

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E SISTEMAS

TIAGO MACHADO E SILVA

**CARGA DE TRABALHO MENTAL: ANÁLISE CRÍTICA DOS MÉTODOS DE
AVALIAÇÃO**

DISSERTAÇÃO

PATO BRANCO
2018

TIAGO MACHADO E SILVA

CARGA DE TRABALHO MENTAL: ANÁLISE CRÍTICA DOS MÉTODOS DE AVALIAÇÃO

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Engenharia de Produção, do Programa de Pós-Graduação em Engenharia de Produção e Sistemas - PPGEPS, da Universidade Tecnológica Federal do Paraná. Área de Concentração: Gestão dos Sistemas Produtivos. Linha de pesquisa: Engenharia Organizacional e do Trabalho.

Orientador: Prof. Dr. Sergio Luiz Ribas Pessa.

PATO BRANCO

2018

S586c Silva, Tiago Machado e.
Carga de trabalho mental: análise crítica dos métodos de avaliação /
Tiago Machado e Silva. – 2018.
233 f. : il. ; 30 cm.

Orientador: Prof. Dr. Sergio Luiz Ribas Pessa
Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná.
Programa de Pós-Graduação em Engenharia de Produção e Sistemas.
Pato Branco, PR, 2018.
Bibliografia: f. 202 - 221.

1. Saúde mental. 2. Cognição. 3. Psicologia industrial. 4. Ergonomia.
I. Pessa, Sergio Luiz Ribas, orient. II. Universidade Tecnológica Federal do
Paraná. Programa de Pós-Graduação em Engenharia de Produção e
Sistemas. III. Título.

CDD 22. ed. 670.42

Ficha Catalográfica elaborada por
Suélem Belmudes Cardoso CRB9/1630
Biblioteca da UTFPR Campus Pato Branco



TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº 41

A Dissertação de Mestrado intitulada "**Carga de trabalho mental: análise crítica dos métodos de avaliação**", defendida em sessão pública pelo candidato **Tiago Machado e Silva**, no dia 10 de outubro de 2018, foi julgada para a obtenção do título de Mestre em Engenharia de Produção e Sistemas, área de concentração Gestão dos Sistemas Produtivos, e aprovada em sua forma final, pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas.

BANCA EXAMINADORA:

Prof. Dr. Sergio Luiz Ribas Pessa - Presidente - UTFPR

Prof^a Dr^a Lizandra Garcia Lupi Vergara - UFSC

Prof. Dr. Sergio Paes de Barros - UTFPR

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Pato Branco, 23 de outubro de 2018.

Carimbo e assinatura do Coordenador do Programa.

DEDICATÓRIA

Ao meu pai, Edson, e a minha mãe, Margarida, que sempre me deram amor, educação e apoio incondicional.

Ao meu irmão Junior, pela companhia de sempre, independente da distância.

A minha irmã Francini, pelo exemplo de irmã mais velha.

Eu amo vocês!

AGRADECIMENTOS

Primeiramente a minha família, por ser a minha base e o meu apoio em todos os momentos da minha vida, por acreditar em mim, pelo incentivo, pelas palavras de carinho, pelas risadas e pelos momentos que guardarei para o resto da minha vida.

Agradeço ao meu pai Edson, por ser o maior exemplo de homem que eu poderia ter e por me ensinar a ser um, pela dedicação em cuidar da nossa família e pelo empenho em buscar algo melhor para si e para os seus filhos.

Agradeço a minha mãe Margarida, pelo amor mais puro que eu poderia receber, por ser o maior exemplo de mulher forte e independente, pela dedicação com a nossa família e por se preocupar diariamente com o meu bem-estar.

Agradeço ao meu irmão Junior, por acompanhar toda a minha trajetória e ser o meu maior companheiro, por me ajudar nos momentos difíceis e por apoiar todas as minhas decisões.

Agradeço a minha irmã Francini, por estar ao meu lado e por cuidar de mim desde o meu primeiro dia de vida, pelo exemplo de mulher guerreira e batalhadora e pelo apoio incondicional.

Agradeço aos meus amigos, que se tornaram a minha segunda família. Agradeço por me mostrarem que eu possuo uma família em Pato Branco, pela companhia, por todos os momentos juntos, pelas jantãs, festas e desabafos. Nunca esquecerei de vocês.

Agradeço, por fim, ao meu orientador, o professor Dr. Pessa, por acreditar em meu potencial, pelo incentivo e pelos ensinamentos.

RESUMO

A carga de trabalho mental (CTM) caracteriza-se como um reflexo da tensão mental, dependendo da tarefa realizada, do ambiente e das condições operacionais específicas, juntamente com a capacidade do trabalhador em responder a essas demandas. A análise da CTM depende da aplicação de ferramentas ou métodos específicos, os quais podem variar dependendo do contexto específico da aplicação, da categoria de trabalhadores analisada, do país de aplicação, da disponibilidade de recursos e/ou do ergonômista. O objetivo do presente estudo consistiu em realizar uma análise crítica das ferramentas utilizadas para avaliar a carga de trabalho mental. Para isso, foi realizada uma revisão sistemática da literatura em artigos publicados no período de 2000 a 2017 disponíveis nas bases eletrônicas *ScienceDirect*, *Scopus* e *Web Of Science*, a qual envolveu uma análise bibliométrica e do conteúdo do portfólio selecionado. Os resultados compilaram 85 artigos alinhados com os objetivos e critérios estabelecidos na pesquisa, dos quais foram destacados os autores, os periódicos, as palavras-chave, as ferramentas utilizadas, os trabalhadores analisados e os países de aplicação das pesquisas. O periódico de maior destaque foi o "*Ergonomics*" e as principais palavras-chave encontradas foram "*mental workload*", "*workload*" e "NASA-TLX". Os métodos mais utilizados foram a análise de medidas fisiológicas e o NASA-TLX, dentre 22 encontrados. Os principais trabalhadores estudados foram os da área da saúde, os pilotos e os controladores de tráfego aéreo e 35% dos trabalhos foram realizados nos Estados Unidos. As 22 ferramentas foram classificadas em medidas fisiológicas, medidas de desempenho e medidas subjetivas, as quais foram ainda rotuladas em medidas subjetivas multidimensionais e unidimensionais. As medidas fisiológicas, de desempenho e algumas medidas subjetivas unidimensionais podem ser medidas continuamente e em tempo real, acompanhando variações na carga de trabalho mental, e não dependem da memória dos participantes. As medidas subjetivas multidimensionais fornecem um diagnóstico sobre as diversas dimensões da CTM, são as ferramentas mais utilizadas, são aplicadas após os procedimentos e não são intrusivas. As medidas fisiológicas possuem procedimentos mais caros e complexos, as medidas de desempenho podem ser intrusivas, as medidas subjetivas unidimensionais fornecem apenas uma pontuação geral para a carga de trabalho e as medidas subjetivas multidimensionais possuem procedimentos mais longos e são dependentes da memória do trabalhador. A grande maioria das ferramentas não possui limitação de países e nem de aplicações em trabalhadores e situações de trabalho. Assim, o critério de escolha do método depende da situação de trabalho, dos objetivos da análise e dos recursos disponíveis, sendo indicado o uso de pelo menos dois métodos categorizados de forma diferente para aumentar a validade dos resultados, idealizando uma análise com as três categorias de métodos.

