

Theoretical Versus Mathematical Approach to Modeling Psychological and Physiological Data

Lauren Reinerman-Jones¹(✉), Stephanie J. Lackey², Julian Abich IV¹,
Brandon Sollins¹, and Irwin Hudson³

¹ Institute for Simulation and Training (IST),
University of Central Florida (UCF), Orlando, FL, USA
{lreinerm, jabich}@ist.ucf.edu

² Design Interactive, Orlando, FL, USA
stephanie.lackey@designinteractive.net

³ U.S. Army Research Laboratory, Orlando, FL, USA
irwin.hudson@us.army.mil

Abstract. Variable selection for predictive modeling has traditionally relied on theory in the psychological domain. Given the recent advancements in computing technology and availability, researchers are able to utilize more sophisticated mathematical modeling techniques with greater ease. The challenge becomes evaluating whether theory or mathematics should be relied upon for model development. The presented analyses compared the use of hierarchical and stepwise variable selection methods during a predictive modeling task using linear regression. The results show that the stepwise variable selection method is able to obtain a more efficient model than the hierarchical variable selection method. Implications and recommendations for researchers are further discussed.

Keywords: Cognitive modeling · Perception · Emotion · Interaction · Electroencephalography · Brain activity measurement · Physiological measuring · Human performance

1 Introduction

Psychological researchers have utilized predictive statistical model building techniques to predict, classify, and further understand the true nature of a variety of behavioral phenomenon. While building predictive statistical models, researchers are faced with an assortment of options to consider in order to construct the most appropriate model from their data, such as (1) whether the model is being used for prediction or understanding the relationship between the independent and dependent variable(s), (2) the algorithm being used to build the model, (3) the specific parameters of the algorithm being used, (4) possible data transformations, (5) sampling techniques to build the model, as well as (6) evaluation of the generated models.

One essential consideration to model development is determining how variables are selected [1]. Traditionally in psychological research, variable selection has “relied on informal or intuitive reasoning or historical precedent” [2]. This method of variable

selection relies strictly on theory, and is efficient and practical when small amounts of well-studied variables are being evaluated as predictors within a model. With improvements in computing power and advancements in measurement techniques, researchers are now able to collect, store, and analyze a much larger amount of data that often times contains an exhaustive amount of potential predictor variables. For example, researchers utilizing physiological measures, most notably the electroencephalogram (EEG), have access to brain activity data from a large amount of sensor sites (sometimes up to 256) over a large amount of frequency bins (1–100+). Physiological variables have been shown to be strongly task dependent [3] and psychophysiological metrics seldom intercorrelate [4], therefore, it may be difficult to assume a predictor for one task type will be applicable to another. Furthermore, more advanced modeling software packages have become available to researchers that allow complex, state-of-the-art statistical methods to be utilized with relative ease. A variety of open source and proprietary software packages are used by researchers for modeling data including SPSS, SAS, R, Python, and WEKA, to name a few. Lastly, the myriad of experimental environments available, including computer-based questionnaires and simulation-based approaches, allow researchers to explore a new range of variables over a large population with fewer cost and resources. As a result, with the influx of large data sets, new variables being collected, and access to more sophisticated modeling tools, researchers are applying objective, mathematical variable selection methods that can be used to reduce the dimensionality of their datasets, facilitate data understanding, discover new patterns or trends in the data, and ultimately improve model prediction [1].

The data and tasks for these present analyses were derived from previous work [5]. The goal for the present analyses is to evaluate a commonly used algorithm by psychological researchers, linear regression, while using a theoretical approach (i.e. hierarchical regression) and mathematical approach (i.e. stepwise regression) for variable selection to develop and compare performance prediction models using subjective and objective measures of workload.

