Journal of the Ergonomics Society of Korea
Vol. 32, No. 2 pp.167-177, April 2013 http://dx.doi.org/10.5143/JESK.2013.32.2.167

A Comparison of Subjective Mental Workload Measures in Driving Contexts

Ji Yeon Kim, Yong Gu Ji

Department of Information & Industrial Engineering, Yonsei University, Seoul, 120-749

ABSTRACT

Objective: This study aims to compare the usefulness of subjective measures which are comprised of existing methods like NASA-TLX, Bedford-scale and ZEIS and newly developed method like DALI in measuring drivers' mental workload in terms of validity, sensitivity and diagnosticity. **Background:** Nowadays, with the development of intelligent vehicle and HMI, mental workload of driver has become more and more important. For this reason, the studies on drivers' mental workload about driving situation and the use of information technology equipment such as mobile phones and navigations were conducted intensively. However, the studies on measuring drivers' mental workload were rarely conducted. Moreover, most of studies on comparison of subjective measures were used with performance based measure. However, performance based measures can cause distraction effect with subjective measures. **Method:** Participants (N=19) were engaged in a driving simulation experiment in 2 driving contexts (downtown driving and highway driving context). The experiment has 2 sessions according to driving contexts. The level of difficulties by driving contexts were adjusted according to existence of intersections, traffic signs and signals, billboards and the number of doublings. Moreover, as criteria of concurrent validity and sensitivity, the EEG data were recorded before and during the sessions. **Results:** The results indicated that all subjective methods were correlates with EEG in high-way driving. On the contrary to this, in downtown driving, all subjective methods were not correlates with EEG. In terms of sensitivity, multi-dimensional scales (NASA-TLX, DALI) were the only ones to identify differences between high way and downtown driving. Finally, in terms of diagnosticity, DALI was the most suitable method for evaluating drivers' mental workload in driving context. **Conclusion:** The DALI as newly developed method dedicated to evaluate driver's mental workload was superior in terms of sensitivity and diagnosticity. However, researchers should consider the characteristics of each subjective method synthetically according to research objective by selecting the method in subjective measures. **Application:** The results of this study could be applied to the intelligent vehicle and next generation of HMI design to decrease mental workload of driver and for the development of new subjective method in vehicle domain.

Keywords: Mental workload, Subjective measures, Vehicle domain, ZEIS, Bedford-scale, NASA-TLX, DALI, U-City

1. Introduction

Recently, with the development of in-vehicle devices such as mobile phones and navigation, there is an increase on the amount of information that drivers' receive during driving. The increased amount of information changes the workload of drivers and distracts them (Patten, Kircher,

Corresponding Author: Yong Gu Ji. Department of Information and Industrial Engineering, Yonsei University, Seodaemun-gu, Seoul, Korea, 120-749.
Mobile: +82-10-9058-9810, E-mail: yongguji@yonsei.ac.kr

Östlund, & Nilsson, 2004). According to Verwey (1993), drivers' distraction is the main reason for approximately 30~50 percent of vehicle accidents. The distraction is related to mental workload because distraction can mean competition over mental attention resources (Patten et al., 2004). Even though, driving can be related to various types of workload, mental workload is the most important factor (Miller, 2001). Thus, mental workload is a very important issue in driving in vehicle domain (Pauzie, 2008).

1.1 Mental workload assessment in driving

Methods of measuring mental workload can be divided into three categories: subjective measures, physiological measures, and performance based measures. The subjective measures have an advantage of usefulness in terms of time and expense. Thus, many researchers measure the drivers' mental workload by existing subjective measures like NASA-Task Load Index (NASA-TLX).

Horberry, Anderson, Regan, Triggs, and Brown (2006) measure effects of in-vehicle sources and driving environmental complexity on distraction by using NASA-TLX and objective measures like average speed, deviation from speed limit, and reaction time from risk factors. Matthews, Legg and Charlton (2003) also indicated that "personal hands free cell phone would interfere cognitive demands of driving" (Matthews et al., 2003) by measuring subjective mental workload and performance using NASA-TLX and Modified Rhyme Test (MRT).

In addition, Lei and Roetting (2011) presented feasibility of Electroencephalography (EEG) in drawing mental workload index for driver. They used not only physiological measures (EEG, Electrooculography (EOG)) but also subjective measure (NASA-TLX), Performance based measures (Reaction Time, error rate, mean deviation), and investigated the correlation of EEG parameters with other parameters. However, "mental workload depends on the type of loading task" (Pauzie, 2008). Moreover, "mental workload is task specific and person specific" (Rouse, Edwards, & Hammer, 1993).

