



Physiological and Subjective Assessment for Training in Emergent Environments

Gerald Matthews¹ and Lauren Reinerman-Jones²
(University of Central Florida)

[1gmatthews@ist.ucf.edu](mailto:gmatthews@ist.ucf.edu) [2lreinerm@ist.ucf.edu](mailto:lreinerm@ist.ucf.edu)



Overview

- Relevance of workload measurement to training
- Subjective assessments of workload and stress
- Psychophysiological metrics for workload
- Workload response in three simulation studies (NPP, UAV, UGV)
- Individual differences: Performance prediction
- Applications
- Conclusion



Training, Workload and Stress

- Training generally mitigates overload and stress
- Two forms of workload/stress reduction:
 1. *Direct effects*
 - Acquisition of more effective strategies for coping with workload and negative emotion
 2. *Indirect effects*
 - Automatization of processing likely to reduce workload and consequent stress (on high workload tasks)
- Benefits are not guaranteed
 - Suboptimal learning, strategic readjustment
 - Individual differences are likely
- Needs for monitoring workload and stress during training



Training for Unexpected Events

- Typical unexpected event
 - Cognitive overload and failure of problem-solving/situation awareness (industrial disasters)
 - Accompanied by stress and fatigue
 - Reciprocal effect: stress feeds back into overload
- But also...
 - Complacency and failure to grasp gravity of event
 - May be exacerbated by automation (aviation incidents)
- Training needs
 - Metrics for gauging psychological impact of events
 - Metrics for evaluating simulated scenarios
 - Interventions matched to individual vulnerabilities





Assessment Challenges for Training

- Multifactorial nature of stress and workload
 - Explicit and implicit components
 - Limited convergence between subjective workload (NASA-TLX) and various physiological metrics (Matthews, Reinerman-Jones, Abich & Barber, 2014)
 - Both explicit and implicit measures are multidimensional
 - e.g., 6 components of NASA-TLX workload
- Multidimensional assessment important for training
 - Changing sources of workload (e.g., task demands vs. effort)
- Two training effects on stress
 - Anticipation of stress (baseline)
 - Acute stress management (task-induced response)



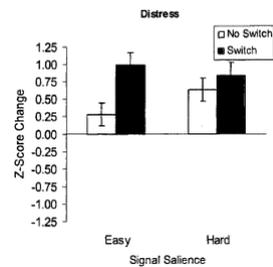
Subjective Metrics: Dimensional Models

- Workload: NASA-TLX (Hart & Staveland, 1988)
 - External sources: Mental, physical, and temporal demands
 - Internal sources: Effort, performance, and frustration
- Subjective state: DSSQ (Matthews et al., 2002, 2013)
 - Task engagement, Distress, Worry
 - Plus subscales
 - Validated in studies using a range of task and environmental stressors
 - Supplement with appraisal and coping measures



Stress States: Workload Transition

- Sudden workload transition
 - Vigilance studies
 - Changes in workload increase distress (Helton et al., 2004)
 - Pre-task warning of workload change does not help (Helton et al., 2008)
 - Low-high and high-low switches both detrimental

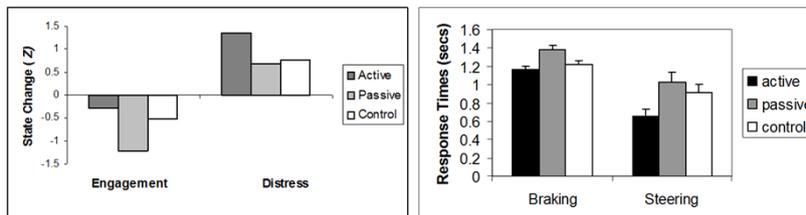


Workload switch effects in vigilance (Helton et al., 2004)



Stress States: Unexpected Events

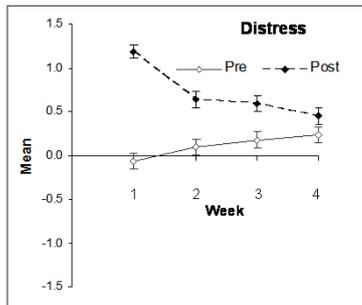
- Simulated vehicle driving (Saxby et al., 2013)
 - Automated-to-manual workload transition
 - Automated driving induces passive fatigue (loss of engagement)
 - Workload manipulation induces active fatigue and stress
 - Passive fatigue: slowed response to emergency (van pulling out)





