Proposal for a Human Factors Fatigue Scale

Uyen Bui

California State University, Long Beach

**Feature at a Glance**

In 2017, the National Highway Traffic Safety Administration (NHTSA) reported 91,000 motor vehicle crashes that involved drowsy driving as well as 795 deaths from drowsy-driving-related crashes. From 2013 to 2017, there were 4,111 fatalities from motor vehicle crashes that involved drowsy driving (Drowsy Driving, 2019). Additionally, workers and students are using virtual meeting platforms (e.g., Zoom, Microsoft Teams) more than ever due to the ongoing COVID-19 pandemic. Having to use these virtual meeting platforms day after day, however, has led to “Zoom fatigue” (Wolf, 2020). With fatigue being so prevalent in everyday life activities, it is important to operationally define and accurately measure fatigue. This paper focuses on the construction of a scale to measure fatigue in a human factors context. Current fatigue scales and their issues as well as recommendations for the psychometric construction of a human factors fatigue scale will be discussed.

**Keywords:** active fatigue, passive fatigue, acute fatigue, chronic fatigue, psychometric scales, validity, reliability

**Relevance of Fatigue**

Fatigue can impair everyday activities, such as driving. According to the National Highway Traffic Safety Administration (NHTSA), there were around 91,000 motor vehicle crashes, 50,000 injured people, and 795 deaths related to drowsy driving in 2017. From 2013 to 2017, there were 4,111 deaths from drowsy-driving-related motor vehicle crashes (Drowsy Driving, 2019). Drowsiness during driving could be due to sleep-related fatigue from sleep deprivation or decreases in performance due to human circadian rhythms, active fatigue due to the large amount of mental resources consumed by driving in bad weather or high traffic conditions, or passive fatigue from long, monotonous drives in low traffic conditions that requires more monitoring (Körber, Cingel, Zimmerman, & Bengler, 2015). Semi-autonomous vehicles (e.g., Tesla) can reduce active fatigue as the automation performs the majority of driving while drivers perform the more monotonous task of monitoring the automation to ensure it is working properly while also remaining vigilant in case of a take-over request. Although drivers may not be actively fatigued from the tasks that come with manually operating a vehicle, the monotony of having to monitor the automation can lead to passive fatigue and vigilance decrements (Schmidt, Schrauf, Simon, Fritzsche, Buchner, & Kincses, 2009), leading to a decreased ability to sustain attention to the monitoring task (Parasuraman, 1998). Ultimately, drivers could have delayed reaction times when expected to take over driving in case of automation failure.

Additionally, workers and students across the globe are increasingly using virtual meeting platforms (e.g., Zoom, Microsoft Teams) due to the ongoing COVID-19 pandemic. Overuse of virtual meeting platforms has led to people experiencing “Zoom fatigue,” which includes feelings of tiredness, anxiety, or worry (Wolf, 2020). Although it should be quick and easy to meet virtually, workers and students may still struggle with needing to make time in their schedules to dress up, finding an appropriate and preferably quiet place, and navigating technology (e.g., making sure cameras and microphones are working) in order to attend a single Zoom meeting. The Zoom meeting itself is also mentally exhausting due to the difference between virtual social interactions and interactions in the real world (Wolf, 2020). Nonverbal cues could be delayed or not as visible, and prolonged eye contact with up close faces on a screen create unnatural social interactions.

**Defining Fatigue**

Due to the prevalence of fatigue in everyday life, it is important to know how to operationally define and accurately measure all types of fatigue. Although there are already many existing scales that measure fatigue according to varying contexts such as certain diseases (e.g., Parkinson’s disease, chronic fatigue syndrome), athletic settings (e.g., athlete fatigue or burnout), and particularly workplace settings (e.g., burnout or fatigue for medical staff, truck drivers, and pilots), fatigue may be defined differently according to the context in which it is being used (Enoka & Duchateau, 2016), or not be defined at all (Friedman et al., 2010). Varying or missing definitions raise the first issue of a general lack of consensus on how to define fatigue across different domains, which then raises a potential validity issue with existing fatigue scales—are the scales really measuring fatigue?