1.1 Workload Metrics

Research has suggested that mental workload plays an essential role in task performance and is an indicator of performance across multiple domains [6, 7]. Although a universally accepted, formal definition of workload does not exist, workload can be regarded as the “perceived evaluation and accompanying physiological response to the experience imposed by task demands” [5]. A significant body of research has investigated the use of subjective and objective methods to quantify an operator’s level of mental workload. The subjective and objective metrics described below will be used in the present analyses as each one has been found to contribute to the explanation of performance.

In regards to subjective measures, the Instantaneous Self-Assessment (ISA) [8] and NASA-Task Load Index (TLX) [9] have been extensively used by researchers to capture an operator’s subjective level of perceived workload. The ISA is a unidimensional measure that provides an immediate subjective rating of workload during a given task [8]. The ISA has the benefit of being minimally intrusive, is able to be

administered in real-time, and has been shown to be a good indicator of workload [10]. Traditionally, the TLX has been used as a “gold-standard” of workload assessment. The TLX is a multidimensional measure that assesses perceived workload during a given task and usually administered post-task [9]. Operators rate their perceived level of workload on six dimensions: three related to the demands on the operator and three related to the interaction with the task [9]. The original measure additionally required pair-wise comparisons to weight the ratings, but research found the weighting is time-consuming and unnecessary [11]. Additionally, the TLX sensitivity is robust to time delays [11]. Although subjective assessment provides valuable insight regarding the operator’s perceived impact of task demands, access to unbiased and objective data could provide critical information that might account for more variance associated with task performance.

Psychophysiological measures, such as the electroencephalogram (EEG), electrocardiogram (ECG), functional near infrared spectroscopy (fNIR), transcranial Doppler (TCD) ultrasonography, and eye tracking, have been extensively used by researchers to objectively assess workload. Several psychophysiological metrics have been identified in the literature to be sensitive to workload variation during task performance [4]. EEG monitors electrical activity in the cerebral cortex. Research found decreased parietal alpha activity [12] and increased frontal lobe theta activity when mental workload increased during a variety of task types [13]. These findings are further supported by functional neuroimaging (fMRI) studies that found psychophysiological responses to workload were associated with both increased thalamic metabolism and a reduction in alpha activity [14], as well as both increased cingulate cortex activation and increased frontal theta activity [15]. Research utilizing ECG to capture cardiac activity found heart rate variability and interbeat-intervals were negatively correlated with workload [16]. The level of regional oxygen saturation (rSO₂) in the pre-frontal cortex gathered from fNIR has been associated with effort [17] and positively correlated with workload [18]. Additionally, research using the TCD to capture cerebral blood flow velocity in the middle cerebral artery found a positive correlation with workload [19]. Finally, eye tracking studies found increased pupil dilation [20], increased randomness in scan patterns as assessed by nearest-neighbor index (NNI) [21], increased fixation durations [22], increased number of fixations [23], and the Index of Cognitive Activity (ICA) [24] were all associated with workload changes. The results of these studies suggest psychophysiological metrics might account for unique variance in task performance unaccounted for by subjective metrics, therefore, investigation of such variables should be included in regression model analyses.

1.2 Theoretical Approach: Hierarchical Regression

Hierarchical regression is a method of variable selection in which variables are user-selected and entered into the model in incremental steps based upon their importance for outcome prediction [25]. The variables chosen and the order in which they are entered into the model are based on the specific research hypotheses, underlying theory, and past research [26]. Within the social sciences, correlated variables are commonly utilized to explain variance on a criterion variable while controlling for

other variables, hence justifying the application of a hierarchical regression approach [27]. The adjusted R^2 helps control the amount of variance accounted for in the dependent variable by adjusting the directional impact of correlated and non-correlated independent variables. Consequently, variables that are considered theoretically important contributors to performance or found to be associated with performance in past research are incorporated into the model first, followed by the addition of new exploratory variables [28]. The limitation of this approach relies on the researcher's theoretical knowledge of the relationships among variables and therefore an unbiased algorithmic approach might be more suitable in some cases.