Driving is a very dynamic task and it can be affected by drivers' behavior and decision (De Waard, 1996). Thus, several subjective measures were developed to evaluate drivers' mental workload in driving task. These measures were differentiated from existing subjective measures because of specializing in vehicle domain.

The Driving Activity Load Index (DALI) was dedicated to evaluate drivers' mental workload with or without in-vehicle systems in vehicle domain (Pauzie, 2008). As a revised version of NASA-TLX, it is focused on driving task while NASA-TLX is focused on pilot task (Pauzie, 2008). The procedure for DALI is the same as those for NASA-TLX. However, some factors of DALI are different in order to adapt to the driving task (Pauzie, 2008). The factors of DALI are effort of attention, visual demand, auditory demand, temporal demand, interference, and situational stress (Pauzie, 2008).

The PSA-Task Load Index (PSA-TLX) is another method which is dedicated to evaluate mental workload in real road context (Chin et al., 2004). This method considers the drivers' state and perceived driving. PSA-TLX evaluates the impact and compatibility of using In Vehicle Information System (IVIS) and Advanced Driving Assistance System (ADAS) (Chin et al., 2004). It also considers negative factors like stress, and fatigue on safety driving (Chin et al., 2004). The Behavioural Markers of Driver Mental Workload (BMDMW) is the other driving task oriented method, designed for considering behavior pattern of driver and it is comprised of 34 items with 5 point scales (Chin et al., 2004).

However, these three methods have rarely been used in researches which are related to mental workload in vehicle domain. Instead, many researchers mainly used existing subjective mental workload measures like NASA-TLX, Sequential Judgement Scale (ZEIS), Bedford scale with performance based measures.

When researchers measure mental workload by using performance-based measures, they have to keep in mind that performance based measures can be risky because of inaccurate evaluation. Yeh and Wickens (1988) indicated the dissociation between subjective measures and performance. The dissociation between performance and subjective ratings was also found by Horberry, et al. (2006). They found that driving in more complex environment showed low level of average speed than other environments, however, the results of NASA-TLX are not significant in environmental complexity. Cantin, Lavallière, Simoneau, and Teasdale (2009) also found that operators can show similar level of performance even though they have the distinct differentiation of mental workload. They indicated that older drivers have higher level of mental workload than younger drivers.

However, there is little difference of performance between those drivers.

1.2 The comparison among subjective measures

Subjective measures were mainly used with performance based measure when comparing subjective measures. In Europe, as Adaptive Integrated Driver-vehicle interface (AIDE) project, Chin et al. (2004) identified the most suitable method to evaluate IVIS, ADAS and AIDE system in vehicle domain among three methods (DALI, BMDMW, PSA-TLX) which are adapted to driving task based on the results of performance. They conducted three experiments in both simulated and on road environments. They found that DALI and BMDMW were sensitive in both simulated and on road environments, however, PSA-TLX was not sensitive in simulated environment. Thus, in the presented study, DALI was only chosen as newly developed subjective measures by considering experimental environment of this study. Besides, in most previous studies about comparison of subjective measures, the comparison of mental workload measures were based on seven requirements including sensitivity, diagnosticity, selectivity/validity, intrusiveness, reliability, implementation and subject acceptability. These requirements can be used when the researchers identify the suitability of the procedures to evaluate mental workload (Eggemeier, Wilson, Kramer & Damos, 1991).

Hill, Iavecchia, Alvah C. Bittner et al. (1992) compared the four subjective mental workload scales (MCH, NASA-TLX, OW, and the subjective workload assessment technique) in terms of sensitivity, operator acceptance, resource requirements, and special procedures. Rubio, Díaz, Martín, and Puente (2004) evaluates three multi-dimensional methods (NASA-TLX, Subjective Workload Assessment Technique (SWAT), Workload Profile (WP) along with six dimensions as sensitivity, diagnosticity, validity, intrusiveness, implementation requirements, and operator acceptability. In evaluating intrusiveness and validity, they also used the performance based measures.

Thus, this study used the physiological measures which are regarded as the most exact measures to evaluate mental workload in driving instead of performance based measures to prevent dissociations. Moreover, we also used the seven requirements like previous studies and especially, we focused on validity, sensitivity and diagnosticity.

1.3 Research objectives

In this study, we have two research objectives below:

(1) Comparing the usefulness between existing subjective mental workload measures and newly developed measures in terms of validity, sensitivity and diagnosticity.