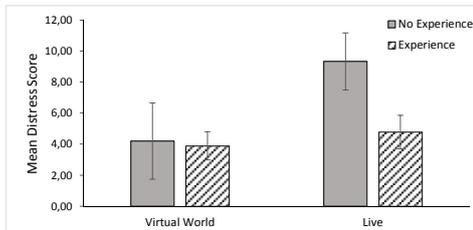
Stress States: Change with Practice

- Four sessions of practice on a speeded working memory task (Matthews & Campbell, 2010: N=112)
 - Practice enhances performance
 - Effects on subjective distress
 - Increasing baseline distress
 - Decreasing post-task distress
 - Consistent distress – working memory correlation across sessions
 - Illustrates tracking of stress and its functional significance during training



Assessment of Training

- Virtual world vs. live training for Army room clearing (Maxwell, Lackey, Salcedo & Matthews, 2014)
- Participants 64 reserve unit Soldiers; only some had prior experience
- Virtual world experience mitigated distress in the no-experience group





Psychophysiology of Workload

- Multiple sensors: EEG, ECG, hemodynamic (CBFV, fNIR), and eye tracking
- Compared with subjective measures (e.g., NASA-TLX)

Pros	Cons
Objective (no reporting bias)	Bias from physiological factors (e.g., metabolism, arousal)
Link to neuroscience	Limited evidence on validity
Continuous measurement	Plethora of data (need metrics)
	Practical issues (training, cost)



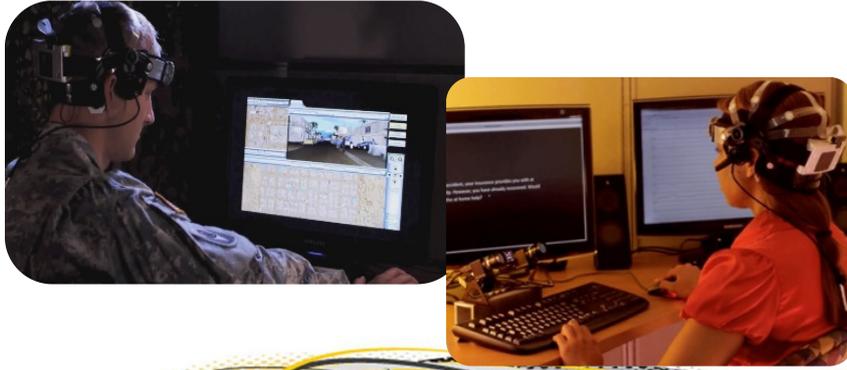
A Suite of Workload Sensors

Sensor	Method	Metrics	Notes on Metrics	Ambulatory Assessment
Electrocardiogram (ECG)	• Typical electrode placement: single-lead electrodes on the center of right clavicle and lowest left rib	• Heart rate (HR) or Inter-beat interval (IBI) • Heart rate variability (HRV)	• HR, IBI: Sensitive to arousal, activity • HRV: Mental effort, emotion regulation	• Straightforward: Ambulatory measures compare well with those from laboratory ECG equipment
Electroencephalogram (EEG)	• Multiple scalp electrodes	• Spectral power densities (SPDs) for frequency bands (delta, theta, alpha, beta, gamma) • Derived indices	• Task effort (frontal theta) • Other bands sensitive to arousal, cognitive activity • Task Load Index - Ratio of theta Fz: alpha Pz	• Challenging: low voltage signal, technical issues • Recent advances: wireless systems, miniaturization, 'dry' electrodes (no gel)
Cerebral bloodflow velocity (CBFV)	• Transcranial Doppler (TCD) ultra-sonography using probes above zygomatic arch	• Bilateral CBFV in medial cerebral arteries • Task-induced response	• Cognitive engagement vs. fatigue • Sensitive to coping	• Not yet practical—but useful for lab-based validation
Functional near-infrared spectroscopy (fNIR)	• Forehead IR light sources and detectors to measure prefrontal blood oxygenation	• Bilateral cortical oxygenation in the prefrontal cortex	• Task-directed effort, executive processing	• Portable and usable in field settings
Eyetracking	• Camera recording of the eye	• Frequency and duration of fixations • Pupillometry: Index of Cognitive Activity (ICA)	• Cognitive load • Also for areas of interest	• Head mounted units are promising



