

日本語版NASA-TLXによるメンタルワークロード測定* —各種室内実験課題の困難度に対するワークロード得点の感度—

芳賀 繁**, 水上直樹***

Workload of three different types of laboratory tasks, each of which has three difficulty levels, was measured with the Japanese version of NASA Task Load Index (TLX), and the result demonstrated that the obtained workload scores were sensitive to the task difficulty in any of the tasks. The weights of the rating scales, which were decided through the pairwise comparison of the importance of the scales, were found to be unique to each task. This indicates that the three experimental tasks had different constitution of loading factors from each other, and that TLX is applicable to various types of workload. Then, correlation and regression analyses were carried out and the results were compared with those reported for the original NASA-TLX. Finally, workload scores which were obtained with no pairwise comparison from the ratings on the same TLX scales were compared with the regular Weighted Workload (WWL) scores, and their validity as alternative indices of workload was discussed.

日本語版 NASA-TLX を用いて 3 種類の室内実験課題のワークロード測定を行ったところ、いずれの課題においても、困難度の変化に感度よく対応した TLX 得点が得られた。また、尺度の重要度に関する一対比較から算出される尺度の重みが課題ごとに異なるパターンを示すことから、3 つの実験課題は作業負荷の面で相互に性質の異なるものであったことが検証され、TLX 得点が性質の異なる種々の作業に適用可能であることが確認された。さらに、相関分析や重回帰分析を行って、その結果をオリジナル版の NASA-TLX に関する報告と比較した。最後に、一対比較を省略した簡便法から得られる各種指標を推定したところ、被験者間の変動は正規の NASA-TLX 得点である WWL と同様またはむしろやや小さく、WWL や“全体的負荷”との相関も高いことが明らかになった。

(キーワード：作業負担、精神的負荷、主観的評価尺度、トラッキング、室内実験)

1. はじめに

メンタルワークロードの測定を目指して様々な手法が研究、開発されているなかであって、NASA-TLX (NASA Task Load Index)¹⁾ は、すでに実用化を達成した数少ない手法のひとつであり、SWAT (Subjective Workload Assessment Technique)²⁾ と並んで、最も広く使われている主観的作業負荷評価尺度である^{3,4)}。芳賀⁵⁾ は、これを日本でも利用できるように日本語版を作成したうえ、パソコン画面上で尺度の一対比較と、作業負荷評定をするプログラムを開発した。

しかしながら、三宅と神代⁶⁾ が指摘しているとおり、翻訳版はオリジナルの尺度で確認された妥当性、信頼性をそのまま有しているとはいえないので、それを改めて確認する必要がある。芳賀⁵⁾ はすでに、上記の日本語版がトラッキング課題の困難度に対して感度が高いこと、下位尺度どうしの独立性が高いこと、算出されたワークロード得点と“全体的負荷”の評定値との間

翻訳版はオリジナルの尺度で確認された妥当性、信頼性をそのまま有しているとはいえないので、それを改めて確認する必要がある。芳賀⁵⁾ はすでに、上記の日本語版がトラッキング課題の困難度に対して感度が高いこと、下位尺度どうしの独立性が高いこと、算出されたワークロード得点と“全体的負荷”の評定値との間

* 1995 年 6 月 29 日受付

** 東和大学
Tohwa Univ.

*** 財団法人鉄道総合技術研究所
Railway Technical Research Institute.

の相関が高いことなどを報告した。この“日本語版 NASA-TLX”が、異なるタイプの作業負荷に対して、より一般的に妥当性をもっていることを検証するために、今回はトラッキングに加えて、視覚探索と暗算を作業課題として同様の結果が得られるか否かを調べることにした。さらに、一対比較を省略した簡便法によって得られる指標の有効性についても検討した。

なお、ISO 6385においては、“作業負荷”(work stress)は、作業システムにおいて人の生理的・心理的状态を乱すように作用する外的条件や要求の総量、“作業負担”(work strain)は、作業負荷が個人の特性や能力と関連して与える影響と定義されている⁷⁾。つまり、作業負荷は作業の側の特性で、作業負担は作業側の反応である。その後制定されたISO 10075においても、“精神的作業負荷”(mental stress)と“精神的作業負担”(mental strain)について同様の定義と日本語訳が行われた⁸⁾。そこで、本論文では、できるだけこの定義に沿って“負荷”と“負担”の語を使い分けるよう努力するが、どちらも含む概念を表したい場合や、どちらかにはっきり分けられない場合には“ワークロード”の語を用いる。また、引用した文献で使われている語はそのまま用いた場合もある。NASA-TLXの開発者は、負荷の要素が特に強い場合に“demand”という語を用いる以外は、ほとんど負荷と負担を区別せずに“workload”で通しているからである。

