Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: An example with driving

Edith Galy, Julie Paxion, Catherine Berthelon

Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score:
An example with driving

Edith GALY[a], Julie PAXION[b], and Catherine BERTHELON[c]

[a] *University Côte d'Azur, LAPCOS, 24 avenue des Diables Bleus, 06357 Nice cedex 04, France.*

[b] *French Armed Forces Biomedical Research Institute, ACSO, BP 73, 91223 Brétigny-sur-Orge Cedex*

[c] *IFSTTAR, TS2, LMA, F-13300 Salon de Provence, France*

Introduction

The objective of this work is to improve current understanding of how the characteristics of an individual, a task or a work situation can influence mental workload.

Mental workload concept is very used but it is currently still difficult to define it. According to Young, Brookhuis, Wickens and Hancock (2015), numerous definitions and interpretations of mental workload exist, but many of them agree on the fact to consider two components, stress and strain. Stress concerns task demand and strain concerns impact on the individual (Schlegel, 1993). Others authors add the attentional resources available to these concepts (e.g. Wickens, 2002). Thus, for Young and Stanton (2005), mental workload would reflect amount of attentional resources necessary to perform task as a function of task demand, environmental context in which the task is performed, and past experience of individual with task. In our study, it is this definition that we consider. Thus, mental workload can be determined by matching individual resources with activity cost, and this cost would depend on task characteristics, environmental context, and individual experience with task. To respond to this definition, we will consider resources of individual through multidimensional activation model of Thayer (1986) and, different factors, responsible for activity cost and determining resources used to perform task, that are considered through cognitive load theory of Sweller (1988). These theories are developed following.

Although definitions of mental workload used in ergonomics account of these different components or dimensions, this principle has yet to form the basis of any theory. Only the field of education develops a model that features different categories of workload and attempts to elucidate their mutual relations. This is Sweller's cognitive load theory (1988) which identifies three categories of cognitive load. *Intrinsic cognitive load* refers to load imposed by the intrinsic nature of the information being processed. The greater the difficulty

of the task is, the greater the intrinsic load will be. This category of workload is unavoidable, in that it is inherent to the task. *Extraneous cognitive load* is induced by external factors (e.g., work organization, time pressure, background noise), which can widely vary, depending on the context (Sweller 1994). *Germane load* is defined as the load imposed on working memory by the construction and automation of schemas (Paas, Renkl, and Sweller 2003; Sweller, van Merriënboer, and Paas 1998). More recently, Schnotz and Kürschner (2007) suggested that germane load corresponds to the conscious application of strategies, thus reflecting metacognitive processes.

Cognitive load theory has been tested on many occasions in the fields of educational psychology and the psychology of learning (for a review, see Chanquoy, Tricot, and Sweller 2007). In recent years, it has at long last been applied to the field of ergonomics (Debué and van de Leemput 2014; Galy et al. 2012; Galy and Mélan 2015; Mélan and Cascino 2014), where authors now generally agree that mental workload has at least two components: task requirements and task context (Collet et al. 2009; Gaillard 1993; Luczak and Göbel 2000). Some aspects of mental workload therefore depend on features of the task. In previous studies (Debué and van de Leemput, 2014; Galy et al. 2012; Galy and Mélan 2015), parallels were drawn with Sweller (1988)'s intrinsic load, which is directly dependent upon task demand, and extraneous load, which is determined by the context in which this task is performed.

The issue is somewhat different when it comes from germane load, for there appears to be nothing in the ergonomics literature on mental workload that corresponds to this category. We do, however, come across this notion in research on expertise. In ergonomics, the term *expertise* refers to the ability of experts to perform some tasks extremely effectively, by applying task-specific strategies that allow them to reduce the cost of processing information. Experts are good at planning and can adopt behaviours that allow them to carry tasks out quickly and efficiently (Hung 2003). To do so, they create schemas to categorize or group

together items of information. These schemas are stored in long-term memory, and the more people are expert in a given field, the more schemas they have and the lower the cognitive cost of performing tasks relating to this field is. Whenever a new problem presents itself, experts retrieve the processes they learned to resolve similar problems in the past, in order to come up with possible solutions (Bradley, Paul, and Seeman 2006). Similarly, the experience they have gained allows experts to construct new schemas more quickly, and their knowledge enables them to apply the most effective and most relevant strategies to each particular task (van Gog, Paas, and van Merriënboer 2005).

According to Schnotz and Kürschner (2007), Sweller's *germane load* is induced by the application of just such metacognitive strategies and processes with the aim of successfully completing the task. It is possibly that this category of mental workload is even more reliant on the availability of cognitive resources than the two others on which it therefore depends. For instance, a specific task performed in a given context necessarily generates a degree of intrinsic and extraneous load, so if the task is difficult and the context unfavourable, these two types of load will be very high, draining cognitive resources and leaving very little left for germane load (Galy et al. 2012; Galy and Mélan 2015; Young, Watcher, Cate, O'Sullivan, and Irby, 2016). Even if intrinsic and extraneous loads do not use up all the resources, the remainder is not necessarily allocated to germane load. For this to happen, the operator must have a certain degree of expertise in the field, and be capable of activating schemas and previously learned strategies.

Germane load is thus dependent upon the overall amount of cognitive resources available. This amount depends on general activation of individuals. Thayer (1986) developed a model of multidimensional activation and described two dimensions of activation. These two dimensions are energetic arousal and tense arousal. According to Thayer (1987), "tense arousal is assumed to be determined by danger, broadly interpreted, and to be largely

cognitively mediated. But variations in energetic arousal are believed to occur naturally, as a function of such factors as time of day, exercise, and nutrition." (Thayer, 1987; p. 21). Thus, energetic arousal corresponds to alertness and would be very depending on biological rhythms, while tense arousal corresponds to tension and would depend on external factors, as the occurrence of stressful events (Matthews, Jones and Graham Chamberlain, 1990; Koscec and Radosevic-Vidacek, 2004).