Physiological Sensor Integration

- Evidence for physiological links with performance supports the integration of Electroencephalography (EEG), Eye Tracking, Electrocardiography (ECG), and Hemodynamics (fNIR, TCD)



Physiology and Simulation Testbeds

- Include tasks that are independent and interact within complex dynamic environments while still supporting the controls demanded of a laboratory.
- Must have built in methods for inclusion of measures collected from multiple computers, third-party applications, and participants through external time synchronization techniques.
- Physiological sensors must be accounted for in any task battery as an alternative means of assessing observer performance and state and for predicting and augmenting within a given environment.



Psychometric Issues

- Individual metrics are internally consistent (reliable)
- Metrics are sensitive to external workload manipulations
 - Simulation studies of nuclear power plants (NPPs), unmanned ground vehicles (UGV)
 - Sensitivity of metrics varies across tasks and manipulations
 - Assessment of workload response as a multivariate profile
- No clear unitary workload response
 - Metrics are only weakly intercorrelated
 - Metrics are weakly related to subjective workload (e.g., NASA-TLX)



Evaluating Metrics for NPP

- Nuclear Regulatory Commission (NRC) has developed a tool to support the evaluation of workload (WL), situation awareness (SA), and teamwork (TW) metrics (Reinerman-Jones et al., 2015)
- Generic Metrics Catalog (GMC)
 - Database of metrics listing psychometric properties and study characteristics
- Decision-Making Wizard (DMC)
 - Decision tree to assist NRC reviewers in evaluating the choice and implementation of a metric.
- Available at <http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr7190/>



Diagnosticity for Sources of Demand

- Multiple sources of excessive demand
 - Multi-tasking
 - Difficult perceptual discriminations
 - Time pressure
 - ...and numerous others
- Can we use physiology to identify sources of demand?
 - e.g., distinctive workload 'signature' for multi-tasking
 - Support targeted group-based or personalized training
- Are diagnostic patterns general across applications?
 - e.g., does the multi-tasking 'signature' generalize across different task combinations and settings?
 - Or does the signature depend on task context?



Summary of Studies

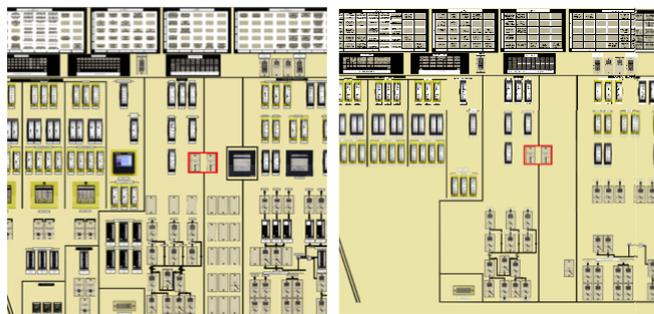
- Three studies
 1. Simulated NPP operation (N=81: Reinerman-Jones et al., submitted)
 - Task type: Checking, detection, response implementation
 2. Simulated UGV task (N=150: Abich et al., in press; Matthews et al., 2014; Reinerman-Jones et al., 2014)
 - Task type : Change detection, human threat detection
 - Multi-tasking: Single vs. dual
 3. Simulated UAV operation (N=68: Wohleber et al., submitted)
 - Multi-tasking: 2 vs. 6 UAVS (+time pressure)
 - Negative feedback: none vs. present)
- Comparison of physiological workload profiles
- A note on stress (UAVs)

Nuclear Power Plant Simulation

- Generic Pressurized Water Reactor (GPWR) simulator (Reinerman-Jones, Guznov, Mercado, & D'Agostino, 2013)
 - Based on GSE plant control room, modified with input from SMEs
 - Modified operating procedure (initiating event: EOP-EPP-001)
 - Sequence of tasks performed by three-person crew using three-way communication
 - Senior Reactor Operator (SRO: experimenter)
 - Reactor Operator (RO 1: participant)
 - Reactor Operator (RO2: confederate/experienced participants)
 - Checking, Detection, Response Implementation tasks
 - Two-hour training session: participants required to reach 80% accuracy criterion
 - 81 participants (45 M, 36F)