1.3 Mathematical Approach: Stepwise Regression

Stepwise regression is a method of variable selection that accounts for the inclusion and deletion of variables during each step of the model building process [29]. The appeal of this approach becomes apparent when a model aims to explain the variance associated with the dependent variable using the least amount of predictor variables [25], which reduces the likelihood of overfitting the model with variables that can result in misleading predictive power [30]. The method begins by first evaluating all possible one-variable models using the following regression Equation (1):

$$E(y) = \beta_0 + \beta_1 x_i \quad (1)$$

where β_0 is a constant, β_1 is the coefficient for the i th variable, and x_i is the i th independent variable. For each i th independent variable, a t-test evaluating the β_1 parameter is conducted (computed by taking the value of the coefficient divided by standard error of the coefficient), and the variable with the largest absolute t-value is retained [31]. The following regression equation evaluates the remaining independent variables (2):

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_i \quad (2)$$

where β_0 is a constant, β_1 is the coefficient for the first variable, x_1 is the first selected independent variable, β_2 is the coefficient for the i th variable, and x_i is the i th independent variable. For each remaining i th independent variable, a t-test evaluating the β_2 parameter is conducted, and the variable with the largest absolute t-value is retained. Once the second variable is selected, the t-value of the β_1 parameter is rechecked to determine if it is still significant within the model. If the β_1 parameter is no longer significant, the β_1 variable is removed and replaced with another variable that results in the most significant t-test with the β_2 variable [31]. This procedure continues until no other independent variables are found to be significant within the model. The limitation of this approach falls on to the type of algorithm used by the statistical software [25], therefore to ensure validity of the outcome, researchers must know the mathematical procedure used to achieve any models.

2 Methods

2.1 Participants

Data were collected from 150 university undergraduates and graduates (age: $M = 19.57$, $SD = 3.45$) with 85 males (age: $M = 19.62$, $SD = 3.72$) and 65 females (age: $M = 19.50$, $SD = 3.09$). All participants were required to be right-handed, have normal or corrected to normal vision, and have no experience with the experimental testbed. Additionally, participants were required not to consume alcohol or sedative medications at least 24 h prior to the study, and caffeine and/or nicotine at least two hours prior to the study.

2.2 Experimental Task

Participants completed the experimental task using the Mixed Initiative eXperimental (MIX) testbed [32]. The MIX testbed simulated an operator control unit (OCU) for an unmanned ground vehicle (UGV) that traveled through a Middle Eastern town. During the task, participants monitored an aerial map located on the bottom of the OCU. The icons on the aerial map exhibited three types of changes: appear (icons added), disappear (icons removed), or move (icons relocated). Participants were required to identify and indicate the type of change by left-clicking on the appropriate corresponding change detection button located above the aerial map as quickly as possible before another change event occurred. The icons were derived from a common war-fighter symbol database [33], but had no associated meaning. During the experimental scenario, participants received three 5-min conditions comprised of 6, 12, or 24 changes per minute. Each event change consisted of two separate icons changing, but only one type of change occurred at a time. Event rates and saliency of event rates were derived from previous research [6]. Performance during the experimental task was calculated by taking the total number of change events correctly detected and dividing by the total number of change events presented collapsed across all three change types to give one total performance score.

2.3 Subjective Measures

Participants were administered the ISA and TLX after each event rate condition. The ISA is based on a 5-point rating scale and consists of a single question to assess how an operator felt during the task. The TLX requires participants to rate their perceived level of workload on six dimensions using a 100-point sliding scale. A global workload score was calculated by averaging each of the six subscales. Ratings from all three event rate conditions were averaged to determine an overall score for each subscale of each questionnaire across the entire scenario.