Previous researches (Galy, Cariou, and Mélan 2012; Galy and Mélan 2015) had shown that performance on tasks such as mental arithmetic or memorization can only be explained if mental workload is broken down into several different categories. In these studies, the authors controlled the participants' level of alertness and/or cognitive appraisal of the task, and manipulated task difficulty and time pressure. Results of these studies have shown that when the task is difficult and the time pressure is high, individuals are only able to allocate resources to germane load if they have a high level of alertness (or energetic arousal) and/or they are challenger with regards to the task (cognitive appraisal of task). On the other hand, individuals haven't sufficient attentional resources or necessary skills (lack of experience) to implement the relevant strategies to ensure a good performance.

In the present study, we set out to test the existence of three categories of mental workload described by Sweller's theory in a more complex task (driving in a car simulator) than the memorization or mental arithmetic tasks used in previous studies. We looked at three factors thought to determine germane load, extraneous load, and intrinsic load taking into account the complexity of the driving situation, individual expertise that is driving experience, and the amount of available resources with drivers' levels of alertness and tension.

According to Verwey (2000), the driving situation is the major determinant of the driver's mental workload. Meister (1976)'s model establishes the relationship between more or less complex situations and performance, according to workload. Driving performance is

good when the task is moderately costly (i.e., moderately complex situation), but poor in situations that are either monotonous and repetitive, lacking in variation (i.e., simple situation), or extremely costly, with large amounts of information to process (i.e., highly complex situation). In moderately complex situations, therefore, a good performance can be maintained by investing effort.

Fastenmeier (1995) classified 134 driving situations according to their complexity, using Frieling and Hoyos (1978)'s Position Analysis Questionnaire to assess images of these situations. Complexity was determined on the basis of three main criteria: type of road (urban, rural, motorway), nature of that road (horizontal with vs. without bends, vertical with vs. without bends, intersections with vs. without signposting), and traffic flow (lane narrowing, traffic jam). The results of his study showed that a driving situation featuring congested roads in and around cities, with bends or intersections is highly complex. Based on Fastenmeier (1995)'s classification, we decided to take three levels of complexity into account in the present study (see the procedure part).

The workload generated by the driving activity can also explain why young drivers have been found to be disproportionately involved in road accidents, despite the introduction of a three-year probationary period after the obtention of the driving license. Learner drivers have been found to have a lower threshold of overload (Paxion, Galy and Berthelon 2015), probably owing to their lack of automated routines (Amalberti 1996; Cegarra and Hoc 2006; De Craen et al. 2008; De Waard 1996) and the higher cognitive cost of driving (Damm et al. 2011). Because of this low level of automatization, they have greater difficulty implementing the relevant strategies and allocate fewer resources to germane load, even if the task is simple and performed in a favourable context. Experienced drivers only go beyond their overload threshold, where increased workload leads to reduced performance, when a situation is particularly complex.

A further factor that has a major effect on germane load is the level of activation, as the individual's activation state is thought to determine the amount of available resources (Galy et al. 2012; Galy and Mélan 2015). According to Thayer (1986)'s model of multidimensional activation, arousal is the product of two separate but interconnected processes, namely alertness and tension. *Alertness* refers to a state of energetic arousal that allows an individual to perform an activity. *Tension*, meanwhile, corresponds to a state of tense arousal or physiological anxiety that increases when the activity becomes more intense and/or the level of alertness falls. Thayer (1986) claimed that increased tension can compensate for a decline in alertness, thus ensuring that there are sufficient energy resources to sustain performance. Alertness and tension must thus be considered to correctly gauge an individual's activation state and thus the amount of resources available to perform a given task.

The mechanisms associated with driving require cognitive resources, energy, and therefore a degree of arousal of the nervous system (i.e., alertness) which varies according to task complexity, situation and time of day (Philip et al. 2005). A lack of alertness can mean that drivers are unable to resist falling asleep, causing road accidents with particularly high levels of mortality and morbidity (Hadj-Mabrouk et al. 2001). According to Campagne et al. (2004), excessive sleepiness could be the main reason for a deterioration in driving performance. It can therefore conclude that a high level of alertness ensures that the driver has enough resources to carry out the task, and because the cost of the driving activity is therefore relatively low, it reduces the size of the cognitive load imposed by the task itself and by the context in which it is performed (intrinsic and extraneous load). By the same token, a low level of alertness results in a high mental workload, insofar as the driver's activation state reduces the amount of cognitive resources available, thus forcing the driver to expend considerable effort on performing the task. We can thus assume that if a hazardous event

occurs while driving, a driver with a high level of alertness will be able to allocate the necessary resources to the germane load, and thus implement the strategies needed to manage the event. With a low level of alertness, however, this will not be the case, possibly resulting in an accident. This problem is, of course, exacerbated if the driving situation is particularly complex or if the driver is not experimented.

However, in link with Thayer (1986)'s multidimensional model, Collet et al. (2005) showed that a good driving performance relies on physiological arousal. In their study, drivers who managed not to hit the obstacle in a critical crash avoidance situation had greater electrodermal activity than those who caused an accident. This means that drivers probably need a degree of anxiety in order to be sufficiently alert to react in time in a critical situation. The lower their level of alertness is, the more anxiety they will need. Thus, it is important to use in this study a measure of activation state which evaluates both dimensions of activation. This is why, in our study, we focalised on Activation Deactivation Adjective Checklist developed by Thayer (1986). This tool allows to evaluate the two dimensions of his multidimensional arousal model, energetic arousal and tense arousal. These evaluations will permit to estimate amount of available resources of individuals at the time of task execution.

Concerning the measurement of mental workload, there is a plethora of subjective, performance-related or psychophysiological indicators of mental workload, but they do not all measure the same component (Galy et al. 2011; Galy et al. 2012; Galy and Mélan 2015; Mélan et al. 2012; Ucelli et al. 2011). The ergonomics literature has long made a distinction between the different components of mental workload, not least in the design of measurement tools. Indeed, the two most widely used tools for measuring workload - the NASA-Task Load Index (NASA-TLX; Hart and Staveland 1988) and the Subjective Workload Assessment Technique (SWAT; Reid and Nygren 1988) – are both based on just such a breakdown. The NASA-TLX comprises six subscales (mental demand, physical demand, time demand,

performance, effort and frustration), while the SWAT only covers three dimensions (time load, mental effort load, and psychological stress load). The vast majority of researchers nevertheless ignore the structure of these scales and only use their overall scores. The few that do consider the subscale scores observe differential effects of workload. One study, for instance, looked at the impact of workload components on radiotherapy incidents (Mazur et al. 2012), while another used these components to assess flight deck design (Zheng et al. 2011). Another study showed also, in virtual reality context, relevance to consider each sub-scale of NASA-TLX. Indeed, as a function of role given to participants (guide or manipulator), it existed or not significant differences between dimensions (Pouliquen-Lardy, Milleville-Pennel, Guillaume and Mars 2016). In the present study, we focus on the multidimensional NASA-TLX to find out what is exactly measured by each of these dimensions, and how they are related to each other. By varying the factors of intrinsic and extraneous load, as well as the factors of germane load, it is expected that the dimensions vary in their sensitivity to these factors, and thus assess specific load categories.

