The Expanded Cognitive Task Load Index (NASA-TLX) applied to Team Decision-Making in Emergency Preparedness Simulation

Denis A. Coelho[1], João N. O. Filipe[1], Mário Simões-Marques[2], Isabel L. Nunes[3,4]
[1]Universidade da Beira Interior, [2]Portuguese Navy,
[3]Universidade Nova de Lisboa, [4]UNIDEMI
Portugal

Abstract

The study demonstrates the use of the expanded TLX instrument (Helton, Funke & Knott, 2014) for cognitive and team-related workload self-assessment of 38 participants, solving the UNISDR – ONU stop disasters game simulation. Subjects in one group (GF; n=30) performed group decision-making without prior individual practice on the simulation. A subset of GF participants (n=6) subsequently reiterated the simulation alone, reassessing their cognitive workload. Another group (IF; n=8) individually performed the simulation and reiterated it in groups. Most GF participants, moving from group to singly conditions, reported decreasing physical and temporal demands, unchanged self-assessed performance, and increased mental demands, effort and frustration. IF participants incurred increasing mental, physical and temporal demands, as well as increased effort, with decreasing frustration and better performance, from singly to group conditions. Team workload results differed across groups; GF had higher levels of reported team dissatisfaction, equivalent assessments of team support and lower assessments of coordination and communication demands coupled with decreased time sharing as well as lower team effectiveness, compared to IF. Results bear implications on training of decision-making teams; singly training team members preceding group training supports team-decision making effectiveness and individual performance within teams going through first stages of a system learning curve.

Introduction

This section presents the interest in studying training for team-decision making and the scope of emergency preparedness. To this follows the presentation of the study aims, a methods section describing participants, the simulation and the experimental procedure, the results and their statistical analysis and, finally, a concluding discussion.

In D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, and H. Hoonhout (Eds.) (2015). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Training for team-decision making

Growing attention has been paid to the need to develop problem-specific models of problem solving, as opposed to traditional phase models articulating single approaches to solving all kinds of problems (Silber & Foshay, 2009). Work has become complex enough to require the use of teams at all hierarchical levels, with organizational success depending to a large extent on the ability of teams to collaborate and work effectively in solving complex problems (DeChurch & Mesmer-Magnus, 2010). Problem solving is also a learning process (Cooke et al. 2000) and team training benefits from a curriculum designed by a task analysis (Hamman, 2004). In the process of researching and understanding new information, the newly acquired understanding is added into the team's knowledge base, accumulating its experience from solving similar types of problems (Hung, 2013). According to DeChurch and Mesmer-Magnus (2010) relatively little is known about how team cognition forms and how to support it, dispite this being a critical issue for those designing teams and using teams in applied settings. The present study contributes to unveiling how to support the individual's performance within a decision-making team as well as team effectiveness.

This study investigates the effect of individual practice taking place prior to an otherwise unprepared group problem solving session (consisting of an emergency preparedness simulation) on individual and team-related workload. Studies focusing on workload measurement as a state should take a within-subjects perspective in their analysis (Helyton, Funke &Knott, 2014), although studies focusing on training evaluation often do not concurrently develop a within-subjects and a between-subjects perspective (Hagemann & Kluge, 2013). In this contribution, both within-subjects and between subjects perspectives are considered.

In this study, it is expected that the effect of training improves individual performance by the time of a second simulation run, irrespective of having done a first simulation run within a group or singly, or having done a second simulation run singly or within a group. This notwithstanding, it is expected at the onset of the study that first handedly and individually acquiring knowledge related to the problem at hand, prior to engaging in team-decision making within the process of solving the problem, will lead to improved team effectiveness. Individual practice following group interaction is used in the experiment as a means of balancing two group conditions, and enabling more extensive between subjects-analyses even if the primary interest of the study is supporting effective team- decision making.

Emergency preparedness and the nature of decision-making therein

Emergencies are unpredictable; needs for resources and information are difficult to define beforehand (Coelho, 2013-b). Emergency management is a mission that in several phases: work to avoid crises, preparation for crises, operative work, and evaluations after an event (Fig. 1).

