<General remarks > Special Issues No. 3: Measurement Technique for Ergonomics, Section 3: Psychological Measurements and Analyses (6), Mental Workload Assessment and Analysis —A Reconsideration of the NASA-TLX—, by Shinji MIYAKE.

特集③人間工学のための計測手法

第3部:心理計測と解析(6)1

メンタルワークロードの計測と解析-NASA-TLX再考-

三宅晋司2

391

1. はじめに

NASA-TLX (以下、TLXと略記する) は最もよく用いられている主観的メンタルワークロード評価手法である. De WinterはGoogle Scholarでメンタルワークロード評価の複数の手法について検索し、TLXが「爆発的」に使用されていることを報告している¹⁾. 図1は彼の論文のTab. 3から主要な3つのメンタルワークロード評価指標を抜き出して著者が図示したもので、1969年から4年間

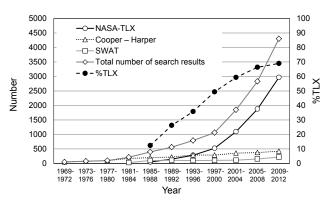


図1 3つのワークロード評価手法の4年毎の報告件数とTLX の使用率 (右軸). De Winter¹⁾ のTable 3を基に筆者が作 図したもの.

Fig. 1 Number of articles per quadrennium in which three major workload questionnaires were used. This graph is drawn by the author after Table 3 in De Winter's article¹⁾.

School of Health Sciences, University of Occupational and Environmental Health, Japan 毎に検出された件数を示している。論文数そのものも急激に増加しているが、TLXの使用割合(右目盛り)が急激に増えており、近年では70%近くに達している。彼は、これはMatthew効果(金持ちはより金持ちになり、貧乏人はより貧乏人になる)によるもので、「みんなが使うから、自分も使う」ためと評している。私自身、実験を行う度にTLXを使っており、本誌でもTLXを用いた報告が多く見られる。しかし、原著論文でもTLXをどのように使用したかについては詳しく記述されておらず、誤った使い方がされている可能性は否定できない。このようなことから、初心に戻って原典を読み直し、正しい(標準的な)使い方を確認する必要性を感じていた。そこで本稿では、「NASA-TLX再考」と題して、考察を含めながら、TLXの基本事項をまとめてみた。

本稿は、特集「人間工学のための計測手法 第3部 心理計測と解析」の一部であり、メンタルワークロードの主観的評価法の総説としての位置づけであるが、上述のようにTLXが世界標準のようになりつつある現状で、それ以外のメンタルワークロード評価手法を紹介する意義は低いと考え、Hart and Stavelandの「原典」 2)とTLX マニュアル 3 4 4 からの引用を中心として、TLXのみについて論述した(一部、Subjective Workload Assessment Technique:SWAT 5 5 にも言及). Hart 6 6 は、この書物 "Human Mental Workload" はすでに絶版となっているので、TLXを使用した論文でこの「原典」を参照しているものは少ないと述べているが、この書物の該当部分 (pp.139-183) のコピー(PDF) はwebsite 7 7 から入手可能である。また、TLXについては以前の本誌 8 8 で詳細を紹介しているが、まずその概要を解説する.

¹ 受付:2015年7月22日 受理:2015年10月23日

² 産業医科大学産業保健学部

Tab. 1 NASA-TLX rating scale definitions⁸⁾.

項 目	端点	
精神的要求	低い/高い	どの程度,精神的かつ知覚的活動が要求されましたか? (例. 思考,意志決定,計算,記憶,観察,検索,等)作業は容易でしたか,それとも困難でしたか.単純でしたか,それとも複雑でしたか. 苛酷でしたか、それとも寛大でしたか.
身体的要求	低い/高い	どの程度、身体的活動が必要でしたか? (例. 押す、引く、回す、操作する等) 作業は容易でしたか、それとも困難でしたか、ゆっくりしていましたか、それともきびきびしていましたか。ゆるやかでしたか、それとも努力を要するものでしたか。落ち着いたものでしたか、それとも骨の折れるものでしたか。
時間切迫感	低い/高い	作業や要素作業の頻度や速さにどの程度,時間的圧迫感を感じましたか? 作業ペースはゆっくり していて暇でしたか,それとも急速で大変でしたか.
作業達成度	良い/悪い	実験者(あるいは、あなた自身)によって設定された作業の達成目標の遂行について、どの程度成功したと思いますか? この目標達成における作業成績にどのくらい満足していますか?
努力	低い/高い	あなたの作業達成レベルに到達するのにどのくらい―生懸命 (精神的および身体的に) 作業を行わなければなりませんでしたか?
不満	低い/高い	作業中, どのくらい, 不安, 落胆, いらいら, ストレス, 不快感, あるいは安心, 喜び, 満足, リラックス, 自己満足を感じましたか?