Palavras-chave: Carga mental. Carga cognitiva. Carga de trabalho. Ergonomia. Revisão sistemática.

ABSTRACT

The mental workload (MW) is characterized as a reflection of mental stress, depending on the task performed, the environment and the specific operating conditions, along with the worker's ability to respond to these demands. The analysis of the MW depend on the application of specific tools or methods, which may vary depending on the specific context of the application, the type of worker analyzed, the country of application, the availability of resources and/or the ergonomist. The main goal of the present study was to perform a critical analysis of the tools used to evaluate the mental workload. For this, a systematic review of the literature was carried out in articles published in the period from 2000 to 2017 available in the electronic databases ScienceDirect, Scopus and Web of Science, which involved a bibliometric analysis and the contents of the selected portfolio. The results compiled 85 articles aligned with the objectives and criteria established in the research, of which the authors, journals, keywords, tools used, workers analyzed and countries of research application were highlighted. The most prominent journal was "Ergonomics" and the main keywords found were "mental workload", "workload" and "NASA-TLX". The most used methods were the analysis of physiological measurements and NASA-TLX, out of 22 found. The main workers studied were health workers, pilots and air traffic controllers and 35% of the studies were performed in the United States. The 22 tools were classified into physiological measures, performance measures and subjective measures, which were further labeled in multidimensional and one-dimensional subjective measures. Physiological measures, performance measures and some one-dimensional subjective measures can be measured continuously and in real time, accompanying variations in mental workload, and it is not depending on the participants' memory. The multidimensional subjective measures provide a diagnosis about the various dimensions of the MW, are the most used tools, are applied after the procedures and are not intrusive. Physiological measurements have more expensive and complex procedures, performance measures may be intrusive, one-dimensional subjective measures provide only a general score for the workload and multidimensional subjective measures have longer procedures and are dependent on worker memory. The vast majority of tools have no country limitation or application in workers and work situations. Thus, the criterion for choosing the method depends on the work situation, the objectives of the analysis and the available resources, and it is indicated to use at least two methods categorized differently to increase the validity of the results, idealizing an analysis with the three categories.

Keywords: Mental workload. Cognitive workload. Workload. Ergonomics. Systematic review.

LISTA DE FIGURAS

Figura 1 - Estrutura da Dissertação.....	21
Figura 2 - Arquitetura Cognitiva de Richard associada ao Conceito de CM.....	32
Figura 3 - Enquadramento Metodológico da Pesquisa.....	37
Figura 4 - Etapas de construção do portfólio bibliográfico.....	40
Figura 5 - Distribuição das publicações por ano.....	44
Figura 6 - Autores com duas ou mais publicações no PB	46
Figura 7 - Publicações das referências e dos artigos do PB por periódico	47
Figura 8 - Principais palavras-chave do PB	48
Figura 9 - Classificação dos trabalhos do PB.....	49
Figura 10 - Indicadores de medidas fisiológicas utilizados para analisar a CTM	53
Figura 11 - Países de aplicação da pesquisa.....	55
Figura 12 - Mapa com países onde foram desenvolvidos os trabalhos do PB	56
Figura 13 - Sistema de rastreamento ocular	64
Figura 14 - Sistema fNIR	68
Figura 15 - Local das temperaturas utilizadas.....	72
Figura 16 - Matriz tridimensional das dimensões do SWAT	97
Figura 17 - Escala de Avaliação do Esforço Percebido (RSME)	111
Figura 18 - Classificação das seis escalas do SURG-TLX.....	117
Figura 19 - Árvore de decisões da Escala de Cooper Harper	155
Figura 20 - Árvore de decisões da MCH	157
Figura 21 - Escala Integrada de Carga de Trabalho (IWS)	164
Figura 22 - Questionário SURG-TLX + GOALS	180
Figura 23 - Diagrama esquemático para cálculo do TLI.....	183
Figura 24 - Linha do tempo dos métodos de avaliação da CTM	187

LISTA DE QUADROS

Quadro 1 - Variáveis analisadas nos primeiros estudos sobre CTM.....	28
Quadro 2 - Comparação entre o nível de carga e o tipo de ação do piloto	30
Quadro 3 - Síntese das características dos conceitos de carga mental.....	31
Quadro 4 - Categorias e definições dos métodos de avaliação da CTM.....	33
Quadro 5 - Vantagens e desvantagens das categorias dos métodos da CTM	33
Quadro 6 - Artigos do PB organizados por ano de publicação.....	43
Quadro 7 - Definição das seis dimensões do NASA-TLX	80
Quadro 8 - Definições dos níveis das dimensões do SWAT	96
Quadro 9 - Borg Scale of Perceived Exertion.....	122
Quadro 10 - BSPE modificada	123
Quadro 11 - Fatores e itens contidos na ESCAM.....	133
Quadro 12 - Descrição da escala ISA	139
Quadro 13 - Principais características do ISTAS 21	144
Quadro 14 - Dimensões psicossociais incluídas na versão média do ISTAS 21	147
Quadro 15 - Avaliação Operacional Global de Habilidades (GOALS).....	174
Quadro 16 - Lista de verificação para dissecação da vesícula biliar	175
Quadro 17 - Classes da carga de trabalho de acordo com o TLI.....	185
Quadro 18 - Quadro resumo dos métodos de avaliação da CTM	196

LISTA DE TABELAS

Tabela 1 - Estatística das variações do FI dos periódicos por ano de publicação dos artigos	45
Tabela 2 - Ferramentas de análise da carga de trabalho mental analisadas por meio de dados estatísticos do FI dos seus periódicos	50
Tabela 3 - Aplicação das ferramentas da carga de trabalho mental por ano	52
Tabela 4 - Trabalhadores analisados nos artigos do PB	57
Tabela 5 - Divisões do ISTAS 21 em seções, escalas e perguntas	145

LISTA DE ABREVIATURAS E SIGLAS

ABERGO	Associação Brasileira de Ergonomia
Ag	Prata
AgCl	Cloreto de prata
AHP	Analytical Hierarchy Process
ANN	Redes neurais artificiais
ANOVA	Análise de variância
ATC	Controlador de Tráfego Aéreo
ATWIT	Air Traffic Workload Input Technique
BIREME	Biblioteca Virtual em Saúde on-line
BSPE	Borg Scale of Perceived Exertion
BVS	Biblioteca Virtual de Saúde
CM	Carga mental
CoPsoQ	Questionário Psicossocial de Copenhague
CH	Cooper Harper Scale
CT	Carga de trabalho
CTM	Carga de trabalho mental
DM	Diabetes Mellitus
DP	Doença Periodontal
D. Padrão	Desvio Padrão
ECG	Eletrocardiograma
EDA	Atividade eletrodérmica
EEG	Eletroencefalograma
EMG	Dados eletromiográficos
EOG	Eletrooculograma
ERPs	Dados eletrocardiográficos e potenciais relacionados a eventos
ESCAM	Escala Subjetiva de Carga Mental de Trabajo
F0	Frequência fundamental
FFT	Transformada Rápida de Fourier
FLS	Fundamentos da cirurgia laparoscópica
FI	Fator de impacto
fNIR	Espectroscopia funcional de infravermelho próximo
Freq.	Frequência