主観的ワークロード評価の場合は、作業負荷の程度を評価するのか、感じた負担の程度を答えるのかがあいまいになりがちであるが、被験者にこれを厳密に区別して答えるよう求めるのにも無理がある。今回用いた日本語版NASA-TLXにおいても、尺度によって、負荷の側の要素の強いものと負担の評価を求めるものが混じっている。しかし、最終的な目的はあくまで負荷、すなわち作業システムの評価にある。

2. 日本語版 NASA-TLX

NASA-TLXは、知的・知覚的要求(mental demand: MD)、身体的要求(physical demand: PD)、タイムプレッシャー(temporal demand: TD)、作業成績(own performance: OP)、努力(effort: EF)、フラストレーション(frustration: FR)の6つの尺度から構成されている。被験者はワークロードを評価する前に、まず、これらの尺度の重要度を一対比較法によって判定することを求められる。6つの尺度の組み合わせは15対できるので、被験者は15回の比較判断を行い、そのなかで何回“より重要”な

ものとして選択されたかを数えて各尺度の重みとする。今回用いた芳賀⁹⁾のパソコンプログラムでは、提示する尺度の対の左右を逆にして再度比較判断を求め、1回目と2回目で判定が分かれた対については、3回目の提示で結果を確定する。

作業のワークロードを評価する際には、被験者は6つの尺度ごとに両極の間に引かれた線分のうちの適当な位置に印をつける。パソコンプログラムでは1画面に1つの尺度をランダム順に選んで提示し、下に説明文を表示し、被験者はマウスを使って線分上のポイントを動かしたうえボタンをクリックして評価を入力する。ポイントの位置は0~100の評定点に置き換えられて記録される。尺度ごとの評価点に、一対比較によって定められた尺度ごとの重みを乗じて平均したものが加重平均ワークロードスコア(Mean Weighted Workload Score)、またはWWL得点である。パソコン画面に表示した尺度名、終端点(両極)、説明文は表1のとおりである。

なお、パソコンプログラムではTLXの6尺度の後に“全体的負荷”(Overall Workload: OW)という尺度を提示して同様に評価を求め、記録した。“全体的負荷”は、他の尺度から合成可能なことなどの理由によりTLXの開発段階で尺度から除かれたものであるが、日本語版は、いわば開発段階にあるので、他の尺度の評定値や、合成されたWWL得点との相関などを調べるために用いた。表にあるとおり、その説明文は“負荷”というより“負担”を聞くものとなっている。これは先行研究⁹⁾において“workload”の訳語として一般的な“負荷”を採用したためで、前述した用語法に従うなら、“全体的負担度”とでも訳すべきであったかもしれない。

3. 作業課題

CRTディスプレイに刺激を提示する3種類の実験課題を用意した(図1)。

3-1. 視覚探索課題

画面上部に横長の窓、中央から下部にかけて正方形の窓が描画されている。実験が始まると、まず上の窓に標的のアルファベットが提示され10秒後から下の窓に4×4に配列された16文字が3秒ごとに次々と表示される(図1a)。被験者の課題は、配列のなかに標的が1つでもあるか1つもないかを判断し、マウスの左右のボタンで答えることである。配列が10回提示された後、新しい標的が変わって作業が繰り返される。作業の困難度は標的文字の数で操作した。すなわち、

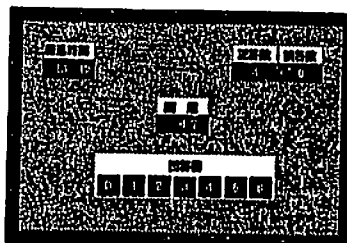
表1 パソコン化された日本語版 NASA-TLX における尺度名とその説明文

Tab.1 Japanese titles of the NASA-TLX scales, their endpoints, and their descriptions.