The hypotheses of present study can be formulated at three levels. Thus, at a first level, physical, mental and temporal demands could be considered as indicators of load factors and be determined by situation complexity. Effort, whose definition corresponds to the cost represented by task execution for individual (Galy et al., 2012; Galy and Mélan, 2015), would be determined by physical, temporal and mental demands, and arousal (alertness and tension), with a lower effort when demands are low and arousal is high. Frustration, corresponding to emotional dimension of NASA-TLX, could be sensitive to tension level of drivers.

Driving performance was assessed during hazardous events, namely pedestrians crossing the road, the criterion being the number of collisions with these pedestrians, and during driving session with standard deviations of lateral position of vehicle (SDLP). Thus, at a second level, objective performance (number of collisions with pedestrians and SDLP)

could be determined by situation complexity, different demands, arousal and experience of driving. Consequently, performance would become poorer as the complexity of the driving situation or/and demand increases, especially in the case of novice drivers or low levels of arousal (alertness and tension). A lack of experience, and a lack of cognitive resources, would make it difficult for participants to engage in processes that generate germane load. The deterioration of novice drivers' performance in a complex situation would be compounded if they had a low level of alertness that was not offset by a high level of tension. Finally, at a third level, own performance would be determined by objective performance and experience of driving. Indeed, novice drivers could have more difficulties to evaluate relevant situation and their own performance.

Method

The experiment took place in a driving simulator (horizontal field of view: 150°; vertical field of view: 40°) at the French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR) in Salon-de-Provence, France. This simulator is equipped with a multi-actor parallel architecture for traffic simulation (ArchiSim) and an object database SIM²-IFSTTAR (simulation software) (Espié, Gauriat and Duraz, 2005). The "ArchiSim" architecture was built on the DR2 traffic simulation model (management of "autonomous" and "enslaved" vehicle with a behavior defined by the scripts for each scenario, simulation generated by captors of punctual and space traffic) and on the 3D SIM2 loop of visualization. The interactive driving station comprised one quarter of a vehicle including a seat, a dashboard, and controls equipped with captors, i.e. pedals and steering wheel.

The different movements and actions exerted on the vehicle were virtually reproduced. The acceleration, braking, and steering values of the simulator were those of an average vehicle. The captors installed on the simulator recorded the values in real time. The simulator

had an automatic gearbox and was not equipped with rearview mirrors. The image projection (30 Hz) surface, placed on three screens in front of the simulator at 1.93 m of the driver's eye, filled an angular opening that spanned 150° horizontally and 40° vertically. Each screen had a 1280 x 1024 pixels resolution. A sound in quadriphonic diffused in the simulator consisted of internal noise (motor, bearing and starter) and external noise (traffic).

Participants

Drivers were divided into three groups according to their driving experience. The first group comprised 15 drivers (6 females and 9 males) who had passed their driving test within the previous 2 months (mean age = 19 years ± 0.84), the second contained 15 drivers (7 females and 8 males) who had passed their driving test at least 3 years earlier (mean age = 21 years ± 0), and had therefore reached the end of their probationary period, and the third contained 15 drivers (5 females and 10 males) who had held a driving licence for more than 5 years (mean age: 26.87 years ± 2.97).

Procedure

All the participants were exposed to three separate driving situations, each over a distance of 22.5 km, but with a different level of complexity. The simple situation involved a straight two-lane A-road without any traffic. The moderately complex situation featured a combination of straight lines and right- or left-hand bends (length: 600 m, radius: 300 m) without any traffic. The highly complex situation had twice as many bends, and these were more difficult than before (length: 300 m, radius: 120 m). There was also oncoming traffic and traffic ahead of the participant. Each of these three situations included three scenarios involving a pedestrian somewhere along the way. Concealed by different types of scenery, these pedestrians stepped out into the road just ahead of the participant, who had approximately 2.7 seconds to take avoiding action. Participants were instructed to drive at a

maximum speed of 90 km/hour for all scenarios. We measured number of collisions with pedestrians and standard deviations of lateral position of vehicle (SDLP) during driving.

Before the driving test, participants completed the Thayer's Activation-Deactivation Adjective Check List (AD-ACL; 1986). After each driving session in one of the three situations (simple, moderately complex, or highly complex), they filled in the NASA-TLX for when the pedestrians crossed the road. Thus, participants were asked to only estimate workload generated by portions of driving involving pedestrians at the end of each situation. The order of the situations was randomly counterbalanced across participants.

Questionnaires

The AD-ACL enabled us to collect participants' subjective levels of alertness and tension. Respondents rated their experience of each of its 20 adjectives as *definitely feel*, *feel slightly*, *cannot decide*, or *definitely do not feel*. These adjectives belong to four different dimensions, and respondents' level of alertness can be estimated by summing their scores on General Activation (GA) and Deactivation-Sleep (DS), and their level of tension by summing their scores on High Activation (HA) and General Deactivation (GD).

The NASA-TLX (Hart and Staveland 1988) was used to measure participants' subjective level of workload. Participants rated five of its six dimensions (mental demand, physical demand, temporal demand, effort, and frustration) on a 20-point scale ranging from 0 (*very low*) to 20 (*very high*). The endpoints of the sixth subscale (own performance) were *success* and *failure*.

Results

Stepwise regression analyses were run to study the effects of driving experience, situation complexity, levels of alertness and tension (Thayer's scale), and the interactions between these factors, on participants' scores on each dimension of the NASA-TLX.