注:これは下位尺度の意味を理解するために、原文をそのまま訳したもので、日本語版として提案したものではない。

2. 概 要

TLXでは表1に示した6項目(精神的要求 Mental Demand: MD, 身体的要求 Physical Demand: PD, 時 間切迫感 Temporal Demand: TD, 作業達成度 Own Performance: OP, 努力 Effort: EF, 不満 Frustration Level: FR. これらを下位尺度と呼ぶ) について、「低 い一高い」または「良い-悪い」を両極とするGraphical scaleを用いて $1 \sim 100$ 点で評定点を得て、その平均値を 算出するものである. 平均値は, 一対比較 (Source - of -Workload Evaluationと呼ばれている)を用いた重み付け係 数による平均値WWL(Weighted Workload)を用いるの が基本的な方法である. これは後述 (4. TLXの使い方 手順2:一対比較)の方法で、6つの下位尺度の全てのペア 15組について、ワークロードへの関わりが大きいと思う方 を選択させ、各下位尺度の選択された回数 (0~5)を その尺度の重み係数 w_i とする. 各下位尺度の評定値を v_i とすると、WWLは次式で算出され、1~100の値とな る. 分母の重み係数 w_i の総和はペアの数と同じ15である.

$$WWL = \frac{\sum_{i=1}^{6} (w_i \times v_i)}{\sum_{i=1}^{6} w_i}$$

なお、一対比較を用いずに、重み付け平均値を算出する簡便法が2つ提案されている。ひとつは著者による AWWL(Adaptive Weighted Workload) 8 であり、6つの下位尺度の評定点の大きい順に、 $6\sim1$ の重み付け係数を定めるものである。もうひとつは芳賀によるCSTLX

(card -sort TLX) である. この方法はSWAT® のcard sort (Time Load, Mental Effort Load, Psychological Stress Loadの3項目それぞれについて、高、中、低の3レベルが設定され、これらの全ての組み合わせ27通りの例示がされた27枚のカードをワークロードが高いと思う順に並べ変える) に類似したもので、6つの下位尺度を重要度の順に並べさせ(順位付け)、その順位を重み付け係数として用いるものである90.

3. 特 徵

3-1. 一対比較による重み付け平均値の算出

TLXの最大の特徴は、前節の概要で述べた尺度の一対比較より重み付け係数を算出し、それを用いた重み付け平均値の算出であろう。一対比較は評定者毎に行われ、評定者のワークロードに対する感受性の差異と作業の特性の差異を平均値に反映させるものであり、次の仮定より重み付けされたワークロードの得点を得る:(1)単一の総合的ワークロード(OW:Overall Workload)評価値を構成する際に考慮される要因は被験者毎に異なる。(2)被験者はそれら全ての要因を評価することができる(たとえワークロードを評価する際に通常はそれらを考慮しないかもしれないとしても)。(3)被験者ははっきりしないOWの概念に対してよりも、より正確で、かつ個人間のばらつきがより小さく、要素要因の量を判定できるであろう。(4)被験者が行う評定は、被験者の自然な推定ルールに対する"生の値"を表す。(5)個々の被験者

自身の推定ルールに従って個々の要素を結合することによって、ワークロードの推定値(WWL)を得ることができ、それはOWの評定よりも被験者間のばらつきが小さいであろう($p.148^2$).

この一対比較は各作業後に行うことになっている(一対比較教示の項参照). SWATでも、このような個人特性の差異を反映させるのに、前節で述べたcard sortという方法を用いた評価値の算出を行っているが、SWATでは、特定の作業についてではなく、「あなたのワークロードについての一般的見方(your general view of workload)に従って」 card sortを行うように教示している(SWATの詳細については文献®を参照のこと). TLXでは、これを作業後に行うことにより、同一作業に対するワークロードの定義の被験者間の差異と、異なる作業間のワークロード要因の差異に起因するばらつきを考慮でき、さらにその作業によって生じるワークロードの特性に関する情報も得ることができると考えられている(p.176²)).