Panel Modification

Original (experts) → Modified (trained novices)

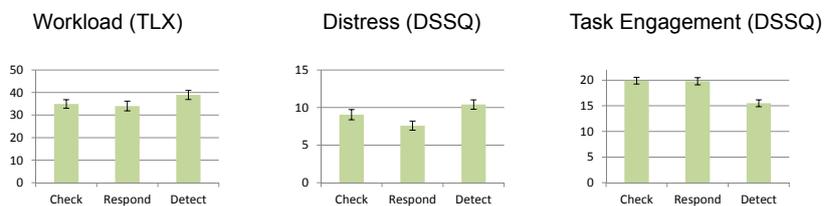


NPP Tasks

- **Checking**
 - Inspection of an instrument or control to verify that it was in the correct state (e.g., “verify valve PCV-444B is shut”)
- **Detection**
 - Locate and monitor a gauge for five minutes (“e.g., verify gauge TI-430 SB and report when less than 400 PSIG”)
 - Expected to impose attentional demands
- **Response Implementation**
 - Identify a control, and then open or shut a switch on that control (e.g., “shut valve 1CS-235B”)
- All tasks made up of steps executed using three-way communication between SRO and RO

Task Effects: Subjective

- Detection induces higher workload, higher distress, lower task engagement

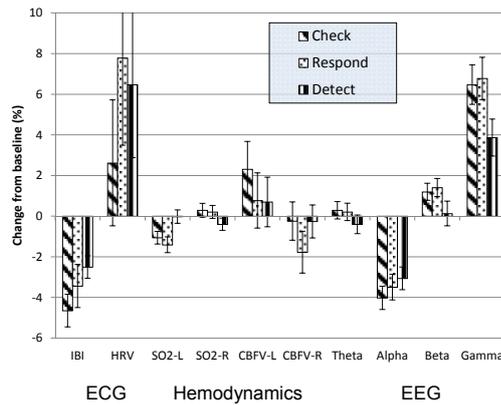


[Performance difficult to compare across tasks: ROs require more clarifications for detection]



Physiological Workload Response

- Change from baseline (%) for each workload metric
- Of most interest: differential task effects

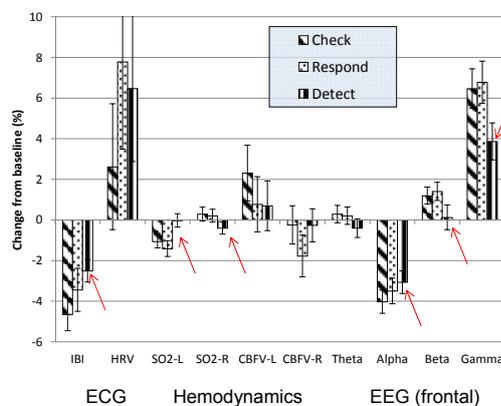


- Also, overall effects of performance
- Cardiac activity
 - Shift to higher frequency EEG



Physiological Workload Response

- Differences between Detection and other two tasks
- Multiple metrics are sensitive



- Patterned workload response
 - Lower arousal (IBI, alpha)
 - Higher effort (rSO₂)
 - Less high-level cognition (beta, gamma)
- Some classic workload metrics insensitive (HRV, frontal theta)

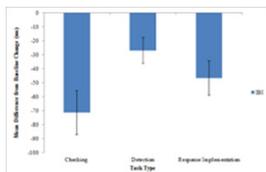
NPP Study: Implications

- Different task components of the OP vary in workload and stress (subjective data)
- High workload components (e.g., detection)
 - A focus for training
 - A vulnerability during unexpected events
- Physiological data
 - No general workload response
 - Patterned, nuanced response
 - May reflect multiple neural systems
- Do these findings generalize to UGV?