(2) Suggesting the most suitable subjective method for a particular driving context.

To achieve these two research objectives, we measured the drivers' mental workload in two different driving context (highway, downtown) using subjective measures and EEG which is well-proven and reliable method (Doyle, 2007) instead of performance based measures. After, subjective measures were analyzed by comparing each with others in terms of sensitivity, concurrent/convergent validity, and diagnosticity. Especially, EEG data was used to compare subjective measures in terms of sensitivity and concurrent validity except for convergent validity and diagnosticity.

2. Method

2.1 Participants and apparatus

Nineteen drivers (11 males, 8 females) participated, and each driver conducted the experiment individually. Their ages were ranged from 22 to 31 years (mean = 27.05). All participants had a driver's license, the average of driving experience was 3.2 years and all participants were reported being free of neurological/psychiatric diseases. Each participant was prohibited from smoking, having alcohol and caffeine before the experiment and, after completing the experiment, they received a cash payment for their participation.

The experiment was conducted using a game simulator. The simulator was composed of steering wheel and pedals from Logitech, a display, a play-station 3, and the driving game software Test Drive unlimited 2 from EDEN GAMES. Test drive unlimited 2 reproduces a real road environment of 2,400km in Hawaii. EEG Data was recorded by using a Poly G-I from LAXHA, and the recorded data were analyzed by Telescan software from LAXHA and the evaluated subjective mental workloads were analyzed by SPSS 18.0.

2.2 Experimental design

With-in subject design was implemented and each participant ran the experiment that consisted of two contexts: highway and downtown. In previous studies, drivers showed different level of mental workload according to the driving contexts. Mental workload for driving on straight roads like highway was lower than mental workload when driving in downtown (Cantin et al., 2009; Teasdale, Cantin, Blouin, & Simoneau, 2004). Teasdale et al. (2004) indicated that mental workload causes by driving on intersection, lane change and doubling. Moreover, Horberry et al. (2006) also indicated that driving in more complex context, causes more distraction.

Moreover, Patten, Kircher, Östlund, Nilsson, and Svenson (2006) defined the levels of demands on information processing and vehicle handling as the levels of complexity. They classified that the most complex environment was city center environment, continued by driving on intersections which are regulated by traffic lights, road signs and finally, motorways. Thus, in the presented study, we presented two driving contexts (highway, downtown) as the complexity of driving to adjust the levels of mental workload which are offered to drivers. In highway driving, participants were asked to drive on straight road with no traffic lights, road sign and with only two doublings. However, in downtown driving, participants were asked to drive roads that include 122 intersections, 62 doublings, 42 road signs like a STOP sign, 13 traffic signs and billboards. Moreover, in this context, other traffics were offered more frequently than in highway driving. Thus, we assumed that driving in downtown will yield increased mental workload than driving highway. This assumption was additionally identified by analyzing sensitivity.

The highway environment was approximately 47.5km and downtown environment was 32.5km in length, and both environments took about 30 minutes. Each participant instructed to drive similarly as they would for on-road conditions and to comply with traffic law including speed limit in different contexts. The course directions were provided with arrow and sound before approaching the intersections. The independent variables in experiment were each subjective method (NASA-TLX, DALI, ZEIS, Bedford scale) and the dependent variables were mental workloads which were evaluated by using each subjective method.

2.3 Procedure

Before the experiment, each participant received instruction about the purpose, procedure, and order of task. Participants were provided with the opportunity to practice the driving task as much as they wanted. In practice, each participant modulated the sensitivity of steering wheel and pedals. After the practice, electrodes of 8 channels were attached to the scalp of each participant.

The experiment was composed of two sessions according to driving contexts. In each session, participants were given a rest for 5 minutes, and the EEG reference data was recorded for 3 minutes before conducting driving task. The main driving task lasted an average of 30 minutes each. The rest break between each session was provided to participants, for about 30 minutes. Before conducting reference and driving task, participants conducted the weighting process for NASA-TLX and DALI. Moreover, after each session, participants were requested to have an in-depth interview to identify factors which were affecting on subjective feeling about stress.

2.4 EEG measurement and analysis

EEG was recorded using electrodes placed at P3, P4, C3, C4, O3, O4, T3 and T4 in 10/20 international systems with A2 as reference. The ground electrode was placed in skull behind the left ear and the EEG data from all electrodes were sampled at 512Hz. Figure 1 presents the locations of electrodes which were used in this experiment, in 10~20 international systems.