Analyses showed that these factors had widely differing effects, depending on the dimension. Mental and temporal demand, for instance, were determined only by the complexity of the situation (r^2 adj = .150, β = .398, $p < .001$ and r^2 adj = .063, β = .269, $p < .01$), whereas the physical demand dimension was determined both by situation complexity and by the interaction between experience and level of tension (r^2 adj = .100, β = .235, $p < .05$ and β = -.251, $p < .01$). Therefore, the more complex the situation was, the greater the perceived mental, physical and temporal demand were. To explain interaction effect between experience and tension level, partial regression analyses were conducted for physical demand as a function of tension level. Thus, low and high levels of tension were considered according to median value of this dimension. Results of these analyses revealed that the physical demand dimension was only determined by experience when tension was low (low tension: β = -.280, $p < .05$; high tension: β = -.163, ns), in which case a greater degree of driving experience resulted in lower physical demand.

As for the frustration dimension, it was determined by the level of tension (r^2 adj = .042, β = .410, $p < .001$), such that the greater the tension was, the greater the perceived frustration was.

As the effort dimension was not explained by any of the analysed factors, additional regression analyses were conducted to test the effects of frustration, mental, physical and temporal demand, alertness, and tension on self-reported effort. Results indicated significant effects of mental and physical demand, alertness and interaction between alertness and mental demand (r^2 adj = .309, respectively β = .357, $p < .001$, β = .258, $p < .05$, β = .359, $p < .05$, β =

-.603, $p < .05$), such that when mental and physical demand and alertness were high, so was also effort. Partial regression analyses were run to study the effect of mental demand on effort as a function of alertness level and to explain interaction effect between alertness and mental demand. When alertness was low, mental demand had significant effect on self-reported effort ($\beta = .613$, $p < 0.001$). This effect ceased to be significant when alertness was high ($\beta = .264$, *ns*).

A further stepwise regression analysis explored the effects of experience, situation complexity, alertness and tension levels, as well as NASA-TLX dimensions (mental demand, physical demand, temporal demand, effort and frustration), on the number of collisions with pedestrians. Results revealed only a marginal effect of interaction between alertness and experience (r^2 adj = .02, $\beta = -.150$, $p = .081$). Although this effect is only marginal, partial regression analyses showed a significant effect of experience when alertness level was high (r^2 adj = .104, $\beta = -.357$, $p < .05$). This effect wasn't significant when alertness level was low ($\beta = .022$, *ns*). Thus, when alertness level of drivers was high, novice drivers presented a mean number of collisions higher (mean = 1.22 ± 0.83) than drivers with 3 years and more than 5 years of experience (respectively, mean = 0.42 ± 0.67; mean = 0.50 ± 0.51).

Another regression analysis explored the effects of experience, situation complexity, alertness and tension levels, as well as NASA-TLX dimensions (mental demand, physical demand, temporal demand, effort and frustration), on SDLP. Results showed a significant effect of experience, self-reported effort, and alertness level (r^2 adj = .422, $\beta = -.211$, $p < .01$; $\beta = .619$, $p < .001$; $\beta = -.150$, $p = .05$). Thus, mean SDLP were higher for novice drivers (mean = 9.51 ± 3.34) than for drivers with 3 years and more than 5 years of experience (respectively, mean = 7.40 ± 2.42; mean = 7.97 ± 2.15), and when self-reported effort was high and alertness was low.

The analyses exploring effects of experience, SDLP, and number of collisions on own performance also showed that the own performance dimension was explained by experience, with the most experienced drivers providing the highest ratings on the 20-point *success* to *failure* scale, by SDLP, and by marginal interaction between experience and number of collisions (r^2 adj = .285; respectively, β = -.248, *p* < .05; β = .453, *p* < .001; β = .365, *p* = .08). Partial regression analyses revealed that number of collisions had a significant effect on own performance for the most experimented drivers (more than 5 years of experience; β = .425, *p* < .01). This effect became marginal for drivers with 3 years of experience (β = .340, *p* = .08), and non-significant for novice drivers (β = -.015, *ns*). All these results are summarized in Figure 1.
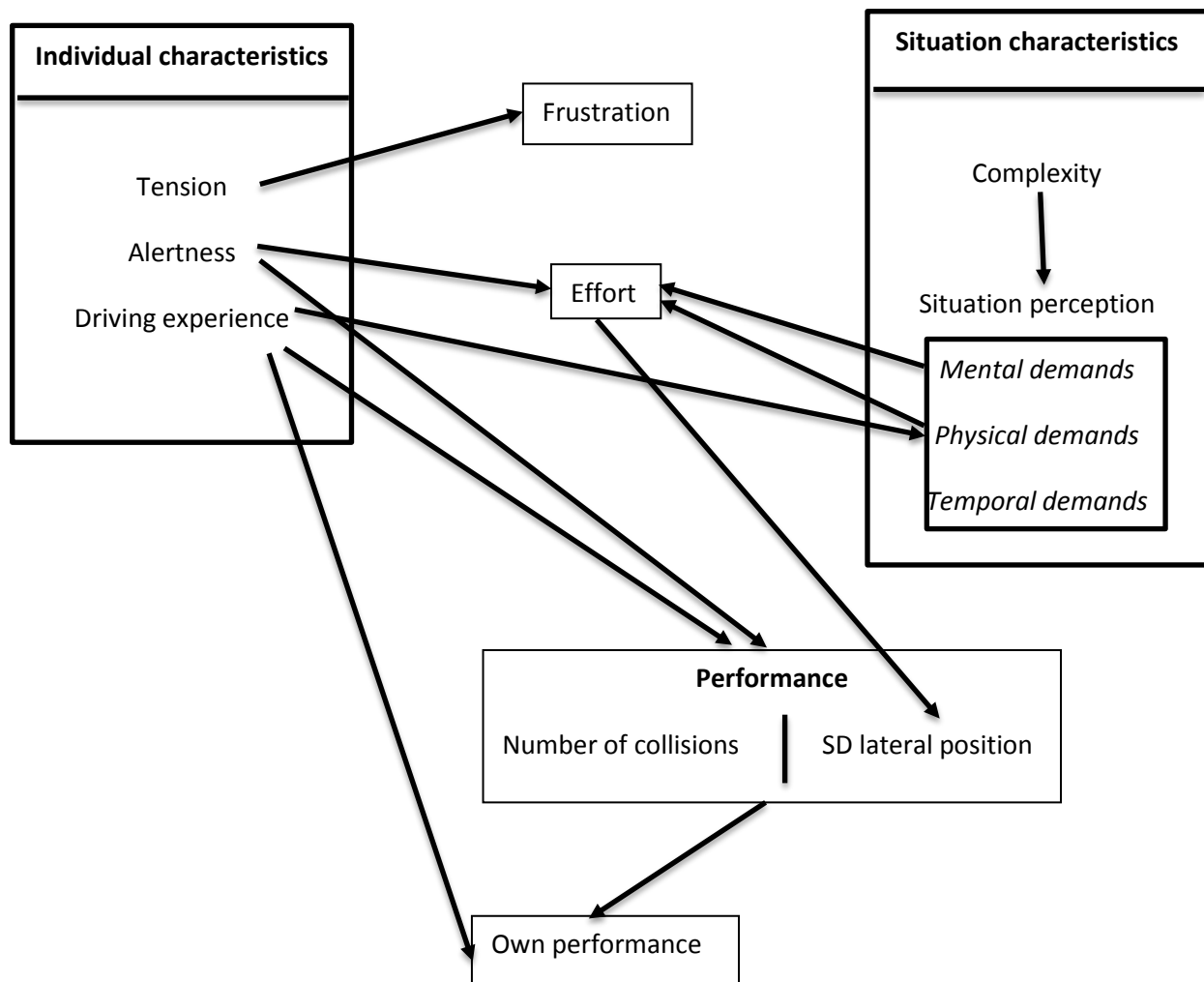
Figure 1: Schematic representation of the significant relationships between the variables entered into the successive regression analyses.

