HHS Public Access

Author manuscript

Ergonomics. Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

Ergonomics. 2022 July ; 65(7): 960–975. doi:10.1080/00140139.2021.2006317.

Validation of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) Adapted for the Whole Day Repeated Measures Context

Raymond Hernandez[a], **Shawn C. Roll**[a], **Haomiao Jin**[b], **Stefan Schneider**[b], **Elizabeth A. Pyatak**[a]

[a]Chan Division of Occupational Science and Occupational Therapy, University of Southern California, 1540 Alcazar St, Los Angeles, CA 90089 United States

[b]Dornsife Center for Self-Report Science and Center for Social & Economic Research, University of Southern California, Los Angeles, California, USA

Abstract

Our objective was to investigate the validity of four-item and six-item versions of the National Aeronautics and Space Administration Task Load Index (NASA-TLX, or TLX for short) for measuring workload over a *whole day* in the *repeated measures* context. We analyzed data on 51 people with type 1 diabetes from whom we collected ecological momentary assessment and daily diary data over 14 days. The TLX was administered at the last survey of every day. Confirmatory factor analysis fit statistics indicated that neither the TLX-6 nor TLX-4 were a unidimensional representation of whole day workload. In exploratory analyses, another set of TLX items we refer to as TLX-4v2 was sufficiently unidimensional. Raw sum scores from the TLX-6 and TLX-4v2 had plausible relationships with other measures, as evidenced by intra-person correlations and mixed-effects models. TLX-6 appears to capture multiple factors contributing to workload, while TLX-4v2 assesses the single factor of “mental strain.”

Practitioner Summary

Using within-person longitudinal data, we found evidence supporting the validity of a measure evaluating whole-day workload (i.e. workload derived from all sources, not only paid employment) derived from the NASA- TLX. This measure may be useful to assess how day-to-day variations in workload impact quality of life among adults.

Keywords

task load; mental strain; workload; patient ergonomics; type 1 diabetes

Corresponding Author: Raymond Hernandez, 1540 Alcazar St., CHP-133, Los Angeles, CA 90089-9003 United States. hray57024@gmail.com.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, RH, upon reasonable request.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

1. Introduction

The National Aeronautics and Space Administration Task Load Index (NASA-TLX, or TLX for short) is a multi-dimensional scale widely used in ergonomics and human factors research to obtain workload estimates (Hart, 2006; Mansikka et al., 2019; Noyes & Bruneau, 2007). The TLX defines workload as the cost (e.g., fatigue, stress, illness) of performing tasks (Hart, 2006). It accounts for the contribution of objective task demands and a person's experience of those demands (Hart & Staveland, 1988). Workload as measured by the TLX has been found to be a significant predictor of overall fatigue (Arellano et al., 2015), emotional stress (Rutledge et al., 2009), and burnout (Ziaei et al., 2015).

The TLX was originally developed to assess perceived demands of discrete activities (Hart & Staveland, 1988) and has more recently been adapted to examine entire work shifts in the nursing literature (Hoonakker et al., 2011; Tubbs-Cooley et al., 2018). In the discrete task context, validity evidence was found for the full-length six item version (TLX-6) (Hart & Staveland, 1988, p. 199). As shown in Figure 1, the TLX-6 considers overall workload as *caused by* six items, namely, mental demand, physical demand, temporal demand, effort, performance, and frustration. It is often assumed that each of the six items has an equal impact (same weight) to justify using a raw sum score to represent the overall workload (Hart, 2006). In the whole work shift context, results of confirmatory factor analysis supported the validity of a four-item TLX version (TLX-4) that omitted performance and frustration (Tubbs-Cooley et al., 2018). Within their conceptual model (also show in Figure 1), workload is *indicated by* the four TLX items. They assume that the four items (mental demand, physical demand, temporal demand, and effort) have equal loadings on the one underlying factor of perceived demands, which then justifies using the sum of the four items as indicators of a latent workload construct (DiStefano et al., 2009).

There is a need to extend the use of the TLX to measure total exposure to workload over an *entire day* from work and/or non-work sources. It comes with the possible disadvantage (depending on the research question) of conflating work and non-work workload, but also has potential utilities. No existing workload measures account for workload experienced during non-work time, when in fact collective workload from all tasks in a day can contribute to the symptoms of high workload. For instance, fatigue and stress were found to be significantly higher for nurses when considering their additional non-work responsibilities of caring for children and elders (Scott et al., 2006). The National Institute for Occupational Safety and Health (NIOSH) has proposed a new worker well-being framework that reflects a holistic paradigm shift and supports consideration of non-work factors that could impact worker health (Chari et al., 2018). This expanded approach to worker well-being integrates work and non-work health promotion, and acknowledges that work and non-work factors may contribute in an additive fashion to health outcomes such as stress (Sauter, 2013). A whole day workload measure is aligned with this paradigm shift. Another potential use is capturing the effects of work-related workload that also impact non-work time. Effects of work are often not confined just to the workplace, and at times spill over to non-work settings (Leiter & Durup, 1996). For instance, fatigue from workload at work may make it difficult to meet the demands of non-work roles (Greenhaus & Beutell, 1985) (e.g. caring for children), which may in turn increase frustration and

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

decrease satisfaction with activity performance, thereby increasing the perception of whole day workload (TLX-6 conceptualization of it).

A whole day measure of workload may also be relevant for adults with chronic conditions. Nearly 45% of all Americans suffer from at least one chronic condition (Raghupathi & Raghupathi, 2018), and approximately 11% of the US population has diabetes (Centers for Disease Control and Prevention, 2020). Relative to people without a chronic condition, individuals with diabetes often spend more time managing their health, possibly increasing total demand exposure (Hansen et al., 2018). Daily health management responsibilities for type 1 diabetes include self- monitoring of blood glucose, calculating and administering insulin in accordance with blood sugar levels, food, activity patterns, and other variables, abiding by dietary recommendations, and treating acute complications such as hypoglycemia (Pyatak et al., 2018). As such, *whole day workload may impact how the daily presentations of chronic conditions like diabetes manifest*. For example, a worker with diabetes experiencing excessive workload may neglect to take his/her insulin and hence present with less controlled blood sugar levels for that day (Hansen et al., 2018).

A whole day workload measure may have utility for the practice of patient ergonomics. This idea that patients perform effortful work towards health-related goals (and thus may have greater whole day workload as a result) has been termed “patient work” (Holden & Abebe, 2021). Patient ergonomics is the application of human factors or related conditions to support the performance of patient work (Holden et al., 2020). It assumes that just like the sociotechnical system around workers influences their ability to perform work tasks, the sociotechnical system around patients impacts their ability to carry out health management tasks (Holden et al., 2015). The patient ergonomics framework has been applied to patients with diabetes and other conditions (Valdez et al., 2016). A whole day workload measure may help serve as one metric of whether a patient ergonomics intervention is effectively able to address the sociotechnical system around a patient, as evidenced by changes in their perceived workload.

Validation of a whole day workload measure requires data collected longitudinally, as opposed to cross-sectionally. Measures are typically validated in the cross-sectional context, which is most appropriate when the measure of interest is static in nature thereby justifying one- time measurement. *Workload however is not static* and can fluctuate greatly from day to day. Proper validation therefore requires collection of longitudinal data to allow investigation of questions such as if, within a particular person, changes in workload have the expected associations with other daily variables (e.g. higher fatigue on high workload days) (Bolger et al., 2003). Furthermore, within-person variation in workload may be important to capture because day-to-day fluctuations could be pertinent to short-term well-being, pain, and functioning (Ansiau et al., 2008). For example, someone exposed to a high workload from working overtime on a particular day may experience increased severity of pain symptoms on that day or on subsequent days.

1.1 Study Aims

We investigated the *intra-person* validity and reliability of the TLX-6 and TLX-4 used in the *whole day* context. Despite the potential utility for research and practice within both

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

general and working populations, there have been no measures validated to evaluate whole day workload. To evaluate construct validity, we investigated the factor structure of both TLX versions using confirmatory factor analysis (CFA), assessed the degree to which TLX sums had the expected associations with other variables (Abma et al., 2016), and examined day-of-week differences in whole day workload. We calculated within-person Cronbach's alpha (i.e., internal consistency) to determine the reliability of the TLX in a whole day context. To address the use of whole day workload as it relates to the well-being, health, and functioning of individuals with chronic conditions, we conducted this validation study with adults with type 1 diabetes (T1D).

2. Methods

2.1 Study Overview

We collected data from adults aged 18 to 75 with T1D who were participating in a longitudinal study on diabetes management (Pyatak et al., 2021). Recruitment criteria included the ability to use a smartphone, oral proficiency in English or Spanish, not currently pregnant, and not undergoing treatments or procedures that could impact blood glucose levels. Data for this analysis were collected across 14 days and included a baseline survey, daily ecological momentary assessments (EMA), end-of-day surveys, and an exit survey at the end of the two weeks. The University of Southern California Institutional Review Board approved the study protocol, and participants provided informed consent through REDCap e-consent (Harris et al., 2009).

