P. Lopengo, students of master "Risques et Environnement" of INU Champollion who contributed to the experiment presented in this paper.

REFERENCES

- [1] Mavor A. S., Pew R. W. Human-System Integration in the System Development Process: A New Look. National Academies Press, 2007.
- [2] Hardman N. S. An Empirical Methodology for Engineering Human Systems Integration. 2009
- [3] Madni A. M. Integrating humans with and within complex systems. CrossTalk, 5, 2011.
- [4] Orellana D. W., Madni A. M. Human System Integration Ontology: Enhancing Model Based System Engineering to Evaluate Human-System Performance. Procedia Computer Science, 28, 19-25. 2014.
- [5] Annett J. Hierarchical task analysis. In: D. Diaper and N.A. Stanton, eds. The handbook of task analysis for human-computer interaction. Mahwah, NJ: Lawrence Erlbaum Associates, 67–82. 2004.
- [6] Naikar N., Hopcroft R., Moylan A. Work domain analysis: Theoretical concepts and methodology. Technical report, DTIC Document, 2005.
- [7] Rasmussen J. Information processing and human-machine interaction. An approach to cognitive engineering, 1986
- [8] Vicente K. J. Cognitive work analysis: Toward safe, productive and healthy computer-based work. Lawrence Erlbaum Associates. 1999.
- [9] Read G. J. M., Salmon P. M., Lenné M. G. Cognitive work analysis and design: current practice and future practitioner requirements. Theoretical Issues in Ergonomics Science, 16(2):154–173, 2015
- [10] Sanderson P., Naikar N., Lintern G., Goss S. Use of cognitive work analysis across the system life cycle: From requirements to decommissioning. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, volume 43, pages 318–322. SAGE Publications Sage CA: Los Angeles, CA, 1999.
- [11] Stanton N. A., Jenkins D. P. Application of cognitive work analysis to system analysis and design. In Cognitive Work Analysis: Applications, Extensions and Future Directions, pages 3–71. CRC Press, 2017.
- [12] Torenvliet G., Hilliard A., Burns C. M., Lintern G., Lamarre J-Y. Modelling and simulation for requirements engineering and options analysis. Technical report, ESTERLINE CMC ELECTRONICS OTTAWA (ONTARIO), 2010.
- [13] Lintern G. Work domain analysis: Tutorial. General Dynamics, Dayton, Ohio, 2013.
- [14] Panchal J. H., Szajnfarber Z. Experiments in systems engineering and design research. System Engineering, 2017; 20:529–541.
- [15] Salmon P. M., Jenkins D. P., Stanton N. A., Walker G. H. Hierarchical task analysis vs. cognitive work analysis: comparison of theory, methodology and contribution to system design. Taylor & Francis Ltd.; Theoretical Issues in Ergonomics Science, 11. 504-531. 2010.
- [16] Szajnfarber Z., Gralla E. Qualitative methods for engineering systems: Why we need them and how to use them. System Engineering. 2017; 20:497–511.
- [17] Raymond C. Extension des modèles d'ingénierie système à l'analyse du domaine de travail. PhD thesis, Université Paul Sabatier, 2018.
- [18] Hart S. G., Staveland L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. Advances in psychology, 52:139–183, 1988.
- [19] Davis F. D. A technology acceptance model for empirically testing new end-user information systems: Theory and results. PhD thesis, Massachusetts Institute of Technology, 1985.
- [20] Brooke J. SUS: a "quick and dirty" usability scale. In P. W. Jordan; B. Thomas; B. A. Weerdmeester; A. L. McClelland (eds.). Usability Evaluation in Industry. London: Taylor and Francis. 1986.
- [21] Mann H. B., Whitney D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Annals of Mathematical Statistics. 1946.