

原著論文

Rating Scale Mental Effort の日本語化に関する研究

大門 樹¹⁾, 土屋 有加¹⁾
¹⁾ 慶應義塾大学理工学部

Study on Adaptation of the Rating Scale Mental Effort (RSME) for use in Japan

Tatsuru DAIMON¹⁾, Yuka TSUCHIYA¹⁾
¹⁾ Faculty of Science and Technology, Keio University

Abstract: The Rating Scale Mental Effort (RSME) is a unidimensional scale used to measure subjective mental workload. Original version of the RSME consists of a line with a length of 150 mm marked with nine anchor points, each accompanied by a descriptive label in Dutch indicating a degree of effort. The RSME has been widely used in Europe and North America. A past study has adopted the RSME to Indonesian and it has also been reported that the sensitivity is related to national culture or to the translation of the scale. This study aims to adapt the RSME to Japanese. In the experiment, the English labels of RSME were translated to Japanese and each Japanese label was rated against the 150mm line. A new seven-anchor point version of the RSME was created and the new version of the RSME was compared with the nine-anchor point version in a questionnaire with 72 participants. Moreover, the new version of the RSME was compared with the NASA-TLX in an experiment in which 10 participants performed a memory and response task of different levels of difficulty. Through these questionnaire and experiment, the adapted Rating Scale Mental Effort (RSME) is expected to be useful for use by Japanese.

Keywords: rating scale, workload, mental effort, evaluation method, questionnaire

キーワード: 評価尺度, 作業負担, 精神的努力, 評価手法, 質問紙

1. はじめに

スマートフォンやタブレット端末などのモバイル機器やカーナビゲーションなどの車載情報機器において多様なサービスが提供される一方で、それに伴う情報コンテンツの増加や選択操作の複雑化を招くことが懸念される。モバイル機器上に提示される情報コンテンツを視認したり、特定の操作を行ったりする際に、ユーザとモバイル機器の間のインタラクションを適切に設計することが重要となる。ヒューマン・マシン・インタフェース(以後、HMIと称す)の仕様によっては、ユーザとモバイル機器とのインタラクションにおいてユーザに過度な負担を強いったり、情報コンテンツから目的の情報を抽出するのにユーザが多くの時間を費やしたりしてしまう場合がある。このような観点から、ユーザとモバイル機器とのインタラクションやモバイル機器のHMIを適切に設計する上で、ユーザの負担を考慮する必要がある。

一方、ユーザとモバイル機器のインタラクションやHMI利用に伴う負担を評価する方法としては、反応時間などを指標として客観的に分析・評価する定量的手法、質問紙などを介して

ユーザの心理面などを主観的に抽出して分析・評価する定量的手法の利用が考えられる。前者の場合、反応時間などを計測するための機器類の実装が必要となることが多く、これらの実験データの収集は必ずしも容易に収集できるとは限らない。一方、後者の場合は、客観的な実験データではないが、計測機器を追加実装する必要がないことが多く、容易に利用することが可能であり、分析・評価対象となるインタラクション設計やHMI仕様の初期的な段階において有用な知見を収集することが可能であると考えられる。

モバイル機器利用時のユーザの負担評価に限らず、ユーザの作業負担を評価するために多くの質問紙調査法が提案されている。ユーザの作業負担を多次元的に捉えて評定する手法としては、NASA-TLX (NASA Task Load Index)[1]やSWAT (Subjective Workload Assessment Technique)[2], WP (Workload Profile)[3]などが提案されており、NASA-TLX やSWAT は日本語化もなされている[4]。これらの手法はユーザの作業負担を多次元的に捉えた複数の項目に基づいて評定するといった特徴がある。その一方で、これらの手法で設定されている多くの項目に対してユーザが評定する必要があることから、作業負担評定のための準備作業や評定作業そのものに多くの時間を費やしてしまったり、手法の利用そのものがユーザに対して作業負担の増加を招いてしまったりすることなど

2017年2月13日受理。(2017年3月9-10日シンポジウム「モバイル17」にて発表)

が懸念されている。NASA-TLX と SWAT については簡便法が提案されているが、オリジナルの手法と同様な感度で作業負担が評価されない場合があることが論じられている[4]。

一方、ユーザの作業負担を単一次元的に捉えて評定する手法としては、Cooper-Harper Rating Scale[5]が提案されており、日本語化もなされている[6]。この手法では、決定木に従ってユーザが質問に「はい」「いいえ」で回答をすすめていくと、評価対象となる作業負担の評定値を 1 から 10 の整数値で与えることができる。NASA-TLX や SWAT, WP などの手法と比較すると、より簡易的に作業負担の評価が可能であり、ユーザが手法を利用することに伴う作業負担もそれほど高まらないという利点がある。しかしながら、Cooper-Harper Rating Scale による作業負担の評定値は 10 段階の分解能であることから、作業負担の小さな違いが同じ評定値に分類されてしまう可能性を含んでいる。ユーザの心理量を単一次元的に捉えて評定する手法としては VAS (Visual Analogue Scale)がある[7]。VAS は痛みの程度を評価する手法として開発され、0 から 100 などのある長さの直線に対して、感じた痛みがそのうちのどこに当たるかを記入させることで評定する方法であり、VAS の評定用の直線には目盛りを付けないといった特徴がある。この VAS を参考にして、作業遂行に対してユーザがどの程度の精神的努力を割当てているのかという観点から、評定用の直線に付した複数のラベルをユーザに参照させて作業遂行に伴う作業負担を評定する手法が Rating Scale Mental Effort (RSME)である。次章では、RSME の概要と本研究の目的について述べる。