2.2 Measures

We examined three broad categories of constructs highly relevant to whole day workload (activity exposure, cognitive performance, and health-related outcomes), to assess the convergent validity of the whole day workload measure. Individual items and measurement frequencies within each construct are presented in table 1. The TLX and work hours were administered in the daily end-of-day survey, and sleep quality was assessed in the first EMA survey each morning. All other items were measured through EMA surveys administered on smartphones up to 6 times per day at 3-hour intervals. All EMA data were converted to daily unweighted averages based on the number of responses received during that day.

The TLX-6 and total work hours were both measures of daily activity exposure, and momentary activity exposure was assessed through reports of activity types. The workload questions were six items derived from the original TLX (Hart, 2006). We altered the wording slightly to simplify phrasing for readability on phone screens and be more applicable to a whole day instead of specific tasks. The TLX-4 contained only four of the six items (mental demand, physical demand, temporal demand, and effort).

The frequencies of strenuous and restful activities relative to the total number of EMA surveys taken in the day were derived from an item asking participants what activity they were doing immediately before each momentary (EMA) prompt, using a method with preliminary evidence of validity (Hernandez et al., 2021). Possible activity responses were based on a taxonomy of activities created by occupational therapists based on their practice

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

framework (American Occupational Therapy Association, 2020) and decided upon with expert review by occupational scientists. The activity response choices were conceptually sorted into strenuous and restful activities by emulating an approach used in prior studies where activities are placed into categories based on their typical characteristics (e.g. watching television typically represents a "sedentary" behavior) (Tudor-Locke et al., 2009). Strenuous (high demand) activities included "work/school" and "caring for others" because of the large body of literature conceptualizing both as high demand engagements (Dich et al., 2015; Meijman & Mulder, 1998; Sonnentag & Fritz, 2007). We defined restful (low demand) activities as engagements often accompanied by increases in parasympathetic activity (Tindle & Tadi, 2021), so "relaxing/chilling" and "sleeping/napping" were included under its umbrella. Admittedly, this sorting scheme may have missed person and context specific sources of workload and rest, and we discuss this further in the limitations section.

To calculate daily activity frequencies within the strenuous and restful categories, we divided the number of times people reported engaging in a particular activity type by the number of EMA assessments taken (Sonnenberg et al., 2011). For example, if a participant completed six surveys in a day and reported engaging in "work/ school" during two surveys and "caring for others" in one, then the relative frequency of strenuous activities for that day would be 3/6=.5. Frequencies were only calculated for days where participants completed four or more surveys.

To assess if the whole day TLX was sensitive to non-work workload, we measured caregiving frequency as the proportion of times in a day participants reported "caring for others". Caregiving is one of the few sources of non-workplace demands that has been frequently investigated and associated with higher demand levels (Dich et al., 2015). Thus, we used it as a rough approximation of exposure to non-work demands.

Cognitive performance was assessed at each EMA survey using a Go/No-Go task, a test of sustained attention ability, and a Symbol Search task, an assessment of visual-spatial attention and processing speed (Sliwinski et al., 2018). These aspects of cognitive performance were chosen because they were hypothesized to be affected by the blood glucose level of adults with T1D. Prior research had provided preliminary evidence indicating that hypoglycemia (low blood glucose) was associated with poorer visual processing speed (Ewing et al., 1998) and decreased attention (Draelos et al., 1995). The Go/No-Go and Symbol Search were also chosen because of their feasibility to administer on a smartphone. Both tests were scored such that higher scores indicate better momentary cognitive performance.

In addition to self-reported health outcomes (fatigue, affect, stress, pain, and sleep), we also explored associations between workload and physiological health parameters. In our sample of adults with T1D, we examined the convergent validity of workload to multiple blood glucose (BG) measures derived from continuous glucose monitoring when participants were awake. These T1D specific measures of BG included mean, standard deviation (SD), coefficient of variation (CV), and the percent of the time in a normal range (70mg/dL<BG≤180 mg/dL), with high blood glucose (BG>180mg/dL), and with low blood glucose (BG<70mg/dL).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2.3 Statistical Power

A sample size of 50 provided sufficient power to conduct a within-person CFA to examine the validity of the TLX as a whole day measure. The largest factor model we planned to investigate had 12 parameters, including the factor loadings for each item and their residual variances. We conservatively estimated a need for 120 (12*10) observations, based on standard guidelines for indicating a need for up to ten observations per estimated parameter (Bentler & Chou, 1987; Bollen, 1989). In the context of a within-person CFA model, each daily measure counts as an observation. With approximately 12 daily measures per person, enrollment of 50 people would result in 600 observations that would be sufficient to conduct the planned analysis.

A sample size of 50 was also sufficient to detect small effects (r of .20) (Cohen, 2013) in within-person correlation tests. Using the "power.rmcorr" function in the statistical software "R" (R Core Team, 2020) that tests for power of repeated measures correlations (Bakdash & Marusich, 2017), a sample of 50 with 12 observations per individual was calculated to provide 81% power to detect effect sizes of r=.12 with α=.05. Based on prior daily diary research, we assumed an intraclass coefficient of .5 for the nesting of repeated observations in individuals (Merz & Roesch, 2011; Roesch et al., 2010).

2.4 Statistical Analyses: Validity

We assessed validity for the TLX as a whole day measure in three ways that all took advantage of the multiple observations per individual to assess *intra-person*, as opposed to inter-person, relationships. Within-person CFA was used to analyze the degree of fit for the TLX to a unidimensional model of whole day workload. Repeated measures correlations were carried out to examine if associations of the TLX with activity exposure, cognitive, and health measures provided evidence supporting its convergent validity. Finally, mixed-effects models were used to identify day-of-week differences in TLX scores.

We determined that having factor structures and significant associations with other variables consistent with theoretical underpinnings (Table 2) would be evidence of construct and convergent validity for use of the TLX-6 and TLX-4 in a whole day context. With activity exposure, we expected the experience of strenuous (high demand) activities to be associated with increased workload while greater rest would decrease it (Meijman & Mulder, 1998). Furthermore, we anticipated that more frequent engagement in caring for others (i.e. non-work workload) would be associated with increased workload. In terms of cognitive performance, the workload stemming from long working hours has often been associated with lower cognitive performance (Olds & Clarke, 2010; Virtanen et al., 2009). Thus, we expected these findings to generalize to the intra-person context as a negative association between job demands and cognitive test scores on the same day and on the next day.
With regards to health- related outcomes, in line with prior literature assessing demands primarily in the work context, positive associations were expected between whole day workload and fatigue, negative affect, stress (Bowling et al., 2015), pain (Ariëns et al., 2001; Linton, 2001), and diabetes stress (Hansen et al., 2019). Negative associations were expected between whole day workload and positive affect (mental well-being) (Bowling et al., 2015) and sleep quality (Lange et al., 2009). Finally, consistent with prior literature, some daily

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

BG metrics were predicted to have relationships with whole day workload (Hansen et al., 2018, 2019), though which BG metrics and directionality have yet to be established in this emerging area of research.

2.4.1 Confirmatory Factor Analysis—

The anticipated structures of the TLX-6 and TLX-4 are shown in Figure 1; we tested how well a single factor model fit both. Because the TLX-6 has been conceptualized as a composite measure, we expected a single factor model to fit poorly. On the other hand, in keeping with previous research, we anticipated that a single factor model would fit the TLX-4 well (Tubbs-Cooley et al., 2018). A single level CFA was examined for both using only the pooled within-groups covariance matrix, which provides information about how items typically covary within individuals from one day to another. This approach is appropriate for within-person analyses of nested data (e.g., multiple TLX measures nested in individuals) (Huang, 2017). The CFA was estimated using the lavaan package in R (Huang, 2017).

Model fit was evaluated using the following fit indices: root mean square error of approximation (RMSEA) of at least <.08 but ideally <.05, comparative fit index (CFI) and Tucker-Lewis Index (TLI) of at least >.90 but ideally >.95, standardized root mean square residual (SRMR) <.08 (Hu & Bentler, 1999), and a non-significant Chi-square at the .05 level (Hooper et al., 2008). Additionally, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the four-item and six-item models were compared where lower values indicate better fit (Gamst-Klaussen et al., 2018). Finally, we examined the extent to which each item loaded on the underlying latent factor. One guideline is that items with factor loadings <.40 have weak associations with the latent variable (Cabrera-Nguyen, 2010).