この一対比較が問題点としてあげられることがある が、それはTLXそのものの問題としてではなく、一対 比較の実施時の問題である. 一対比較そのものは極めて 短時間で実施できるので、実験遂行上の障害になること は少ないと思われる (SWATのcard sortは30分~1時間 を要する)が、そもそも、教示されている「ワークロー ドについてより重要な関りを示している方を選べ」とい う判断が難しいという指摘もできる. 一般の人を被験者 とした場合は、ワークロードそのものの意味がわからな いこともあり(そのため、前述の仮定のように、ワーク ロードそのものではなく、6つの下位尺度について例を 示して回答させる)、一対比較における判断基準が曖昧 になっている可能性がある. この一対比較の意図(意味) がよくわからないと、「どちらの項目の方がワークロー ドへの関わりが大きいか」に対して、「どちらの項目の方 を大きく感じたか」の判断をしているかもしれない. こ れが一対比較を用いない重み付け平均値AWWL®の算出 の根拠であり、6つの尺度について評定点が最も大きい 尺度は100点満点とし、2番目に大きい尺度は83点満点 (100点×5/6),3番目は66.7点満点,4番目は50点満点, 5番目は33.3点満点、評定点が最小の尺度は16.7点満点 とした6尺度の平均点と等価であり、下位尺度における 評定点差を拡大したものに相当する.

Hart and Staveland²⁾ の実験では、TDとFRについては、重み付け係数と評定点が高い相関を示し、タイムプレッシャーの重み係数が高いのは、その評定が高いときのみであった。しかし、MDまたはPDがワークロードの主要な要因であった場合は、それらの評定点は必ず

しも高くなかった。例えば、コントローラの操作が緩やかなトラッキング課題では、PDはワークロードの第一の要因と考えられるが、PDの評定値は極めて低かった(26点)。多くの課題ではMDはワークロードの第一の要因と考えられたが、それぞれの課題において被験者へ掛かったMDの大きさに応じて、MDの評定値は20から60の範囲であった(p.174)。

WWLでは、6つの下位尺度それぞれを100点満点として評定するのではなく、重みの最も大きいもの(係数5)については、33.3点満点(100点×5/15)、以下同様に、係数4では26.7点満点、係数3では20点満点、係数2では13.3点満点、係数1では6.7点満点で評定させた6つの下位尺度の合計点と等価である(満点はVASの線分の極の値である).

RTLX (Raw TLX, 単純平均値)とWWLの相関が高 いのに、2分を費やして一対比較を行い、重み係数を得 るのはムダだという指摘(文献¹⁾のTable 4)に対し、前 述のように、重み係数より、作業特性に関する有用な 情報が得られると記載されている2. 図2は特性の異な る2つの作業に対するTLXの値である10). MA (Mental Arithmetic:暗算課題) は5秒毎にPC画面に表示される計 算式(足し算または引き算)に対して、その後に表示さ れた答えが正しいか、間違っているかをマウスボタンの クリック(正しければ左クリック、間違っていれば右ク リック)で回答するものである. 計算そのものはそれほ ど難しくはないが、5秒以内に回答しなければならない というマシンペースの課題であるため、多くの被験者は 高いタイムプレッシャーを感じる. 一方, MT (Mirror Trace: 鏡映描写課題) は、PC画面に表示された屈曲路 をマウスを使って辿る課題であり、時間制限のないセル

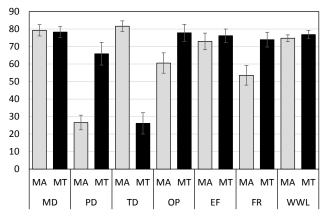


図2 暗算課題 (MA) と鏡映描写課題 (MT) に対するTLX 評価 (*n*=19, means±SE).

Fig. 2 TLX ratings and WWL scores for Mental Arithmetic (MA) task and Mirror Trace (MT) task (n = 19. Bars represent standard errors of the mean.

フペースの作業である. ただし、マウスのX軸とY軸が入れ替えてあり、屈曲路から外れずに辿ることは「ゆっくり」であっても大変難しい. これら2つの作業に対するTLXの下位尺度は、作業の特性を表している. すなわち、MAではTDは極めて高いが、MTではMAの1/3程度である. 一方、PDとFRはMTで高いが、MAではそれほど高くない. これらの「逆転」の結果、2つの作業に対するWWLはほぼ同一の値を示している. 下位尺度を検討せず、平均値のみを表記したら、このような作業間の差異は示されない. 心拍数や心拍変動性の指標はこれらの2つの作業で異なり、WWLが同一でも、生理反応は異なるという例である11).