Emergency management is a complex process requiring coordination of different actors, with different cultures, goals and views of the world. It aims to provide efficient and effective responses to multiple and often conflicting needs in situations of scarce resources, considering several complementary functional elements, such as supply, maintenance, personnel, health, transport and construction. In all these elements the decision-making issues relate to basic questions: what, where, when, who, why, how, how much? These questions become particularly difficult to answer in critical situations, such as disaster relief, especially sensitive to the urgency and impact of decisions (Simões-Marques & Nunes, 2013). The commonly accepted phases of the management of the response to emergent events and critical disasters can be further characterized as follows: mitigation - preventing future emergencies or minimizing their effects, preparedness - preparing to handle an emergency, response - responding safely to an emergency, and, recovery - recovering from an emergency. The preparedness phase allows the development of an adequate level of resilience which enables effective emergency response and faster recovery, namely through a continuous cycle of planning and training (Fig. 2), as well as through public information, education and communication.

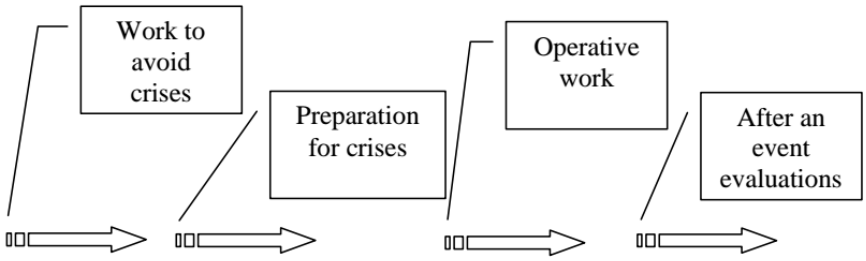

Figure 1. Phases in the management of the response to emerging events and critical disasters (Coelho, 2013-c).

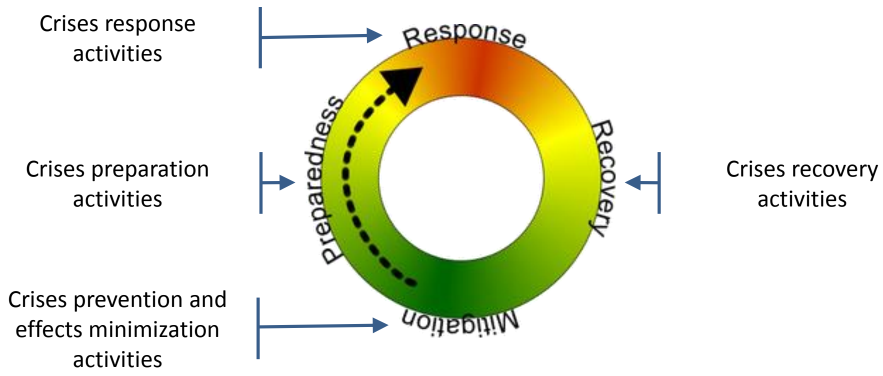

Figure 2. The continuous cycle of mitigation, preparedness, response and recovery.

According to Helton, Funke and Knott (2014) there is a growing interest in developing collaborative ways of teaching students about natural disasters (Berson & Berson, 2008; Gaillard & Pangilinan, 2010) as well as using simulation games to understand human behaviour in regard to disasters (Brigantic et al., 2009). The simulation that is used in the experimental study deals with natural disaster preparedness, as a means of taking actions and altering the built environment as a way of mitigating the severity of the consequences of the disaster when it strikes, even if in reality it is uncertain when in the future it will occur.

Aims

Overall, this study is oriented towards empirically inducting knowledge contributing to support effectiveness of team decision-making and the individual's performance therein. The main aim of the experiment is to analyse the effect of individual problem-specific training on individual and team-related workload and performance/effectiveness in the course of a group decision-making activity.

Aditionally, an assumption was established in the design phase of the study. It was that practice leads to improved individual performance, irrespective of the order in which its two experimental conditions (group and solo) are experienced by the participant.

Method

Participants

Thirty-eight engineering students (13 women, 25 men), divided into two groups participated in the study for course credit. Their age ranged from 20 to 25 years. All study participants had normal or corrected-to-normal vision and hearing and none had any upper-body impairments limiting the use of a keyboard coupled with a computer pointing device (mouse) as interface. Participants were assigned to two groups. Table 1 presents participants count and sex by group, as well as subgroup size and gender mix.