NPP Workload Response: Stability over Time

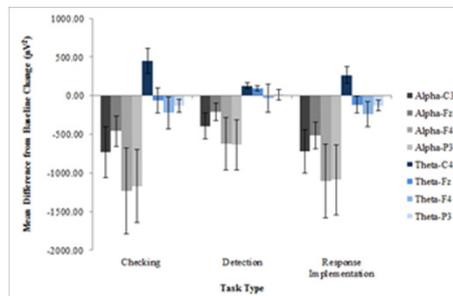
- Leis et al. (2014): Similar OP to previous study
- Three experienced participants. Seven training sessions, plus five experimental sessions (over three months)

ECG: Smaller IBI response



- No session effects: stable response

EEG: Less alpha blocking



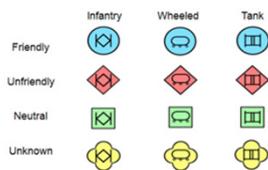


MIX: Stimuli for Detection Tasks

- For Threat Detection



- For Change Detection



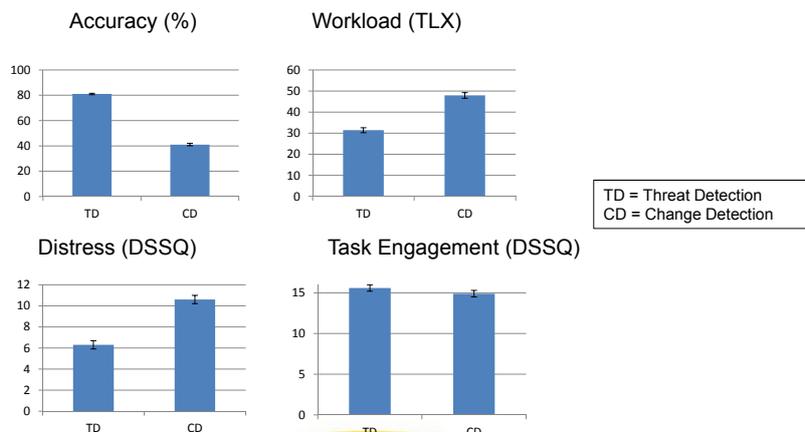
Change detection. Icons on map display may:

- appear
- disappear
- move



Threat Detection vs. Change Detection

- Threat detection is easier, lower in workload and less stressful



INSTITUTE for SIMULATION & TRAINING.

Physiological Workload Response

- Change detection vs. threat detection (single task)
 - For change detection (*higher workload*)
 - ECG: Lower HRV
 - Hemodynamics: Higher SO_2 (bilateral)
 - EEG: higher theta, lower beta
 - Eyetracking: higher pupillary Index of Cognitive Activity (ICA)
- Different pattern of workload response to NPP tasks
 - Reflecting differing attentional demands (multi-stimulus monitoring vs. vigilance)

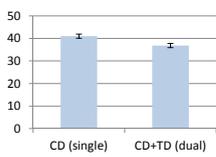


INSTITUTE for SIMULATION & TRAINING.

Single Task vs. Dual Task

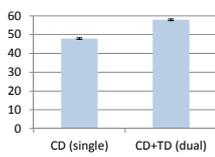
- Single task is easier, lower in workload and less stressful

Accuracy (%)



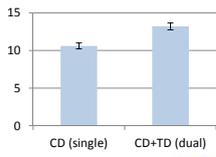
Task Condition	Accuracy (%)
CD (single)	~40
CD+TD (dual)	~35

Workload (TLX)



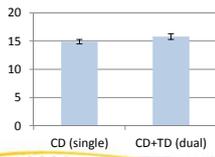
Task Condition	Workload (TLX)
CD (single)	~45
CD+TD (dual)	~55

Distress (DSSQ)



Task Condition	Distress (DSSQ)
CD (single)	~10
CD+TD (dual)	~13

Task Engagement (DSSQ)



Task Condition	Task Engagement (DSSQ)
CD (single)	~15
CD+TD (dual)	~16

Physiology:

Only sensitive metric is eye fixation duration (shorter for dual task)

Legend: TD = Threat Detection, CD = Change Detection





Multi UAV Simulation (RESCHU)