The measured EEG signal was band-pass filtered at 4~50Hz to eliminate Delta power which is affected by artifacts generated by eye blink, movement of body and equipment. Moreover, the relative power analysis was conducted to extract data for analyzing sensitivity and concurrent validity.

3. Results

3.1 Concurrent validity

Correlation analysis was conducted and Pearson correlation coefficients were computed between EEG and the global scores obtained from four subjective measures.

Results from each driving contexts are shown in Table 1 and 2.

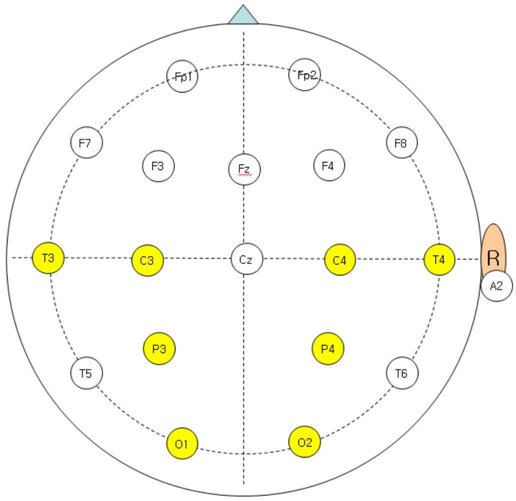

Figure 1. Electrodes locations in international 10~20 system for EEG recordings

Table 1. Pearson correlation coefficients between EEG and subjective measure in highway driving

	P3	P4
ZEIS	0.331	0.548*
Bedford-scale	0.328	0.502*
NASA-TLX	0.483*	0.398
DALI	0.492*	0.473*

*$p < .05$

The concurrent validity was different according to driving contexts. In highway driving (Table 1), which is simple and less complex driving, all coefficients were positive and statistically significant ($p<.05$). Whereas, in downtown driving, which is more complex driving, all correlated coefficients were positive, however not significant (Table 2).

Table 2. Pearson correlation coefficients between EEG and subjective measure in downtown driving

	P3	P4
ZEIS	0.219	0.208
Bedford-scale	0.154	0.203
NASA-TLX	0.167	0.056
DALI	0.284	0.079

That is, in simple and less complex task condition, all subjective measures were effective to evaluate mental workload which was yielded from EEG. However, in complex task condition, all subjective measures were differed from mental workload which was yielded from EEG.

3.2 Convergent validity

Pearson correlation coefficients were also computed among four subjective measures. Table 3 shows the Pearson correlation coefficients in highway driving. Convergent validity was shown according to dimension of measures. Bedford scale was positively correlated with ZEIS which is another uni-dimensional scale and NASA-TLX was also positively correlated with DALI ($p<.01$). Moreover, coefficients between ZEIS as uni-dimensional scale and DALI as multi-dimensional scale was also positive and statistically significant ($p<.05$).

Table 3. Pearson correlation coefficients among subjective measures in highway driving

	ZEIS	Bedford -scale	NASA-TLX	DALI
ZEIS	1.000	0.773**	0.277	0.477*
Bedford-scale	0.773**	1.000	0.275	0.234
NASA-TLX	0.277	0.275	1.000	0.615**
DALI	0.477*	0.234	0.615**	1.000

*$p < .05$; **$p < .01$

However, in downtown driving (Table 4), all coefficients among subjective measures were positive and statistically significant ($p<.01$).

Table 4. Pearson correlation coefficients among subjective measures in downtown driving

	ZEIS	Bedford -scale	NASA-TLX	DALI
ZEIS	1.000	0.819**	0.611**	0.804**
Bedford-scale	0.819**	1.000	0.641**	0.811**
NASA-TLX	0.611**	0.641**	1.000	0.633**
DALI	0.804**	0.811**	0.633**	1.000

**$p < .01$

Therefore, convergent validity appears to be very high for the four subjective measures in more complex driving contexts.

3.3 Sensitivity

Wilcoxon signed ranks test was conducted for EEG data and Paired *t*-test was conducted for global scores obtained from four subjective measures. Many studies of mental workload using EEG indicated that alpha power decreases (Brookings, Wilson, & Swain, 1996; Sterman, Mann, Kaiser, & Suyenobu, 1994; Valentino, Arruda, & Gold, 1993) and theta power increases (Brookings et al., 1996) when the mental workload increases.

The results of Wilcoxon signed ranks test for EEG data, mental workloads were increased after driving when compared to reference both highway and downtown contexts. Theta power was increased and alpha power was decreased significantly (Table 5).