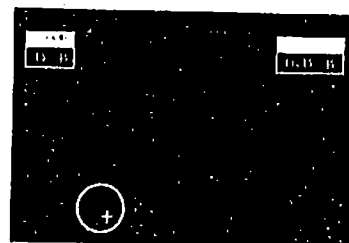
尺度名(終端点)	説明文(上段は対比較, 下段は負担評価の際に用いる)
知的・知覚的要求 (小さい/大きい)	どの程度の知的・知覚的活動(考える, 決める, 計算する, 記憶する, 見るなど)を必要とするか。課題がやさしいか難しいか, 単純か複雑か, 正確さが求められるか大ざっぱでよいか どの程度の知的・知覚的活動(考える, 決める, 計算する, 記憶する, 見るなど)を必要としましたか。課題はやさしかったですか難しかったですか, 単純でしたか複雑でしたか, 正確さが求められましたか大ざっぱでよかったですか
身体的要求 (小さい/大きい)	どの程度の身体的活動(押す, 引く, 回す, 制御する, 動き回るなど)を必要とするか。作業がラクかキツイか, ゆっくりできるかキビキビやらなければならないか, 休み休みできるか働きづめか どの程度の身体的活動(押す, 引く, 回す, 制御する, 動き回るなど)を必要としましたか。作業はラクでしたかキツかったですか, ゆっくりできましたかキビキビやらなければなりませんでしたが, 休み休みできましたか働きづめでしたか
タイムプレッシャー (弱い/強い)	仕事のペースや課題が発生する頻度のために感じる時間的切迫感がどの程度か。ペースはゆっくりとして余裕があるものか, それとも速くて余裕のないものか 仕事のペースや課題が発生する頻度のために感じる時間的切迫感はどの程度でしたか。ペースはゆっくりとして余裕があるものでしたか, それとも速くて余裕のないものでしたか
作業成績 (良い/悪い)	作業指示者(またはあなた自身)によって設定された課題の目標をどの程度達成できたか。目標の達成に関して自分の作業成績にどの程度満足しているか 作業指示者(またはあなた自身)によって設定された課題の目標をどの程度達成できたか。目標の達成に関して自分の作業成績にどの程度満足していますか
努力 (少ない/多い)	作業成績のレベルを達成・維持するために, 精神的・身体的にどの程度いっしょうけんめいに作業しなければならぬか 作業成績のレベルを達成・維持するために, 精神的・身体的にどの程度いっしょうけんめいに作業しなければならぬものでしたか。
フラストレーション (低い/高い)	作業中に, 不安感, 落胆, いらいら, ストレス, 悩みをどの程度感じるか。あるいは逆に, 安心感, 満足感, 充足感, 楽しさ, リラックスをどの程度感じるか 作業中に, 不安感, 落胆, いらいら, ストレス, 悩みをどの程度感じましたか。あるいは逆に, 安心感, 満足感, 充足感, 楽しさ, リラックスをどの程度感じましたか
全体的な負荷 (低い/高い)	さまざまな負荷要因, 負荷原因, 部分部分の課題内容を総合すると, 全体としてどの程度の作業負担を感じましたか



(a) 視覚探索課題



(b) 暗算課題



(c) トラッキング課題

図1 作業課題の刺激提示画面

Fig.1 Computer display layouts for the experimental tasks.

a : visual search, b : mental arithmetic, c : tracking.

L (易) 条件が1, M (中困難度) 条件が2, H (難) 条件が4である。

3-2. 暗算課題

画面中央に数字が提示され、被験者はそれを7で割った剰余を計算して、下部の回答欄に表示されている0～6のいずれかの枠にマウスカーソル(矢印)をもっていき左クリックする(図1b)。正答の場合は澄んだピープ音とともに画面右上の“正答数”の表示が1つ増え、次の問題が提示される。誤答の場合は濁った音がして“誤答数”が1つ増え、問題はそのまま残る。被験者は正答が得られるまで計算をやり直す。困難度条件は、割り算をしたときの商が1桁(問題は7～69の範囲=L条件), 2桁(70～699=M条件), 3桁(700～6999=H条件)により定義された。なお、被験者には“できるだけ速く、しかしできるだけ間違えないよう回答する”旨が教示された。

3-3. トラッキング課題

Microsoftのフライトシミュレータ用に開発された“Mouse York”(Colorado Spectrum製)という操縦桿様の器具を用いて、画面上を動き回る円のなかに十字形のカーソルを留めていることが課題である(図1c)。操縦桿はハンドルを回すとカーソルが左右に動き、軸を水平方向に押し込むと上、引くと下にカーソルが動く。円は斜めに運動するので、この2種類の動作を協調させて操作しなければカーソルは円からはみ出してしまふ。カーソルが円の外に出ている間、カーソルの色は白から赤に変わり、ピープ音が鳴り続ける。困難度は円の直径と動きの速さで3段階(L, M, H条件)に設定した。

4. 実験手続き

被験者は22～31歳の男女12名で、日を替えて3種類すべての課題を行った。1回の実験セッションでは、まずM条件のもとで課題の練習を3分間行い、次にTLX尺度の重みづけのための対比較を行い、その後被験者間でカウンタバランスされた順序で3つの困難度条件のもとでの作業を5分(視覚探索課題は5分20秒)ずつ行った。各作業後に、負担度についてTLXの6尺度による評価と“全体的負担”の評価を求めた。なお、一部の被験者には心拍と呼吸のデータをとるための電極、呼吸サーミスタなどを装着したが、生理データの分析結果は別報で報告する。