- Control of multiple UAVs on map display
 - Direct each UAV to a target location
 - Find target at location in camera view
 - Monitor messages in text box
 - Avoid hazards
- 68 undergraduate participants (31 men, 37 women)
- Within-subjects manipulations (Panganiban & Matthews, 2014)
 - (focus on stress as well as workload)
 - Task load (#UAVs, time pressure)
 - Negative feedback (non-contingent)
- 30 min training followed by series of 10-min tasks



Workload Manipulation

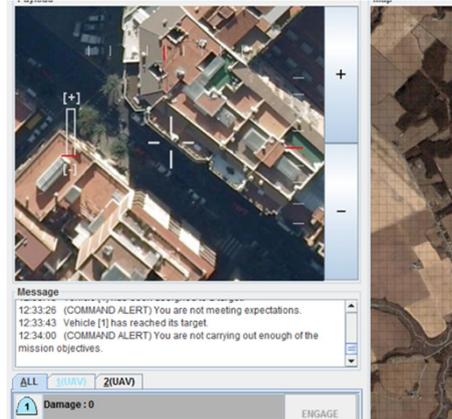
- 2 vs. 6 UAVs (+time pressure variation, #hazards)





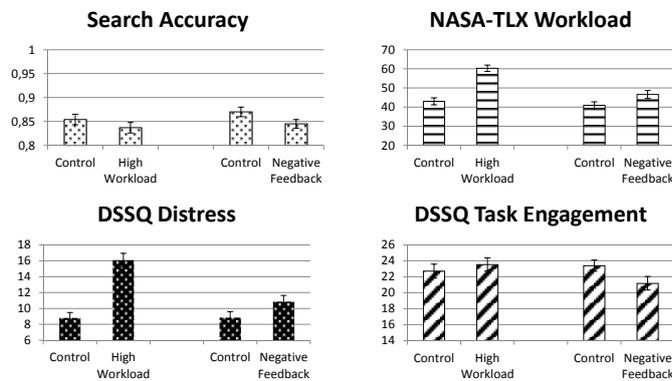
Negative Feedback Manipulation

- Periodic negative messages (versus neutral messages)



Cognitive Load/Stress Manipulations

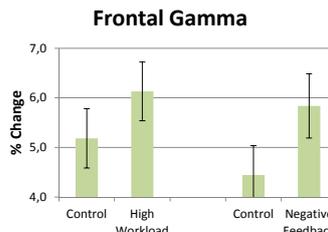
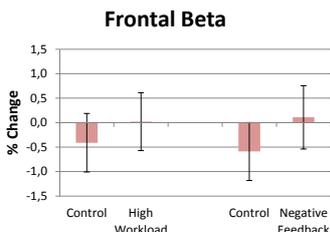
- Both factors: Poorer performance, higher workload, higher distress
- Negative feedback: Also reduces subjective engagement





Physiological Workload Response

- Different response pattern to both NPP and UGV studies
- Strongest effects of manipulations on EEG
 - Increased high frequency response



- Increased HRV (both)
- Decreased CBFV (feedback only)



Summary of Workload Effects

- Patterns of response specific to task/manipulation
- Subjective workload may signal a variety of neuro responses
- Metrics must be validated for a specific training context

Study	Task or Manipulation	NASA-TLX effect (points)	Metrics for elevated workload (expected)	Metrics for reduced workload (unexpected)
NPP simulation	Detection (vs. other tasks)	4.9	Higher rSO ₂ (fNIR)	Lower beta, gamma Lower heart rate Higher alpha
UGV simulation	Change detection (vs. threat detection)	16.1	Lower HRV, higher rSO ₂ (fNIR), higher frontal theta, higher ICA	
UGV simulation	Dual-task (vs. single task)	10.0	Shorter fixation duration	
UAV simulation	Cognitive load (vs. low cog. load)	16.7	Higher beta Higher gamma	Higher HRV

- N.B. Comparisons also affected by power of statistical analyses and strength of manipulations



Implications for Stress Metrics

- Increased distress common response to task demands
- DSSQ distress correlates at $\sim .5$ with NASA-TLX workload
 - Distress depends also on appraisal and coping strategy
- Changes in distress may be accompanied by greater or lower task engagement
 - e.g., negative feedback: distress + disengagement
 - Game-like tasks provoke both distress and engagement
- Physiology can identify different forms of distress
 - e.g., distress on RESCHU linked to high-level cognitive activity (increased high frequency EEG)