Table 5. The results of Wilcoxon-signed ranks test between reference and driving task

Condition		P3		P4	
		Theta	Alpha	Theta	Alpha
Reference -Highway driving	Z	-3.340	-3.823	-3.099	-3.542
	p	0.001	0.000	0.002	0.000
Reference -Downtown driving	Z	-2.978	-3.743	-2.073	-3.541
	p	0.003	0.000	0.014	0.000

Moreover, alpha power was significantly decreased in downtown driving when compared to highway driving. Thus, downtown driving requires higher level of mental workload than highway driving (Table 6).

Table 6. The results of Wilcoxon-signed ranks test between highway driving and downtown driving

	P3		P4	
	Theta	Alpha	Theta	Alpha
Z	-0.845	-2.052	0.000	-2.093
p	0.398	0.040	1.000	0.036

Based on these results, sensitivity was analyzed by identifying the ability of distinction between three conditions. As the results of paired *t*-test, all subjective measures can show the differentiation of mental workload between reference and highway or downtown driving. However, between highway and downtown driving contexts, only multi-dimensional measures (NASA-TLX, DALI) showed the difference of mental workload (Table 7).

Table 7. The results of paired *t*-test among subjective measures

Measure	Condition	Mean (SD)	*t*	*p*
ZEIS	Reference -Highway driving	-2.368 (3.685)	-2.802	0.012*
	Reference -Downtown driving	-3.184 (4.457)	-3.114	0.006**
	Highway driving -Downtown driving	-0.816 (4.488)	-0.792	0.439
Bedford -scale	Reference -Highway driving	-1.263 (1.881)	-2.927	0.009**
	Reference -Downtown driving	-1.526 (2.118)	-3.141	0.006**
	Highway driving -Downtown driving	-0.263 (2.663)	-0.431	0.672
NASA -TLX	Reference -Highway driving	-9.737 (17.236)	-2.463	0.024*
	Reference -Downtown driving	-20.089 (19.075)	-4.591	0.000**
	Highway driving -Downtown driving	-10.352 (16.987)	-2.656	0.016*
DALI	Reference -Highway driving	-0.827 (0.992)	-3.635	0.002**
	Reference -Downtown driving	-1.658 (1.031)	-1.165	0.000**
	Highway driving -Downtown driving	-0.831 (0.971)	-0.363	0.002**

$^*p < .05$; $^{**}p < .01$

3.4 Diagnosticity

Multiple regression analysis and participant interview were conducted to figure out to what extent mental workload profiles allow discrimination among two contexts. First, the factors which could affect drivers' mental workload were drawn according to characteristics of the task and driving environments. Moreover, the factors which were affecting stresses of participants were drawn by interviewing participants. Next, we identified the relationships between drawn factors and sub-dimensions of NASA-TLX and DALI. Finally, based on these relations, we made a distinction between the results of multiple regression and drawn sub-dimensions which can cause or caused driver's mental workload in two contexts.

The drawn factors which can cause or caused mental workload were negative emotions due to speed adjustment in driving on highway and repeated sound for course direction, decision making, delay of driving time, interference of other traffics, allow and sound for course direction in driving on downtown. The relationships between these factors and sub-dimensions of NASA-TLX and DALI were shown in Table 8.

Table 8. The relationships between the mental workload causing factors and sub-dimensions of NASA-TLX and DALI

Driving environment	Mental workload causing factors	NASA-TLX	DALI
Highway	Negative emotion due to speed adjustment	Frustration	Stress
Downtown	Negative emotion due to repeated sound for course direction	Frustration	Stress
	Decision making	Mental demand	Global attention
	Delay of driving time	Temporal demand	Temporal demand
	Interference of other traffics	–	Interference
	Course direction (Presented allow and sound)	–	Auditory demand Visual demand

In Table 9 and 10, the results of multiple regression analysis on the highway driving were shown. The global score of NASA-TLX was related to physical demand, effort, mental demand, and performance. These sub-dimensions and constant can describe 84.2% of global workload. Among these sub-dimensions, physical demand was the most describable dimension (B=1.279) and effort was also describable similarly (B=1.204). Physical demand, effort and mental demand were statistically significant (p<.01). Performance was also significant (p<.05) but less describable among four drawn factors (Table 9).

Next, the global score of DALI was related to temporal demand, stress, visual demand, auditory demand and interference. These sub-dimensions and constant can describe 91.6% of global workload. The auditory demand was the most describable (B=1.378) and other sub-dimensions except interference were describable similarly. The "interference" was the less describable with 0.795 and all sub-dimensions were significant with p<.01.