2.4.2 Within-person Correlations—

We investigated if workload had plausible correlations with other *continuous* measures as an indicator of convergent validity. There are a variety of methods to calculate workload with the TLX, including a raw sum, weighting items by factor loading, only summing items with factor loadings above a specific cutoff, and calculating factor scores (DiStefano et al., 2009). We opted for the raw sum approach for both the TLX-6 and TLX-4 as it is the most widely used method (Hart, 2006; Hoonakker et al., 2011), thereby allowing for ease of interpretation and comparison with other studies. We calculated *within-person* correlations between workload and other daily measures using the “rmcorr” function in R, which calculated within-person correlations while accounting for the non-independence of data due to multiple observations per individual (Bakdash & Marusich, 2017). For the cognitive variables, we examined correlations with the *same day* and *day after* TLX measurement because we hypothesized potential temporal effects in their relationships. We also tested the within-person correlation between workload and scores on the *last* cognitive tests of each day. Ratings of sleep quality were compared to TLX measures completed both later that evening and in the evening of the day prior.

Because the planned analyses involved a large number of statistical comparisons, we considered whether adjustment for multiple comparisons was needed. Instead of accounting for multiple comparisons of within-person correlations with statistical corrections of alpha, we opted to report all comparisons along with the individual 95% confidence intervals

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Rothman, 1990; Saville, 1990). With this complete information, readers can account for the multiple comparisons when interpreting the results, making the process less susceptible to type II error than if statistical corrections were applied (Rothman, 1990).

2.4.3 Mixed Modeling—We assessed if workload had plausible relationships with day of the week and work vs. non-workdays through mixed modeling to account for the categorical and nested nature of the data (multiple days nested in an individual) while testing hypotheses involving within-person categorical predictors. We defined workdays as days where participants reported at least one period of engaging in work activities. Models had the categorical variable as the only predictor, the intercept specified as the random effect, and the TLX sums as the dependent variable. Non-workdays and Sunday served as the reference category for each respective model. Given that p-values of parameters from mixed models may be biased because of the inherent uncertainty in the degrees of freedom to specify for calculations (Luke, 2017), especially when the data structure is complex or the dataset is unbalanced, we used bootstrapping to derive standard error and perform significance testing (Appiah, 2018; Luke, 2017). Bootstrapping was also done to account for potential non-normality in the mixed model residuals.

2.5 Statistical Analyses: Reliability

2.5.1 Cronbach's alpha—To test the reliability of the four-item and six-item versions of the TLX, we calculated within- and between-person Cronbach's alpha using the "mcfa.input" function in R (Huang, 2017). Alpha values of .7 or above are widely considered desirable (Taber, 2018), with values of .6 described as acceptable (Ursachi et al., 2015; van Griethuijsen et al., 2015).

3. Results

3.1 Participant Demographics

We analyzed data from a total of 51 participants with a mean age of 38.9 years (SD 14.36, range 18 to 75 years), who were 61% female and predominantly white (45%) or Latino/x (37%). Most of our sample were working either full-time (45%) or part-time (12%), and 65% of the sample had an associate's degree or higher. Participants were not required to report their vocations, but a wide variety of occupations were reported including housekeeper, teacher, investigator, dental assistant, and software engineer. The predominant annual income categories were "less than 25,000" (22%), "didn't provide" (20%), "100,000 to 199,000" (16%), and "200,000 or more" (12%). The median EMA completion percentage was 94%, with an interquartile range of 9.5%. On average, participants completed at least four EMA surveys per day on 12.6 days out of the possible 14 days.

3.2 Construct Validity

3.2.1 Confirmatory Factor Analysis—CFA results are presented in Table 3. It is likely that our large sample size led to all chi-square values being statistically significant (Hooper et al., 2008); therefore, we focused on other fit metrics. Both the TLX-6 and TLX-4 had a poor fit in the unidimensional CFA model testing, each with RMSEA >.10.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The TLX-4 fit indices suggested that a slight adjustment to the items included had potential to create a TLX version that was unidimensional. We decided to conduct exploratory analyses to investigate if a unidimensional version was possible to help us understand the extent of the similarities between the various workload dimensions, which we thought would help provide greater general clarity on the whole day workload construct. Exploratory analyses indicated that substituting physical demands with frustration in the TLX-4 had the best fit to a unidimensional model, with all fit statistics except chi-square within a recommended range and a lower AIC and BIC than either the TLX-6 or TLX-4. Conceptually, we consider these four items to commonly assess workload in terms of "mental strain," and we refer to this as the TLX-4 version 2 (TLX-4v2; Figure 2).

Standardized within-person factor loadings for single-factor models fitted on the three TLX versions are shown in Table 4. "Performance," "frustration," and "physical" had loadings of <.40 in all models suggests weak associations with the latent variables. Further analyses were conducted only with TLX versions that had the expected results from a test of unidimensional model fit (i.e., TLX-6 and TLX-4v2). We anticipated that the original TLX-4 hypotheses would also apply to TLX-4v2.

3.2.2 Within-person Correlations—Within-person correlations between the TLX sums and other measures are shown in Table 5. Correlations of the TLX-6 and TLX-4v2 were in the same direction for all measures. Of the 19 within-person correlations, for the TLX-6 and TLX-4v2 10 and 9 of the tests respectively were consistent with our predictions. For the TLX-6 four of four activity exposure hypotheses were confirmed, while for the TLX-4v2 the figure was three of four. With both the TLX-6 and TLX-4v2, none of the six cognitive hypotheses, and six of nine health hypotheses were confirmed. Associations with positive affect (TLX-6: r=−.08, p=.065; TLX-4v2: r=−.07, p=.082) and sleep quality the same day the TLX was measured (TLX-6: r=0, p=.98; TLX-4v2: r=.02, p=.607) did not align with our hypotheses. None of the correlations with either cognitive test were significant. TLX sums were significantly negatively correlated with average BG during waking hours of the same day (TLX-6: r=−.14, p=.02; TLX-4v2: r=−.09, p=.041). Compared to the TLX-4v2, the TLX-6 had slightly greater correlations with fatigue, negative affect, diabetes stress, pain, and mean daily BG.

3.2.3 Mixed Modeling—Table 6 shows the results of mixed models for day of the week and mean TLX-6/TLX-4v2 values by day. Sunday had the lowest mean overall task load for both TLX-6 and TLX-4v2. Mean TLX scores for all weekdays were significantly higher than those on Sundays, with scores being the highest toward the end of the week (Thursdays), while TLX scores on Saturdays did not significantly differ from those on Sundays. Results shown at the bottom of Table 6 indicate that workdays had significantly higher mean TLX scores than non-workdays. These results are consistent with our hypotheses.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

3.3 Reliability

3.3.1 Cronbach's alpha—TLX-6 had a within-person Cronbach's alpha of .515 while TLX-4v2 had alpha=.696. Since TLX-6 was not unidimensional per its CFA fit indices, its lower Cronbach's alpha relative to TLX-4v2 was expected.

4. Discussion

4.1 Principal Findings: Validity

4.1.1 Confirmatory factor analysis—*Contrary to prior research* in the between-person context (Tubbs-Cooley et al., 2018) and to what we hypothesized, the original TLX-4 did not appear unidimensional. One possible reason for the discrepancy may be differences in participant characteristics. Our study had participants with various employment statuses, whereas the prior TLX-4 study only had nurses in their sample (Tubbs-Cooley et al., 2018). Another possible reason is that we conducted CFA in the within-person context, whereas prior CFA analyses were conducted in the between-person context. The lack of unidimensionality of the original TLX-4 prompted us to engage in exploratory analyses to test a set of four items that seemed more theoretically likely to load onto a single factor. With TLX-4v2, we hypothesized that the physical demand item captured something different from the other items, and we replaced it with the frustration item. The resultant set of four items fit a single factor model well, which might be best described as "mental strain."

Figure 3 illustrates the relationship between the TLX-4v2, TLX-6, and workload implied by our CFA results. As hypothesized, CFA results indicated the six-item TLX did not appear to represent a single construct; instead, in this context, the TLX-6 portrays some combination of mental strain, physical demand, and satisfaction with activity performance, all of which are theorized to collectively contribute to overall workload. Directional arrows to the TLX-6 from the construct of mental strain in the TLX-4v2, along with physical demand and performance, denote that these factors are *causing* workload and may not necessarily covary strongly with one another. Arrows from the TLX-4v2 to the four items on the left signify that the TLX-4v2 (daily task demands) is being *indicated* by them, as evidenced by the high covariance they share.

To summarize, CFA results supported the construct validity of the TLX-6 and TLX-4v2 but not TLX-4. We draw this conclusion because the single factor model CFA fit results were consistent with the theory for TLX-6 (composite measure) and TLX-4v2 (unidimensional), but not the theory for TLX-4 (supposed to be unidimensional). The TLX-4v2 was not a model we specified a priori but emerged in our analyses. Thus, there is a possibility that its factor structure seen here is specific to our dataset. Replication of the CFA results for TLX-4v2 in another population would strengthen the argument that its factor structure aligns with its theoretical underpinnings.

