

## Validity and Reliability Testing on Train Driver Performance Model Using A PLS Approach

Jalil Azlis-Sani<sup>1,2 a</sup>, SitiZawiahMd Dawal<sup>2,b</sup> and NorhayatiMohmadZakuan<sup>3,c</sup>

<sup>1</sup>Faculty of Mechanical & Manufacturing Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), Parit Raja, Batu Pahat, Johor, Malaysia

<sup>2</sup>Centre for Product Design & Manufacturing (CPDM), Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia

<sup>3</sup>Faculty of Management and Human Resource Development, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia

<sup>a</sup>azlis@uthm.edu.my, <sup>b</sup>sitizawiahmd@um.edu.my, <sup>c</sup>norhayatimz@utm.my

**Keywords:** human performance, structural equation modeling, reliability, validity, railway

**Abstract.** The purpose of this paper is to explain the assessment of validity and reliability of measured items used in survey research. The structural equation modeling techniques and Partial Least Square (PLS) will be adopted to assess the goodness of measures of constructs used in a model to examine the performance of the Malaysian train drivers. The measurement process involves assessment of construct validity of the items and followed by convergent validity. Then, the composite reliability was assessed with internal consistency measure of Cronbach's alpha. Lastly, discriminant validity was tested to assess the validity of the measurement. The constructs are not supposed to measure other constructs or overlapping constructs.

### Introduction

**Human performance.** Performances are related to the job or task assigned to the operators unless it is conducted fully automatic by machines or robots. But, most of industries still depend on their human operators including railway, where locomotives are still operated by the train drivers.

Previously, human factors was left behind and be considered as independent factors with very less contribution to the overall system performance[1]. Researchers conducted their studies on individual variable to measure performance, factors affecting performance and effect of performance; either on individual level or organizational level. Most of them relate one or two variables or factors affecting performance with the effect or consequences of the level of performance.

To date, investigations on human performance factors were conducted and influences the overall performance measurement and study ranging from the air traffic control system, design process at the factories, train driving activities and ship navigation [2-5]. Awareness on importance of human performance studies are increasing and consideration on human performance factors come earlier during design stage.

Therefore, the development of human performance will enable the consideration of potentially conflicting task demands in a systematic and structured ways as the earliest stages in a design [6].

This study was conducted through survey research method, where it involves the usage of questionnaires and scales. Therefore, this paper will look at the goodness of measures assessment used in the analysis by looking at the validity and reliability. The structural equation modeling techniques and Partial Least Square (PLS) will be adopted to assess the goodness of measures of constructs used in a model to examine the performance of the Malaysian train drivers.

### Research Context and Research Model

This paper includes a big part of a research which generally aims to study human performance of train driver in Malaysia.

**Factors affecting human.** Stress is a descriptive term to describe feeling of pressure and strain, in psychological point of view. This term also applies for biological sciences which includes stress on physiology or biological nature which the physical respond to the stressors such as extreme environments; but are not limited to temperature, noise, diseases, sleepiness and emotions [7].

Occupational stress describes stress caused by working and job environment. Workers feel the pressure and strain can be because of the environmental factors. Poor job design, layout setting and high workload are highly potential stressors. These stressors contribute to the increasing level of stress among workers. The worker interacts with the work environment when conducting their tasks or job. Interaction between these two elements (worker – work environment) requires proper planning; that is why job design is used to identify the suitable conditions for the worker to work in best condition and environment [8]. The interaction between workers – work environment is critical especially in transportation industry. In fact, in an earlier subtopic of literatures (subtopic number), all three researchers [1, 2, 9] indicate the interaction between human (worker) and the environment. However, failure to meet this general requirement usually creates discomfort environment and lead to stress of the worker. In this particular problem, it leads to occupational stress.

**Workers activities.** Study on Fatigue and Shift Work in UK train drivers by Stone B [10] revealed that the duration of the shift length is a key factor leading to fatigue. The research indicates that accident rates in shift workers are 25% higher on 12-hour compared with eight-hour shifts. It is therefore important to restrict the amount of overtime taken at the end of a shift. Moreover, the impacts of long duty periods on fatigue are likely to be most severe on night and early morning shifts. Continuous driving time in particular, long periods of continuous duty without a break are particularly fatiguing. It is therefore important to ensure that rest breaks at appropriate times are taken in order to reduce risk. Ideally, breaks should be of at least 15 minutes duration and free from any work-related activities.