Table 9. The results of multiple regression analysis of NASA-TLX in highway driving

Sub -dimension	B	Std E	β	t	R^2
Physical demand	1.279	0.265	0.541	4.826**	0.842
Effort	1.204	0.220	0.606	5.481**	
Mental demand	0.953	0.183	0.593	5.192**	
Performance	0.428	0.177	0.281	2.416*	
Constant	235.298	75.333		3.123**	

F=18.589, Adjusted R^2=0.796
*p < .05; **p < .01

Table 10. The results of multiple regression analysis of DALI in highway driving

Sub -dimension	B	Std E	β	t	R^2
Temporal demand	1.192	0.253	0.513	4.704**	0.916
Stress	1.158	0.183	0.565	6.313**	
Visual demand	1.043	0.211	0.444	4.592**	
Auditory demand	1.378	0.398	0.407	3.461**	
Interference	0.795	0.239	0.286	3.335**	

F=28.210, Adjusted R^2=0.883
**p < .01

Generally, in simple task and less complex environments, both NASA-TLX and DALI were less diagnostic. However, DALI had only related sub-dimension "Stress" due to speed adjustment. Thus, DALI was a little more diagnostic than NASA-TLX in this context.

In downtown driving, the result of NASA-TLX is shown in Table 11. A total of five sub-dimensions were related to global workload and these dimensions and constant took into account about 85% of global workload. The "effort" was best describable with B=0.950. The "mental demand" and "Temporal demand" were describable similarly. Next, "physical demand" was describable. Above mentioned sub-dimensions were all significant with p<.01. However, "Performance" was significant with p<.05 and the less describable (B=0.48).

Table 11. The results of multiple regression analysis of NASA-TLX in downtown driving

Sub -dimension	B	Std E	β	*t*	R^2
Effort	0.950	0.198	0.597	4.800**	0.850
Mental demand	0.784	0.213	0.479	3.690**	
Temporal demand	0.704	0.199	0.431	3.539**	
Physical demand	0.644	0.164	0.412	3.936**	
Performance	0.480	0.202	0.305	2.376*	

F=14.680, Adjusted R^2=0.792
*p < .05; **p < .01

However, as the sub-dimensions of the DALI, "Global attention" which is related to mental activity was the most describable (B=1.516). Next, "Interference" was describable with B=-.794. These 2 sub-dimensions were significant (p<.01). However, Temporal demand was significant with p<.05 and the less describable (B=0.674). Three sub-dimensions and constant could take into account 73.6% of global workload (Table 12).

Table 12. The results of multiple regression analysis of DALI in downtown driving

Sub -dimension	B	Std E	β	t	R^2
Global attention	1.516	0.286	0.712	5.292**	0.736
Interference	0.794	0.232	0.459	3.422**	
Temporal demand	0.674	0.252	0.359	2.671*	
Constant	28.800	5.995		4.804**	

F=13.922, Adjusted R^2=0.683
*p < .05; **p < .01

In more complex task and environment, DALI was more diagnostic because of including main factors which were caused or could cause mental workload of driver. NASA-TLX had main factors but also had unrelated factors like physical demand, effort and performance. However, DALI had only related factors like global attention, interference and temporal demand.

4. Discussion

In this paper, four subjective measures (NASA-TLX, DALI, ZEIS, and Bedford-scale) were compared with respect the following issues: concurrent validity, convergent validity, sensitivity, and diagnosticity.

Validity

As a result of correlation analysis, concurrent and convergent validity were different according to driving contexts (Figure 2). In simple task and environment like highway driving, subjective measures had a relationship according to dimension. As convergent validity, uni-dimensional scale (ZEIS) was related to another uni-dimensional scale (Bedford scale) and multi-dimensional scale (NASA-TLX) was also related to same-dimensional scale (DALI). DALI was also related to different dimensional scale (ZEIS). Moreover, all subjective measures were significantly related to EEG as concurrent validity.

On the other hand, in more complex task and environment like downtown driving, subjective measures were related to each other regardless of dimension. However, these measures were not related to EEG.