以上のように、TLXの平均値には、下位尺度の比較 から得られる重み係数を用いた重み付け平均値WWLと CSTLX (および下に述べた順位法による方法),一対比 較を用いない単純平均値RTLX, 一対比較を用いない重 み付け平均値AWWLといった複数のものが存在し、そ れら相互は高い相関係数を示す12). 下位尺度の比較をさ せて重み付け平均値WWLを算出するのが好ましいと思 われるが、例えば、ある条件の自動車運転直後で、さら に運転を続けているような場合や3.5章の例(手術直後 等) の場合は、一対比較を行うのは困難であり、TLXの 評定は発話等により行うことになるであろう. その場合 はRTLXとAWWLしか算出できない(発話での一対比 較は不可能ではないが). すなわち、複数のTLXの平均 値のどれを算出するかは、実験環境に依存すると考えら れ、繰り返しになるが、WWLの算出が基本であろう. なお、一対比較を行えば、WWLのみならず、RTLXも AWWLも算出できるので、それらについても検定を行 い、これらの指標間の差異について検討するとよいであ ろう. なお, 一対比較を行わずに重み係数を得る方法と して、Hart and Stavelandは6つの項目を順位付けさせる 方法が最も簡単だろうと述べている (p.1702). この方 法はCSTLXとほぼ等価である.

3-2. Graphical scale

次にTLXの特徴と言えるのが、Graphical scale、すなわちVisual Analog Scale: VASによる評定である。量的尺度を求める主観評価では、通常、Likertスケールと呼ばれる何段階かの評定尺度法が用いられるが、TLXでは両極を、高い(High) – 低い(Low)(OPのみ、悪い(Bad) – 良い(Good))とするVASを用いている。Hart and Staveland²⁾ は、「これで得られる値は比例尺度でも、間隔尺度でもないかもしれないが、評定のばらつきは許容できるほど小さく、尺度範囲のほとんどが様々な作

業への評価で用いられ、その数値は実験操作と高い相関 を示した」と述べている (p.170). 評定尺度では離散値 しか得られないのに対し、VASでは、たとえそれが21 目盛りによる5刻みのものであっても、より連続量に近 い (分解能が高い?) 値が得られる. パソコン画面に表 示されたスライダ (図4参照) を用いて評定 (Computer 版,以下,PC版)させれば,読み取りの分解能を上げる ことができ、記入用紙を用いる場合(Paper & Pencil版、 以下、PP版)でも、チェックされたマークを1mmの精度 (1/120) で読み取ることは可能であるが、いずれも分解 能が5に「丸められて」使用される、PP版では、あとか ら数値を読み取らなければならないので、目盛りをつけ ておくと読み取り作業が楽になるというメリットがある が、PC版では、数値が保存されるため、目盛りをつけ るメリットはない. むしろ、両端を示すだけで、無目盛 りの線分に対して評定させるほうが良いように思われ、 Hart and Staveland²⁾ は無目盛りが好ましいと述べている (p.171) が、NASAのPC版はPP版と同じように、5刻み の21の目盛りがある.

3-3. 感度

TLXが多く用いられている理由は、その評価値の感度 の良さであろう. これは難易度を実験的に調整した複数の 作業に対する評価値が、難易度と相関を示すということ である. 言い換えれば、作業難易度への感度が高いとい うことである. しかし、被験者が遂行をあきらめるよう な高い難易度の作業、すなわちオーバーロード領域では 相関が悪くなることがある13). また, 先行する作業難易度 が、その後に実施する難易度が同一の作業に対するTLX 評価へ影響する(例えば、同じ作業であっても、難しい 作業の後は易しく感じる) ことが報告されている14,15). さ らに、6つの下位尺度のうちの作業そのものに関連する3 つ、すなわちMD、PD、TDについては、作業を遂行しな くても、誰かが作業を遂行している様子を端から見てい るだけで評価できる可能性がある. そのことが実際に作 業を行った後のTLX評価へ何らかの影響を与えるのでは ないと考えられるが、実験的検証は行っていない.