2.4 Objective Measures

Participants were attached to EEG, ECG, fNIR, TCD, and eye tracking sensors that monitored their physiological responses during the task. Similar to the subjective metrics, all three event rate conditions were averaged to determine an overall score for each metric across the entire scenario. Advanced Brain Monitoring's B-Alert X10 EEG nine channel system was used to record participant's brain and cardiac activity. The EEG was sampled at 256 Hz from F3, F4, Fz, C3, C4, Cz, P3, P4, and POz sensors sites using the international 10–20 system with references at each mastoid. Power spectral density analysis was used to extrapolate alpha (8–13 Hz), beta (14–26 Hz), and theta (4–7 Hz) wavelengths from each individual sensor site. Individual sensor sites were further combined to generate values for lobes (frontal, temporal, parietal) and hemispheres (left and right). Participant's heart rate and heart rate variability were calculated using the So and Chan method [34]. Somantics' Invos Cerebral/Somatic Oximeter was used to record participant's regional cerebral oxygen saturation (rSO₂). The fNIR sensors were placed on the participant's left and right hemisphere prefrontal cortex and measured changes in the levels of oxygenated hemoglobin and deoxygenated hemoglobin. Spencer Technologies' ST³ Digital Transcranial Doppler was used to record participant's cerebral blood flow velocity in the middle cerebral artery. TCD probes were carefully positioned on the participant's temples using the Marc 600 head frame set. Seeing Machine's FaceLAB 5 system was used to record participants' eye tracking data. Two desk-mounted cameras and an infrared light source were positioned in front of the participant, and were individually calibrated for each participant.

3 Results

In the present analyses, each metric previously described will be considered for building the model. In total, a mix of 43 objective and subjective variables are under consideration as contributors to the prediction of task performance. RStudio software was used to conduct hierarchical and stepwise regression analyses. Due to listwise deletions, 107 participants were included in the hierarchical regression analysis, and 94 participants were included in the stepwise regression analysis. Models were evaluated utilizing 5-fold cross-validation to accurately determine their performance with new data.

3.1 Hierarchical Regression

Subjective measures were entered at Step 1 and physiological measures were entered at Step 2 based on the theoretical assumption that subjective measures are more standardized and have been strongly correlated with task performance, specifically the TLX, and should therefore be entered into the model first. The subjective and objective variables entered in each step of the model can be found in Table 1.

In Step 1 of the analysis, the subjective variables resulted in a significant model for each fold that was evaluated with an average adjusted R^2 of .052. The Performance

Table 1. Subjective and objective variables entered into each step of the hierarchical regression.

Variables	Step entered
ISA	1
TLX_mental demand	1
TLX_physical demand	1
TLX_temporal demand	1
TLX_frustration	1
TLX_performance	1
EEG_frontal lobe theta	2
EEG_parietal lobe alpha	2
ECG_interbeat interval	2
ECG_heart rate variability	2
fNIR_left mean rSO ₂	2
fNIR_right mean rSO ₂	2
TCD_left mean velocity	2
TCD_right mean velocity	2
Eyetracker_ICA	2
Eyetracker_number of fixations	2
Eyetracker_average fixation duration	2
Eyetracker_square NNI	2

subscale from the TLX resulted in a significant coefficient for each of the five folds, and the ISA measure resulted in a significant coefficient for one of the folds. No other subjective variables resulted in significant coefficients.

In Step 2 of the analysis, the inclusion of the objective measures resulted in a significant model for each fold that was evaluated with an average adjusted R^2 of .207. The number of fixations and average fixation duration variables resulted in significant coefficients for each of the five folds. The ICA metric was a significant coefficient for four of the folds. Lastly, the right mean rSO₂ variable resulted in a significant coefficient for one of the folds. No other objective variables resulted in significant coefficients.

3.2 Stepwise Regression

The subjective and objective variables entered into the stepwise analysis can be found in Table 2.

The BIC information criteria was used to determine the addition and removal of variables into the model during the stepwise procedure [35]. According to the results, each fold resulted in a significant model with an average adjusted R^2 of .323. A summary of the variables entered into the model can found in Table 3. Given the nature of the stepwise procedure, each variable entered into the model resulted in a significant standardized coefficient.