5. 結果と考察

5-1. 作業成績

各課題の困難度段階ごとの作業成績の平均値を、表2～4に示す。各課題の作業成績を表す代表的な指標は、視覚探索課題では正答率、暗算課題では正答数、トラッキング課題では“はみだし時間”と考えられるが、表にみられるとおり、どの課題も困難度の上昇に伴って作業成績は有意に低下している(分散分析, $p < .01$)。

5-2. WWL得点

被験者によるワークロード評価(WWL得点)は図2(個人ごと)および図3(平均値)のとおりであり、どの課題においても有意な条件差が見いだされた(分散分析, $p < .01$)。

表2 視覚探索課題の作業成績(12名の平均値, カッコ内は標準偏差)

Tab.2 Visual search task performance; mean (number of subjects=12) and standard deviation (in parentheses).

条 件	ヒット*	正棄却*	誤警報*	ミ ス*	無回答	正答率(%)**
L	35.33 (4.55)	41.58 (5.31)	0.17 (0.37)	2.75 (1.53)	0.17 (0.55)	96.15 (2.31)
M	33.08 (5.51)	38.83 (5.26)	0.17 (0.37)	7.17 (3.67)	0.75 (1.30)	89.90 (4.49)
H	20.25 (6.44)	35.83 (7.01)	1.75 (1.53)	15.42 (3.82)	6.75 (5.57)	70.10 (9.15)

* ヒット, 正棄却, 誤警報, ミスは, それぞれ標的があるときに“ある”, 標的がないときに“ない”, 標的がないのに“ある”, 標的があるのに“ない”と反応したことを表す。

** 正答率(%) = $100 \times (\text{ヒット} + \text{正棄却}) / 80$ 試行

表3 暗算課題の作業成績(12名の平均値, カッコ内は標準偏差)

Tab.3 Mental arithmetic task performance ; mean (number of subjects=12) and standard deviation (in parentheses).

条 件	正答数	誤答数	試行数	正答率(%)
L	134.00 (32.23)	4.58 (3.99)	138.58 (31.03)	96.34 (3.20)
M	65.92 (19.59)	4.67 (2.49)	70.58 (20.21)	93.25 (3.79)
H	42.92 (11.57)	3.92 (2.25)	46.83 (11.25)	91.13 (5.16)

表4 トラッキング課題の作業成績(12名の平均値, カッコ内は標準偏差)

Tab.4 Tracking task performance ; mean (number of subjects=12) and standard deviation (in parentheses).

条 件	はみだし時間(秒)	追跡率(%)*
L	6.00 (5.42)	98.00 (1.81)
M	15.29 (13.51)	94.90 (4.50)
H	27.96 (19.84)	90.68 (6.61)

* 追跡率(%) = $100 \times (300 - \text{はみだし時間}) / 300$

TLX 尺度のひとつに“作業成績”があり, 今回はどの課題においても成績に条件差があったことがWWL 得点の差に直結したのではないかと懸念される。そこで, “作業成績”の尺度を除いた5尺度から得点(尺度の重みはそのまま使用)を算出して再度検定してみたところ, なお有意な条件差が検出された(分

散分析, $p < .01$).

5-3. 尺度の重みと尺度選択の妥当性

一対比較の結果から決められた尺度の重みを被験者間で平均すると, 図4のとおりとなった。視覚探索課題で最大の重みがつけられた“タイムプレッシャー”

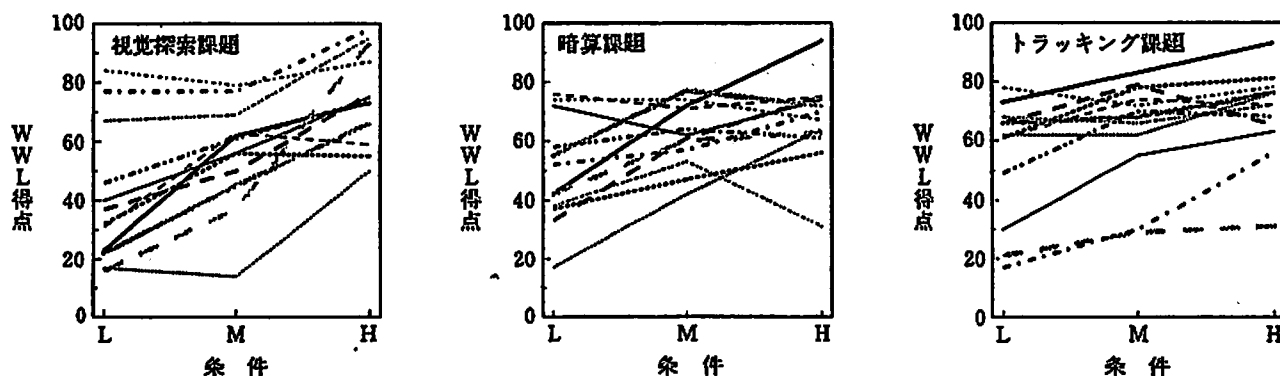


図2 各課題の各作業負荷条件に対する12名の被験者のWWL得点

Fig.2 Weighted Workload Scores calculated from the ratings by 12 subjects for the different levels of workload (Low/Medium/High) in the three different tasks.