Performance Prediction

- Pronounced individual differences in both workload/ stress response and performance
- Abich, Matthews, & Reinerman-Jones (submitted): data from UGV study
 - All measures are internally consistent across task conditions (α range from .857 - .997)
 - Measures averaged across scenarios
- Two perspectives on prediction
 1. Use of pre-task measures to predict performance
 - Personnel selection, fitness-for-duty applications
 2. Use of concurrent measures to predict performance
 - Diagnostic monitoring, including training





Prediction from Subjective and Physio Metrics

- Prediction from baseline/pre-task metrics
 - Criteria are accuracy on change (CD) and threat detection (TD)

	HRV	EEG		DSSQ-Pre	
		Theta	Beta	Distress	Eng.
CD	-.225**	-.210**	-.261**	-.261**	.209*
TD	-.079	-.087	-.057	-.277**	.154

Note: *p < .05, **p < .01

- Regression statistics

Criterion	Predictor Set			
	Physio (R ²)	Subjective (ΔR ²)	Final R	Adj R ²
CD	.17**	.05*	.48**	.18
TD	.02	.08*	.32*	.05

Note: *p < .05, **p < .01

- Both types of metric necessary to optimize prediction



Prediction from Subjective and Physiological Metrics

- Concurrent/post-task metrics
 - Criteria are accuracy on change (CD) and threat detection (TD)

	HRV	EEG			DSSQ-Post		
		Alpha	Beta	Fix. Dur.	Dist.	Eng.	Worry
CD	-.258**	-.149	-.259**	.262**	-.392**	.451**	-.302**
TD	-.074	-.172*	-.154	.161	-.214**	.280**	-.247**

Note: *p < .05, **p < .01

- Regression statistics

Criterion	Predictor Set			
	Physio (R ²)	Subjective (ΔR ²)	Final R	Adj R ²
CD	.238**	.148**	.621**	.314
TD	.095	.065*	.400	.063

Note: *p < .05, **p < .01

- Both types of metric necessary to optimize prediction



Applications

- Multidimensional tracking of workload during training
 - Augmented cognition: 'closing the loop'
 - Identifying individual vulnerabilities
 - Personnel selection
- 



Multidimensional Tracking of Workload during Training

- Workload metrics may be diagnostic of overload
 - Diagnosticity may depend on context
 - Different metrics may identify different sources of overload
 - Use in scenario-based training
 - What are the events during training that provoke overload?
 - Physiological metrics better suited to tracking specific events than subjective ones
 - Design of scenarios to regulate workload
 - Use for training handling of unexpected events
 - Design of scenarios to provoke overload
- 



Augmented Cognition

- Human-Robot Teaming
 - How to make the robot sensitive to human workload?
 - Robot could be more supportive or proactive when human is overloaded
 - Use of physiology to monitor human functioning
- Closing the loop
 - Algorithm for continuous monitoring of human workload
 - Criterion value triggers adaptive response from robot
 - Work in progress using UGV simulation



Identifying Individual Vulnerabilities

- UAV (RESCHU) study
 - Resilience factors such as grit, hardiness, and meta-worry predict stress response
 - Systematic individual differences in vulnerability to overload and to negative evaluation
- Resilience training
 - Train for the vulnerabilities of the individual
 - Focus on performance and/or affective response
 - Cognitive strategies for overload
 - Detached mindfulness for negative evaluation stress



Personnel Selection

- Challenges for selection – how to identify...
 - Aptitude for working with complex tech (e.g., NPP, UAV interfaces)
 - Including automated and autonomous systems
 - Aptitude for maintaining performance during unexpected events
 - Aptitude for maintaining performance under multiple stressors
- Maximize prediction of performance and stress, in-context
 - Use of simulated environments
 - Use of multiple physiological and subjective indices
 - Multivariate statistical modeling