Table 2. The subjective and objective variables entered into the stepwise regression analysis

Source	Variables
ISA	ISA
TLX	Mental Demand, Physical Demand, Temporal Demand, Frustration, Performance, Global Workload
EEG	Frontal Lobe Alpha, Frontal Lobe Theta, Frontal Lobe Beta Parietal Lobe Alpha, Parietal Lobe Beta, Parietal Lobe Theta Occipital Lobe Alpha, Occipital Lobe Beta, Occipital Lobe Theta Midsagittal Alpha, Midsagittal Beta, Midsagittal Theta Left Hemisphere Alpha, Left Hemisphere Beta, Left Hemisphere Theta Right Hemisphere Alpha, Right Hemisphere Beta, Right Hemisphere Theta
ECG	Inter-beat Interval, Heart Rate Variability, Heart Rate
fNIR	Left Mean rSO ₂ , Left Median rSO ₂ , Right Mean rSO ₂ , Right Median rSO ₂
TCD	Left Mean Peak Velocity, Left Mean Dias Velocity, Left Mean Velocity Right Mean Peak Velocity, Right Mean Dias Velocity, Right Mean Velocity
Eye tracker	ICA, Number of Fixations, Average Fixation Duration, Number of Saccades, Square NNI, Convex-hull NNI

Table 3. A summary of the variables entered into the model based on the stepwise regression analysis.

Variable	Number of times selected
Number of fixations	4
Average fixation durations	4
ICA	3
fNIR_right mean rSO ₂	1
fNIR_right median rSO ₂	1
Number of saccades	1
Square NNI	1
TLX_frustration	1
TLX_performance	1

4 Discussion

The goal for the present analyses was to evaluate two linear regression approaches, theoretical (hierarchical regression) and mathematical (stepwise regression), for variable selection to develop and compare performance prediction models using subjective and objective measures of workload. The analyses showed that the stepwise method resulted in better model performance than the hierarchical method in terms of adjusted R^2 when investigating the addition of psychophysiological metrics, as well as differing in the number of variables selected within the model. These differences suggest that the mathematical approach was more efficient compared to the theoretical approach for variable selection.

According to the results, the theoretical approach resulted in an average adjusted R^2 of .207 and the mathematical approach resulted in an average adjusted R^2 of .323. Both

of these results are deemed to be very weak effects for social science data and potentially due to the ratio of sample size to independent variables [36], however the performance difference between the two models are substantial. The theoretical approach included 18 variables while the mathematical approach included 4 to 7 variables into the final model. With such a high variable set included in the theoretical approach, multicollinearity becomes a concern [26]. Although 18 variables were entered into the final model using the theoretical approach, only three of those variables consistently resulted in significant coefficients including the ICA, number of fixations, and average fixation durations. The mathematical approach resulted in similar findings in which the ICA, number of fixations, and average fixation durations also consistently resulted in significant coefficients, however with substantially less variables entered into the final model. For both approaches, the majority of the variables selected into the final model were eye tracking metrics which is consistent with past research on the effectiveness of using the eye tracker for discriminating between levels of workload during a change detection task [10]. These results suggest that the mathematical approach is consistent with the theoretical approach, however the mathematical approach was more stringent as it was able to objectively identify and ignore non-contributing extraneous variables while selecting only the most relevant variables into the final model.

Variables entered into the model through the theoretical approach were the workload variables that have been backed by a significant body of research relating those variables with task performance. Several of these variables, most notably from the TLX, are considered standard metrics in the workload literature and have been consistently used by researchers to assess performance during a variety of tasks across a variety of domains [6, 9, 11, 37]. Although these variables were entered into the model using the theoretical approach and had the opportunity for being entered into the model through the mathematical approach, none of these variables resulted in significant coefficients for either of the final models. Variables that were selected included those that were associated with task performance, but do not have as much theoretical support compared to the TLX and EEG variables. These results suggest that utilizing a strict theoretical approach for variable selection can introduce bias early into the model building process in which variables are ignored and not properly utilized despite potential for significant prediction. Furthermore, these results suggest using a mathematical approach might help improve and contribute to theory by providing objective outcomes with limited bias to assist in evaluating the potential contribution of new exploratory variables.