4.1.2 Within-person correlations—Sums of the TLX-6 had slightly stronger correlations with manifestations of whole day workload (e.g., fatigue, negative affect, pain) compared to the TLX-4v2 sums, likely because it accounts for additional contributors (e.g., physical demand and satisfaction with performance) to workload. This finding supports the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

construct validity of the TLX-6 within a whole day context as it is consistent with the theoretical formulation of the TLX-6 as a measure of multiple factors contributing to overall workload. Given these results, the TLX-6 may be more useful as a whole day measure for populations where physical demands and satisfaction are more significant contributors to workload, or these components are key factors relative to outcomes of interest. For example, physical demands would likely be much greater contributors to overall workload in construction workers compared to office workers or other sedentary occupations. Physical demands may also be particularly relevant for people with chronic conditions that make performance of physical activities difficult, such as rheumatoid arthritis (Carandang et al., 2020) or multiple sclerosis (Krupp et al., 1988).

We anticipated that exposure to higher workload on one day would decrease cognitive performance on the following day; instead, we found no significant association between the two. When examining the association between workload and *same day* cognitive performance and between workload and cognitive performance on the *last* survey of each day, no significant associations were again seen. The lack of relationships between TLX sums and cognitive performance may be because *long-term* exposure to high workload may be the precursor to cognitive performance deficits (e.g., routinely high working hours) (Virtanen et al., 2009). The TLX measures used here captured *short-term* exposure to workload.

Positive affect and sleep quality were the two other variables with non-significant correlations with the TLX, contrary to our hypotheses. The correlations with positive affect approached significance (.065 for TLX-6 and .082 for TLX-4v2) and were in hypothesized directions (r=−.08 for TLX-6 and r=−.07 for TLX-4v2). Correlations with sleep quality on the same day or the next were near zero and not close to significance. One possible explanation is that the question “How well rested do you feel?” does not reliably capture sleep quality. People may often take time to “feel rested” after waking. They may feel groggy right after rising from bed, resulting in lower “rested” ratings, but may have had acceptable sleep quality.

The TLX-6 was found to have a significant (and small) correlation with caregiving frequency, while a non-significant association with the TLX-4v2 was seen. Caregiving frequency was our only measure of non-work workload, so the results would seem to imply that only the TLX-6 and not the TLX-4v2 is sensitive to non-work workload. Perhaps the TLX-6 had a significant association with caregiving frequency because it accounted for physical demands and performance satisfaction, whereas the TLX-4v2 did not. Further research may be needed to investigate the sensitivity of both versions of the whole day TLX to non-work workload. While caregiving is frequently a source of non-work workload, many other sources of non-work workload exist. Comparison of the TLX to measures that better capture the totality of non-work workload experienced may be a more valid way of determining the sensitivity of both versions of the TLX to non-work workload.

Mean daily BG had a slightly more negative correlation with the TLX-6 (r=−.14, p=.02) as compared to the TLX-4v2 (r=−.09, p=.041). This may have been because the TLX-6 also accounts for physical demands. Blood sugar typically lowers after exercise/physical activity,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

so adding the variance of physical demands to overall workload may have increased the magnitude of the negative correlation between mean BG and whole day workload.

Overall, the within-person correlation results appeared to support the convergent validity of the TLX-6 and TLX-4v2. Except for cognitive performance, workload was associated with most of the activity and health variables measured in our sample of individuals with T1D. The TLX-6 showed slightly stronger correlations with hypothesized manifestations of workload, consistent with the theory that it encompasses more task load contributors than the TLX-4v2.

4.1.3 Mixed modeling—Sums of both TLX-6 and TLX-4v2 had expected relationships with the day of the week and work versus non-workdays as per the results of mixed modeling, further supporting their validity. Though not hypothesized a priori, the mean workload steadily increased from Monday to Thursday and slightly decreased on Friday. This pattern is consistent with prior literature. Due to a greater proportion of demands relative to recovery experiences in the early part of the work week, ratings of fatigue peaked midweek then decreased (Rook & Zijlstra, 2006).

4.2 Principal Findings: Reliability

4.2.1 Cronbach's alpha—TLX-4v2 had acceptable within-person reliability as indicated by its Cronbach's alpha value (.696), consistent with the unidimensional nature of that scale. The TLX-6 had poorer reliability (under .60). Given that the TLX-6 does not follow the traditional effect indicator model (items indicating a factor) that classical reliability measures typically assume (Bollen & Bauldry, 2011), the poor Cronbach's alpha was expected. Statistical methods to find the optimal way to measure reliability of causal indicator models such as the TLX-6, where items that cause the latent variable are still unclear (Bollen, 2017). Once such a method is further developed, it may be a more accurate way to assess the reliability of assessments like the TLX-6.

4.3 Implications

If interested in measuring mental strain as a single aspect of whole day workload, our findings suggest that the TLX-4v2 could be used. In contrast, the TLX-6 may be a useful whole day measure that accounts for a mix of factors contributing to overall workload. A whole day TLX may have the greatest utility as a repeated measure. In the repeated measures context, the whole day TLX would enable investigation of how other repeated measures relate to daily workload. For example, whole day workload could be associated with daily fluctuations in abdominal pain for people with irritable bowel syndrome. If the whole day TLX is used as a one-time measure, then results may be aggregated across multiple individuals to investigate whole day workloads within larger groups.

We hope this work can serve as a contribution to both worker and patient ergonomics. One of the gaps identified in the patient ergonomics literature was lack of longitudinal studies (Holden et al., 2020), so perhaps one contribution of this study was a longitudinal investigation of the human factor of workload in a type 1 diabetes population. One finding was a within-person relationship between blood glucose and whole day workload, such

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

that higher workload was associated with lower average BG on the same day. Our study may serve as a template for future research investigating the within-person relationships between whole day workload and the presentations of other chronic conditions, in case the results inform management strategies for these conditions. This study may also offer another metric, whole day workload, to assess patient ergonomic interventions. Chronic conditions often come with a host of additional responsibilities including following complex treatment regimens, self-monitoring one's condition(s), and making decisions about when/how to receive professional treatment (Dixon et al., 2009) that can all serve as additional workload sources. Ensuring that whole day workload is not excessive may be particularly relevant for preventing burnout from these health management practices. In terms of a worker population, a whole day workload measure may aid in studies/interventions operating under NIOSH's more holistic well-being framework that supports consideration of non-work factors that could impact worker health (Chari et al., 2018).

4.4 Limitations

One limitation of this study is that the relatively small number of participants in our sample did not afford meaningful tests of the reliability and validity of whole day workload measures in the between-person context. The within-person evidence presented here supports the argument that the TLX-4v2 and TLX-6 are valid measures of the workload of a single day. However, we provided no evidence regarding whether the average of a person's TLX measures across all testing instances was a valid indicator of typical whole day perceived demand (between-person setting).

Our sample included individuals with T1D who were experiencing the COVID pandemic during data collection. Measures of daily workload and findings of all analyses using the TLX-6, TLX-4, and TLX-4v2 items based on our sample may differ from adults without diabetes or measures obtained in the absence of a global health pandemic. However, given that the TLX-6 and TLX-4v2 demonstrated expected associations with our other variables, we anticipate these findings persist under different circumstances. Replication in larger, heterogeneous samples and further exploration of whole day workload measures within other contexts can be used to validate these findings.

Implicit in our sorting scheme for our strenuous and restful activity frequency measures was a certain degree of error that likely created a deviation from true strenuous/restful activity exposure. This is because misclassification during conceptual sorting of activity types into strenuous or restful may have occurred. For instance, if a participant reported engaging in "housework/errands", this activity would never fall into the strenuous or restful categories. However, housework can be strenuous or restful depending on the person and situation, such as an individual that may find cleaning the house after a long day of work therapeutic (restful). Even with the error implicit in our sorting approach, we anticipated that it was still sufficient to roughly capture strenuous and restful activity frequencies by virtue of accounting for activities most commonly identified in the literature as falling into those categories (e.g. work/school and caring for others for strenuous, and relaxing/chilling and sleeping/napping for restful). In future studies, increased accuracy could be achieved using a more refined sorting of activities into strenuous and restful bins if, in addition to a question

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

about type of activity engaged in, there is also a rating of the perceived demand associated with that activity. The downside of this approach is the added respondent burden with the additional question. If this method is taken however, the calculated strenuous and restful activity frequencies would likely be more accurate, and by extension correlations of larger magnitude with whole day workload may be seen.

Average (middling amount of) workload was not captured in the exposure measures against which we validated the TLX, and thus we cannot say whether the whole day TLX is sensitive to frequency of engagement in average workload tasks. Theoretically, perhaps we would expect a correlation between frequency of “average” demand activities and the TLX that is of lower magnitude compared to strenuous activities, but still positive and significant. However, we could not test this because we could not, with confidence, assert that any particular activity type was consistently associated with “average” workload (like we asserted that work is associated with higher demands). If, in a future study, participants reported the activity just engaged in on EMA surveys as well as associated ratings of demands, then frequency of engagement in average workload activities may be more readily assessed. This measure could then be used to investigate how sensitive the TLX is to a middling level of workload.