2. RSME の概要と本研究の目的

Zijlstra はユーザの作業負担にかかる「精神的な努力」について単一次元的に評定するための評価尺度 Rating Scale Mental Effort (RSME)を開発した[8]。オランダ語で作成されたオリジナルの RSME(以後、オランダ語版 RSME)の評定スケール[8]ならびに英語で作成された RSME(以後、英語版 RSME)の評定スケール[9]を図 1 に示す。評定スケールは、150mm の縦方向の線分に 10mm 間隔の目盛りと精神的努力の程度を言語で表現した 9 つのラベルを付した基準点が設定され、英語の評定スケールを例に挙げると、“Absolutely no effort”(評定スケール上の 0 点付近)から“Extreme effort”(評定スケール上の約 112 点付近)に至る表現となっている。これらのラベルの基準点は必ずしも等間隔に配置されていない。モバイル機器とのインタラクションや HMI 仕様に伴う精神的努力の程度を評価する場合を例にとると、ユーザはこれらの評定スケールと精神的努力の程度を表現したラベルに基づいて、そのインタラクションや HMI 仕様に対して費やした精神的努力の程度に該当する位置に印を付け、その印された位置が評定値として 0 から 150 までの数値で抽出される。RSME の実施に要する時間は、単一次元的評定スケールであることから、精神的努力を

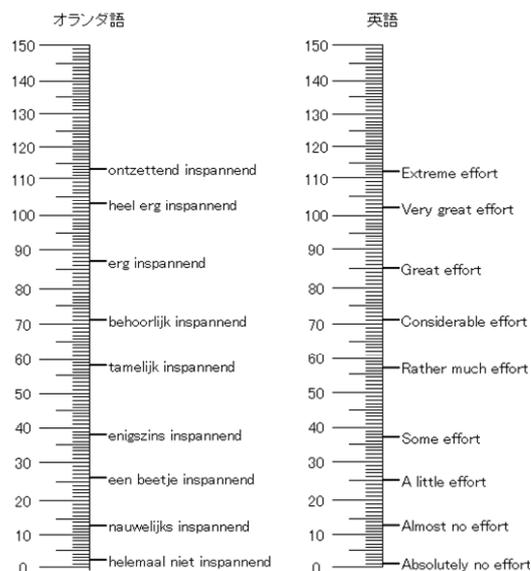


図 1 Rating Scale Mental Effort (RSME)

作業負担として捉えた場合、作業負担評価に要する時間が短くなるものと期待される。その一方で、RSME に基づく評定作業は、評定スケールに付されたラベルの解釈やその基準値によって大きく影響される。このラベルの解釈や基準値については、国や地域、文化などの違いによって異なることが報告されており[10]、インドネシア人を対象にした RSME の開発では、オリジナルの RSME と比較してラベル数の削減や基準値の変更などが行われている。これと同様に、日本人を対象にした RSME の開発、つまり RSME の日本語化にあたっては、単に英語を日本語に翻訳するのではなく、日本人が感じる精神的努力の程度に対して適切な日本語ラベルを複数設定することが必要となる。

そこで本研究では、Zijlstra の RSME の開発方法[8]ならびに Widyanti らの RSME のラベル作成方法[10]を参考にして、RSME の日本語化に関して具体的な評定スケール、つまり、日本語ラベルとその基準値を提案することを目的とする。まず RSME の具体的な日本語ラベルの候補を抽出し、それら日本語ラベルの候補に対する基準値を確認する。抽出された各基準値に基づいて日本語ラベルの候補を再検討する。さらに記憶課題と選択課題を伴う検証実験を通じて、日本語化した RSME の評定スケールの妥当性について確認する。

3. RSME における日本語ラベルの候補検討

3.1 日本語ラベルの候補抽出方法

Widyanti らの RSME のラベル作成方法[10]と同様に、表 1 に示される英語版 RSME の 9 つの英語ラベルの書かれた A4 サイズの質問紙を用意して実験参加者に日本語に翻訳するよう依頼した。英語版 RSME の 9 つの英語ラベルは、ラベルの配置順序による影響を考慮して、質問紙の左側でランダムな

表 1 英語版 RSME ラベル

英語版 RSME
Extreme effort
Very great effort
Great effort
Considerable effort
Rather much effort
Some effort
A little effort
Almost no effort
Absolutely no effort

表 2 英語版 RSME ラベルに対する日本語ラベルの結果

英語版 RSME	日本語ラベル
Extreme effort	極めて大きな努力
Very great effort	とても大きな努力
Great effort	大きな努力
Considerable effort	相当な努力
Rather much effort	かなり大きな努力
Some effort	多少の努力
A little effort	少しの努力
Almost no effort	ほとんど努力なし
Absolutely no effort	まったく努力なし

順序で並べられた。実験参加者は質問紙上に配置された各英語ラベルの右側部分に日本語ラベルを記入するよう指示された。実験参加者が英語版 RSME の英語ラベルを日本語に翻訳するにあたっては、英語辞書電子版(リーダーズ英和辞典第 2 版:株式会社研究社出版, 電子辞書:CASIO 製)の利用が可能であった。実験参加者は TOEIC スコア 690 点以上の日本人大学生 11 名(平均年齢 22.1 歳(標準偏差 0.51), 男性 9 名, 女性 2 名)であった