This paper will use the definition forwarded by Soames-Job and Dalziel (2001): “Fatigue refers to the state of an organism’s muscles, viscera, or central nervous system, in which prior physical activity and/or mental processing, in the absence of sufficient rest, results in insufficient cellular capacity or system-wide energy to maintain the original level of activity and/or processing by using normal resources” (pp. 469). This definition specifies the many causes of fatigue. For example, fatigue could result from muscular exertion (“physical activity”), or prolonged attention to a particular stimuli (“mental processing”). As attention has been described as a limited resource by both Kahneman (1973) and Wickens (1980), it is logical that exhausting the resource during information processing leads to fatigue. Addressing the “absence of sufficient rest” (Soames-Job & Dalziel, 2001, pp. 469) acknowledges that breaks during a task would help with fatigue and including “cellular capacity and system-wide energy” takes into account there are physiological inter- and intra-cellular processes that are influenced by fatigue, which will then impact task performance. Past research has indeed shown that regular rest breaks during prolonged tasks in industrial work settings allow workers to maintain performance levels, manage fatigue, and limit accident risks (Tucker, 2003), and that muscular fatigue impairs reaction time and successful completion of physical throwing tasks (Forestier & Nougier, 1998). The use of the inter- and intra-cellular processes “to maintain the original level of activity and/or processing by using normal resources” both acknowledges there is a change in resources to meet task demands and implies there are performance decrements, but does not equate fatigue with those decrements. Lastly, fatigue is a “state” that can produce a feeling, but is not only a feeling unto itself.

There are several types of fatigue, but for the sake of discussion, this paper will only discuss active, passive, acute, and chronic fatigue as they are the most relevant to human factors systems. Active fatigue is caused by active engagement with a task which consumes mental resources (Körber et al., 2015) due to perceptual and motor systems needing to continuously adjust to the task for a prolonged period of time (Desmond & Hancock, 2001). When a person is actively fatigued, they are overwhelmed by the demands of the task. Passive fatigue, on the other hand, is caused by a low workload consisting of merely monitoring a system, which requires little to no response from the perceptual and motor systems (Desmond & Hancock, 2001). The low workload and overall monotony of the monitoring task can lead to loss of attention and decrements in performance (Körber et al., 2015). Acute fatigue describes fatigue that is rather normal, as it has a quick onset and lasts for a short amount of time, can be traced back to a single cause, and be reduced with proper rest (Piper, 1989). Chronic fatigue, however, is seen as abnormal, cannot be traced back to a single cause as there can be multiple or even unknown causes, remains present over time, and cannot be reduced with proper rest (Piper, 1989).

**Measuring Active, Passive, Acute, and Chronic Fatigue**

Scales that measure task engagement tend to be used to measure active and passive fatigue. These scales include the Dundee Stress State Questionnaire (DSSQ) developed by Matthews and colleagues (1999, 2002) and the Short Stress State Questionnaire (SSSQ; Helton, 2004). For the sake of discussion, only the DSSQ will be described in further detail given that the SSSQ is the same scale in an abbreviated form. The DSSQ is an 11-item scale that measures task engagement, distress, and worry on a 5-point Likert rating ranging from “definitely false” to “definitely true.” Scoring low on task engagement items most closely correlates with fatigue, as task engagement indicates the amount of energy, interest in, and concentration on the task at hand (Saxby, Matthews, Warm, Hitchcock, & Neubauer, 2013). Though there may be task disengagement both active and passive fatigue, the effect should be larger for passive fatigue—if the task is monotonous and does not allow for much control from the operator, task engagement greatly declines as the operator will redirect their attention towards another task (Matthews et al., 2002). Elevated distress is expected to be associated with active fatigue, since workload is highly correlated with distress, and worry is expected to be associated with active fatigue as the operator has more time for self-reflection on performance due to the low workload of the task (Matthews et al., 1999). The DSSQ has shown internal consistency with alpha scores that range from .77 to .89 (Matthews et al., 2002).

Existing scales for acute and chronic fatigue mainly use subjective assessments. The Fatigue Assessment Scale (FAS) is a 10-item unidimensional scale that measures how a person typically feels using a 5-point Likert rating ranging from “never” (1) to “always” (5; Michielsen, De Vries, & Jolanda, 2003). Although the scale treats fatigue as one dimension, items of the scale measure both physical and mental symptoms of fatigue (Shahid, Wilkinson, Marcu, & Shapiro, 2011) as can be seen by the items “Physically, I feel exhausted” and “Mentally, I feel exhausted,” respectively. The FAS is used to assess chronic fatigue and adapts items from other existing fatigue scales: Checklist Individual Strength (CIS), World Health Organization Quality of Life assessment instrument (WHOQOL), and Fatigue Scale (FS; Michielsen et al., 2003). Internal consistency of the scale is high with a score of .90, and correlations are high between the FAS and subscales of other fatigue questionnaires (Michielsen et al., 2003), making it a useful scale for measuring fatigue.