3-4. 規準課題

上述のように、この種の主観評価では、評価者は何らかのものとの相対評価を行う傾向にある。従って、それらの「比較対象」がない状況で、いきなり「高い一低い」のVASの評定を求められても、判断基準がないので、評定値の信頼性が落ちる可能性がある。例えば、最初に行ったある作業に対して、MDがかなり高いと感じ、100

点に近い評定をした被験者は、それよりももっとMDが高い作業に対しては、100点以上の評定をつけることができない。したがって、それ以降にさらに何段階か作業要求度が上がっていっても、その差異は評価できず、全て100点になってしまう。そこで、実験課題以外に、何らかの規準作業(reference task)を遂行させ、それを判断基準とした相対評価を行わせる方法も考えられる。Hart and Staveland²⁾は、規準課題を示さなくても、実験の被験者に、評価すべき作業の難易度の範囲と平均の例示を示せば、安定した判定値を得るのに役立ち、遂行すべき作業のタイプを被験者に示すことができると述べている(P.171)。

3-5. 音声版 (verbal) TLX

作業直後に評価を行う際に、手が使えず、PC版もPP版の利用できない場合は、発話により回答を得ることができる。Hart and Staveland²⁾ は、PC版、PP版、音声版を比較し、PC版は音声版より2ポイント高く、PP版より7ポイント高かったことを報告しており、PC版でPP版よりかなり高い値が得られていることについては、評定点のパターンに一貫性があるので、数値の差異はそれほど重要ではないと述べている(p.175)。Grantら 16 は腹腔鏡手術の練習時にPP版と音声版のTLXの比較をし、両者の相関が高い($\mathbf{r}>0.80$)ことを示し、手術技術の評価では、音声版はPP版の代用として同等な感度を有していることを示した。

3-6. 日本語版TLX

TLXの日本語版は芳賀によるものが提案されている⁹. 本稿表1および過去のTLX紹介論文⁸⁾の付表1で示した訳は、TLXの内容を理解するために翻訳したもので、日本語版として提案したものではない. 私はこの翻訳版を用いているが、言語の異なるオリジナル版を翻訳して用いる際は、充分な検証が必要であろう.

TLXはメンタルワークロードの評価手法であるが、これを製造工場における組み立て作業者のワークロード評価に用いることができないかの検討が行われており、ライン作業者用に改訳したTLX-L(Lはline workerを表す)が提案されている「7.18)。これは先に述べた「直訳」では、現場作業者が下位尺度の意味を理解できないため、現場の作業者にヒアリングをしながら訳語を定めたものである。例えば、「身体的要求」は「肉体的負担」と改訳されている。「要求」と「負担」は異なる概念であるが、現場作業者は「身体的要求」と聞かれても、意味が理解できない。また、TDは「時間的圧迫感」とし、説明文章では

「手待ち」や「余裕」というライン作業者が普段用いる用語が取り入れられている。このTLX-Lについては、再現性が確認されている。

4. TLXの使い方

4-1. 手順

TLXの標準的な使用方法をまとめた.

手順1:練習

実験課題の説明を行ったり、練習や習熟をさせる際に、その課題についての評価を「練習」として実施し、TLXの評定の仕方および6つの下位尺度の意味について理解させる.

手順2:一対比較

実験課題遂行後(あるいは評価したい作業後)に一対比較を実施する. 紙媒体(PP版)で実施の場合は、下位尺度15対の一対ずつを1枚のカードに記載した15枚のカード(写真1)を準備し、これを被験者および課題ごとにシャッフルして被験者に渡し、前に記入した内容を見ないように指示して、1枚ずつ記入させる. 教示は次項に示した. PC画面へ表示して、マウスやキーボードで実施する場合(PC版)も、1画面に1つのペアのみをランダムな順に提示(図3)して回答させ、回答が終了したら次の画面に進み、前に回答した内容を参照できないようにする.

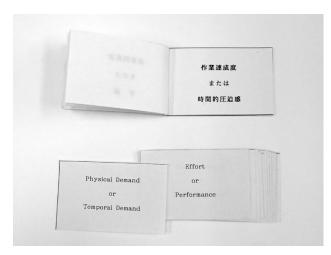


写真1 一対比較用カード.上は綴じたもの.下は綴じずに使用するもの.いずれも前に記入したカードを見ないように教示する.

Photo 1 Cards for the Source-of-Workload pair wise comparisons. Booklet type (above) or an unbound card set (below) are used. Participants are instructed not to look at the previously marked card.