We used the standard approach in longitudinal data analyses of assuming that our missing data were missing at random (Cursio et al., 2019), which may have biased our results. There is a possibility that data was not missing at random. For instance, high workload engagements may have been undercounted because some participants may be less likely to answer surveys during those activities. Work on modeling data not missing at random in the longitudinal context is ongoing (Cursio et al., 2019).

Finally, a limitation associated with use of the whole day workload measure in any population is conflation between workload from professional work and workload from other sources. Whole day workload cannot distinguish between the two, and should not be used for studies or interventions where that discernment is required. However, as stipulated in NIOSH’s updated worker well-being framework, different areas of people’s lives (e.g. professional versus private) often overlap, so broad based programs (and perhaps by extension assessments) covering work and non-work may at times have utility (Chari et al., 2018). For instance, if the goal of an intervention is to address a worker’s experience of stress generally rather than stress only from work, then a whole day workload measure may provide a fuller picture of the collective demands on the worker that may be contributing to stress. In some contexts, the assessment of both whole day workload and workload from specific sources (e.g. paid work, patient work, or other strenuous non-work activities) may be indicated to reap the benefits of both, albeit at the cost of additional participant burden.

5. Conclusion

In this study, we found evidence supporting the reliability and validity of a new combination of four items from the TLX for use as a whole day within-person measure of a single factor of mental strain. The TLX-6 was also validated as a whole day measure of overall workload that combines three factors of mental strain, physical demand, and performance satisfaction.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

While not unidimensional, the TLX-6 had a slightly greater magnitude of association with activity exposure and health factors in a sample of individuals with T1D relative to the TLX-4v2, and a significant association with non-work workload (caregiving frequency) that the TLX-4v2 did not. Contrary to prior research, the original TLX-4 was not unidimensional and may not be suitable for use in the within-person context. Choice of which version of the TLX to employ may depend on the purpose of its use. Researchers or practitioners primarily interested in examining mental strain as a contributor to health or well-being may use the TLX-4v2. Conversely, researchers investigating populations for which a range of task load contributors are more relevant may use the TLX-6.

Acknowledgments

Funding

Data collection was supported by the National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK 1R01DK121298–01). Costs for data analysis support were covered by the Pilot Project Research Training Program of the Southern California NIOSH Education and Research Center (SCERC) (T42 OH008412) from the Centers for Disease Control and Prevention (CDC).

References

Abma IL, Rovers M, & van der Wees PJ (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: Constructing hypotheses and interpreting outcomes. BMC Research Notes, 9. 10.1186/s13104-016-2034-2

American Occupational Therapy Association. (2020). Occupational Therapy Practice Framework: Domain and Process—Fourth Edition. American Journal of Occupational Therapy, 74(Supplement_2), 7412410010p1–7412410010p87. 10.5014/ajot.2020.74S2001

Ansiau D, Wild P, Niezborala M, Rouch I, & Marquié JC (2008). Effects of working conditions and sleep of the previous day on cognitive performance. Applied Ergonomics, 39(1), 99–106. 10.1016/j.apergo.2007.01.004 [PubMed: 17434440]

Appiah AK (2018). Bootstrap Linear Mixed-Effects Models Using SAS® Procedures. 17.

Arellano JLH, Castillo Martínez JA, & Serratos Pérez JN (2015). Relationship between Workload and Fatigue among Mexican Assembly Operators. International Journal of Physical Medicine & Rehabilitation, 03(06). 10.4172/2329-9096.1000315

Ariëns GAM, Bongers PM, Hoogendoorn WE, Houtman ILD, van der Wal G, & van Mechelen W. (2001). High Quantitative Job Demands and Low Coworker Support As Risk Factors for Neck Pain: Results of a Prospective Cohort Study. Spine, 26(17), 1896–1901. [PubMed: 11568702]

Bakdash JZ, & Marusich LR (2017). Repeated Measures Correlation. Frontiers in Psychology, 8, 456. 10.3389/fpsyg.2017.00456 [PubMed: 28439244]

Bentler PM, & Chou C-P (1987). Practical Issues in Structural Modeling. Sociological Methods & Research, 16(1), 78–117. 10.1177/0049124187016001004

Bolger N, Davis A, & Rafaeli E. (2003). Diary methods: Capturing life as it is lived. Annual Review of Psychology, 54(1), 579–616.

Bollen KA (1989). Structural equations with latent variables Wiley. New York.

Bollen KA (2017). Notes on measurement theory for causal-formative indicators: A reply to Hardin. Psychological Methods, 22(3), 605–608. 10.1037/met0000149 [PubMed: 28891664]

Bollen KA, & Bauldry S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. Psychological Methods, 16(3), 265–284. 10.1037/a0024448 [PubMed: 21767021]

Bowling NA, Alarcon GM, Bragg CB, & Hartman MJ (2015). A meta-analytic examination of the potential correlates and consequences of workload. Work & Stress, 29(2), 95–113. 10.1080/02678373.2015.1033037

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cabrera-Nguyen P. (2010). Author Guidelines for Reporting Scale Development and Validation Results in the Journal of the Society for Social Work and Research. Journal of the Society for Social Work and Research, 1(2), 99–103. 10.5243/jsswr.2010.8

Carandang K, Vigen CLP, Ortiz E, & Pyatak EA (2020). Re-conceptualizing functional status through experiences of young adults with inflammatory arthritis. Rheumatology International, 40(2), 273–282. 10.1007/s00296-019-04368-8 [PubMed: 31300847]

Centers for Disease Control and Prevention. (2020). National Diabetes Statistics Report 2020. Estimates of diabetes and its burden in the United States. 32.

Chari R, Chang C-C, Sauter SL, Sayers ELP, Cerully JL, Schulte P, Schill AL, & Uscher-Pines L. (2018). Expanding the Paradigm of Occupational Safety and Health A New Framework for Worker Well-Being. Journal of Occupational and Environmental Medicine, 60(7), 589–593. 10.1097/JOM.0000000000001330 [PubMed: 29608542]

Cohen J. (2013). Statistical power analysis for the behavioral sciences. Academic press.

Cursio JF, Mermelstein RJ, & Hedeker D. (2019). Latent trait shared-parameter mixed models for missing ecological momentary assessment data. Statistics in Medicine, 38(4), 660–673. 10.1002/sim.7989 [PubMed: 30318637]

Dich N, Lange T, Head J, & Rod NH (2015). Work Stress, Caregiving and Allostatic Load: Prospective results from Whitehall II cohort study. Psychosomatic Medicine, 77(5), 539–547. 10.1097/PSY.0000000000000191 [PubMed: 25984826]

DiStefano C, Zhu M, & Mîndril D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. 14(20), 11.

Dixon A, Hibbard J, & Tusler M. (2009). How do People with Different Levels of Activation Self-Manage their Chronic Conditions? The Patient: Patient-Centered Outcomes Research, 2(4), 257–268. 10.2165/11313790-000000000-00000 [PubMed: 22273246]

Draelos MT, Jacobson AM, Weinger K, Widom B, Ryan CM, Finkelstein DM, & Simonson DC (1995). Cognitive function in patients with insulin-dependent diabetes mellitus during hyperglycemia and hypoglycemia. The American Journal of Medicine, 98(2), 135–144. 10.1016/S0002-9343(99)80397-0 [PubMed: 7847430]

Ewing FM, Deary IJ, McCrimmon RJ, Strachan MW, & Frier BM (1998). Effect of acute hypoglycemia on visual information processing in adults with type 1 diabetes mellitus. Physiology & Behavior, 64(5), 653–660. [PubMed: 9817577]

Gamst-Klaussen T, Gudex C, & Olsen JA (2018). Exploring the causal and effect nature of EQ-5D dimensions: An application of confirmatory tetrad analysis and confirmatory factor analysis. Health and Quality of Life Outcomes, 16(1), 153. 10.1186/s12955-018-0975-y [PubMed: 30064432]

Greenhaus JH, & Beutell NJ (1985). Sources of conflict between work and family roles. Academy of Management Review, 10(1), 76–88.

Hansen UM, Cleal B, Willaing I, & Tjørnhøj-Thomsen T. (2018). Managing type 1 diabetes in the context of work life: A matter of containment. Social Science & Medicine, 219, 70–77. 10.1016/j.socscimed.2018.10.016 [PubMed: 30391872]

Hansen UM, Skinner T, Olesen K, & Willaing I. (2019). Diabetes Distress, Intentional Hyperglycemia at Work, and Glycemic Control Among Workers With Type 1 Diabetes. Diabetes Care, 42(5), 797–803. 10.2337/dc18-1426 [PubMed: 30765430]

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform, 42(2), 377–381. 10.1016/j.jbi.2008.08.010 [PubMed: 18929686]

Hart SG (2006). NASA-task load index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50, 904–908.