The discussion of influential factors support the development of research model to identify relationship between human performance, workers activities, human factors and the working context of the train drivers in Malaysia. Goodness of measure assessment of the constructs will be discussed in the next section in terms of validity and reliability of the research model.

## Research Method

The survey was conducted among train drivers of the train operating company (TOC) in Malaysia. The self-administered survey was voluntarily completed by the locomotive drivers and junior drivers. Surveys were conducted in 5 different railway depots (Pera, Ipoh, Kuala Lumpur, Gemas and Kuala Lipis) across Peninsular Malaysia. Each respondent who was off duty driver from different depots was picked randomly; received the questionnaire and filled it immediately. Researcher was always at the survey location for any assistance during the answering session. There were 229 total responses returned.

## Measures and Assessment of Goodness of Measures

An adapted questionnaire using 5-point Likert type scale was used to collect data from the respondents for each of constructs. Adaptation of questionnaire was taken from previous literatures namely Ryan Wilson et. al (2009) [11], Strahan, Watson et al. (2008) [12], Johns (1991, 1992, 1993) [13-15], Johns and Hoaking (1997) [16], Gradisar, Leon Lack et al. (2007) [17], Austin and Drummond (1986) [18] and Dawal (2005) [19].

**Goodness of Measures.** PLS analysis; influenced by CBSEM approach suggested to present results in two steps [20, 21]. The first step was to analyse the measurement model by examining validity and reliability of the survey items.

**Construct validity.** To measure construct validity, first by determining individual reliability of the items to their respective constructs. Items (or composite items) were assessed and cut-off value of 0.5 was used as significant [22]. As such, any item below cut-off value will be deleted. Item with lowest loading in a particular construct or latent variable (LV) was deleted and the cross-

loadings were computed. Table 1 shows the final result of composite items loading. This cross-loadings assessment was to measure how the relationship of each item to other constructs. Loading of the item on its particular construct should be higher, but not to the other constructs [21]. Therefore, construct validity was confirmed by observing all items, where all items with high loading were measuring a particular construct but loaded lower on other constructs [23]. Loadings with bold marking were measured the particular construct while others were cross-loadings.

Table 1: Result of construct validity

Items	Activity	Human	Performance	Safety	Work environment
DT1	<b>0.714</b>	0.395	0.117	-0.113	0.116
DT2	<b>0.767</b>	0.496	0.109	-0.174	0.165
DT3	<b>0.798</b>	0.354	0.138	-0.163	0.090
FF	0.389	<b>0.754</b>	0.160	-0.286	0.272
JRTa	0.418	<b>0.745</b>	0.107	-0.324	0.359
JRTb	0.414	<b>0.674</b>	0.094	-0.283	0.270
JRTc	0.420	<b>0.762</b>	0.160	-0.324	0.383
STRb	0.303	<b>0.641</b>	0.093	-0.258	0.219
STRd	0.367	<b>0.705</b>	0.099	-0.263	0.224
STRe	0.366	<b>0.710</b>	0.042	-0.335	0.244
PMS2008	0.091	0.130	<b>0.727</b>	-0.160	0.096
PMS2009	0.182	0.217	<b>0.828</b>	-0.173	0.162
PMS2010	0.110	0.096	<b>0.733</b>	-0.128	0.106
PMS2011	0.077	0.027	<b>0.698</b>	-0.142	0.154
SC05	-0.177	-0.433	-0.180	<b>0.799</b>	-0.339
SC08	-0.115	-0.209	-0.164	<b>0.802</b>	-0.241
SC10	-0.169	-0.289	-0.111	<b>0.700</b>	-0.219
WC5	0.129	0.324	0.172	-0.309	<b>0.737</b>
WC6	0.129	0.216	0.125	-0.219	<b>0.664</b>
WC7	0.134	0.308	0.146	-0.204	<b>0.719</b>

Table 2: Result of measurement model

Model construct	Measurement item	Loading	CR	AVE	Cronbach's Alpha
Activity	DT1	0.714	0.804	0.578	0.637
	DT2	0.767			
	DT3	0.798			
Human	FF	0.754	0.879	0.509	0.846
	JRTa	0.745			
	JRTb	0.674			
	JRTc	0.762			
	STRb	0.641			
	STRd	0.705			
	STRe	0.710			
Performance	PMS2008	0.727	0.836	0.561	0.739
	PMS2009	0.828			
	PMS2010	0.733			
	PMS2011	0.698			
Safety	SC05	0.799	0.812	0.590	0.661
	SC08	0.802			
	SC10	0.700			
Work environment	WC5	0.737	0.750	0.501	0.508
	WC6	0.664			
	WC7	0.719			