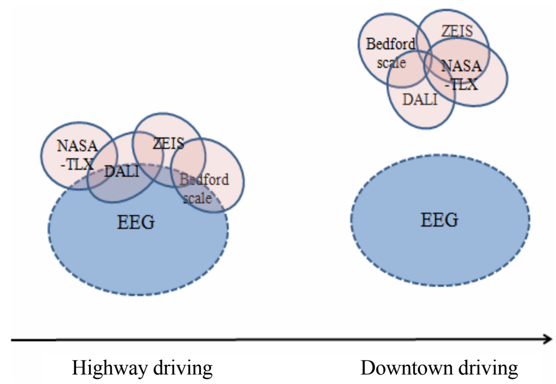

Figure 2. Validity of subjective measures in highway and downtown driving

Sensitivity

Multi-dimensional scales (NASA-TLX, DALI) were only ones to identify differences between highway and downtown driving. Thus, multi-dimensional scales including DALI were more superior to uni-dimensional scale in terms of sensitivity.

Diagnosticity

This index was identified that the sub-dimensions which were drawn from multi-dimensional analysis were discriminated whether these dimensions were related to cause mental workload in experiment. As a result, generally, in low complex task and environment like highway driving, both subjective measures were little diagnostic. However, in high task and environment like downtown driving, DALI was slightly superior to NASA-TLX by including main sub-dimensions except unnecessary sub-dimensions. That is, sub-dimensions of DALI were more suitable than NASA-TLX to reflect the factors which are yieldable mental workload on driving in vehicle domain.

In summary, the present results demonstrate that, according to task and environment complexity, there was more or less usefulness of subjective measures. However, DALI was the superior in both contexts.

However, Researchers have to keep in mind that DALI isn't going to be a perfect subjective measure in vehicle domain. The DALI had little concurrent validity in the more complex task and environment like downtown driving. In addition, in terms of convergent validity, DALI had only relationship with ZEIS as uni-dimensional scale, but not Bedford scale.

Thus, for more precise evaluating of mental workload, physiological measures like EEG, EOG, and ECG have to be used simultaneously with Subjective measures. Otherwise, these are needs of developing a new and the most suitable subjective measure in vehicle domain.

5. Conclusion

The main goal of this study was to compare the properties of four subjective measures (two uni-dimensional scale and two multi-dimensional scale; three rating scales which early developed and could use in vehicle domain (NASA-TLX, ZEIS, Bedford scale) and newly developed scale which is oriented to vehicle domain (DALI) to identify the most suitable subjective measure for vehicle domain among these measures.

The results of this study show that DALI was the superior to other subjective measures which were early developed for aviation domain and could be used in vehicle domain in terms of validity, sensitivity and diagnosticity. However, DALI, in common with other subjective measures used in this study, had low level of concurrent validity in downtown driving. That is, if researchers only use the subjective measures, they will be unable to measure actual mental workload of driver. Thus, researchers have to use the other measures like physiological measures or develop new subjective measures specialized in vehicle domain.

In this study, the driving game simulator was used instead of driving simulator. In highway driving environment, the roads were inflective, thus, participants felt difficulty for speed adjustment. These negative feelings could affect to the results of subjective measures of mental workload. In the previous study, Ryu and Myung (2005) indicated that the mental workload which was evaluated by NASA-TLX was increased as speed increases. Thus, participants were able to evaluate the subjective mental workload high in highway driving, and this result can lead to low level of concurrent validity

Although, the simulator conditions were limited, this study contributes considerations and guidance for the selection of subjective measures in vehicle domain. In vehicle driving task, researchers can select the subjective measures according to requirements and purpose of the study.

Moreover, this study is quite distinct from other comparison research of subjective measures. Because, this study choose the subjective measures which can be used in specific domain as vehicle domain. Moreover, in terms of the comparison of concurrent validity and sensitivity, EEG which was the most well proven measure, was used, instead of performance based measures.

Moreover, we ultimately expect that the results of this study could be applied to design of intelligent vehicle and next generation of HMI to decrease mental workload of drivers in various environments such as U-city, Smart City, and Intelligent Road and development of another new subjective measure in vehicle domain. For this, the conclusions of this study need to be verified in more realistic driving situations through field study and further research like context analysis are required.

Acknowledgements

This work is financially supported by Korea Minister of Ministry of Land, Transport and Maritime Affairs (MLTM) as U-City Master and Doctor Course Grant Program.

References

Brookings, J.B., Wilson, G.F. & Swain, C.R., Psychophysiological responses to changes in workload during simulated air traffic control, *Biological Psychology*, 42(3), 361-377, 1996.

Cantin, V., Lavallière, M., Simoneau, M. & Teasdale, N., Mental workload when driving in a simulator: Effects of age and driving complexity, *Accident Analysis & Prevention*, 41(4), 763-771, 2009.