Discussion

Results obtained in this study allow us to partially confirm our hypotheses. Indeed, physical, mental and temporal demand was determined by situation complexity and these dimensions of NASA-TLX seem to be as assumed indicators of intrinsic and extraneous workload factors. Thus, when driving situation is more complex, drivers perceive this

complexity and estimate that task demand is more important. However, even if perceived demand is high, perceived effort is not necessarily high. Indeed, results reveal that effort is the resultant of interaction between alertness and task demand. Thus, effort is estimated high when mental and physical demand (intrinsic load) is high and alertness is low. Consequently, if drivers have sufficient amount of cognitive resources to perform the task, they are conscious that the task is complex and so demanding but they estimate that effort to provide is not very intense because their alertness level allows them to easily mobilize necessary resources (Galy et al. 2012; Galy and Mélan 2015).

Otherwise, according to our hypotheses, frustration dimension of the NASA-TLX is determined by tense arousal. Thus, when drivers begin their task with a high level of tension, they feel a frustration with regard to the task. Drivers' feeling during driving depends on their tension state before the task. This result is not surprising since tense arousal can be compared to physiological anxiety (Thayer 1986).

Concerning the determinants of performance (number of collisions and SDLP), results do not agree with our hypotheses. Indeed, contrary to what was supposed, no effect of situation complexity or task demand is observed. Performance seems determined by individual characteristics, particularly alertness and driving experience. Whatever the situation complexity or task demand, the number of collisions and standard deviations of lateral position are lowest when drivers are experimented and their alertness level is high. Consequently, according to previous studies (Galy et al 2012; Galy and Mélan 2015; Schnotz and Kürsner 2007), driving performance depends on germane load. Indeed, experience only affected performance when alertness was high, that is when sufficient resources were available at the outset for some to be allocated to germane load, once intrinsic load had taken its share. Germane load is induced by the implementation of relevant strategies to maintain performance (Debué and van de Leemput, 2014; Galy et al. 2012; Galy and Mélan 2015;

Schnötz and Kurschner 2007), and experimented drivers have less difficulty in engaging the right strategies than novices (van God, Paas and van Merriënboer 2005; Vessey 1994) because they have constructed and automated more schemas.

In our hypotheses, we assumed that own performance would be due to performance indicators and driving experience. Results confirm this assumption. Thus, only the most experimented drivers have an estimation of their performance reflecting their actual performance. The other drivers, and particularly novice drivers, present an estimation of their performance that does not correspond to the actual performance evaluated by the number of collisions with pedestrians and SDLP. Consequently, this mismatching between actual performance and own performance by novice drivers could be due to a failing situation awareness (Endsley 1995). Indeed, this author defines attention and working memory as critical factors limiting individuals for acquiring and interpreting information from the environment to form situation awareness, and specifies that mental models (schemas stored in long-term memory) are important mechanisms to maintain satisfactory situation awareness when mental workload due to the task is high.

Another result, not expected but interesting, concerns determinants of physical demand. Indeed, besides situation complexity, driving experience and tension determine physical demand. This result can be explained by the fact that novice drivers are characterised by a lack of automated routines (Amalberti 1996; Cegarra and Hoc 2006; De Craen et al. 2008; De Waard 1996). In driving, these routines are essentially procedural and would be the origin of a higher physical demand by novice drivers.

In the other hand, analyses of this study showed that the six NASA-TLX dimensions assessing workload do not all measure the same thing, as they were not all sensitive to the same factors. Thus, our results revealed a clear difference between effort and task demand,

even though these dimensions were all subjectively estimated. We should not, therefore, consider the dimensions of the NASA-TLX on an equal footing. The participants did not appear to have any difficulty distinguished between the demand imposed by the task and the effort required to perform the task. Task demand may not vary, but the amount of effort required is contingent upon several features, not least the level of alertness. For this reason, the temptation of grouping all the items scores of the NASA-TLX into a single latent variable must not be yield, as its dimensions appear to reflect independent - albeit interrelated - processes. This distinction has already been made in ergonomics by Collet et al. (2009), who talk about two components of mental workload: task demand and cognitive effort. The dichotomy that emerged in our results has therefore already been highlighted, although rather than viewing task demand and effort as two equal components of mental workload, we would argue that the present results show that they cannot be placed at the same level, as effort is a *consequence* of task demand. Furthermore, the latter can be broken down into mental, physical and temporal demand. Finally, here, effort was determined by task demand (mental and physical) and by the drivers' level of alertness - a result that is in line with the most widespread definition of mental workload: the operator's ability to meet task demand with his/her available resources (Leplat 2002; Prichard, Stratford, and Bizo 2011).