3.2 日本語ラベルの候補抽出の結果

日本語への翻訳の結果, 各英語ラベルは実験参加者により多様な日本語ラベルに翻訳され, 英語ラベルの各々に対して, 抽出された候補は 2 種類以上となった。それらは, 「努力」という単語を同程度に修飾する形容詞や関連した副詞などが利用されたものの他, 英語ラベルを上手く翻訳できなかったと考えられるもの(常軌を逸した, 崇高な, 素晴らしい, …, など)も見られた。これらに関して, 国語辞典を用いて精神的努力を表現する形容詞, 関連した副詞として適切かどうかをある程度確認した。その結果, 適切でないと判断された日本語ラベルは少数であったため, それらについては削除した。この段階で多く抽出された形容詞ならびに副詞を伴った日本語ラベルを表 2 に示す。実験参加者から得られた英語ラベルの翻訳結果に配慮して, 「大きな」「相当な」「多少の」「少し」「(…)なし」などの形容詞, また「極めて」「とても」「かなり」「ほとんど」「まったく」など副詞を日本語ラベルの修飾語句として抽出した。これらを初期段階での日本語ラベルの候補として設定し, 各ラベルの基準値について検討を行うこととした。

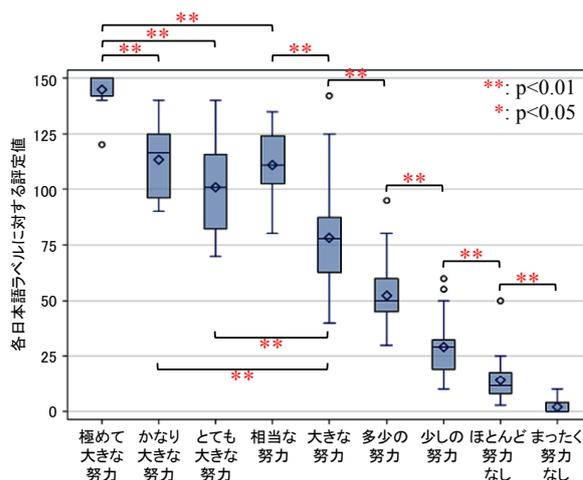


図 2 日本語ラベル候補の基準値結果

4. 日本語ラベルの各候補に対する基準値検討

4.1 日本語ラベルの各候補に対する基準値抽出方法

図 1 に示した RSME の縦型 150mm の評定スケールからすべてのラベルを取り除いた評定スケールを用意した。実験参加者には, 初期段階での候補となった表 2 に示される日本語ラベルがこの評定スケール上のどの位置に相当するかを印すように依頼した。その際, 実験参加者において最大限の精神的努力を行っている場合を 150 点に, 実験参加者において精神的努力を全く行っていない場合を 0 点にそれぞれ設定すること, 日本語ラベルのいずれかの候補を必ず 0 点または 150 点の位置に印す必要がないこと, 評定スケールの目盛りちょうどどの位置に印す必要がないことを実験参加者に教示した。実験参加者は, 日本語ラベルの翻訳に参加した日本人大学生 11 名であった。

4.2 日本語ラベルの各候補に対する基準値結果

日本語ラベルの各候補に対する基準値の結果を箱ひげ図としてまとめた。その結果を図 2 に示す。評定値の結果がおおよそ近い日本語ラベル同士を比較の一組として設定して, Brunner-Munzel 検定を実施したところ, 一部の比較の組を除いて有意差が見られた。また英語版 RSME のラベルと比較すると, 精神的努力を多く要する日本語ラベルの候補では, 評定値が高く評価される傾向が見られ, 英語版「Extreme effort」に相当する「極めて大きな努力」の評定値の結果は 145 点付近であり, 英語版 RSME ラベルの基準値 112 点付近とは大きく異なる結果となった。「かなり大きな努力」「とても大きな努力」「相当な努力」の日本語ラベルについては, 異なる日本語ラベルであってもほぼ同様な評定値となった。

4.3 日本語ラベルの各候補に対する基準値追加検討

前節にて抽出された日本語ラベルの候補と基準値について, 評定値が 10 から 120 までの日本語ラベルに対する評定値の特徴を確認するため, 英語版 RSME の「Extreme effort」と「Absolutely no effort」を除いた 7 つの英語ラベルについて実

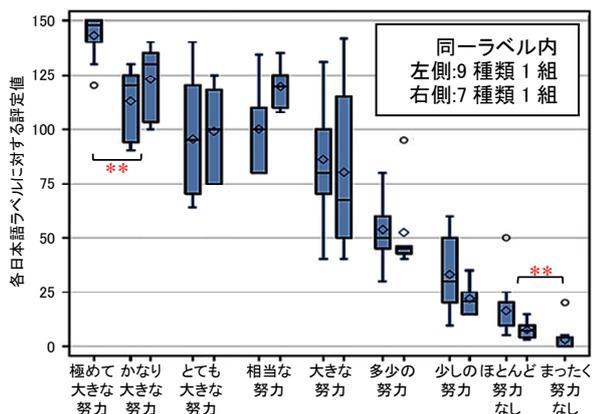


図3 日本語ラベル(9種類1組, 7種類1組)の各候補と基準値の結果

表3 日本語化 RSME ラベルの候補

日本語化 RSME ラベル (9種類1組の場合)	日本語化 RSME ラベル (7種類1組の場合)
極めて多くの努力	極めて多くの努力
とても多くの努力	
かなり多くの努力	かなり多くの努力
相当な努力	
多くの努力	多くの努力
多少の努力	多少の努力
少しの努力	少しの努力
ほとんど努力なし	ほとんど努力なし
まったく努力なし	まったく努力なし