*Table 1. Case counts for subgroup size and sex mix (legend: M - male sex; F - female sex; one of the subgroups in each category marked with * had 2 participants subsequently performing the simulation alone, for a total of 6 participants – 4 men and 2 women).*

Group	*Subgroup size*	*Quantity*	*Subgroup composition*		
			All male	*All female*	*Mixed*
GF – Group Simulation First (n=30; 8F; 22M)	2	2	1	—	1*
	3	3	1*	—	2*
	4	3	1	—	2
	5	1	—	—	1
IF – Individual Simul. First (n=8; 5F; 3M)	2	4	1	1	2

Simulation

The Stop Disasters game (www.stopdisastersgame.org) was developed by Playerthree© for the United Nations International Strategy for Disaster Reduction (UN/ISDR). In the Stop Disasters game (Fig. 3), players attempt to build disaster-resilient communities while also achieving development goals (e.g., building infrastructure). In this study, we focused on an earthquake simulation, as it represents a regional interest for participants in Portugal. Because of course administration constraints, the time available for reiterations of the simulations was very limited (allowing only one to two per participant), which led to choosing the easiest setting. While most participants chose English, they were given the possibility of opting for the interface language that they felt most confident with of those available in the simulation game (English, Spanish or French). This game had previously been used for research (e.g. Khalid & Helander, 2013), but no team task analysis was available. The game yields a simulation performance score at the end of the simulation, which was not retained by the researchers.

Figure 3. Screenshot taken from the Stop Disasters Earthquake simulation game.

Expanded NASA TLX instrument for cognitive and team workload

NASA-TLX was established after an extensive three-year research effort and it sits properly in a web of correlations with external variables (Hart & Staveland, 1988). Workload has now become almost synonymous with the TLX (De Winter, 2014). Helton, Funke and Knott (2014) presented a modified version of the NASA-TLX that includes six additional team workload measures (Table 2). The additional team workload items were developed on the basis of literature review on teams carried out by Funke et al. (2012). The expanded version was used in this study in the decision-making in teams condition, while the standard version was used for the singly condition.

Procedure

The expanded version of the NASA-TLX instrument (Helton, Funke & Knott, 2014) was used to assess cognitive and team-related workload of a total of 38 students, divided into two groups (Fig. 4). Participants joined in teams of 2 to 5 people, solved the UNISDR – ONU stop disasters game simulation (earthquake challenge - easy mode) in a classroom setting. After the group simulation, each individual assessed his or her workload as well as the team-related workload using the expanded NASA-TLX. Subjects had no previous contact with the simulation and completed it within the allotted 25 minutes. A subset of 2 female and 4 male participants, who had made part of one of the two-person groups and of two of the three-person groups subsequently reiterated the simulation on their own, reassessing their cognitive task load, using the standard NASA-TLX.

Table 2.Rating scale definitions of the expanded (*) Task Load Index (TLX) (NASA, 1986, 2014; Helton, Funke & Knott, 2014) (these items were measured on 0-to-20 scales and multiplied by 5 to create comparable 0-to-100 scales).

Title	*Descriptions*
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?
Performance	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?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
*Coordination Demand	How much coordination activity was required (e.g., correction, adjustment)? Were the coordination demands to work as a team low or high, infrequent or frequent?
*Communica-tion Demand	How much communication activity was required (e.g. discussing, negotiating, sending and receiving messages)? Were the communication demands low or high, infrequent or frequent, simple or complex?
*Time Sha-ring Demand	How difficult was it to share and manage time between taskwork (work done as a team)? Was it easy or hard to manage individual tasks and those tasks requiring work with other team members?
*Team Effectiveness	How successful do you think the team was in working as a team? How satisfied were you with the team-related aspects of performance?
*Team Support	How difficult was it to provide and receive support (providing guidance, helping team members, providing instructions, etc.) from team members? Was it easy or hard to support/guide and receive support/guidance from other team members?
*Team Dis-satisfaction	How emotionally draining and irritating versus emotionally rewarding and satisfying was it to work as a team?

An unrelated group of 5 female and 3 male subjects individually performed the simulation (assessing their individual workload afterwards), and later, reiterated it in groups of 2 (assessing both their individual and team-related workload after the group simulation with the use of the expanded NASA-TLX). All assessments were made in the original language of the instrument. Statistical analysis was made with the assistance of IBM™ SPSS© 20 and using the approach described by Coelho et al. (2013-a).