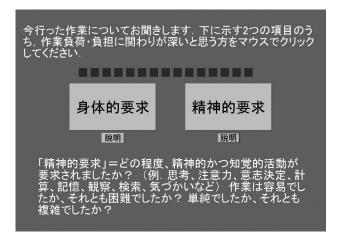


図3 TLXコンピュータ版 (著者作成) ¹⁹⁾ での一対比較画面. 15個の■がひとつずつ減っていき,進捗状況を表す.このプログラムでは尺度名の下部に「説明」ボタンを配し,必要に応じて,随時,尺度の説明を表示することができる. 右または左の尺度名をクリックすると,直ちに次のペアが表示される.ペアの表示順序はランダマイズされていない.

Fig. 3 Screen image of the source-of-workload paired comparisons in a computerized version programmed by the author¹⁹⁾. Squares indicate the progress by disappearing one by one. "Description" buttons are arranged below the subscale title buttons so that participants can refer the description of the subscale anytime they want. When the right or the left button of the subscale is pressed, the next pair will be appeared immediately. The order of the pairs is not randomized in this program.

一対比較教示

以下の文はTLXマニュアル³⁾ からの引用である(下線は原文に示されている通り).

「この実験を通じて、異なる作業条件におけるあなた の体験を評価するのに、評定尺度が使われます. この種 の尺度は極めて役に立つのですが、その有用性は人々が それらを個人個人の方法で解釈する傾向の影響を受けま す. 例えば、ある人はどれだけ努力をしたかとか、どの くらいよくできたかに関係なく、精神的要求や時間切迫 感をワークロードの主要な要素と感じるでしょう. 他の 人は、もし作業成績がよければ、ワークロードは低かっ たに違いない (またはその逆) と感じるでしょう. さら に努力や不満がワークロードの最も重要な要素と感じる 人もいるでしょう. これまでの研究は、考えられるあ りとあらゆる数値パターンをすでに見出しています. さ らにワークロードのレベルを生み出す要因は作業に依存 して異なります. 例えば、ある作業はとてもすばやく 行わなければならないので、難しいかもしれません。別 の作業は、要求されている精神的または身体的努力の程 度によって、難しかったり、易しく見えるかもかもし

れません. さらに、ある作業では、どれだけ努力したか に拘わらず、うまく遂行できないために難しく感じるで しょう.

あなたが今行なおうとしている評価は、NASAによって開発されたもので、どの程度あなたがワークロードを体験したかを決定するのに用いる6つの要素の相対的な重要度を調べるものです。手順は簡単です:あなたは一連の、対になった尺度の名称(例えば、努力ー精神的要求)を見せられます。そして、今行なった作業であなたが体験したワークロードについて、より重要な方を選んでください。それぞれの尺度の名称の対は、別々に画面に表示されます。 あなたがこの実験で行なった作業のワークロードについてより重要な関りを示している尺度の名称を選んでください。」

PP版では、上記最後の3行の代わりに次の文が用いられる⁴).

「それぞれの尺度の名称の対は、別々のカードに示されています。<u>あなたがこの実験で行なった作業のワークロードについてより重要な関りを示している尺度の名称に</u>○をつけてください.」

この一対比較の教示では、上記の下線部分が重要であり、したがって、このような教示の前に、「どちらの尺度を大きく感じたかではなく」の説明を入れるとよい.

手順3:評定

PP版の場合は「高い一低い」または「良い一悪い」を両極とする12 cmの線分を記載した用紙を用いる. 無目盛りのものが推奨されているが、目盛りをつける場合は20刻み (5間隔) として、目盛り上に○などを記入させる. 目盛りの間にしるしを書いた場合は、大きいほうの数値を読み取る. 6つの尺度は1枚の用紙 (PP版) または1画面 (PC版:図4) に並べて表示する.