験参加者に日本語ラベルへの翻訳と各日本語ラベルに対する評定を依頼した。これら 7 つの英語ラベルの翻訳や評定が残り 2 つの英語ラベルの存在によってどのように影響されるのかも確認するために、9 つの英語ラベルの翻訳と評定も別途依頼し、それらの特徴を比較した。実験参加者は TOEIC スコアが 690 点以上の日本語を母国語とする 8 名の学生(平均 22.3 歳(標準偏差 0.89), 男性 6 名, 女性 2 名)であった。

7 つの英語ラベルの翻訳と評定の結果ならびに、比較のために収集した 9 つの英語ラベルの翻訳と評定の結果を図 3 に示す。「Extreme effort」と「Absolutely no effort」を取り除いた 7 つの英語ラベルの場合と 9 つの英語ラベルの場合で、英語ラベルが共通なものについては同じ日本語ラベルが得られた。7 つの英語ラベルと 9 つの英語ラベルから得られた共通の日本語ラベルのいくつかの評定値については同様な値であったが、隣接する日本語ラベルの組の一部の評定値について Brunner-Munzel 検定を実施したところ、9種類1組の日本語ラベルにおける「極めて大きな努力」と7種類1組の日本語ラベルにおける「かなり大きな努力」の間、9種類1組の日本語ラベルにおける「まったく努力なし」と7種類1組の日本語ラベルにおける「ほとんど努力なし」の間にそれぞれ有意差が見られた ($p < 0.01$, $p < 0.01$)。150mm の評定スケールに対して、9種類1組の日本語ラベル、7種類1組の日本語ラベルともに「かなり大きな努力」「とても大きな努力」「相当な努力」「大きな努力」に対する評定値の幅が大きく、特に「かなり大きな努力」「とても大きな努力」「相当な努力」の日本語ラベルでは精神的努力の程度に対して異なる評定値への分解が困難である可能性が示唆された。

5. 日本語 RSME ラベルの再検討と評定実験

精神的努力の程度を表現する日本語ラベルのうち、「かなり大きな努力」「とても大きな努力」「相当な努力」「大きな努力」については評定値の幅が大きく、特に「かなり大きな努力」と

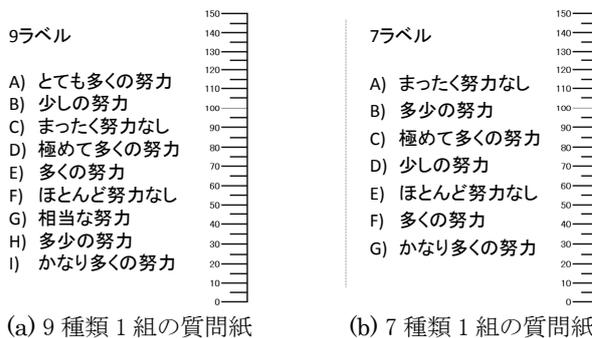


図4 本実験で使用した2種類の日本語ラベルの質問紙

でも大きな努力」「相当な努力」の日本語ラベルでは評定値にあまり違いが見られなかったことから、これら 3 つのラベルを一つにまとめ、「かなり大きな努力」という表現のラベルを設定した。一方、精神的努力の程度を表現する形容詞に関して、前述の実験から得られた結果では「大きい」などの表現が抽出されていた。心理尺度としてのラベル表現の適切性・妥当性について、人間工学分野の研究者 2 名に意見収集を行ったところ、「精神的努力」の量や程度を形容する表現に関しては「大きい」よりも「多い」という表現が適当であると考えられること、科学技術文献情報データベース JDreamIII のアドバンスドサーチでの検索ヒット件数では、「大きな努力」に対して 584 件、「多くの努力」に対して 1467 件の検索結果(いずれも 2016 年 1 月 28 日アクセス)が得られたことなどから、日本語版 RSME の日本語ラベルを再設定することとし、表 3 に示されるように、日本語ラベルを 9種類1組とする場合と 7種類1組とする場合の 2 種類を比較検討することとした。本章では、日本語化 RSME ラベルの 9種類1組の場合と 7種類1組の場合の各ラベルから認識される精神的努力の程度について評定値を収集・確認するとともに、9種類1組の場合と 7種類1組の場合の妥当性について比較検討することを目的とする。

5.1 日本語ラベル 2 種類に関する実験方法

実験参加者 72 名全員が一度に着座可能な大教室にて、隣の実験参加者と 1 席以上間を空けて座るよう指示した。日本

語ラベルの評定のための質問紙の例を図 4 に示す。質問紙は 2 種類の日本語ラベル(9 種類 1 組の日本語ラベルと 7 種類 1 組の日本語ラベル)のそれぞれに対して 1 枚の質問紙を用意し、各質問紙の左側部分に該当する日本語ラベルをランダムに配置し、右側部分に 0~150mm の縦型の評定スケールをそれぞれ配置した。4 章の調査実験と同様に、実験参加者には質問紙上に記載された各日本語ラベルが評定スケール上のどの位置に相当するかを印すように依頼した。日本語を母国語とする理工学系の学部生 72 名を対象に、9 種類 1 組の日本語ラベルの調査に対して実験参加者 36 名(19 から 22 歳までで平均 19.9 歳(標準偏差 0.67), 男性 29 名, 女性 7 名)が、7 種類 1 組の日本語ラベルの調査に対して実験参加者 36 名(19 から 23 歳までで平均 21.1 歳(標準偏差 0.97), 男性 28 名, 女性 8 名)がそれぞれ参加した。