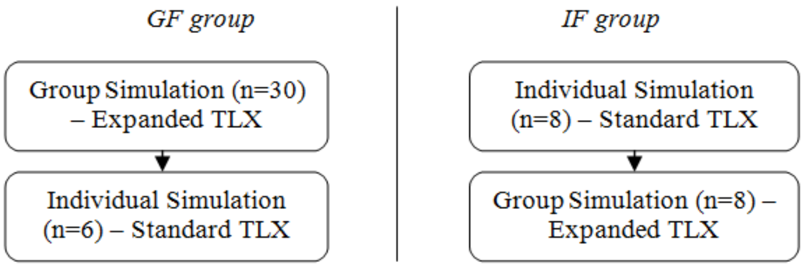

Figure 4. Diagram of experimental procedure.

Results and analysis

This section begins with the descriptive presentation of the results followed by their analysis (between subjects, within subjects and association of scales).

Presentation of results

Aggregated overall results are shown in Table 3, considering the condition that was rated and the order of the conditions in each group. The results overview suggests that within GF, effort and all types of demands increased for the participants involved in the two conditions, while performance and frustration remained almost unchanged. Conversely, for IF, performance increased and frustration decreased, while effort and all demands (mental, physical and temporal) increased. Looking across the team-related scales suggests higher coordination, communication and time sharing demands in the 2^{nd} group, with much higher team effectiveness and equivalent team support. Selecting all participants in GF for comparison with IF, would suggest lower team dissatisfaction in IF, but the opposite ensues when selecting only the six participants in GF who reiterated the simulation alone.

Between subjects workload comparison (across both groups - group condition)

The independent samples Mann-Whitney test only yielded significant differences (significance threshold lowered to 0.001 to account for multiple comparisons – 12) across both complete groups in the group condition for communication demands (U=10; p<0.001) and for time sharing demands (U=14; p<0.001), both higher on average for IF. This would suggest that having more knowledge of the problem domain would require more communication and time sharing within the problem-solving setting in groups, even if groups are significantly smaller (p<0.001).

*Table 3. Mean and standard deviations obtained for each rating scale and group condition (legend: * - expansion team work related TLX rating scales; ** - subgroup of participants from GF who were subjected to the two experimental conditions).*

Rating scale	*GF (n=30)* Group 1st n=30	*GF (n=30)* Group 1st **n=6	*GF (n=30)* Solo 2nd **n=6	*IF n=8* Group 2nd	*IF n=8* Solo 1st
Mental Demand	56 (19)	44 (27)	64 (12)	66 (18)	56 (23)
Physical Demand	36 (19)	25 (21)	33 (21)	48(26)	33 (21)
Temporal Demand	50 (16)	34 (17)	41 (11)	61 (26)	43 (18)
Performance	50 (22)	48 (28)	50 (28)	60 (30)	44 (31)
Effort	54 (20)	51 (29)	62 (23)	58 (29)	53 (18)
Frustration Level	52 (25)	40 (26)	42 (30)	45 (29)	64 (29)
*Coordination Demand	61 (19)	60 (25)	—	71 (16)	—
*Communication Demand	64 (16)	68 (17)	—	94 (8)	—
*Time Sharing Demand	54 (17)	48 (29)	—	88 (13)	—
*Team Effectiveness	54 (20)	37 (23)	—	73 (17)	—
*Team Support	64 (16)	67 (20)	—	66 (29)	—
*Team Dissatisfaction	35 (22)	14 (18)	—	24 (22)	—
Group size	3.6 (0.9)	2.7 (0.5)	—	2.0 (0.0)	—

When selecting only the sub-set of participants in GF with smaller average team size, closer to the team size in IF, for comparison, more of the differences show significance, as the data summarised in the 2nd and the 4th columns of Table 3 are compared between each other. The differences that had been previously found when considering the whole GF are comfirmed for communication demands (U=3.5; p=0.00).

Association of scales (within subjects) for expanded instrument (both groups)

The 12 expanded NASA-TLX rating scales were correlated against each other yielding the significant results depicted in Table 4 (considering both groups below the diagonal and only GF above the diagonal, which may emphasize which associations are tied in part to differing experimental conditions and which are not; an association shown above and below the diagonal is deemed more robust). The positive moderate association between performance and mental demand shows up consistently in the top left quadrant of Table 4 (correlations amongst the standard TLX scales). Crossing the standard and expansion TLX rating scales shows that temporal demand is consistently positively correlated with team effectiveness and team dissatisfaction (but team effectiveness and team dissatisfaction do not correlate amongst each other). Within the new team workload scales, correlations are plentiful. Those significant and consistent below and above the diagonal of Table 4 lay between communication and coordination demands, as well as between team support and both communication and coordination demands. Team effectiveness was found to be consistently moderately and positively correlated with both communication and time sharing demands.