GOALS	Global Operative Assessment of Laparoscopic Skills
GSW	Global Subjective Workload
HR	Frequência cardíaca
HRV	Variabilidade da frequência cardíaca
IBI	Intervalo entre batimentos
ICC	Coeficiente de correlação intraclasse
ISA	Instantaneous Self-Assessment
IWS	Integrated Workload Scale
JCQ	Job Content Questionnaire
JDS	Job Demands Scale
MCH	Modified Cooper Harper Scale
Mdn.	Mediana
MLR	Regressão Multilinear
MW	Mental Workload
N	Número de aeronaves
NASA-TLX	National Aeronautics and Space Administration Task Load Index
NN	Normal ao normal
NNI	Índice de vizinho mais próximo
OSATS	Avaliação Estruturada Objetiva de Habilidades Técnicas
PAHO	Organização Pan-Americana de Saúde
PB	Portfólio Bibliográfico
PHPMS	Pitch high pass mean square
POWER	Performance and Objective Workload Evaluation Research
RHPMS	Roll high pass mean square
RR	Taxa de respiração
RSME	Rating Scale of Mental Effort
RCAT	Ferramenta de Avaliação Cognitiva Rápida
RPE	Rating of Perceived Exertion
RTLX	Raw-NASA-TLX
SC	Condutância cutânea tônica
SDNN	Desvio padrão do intervalo normal ao normal
SOI	Simple Objective Index
SUSESO-ISTAS 21	Questionário de Avaliação de Risco Psicossocial
SURG-TLX	Surgical Task Load Index

SWAT	Subjective Workload Assessment Technique
Tf	Temperatura da testa
TJDFT	Tribunal de Justiça do Distrito Federal e dos Territórios
TLI	Traffic Load Index
Tn	Temperatura nasal
UAS	Unmanned Aerial System
VAS	Escala de analogia visual
W	Coeficiente de Concordância de Kendall
WAK	Workload Assessment Keypad
WCI/TE	Workload-Compensation-Interference/ Technical Effectiveness scale
XTL	Carga de Tarefa Cognitiva Externa

SUMÁRIO

1 INTRODUÇÃO	16
1.1 CONTEXTUALIZAÇÃO	16
1.2 PROBLEMA DE PESQUISA	18
1.3 OBJETIVOS	18
1.3.1 Objetivo Geral	18
1.3.2 Objetivos Específicos	19
1.4 JUSTIFICATIVA	19
1.5 ESTRUTURA DA DISSERTAÇÃO	20
2 REFERENCIAL TEÓRICO	22
2.1 ERGONOMIA	22
2.2 CARGA DE TRABALHO	23
2.3 CARGA DE TRABALHO MENTAL.....	27
2.3.1 Métodos de mensuração da carga de trabalho mental	32
2.4 TRABALHOS DE REVISÃO DE LITERATURA.....	34
3 METODOLOGIA	37
3.1 ENQUADRAMENTO METODOLÓGICO.....	37
3.2 REVISÃO SISTEMÁTICA DE LITERATURA	38
3.3 ANÁLISE DOS MÉTODOS DE AVALIAÇÃO DA CTM	41
4 RESULTADOS E DISCUSSÕES	43
4.1 REVISÃO SISTEMÁTICA DE LITERATURA	43
4.1.1 Análise Bibliométrica	43
4.1.2 Análise do Conteúdo	48
4.2 MÉTODOS DE AVALIAÇÃO DA CARGA DE TRABALHO MENTAL.....	58
4.2.1 Medidas Fisiológicas	58
4.2.1.1 Coração.....	60
4.2.1.2 Olhos	63
4.2.1.3 Cérebro	66
4.2.1.4 Pele	69
4.2.1.5 Respiração	70
4.2.1.6 Temperatura facial.....	71
4.2.1.7 Fala	72
4.2.1.8 Movimentos corporais	73

4.2.1.9 Urina.....	73
4.2.1.10 Visão dos autores sobre as medidas fisiológicas	74
4.2.2 National Aeronautics and Space Administration Task Load Index (NASA-TLX)	79
4.2.2.1 História e evolução	79
4.2.2.2 Análise dos dados e interpretação dos resultados	80
4.2.2.3 Procedimento experimental.....	81
4.2.2.4 Coleta de dados	82
4.2.2.5 Adaptações do NASA-TLX.....	82
4.2.2.6 NASA-TLX relacionado com outros métodos	84
4.2.2.7 Análises estatísticas	85
4.2.2.8 Benefícios do NASA-TLX	85
4.2.2.9 Limitações do NASA-TLX.....	87
4.2.3 Escala Subjetiva	88
4.2.4 Questionário	91
4.2.5 Subjective Workload Assessment Technique (SWAT)	95
4.2.6 Tarefa secundária.....	100
4.2.7 Air Traffic Workload Input Technique (ATWIT).....	106
4.2.8 Rating Scale of Mental Effort (RSME).....	110
4.2.9 Surgical Task Load Index (SURG-TLX).....	114
4.2.10 Borg Scale of Perceived Exertion (BSPE).....	121
4.2.11 Job Demands Scale (JDS)	125
4.2.12 Escala Subjetiva de Carga de Trabalho Mental (ESCAM).....	131
4.2.13 Global Subjective Workload Scale (GSW).....	136
4.2.14 Simple Objective Index (SOI).....	137
4.2.15 Instantaneous Self-Assessment (ISA).....	139
4.2.16 ISTAS 21	143
4.2.17 Modified Cooper Harper Scale (MCH).....	153
4.2.18 Tarefa primária	159
4.2.19 Integrated Workload Scale (IWS)	163
4.2.20 Raw-NASA-TLX (RTLX).....	168
4.2.21 Surgical Task Load Index (SURG-TLX) + Global Operative Assessment of Laparoscopic Skills (GOALS).....	172
4.2.22 Traffic Load Index (TLI).....	182

4.3 ANÁLISE GERAL DOS MÉTODOS DE AVALIAÇÃO DA CTM	186
5 CONSIDERAÇÕES FINAIS	200
REFERÊNCIAS BIBLIOGRÁFICAS	202
APÊNDICE A - Dissertações de Revisão de Literatura.....	222
APÊNDICE B – Artigos do PB organizados conforme as ferramentas de análise da CTM	229
APÊNDICE C – Artigos do PB organizados conforme os indicadores fisiológicos utilizados para análise da CTM	231
APÊNDICE D – Artigos do PB organizados conforme os países de aplicação das pesquisas.....	232
APÊNDICE E – Artigos do PB organizados conforme os grupos estratificados de trabalhadores.....	233

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A ergonomia consiste no estudo das interações entre as pessoas e a tecnologia, a organização e o ambiente, buscando intervenções e projetos para melhorar a segurança, o conforto, o bem-estar e a eficácia das atividades humanas, de forma integrada (IIDA; GUIMARÃES, 2016).

O surgimento da Ergonomia data de 12 de julho de 1949, quando cientistas e pesquisadores se reuniram na Inglaterra para discutir e formalizar esse novo ramo de aplicação da ciência, mas o seu desenvolvimento ocorreu principalmente durante a II Guerra Mundial, por meio dos esforços conjuntos entre tecnologia, ciências humanas e biológicas, usando de médicos, psicólogos, antropólogos e engenheiros, para resolver problemas causados pela operação de equipamentos militares complexos (DUL; WEERDMEESTER, 2004; IIDA; GUIMARÃES, 2016).