The measures that are commonly used to estimate mental load actually assess either task demand or mental effort. Using a single measure therefore provides only a partial picture, as already emphasised by Collet et al. (2009), and explains why some authors recommend using a combination of load measures instead of looking for one ideal measure (Cegarra and Chevalier 2008; Miyake 2001). Actually, a single measure does not take account of the asymmetric relations between the various components of mental load. People do not simply have task demand on one side and effort on the other, each measurable with different tools, as the demand has an impact on effort that is moderated by the individual's functional state. If

we assume that mental load corresponds to the cognitive cost that the individual must incur in order to perform a task, effort would appear to be the only variable that truly constitutes an element of mental workload, the task demand being subjectively assessed factors for intrinsic or extraneous load.

All of these results can be modelled by the schema in figure 2. This schema represents putative relationships between cognitive load factors and mental load categories. Thus, complexity of driving situation is a load factor that can be estimated by three NASA-TLX dimensions (physical, mental and temporal demand). This factor determines intrinsic and extraneous cognitive loads. These loads represent cognitive resources obligatorily used by drivers during driving. As a function of their functional state (alertness and tension), drivers have a certain amount of available cognitive resources. The mapping between resources obligatorily used and available resources allows to obtain amount of remaining cognitive resources. The mismatching between these two elements can be estimated by effort dimension of NASA-TLX. Thus, when cognitive resources obligatorily used (intrinsic and extraneous loads) are great, they represent a very important part of available cognitive resources and the mismatching is low, but perceived effort is high. According to effort, it exists a certain amount of remaining cognitive resources. If these resources are sufficient and if drivers have the necessary skills (experimented drivers), they would be allocated to germane cognitive load and drivers could implement the relevant strategies to ensure a good performance and situation awareness evaluated by own performance dimension of NASA-TLX.
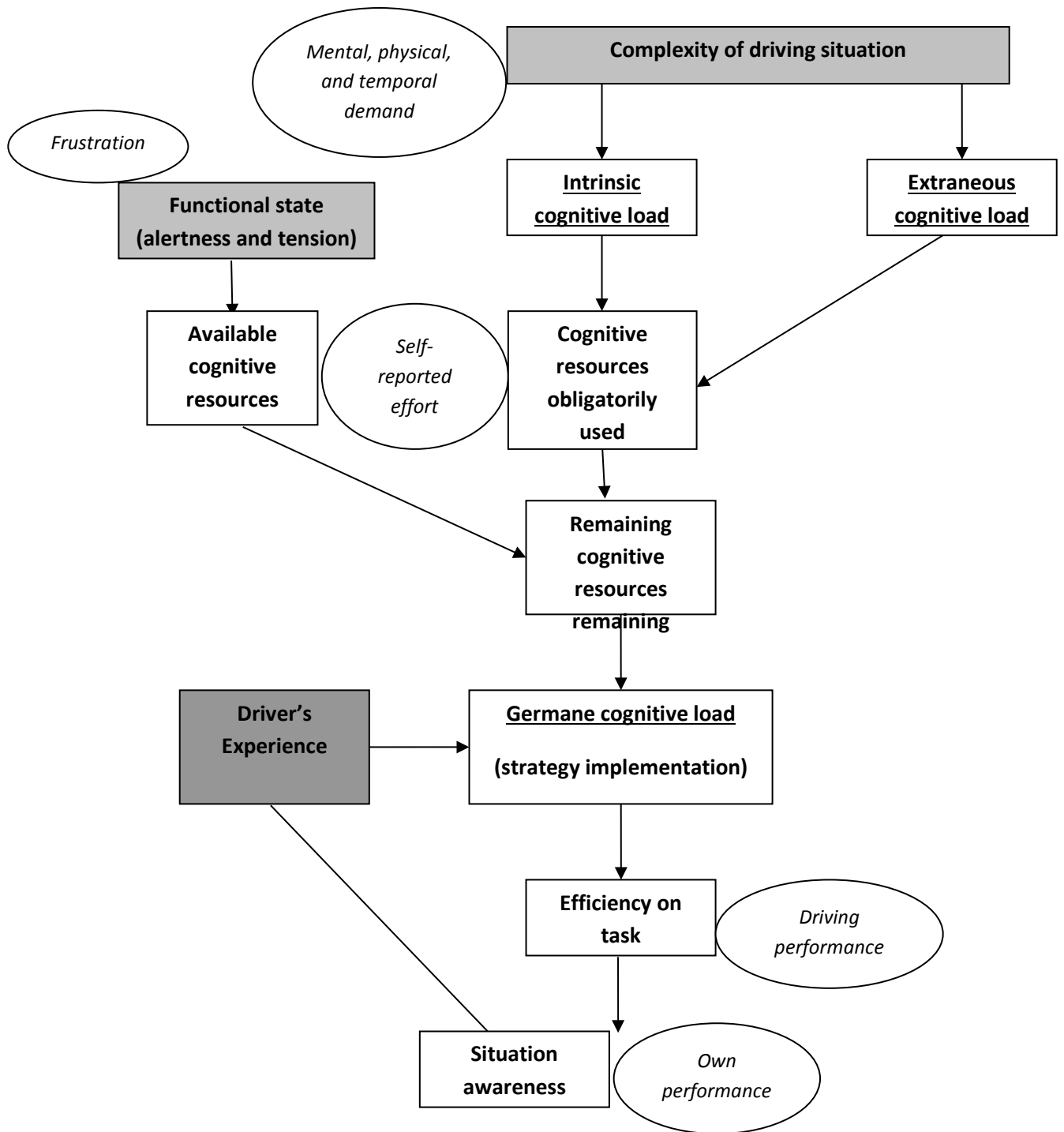
Figure 2: Graphical representations of putative relationships between cognitive load factors and mental load categories.

Conclusion

This study allowed to highlight relevance of Galy and colleagues' (Galy et al. 2012; Galy and Mélan 2015) model to explain performance to a complex task like driving. Thus, task performance can be explained by matching of three categories of mental workload, i.e. intrinsic, extraneous and germane load. Therefore, the measurement of subjective mental workload via the NASA-TLX questionnaire needs to take into account each of its dimensions rather than a global score. Other mental workload measurements could also be studied by testing the independence of their dimensions, as time load, mental effort load and psychological stress load for the SWAT questionnaire (Reid and Nygren 1988). In driving studies, it is thus important to moderate the results that imply a global score of mental workload, or at least to be aware that this score can refer to different categories (intrinsic, extraneous and germane load) that are not directly influenced by the same variables (driving situation, driving experience and driver's functional state). Furthermore results of this study suggested that lower performance of novice drivers actually come from a lack of experience making difficult implementation of appropriate strategies to the situation (van Gog, Paas and van Merriënboer 2005; Vessey 1994).