Within subjects workload scale change (controlled for order of simulation type)

Aggregate change in each rating scale (the workload scale shown was obtained for each participant and condition by summing the ratings for mental, physical and temporal demands together with effort) is shown in Table 5. No statistical significance was found in the differences between the level of change that was incurred on the standard TLX and the compounded workload scales moving from the first simulation to the second one, across groups. Moreover, the one sample T-test, with test value zero, in GF, only showed significance (p=0.04) for mental demand change and workload change (p=0.02), while approaching significance (p=0.06) for effort change. In IF, tests did not yield significance.

The assumption that practice leads to improved individual performance, irrespective of the order in which its two experimental conditions (group and solo) are experienced by the participant, was further tested by joining both groups (last column in Table 5) and performing the one sample T-test for the test value of zero. This yielded significance for mental demands change (p=0.02), for physical demands change (p=0.03) and for workload change (p=0.02), but not for performance. Hence, the aforementioned assumption was not confirmed in the analysis.

*Table 4. Significant correlations (Spearman) encountered among the rating scales of the expanded TLX (legend: * - p< 0.05; ◊ - p< 0.01) joining both groups in the group condition (n=38) below the diagonal, and considering only GF above the diagonal (n=30).*

Rating scale	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1.Mental Demand	1			+.5*								
2.Physical Demand		1										+.4*
3.Temporal Demand	+.5◊		1							+.4*		+.4*
4.Performance	+.4*			1								
5.Effort					1							
6.Frustration Level						1						
7.Coordination Dem.						-.3*	1	+.5◊			+.6◊	
8.Communication D.							+.5◊	1		+.4*	+.5◊	
9.Time Sharing Dem.								+.6◊	1	+.4*		
10.Team Effectiven.			+.5◊				+.4*	+.5◊	+.4◊	1		
11.Team Support						-.3*	+.4*	+.4◊	+.3*		1	
12.Team Dissatisfact.			+.4*			+.4*		-.4*			-.4*	1

Table 5. Change in ratings of the standard TLX scales, from the first to the second simulation run, across groups (mean and standard deviation in parentheses; workload score obtained from adding effort to mental, physical and temporal demands ratings).

Standard TLX rating scale	*GF (n=6)*	*IF (n=8)*	*Both groups (n=14)*
Mental Demand (change)	20 (17)	11(24)	15 (21)
Physical Demand (change)	8 (19)	15 (20)	12 (19)
Temporal Demand (change)	7 (16)	18 (31)	13 (25)
Performance (change)	3 (30)	12 (51)	8 (42)
Effort (change)	11 (11)	4 (34)	7 (26)
Frustration Level (change)	2 (33)	-19 (46)	-10 (41)
Workload (change)	46(33)	48 (88)	47 (68)

Discussion

Effect of individual practice on group activity

Significant differences in the outcomes across two groups appeared for team communication and team time-sharing demands, which were higher for participants who had undergone singly practice prior to group activity. No significant differences were found across groups for individual performance and team effectiveness in the group condition.

Verification of assumption that practice leads to improved performance

Although on average there was an overall self-assessed performance increase of 8 percentage points (only 3% in GF and as much as 12% in IF) it was not significantly different from zero. Moreover, the conditions in GF may have increased the likelihood of a more intensified workload in the second simulation (carried out alone), for a marginal improvement in performance, compared to IF. Interestingly, workload (obtained from adding effort with mental, physical and temporal demands ratings) increased significantly from the first to the second experimental condition considering both groups united.

Conclusion

The results bear implications on training of decision-making teams, suggesting that singly practice of team members preceding group practice supports team-decision making effectiveness within teams going through the first stages of a system or problem-solving learning curve.

Limitations of the study

The study was based on a video-game based simulation. Kühn et al. (2014) reported on an anatomically based corroboration for association between frequent video-game playing and improvement in cognitive functions. Although participants had not previously interacted with the simulation used, previous experience with video-games at large was not controlled in this study. Hence, the evolution of each participant's individual workload and performance assessments from the first to the second simulation run could have been influenced by general video-gaming experience.

References

Berson, I.R., & Berson, M.J. (2008). Weathering natural disasters with a net of safety. *Social Education*, *72*, 27–30.