A abordagem ergonômica deve basear-se no princípio de que o trabalho necessita ser adaptado ao homem e não o contrário, buscando maior humanização, criando um ambiente confortável e que encoraje o colaborador a desenvolver suas habilidades (SILVA; DE OLIVEIRA LUCAS, 2009). Segundo os autores, as condições de trabalho podem ser avaliadas por muitos fatores, como pela jornada e carga de trabalho, ambiente físico, material e equipamentos, ambiente saudável e estresse.

O desenvolvimento dos conceitos da ergonomia deu-se na direção da redução de esforços, doenças ocupacionais e acidentes, além de basear-se na capacidade de mobilizar conhecimentos e métodos adaptados a cada situação e/ou ação relacionada ao trabalho (DANIELLOU; BÉGUIN, 2007; IIDA; GUIMARÃES, 2016).

Laville (1997) divide a aplicação ergonômica em três etapas: (i) análise do quadro técnico, econômico, social e político da empresa, (ii) análise da demanda e (iii) análise ergonômica do problema. É na terceira etapa que se estabelece um diagnóstico especializado para cada demanda, por meio de métodos de estudos sistemáticos.

Desde a década de 70 são desenvolvidos roteiros para execução de análises ergonômicas, os quais passaram a ser modelos e serviram de bases para

outras propostas metodológicas, tornando-se os métodos, as ferramentas e as normas empregados na análise ergonômica e que utilizam recursos para averiguar, levantar, analisar e sistematizar o trabalho, por meio de instrumentos qualitativos e quantitativos (LIMA, 2004; MASCULO; VIDAL, 2013).

Como a ergonomia busca estudar o relacionamento entre o homem e o seu trabalho, ela contempla todas as variáveis intervenientes desse processo (DE SOUZA CARDOSO; GONTIJO, 2012).

Visto que as exigências físicas e mentais de trabalho são algumas das principais responsáveis por problemas no trabalhador, gerando doenças e, em casos extremos, invalidez, torna-se importante o estudo da carga de trabalho, seja ela física, mental ou psíquica, conforme classificação de Velázquez et al. (2015).

A carga física geralmente é descrita em termos de postura errada ou restrita, uso de força, movimentos repetitivos, falta de recuperação e combinação desses fatores (HANSSON et al., 2009) e a carga psíquica relaciona-se ao componente afetivo da tarefa (VELÁZQUEZ et al., 1995).

Já a carga de trabalho mental, que é foco desta pesquisa, pode ser caracterizada como um reflexo da tensão mental, resultado da realização de uma tarefa, em determinado ambiente e em condições operacionais específicas, juntamente com a capacidade do trabalhador em responder a essas demandas, tendo como objetivo de mensuração o custo mental de executar atividades para presumir o desempenho do operador e do sistema (CAIN, 2007). Além disso, a avaliação da carga de trabalho mental busca níveis mais altos de conforto, satisfação, eficiência e segurança no local de trabalho, sendo um ponto chave na pesquisa e desenvolvimento de interfaces homem-máquina (RUBIO; MARTÍN; PUENTE, 2004).

Diversos autores classificam em categorias as ferramentas utilizadas para avaliar a carga de trabalho mental: Lean e Shan (2012) dividem os métodos em subjetivos, por meio de questionários, e objetivos; para Byrne (2011) a avaliação deve incluir no mínimo uma medida da carga de trabalho primária, uma medida objetiva e uma medida subjetiva.

Moray (1988) utiliza uma classificação com três categorias: métodos ou medidas subjetivas, medidas comportamentais e medidas fisiológicas. De forma semelhante, Galy, Cariou e Mélan (2012) separam as categorias em medidas subjetivas, medidas de desempenho e medidas psicofisiológicas.

Já para Meshkati et al. (1995) existem quatro categorias, sendo elas medidas de tarefa primária, medidas de tarefa secundária, medidas de classificação subjetivas e medidas fisiológicas. Da mesma forma, Jorgensen et al. (1999), Sanders e McCormick (1993) e Schonblum (2004) utilizam esse mesmo formato de classificação.

Para as análises realizadas posteriormente neste trabalho, os métodos que analisam a carga mental serão divididos em três categorias, com base no trabalho de Moray (1988): (i) medidas subjetivas, nas quais se enquadram os questionários autoaplicáveis; (ii) medidas fisiológicas e (iii) medidas de desempenho, por meio da avaliação de tarefas primárias e/ou secundárias.

1.2 PROBLEMA DE PESQUISA

Existe uma grande variedade de métodos de avaliação da carga de trabalho mental, alguns mais conhecidos e mais empregados que outros. Nesse contexto, procuram-se os métodos mais utilizados na avaliação da CTM, com base em uma revisão sistemática de literatura, buscando aplicações em grupos estratificados de trabalhadores.

Sendo assim, o trabalho tem foco na prospecção das ferramentas de avaliação da carga de trabalho mental mais aplicadas atualmente e suas características.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Realizar uma análise crítica das ferramentas de avaliação da carga de trabalho mental, destacando suas particularidades, vantagens e desvantagens, atualizações e abrangência.

1.3.2 Objetivos Específicos

- Selecionar um portfólio bibliográfico com artigos científicos reconhecidos que abordam a aplicação de uma ou mais ferramentas para avaliação da carga de trabalho mental;
- Caracterizar as particularidades de cada método de avaliação da carga de trabalho mental;
- Organizar uma linha do tempo com os métodos de avaliação da carga de trabalho mental encontrados no portfólio bibliográfico.

1.4 JUSTIFICATIVA

A carga de trabalho pode ser dividida em física, cognitiva e psíquica, sendo que a carga física se relaciona ao esforço muscular, a carga cognitiva ao esforço mental e a carga psíquica ao componente afetivo da tarefa (VELÁZQUEZ et al., 1995).

Muitas vezes a atenção é somente voltada para a carga de trabalho física, sendo esquecidas as cargas cognitiva e psíquica. Porém, de acordo com o 1º Boletim Quadrimestral sobre Benefícios por Incapacidade de 2017, que trata de adoecimento mental e trabalho, os transtornos mentais e comportamentais foram a terceira causa de incapacidade para o trabalho, considerando a concessão de auxílio-doença e aposentadoria por invalidez, no período de 2012 a 2016, representando 9% das concessões desse período.

Nos dias de hoje, a organização do trabalho é marcada por um ambiente corporativo e pautada por uma lógica capitalista globalizada, em que o aumento de produtividade e o lucro máximo são os objetivos principais, contribuindo para o agravamento do adoecimento mental do trabalhador (MINISTÉRIO DA FAZENDA, 2017). Além disso, situações de banalização da violência, relações interpessoais norteadas por autoritarismo e competitividade e a desvalorização da subjetividade dos trabalhadores contribuem para esse cenário (BAHIA, 2014).

Dessa forma, é importante avaliar a carga mental do trabalhador, buscando melhorias na organização do trabalho, em função dos resultados apresentados.