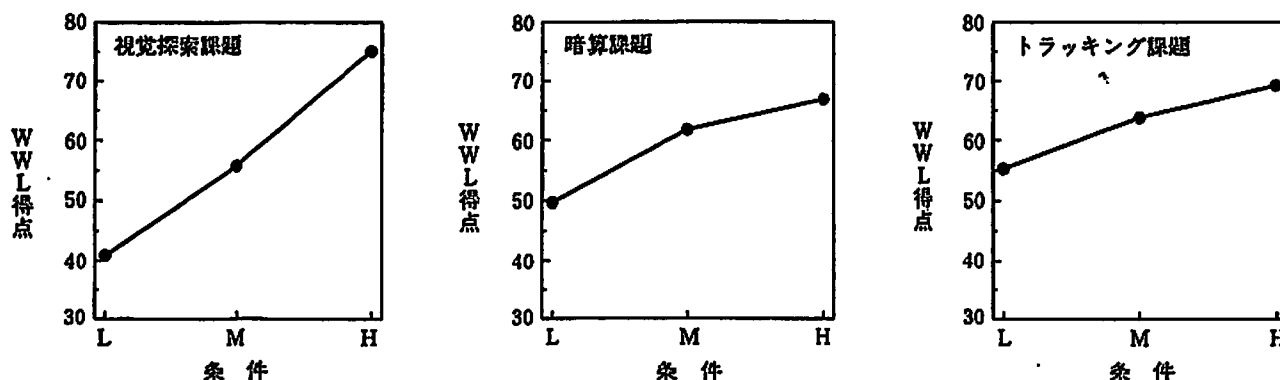


図3 各課題の各作業負荷条件に対するWWL得点の平均値(被験者各12名)

Fig.3 Weighted Workload Scores averaged over 12 subjects for the different levels of workload (Low/Medium/High) in the three different tasks.

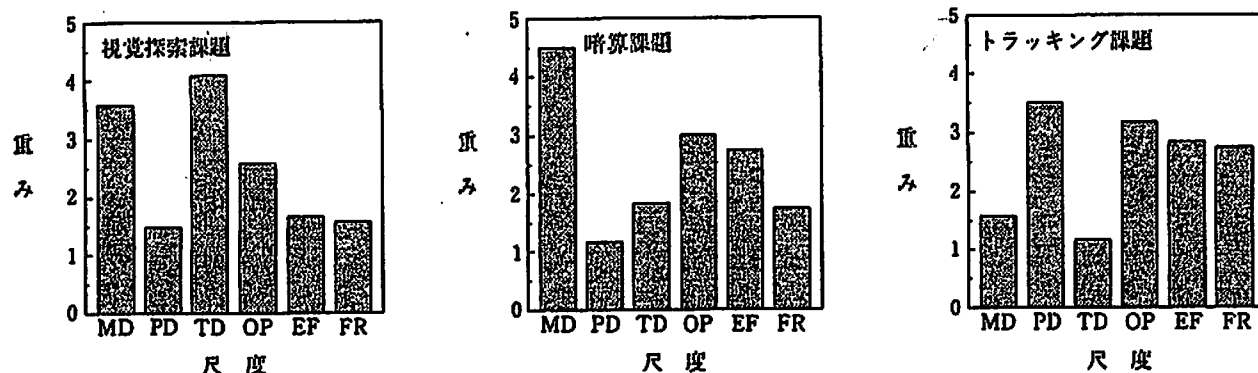


図4 一対比較の結果、各尺度につけられた重みの平均値。MD：知的・知覚的要求、PD：身体的要求、TD：タイムプレッシャー、OP：作業成績、EF：努力、FR：フラストレーション

Fig. 4 Weights given for the six TLX scales: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Own Performance (OP), Effort (EF), and Frustration (FR). The weights shown in the graph are averaged over 12 subjects.