Chin, E., Nathan, F., Pauzie, A., Manzano, J., Nodari, E., Cherri, C, et al. Subjective Assessment methods for Workload. http://www.aide-eu.org/pdf/sp2_deliv_new/aide_d2_2_6.pdf, 2004.

De Waard, D., The measurement of Driver's Mental workload. *Ph.D. Thesis. University of Groningen. Traffic research center. Haren*, The Netherlands, 1996.

Doyle, M.J., Psychophysiological correlates of cognitive workload during a satellite management decision-training task. *In proceedings of Human Systems Integration Symposium,* Annapolis, Maryland, http://navalengineers.net/Proceedings/HSIS2009/Index.html, 2007.

Eggemeier, F.T., Wilson, G.F., Kramer, A.F. & Damos, DL., General considerations concerning workload assessment in multi-task environments. Taylor & Francis, London, 1991.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Byers, J.C., Zaklade, A.L. & Christ, R.E., Comparison of Four Subjective Workload Rating Scales. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 34(4), 429-439, 1992.

Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J. & Brown, J., Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance, *Accident Analysis & Prevention*, 38(1), 185-191, 2006.

Lei, S. & Roetting, M., Influence of Task Combination on EEG Spectrum Modulation for Driver Workload Estimation, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(2), 168 -179, 2011.

Matthews, R., Legg, S. & Charlton, S., The effect of cell phone type on drivers subjective workload during concurrent driving and conversing, *Accident Analysis & Prevention*, 35(4), 451-457, 2003.

Miller, S., LITERATURE REVEIW Workload Measures. http://www.nads-sc.uiowa.edu/publicationStorage/200501251347060.N01-006.pdf, 2001.

Patten, C.J.D., Kircher, A., Östlund, J. & Nilsson, L., Using mobile telephones: cognitive workload and attention resource allocation, *Accident Analysis & Prevention*, 36(3), 341-350, 2004.

Patten, C.J.D., Kircher, A., Östlund, J., Nilsson, L. & Svenson, O., Driver experience and cognitive workload in different traffic environments, *Accident Analysis & Prevention*, 38(5), 887-894, 2006.

Pauzie, A., A method to assess the driver mental workload: The driving activity load index (DALI), *IET Intelligent Transport Systems*, 2(4), 315-322, 2008.

Rouse, W.B., Edwards, S.L. & Hammer, J.M., Modeling the dynamics of mental workload and human performance in complex systems, *Systems, Man and Cybernetics*, 23(6), 1662-1671, 1993.

Rubio, S., Díaz, E., Martín, J. & Puente, J.M., Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods, *Applied Psychology*, 53(1), 61-86, 2004.

Ryu, K. & Myung, R., Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic, *International Journal of Industrial Ergonomics*, 35(11), 991-1009, 2005.

Sterman, M.B., Mann, C.A., Kaiser, D.A. & Suyenobu, B.Y., Multiband topographic EEG analysis of a simulated visuomotor aviation task, *International Journal of Psychophysiology*, 16(1), 49-56, 1994.

Teasdale, N., Cantin, V., Blouin, J. & Simoneau, M., Attentional demands while driving in a simulator: effects of driving straights on open roads, approaching intersections and doubling maneuvers, *Advances in Transportation Studies an international Journal*, *2004 Special issue*, 75-84, 2004.

Valentino, D.A., Arruda, J.E. & Gold, S.M., Comparison of QEEG and response accuracy in good vs poorer performers during a vigilance task, *International Journal of Psychophysiology*, 15(2), 123-133, 1993.

Verwey, W.B., How can we prevent overload of the driver?, in: Parkes, A.M., S.Franzen. (Eds.), Driving future vehicles. Taylor&Francis, London, pp. 235-244, 1993.

Yeh, Y.-Y. & Wickens, C.D., Dissociation of Performance and Subjective Measures of Workload, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 30(1), 111-120, 1988.

Author listings

Ji Yeon Kim: jykim29@yonsei.ac.kr

Highest degree: Master's degree, Department of Information and Industrial Engineering, Yonsei University

Position title: GM Korea

Areas of interest: Human Factors in Vehicle, Mental Workload, HCI

Young Gu Ji: yongguji@yonsei.ac.kr

Highest degree: PhD, Industrial Engineering, Purdue University
Position title: Associate Professor, Department of Information and Industrial Engineering, Yonsei University
Areas of interest: Usability, Emotional Design, Accessibility, and Elderly in HCI

Date Received : 2012-09-23
Date Revised : 2013-03-11
Date Accepted : 2013-03-11