Existem muitas ferramentas para a avaliação da carga de trabalho mental, com métodos diferentes de coleta de dados e com aplicações mais usuais em

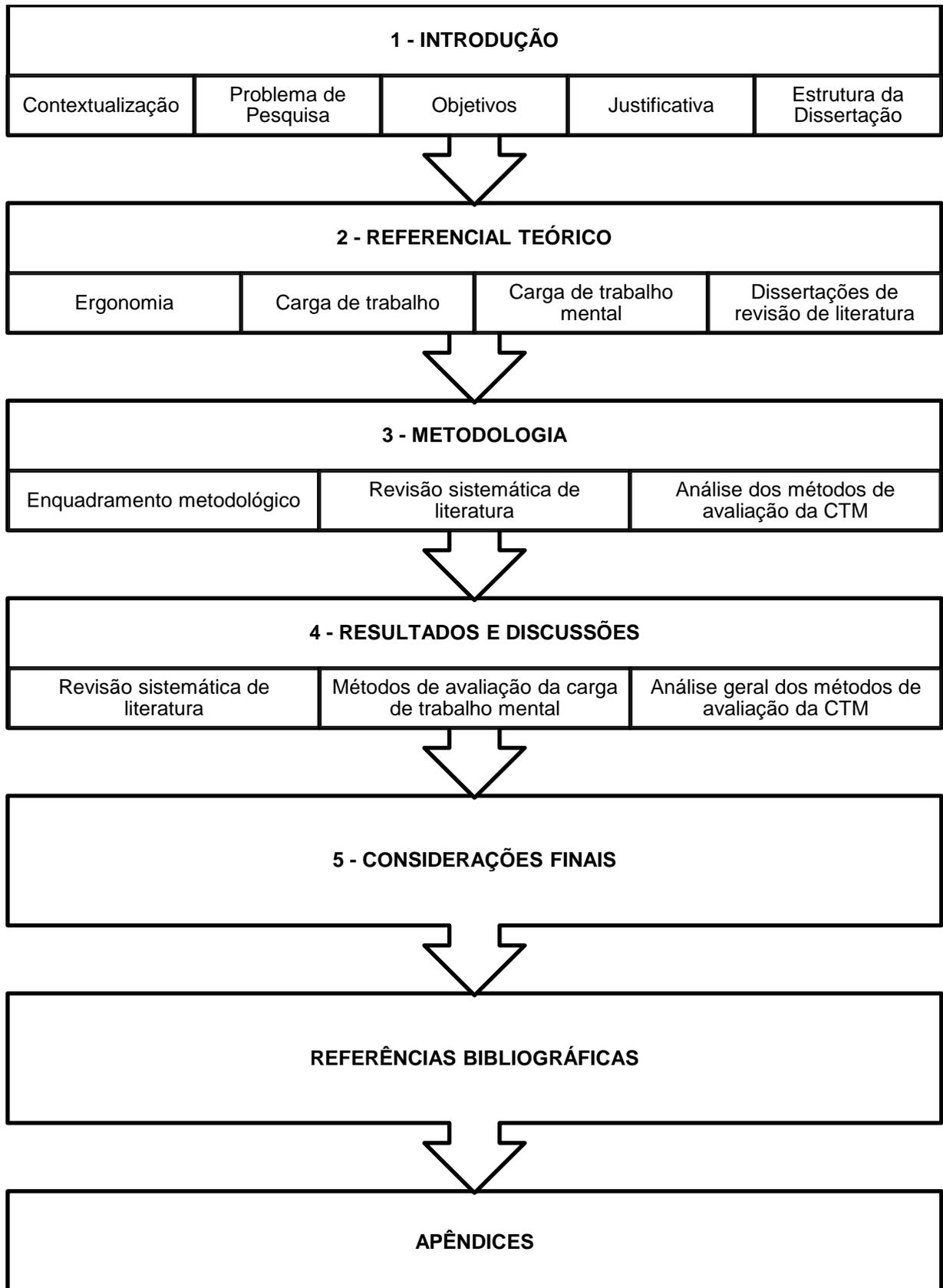
diferentes grupos estratificados de trabalhadores. Dessa forma, caracteriza-se como uma demanda importante reunir essas ferramentas em um único documento, descrevendo a metodologia de coleta de dados e a forma de análise dos resultados de cada ferramenta, além dos principais trabalhadores analisadas por cada método e suas vantagens e desvantagens.

Sendo assim, o meio acadêmico e profissional ganhará um material compilado com as ferramentas utilizadas para avaliar a carga de trabalho mental, com base em artigos científicos aplicados, juntamente com a descrição e a análise crítica de cada uma delas.

1.5 ESTRUTURA DA DISSERTAÇÃO

A presente dissertação está estruturada em sete capítulos: (i) Introdução, (ii) Referencial Teórico, (iii) Metodologia, (iv) Resultados e Discussões, (v) Considerações Finais, (vi) Referências Bibliográficas e (vii) Apêndices, conforme ilustra a Figura 1.

Figura 1 - Estrutura da Dissertação



Fonte: Elaborado pelo autor (2018).

2 REFERENCIAL TEÓRICO

2.1 ERGONOMIA

A definição mais antiga de ergonomia foi formulada em 1950, na Inglaterra, pela *Ergonomics Research Society*:

Ergonomia é o estudo do relacionamento entre o homem e seu trabalho, equipamento, ambiente e, particularmente, a aplicação dos conhecimentos de anatomia, fisiologia e psicologia na solução dos problemas que surgem desse relacionamento. (IIDA; GUIMARÃES, 2016, p. 4).

Utilizando a definição aprovada em 2000 pela Associação Internacional de Ergonomia, a Associação Brasileira de Ergonomia (ABERGO) adota a definição:

Ergonomia (ou Fatores Humanos) é a disciplina científica que estuda as interações entre os seres humanos e outros elementos do sistema de trabalho, aplicando os princípios teóricos, dados e métodos, a fim de realizar projetos para otimizar o bem-estar humano e o desempenho geral desse sistema. (IIDA; GUIMARÃES, 2016, p. 4).

Ainda segundo os mesmos autores, os objetivos da ergonomia abrangem a saúde, a segurança e a satisfação do trabalhador, tendo como consequência o aumento da eficiência.

A ergonomia no ambiente de trabalho pode envolver o ambiente físico e aspectos organizacionais, incluindo ferramentas e materiais. Assim, ela analisa a atividade real, descobrindo pontos críticos e inadequações, a fim de propor mudanças para melhorar as condições de trabalho (SILVA; DE OLIVEIRA LUCAS, 2009).

Apesar das condições ergonômicas estarem inadequadas ao trabalho, Martins (2015) acredita que é possível prevenir lesões e doenças, desde que o local e a organização do trabalho estejam ajustados às necessidades individuais, físicas e mentais, de cada trabalhador. Ainda, podem ser feitas atividades para contribuir na melhoria da saúde, do bem-estar e da qualidade de vida dos trabalhadores, como acompanhamentos médicos, por meio de exames de saúde, avaliação da capacidade

funcional dos trabalhadores e das exigências do trabalho, treinamentos e manutenção do ambiente de trabalho (SILVA; DE OLIVEIRA LUCAS, 2009).

Embora deve-se buscar a adaptação do trabalho ao homem, para lida e Guimarães (2016) isso nem sempre é resolvido de forma fácil, principalmente quando a ergonomia não é aplicada desde as etapas iniciais dos projetos, não incluindo o ser humano como um dos componentes.