**Convergent validity.** Convergent validity were assessed through factor loadings, composite reliability and average variance extracted (AVE) [22]. Loading of the items should more than 0.5 with construct validity (CR) value of 0.7 [22]. The average variance extracted (AVE) proposed by Farnell and Lacker (1981) was to measure variance amount by the indicators relative to measurement error. AVE should more than 0.5[21]. Table 2 presents the results of the measurement

model. It shows that all constructs activity, human, safety, work environment and performance are more than cut-off value of loadings, composite reliability (CR), average variance extracted (AVE) and Cronbachs alpha. The results indicate that all the constructs met the recommended values and thus are reliable and valid.

**Discriminant validity.** Discriminant validity was tested to assess the validity of the measurement. The constructs are not supposed to measure other constructs or overlapping constructs. The cross-construct correlations should be very low but it measures strongly the construct it attempts to reflect [21]. As shown in Table 3, the average variance extracted (AVE) of the construct is much higher than the squared correlations for each constructs. This indicates adequate discriminant validity. Therefore, the measurement model demonstrated adequate convergent validity and discriminant validity.

Table 3: Discriminant validity of construct

	Activity	Human	Performance	Safety	Work environment
Activity	<b>0.578</b>				
Human	0.289	<b>0.509</b>			
Performance	0.025	0.027	<b>0.561</b>		
Safety	0.039	0.167	0.041	<b>0.590</b>	
Work environment	0.034	0.164	0.045	0.123	<b>0.501</b>

\*Diagonals (in bold) represent the AVE while the off diagonals represent the squared correlations

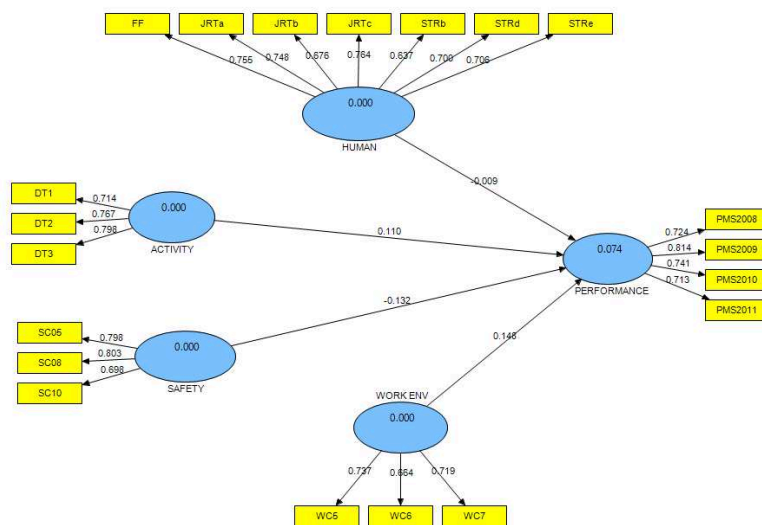


Fig. 1: Research model

## Discussion

This paper was to explain how PLS could be used to assess goodness of measure by looking at the validity and reliability of the measure. Construct validity was measured which consisted of convergent and discriminant validity. At the same time, Cronbach's alpha values and composite reliability were identified to assess the reliability of the measures. As such, the measures were shown to be reliable.

This paper could benefit other researchers how to conduct and assess validity and reliability of measures in survey research using SEM & PLS approaches.

## Acknowledgement

The main author is supported by the Office of Research, Innovation, Commercialization and Consultation (ORICC) Universiti Tun Hussein Onn Malaysia. We also would like to thank Department of Safety, Health and Environment, Keretapi Tanah Melayu Berhad (KTMB) on their contribution and participation in this study.