Acknowledgements. This work was in part supported by the US Army Research Laboratory (ARL) (W91CRB-08-D-0015). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL or the US Government.

References

1. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003)
2. Hogarty, K.Y., Kromrey, J.D., Ferron, J.M., Hines, C.V.: Selection of variables in exploratory factor analysis: an empirical comparison of a stepwise and traditional approach. *Psychometrika* **69**(4), 593–611 (2004)
3. Otten, L.J., Rugg, M.D.: Electrophysiological correlates of memory encoding are task-dependent. *Cogn. Brain. Res.* **12**(1), 11–18 (2001)
4. Reinerman-Jones, L.E., Matthews, G., Barber, D.J., Abich IV, J.: Psychophysiological metrics for workload are demand-sensitive but multifactorial. In: *Proceedings of the 58th Human Factors and Ergonomics Society 2014*, pp. 974–978. SAGE Publications, Chicago
5. Abich IV, J.: Investigating the universality and comprehensive ability of measures to assess the state of workload. Doctoral dissertation, University of Central Florida, Orlando, FL (2013)
6. Abich IV, J., Reinerman-Jones, L., Taylor, G.S.: Establishing workload manipulations utilizing a simulated environment. In: Shumaker, R. (ed.) *Proceedings of the 8th International Conference of HCI International 2013*. Springer, Las Vegas, NV (2013)
7. Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kineses, W.: Improving human performance in a real operating environment through real-time mental workload detection. In: Dornhege, G., Millan, J.R., Hinterberger, T., McFarland, D.J., Muller, K.R. (eds.) *Toward Brain-Computer Interfacing*, pp. 409–422. MIT Press, Cambridge (2007)
8. Tattersall, A.J., Foord, P.S.: An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics* **39**(5), 740–748 (1996)
9. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, pp. 139–184. Elsevier Science Publishers, North-Holland (1988)
10. Abich, J., Reinerman-Jones, L., Taylor, G.S.: *Investigating Workload* (2013)
11. Measures for adaptive training systems. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **57**(1), 2091–2095. SAGE Publications, San Diego
12. Moroney, W.F., Biers, D.W., Eggemeier, F.T., Mitchel, J.A.: A comparison of two scoring procedures with the NASA-Task Load Index in a simulated flight task. In: *Proceedings of the 1992 IEEE Aerospace and Electronics Conference*, vol. 2, pp. 734–740 (1992)
13. Ergenoglu, T., Demiralp, T., Bayraktaroglu, Z., Ergen, M., Beydagi, H., Uresin, Y.: Alpha rhythm of the EEG modulates visual detection performance in humans. *Cogn. Brain. Res.* **20**(3), 376–383 (2004)
14. Smith, M.E., Gevins, A., Brown, H., Karnik, A., Du, R.: Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Hum. Factors* **43**(3), 366–380 (2001)
15. Larson, C.L.: Relations between PET-derived measures of thalamic glucose metabolism and EEG alpha power. *Psychophysiology* **35**(2), 162–169 (1998)
16. Asada, H., Fukuda, Y., Tsunoda, S., Yamaguchi, M., Tonoike, M.: Frontal midline theta rhythms reflect alternative activation of prefrontal cortex and anterior cingulate cortex in humans. *Neurosci. Lett.* **274**(1), 29–32 (1999)
17. Veltman, J.A., Gaillard, A.K.: Physiological workload reactions to increasing levels of task difficulty. *Ergonomics* **41**(5), 656–669 (1998)
18. Warm, J.S., Tripp, L.D., Matthews, G., Helton, W.S.: Cerebral hemodynamic indices of operator fatigue in vigilance. In: Matthews, G., Desmond, P.A., Neubauer, C., Hancock, P. A. (eds.) *Handbook of Operator Fatigue*, pp. 197–207. Ashgate Press, Aldershot (2012)

19. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B.: Functional optical brain imaging using near-infrared during cognitive tasks. *Int. J. Hum.-Comput. Interact.* **17**(2), 211–227 (2004)
20. Shaw, T.H., Guagliardo, L., de Visser, E., Parasuraman, R.: Using transcranial Doppler sonography to measure cognitive load in a command and control task. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **54**(3), 249–253 (2010)
21. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* **91**(2), 276 (1982)
22. Di Nocera, F., Terenzi, M., Camilli, M.: Another look at scanpath: distance to nearest neighbour as a measure of mental workload. In: de Waard, D., Brookhuis, K.A., Toffetti, A. (eds.) *Developments in Human Factors in Transportation, Design, and Evaluation*, pp. 1–9. Shaker Publishing, Herzogenrath (2006)
23. Jacob, R.J., Karn, K.S.: Eye tracking in human-computer interaction and usability research: ready to deliver the promises. In: Hyona, J., Radach, R., Deubel, H. (eds.) *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, pp. 573–605. Elsevier Science, Amsterdam (2003)
24. Van Orden, K.F., Limbert, W., Makeig, S., Jung, T.P.: Eye activity correlates of workload during a visuospatial memory task. *Hum. Factors* **43**(1), 111–121 (2001)
25. Marshall, S.P.: The index of cognitive activity: measuring cognitive workload. In: *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants*, pp. 7–9 (2002)
26. Lewis, M.: Stepwise versus hierarchical regression: pros and cons. Paper Presented at the Annual Meeting of the Southwest Education Research Association, San Antonio, TX, 7 February 2007
27. Field, A.: *Discovering Statistics Using SPSS*, 3rd edn. SAGE, Los Angeles (2009)
28. Pedhazur, E.J.: *Multiple Regression in Behavioral Research*, 3rd edn. Harcourt Brace, Orlando (1997)
29. Kerlinger, F.N.: *Foundations of Behavioral Research*, 3rd edn. Holt, Rinehart and Winston, New York (1986)
30. Thompson, M.L.: Selection of variables in multiple regression: part I. A review and evaluation. *Int. Statist. Rev.* **46**(1), 1–19 (1978)
31. Babyak, M.: What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med.* **66**(3), 411–421 (2004)
32. McClave, J.T., Sincich, T.: *Statistics*. Prentice Hall, Upper Saddle River (2009)
33. Reinerman-Jones, L., Barber, D., Lackey, S., Nicholson, D.: Developing methods for utilizing physiological measures. In: Tadeusz, M., Waldemar, K., Rice, V. (eds.) *Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations*. CRC Press, Boca Raton (2010)
34. Department of Defense: *Department of Defense Interface Standard: Common Warfighting Symbology (MIL-STD-2525B)* (2005)
35. Tan, K., Chan, K., Choi, K.: Detection of the QRS complex, P wave and T wave in electrocardiogram. In: *First International Conference on Advances in Medical Signal and Information Processing*, No. 476, pp. 41–47 (2000)
36. Beal, D.: Information criteria methods in SAS for multiple linear regression models. In: *Proceedings of the 15th Annual SouthEast SAS User Group (SESUG)* (2007)
37. Ferguson, C.J.: An effect size primer: a guide for clinicians and researchers. *Prof. Psychol.: Res. Pract.* **40**(5), 532–538 (2009)
38. Cinaz, B., Arnrich, B., La Marca, R., Troster, G.: Monitoring of mental workload levels during an everyday life office-work scenario. *Pers. Ubiquit. Comput.* **17**(2), 229–239 (2011)