5.2 日本語ラベルの各候補に対する基準値結果

日本語化 RSME における 9 種類 1 組および 7 種類 1 組の日本語ラベルに対して評定された評定値の結果を図 5 に示す。各日本語ラベルに対して左側の箱ひげは 9 種類 1 組での結果、右側の箱ひげは 7 種類 1 組の結果をそれぞれ示している。日本語ラベルによって評定値のばらつきが異なり、それらは 9 種類 1 組の場合と 7 種類 1 組の場合で異なる傾向が見られた。各日本語ラベルに対する評定値の違い、ならびに 9 種類 1 組の場合と 7 種類 1 組の場合に対する評定値の違いを検討するためにノンパラメトリック検定である Steel-Dwass の方法を実施した。その結果を表 4 および表 5 にそれぞれ示す。日本語ラベルの 7 種類 1 組の評定値では、任意の異なる日本語ラベル間において有意差(p<0.01)が見られたのに対して、日本語ラベルの 9 種類 1 組の評定値では、「かなり多くの努力」「とても多くの努力」「相当な努力」の 3 つの評定値の間には有意差は見られなかった。これらの結果に基づくと、9 種類 1 組の日本語ラベルにおいて「とても多くの努力」「かなり多くの努力」「相当な努力」は精神的努力の程度を基準値として設定した場合でも明確に異なる基準値を設定することは困難であると考えられることから、評定スケール上で一つの日本語ラベルにまとめ

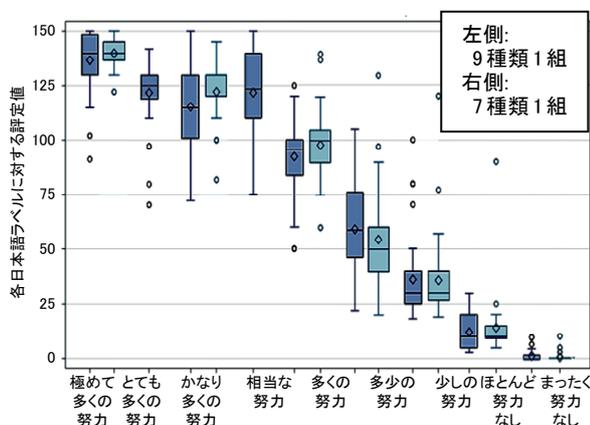


図 5 2 種類の日本語化 RSME ラベルの結果

表 4 9 種類 1 組のラベルの評定値に対する検定結果

ラベル数	①極めて多くの	②かなり多くの	③とても多くの	④相当な	⑤多くの	⑥多少の	⑦少しの	⑧ほとんど努力なし	⑨まったく努力なし
①		**	**	*	**	**	**	**	**
②	**				**	**	**	**	**
③	**				**	**	**	**	**
④	*				**	**	**	**	**
⑤	**	**	**	**		**	**	**	**
⑥	**	**	**	**	**		**	**	**
⑦	**	**	**	**	**	**		**	**
⑧	**	**	**	**	*	**	**		**
⑨	**	**	**	**	**	**	**	**	

*:p<0.05 **:p<0.01

表 5 7 種類 1 組のラベルの評定値に対する検定結果

ラベル数	①極めて多くの	②かなり多くの	③多くの	④多少の	⑤少しの	⑥ほとんど努力なし	⑦まったく努力なし
①		**	**	**	**	**	**
②	**		**	**	**	**	**
③	**	**		**	**	**	**
④	**	**	**		**	**	**
⑤	**	**	**	**		**	**
⑥	**	**	**	**	**		**
⑦	**	**	**	**	**	**	

*:p<0.05 **:p<0.01

た上で設定するのが適当であり、日本語化 RSME では 7 種類 1 組の日本語ラベルを設定するのが妥当であると考えられる。本実験で抽出された 7 種類 1 組の日本語ラベルに対して得られた評定値の平均値を表 6 に、これらの評定値を各日本語ラベルの基準値として採用した場合の日本語化 RSME の評定スケールを図 6 にそれぞれ示す。図 1 に示したオランダ語版 RSME や英語版 RSME と比較すると、ラベル数の違いがあり、また大きく異なる傾向がうかがえる。

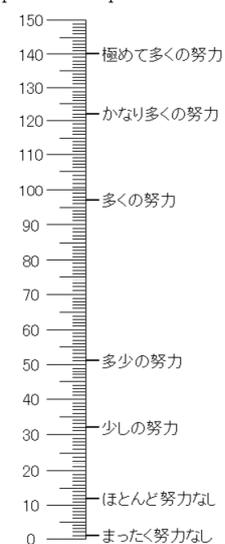


図 6 本研究で得られた日本語化 RSME の評定スケール

表 6 日本語化 RSME ラベルと基準値

日本語化 RSME ラベル	基準値
極めて多くの努力	139.2
かなり多くの努力	121.5
多くの努力	96.8
多少の努力	51.1
少しの努力	32.7
ほとんど努力なし	12.1
まったく努力なし	0.6

6. 日本語化 RSME の適応性と有用性の確認実験

前章で得られた日本語化 RSME および日本語ラベルの基準値が精神的努力の程度を感度よく評価可能であることを検証することを目的とする。NASA-TLX のワークロード得点 (WWL 得点) や NASA-TLX の努力項目の評定値と RSME の評定値を比較し、日本語化 RSME の適応可能性についても検討する。また、実験参加者が NASA-TLX を用いてワークロードを評価するのに要した時間と日本語化 RSME のそれらを比較し、日本語化 RSME の効率性についても検討する。