Hart SG, & Staveland LE (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Advances in Psychology (Vol. 52, pp. 139–183). Elsevier. 10.1016/S0166-4115(08)62386-9

Hernandez R, Pyatak EA, Vigen CLP, Jin H, Schneider S, Spruijt-Metz D, & Roll SC (2021). Understanding Worker Well-Being Relative to High-Workload and Recovery Activities across a

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Whole Day: Pilot Testing an Ecological Momentary Assessment Technique. International Journal of Environmental Research and Public Health, 18(19), 10354. 10.3390/ijerph181910354 [PubMed: 34639654]

Holden RJ, & Abebe E. (2021). Medication transitions: Vulnerable periods of change in need of human factors and ergonomics. Applied Ergonomics, 90, 103279. 10.1016/j.apergo.2020.103279 [PubMed: 33049545]

Holden RJ, Cornet VP, & Valdez RS (2020). Patient ergonomics: 10-year mapping review of patient-centered human factors. Applied Ergonomics, 82, 102972. 10.1016/j.apergo.2019.102972 [PubMed: 31654954]

Holden RJ, Schubert CC, & Mickelson RS (2015). The patient work system: An analysis of self-care performance barriers among elderly heart failure patients and their informal caregivers. Applied Ergonomics, 47, 133–150. 10.1016/j.apergo.2014.09.009 [PubMed: 25479983]

Hoonakker P, Carayon P, Gurses A, Brown R, McGuire K, Khunlertkit A, & Walker JM (2011). Measuring Workload of ICU Nurses with a Questionnaire Survey: The NASA Task Load Index (TLX). IIE Transactions on Healthcare Systems Engineering, 1(2), 131–143. 10.1080/19488300.2011.609524 [PubMed: 22773941]

Hooper D, Coughlan J, & Mullen MR (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. 6(1), 8.

Hu L, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55.

Huang FL (2017). Conducting Multilevel Confirmatory Factor Analysis Using R. 20.

Krupp LB, Alvarez LA, LaRocca NG, & Scheinberg LC (1988). Fatigue in Multiple Sclerosis. Archives of Neurology, 45(4), 435–437. 10.1001/archneur.1988.00520280085020 [PubMed: 3355400]

Lange AHD, Kompier M.a. J. , Taris TW, Geurts S.a. E. , Beckers DGJ, Houtman ILD, & Bongers PM (2009). A hard day's night: A longitudinal study on the relationships among job demands and job control, sleep quality and fatigue. Journal of Sleep Research, 18(3), 374–383. 10.1111/j.1365-2869.2009.00735.x [PubMed: 19493298]

Leiter MP, & Durup MJ (1996). Work, Home, and In-Between: A Longitudinal Study of Spillover. The Journal of Applied Behavioral Science, 32(1), 29–47. 10.1177/0021886396321002

Linton SJ (2001). Occupational Psychological Factors Increase the Risk for Back Pain: A Systematic Review. Journal of Occupational Rehabilitation, 11(1), 53–66. 10.1023/A:1016656225318 [PubMed: 11706777]

Luke SG (2017). Evaluating significance in linear mixed-effects models in R. Behavior Research Methods, 49(4), 1494–1502. 10.3758/s13428-016-0809-y [PubMed: 27620283]

Mansikka H, Virtanen K, & Harris D. (2019). Comparison of NASA-TLX scale, modified Cooper–Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. Ergonomics, 62(2), 246–254. 10.1080/00140139.2018.1471159 [PubMed: 29708054]

Meijman TF, & Mulder G. (1998). Psychological aspects of workload. In Handbook of work and organizational: Work psychology, Vol. 2, 2nd ed (pp. 5–33). Psychology Press/Erlbaum (UK) Taylor & Francis.

Merz EL, & Roesch SC (2011). Modeling trait and state variation using multilevel factor analysis with PANAS daily diary data. Journal of Research in Personality, 45(1), 2–9. 10.1016/j.jrp.2010.11.003 [PubMed: 21516166]

Noyes JM, & Bruneau DPJ (2007). A self-analysis of the NASA-TLX workload measure. Ergonomics, 50(4), 514–519. 10.1080/00140130701235232 [PubMed: 17575712]

Olds DM, & Clarke SP (2010). The effect of work hours on adverse events and errors in health care. Journal of Safety Research, 41(2), 153–162. 10.1016/j.jsr.2010.02.002 [PubMed: 20497801]

Pyatak EA, Carandang K, Vigen CLP, Blanchard J, Diaz J, Concha-Chavez A, Sequeira PA, Wood JR, Whittemore R, Spruijt-Metz D, & Peters AL (2018). Occupational Therapy Intervention Improves Glycemic Control and Quality of Life Among Young Adults With Diabetes: The Resilient, Empowered, Active Living with Diabetes (REAL Diabetes) Randomized Controlled Trial. Diabetes Care, 41(4), 696–704. 10.2337/dc17-1634 [PubMed: 29351961]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Pyatak EA, Hernandez R, Pham L, Mehdiyeva K, Schneider S, Peters A, Ruelas V, Crandall J, Lee P-J, Jin H, Hoogendoorn CJ, Crespo-Ramos G, Mendez-Rodriguez H, Harmel M, Walker M, Serafin-Dokhan S, Gonzalez JS, & Spruijt-Metz D. (2021). Function and Emotion in Everyday Life with Type 1 Diabetes (FEEL-T1D): A fully remote intensive longitudinal study of blood glucose, function, and emotional well-being in adults with type 1 diabetes. JMIR Research Protocols. 10.2196/30901

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/

Raghupathi W, & Raghupathi V. (2018). An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health. International Journal of Environmental Research and Public Health, 15(3). 10.3390/ijerph15030431

Roesch SC, Aldridge AA, Stocking SN, Villodas F, Leung Q, Bartley CE, & Black LJ (2010). Multilevel Factor Analysis and Structural Equation Modeling of Daily Diary Coping Data: Modeling Trait and State Variation. Multivariate Behavioral Research, 45(5), 767–789. 10.1080/00273171.2010.519276 [PubMed: 21399732]

Rook JW, & Zijlstra FRH (2006). The contribution of various types of activities to recovery. European Journal of Work and Organizational Psychology, 15(2), 218–240. 10.1080/13594320500513962

Rothman KJ (1990). No Adjustments Are Needed for Multiple Comparisons: Epidemiology, 1(1), 43–46. 10.1097/00001648-199001000-00010 [PubMed: 2081237]

Rutledge T, Stucky E, Dollarhide A, Shively M, Jain S, Wolfson T, Weinger MB, & Dresselhaus T. (2009). A real-time assessment of work stress in physicians and nurses. Health Psychology, 28(2), 194–200. 10.1037/a0013145 [PubMed: 19290711]

Sauter SL (2013). Integrative Approaches to Safeguarding the Health and Safety of Workers. Industrial Health, 51(6), 559–561. 10.2486/indhealth.MS5106ED [PubMed: 24292810]

Saville DJ (1990). Multiple Comparison Procedures: The Practical Solution. 8.

Scott LD, Hwang W-T, & Rogers AE (2006). The Impact of Multiple Care Giving Roles on Fatigue, Stress, and Work Performance Among Hospital Staff Nurses: JONA: The Journal of Nursing Administration, 36(2), 86–95. 10.1097/00005110-200602000-00007

Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, & Lipton RB (2018). Reliability and Validity of Ambulatory Cognitive Assessments. Assessment, 25(1), 14–30. 10.1177/1073191116643164 [PubMed: 27084835]

Sonnenberg B, Riediger M, Wrzus C, & Wagner GG (2011). Measuring Time Use in Surveys – How Valid are Time Use Questions in Surveys? Concordance of Survey and Experience Sampling Measures. SSRN Electronic Journal. 10.2139/ssrn.1895307

Sonnentag S, & Fritz C. (2007). The Recovery Experience Questionnaire: Development and validation of a measure for assessing recuperation and unwinding from work. Journal of Occupational Health Psychology, 12(3), 204–221. 10.1037/1076-8998.12.3.204 [PubMed: 17638488]

Taber KS (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Research in Science Education, 48(6), 1273–1296. 10.1007/s11165-016-9602-2

Tindle J, & Tadi P. (2021). Neuroanatomy, Parasympathetic Nervous System. In StatPearls. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK553141/

Tubbs-Cooley HL, Mara CA, Carle AC, & Gurses AP (2018). The NASA Task Load Index as a measure of overall workload among neonatal, paediatric and adult intensive care nurses. Intensive and Critical Care Nursing, 46, 64–69. 10.1016/j.iccn.2018.01.004 [PubMed: 29449130]

Tudor-Locke C, Washington TL, Ainsworth BE, & Troiano RP (2009). Linking the American Time Use Survey (ATUS) and the compendium of physical activities: Methods and rationale. Journal of Physical Activity and Health, 6(3), 347–353. [PubMed: 19564664]

Ursachi G, Horodnic IA, & Zait A. (2015). How Reliable are Measurement Scales? External Factors with Indirect Influence on Reliability Estimators. Procedia Economics and Finance, 20, 679–686. 10.1016/S2212-5671(15)00123-9