Dessa forma, muitas vezes é necessário fazer o possível dentro das restrições existentes, mesmo não sendo o ideal. Entretanto, apesar dessa dificuldade, existem estudos da aplicação da ergonomia na indústria, na agricultura, na mineração, na construção civil, no setor de serviços e até mesmo na vida diária (IIDA; GUIMARÃES, 2016).

Além disso, vários trabalhos trazem a análise ergonômica do trabalho aplicada em empresas, visto que uma característica da ação ergonômica é que além dela se basear na capacidade de mobilizar conhecimentos e métodos adaptados a cada situação, ela visa uma ação, ou seja, não se contenta apenas com um conhecimento sobre as situações do trabalho (DANIELLOU; BÉGUIN, 2007).

Para Wisner (1987) podem-se distinguir dois campos principais na ergonomia: ergonomia do produto e ergonomia da produção. Na ergonomia do produto o ergonomista colabora com o setor comercial, nos estudos de mercado, com o setor de fabricação, para avaliar os custos de fabricação e a finalidade, e com outros especialistas da concepção do produto, como os desenhistas industriais e especialistas no controle da qualidade. Já a ergonomia da produção relaciona-se à crescente preocupação com as condições de trabalho, em função da maior ou menor divisão do trabalho.

2.2 CARGA DE TRABALHO

O conceito de carga de trabalho surgiu nos estudos de Leplat e Cuny (1983), sobre psicologia no trabalho, e posteriormente foi retomado e desenvolvido por Alain Wisner, um dos principais autores da área da ergonomia (CORRÊA, 2003).

Para Leplat e Cuny (1983, p. 180) “a noção de carga de trabalho será, pois, sempre relativa à interação entre um sujeito e as exigências de determinado meio”.

Wisner (1987, 1994) acredita que a carga de trabalho sempre estará presente durante o processo de trabalho e relaciona-a com condicionantes internos e externos ao trabalhador, determinando as exigências físicas, cognitivas e psíquicas do trabalho, as quais, para o autor, estão inter-relacionadas, embora uma forte sobrecarga de um desses aspectos não é necessariamente acompanhada de uma sobrecarga nos demais domínios.

Laurell e Noriega acreditam que:

O conceito de carga de trabalho possibilita uma análise do processo de trabalho de que extrai e sintetiza os elementos que determinam de modo importante o nexos biopsíquico da coletividade operária e conferem a esta um modo específico de andar a vida. (LAURELL; NORIEGA, 1989, p. 110).

Para Facchini, carga de trabalho é:

Um atributo de um determinado processo de trabalho cuja presença no ambiente de trabalho pode aumentar a probabilidade de que um grupo de trabalhadores expostos experimente uma deterioração psicológica comparada com aqueles que não estiveram expostos a tal atributo. (FACCHINI, 1994, p. 180).

Seligmann-Silva (1994) diferencia a carga de trabalho em esforços físicos, cognitivos e psicoafetivos (ou emocionais). Para ele, a carga de trabalho representa o conjunto de esforços para atender as exigências das tarefas e ele acredita que os três tipos de carga estão intimamente relacionados, ou seja, uma sobrecarga em um dos tipos, pode refletir em mudanças nos outros tipos.

Moura (1998, p. 80) afirma que a carga de trabalho é a “relação entre constrangimentos impostos pela tarefa, pela interface, pelos instrumentos e pelo ambiente (carga funcional), em conjugação com as atividades desempenhadas e a capacidade de trabalho do operador”.

Greco, Queiros e Gomes, acreditam que:

As cargas de trabalho são definidas como exigências ou demandas psicobiológicas do processo de trabalho, gerando ao longo do tempo as particularidades do desgaste do trabalhador. Em outras palavras, as cargas

são mediações entre o processo de trabalho e o desgaste psicobiológico. (GRECO; QUEIROS; GOMES, 1999, p. 61).

Sell fala sobre a origem das cargas de trabalho:

Da relação entre os elementos de um sistema de trabalho, decorre que, durante a execução da tarefa, a pessoa está sujeita a diversas influências malélicas e/ou benéficas, cujas causas estão na tarefa de trabalho e no ambiente de trabalho. Estas influências malélicas recebem o nome de cargas. (SELL, 2003, p. 64).

Para Corrêa:

As cargas de trabalho estão relacionadas entre as categorias definidas para estudar e avaliar o impacto produzido pelos elementos constitutivos do processo de trabalho – tanto do objeto e da tecnologia, como de sua organização e divisão – sobre a saúde física e mental dos trabalhadores. (CORRÊA, 2003, p. 18).

O mesmo autor divide a carga de trabalho em física e psíquica:

As cargas físicas podem estar presentes tanto no ambiente de trabalho como fora dele e são decorrentes principalmente das exigências técnicas para a transformação do objeto de trabalho. São exemplos de cargas físicas o ruído, temperatura e a iluminação. As cargas físicas são ainda diferenciadas em cargas químicas, biológicas, mecânicas e fisiológicas. (CORRÊA, 2003, p. 19).

As cargas químicas são decorrentes dos objetos de trabalho e dos meios e instrumentos necessários para a sua transformação. As cargas biológicas estão relacionadas à condição de higiene ambiental, sendo causadas por qualquer organismo animal ou vegetal. As cargas mecânicas são decorrentes da tecnologia e das condições de instalação e manutenção da produção. Já as cargas fisiológicas estão relacionadas com a utilização do corpo no trabalho, por meio de correção na postura, repetição de movimentos ou esforço físico (CORRÊA, 2003).

Frutuoso e Cruz definem a carga de trabalho como:

[...] uma construção teórica resultante da necessidade de compreender que, para uma determinada situação de trabalho, há uma tensão permanente entre as exigências do processo e as capacidades biológicas e psicológicas dos trabalhadores para respondê-las. (FRUTUOSO; CRUZ, 2005, p. 29).

Ainda segundo eles, as cargas de trabalho são interferidas pela dinâmica do trabalho, e o equilíbrio é resultado das compatibilidades entre as exigências implícitas e explícitas da tarefa, juntamente com a capacidade da pessoa para realizá-la.

De Rivecourt et al. (2008) acreditam que a carga mental é frequentemente usada para indicar o grau em que as demandas das tarefas afetam a capacidade de processamento de informações pelo trabalhador.

Para Oliveira Secco et al.:

Cargas de trabalho são definidas como os elementos do processo de trabalho que interagem entre si e com o corpo do trabalhador, desencadeando alterações nos processos biopsíquicos que se manifestam como desgastes físicos e psíquicos potenciais ou efetivamente apresentados. (OLIVEIRA SECCO et al., 2010, p. 3).

De forma diferente, Kirchof et al. (2011), dividem as cargas em físicas, químicas, biológicas, mecânicas, fisiológicas e psíquicas. As fisiológicas e psíquicas acontecem internamente no corpo do trabalhador, enquanto as outras externamente. Assim, as cargas fisiológicas e psíquicas manifestam-se somente por meio de um distúrbio ou de uma doença e as demais podem ser observadas no ambiente de trabalho.

Já Waard e Lewis-Evans (2014) acreditam que a carga de trabalho é a interação entre dois fatores: as propriedades físicas e mentais da tarefa, ou tarefa exigida, e a capacidade do operador de executá-la.