6.1 確認実験の実験方法

実験参加者は日本語を母国語とする 10 名の理工系学生 (21 から 24 歳までで平均 22.2 歳 (標準偏差 0.84)、男性 5 名、女性 5 名) であった。実験者が用意した 3 つの異なる難易度を有する記憶・選択反応課題 (以後、実験タスクと呼ぶ) を実施した。実験参加者に実験タスクの実施時に要求される精神的努力の程度を日本語化 RSME、NASA-TLX に基づいて評価させた。計測項目は、実験タスクの反応時間、実験タスクの正答率、提案作成した日本語化 RSME の評定値、NASA-TLX のワークロード得点および努力項目の評定点、RSME および NASA-TLX の評定時間であった。実験参加者には長机の右側に着座させ、質問紙への回答および実験タスクに集中できる静かな環境を用意した。実験者は実験参加者の右側に一席空けて着座して実験参加者に指示を与えた。実験参加者は、机の端から約 20cm 離れた場所にノート PC を設置して各実験タスクを実施した。実験タスクとして難度の異なる下記(a) から(c)のタスクを設定した。

(a) 1 文字タスク ノート PC 画面上で最大 2 秒間提示される数字 (1 文字) が 2 秒ごとに次々と切り替わる。ただし、最初の提示のみ 0.8 秒である。提示された数字と直前に提示された数字を切り替わり前後で比較して一致する数字の個数 (文字の個数) を該当するボタンを押下して解答する (図 7 参照)。実験参加者は直前に提示された数字を常に記憶し、新たに提示される数字と入れ替える必要がある。

(b) 3 文字タスク ノート PC 画面上で最大 2 秒間提示される数字の組合せ (3 文字) が 2 秒ごとに次々と切り替わる。ただ

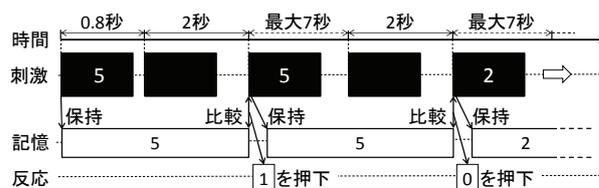


図 7 1 文字タスクの概要図

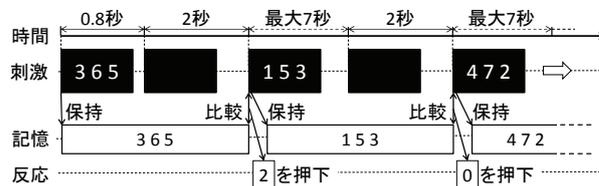


図 8 3 文字タスクの概要図

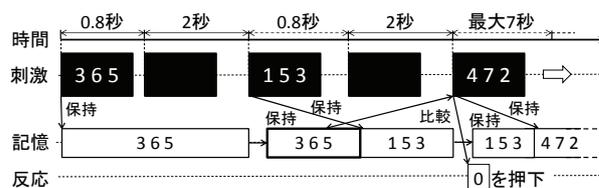


図 9 3 文字タスク+2Back の概要図

し、最初の提示のみ 0.8 秒である。提示された数字の組合せと直前に提示された数字の組合せを切り替わり前後で比較して一致する数字の個数 (文字の個数) を該当するボタンを押下して解答する (図 8 参照)。実験参加者は直前に提示された数字の組合せ (3 文字) を常に記憶し、新たに提示される数字の組合せ (3 文字) と入れ替える必要がある。

(c) 3 文字 2Back タスク ノート PC 画面上で最大 2 秒毎に切り替わる数字の組合せ (3 文字) が 2 秒ごとに次々と切り替わる。ただし、最初と 2 つ目の提示のみ 0.8 秒である。提示された数字の組合せと 2 つ前に提示された数字の組合せを比較して一致する数字の個数 (文字の個数) を該当するボタンを押下して解答する (図 9 参照)。実験参加者は 2 つ前および直前に提示されたそれぞれの数字の組合せ (3 文字) を常に記憶し、新たに提示される数字の組合せ (3 文字) と記憶されている数字を一部入れ替える必要がある。

実験タスクの内容や回答方法に習熟させるため、実験参加者には練習用の実験タスクを 1, 2 回体験させた後、データ計測のための実験を実施した。各実験タスクは 29 間でこれを 1 セットとし、各実験タスクに対して 2 セットずつ実施した。各実験タスクの実施順序はランダムとしたが、実験参加者が各実験タスクを平均的に経験するように配慮して、3 種類の実験タスクを 1 セットずつ実施した後に 2 回目の実験タスクを実施するように設定した。

6.2 確認実験の実験結果

各実験タスクにおける反応時間および正答率の結果を図 10 に示す。1 文字タスクでは反応時間は約 0.5 秒で正答率も 98% であったのに対して、3 文字タスクでは反応時間は約 1.7