Valdez RS, Holden RJ, Caine K, Madathil K, Mickelson R, Lovett Novak L, & Werner N. (2016). Patient Work as a Maturing Approach Within HF/E: Moving Beyond Traditional Self-Management

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Applications. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1), 657–661. 10.1177/1541931213601151

van Griethuijsen RALF, van Eijck MW, Haste H, den Brok PJ, Skinner NC, Mansour N, Savran Gencer A, & BouJaoude S. (2015). Global Patterns in Students' Views of Science and Interest in Science. Research in Science Education, 45(4), 581–603. 10.1007/s11165-014-9438-6

Virtanen M, Singh-Manoux A, Ferrie JE, Gimeno D, Marmot MG, Elovainio M, Jokela M, Vahtera J, & Kivimäki M. (2009). Long Working Hours and Cognitive Function. American Journal of Epidemiology, 169(5), 596–605. 10.1093/aje/kwn382 [PubMed: 19126590]

Ziaei M, Yarmohammadi H, Moradi M, & Khandan M. (2015). Level of Workload and Its Relationship with Job Burnout among Administrative Staff. 8.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

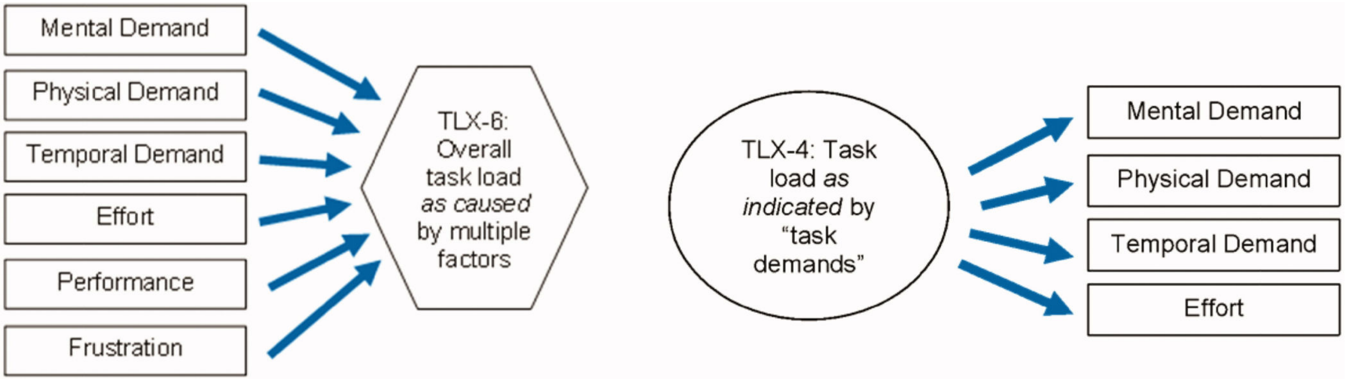

Figure 1.
Models for TLX-6 and TLX-4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

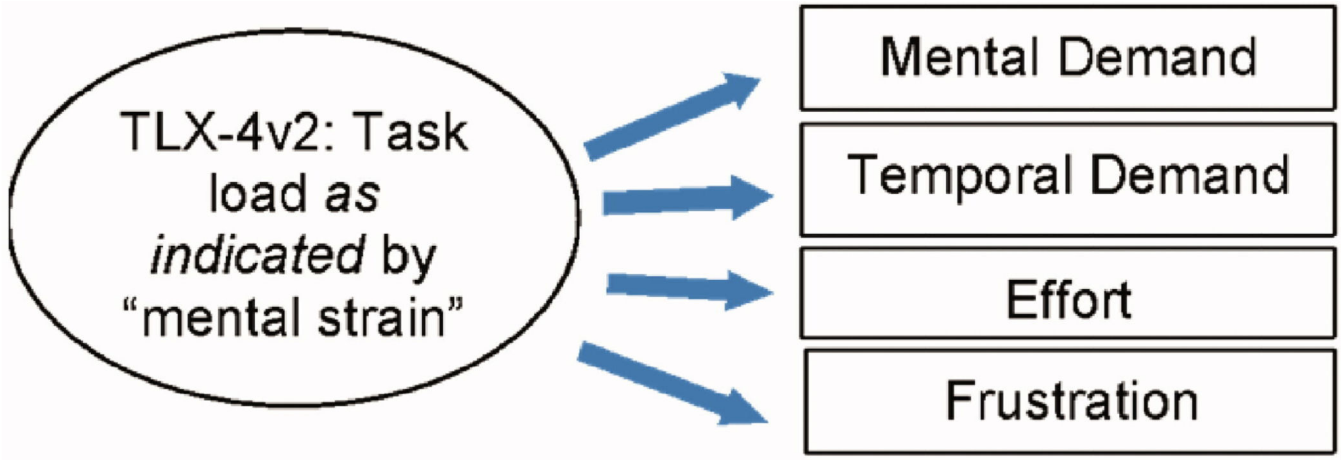

Figure 2.
TLX-4 version 2 (TLX-4v2) that was found in exploratory analysis to have acceptable single factor model fit metrics.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

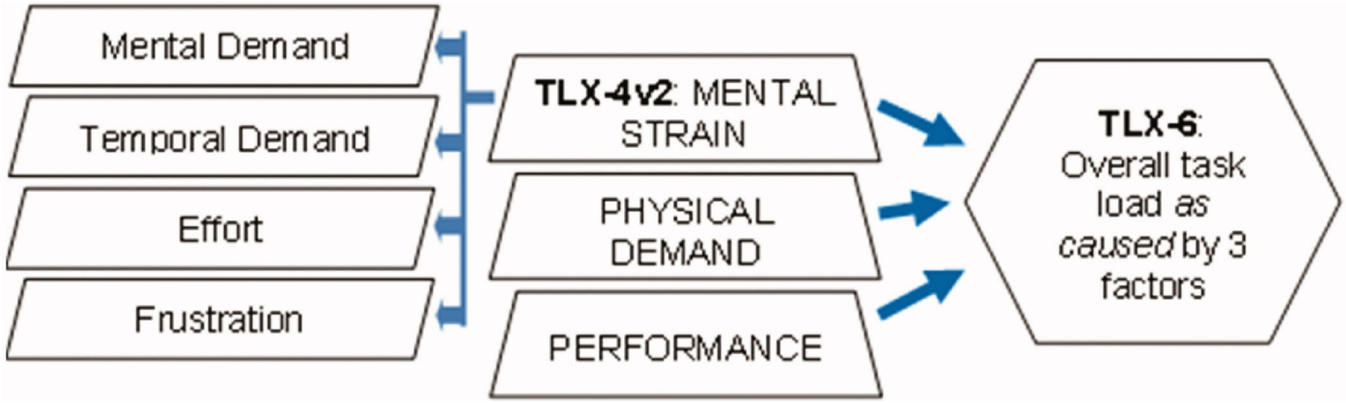

Figure 3.
Relationship between TLX-4 and TLX-6 implied by CFA results

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1:

Ecological momentary assessment measures administered over up to 14 days of data collection.

Construct	Item(s)	Response Option(s)	Time/Frequency
ACTIVITY EXPOSURE			
Workload	TLX-6 Mental demand: How much mental activity was required for your whole day? (thinking, deciding, remembering, etc.) Physical demand: How much physical activity was required for your whole day? (e.g., pushing, pulling, walking, etc.) Time Pressure: How much time pressure did you feel from activities over your whole day? Effort: How hard did you have to work (with your body or your mind) over your whole day? Performance: How pleased were you with your performance of activities over your whole day? Frustration level: How frustrated were you from activities over your whole day?	0 to 100 sliding scale for each item	End of day
Work hours	(If worked) About how many hours did you work?	Hours	End of day
Activity engaged in	What were you doing right before starting this survey?	Work/school, traveling, relaxing/chilling, sleeping/napping, socializing, caring for myself, caring for others, doing housework/errands, fun/play/leisure, other	All survey times
Strenuous (demanding) activity frequency	Proportion of times in a day “work/school” or “caring for others” were answered to the question “What were you doing right before starting this survey?”	N/A	All survey times
Restful activity frequency	Proportion of times in a day “relaxing/chilling” or “sleeping/napping” were answered to the question “What were you doing right before starting this survey?”	N/A	All survey times
Caregiving frequency	Proportion of times in a day “caring for others” was answered to the question “What were you doing right before starting this survey?”	N/A	All survey times
COGNITIVE PERFORMANCE			
Sustained attention ability	Go No Go test	Tapping picture of cities on a smartphone	All survey times
Processing speed	Symbol Search	Tapping matching images on a smartphone	All survey times
HEALTH-RELATED OUTCOMES			
Fatigue	At this moment, how tired do you feel?	0 (Not at all) to 100 (Extremely)	All survey times
Negative affect	Sum of mood ratings for “tense”, “upset”, “sad”, “disappointed”	For each mood, 0 (not at all) to 100 (extremely)	All survey times
Positive affect	Sum of mood ratings for “happy”, “content”, “enthusiastic”, “excited”	Same as above	All survey times
Stress	How stressed are you right now?	0 (Not at all stressed) to 100 (Extremely stressed)	All survey times
Diabetes-related stress	How stressed do you feel about your diabetes or diabetes management right now?	0 (Not at all stressed) to 100 (Extremely stressed)	All survey times
Pain	At this moment, how much bodily pain do you have?	0 (None) to 100 (Extreme pain)	All survey times
Sleep quality	How well rested do you feel?	0 (Not at all rested) to 100 (Extremely rested)	First (morning) survey

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2:

Hypotheses tested to investigate validity of the TLX as a measure of whole day workload.