Brigantic, R.T., Muller, G.A., Taylor, A.E., & Papatyi, A.F. (2009). Gaming to predict human responses to mass causality events. In *Proceedings of the International Conference on Computational Science and Engineering*, Vancouver: IEEE, 1194-1198.

Coelho, D.A., Harris-Adamson, S., Lima, T.M., Janowitz, I., & Rempel, D.M. (2013-a). Correlation between different hand force assessment methods from an epidemiological study. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *23*, 128-139.

Coelho, D.A. (2013-b). Cognitive engineering and emergency management. In *Proceedings of the 10th international conference on Engineering Psychology and Cognitive Ergonomics: understanding human cognition*-Volume Part I (pp. 197-204). Springer-Verlag.

Coelho, D.A. (2013-c). Editorial: Supporting cognition in the management of disasters and emergencies. *Int. J. Human Factors and Ergonomics*, *2*, 1-10.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). Measuring team knowledge. *Human Factors*, *42*, 151-173.

De Winter, J.C.F. (2014). Controversy in human factors constructs and the explosive use of the NASA-TLX: a measurement perspective. *Cognition, Technology & Work*, 1-9.

DeChurch, L.A., & Mesmer-Magnus, J.R. (2010). The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of Applied Psychology*, *95*, 32-53.

Funke, G.J., Knott, B.A., Salas, E., Pavlas, D., & Strang, A.J. (2012). Conceptualization and measurement of team workload: A critical need. *Human Factors*, *54*, 36-51.

Gaillard, J.C., & Pangilinan, M.L. (2010). Participatory mapping for raising disaster risk awareness among the youth. *Journal of Contingencies and Crisis Management*, *18*, 175-179.

Hagemann, V., & Kluge, A. (2013). The effects of a scientifically-based team resource management intervention for fire service teams, *Int. J. Human Factors and Ergonomics*, *2*, 196-220.

Hamman, W. R. (2004). The complexity of team training: what we have learned from aviation and its applications to medicine. *Quality and Safety in Health Care*, *13(suppl 1)*, i72-i79.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Hancock, P.A., Meshkati, N. (eds.) Human mental workload*, North Holland Press, Amsterdam, 139-183.

Helton, W.S., Funke, G.J., & Knott, B.A. (2014). Measuring Workload in Collaborative Contexts: Trait Versus State Perspectives. *Human Factors*, *56*, 322-332.

Hung, W. (2013). Team-based complex problem solving: a collective cognition perspective. *Educational Technology Research and Development*, *61*, 365-384.

Khalid, H. M., & Helander, M. G. (2013). *Psycho-Cultural Analysis of Disaster Risk Attitudes in Situation Awareness*. Damai sciences sdn bhd, Kuala Lumpur (Malaysia).

Kühn, S., Lorenz, R., Banaschewski, T., Barker, G.J., Büchel, C., Conrod, P.J., Flor, H., Garavan, H., Ittermann, B., Loth, E., Mann, F., Nees, F. Artiges, E., Paus, T., Rietschel, M., Smolka, M.N., Ströhle, A., Walaszek, B., Schumann, G., Heinz, A., & Gallinat, J., The IMAGEN Consortium (2014). Positive association of video-game playing with left frontal cortical thickness in adolescents. *PloS one*, *9*, e91506.

NASA (1986). *Task Load Index (NASA-TLX). v. 1.0. Paper and pencil package (instruction manual)*. NASA Ames Research Center, Moffett Field, CA. http://humanfactors.arc.nasa.gov/groups/TLX/downloads/TLX.pdf

NASA (2014). *Task Load Index paper and pencil version*. NASA Ames Research Center, Moffett Field, CA. http://humansystems.arc.nasa.gov/groups/TLX/downloads/TLXScale.pdf

Silber, K.H., & Foshay, W.R. (eds.) (2009). *Handbook of improving performance in the workplace, instructional design and training delivery*. Vol. 1. John Wiley & Sons.

Simões-Marques, M., & Nunes, I.L. (2013). A Fuzzy multicriteria methodology to manage priorities and resource assignment in critical situations. In: *V. Zeimpekis, S. Ichoua, I. Minis (Eds.) Humanitarian & Relief Logistics: Research issues, case studies and future trends*, Operations Research/ Computer Science Interfaces Series (ORCS), Vol. 54, chapter 7, 129-153, Springer.