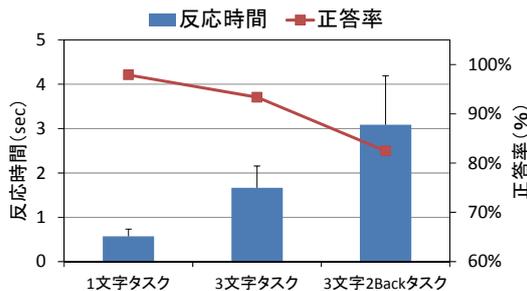


図 10 各実験タスクに対する反応時間および正答率

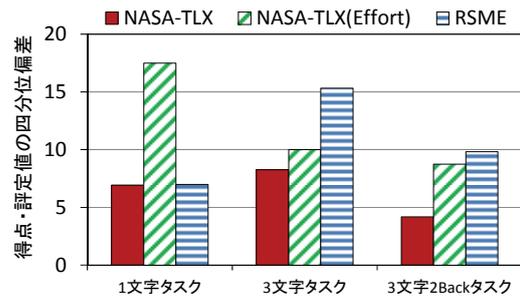


図 12 各実験タスクに対する評定値・得点の四分位偏差

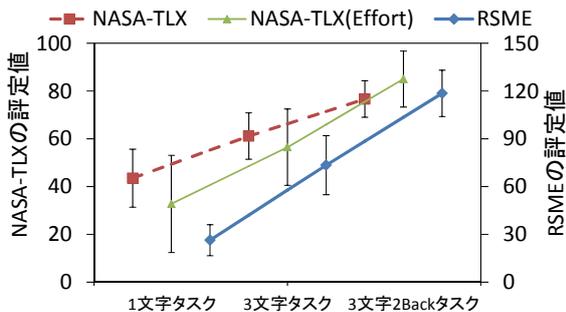


図 11 日本語 RSME と NASA-TLX の評定値・得点

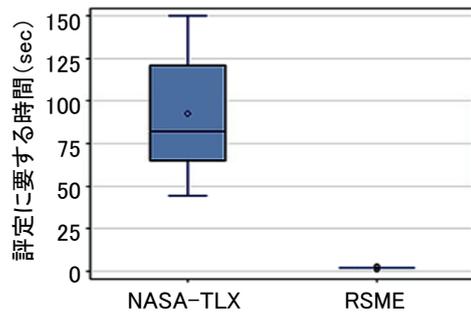


図 13 日本語 RSME と NASA-TLX の評定時間

表 7 実験タスク間での評定値・得点の検定結果

	1文字タスク	3文字タスク	3文字2Backタスク
1文字タスク		***/**/***	***/***/***
3文字タスク	***/**/**		***/***/***
3文字2Backタスク	***/***/**	***/***/***	

： $p < 0.01$ *： $p < 0.001$ (NASA-TLX のワークロード得点 / NASA-TLX の努力項目の評定値 / RSME の評定値)

秒、正答率は 93%、さらに 3 文字 2Back タスクでは反応時間は約 3 秒、正答率は 85%であり、記憶された数字と提示された数字の比較などで費やされた時間が増加したことがうかがえ、タスクの難度が増加した状況が推察される。これらの反応時間に対して、ノンパラメトリック検定である Steel-Dwass の方法を実施した結果、3 種類の実験タスク間で有意差が見られた ($p < 0.001$, $p < 0.001$, $p < 0.001$)。各実験タスクの遂行に対して、日本語化 RSME による評定値ならびに NASA-TLX によるワークロード得点、努力項目の評定値の結果を図 11 に示す。図中の点線(赤色)は NASA-TLX によるワークロード得点、実線(緑色)は NASA-TLX の努力項目の評定値、実線(青色)は日本語化 RSME による評定値の結果をそれぞれ示している。いずれの場合でも 1 文字タスク、3 文字タスク、3 文字 2Back タスクの順に得点や評定値が高まり、精神的な努力あるいは精神的な負担が増加している状況がうかがえた。1 文字タスクと 3 文字 2Back タスクに対する NASA-TLX の得点や評定値、日本語化 RSME の評定値の差を比較した場合、NASA-TLX、日本語化 RSME の各評定スケールの最大値を考慮に入れると、日本語化 RSME の評定値は、各実験タスクの差を大きく捉えている状況がうかがえた。また各実験タスクに対する NASA-TLX

の得点や評定値、日本語化 RSME の評定値の分布の特徴を確認したところ、正規性や等分散性を有さないものが見られた。3 種類の実験タスク間における得点や評定値の違いを検討するために Steel-Dwass の方法を実施した。その結果を表 7 に示す。実験タスク間でそれぞれ有意差が見られ、各実験タスクに対して異なる得点、異なる評定値として評価されている状況がうかがえた。各実験タスクに対する NASA-TLX の得点や評定値、日本語化 RSME の評定値のばらつきの特徴を検討するために四分位偏差を算出した。その結果を図 12 に示す。NASA-TLX のワークロード得点の四分位偏差(第 1 四分位点と第 3 四分位点の差の 1/2)は、1 文字タスクを除いて他よりも小さい値を取る傾向が見られたが、NASA-TLX の得点や評定値では最大値が 100 であること、日本語化 RSME の評定値では最大値が 150 であることを考慮に入れると、全体的に日本語化 RSME の評定値の四分位偏差は小さく、1 文字タスクや 3 文字 2Back タスクでは NASA-TLX の努力項目の評定値の四分位偏差よりも小さい傾向が見られた。NASA-TLX、日本語化 RSME での評定に要した時間の箱ひげ図を図 13 に示す。使用したデータは NASA-TLX、日本語化 RSME での評定に実験参加者が慣れたと考えられる 4 セット目から 6 セット目までの実験タスクの評定に要した時間を対象とした。NASA-TLX では 6 種類の評価項目に対する一対比較および評定作業を実施する必要があったことから平均 92.5 秒を要したのに対して、日本語化 RSME では単一の評価項目での評定作業であったため、平均で 2.13 秒程度と短く、NASA-TLX と比較して大幅に評定時間が短縮される結果となった。