Conclusions

- Unexpected events often (not always) provoke cognitive overload and stress
- Scenario based-training can accommodate unexpected events
 - Assessment of training raises psychometric challenges
- Context-bound multivariate assessment of workload and stress meets these challenges
 - Both subjective and physiological assessment is of value
 - Assessment of both scenarios and individual response
- Further applications
 - Resilience training, fitness-for-duty assessment, selection

References

- Abich, J., IV, Matthews, G., & Reinerman-Jones, L.E. (submitted) *Individual differences in UGV operation: A comparison of subjective and psychophysiological predictors.*
- Abich, J., IV, Reinerman-Jones, L.E., & Matthews, G. (in press). Impact of workload factors on unmanned system Intelligence, Surveillance, and Reconnaissance operations. *Ergonomics*.
- Abich, J., IV, Taylor, G., & Reinerman-Jones, L. (2013, July). Establishing workload manipulations utilizing a simulated environment. Paper presented at the 8th International Conference of HCI International 2013, Las Vegas, NV.
- Helton, W. S., Shaw, T., Warm, J. S., Matthews, G., & Hancock, P. (2008). Effects of warned and unwarned demand transitions on vigilance performance and stress. *Anxiety, Stress, & Coping, 21*, 173-184.
- Helton, W.S., Shaw, T.H., Warm, J.S., Matthews, G., Dember, W.N., & Hancock, P.A. (2004.) Workload transitions: Effects on vigilance performance, and stress. In *Human performance, situation awareness and automation: Current research and trends*, Edited by: Vincenzi, D.A., Mouloua, M. and Hancock, P.A. (pp. 258-262). Mahwah, NJ: Erlbaum.
- Leis, R., Reinerman-Jones, L., Mercado, J., Barber, D., & Sollins, B. (2014). Workload from nuclear power plant task types across repeated sessions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, 210-214.
- Matthews, G., & Campbell, S. E. (2010). Dynamic relationships between stress states and working memory. *Cognition and Emotion, 24* 357-373.
- Matthews, G., Campbell, S., Falconer, S., Joyner, L., Huggins, J., Gilliland, K., Grier, R., et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2*, 315-340.
- Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2014). The psychometrics of mental workload multiple measures are sensitive but divergent. *Human Factors, 57*, 125-143.
- Matthews, G., Szalma, J., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). Profiling task stress with the Dundee Stress State Questionnaire. In L. Cavalcanti & S. Azevedo (Eds.), *Psychology of stress: New research* (pp. 49-91). Hauppauge, NY: Nova Science.
- Maxwell, D., Matthews, G., Salcedo, J., & Lackey, S. (2014). Virtual world room clearing: A study in training effectiveness. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL, 1-4 December 2014.*

References

- Panganiban, A. R., & Matthews, G. (2014). Executive functioning protects against stress in UAV simulation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, 994-998.
- Reinerman-Jones, L. E., Matthews, G., Barber, D. J., & Abich, J. (2014). Psychophysiological metrics for workload are demand-sensitive but multifactorial. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58, 974-978.
- Reinerman-Jones, L., Guznov, S., Mercado, J., & D'Agostino, A. (2013). Developing methodology for experimentation using a nuclear power plant simulator. In D. D. Schmorow, & C. M. Fidopiastis (Eds.), *Foundations of augmented cognition* (pp. 181-188). New York: Springer.
- Reinerman-Jones, L., Guznov, S., Tyson, J., D'Agostino, A., & Hughes, N. (2015). *Workload, situation awareness, and teamwork* (NUREG/CR-7190). Washington, DC: NRC.
- Reinerman-Jones, L., Mercado, J., Sollins, B., D'Agostino, A., & Hughes, N. (submitted). Workload Associated with Nuclear Power Plant Main Control Room Tasks. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.
- Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied, 19*, 287-300.
- Taylor, G. S., Reinerman-Jones, L. E., Szalma, J. L., Mouloua, M., & Hancock, P. A. (2013). What to automate addressing the multidimensionality of cognitive resources through system design. *Journal of Cognitive Engineering and Decision Making, 7*, 311-329.
- Wohleber, R.W., Matthews, G., Reinerman-Jones, L. E., Sollins, B., Panganiban, A. R., & Scribner, D.R. (submitted). *Individual differences in resilience and affective response during simulated UAV operations.*