手順4:結果の集計

WWLのみではなく、6つの下位尺度の平均値や、各課題に対する重み付け係数の平均値を算出すると、課題特性の検討が行える。6尺度について、重み係数×評定値の面積を示すこともある(図5)。何らかの理由で、一対比較(およびCSTLXのカードソート)を行わない場合は、6尺度の単純平均RTLXに加え、AWWLを算出できる。

4-2. 補足1 練習

練習および習熟は実験遂行上の基本 (実験目的により、練習させずにいきなり課題を遂行させるということもある) であるが、練習をさせずに充分内容を理解させないまま実験を行った場合、被験者の取り組みは主に

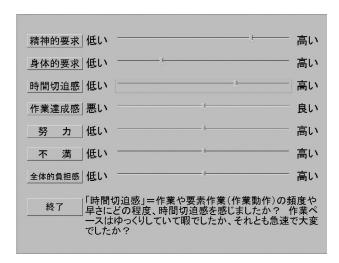


図4 TLXコンピュータ版 (著者作成) ¹⁹⁾ でのグラフィカル表示. スライダの初期位置は中央に配置してある. 左の尺度名をクリックするかスライダを動かすと, 下部に尺度の説明が表示される. TLXの6尺度以外に全体的負担感 (OW) の尺度も追加してある. すべて評価し終えると,終了ボタンをクリックするように教示する.

Fig. 4 Screen image of TLX graphical ratings¹⁹⁾. All initial points of the sliders are the center of scales. When subscale titles placed on the left side are clicked or the slider is moved, the description of the subscale is presented. Overall Workload (OW) scale is also arranged. Participants are instructed to press FINISH button when they complete the ratings.

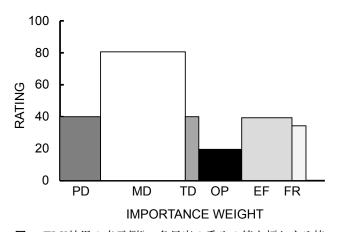


図5 TLX結果の表示例⁶. 各尺度の重みの値を幅とする棒グラフ表示にすることにより、重み係数×評定値が面積として表示され、下位尺度のWWLへの寄与度を理解しやすい.

Fig. 5 Graphic example of a TLX result⁶⁾. Since the bar width is the weight of its subscale, the product of the weight and the rating score is shown as the area of the bar. Therefore, the contribution of each subscale to WWL is understandable.

2つあると思われる. 一つは, 実験が始まってもやりかたがよく分からないので, 実験者に質問する場合である. この場合は, 実験開始後に質問により実験が中断されるということが生じるが, TLXなどの主観評価を行っ

ている際の中断は問題にならないであろう。もう一つは、よく分からないまま、自分なりの勝手な判断をするか、あるいは「いい加減」に遂行する場合である。この場合は、結果の信頼性が大きく損なわれる可能性がある。実験者からみれば、被験者はスムースに遂行しているので、充分理解した上で実行していると思ってしまうが、実はいい加減にやっているのかもしれない。したがって、練習させずに評定をさせる場合は、「分からないことがあったら、途中で中断しても構わないので、必ず質問してください」と指示しておくべきであろう(練習させていても、この指示はしておいたほうがよいのは言うまでもない)。

4-3. 補足2 一対比較

PP版で実施の場合、15枚のカードを用いるのではな く、それらを一枚の用紙に記載したものを用いる場合が ある. この場合は、前に記入したところを見ずに、回答 してくださいと教示しても,前に記入したものを参照す ることが可能となる. そうすると一対比較における判断 の矛盾(A>B, B>CならA>Cとなるはずであるが、C >Aとなることがある. これを一巡三角形という) が生 じる可能性が少なくなり、回答内容の論理的な整合性は 高まるかもしれない. ただし, TLXの一対比較では, 必 ずどちらかを選ばせ,両者に差がないという選択技はな いため、一巡三角形が生じた場合は、それらの尺度間に 差がないことを示しているだけで,「矛盾」 とはいえない かもしれず、重み係数が同一の値をとることに不都合は 生じない. 上述のようなPC版(図2)では,前の回答が 参照できず、前の記入を参照できるPP版とは異なった 結果が得られるかもしれない. この差異については検討 していないが、6項目しかないので、前に回答した内容 をある程度覚えておくことは可能なので、結局は前の回 答を勘案しながら回答しているかもしれない. 問題は, 前の回答に基づき、それ以降のペアの比較を機械的(論 理的?) に行い、比較しなければならない内容について の判断を怠っている可能性である. そのような可能性を 避けるためにも、前の回答内容の参照は禁じた方がよい と考えられる. TLXのマニュアル (次項参照) ではPP 版にもPC版にも、前の回答を見ないようにとの教示は 記載されていない.しかし,PC版の教示では,「それぞ れのペアは別々の画面に表示される」「回答したら、新し いペアが表示される」(図2と同様) と説明され、PP版は 「それぞれのペアは別々のカードに表示される」と教示 されるため、暗に「前の回答を見ないように」と示唆し ているものと思われる.