Assim, ressalta-se que a partir desses conceitos observam-se duas interpretações principais do fenômeno: uma corrente mais psicológica, que vê a carga como uma interação de fatores diversos e uma corrente que vê a carga como uma espécie de relação de causa e efeito (CORRÊA, 2003).

De forma geral, a maioria dos autores acima citados consideram a carga de trabalho decorrente do processo de trabalho, incluindo ambiente, processo,

tecnologia, organização, divisão, tarefa e exigências (CORRÊA, 2003; FACCHINI, 1994; FRUTUOSO; CRUZ, 2005; GRECO; QUEIROS; GOMES, 1999; LAURELL; NORIEGA, 1989; LEPLAT; CUNY, 1983; MOURA, 1998; OLIVEIRA SECCO et al., 2010; SELL, 2002; WAARD; LEWIS-EVANS, 2014; WISNER, 1994). Além disso, para alguns autores a carga de trabalho também é relacionada com as características individuais do trabalhador e sua capacidade para realizar o trabalho (FRUTUOSO; CRUZ, 2005; MOURA, 1998; WAARD; LEWIS-EVANS, 2014).

2.3 CARGA DE TRABALHO MENTAL

Estudos realizados na Suécia, a partir dos anos 70, consideraram a carga de trabalho mental contendo aspectos qualitativos e quantitativos, com a carga quantitativa dizendo respeito ao volume de trabalho mental exigido e a carga qualitativa referente ao nível de complexidade do trabalho (ESPÍNDOLA, 2013). Assim, uma pessoa pode estar ao mesmo tempo realizando um trabalho que apresente aspectos qualitativos e quantitativos de carga mental.

Othman e Romli (2016) consideram a carga de trabalho mental uma interação entre as demandas das tarefas e a capacidade de atenção do operador. Assim, quanto mais complexa a tarefa, maior a demanda da capacidade cognitiva da pessoa para conseguir executá-la com sucesso.

Já Ghanbary Sartang et al. (2016) consideram a carga de trabalho mental um conjunto de fatores que afetam o processamento mental de informações, com tomadas de decisões e reações individuais no local de trabalho.

Os primeiros estudos relacionados à carga de trabalho mental são os de Wierwille e Casali (1983, 1984), Wierwille e Connor (1983) e Wierwille et al. (1985) (MORAY, 1988). As variáveis analisadas nesses estudos estão apresentadas no Quadro 1, onde foram divididas em medidas de opinião, de tarefa secundária, fisiológicas, comportamento dos olhos e tarefa primária. Vale ressaltar que as variáveis em *itálico* foram sensíveis a mudanças impostas na carga de trabalho.

Quadro 1 – Variáveis analisadas nos primeiros estudos sobre CTM

Método/ Artigo	Wierwille e Casali (1983)	Wierwille e Connor (1983)	Wierwille e Casali (1984)	Wierwille et al. (1985)
Opinião	<i>Cooper Harper Modificada Multidescrição</i>	<i>Cooper Harper WCI/TE</i>	<i>Cooper Harper Modificada Multidescrição WCI/TE</i>	<i>Cooper Harper Modificada Multidescrição WCI/TE</i>
Tarefa secundária	<i>D. Padrão estimativa tempo Tocando com regularidade</i>	Sombra de dígitos Escanemamento de memória Aritmética mental Estimativa de tempo média D. Padrão estimativa tempo Erro absoluto estimativa tempo Erro Rms estimativa de tempo	<i>D. Padrão estimativa tempo Tocando com regularidade</i>	<i>Estimativa de tempo Obtenção de regularidade</i>
Fisiológicas	Taxa de respiração Frequência cardíaca média D. Padrão Freq. Cardíaca <i>Diâmetro da pupila</i>	Taxa de pulso média Variabilidade da taxa de pulso Taxa de respiração Diâmetro da pupila Padrão Voz - tarefa digital Padrão Voz - aritmética mental	<i>Taxa de respiração Diâmetro da pupila</i>	Taxa de respiração Frequência cardíaca média D. Padrão Freq. Cardíaca Diâmetro da pupila
Comportamento dos olhos	Movimento de piscar Fixação dos olhos	Frequência mudança do olhar Frequência de piscar de olhos	Movimento de piscar	<i>Movimento de piscar Fração de Fixação</i>
Tarefa primária	Movimentos de controle PHPMS RHPMS <i>Erros de omissão Erros de comissão Tempo resposta comunicação</i>	Localização do erro RSM Inclinação do erro RMS <i>Controle de Movimentos RMS</i>	Movimentos de controle PHPMS RHPMS <i>Tempo resposta ao perigo</i>	Movimentos de controle PHPMS RHPMS <i>Taxa de erro mediático Tempo de reação mediático</i>

Fonte: Elaborado pelo autor (2018).

As medidas de opinião envolviam a opinião do trabalhador sobre as cargas de trabalho, expressas pelo preenchimento de formulários. As medidas fisiológicas envolviam as respostas fisiológicas do trabalhador decorrentes das mudanças na carga de trabalho, envolvendo principalmente dados da respiração, do coração e da voz e o comportamento dos olhos analisava a resposta dos olhos quando variada a carga de trabalho. Já a tarefa primária relacionava-se com o desempenho da tarefa em si e a tarefa secundária utilizava duas tarefas para medir a o desempenho do trabalhador, analisando dados da segunda tarefa para a carga de trabalho (ESPÍNDOLA, 2013).

No trabalho de Wierwille e Casali (1983) apenas as variáveis envolvendo o comportamento do olhar não mostraram significância estatística quando existiram mudanças na carga de trabalho, e os destaques foram as medidas de opinião e de tarefa primária. Da mesma forma, o estudo dos mesmos autores, de 1984, também não possuiu nenhuma medida de comportamento dos olhos com sensibilidade às mudanças na carga de trabalho, tendo como destaque as medidas de opinião e de tarefa secundária.

No artigo de Wierwille e Connor (1983) apenas três medidas apresentaram sensibilidade quando mudanças na carga de trabalho foram impostas: ambas as medidas de opinião e uma das medidas de tarefa primária.

Já no trabalho de Wierwille et al. (1985) nenhuma medida fisiológica foi sensível às mudanças na carga de trabalho, tendo as medidas de opinião o maior destaque.

Porém, com esses estudos, não foi obtida uma coerência para predições teóricas da carga de trabalho mental, não conseguindo generalizações empíricas (CORRÊA, 2003). Dessa forma, os pesquisadores continuaram buscando medidas que pudessem ser generalizadas, ou seja, elas deveriam ser sensíveis a mudanças na carga de trabalho dentro de uma mesma população (CORRÊA, 2003).

Com um estudo de 1999, Hart apresentou uma relação dos comportamentos dos pilotos submetidos aos diversos níveis de CTM, conforme apresentado no Quadro 2.