は、他の2課題では小さな重みしかもたず、トラッキング課題で最重要視されている“身体的要求”は、他の課題では最も軽い要因と判断されている。

このように課題によって重みのパターンに大きな違いがみられることから、今回の実験課題が作業負荷の構成要素の面で相互に性質の異なるものであったこと

が検証された。しかもなお、どの作業にもWWL得点が負荷に対して高感度であったことは、日本語版NASA-TLXが英語版と同様、様々なタイプの作業に適用できることを示すものである。

最後に、オリジナル版のNASA-TLX開発報告にならい(文献1のTable 11, 12, p. 173)、尺度の評定値

表5 6尺度の評定値相互間の相関係数* (N=108**)

Tab. 5 Correlations among workload ratings on the six scales (N=108).

	MD	PD	TD	OP	EF	FR
MD：知的・知覚的要求	—					
PD：身体的要求	-.04	—				
TD：タイムプレッシャー	.46	.17	—			
OP：作業成績***	.36	.38	.33	—		
EF：努力	.38	.29	.29	.52	—	
FR：フラストレーション	.28	.45	.26	.47	.40	—

* ピアソンの積率相関係数

** 3つの課題それぞれに12名ずつの被験者が3条件のもとで評定した結果をすべて込みにして算出した

*** “悪い”ほど得点が高い

表6 6尺度の評定値で“全体的負荷”の評定値を回帰したときの重決定係数(r^2)と標準偏回帰係数(β 係数)

Tab. 6 Coefficients of multiple determination (r^2) and standardized partial regression coefficients (β -weights) for the six rating scales regressed on “Overall Workload”

課題	r^2	MD	PD	TD	OP	EF	FR
視覚探索	.83	.11	.09	.30*	.15	.34**	.18
暗算	.80	.41**	-.03	-.15	.15	.31**	.37**
トラッキング	.72	.13	.51**	-.03	-.03	.38**	.11
全課題	.73	.25**	.15*	.08	.08	.36**	.31**

* $p < .05$, ** $p < .01$

尺度名の略号は表5参照

相互間の相関係数（ピアソンの積率相関係数）を表5に、6尺度の評定値で“全体的負荷”の評定値を回帰したときの重決定係数(r^2)と標準偏回帰係数(β 係数)を表6に掲げる。

相関係数がオリジナル版に比較してむしろ低いうえ、重決定係数はオリジナル版よりも低めながら、なお.72～.83という高水準にあることから、日本語版の下位尺度が相互にある程度独立であり、かつ、全体として精神的作業負荷の諸変動要因をよくカバーしていることが示唆された。しかし、NASAにおける研究と本研究とでは、用いた課題の種類も数も異なるうえ、表5と表6にあげた数値は、12名の被験者が各課題を3回ずつ行った繰り返しデータを込みにして計算した結果である。したがって、尺度選択の妥当性を検討するためには、さらに多くの被験者の測定データの分析結果を待つ必要がある。

5-4. TLXの簡便法

尺度の重要度に関する一対比較は時間がかかるうえに、被験者にとって判断が難しく、負担となるものである。被験者の判断が安定しない場合があるため、本研究で用いたパソコンプログラムでは最大3回まで同じ対を提示し、算出される重みの信頼性を確保しているが、逆に、このステップを省略してしまう簡便法を用いた代替指標も提案されている。

最も簡単なのは、尺度に重みをつけずに単純平均するRTLX (raw-TLX) 得点であり、これでも加重平均によるWWL得点との相関が.93～.98あると報告されている^{6,9,10)}。

もうひとつは、三宅と神代が提案したAWWL (adaptive weighted workload) 得点である。これはワークロード評価のロースコアの高さの順に、尺度に6, 5, 4, 3, 2, 1の重みをつけて平均した値であり、WWL得点との相関が.989と高いうえ、被験者間のバラツキがWWL得点よりも小さいという利点のあることが報告されている⁹⁾。しかしながらTLX尺度の重みとは、作業者が経験するワークロードの構成要素と、各構成要素がトータルな主観的ワークロードに寄与する程度が、作業間や個人間で異なっていることを考慮して考案されたものである。したがって、評価結果をもとに後から重みをつけて得点を計算するのでは、そもそも何のために重みづけを行うのかの論理がなくなってしまうのではないだろうか。また、同じ作業を同じ被験者が行っても、負荷の大きさが変われば尺度の重みも変わる可能性があり、不自然である。

そこで筆者は、SWATのカードソートに類似した手法でNASA-TLXの6つの尺度を重要度の順に並

べてもらい、その順に6～1の重みをつければどうかと考える。この際、同順位も認めることにすれば、一対比較の際にしばしばみられるように、被験者がどちらを選んだらよいのか決めかねて、最後はほとんど苦しまぎれに一方を選ぶというような問題も生じないだろう。ただし、今回の実験ではこの手法を試みていないので、一対比較で“より重要”と選択する回数が多かった尺度ほど上位におかれるであろうと仮定して、順位を推定し、重みを決定した。同順位の場合はAWWL同様、平均順位を両者に等しく与える。このようにして算出した得点をCSTLX (card-sort TLX) 得点と名づけた。以下にWWL, OW, AWWL, RTLXとともに報告したい。