Tests	Hypotheses
1) Confirmatory factor analysis of *unidimensional* model to test factor structure	Good fit metrics with TLX-4, poor fit with TLX-6
2) Within-person correlation tests with the measures below (convergent validity):	Note: We expect minor differences in magnitude here between TLX-4 and TLX-6, but same directions.
ACTIVITY EXPOSURE	
Strenuous Activity Frequency	Positive association
Restful Activity Frequency	Negative association
Caregiving Frequency	Positive association
Reported Work Hours	Positive association
COGNITIVE PERFORMANCE	
Go No Go Outcome (dprime), next day average	Negative association
Symbol Search, next day average	Negative association
Go No Go (dprime), same day average	Negative association (small magnitude)
Symbol Search, same day average	Negative association (small magnitude)
Go No Go (dprime), last survey of day	Negative association
Symbol Search, last survey of day	Negative association
HEALTH-RELATED OUTCOMES	
Fatigue	Positive association
Negative Affect	Positive association
Positive Affect	Negative association
Stress	Positive association
Sleep quality (morning, same day as TLX measure)	Negative association
Sleep quality (morning, day after TLX measure)	Negative association
Pain	Positive association
Diabetes Stress	Positive association
Various blood glucose (BG) measures	Some association with the BG measures, though difficult to predict a direction because research in this area is ongoing
3) Mixed modeling to test the association with days of the week (convergent validity)	Lower task load on weekends relative to weekdays
4) Mixed modeling to test the association of work versus non-work days (convergent validity)	Lower task load on non-work days

TLX- task load index

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Single Factor CFA Model Fit Statistics

TLX Version	CFI	TLI	RMSEA	SRMR	AIC	BIC	χ^2 (p-value)
TLX-6	0.857	0.761	0.130	Within:0.078[*]	33081.467	33134.929	106.23 (p<.001)
TLX-4	0.973[*]	0.920[*]	0.104	Within:0.037[*]	22154.727	22190.368	15.813 (p<.001)
TLX-4v2	0.988[*]	0.965[*]	.067[*]	Within:0.024[*]	21962.685[**]	21998.327[**]	7.760 (p=.021)

AIC- Akaike information criterion; BIC- Bayesian information criterion; CFI- comparative fit index; RMSEA- root mean square error of approximation; SRMR- standardized root mean squared residual; TLX- task load index; Tucker-Lewis index

[*] Fit statistics are in the recommended range. The standards for good fit are: CFI and TLI at least >.90 but ideally >.95, RMSEA at least <.08 but ideally <.05, SRMR<.08

[**] Lower values of AIC or BIC relative to TLX-6 and TLX-4 indicative of better fit

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Standardized factor loading estimates (within-person)

	Item	Standardized Estimate	Standard Error
TLX-6	Mental	.600	.032
	Physical	.335	.039
	Temporal (pressure)	.584	.032
	Effort	.913	.028
	Performance	−.256	.040
	Frustration	.273	.040
TLX-4	Mental	.579	.034
	Physical	.329	.039
	Temporal (pressure)	.563	.034
	Effort	.946	.034
TLX-4v2	Mental	.639	.032
	Temporal (pressure)	.632	.032
	Effort	.840	.030
	Frustration	.313	.041

TLX- task load index

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5:

Within-person correlations between TLX sums and other daily measures. The "hypothesis aligned" column indicates if results are consistent with table 2 hypotheses (yes or no).

Other Measure	TLX-6				TLX-4 v2			
	R	**p**	**95%CI**	**Hypoth. aligned**	**r**	**p**	**95%CI**	**Hypoth. aligned**
ACTIVITY EXPOSURE								
Strenuous Activity Frequency	0.40	p<.001*	(0.33, 0.47)	Y	0.46	p<.001*	(0.39, 0.52)	Y
Restful Activity Frequency	−0.32	p<.001*	(−0.39, −0.25)	Y	−0.31	p<.001*	(−0.38, −0.23)	Y
Caregiving Frequency	.08	.046	(.01, .16)	Y	.06	.120	(−.02, .15)	N
Work Hours	0.28	p<.001*	(0.18, 0.38)	Y	0.31	p<.001*	(0.21, 0.41)	Y
COGNITIVE PERFORMANCE								
Go No Go, next day average	0.07	0.120	(−0.02, 0.15)	N	0.06	0.160	(−0.02, 0.14)	N
Symbol Search, next day average	−0.04	0.326	(−.12,.04)	N	0.01	0.901	(−0.08, 0.09)	N
Go No Go, same day average	−0.01	0.887	(−0.09, 0.08)	N	0	0.938	(−0.09, 0.08)	N
Symbol Search, same day average	−0.01	0.749	(−0.1, 0.07)	N	−0.01	0.901	(−0.09, 0.08)	N
Go No Go, last survey of day	−0.07	0.071	(−0.15, 0.01)	N	−0.06	0.120	(−0.14, 0.02)	N
Symbol Search, last survey of day	−0.03	0.52	(−0.11, 0.05)	N	−0.03	0.465	(−0.11, 0.05)	N
HEALTH RELATED OUTCOMES								
Fatigue	0.17	p<.001*	(0.09, 0.25)	Y	0.16	p<.001*	(0.08, 0.24)	Y
Negative Affect	0.32	p<.001*	(0.24, 0.39)	Y	0.30	p<.001*	(0.22, 0.37)	Y
Positive Affect	−0.08	0.065	(−0.16, 0.01)	N	−0.07	0.082	(−0.15, 0.01)	N
Stress	0.43	p<.001*	(0.36, 0.49)	Y	0.43	p<.001*	(0.36, 0.49)	Y
Sleep Quality (morning, same day as TLX measure)	0	0.98	(−0.08, 0.09)	N	0.02	0.607	(−0.06, 0.11)	N
Sleep Quality (morning, day after TLX measure)	.02	.640	(−.06, .10)	N	.03	.491	(−.05, .11)	N
Pain	0.18	p<.001*	(0.1, 0.26)	Y	0.15	p<.001*	(0.07, 0.23)	Y
Diabetes Stress	0.18	p<.001*	(0.1, 0.26)	Y	0.16	p<.001*	(0.08, 0.24)	Y
Mean BG	−.14	.020*	(−.23,−.05)	Y**	−.09	.041*	(−.18, −.004)	Y**
SD BG	−.02	.741	(−.11, .08)		−.03	.458	(−.12, .06)	
CV BG	.07	.111	(−.02, .16)		03	.514	(−.06, .12)	
Time in range (BG)	.05	.242	(−.04, .14)		.02	.606	(−.07, .11)	
Time high (BG)	−.09	.047*	(−.18, −.001)		−.05	.283	(−.14, .04)	
Time low (BG)	.08	.082	(−01, .17)		.05	.237	(−.04, .14)	

BG- blood glucose; CV- coefficient of variation; N-no; SD- standard deviation; TLX- task load index; Y-yes

*p<.05

[**] Some association was hypothesized with BG metrics, so the correlation with “mean BG” is consistent with this.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 6:

Betas for Days of the Week with TLX sums as outcomes

Day	# obs	TLX-6 Means	Betas with TLX-6 Outcome (95% CI)	TLX-4 v2 Means	Betas with TLX-4 v2 Outcome (95% CI)
Sunday	100	37.16	Reference	36.24	Reference
Monday	105	42.55	5.29, (2.18, 7.92) *	44.23	8.1, (4.4, 12.36) *
Tuesday	115	45.43	7.76, (4.98, 10.69) *	48.57	11.83, (7.48, 16.09) *
Wednesday	114	46.16	8.87, (5.84, 11.79) *	49.17	13.06, (9.02, 16.77) *
Thursday	114	46.53	9.45, (6.59, 12.69) *	49.91	13.79, (9.75, 17.79) *
Friday	122	46.23	8.72, (5.64, 11.69) *	49.22	12.8, (9.13, 16.76) *
Saturday	100	39.11	2.31, (−0.59, 5.44)	38.65	3.3, (−0.72, 7.41)
No work reported	435	38.79	Reference	38.08	Reference
Workday	335	49.06	9.29, (7.37, 11.06) *	54.10	14.59, (12.33, 17.06) *

*CIs don’t contain 0, indicating statistical significance

TLX- task load index