Both chronic and acute fatigue are measured in the Occupational Exhaustion/Recovery scale (OFER), a 20-item scale that measures work-related fatigue on a 7-point Likert rating ranging from “completely disagree” (0) to “completely agree” (6) (Winwood, Winefield, Dawson, & Lushington, 2005). The chronic fatigue section of the scale measures both the mental and physical elements of persistent fatigue or exhaustion, including depressive elements, with regards to work, rather than a general fatigue state. The chronic fatigue subscale items measure both the psychological and physical outcomes of balancing both energy usage and recovery over a period of time, as can be seen from “I use a lot of my spare time recovering from work.” This section has an internal consistency of .93. The acute fatigue section measures the amount of residual energy the worker has remaining after a work shift, as can be seen from the item “I usually have lots of energy to give my family or friends.” The items in this section correlate negatively with the acute fatigue construct, meaning that higher scores on these items represent a lower level of acute fatigue. Therefore, these items need to be reverse-coded in order to show that higher scores equal a higher level of acute fatigue. The internal consistency of the acute fatigue section is .82.

Although all of these scales have good internal consistency, and therefore prove to be valuable items for measuring fatigue, each scale addresses only a few fatigue dimensions. For example, the DSSQ can be used to measure active and passive fatigue but not acute or chronic fatigue, and the FAS is unidimensional by design. Additionally, scales such as the OFER specifically focus on the workplace setting, which makes it difficult to generalize the scale to other contexts of fatigue. A human factors fatigue scale will need to incorporate multiple dimensions of fatigue and be capable of being generalized to multiple operational settings wherein fatigue can occur. For the sake of discussion, the dimensions to be included in a human factors scale are active, passive, acute, and chronic fatigue.

**Recommendations for Creating a Human Factors Fatigue Scale**

When developing a human factors fatigue scale, researchers should first define fatigue (Dawis, 2000). Lack of a clear definition may lead to further fragmentation of the operational definition and reduce validity of the scale in development. The definition of fatigue from Soames-Job and Dalziel (2001) is recommended as a foundation given that it includes many aspects of fatigue, such as the cause, physiological processes underlying fatigue, and benefits of adequate rest. As the scale will cover multiple dimensions of fatigue, it will also be beneficial to include the definitions of active, passive, chronic, and acute fatigue by Desmond and Hancock (2001), Körber et al. (2015), and Piper (1989). Researchers will see the definitions of fatigue and be able to determine whether the scale is appropriate for their usage.

Next, the fatigue scale should cover active, passive, chronic, and acute fatigue. Combining several types of fatigue into one scale will ensure the scale captures the multidimensional nature of fatigue, rather than being unidimensional as with the FAS. The items related to each type of fatigue, however, should be contained in unidimensional subscales to ensure the homogeneity of each subscale and that the scores of each subscale can be accurately interpreted (Dawis, 2000). It may also be beneficial to create an initial general scale that is applicable to different contexts of fatigue rather than immediately narrowing the scope by starting out with a specific context, such as the workplace setting used in the OFER. Variants of the scale can be created later on (e.g., Maslach Burnout Inventory and variants; Maslach & Jackson, 1981).

In the psychometric construction of the scale, it is important to minimize bias with regards to unchangeable participant characteristics such as age and ethnicity (Dawis, 2000). Two of the items in the FAS showed a gender bias, though the overall score of the FAS was not significantly affected and therefore the items were not adjusted (Michielsen et al., 2003). Bias should be taken into account while constructing the items of the scale, and it may be beneficial in the long run to adjust any items that are biased against or towards a population even if overall scores are unaffected. The bias could ultimately affect variants of the scale or other scales if items are adopted.

The aforementioned fatigue scales all use an odd-number of Likert rating responses, though more careful thought should be put into determining the number of responses, as choosing an odd-number of responses creates a “neutral” response that tends to be selected by many participants (Dawis, 2000). As the aim of fatigue scales is to determine whether or not the participant is fatigued, it may be beneficial to include an even amount of responses as it forces participants to indicate they are either fatigued or not fatigued. Including a “neutral” response in fatigue scales may lead to too many ambiguous responses that ultimately give no insight into the current state of the participant (i.e., cannot determine whether participant is fatigued or not).