表7に、3種の実験課題の各困難度条件に対する得点をまとめたものを示す。どの指標もワークロードの変化に敏感に対応した値となっている。被験者間のバラツキについて、三宅と神代は中困難度条件に限定したうえで変動係数による比較を行っているが、ここでは3種の課題の3困難度条件すべてについて、標準偏差による比較検討を行いたい。変動係数は平均値の違いによって大きな影響を受けるからである。

まず視覚探索課題では、AWWLとTLXがどの条件でも他の指標よりバラツキが小さい。暗算課題ではL条件でOW, M条件でWWLとCSTLX, H条件でRTLXのバラツキが小さかった。トラッキング課題では、RTLXの標準偏差がどの条件でも最小である。やや乱暴ではあるが、全課題の全条件について標準偏差の平均値を計算すると、WWL, OW, AWWL, RTLX, CSTLXの順に16.43, 19.05, 15.38, 14.88, 15.83となった。

以上の結果から、AWWLはWWLに比べて被験者間の変動を抑えられるという三宅と神代の主張が裏づけられたが、重みづけを行わないRTLXのほうがバラツキはより小さかった。また、CSTLXはWWLとAWWLの中間に位置した。

次に、指標間の相関を調べた結果を表8に掲げる。WWLの代替指標としては、AWWL, RTLX, CSTLXともに十分その資格があるといえる。また“全体的負荷”との相関はこの3者のほうがWWLよりもかえって高かった。

こうしてみると、わずらわしい一対比較を行ってWWL得点を算出するよりも、簡便指標で十分代替できるばかりか、被験者間のバラツキや“全体的負荷”との相関の点ではかえって優れた特性を有すると思われる。しかし、今回WWLと比較された簡便指標は、一対比較を経験した被験者から得られたものであるという事実には留意しなければならない。この事情は

三宅, 神代⁹⁾ および三宅ら¹⁰⁾ の場合も同様である。一対比較のために尺度の説明を繰り返し読み, 重要度の判断を各尺度について5回も行うことにより, 尺度に対する理解は格段に深まったものと推測される。このプロセスを抜きにして, いきなり6尺度に対するワークロード評価を求めた場合にどのような結果になるか

を確かめるまでは, 簡便法を推奨することは控えたい。なお, 尺度を重要度の順に並べる“カードソート”の手順を踏んで重みを決めるCSTLXの場合は, AWWLとRTLXにみられるこのような弱点がやや小さいものと期待されるが, これも今後の研究で実際に確かめる必要があることはいうまでもない。

表7 各課題の各条件における“全体的負荷”(OW)の評価, および各種 TLX 指標の平均(AVG), 標準偏差(STD), 変動係数(COV)。ただし, $COV = (STD/AVG) \times 100$

Tab.7 Means (AVG), standard deviations (STD), and coefficients of variation (COV) of "Overall Workload" ratings and various TLX scores.

視覚探索課題						
条件		WWL	OW	AWWL	RTLX	CSTLX
L	AVG	41.00	37.42	41.40	38.28	40.22
	STD	22.23	21.86	19.62	19.53	21.29
	COV	54.22	58.43	47.40	51.03	52.93
M	AVG	55.75	54.00	53.87	51.60	54.56
	STD	17.15	21.13	16.90	16.37	16.63
	COV	30.77	39.12	31.38	31.72	30.48
H	AVG	75.08	71.25	73.68	69.07	73.46
	STD	15.18	23.75	12.60	13.37	14.51
	COV	20.22	33.33	17.11	19.35	19.76
暗算課題						
条件		WWL	OW	AWWL	RTLX	CSTLX
L	AVG	49.67	49.50	48.81	44.47	48.17
	STD	17.43	16.46	16.87	16.77	17.20
	COV	35.10	33.26	34.56	37.72	35.71
M	AVG	61.67	62.75	63.21	55.86	60.02
	STD	10.36	14.80	11.19	12.03	10.73
	COV	16.80	23.59	17.71	21.53	17.87
H	AVG	66.83	71.25	66.98	59.86	64.85
	STD	14.19	14.13	13.01	12.82	13.50
	COV	21.24	19.82	19.43	21.41	20.82
トラッキング課題						
条件		WWL	OW	AWWL	RTLX	CSTLX
L	AVG	54.33	52.17	52.38	46.74	52.38
	STD	19.67	19.70	19.71	16.61	18.64
	COV	36.19	37.77	37.63	35.55	35.59
M	AVG	63.83	62.42	61.70	55.24	61.54
	STD	17.00	20.00	13.73	13.02	15.68
	COV	26.63	32.05	22.25	23.58	25.49
H	AVG	69.25	67.67	66.56	59.14	66.48
	STD	14.67	19.60	14.77	13.61	14.25
	COV	21.18	28.96	22.19	23.01	21.43