Finally, it appears interesting to transfer knowledge and tools derived from cognitive load theory to understand performance in complex task such as driving. Indeed, the adaptation of instrument developed by Leppink, Paas, van der Vleuten, van Gog and van Merrinboër (2013) in the learning domain could permit to study more precisely relationships between different cognitive load categories. That was initiated by Debué and van de Leemput (2014) for studying the information retention after reading of an online newspapers.

References
Amalberti, R. 1996. *La Conduite des Systèmes à Risques*. Paris: Presses Universitaires de France.

Bradley, J. H., R. Paul, and E. Seeman. 2006. "Analyzing the Structure of Expert Knowledge." *Information and Management* 43 (1): 77-91.

Campagne, A., T. Pebayle, and A. Muzet. 2004. "Correlation Between Driving Errors and Vigilance Level: Influence of the Driver's Age." *Physiology & Behavior* 80 (4): 515-524.

Cegarra, J., and A. Chevalier. 2008. "The Use of Tholos Software for Combining Measures of Mental Workload: Towards Theoretical and Methodological Improvements." *Behavior Research Methods* 40 (4): 988-1000.

Cegarra, J., and J.-M. Hoc. 2006. "Cognitive Styles as an Explanation of Experts' Individual Differences: A Case Study in Computer-Assisted Troubleshooting Diagnosis." *International Journal of Human-Computer Studies* 64 (2): 123-136.

Chanquoy, L., A. Tricot, and J. Sweller. 2007. *La Charge Cognitive: Théories et Applications.* Paris: Armand Colin.

Collet, C., P. Averty, and A. Dittmar. 2009. Autonomic Nervous System and Subjective Ratings of Strain in Air-Traffic Control. *Applied Ergonomics* 40: 23-32.

Collet, C., C. Petit, A. Priez, and A. Dittmar. 2005. "Stroop Color–Word Test, Arousal, Electrodermal Activity and Performance in a Critical Driving Situation." *Biological Psychology* 69 (2): 195-203.

Damm, L., C. Nachtergaële, M. Meskali, and C. Berthelon. 2011. "The Evaluation of Traditional and Early Driving Learning with Simulated Accident Scenarios." *Human Factors* 53 (4): 323-337.

Debue, N., and C. van de Leemput, C. 2014. "What Does Germane Load Mean? An Empirical Contribution to the Cognitive Load Theory." *Frontiers in Psychology* 5: 1099.

De Craen, S., D. A. M. Twisk, M. P. Hagenzieker, H. Elffers, and K. A. Brookhuis. 2008. "The Development of a Method to Measure Speed Adaptation to Traffic Complexity: Identifying Novice, Unsafe, and Overconfident Drivers." *Accident Analysis and Prevention* 40 (4): 1524-1530.

De Waard, D. 1996. "The Measurement of Drivers' Mental Workload." PhD diss., University of Gröningen.

Endsley M. R. 1995. "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors* 37 (1): 32-64.

Espié S., P. Gauriat and M. Duraz. 2005. Driving simulators validation: The issue of transferability of results acquired on simulator. *In National Advanced Driving Simulator, University of Iowa* (Eds) *Proc. Driving Simulation Conference DSC North-America'2005,*

october 30th - november 2nd, Orlando, FL., Center for Advanced Transportation Systems Simulation, 149-156.

Fastenmeier, W. 1995. "Die Verkehrssituation als Analyseeinheit im Verkehrssystem [The road traffic situation as analysis unit in the road traffic system]." In *Autofahrer und Verkehrssituation Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme*, edited by W. Fastenmeier, 27–78. Cologne: Verlag TÜV Rheinland.

Frieling, E., and C. G. Hoyos. 1978. *Fragebogen zur Arbeitsanalyse (FAA)* (German version of the Position Analysis Questionnaire). Bern: Huber.

Gaillard, A. W. K. 1993. "Comparing the Concepts of Mental Load and Stress." *Ergonomics* 36: 991-1005.

Galy, E., M. Cariou, and C. Mélan. 2012. "What is the Relationship Between Mental Workload Factors and Cognitive Load Types?" *International Journal of Psychophysiology* 83: 269-275.

Galy, E., C. Lapalus, C., J. Paxion, and J. Rivière. 2011. "Quelles Sont les Relations Entre Vigilance, Tension, Charge de Travail Effective et Charge de Travail Subjective Chez des Travailleurs Postés?" Paper presented at the Journées d'Etudes en Psychologie Ergonomique (EPIQUE), Nantes, September 5-7.

Galy, E., and C. Mélan, C. 2015. Effects of cognitive appraisal and mental workload factors on performance to an arithmetic task. *Applied Psychophysiology and Biofeedback* 40: 313-325.

Galy, E., C. Mélan, and M. Cariou. 2008. "Investigation of Task Performance Variations According to Task Requirements and Alertness Across the 24-h Day in Shift Workers. » *Ergonomics* 51 (9): 1338-1351.

Hadj-Mabrouk, A., H. Hadj-Mabrouk, and M. Dogui. 2001. "Chronobiologie de la Vigilance Approche d'Application dans le Domaine de la Sécurité Routière." *Recherche-Transports-Sécurité* 73: 3-26.

Hart, S. G., and L. E. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In *Human Mental Workload*, edited by P. A. Hancock and N. Meshkati, 139-183. Amsterdam: Elsevier.

Hung, S. Y. 2003. "Expert Versus Novice Use of the Executive Support Systems: An Empirical Study." *Information and Management* 40: 1777-1189.

Koscec. A., and B. Radosević-Vidacek. 2004. Circadian components in energy and tension and their relation to physiological activation and performance. *Chronobiology International* 21: 673–690.

Leplat, J. 2002. "Eléments Pour une Histoire de la Notion de Charge Mentale." In *Charge Mentale: Notion Floue et Vrai Problème*, edited by M. Jourdan and J. Theureau. Toulouse: Octarès.