Lastly, it is important to test the validity of the fatigue scale as assessing validity will also allow for assessing reliability (Dawis, 2000), both of which are very important for a psychometric assessment tool. When testing for validity, it is recommended to use a dataset that is different from the dataset used to select the items for the scale, since the response to items in the scale can be influenced by where the item is placed in the scale. Therefore, responses cannot always be generalized to the original dataset from which the final scale items were selected. Instead, researchers should continuously test hypotheses regarding the scale and fatigue itself (e.g., “If this scale is a measure of fatigue, then…”) to ensure the scale is actually measuring fatigue.

**Applications in Summary**

Creating a fatigue scale that includes multiple dimensions allows researchers to accurately assess the multidimensional nature of fatigue and could potentially facilitate the assessment process by using one scale versus multiple scales that each include only one dimension. Adequately defining fatigue and its subtypes in a single scale helps ensure the questionnaire’s validity and prevents further fragmentation of the fatigue definition. Additionally, starting out with construction of a more generic multidimensional fatigue scale can help form the basis of scale variants for more specific contexts, which will be more efficient than starting out in a specific context and working backwards to develop a general fatigue scale.

Concerning psychometric construction, taking into account bias during development of the items will ensure there is minimal bias in the final scale. Adjusting for bias even if the overall score is unaffected will also help in the long run when variants of the scale are created or items of the scale are adopted for use in another fatigue scale, as bias could appear later on in those different scales. Additionally, having an even number of responses to items will eliminate the “neutral” response in scale construction, forcing participants to indicate they are either fatigued or not fatigued for each item. As the scale is focused on assessing fatigue, eliminating the “neutral” response will remove ambiguity in scoring and produce relevant results either do or do not indicate a state of fatigue in the participant. Lastly, with copious amounts of validity testing, the fatigue scale will be both reliable and valid for measuring fatigue.

**Limitations and Future Areas of Consideration**

This paper discusses the development of a single human factors fatigue scale, thoughmore scales will need to be developed in the future. Additionally, only a few types of fatigue are included in the proposed scale when there are still many other types of fatigue to be measured, such as mental and cognitive fatigue. Future scales should consider incorporating more subscales that address the various types of fatigue to ensure every dimension of fatigue is captured. However, incorporating several types of fatigue into one scale could potentially be difficult to administer due to the scale length. Therefore, it is important to find the proper balance between capturing many dimensions of fatigue while also ensuring participants do not become tired and begin quickly selecting the same response towards the end (Galesic & Bosnjak, 2009). The overall practicality of creating multidimensional fatigue scales should also be further explored.

Although definitions are to be included in this proposed fatigue scale, it will be difficult to fix the current fragmentation of the fatigue definition. Hopefully, the recommendation to include definitions of fatigue in scales will become more widespread, though there might not be much that can be done regarding the existing widely used fatigue scales that do not have overall definitions and/or subscale definitions. Researchers should continue to work towards a singular definition of fatigue to ensure valid and reliable scales are being used to measure fatigue.

Regarding the psychometric construction of the proposed scale, only a few suggestions were given. Future areas to explore include further studying of the selection process of the items and absolute versus relative measurements. The selection process of items includes internal criterion methods (e.g., factor analysis and cluster analysis) or external criterion methods, (e.g., single-item or multiple-items-in-combination approach) (Dawis, 2000). When using absolute measurements, researchers are interested in the measure of the construct itself without comparing the construct measure to anything else (Hintze, 2005). When using relative measurements, researchers may compare the construct measure with another measure (Hintze, 2005), which could lead to measuring two different constructs in the same scale. Therefore, researchers should carefully consider and further investigate which measurement is most appropriate for fatigue scales.

References

Dawis, R. V. (2000). Scale construction and psychometric considerations. In H. E. A. Tinsley &

S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 65–94). Cambridge, MA: Academic Press. <https://doi.org/10.1016/B978-012691360-6/50004-5>

*Drowsy driving*. (2019, July 22). [Text]. NHTSA National Highway Traffic Safety

Administration. <https://www.nhtsa.gov/risky-driving/drowsy-driving>

Desmond, P. A., & Hancock, P. A. (2001). Active and passive fatigue states*.* In P. A. Hancock &

P. A. Desmond (Eds.), *Human factors in transportation. Stress, workload, and fatigue* (pp. 455–465). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Enoka, R. M., & Duchateau, J. (2016). Translating fatigue to human performance. *Medicine &*

*Science in Sports & Exercise*, *48*(11), 2228–2238. <https://doi.org/10.1249/MSS.0000000000000929>