397

おわりに

本稿では、世界標準になりつつあるNASA-TLXについて、Hart and Staveland²⁾ の原典とNASAによるマニュアル類 $^{3,4)}$ を再読し、改めて、その使用法を記述するとともに、留意点をまとめた。本稿がNASA-TLXの使用時の一助になれば、特集記事としての役割は全うされるであろう。

謝辞

特集号として、本稿を書く機会を与えていただいた編 集委員会および有用なコメントをいただいた匿名の査読 者の方へ深謝いたします.

参考文献

- De Winter JCF: Controversy in human factors constructs and the explosive use of the NASA-TLX: a measurement perspective, Cognition, Technology & Work, 16, 289-297, 2014, doi: 10.1007/s10111-014-0275-1
- Hart SG, Staveland LE: Development of NASA-TLX (Task. Load Index): Results of Empirical and Theoretical Research. In Hancock PA, Meshkati M (Eds.): Human Mental Workload, Elsevier Science Publishers B. V., North-Holland, 139-183, 1988
- 3) Human Performance Research Group, NASA Ames Research Center: NASA TASK LOAD INDEX (TLX) v.1.0 Computerized version, NASA Ames Research Center. Available from http://humansystems.arc.nasa.gov/groups/tlx/downloads/TLX_comp_manual.pdf
- 4) Human Performance Research Group, NASA Ames Research Center: NASA TASK LOAD INDEX (TLX) v.1.0 Paper and Pencil Package, NASA Ames Research Center. Available from http://humansystems.arc.nasa.gov/groups/tlx/downloads/TLX_pappen_manual.pdf
- 5) Reid BR, Nygren TE: The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In Hancock PA, Meshkati M (Eds.): Human Mental Workload, Elsevier Science Publishers B. V., North-Holland, 185-218, 1988
- Hart SG: NASA-TASK LOAD INDEX (NASA-TLX);
 Years Later. Available from http://www.stavelandhfe.com/images/TLX_20_years_later_2006_Paper.pdf
- 7) http://www.stavelandhfe.com/images/NASA-TLX_

paper.pdf

- 8) 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
- 9) 芳賀繁: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-,人間工学,32(2),71-79,1996
- 10) Teshima M, Miyake S: unpublished data, 2003
- 11) Kuraoka H, Tsuruhara K, et al: Effects of a sensory intake task on heart rate and heart rate variability. Proceedings of 19th Triennial Congress of the IEA, paper number 2002, 2015
- 12) 三宅晋司, 佐々木十太, 他: NASA-TLX簡便法の有用性の検討, 人間工学(suppl), 31, 404-405, 1995
- 13) Miyake S, Yamada S, et al: Physiological Responses to Workload Change. A Test/Retest Examination. Journal of Applied Ergonomics, 40 (6), 987-996, 2009
- 14) Hancock PA, Williams G, et al: The Influence of Task Demand Characteristics on Workload and Performance, International Journal of Aviation Psychology, 5, 63-85, 1995
- 15) Miyake S: Factors influencing mental workload indexes, Journal of UOEH, 19(4), 313-325, 1997
- 16) Grant R, Carswell CM, et al: Equivalent-Forms Reliability of Printed and Spoken Versions of the NASA-TLX. Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting, 1532-1535, 2008
- 17) 三宅晋司, 神代雅晴: NASA-TLXを用いた現場作業者の負担感調査の試み, 人間工学, 32(suppl), 208-209, 1006
- 18) Miyake S, Kumashiro M, et al: Is NASA-TLX a useful method for the evaluation of workload in line workers?, Proceedings of the 4th Pan-Pacific Conference on Occupational Ergonomics, 438-441, 1996
- 19) 三宅晋司:実験室実験用主観指標と精神課題プログラムS. I. M. P. L. E.について, 人間工学, 38(suppl), 332-334, 2002

著者情報



三宅晋司 (みやけしんじ, 正会員, 連絡著者) 1983年北海道大学大学院環境科学研究科修了. 学術博士. 1996年より産業医科大学教授. 専門領域:心理生理学, 癒しの評価. 日本人間工学会財務担当理事, 同PIE 研究部会長ほか.

連絡先: myk@health.uoeh-u.ac.jp