Leppink J., F. Paas, C.P. van der Vleuten, T. van Gog, and J.J van Merrïenboer. 2013. "Development of an instrument for measuring different types of cognitive load." *Behavior research methods* 45 (4): 1058-1072.

Luczak, H., and M. Göbel. 2000). "Signal Processing and Analysis in Application." In *Engineering Psychophysiology: Issues and Applications*, edited by R. W. Backs and W. Boucsein, 79-110. Mahwah, NJ: Lawrence Erlbaum Associates.

Matthews G., D.M. Jones, and A. Graham Chamberlain. 1990. Refining the measurement of mood : the UWIST Mood Adjective Checklist. *British Journal of Psychology* 81: 17-42.

Mazur, L., P. R. Mosaly, M. Jackson, S. X. Chang, K. Deschesne Burkhardt, R. D. Adams, E. L. Jones et al. 2012. "Quantitative Assessment of Workload and Stressors in Clinical Radiation Oncology." *International Journal of Radiation Oncology Biology Physics* 83 (5): 571-576.

Meister, D. 1976. *Behavioral Foundations of System Development.* New York: Wiley.

Mélan, C., and N. Cascino. 2014. "A Multidisciplinary Approach of Workload Assessment in Real-Job Situations: Investigation in the Field of Aerospace Activities." *Frontiers in Psychology* 5: 964.

Mélan, C., N. Cascino, B. Barthe, and E. Galy. 2012. "Mesurer la Charge de Travail: Une Approche Pluridisciplinaire." In *Risques du Travail, la Santé Négociée*, edited by C. Courtet and M. Gollac, 189-204. Paris: Editions la Découverte.

Mélan, C., E. Galy, and M. Cariou, M. 2007. "Mnemonic Processing in Air Traffic Controllers (ATCs): Effects of Task Parameters and Work Organization." *International Journal of Aviation Psychology* 17 (4): 391-409.

Miyake, S. 2001. "Multivariate Workload Evaluation Combining Physiological and Subjective Measures." *International Journal of Psychophysiology* 40: 233-238.

Paas, F. G. W. C., A. Renkl, and J. Sweller. 2003. "Cognitive Load Theory and Instructional Design: Recent Developments." *Educational Psychologist* 38(1): 1-4.

Paxion, J., E. Galy, and C. Berthelon. 2015. "Overload Depending on Driving Experience and Situation Complexity: Which Strategies Faced with a Pedestrian Crossing?" *Applied Ergonomics* 51: 343-349.

Philip, P., P. Sagaspe, N. Moore, J. Taillard, A. Charles, C. Guilleminault, and B. Bioulac. 2005. "Fatigue, Sleep Restriction and Driving Performance." *Accident Analysis and Prevention* 37(3): 473-478.

Pouliquen-Lardy L., I. Milleville-Pennel, F. Guillaume and F. Mars. 2016. "Remote collaboration in virtual reality: asymmetrical effect of task distribution on spatial processing and mental workload." *Virtual reality* 20: 213-220.

Prichard, J. S., R. J. Stratford, and L. A. Bizo. 2011. "Evaluating the Effects of Team-Skills Training on Subjective Workload." *Learning and Instruction* 21 (3): 429–440.

Reid, G. B., and T. E. Nygren. 1988. "The subjective workload assessment technique: A scaling procedure for measuring mental workload." In *Human Mental Workload*, edited by P. A. Hancock and N. Meshkati, 185-218. Amsterdam: Elsevier.

Schnotz, W., and C. Kürschner. 2007. "A Reconsideration of Cognitive Load Theory." *Educational Psychology Review* 19: 469-508.

Schlegel, R. E. (1993). "Driver mental workload." In *Automotive Ergonomics,* edited by B. Peacock, and W. karwowski, 359-382. London: Taylor & Francis.

Sweller, J. 1988. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science* 12: 257-285.

Sweller, J. 1994). "Cognitive Load Theory, Learning Difficulty and Instructional Design." *Learning and Instruction* 4: 295-312.

Sweller, J., J. J. G. van Merriënboer, and F. G. W. C. Paas. 1998. "Cognitive Architecture and Instructional Design." *Educational Psychology Review* 10: 251-296.

Thayer, R. E. 1986. "The Activation-Deactivation Adjective Check List: Current Overview and Structural Analysis." *Psychological Reports* 58: 607-614.

Ucelli, M., C. Mélan, M. Cariou, N. Cascino, and E. Galy. 2011. "Contrôle de Satellite: Relations entre Activité de Travail et Perception de la Charge de Travail en Horaires Atypiques." Paper presented at the Congrès de la Société Française de Psychologie, Metz, September 7-9.

van Gog, T., F. G. W. C. Paas, and J. J. G. Van Merriënboer. (2005). "Uncovering Expertise-Related Differences in Troubleshooting Performance: Combining Eye Movement and Concurrent Verbal Protocol Data." *Applied Cognitive Psychology* 19 (2): 205-221.

Verwey, W. B. 2000. "On-Line Driver Workload Estimation. Effects of Road Situation and Age on Secondary Task Measures." *Ergonomics* 43 (2): 187–209.

Vessey, I. 1994. "The Effect of Information Presentation on Decision Making: A Cost-Benefit Analysis." *Information and Management* 27: 103–119.

Wickens, C. D. 2002. "Multiple resources and performance prediction." *Theoretical Issues in Ergonomics Science* 3: 159-177.

Young, M. S., K. A. Brookhuis, C. D. Wickens, and P. A. Hancock. 2015. "State of science: mental workload in ergonomics." *Ergonomics* 58 (1): 1-17.

Young, J. Q., R. M. Wachter, O. ten Cate, P. S. O'Sullivan, and D. M. Irby. 2016. "Advancing the next generation of handover research and practice with cognitive load theory." *BMJ Quality and Safety* 25: 66-70.

Young, M. S., and N. A. Stanton. 2005. "Mental workload." In *Handbook of Human factors and Ergonomics Methods,* edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. W. Hendrick. Chap. 39. London: Taylor & Francis.

Zheng, Y., T. Yin, D. Dong, and S. Fu. 2011. "Using NASA-TLX to Evaluate the Flight Deck Design in Design Phase of Aircraft." *Procedia Engineering* 17: 77-83.