Forestier, N., & Nougier, V. (1998). The effects of muscular fatigue on the coordination of a

multijoint movement in human. *Neuroscience Letters*, *252*(3), 187–190. https://doi.org/10.1016/S0304-3940(98)00584-9

Friedman, J. H., Alves, G., Hagell, P., Marinus, J., Marsh, L., Martinez-Martin, P., Goetz, C. G.,

Poewe, W., Rascol, O., Sampaio, C., Stebbins, G., & Schrag, A. (2010). Fatigue rating scales critique and recommendations by the Movement Disorders Society task force on rating scales for Parkinson’s disease. *Movement Disorders*, *25*(7), 805–822. <https://doi.org/10.1002/mds.22989>

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and

indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Helton, W. S. (2004). Validation of a short stress state questionnaire. *Proceedings of the*

*Human Factors and Ergonomics Society Annual Meeting*, *48*(11), 1238–1242. <https://doi.org/10.1177/154193120404801107>

Hintze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review*, *34*(4),

507–519. <https://doi.org/10.1080/02796015.2005.12088012>

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance decrement and

passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, *3*, 2403–2409. <https://doi.org/10.1016/j.promfg.2015.07.499>

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of*

*Organizational Behavior*, *2*(2), 99–113. <https://doi.org/10.1002/job.4030020205>

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliand, K., Grier, R., &

Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, 2, 315-340.

Matthews, G., Joyner, L., Gilliland, K., Huggins, J., & Falconer, S. (1999). Validation of a

comprehensive stress state questionnaire: Towards a state big three? In I. Merville, I. J. Deary, F. DeFruyt, and F. Ostendorf (Eds.), *Personality psychology in Europe* (vol. 7) pp. 335-350. Tilburg: Tilburg University Press.

Michielsen, H. J., De Vries, J., & Van Heck, G. L. (2003). Psychometric qualities of a brief self-

rated fatigue measure: The Fatigue Assessment Scale. *Journal of Psychosomatic Research*, *54*(4), 345–352. <https://doi.org/10.1016/S0022-3999(02)00392-6>

Parasuraman, R. (1998). *The attentive brain*. Cambridge, MA: MIT Press.

Piper, B. F. (1989). Fatigue: Current bases for practice. In S. G. Funk, E. M. Tornquist, M. T.

Champagne, L. A. Copp, & R. Wiese (Eds.), *Key aspects of comfort* (pp. 187-198). New York: Springer.

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*(4), 287–300. <https://doi.org/10.1037/a0034386>

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009).

Drivers’ misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, *41*(5), 1087–1093. <https://doi.org/10.1016/j.aap.2009.06.007>

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Fatigue assessment scale (FAS).

In A. Shahid, K. Wilkinson, S. Marcu, & C. M. Shapiro (Eds.), *STOP, THAT and One Hundred Other Sleep Scales* (pp. 161–162). New York: Springer. <https://doi.org/10.1007/978-1-4419-9893-4_33>

Soames-Job R. F., & Dalziel J. R. (2001). Defining fatigue as a condition of the organism and

distinguishing it from habituation, adaptation and boredom. In P. A. Hancock, & P. A. Desmond (Eds.), *Stress, workload and fatigue* (pp. 466-475). Mahwah, NJ: Erlbaum.

Tucker, P. (2003). The impact of rest breaks upon accident risk, fatigue and performance: A

review. *Work & Stress*, *17*(2), 123–137. https://doi.org/10.1080/0267837031000155949

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention*

*and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.

Winwood, P. C., Winefield, A. H., Dawson, D., & Lushington, K. (2005). Development and

validation of a scale to measure work-related fatigue and recovery: The Occupational Fatigue Exhaustion/Recovery Scale (OFER). *Journal of Occupational and Environmental Medicine*, *47*(6), 594–606. <https://doi.org/10.1097/01.jom.0000161740.71049.c4>

Wolf, C. R. (2020, May 14). *Virtual platforms are helpful tools but can add to our stress*.

Psychology Today. <https://www.psychologytoday.com/blog/the-desk-the-mental-health-lawyer/202005/virtual-platforms-are-helpful-tools-can-add-our-stress>

**Uyen Bui** is a student in the Master of Science in Human Factors program at California State University, Long Beach (CSULB). She received her Bachelor of Arts in Psychology at CSULB. Her research interests currently lie in the areas of usability testing, accessibility, and user behavior surrounding passwords. Uyen can be reached at [uyen.bui@student.csulb.edu](mailto:uyen.bui@student.csulb.edu).