6.3 考察

確認実験では、難度の異なる 3 種類の記憶・選択反応課題

を実験タスクとして実験参加者に実施させ、日本語化 RSME および NASA-TLX により評価させた。その結果、日本語化 RSME の評定値は、NASA-TLX のワークロード得点や努力項目の評定値と比較して、実験タスク間の違いを大きく捉える傾向があり、また NASA-TLX の努力項目の評定値と比較して日本語化 RSME の評定値そのもののばらつきも少ない傾向が見られた。実験タスクの遂行に伴う精神的努力の程度が評定スケール上のどの位置に相当するかを印すという点においては、評定スケールの長さや方向は異なるが、日本語ラベルの有無を除き、日本語化 RSME による評定と NASA-TLX の努力項目における評定はほぼ同等であったと考えられる。しかしながら、各実験タスクに対する評定値の差や評定値のばらつきの特徴を考慮すると、精神的努力の程度を表現した日本語ラベルを評定スケールに付与したことが、評定スケール上において実験タスク間の評定値の差の拡大と評定値のばらつきの減少に寄与したものと考えられる。また、確認実験において設定した 3 種類の実験タスクのうち、難度が最も高い 3 文字 2Back タスクでは、NASA-TLX のワークロード得点や努力項目の評定値、日本語化 RSME の評定値はそれぞれの評定スケールの長さに対して相対的にほぼ同様な位置が印されていたのに対して、3 文字タスク、1 文字タスクでは、日本語化 RSME の評定値が NASA-TLX の得点や評定値よりも相対的に小さい値を示す位置に印されていたことを考慮に入れると、NASA-TLX の努力項目と比較して、日本語 RSME は、精神的努力の程度が小さい実験タスクに対して感度が高い可能性が示唆される。

日本語化 RSME は、単一の評価項目であることから短時間で評定可能であることが確認されたが、NASA-TLX の努力項目の評定についても同じ理由により短時間で評定可能であると考えられる。しかしながら、前述の通り、NASA-TLX の努力項目の評定値は大きくばらつく傾向があるため、日本語化 RSME の方が有用であると考えられる。NASA-TLX のワークロード得点は、心理的、精神的な要素だけでなく、身体的な負担なども含めた多次元的な評価項目に基づいて推定される実績のある評価値であるが、NASA-TLX に要する評定時間は一般的に長い傾向にある。一対比較を実施しないなどの簡便法の適用により短縮化が図れるが、その場合でも 6 種類の評価項目について評定する必要があり、評定時間の短縮化には限界があることから、評価対象となるタスクによっては、短時間で評定可能な日本語化 RSME が有用であると考えられる。

7. まとめ

本研究では、Rating Scale Mental Effort (RSME) の日本語化に関する実験検討ならびに提案した日本語 RSME の評定スケールの妥当性について検討した。その結果、7 種類 1 組の日本語ラベルを利用した日本語化 RSME が提案され、検証実験でも NASA-TLX と同様な感度、タスクの種類によっては高い

感度であったことから、本研究で提案した日本語 RSME の評定スケールは妥当であることが示唆される。今後は、基準となりうる様々なタスクに適用し、他の作業負担の指標と比較しつつ、日本語化 RSME の妥当性について検討する予定である。

参考文献

- [1] S.G. Hart et al.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, P.A. Hancock et al. (Eds.), Human Mental Workload, pp.139-183, North-Holland (1988)
- [2] G.B. Reid et al.: The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload, P.A. Hancock et al. (Eds.), Human Mental Workload, pp.185-218, North-Holland (1988)
- [3] P.S. Tsang et al.: Diagnosticity and multidimensional subjective workload ratings, Ergonomics, Vol.39, No.3, pp.358-381 (1996)
- [4] 宮家晋司ほか: 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, Vol.29, No.6, pp.399-408 (1993)
- [5] G.E. Cooper et al.: The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, The use of pilot rating in the evaluation of aircraft handling qualities, Report TN-D-5153, NASA, Ames Research Center, Moffett Field, California (1969)
- [6] 芳賀繁: メンタルワークロードの理論と測定, 日本出版サービス (2001)
- [7] T.G. Kantor: Chemical mediators and treatment of pain in rheumatic diseases, Baillieres Clin Rheumatol, Vol.1, No.1, pp.57-70 (1987)
- [8] F.R.H. Zijlstra: Efficiency in Work Behavior: a Design Approach for Modern Tools, doctoral thesis, Delft University of Technology, <http://repository.tudelft.nl/view/ir/uuid%3Ad97a028b-c3dc-4930-b2ab-a7877993a17f/> (1993) [Accessed 2017.2.13]
- [9] D. De Waad: The Measurement of Drivers' Mental Workload, Doctoral thesis, University of Groningen, <http://www.home.zonnet.nl/waad2/mwl.html> (1996) [Accessed 2017.2.13]
- [10] A. Widyanti et al.: Adaptation of the Rating Scale Mental Effort (RSME) for use in Indonesia, International Journal of Industrial Ergonomics, Vol.43, pp.70-76 (2013)

著者紹介

大門 樹(正会員)

慶應義塾大学理工学部教授。自動車のヒューマンマシンインターフェース、ドライバー行動・特性などの研究に従事。

土屋 有加(非会員)

2016年3月 慶應義塾大学理工学部卒業, 2016年4月より、情報系企業に入社。