表8 “全体的負荷” (OW) の評価および TLX の各種指標間の相関係数*

Tab. 8 Correlations among ratings of “Overall Workload” and various TLX scores.

	WWL	OW	AWWL	RTLX	CSTLX
WWL	—				
OW	.78	—			
AWWL	.94	.86	—		
RTLX	.95	.80	.97	—	
CSTLX	.99	.79	.96	.97	—

* ピアソンの積率相関係数

6. おわりに

日本語版 NASA-TLX を用いて3種類の室内実験課題のワークロード測定を行ったところ、いずれの課題においても、困難度の変化に感度よく対応したワークロード得点が得られた。

また、尺度の重要度に関する一対比較から算出される尺度の重みは、課題ごとに大きく異なることが明らかになった。このことは、3つの実験課題が作業負荷の面で相互に性質の異なるものであった証拠であり、TLX得点が性質の異なる種々の作業に適用可能であることを示唆するものである。

さらに、尺度選択の妥当性を検討するために相関分析および重回帰分析を行ったところ、オリジナル版の NASA-TLX とほぼ同様の結果が得られた。

最後に、一対比較を省略した簡便法から得られる各種指標を推定したところ、正規の NASA-TLX 得点である WWL に比べて被験者間の変動が小さく、WWL や “全体的負荷” との相関も高いことがわかった。このなかにはワークロードの評定結果に従って各尺度の重みを後から決定する AWWL や、尺度を重要度の順に並べさせた結果に基づいて重みを決める CSTLX が含まれる。後者は本論文において初めて提案された手法である。しかし、今回分析に供した CSTLX 得点の値は、あくまで一対比較の結果からの推定値であるため、今後の研究でこの方法の有効性を検証する必要がある。

日本語版 NASA-TLX の訳語や実施手続き、パソコンプログラムのユーザインタフェースなどについても、産業界における実作業を含む広範囲の課題と、年齢、性別、教育水準の異なる各層の被験者を使った実験を繰り返して、妥当性や信頼性の確認、あるいは改良を行わなければならない。その際、被験者の職種などにあわせて語句を若干修正する必要もでてこようが、基本的なかたちの日本語版 NASA-TLX を1つ標準化

しておくことも、意義のあることと考える。そのため今後、関係研究者が連絡をとりあって訳語と標準的な実施手順、教示などについて協議し、妥当性などを検証する実験を協力して進めることが望まれる。今回提案した日本語版 TLX をその“たたき台”にさせていただければ幸いである。

参考文献

- 1) Hart, S. G. and Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (eds.), Human Mental Workload, North-Holland, 139~183, 1988.
- 2) Reid, G. B. and Nygren, T. E.: The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (eds.), Human Mental Workload, North-Holland, 185~218, 1988.
- 3) Moray, N.: Subjective mental workload, Human Factors, 24(1), 25~40, 1982.
- 4) Wierwille, W. W. and Eggemeier, F. T.: Recommendations for mental workload measurement in a test and evaluation environment, Human Factors, 35(2), 263~281, 1993.
- 5) 芳賀 繁: NASA タスクロードインデックス日本語版の作成と試行, 鉄道総研報告, 8(1), 15~20, 1994.
- 6) 三宅晋司, 神代雅晴: メンタルワークロードの主観的評価法: NASA-TLX と SWAT の紹介および簡便法の提案, 人間工学, 29(6), 399~408, 1993.
- 7) 日本人間工学会標準化委員会: 作業システム設計のための人間工学の原則, 人間工学, 18(6), 333~336, 1982.
- 8) 青木和夫: ISO/TC 159 におけるメンタルワークロードの概念と定義および設計の指針, 人間工学, 29(6), 339~342, 1993.
- 9) Byers, J. C., Bittner, A. C. Jr. and Hill, S. G.: Traditional and raw Task Load Index (TLX) correlations: Are paired comparisons necessary? In A. Mital (ed.), Advances in Industrial Ergonomics and Safety, 481~485, Taylor & Francis, London, 1989.
- 10) 三宅晋司ほか: NASA-TLX簡便法の有用性の検討, 人間工学, 31(特別号), 404~405, 1995.