## References

- [1] Chang, Y.H. and C.H. Yeh, *Human performance interfaces in air traffic control*. Applied Ergonomics, 2010. **41**(1): p. 123-129.
- [2] Baines, T.S., et al., *Towards a theoretical framework for human performance modelling within manufacturing systems design*. Simulation Modelling Practice and Theory, 2005. **13**(6): p. 486-504.
- [3] Gore, B.F. and J.D. Smith, *Risk assessment and human performance modelling: the need for an integrated systems approach*. International Journal of Human Factors Modelling and Simulation, 2006. **1**(1): p. 119-139.
- [4] Gould, K.S., et al., *Effects of navigation method on workload and performance in simulated high-speed ship navigation*. Applied Ergonomics, 2009. **40**(1): p. 103-114.
- [5] Mason, S., et al., *Improving the Design Process for Factories: Modeling Human Performance Variation*. Journal of Manufacturing Systems, 2005. **24**(1): p. 47 - 54.
- [6] Clarke, T., *The Ergonomics Programme at Network Rail*, in *Rail Human Factors: Supporting the Integrated Railway*, J.R. Wilson, et al., Editors. 2005, Ashgate. p. 19 - 21.
- [7] Bourne, L.a.Y., *RA Stress and cognition: A cognitive psychological perspective*, in *Final Report National Aeronautics and Space Administration Grant Number NAG2-15612003*.
- [8] Genaidy, A., et al., *The work compatibility improvement framework: An integrated perspective of the human-at-work system*. Ergonomics, 2007. **50**(1): p. 3-25.
- [9] Bailey, R.W., *Human performance engineering : designing high quality, professional user interfaces for computer products, applications, and systems*. 3rd ed 1996, New Jersey: Prentice Hall.
- [10] Stone B, A.M., Mick Spencer, Claire Turner & Ann Mills, *Fatigue and shiftwork in UK train drivers*, 2005, QinetiQ Centre for Human Sciences, Farnborough, United Kingdom and Railway Safety and Standards Board London, United Kingdom.
- [11] Ryan, B., et al., *Developing a Rail Ergonomics Questionnaire (REQUEST)*. Applied Ergonomics, 2009. **40**(2): p. 216-229.
- [12] Strahan, C., B. Watson, and A. Lennonb, *Can organisational safety climate and occupational stress predict work-related driver fatigue?* Transportation Research Part F: Traffic Psychology and Behaviour, 2008. **11**(6): p. 418-426.
- [13] Johns, M.W., *A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale*. Sleep, 1991. **14**(6): p. 540-545.
- [14] Johns, M.W., *Reliability and Factor Analysis of the Epworth Sleepiness Scale*. Sleep, 1992. **15**(4): p. 376-381.
- [15] Johns, M.W., *Daytime Sleepiness, Snoring and Obstructive Sleep Apnea: The Epworth Sleepiness Scale*. Chest, 1993. **103**: p. 30-36.
- [16] Johns, M.W. and B. Hoaking, *Excessive Daytime Sleepiness: Daytime Sleepiness and Sleep Habits of Australian Workers*. Sleep, 1997. **20**(10): p. 844-849.
- [17] Gradisar, M., et al., *The Flinders Fatigue Scale: Preliminary Psychometric Properties and Clinical Sensitivity of a New Scale for Measuring Daytime Fatigue associated with Insomnia*. Journal of Clinical Sleep Medicine, 2007. **3**(7): p. 722-728.
- [18] Austin, A. and P.D. Drummond, *Work problems associated with suburban train driving*. Applied Ergonomics, 1986. **17**(2): p. 111-116.

- [19] Dawal, S.Z., Z. Taha, and Z. Ismail, *Effect of job organization on job satisfaction among shop floor employees in automotive industries in Malaysia*. International Journal of Industrial Ergonomics, 2009. **39**(1): p. 1-6.
- [20] Anderson, J.C. and D.W. Gerbing, *Predicting the Performance of Measures in a Confirmatory Factor Analysis With a Pretest Assessment of Their Substantive Validities*. Journal of Applied Psychology, 1991. **Vol. 76**(No. 5): p. 732 - 740.
- [21] Chin, W.W., *How to Write Up and Report PLS Analyses: Handbook of Partial Least Squares*, V. Esposito Vinzi, et al., Editors. 2010, Springer Berlin Heidelberg. p. 655-690.
- [22] Hair, J.F., et al., *Multivariate Data Analysis*. 7th ed2010: Prentice Hall.
- [23] Ramayah, T., J. Lee, and J. In, *Network collaboration and performance in the tourism sector*. Service Business, 2011. **5**(4): p. 411-428.