Quadro 2 - Comparação entre o nível de carga e o tipo de ação do piloto

Nível de carga	Tipo de ação do operador
Subcarga	Quando as demandas de tarefa são muito baixas, os pilotos podem executar atividades que em nada se relacionem com a missão (por exemplo, conversas casuais). Apesar destas atividades poderem ajudar a manter a vigilância, não mantêm o piloto concentrado unicamente em seu objetivo. Portanto, quando ocorrerem os eventos relativos à tarefa, os pilotos adotam uma estratégia de reação.
Carga de trabalho moderada	Durante os períodos de baixa demanda de tarefas, os pilotos habilidosos executam tarefas antecipadas ao cronograma (para evitar períodos futuros de alta carga), monitoram o ambiente para antecipar problemas em potencial e desenvolvem planos de contingência. Esses comportamentos podem elevar o nível atual da carga de trabalho a fim de reduzir a carga de trabalho posterior no voo.
Alta carga de trabalho	Em muitas situações constrangedoras, quando os eventos são inesperados, ou quando as demandas de tarefas são relativamente altas, os operadores tendem a adotar uma estratégia de reação. Isto é, eles respondem a cada demanda de tarefa assim que ela ocorre. Em voo, essa estratégia é um sintoma de que o piloto está começando a perder terreno para a aeronave.
Sobrecarga	A medida em que as demandas de tarefas aumentam, os pilotos adiarão as tarefas menos críticas (até um período de carga de trabalho menor), transferirão as tarefas para um membro da tripulação mais livre, ou simplesmente decidirão não executar uma tarefa. Tais estratégias são necessárias em situação de sobrecarga, pois permitem ao piloto concentrar sua atenção em requisitos críticos. Todavia a qualidade do desempenho geral depende da habilidade do piloto em selecionar quais tarefas são prioritárias.

Fonte: Hart (1999).

Outro estudo, de Xie e Salvendy (2000) buscou desenvolver e validar um método prático para predizer a carga de trabalho mental, em ambientes com tarefas múltiplas e singulares (CORRÊA, 2003). Para isso, eles desenvolveram um modelo que contemplava medidas de carga instantânea, média, acumulada e máxima, como variáveis independentes, além de carga geral, como variável dependente. Eles ainda acrescentaram duas variáveis, carga efetiva e carga não efetiva, para modelar os fatores individuais. Aproximação multidimensional foi utilizada no modelo para quantificar diferentes aspectos de carga, o qual foi validado por meio de dois estudos experimentais (XIE; SALVENDY, 2000).

Os resultados revelaram que a estrutura, que consistiu em carga de trabalho média, carga de trabalho acumulada e carga de trabalho instantânea, poderia descrever a carga de trabalho com mais precisão do que uma carga de trabalho geral única. Além disso, as relações entre a carga de trabalho efetiva/não efetiva e as quatro variáveis independentes foram parcialmente validadas. Os resultados mostraram que tanto os fatores relacionados à tarefa quanto os fatores relacionados a indivíduos

poderiam significativamente afetar a carga de trabalho mental (XIE; SALVENDY, 2000).

Esses estudos continuam até os dias de hoje e deram origem a todas as ferramentas existentes para analisar a carga de trabalho mental. Além disso, novos estudos buscam aprimorar essas ferramentas e até mesmo criar novos métodos, objetivando melhorar a confiabilidade das medidas.

Quanto ao conceito da carga mental, Corrêa (2003) propõe uma divisão, contendo carga psíquica, carga cognitiva ou informacional e carga mental, conforme Quadro 3.

Quadro 3 - Síntese das características dos conceitos de carga mental

Conceito	Características
Carga Psíquica	Refere-se às cargas relacionadas ao fator afetivo no trabalho ou a significação do trabalho para quem o faz.
Carga Cognitiva ou Informacional	Refere-se às cargas advindas das exigências cognitivas das tarefas. O uso da memória, as decisões, os raciocínios, as regras relacionadas à tarefa.
Carga Mental	Reúne aspectos psíquicos e cognitivos integrantes dos dois outros conceitos.

Fonte: Corrêa (2003).

Assim, para Corrêa (2003), a carga mental engloba ambos os aspectos, psíquico e cognitivo, ou seja, refere-se às cargas relacionadas ao fator afetivo, ou a significância do trabalho, e também às cargas advindas das exigências cognitivas.

Para O'Donnell e Eggemeier (1986) a carga mental pode ser associada a três fatores principais: pressão temporal da tarefa, envolvendo o tempo disponível e o tempo necessário, quantidade e tipos de recursos exigidos pela tarefa e aspectos emocionais, como fadiga, estresse e ansiedade.

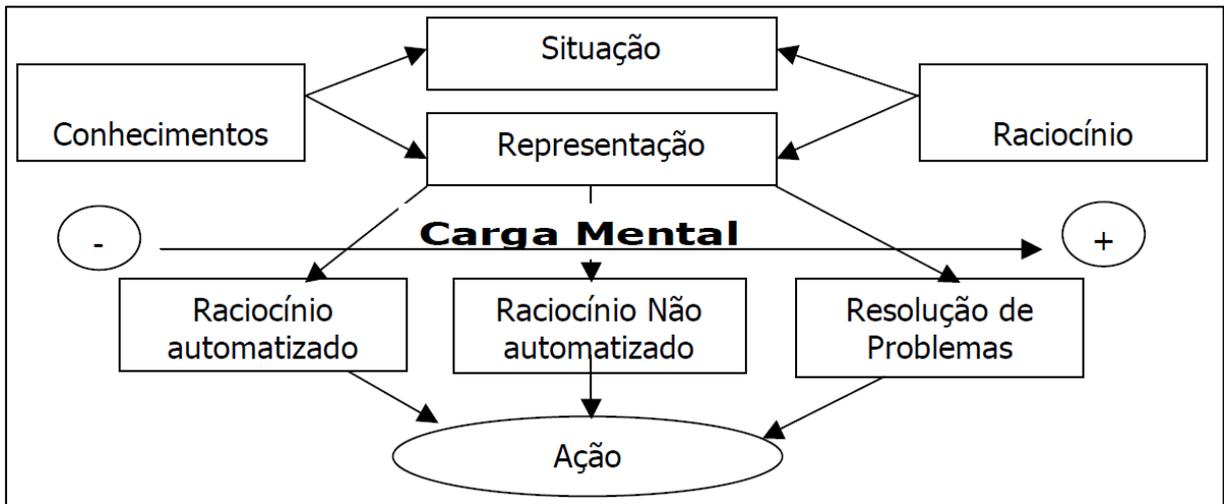
Montmollin (1995) considera a carga mental uma quantidade contínua e homogênea, sendo importante medir a evolução, visando determinar um limite de sobrecarga.

Já para Ayaz et al. (2012) uma característica fundamental do conceito de carga de trabalho mental, a qual reflete o quanto o cérebro está trabalhando para atender às demandas das tarefas, é que ela pode ser dissociada do desempenho.

Para Corrêa (2003) o ramo da ergonomia chamado Ergonomia Cognitiva, o qual estuda aspectos cognitivos da tarefa, desenvolveu-se muito devido a crescente

informatização do trabalho e distanciamento cada vez maior dos trabalhadores em relação aos seus objetos de trabalho. Com base no trabalho de Richard (1990), Corrêa (2003) discutiu as possíveis relações entre o conceito de carga mental, psíquica ou cognitiva, e a Ergonomia Cognitiva.

Figura 2 - Arquitetura Cognitiva de Richard associada ao Conceito de CM



Fonte: Corrêa (2003).

Conforme apresentado na Figura 2, Corrêa acrescentou a linha de carga mental à Ergonomia Cognitiva de Richard e considerou que a resolução de problemas exige mais carga mental que o raciocínio automatizado. O mesmo autor citou o exemplo do serviço de teleatendimento, com a hipótese que serviços exclusivos de auxílio à lista que apenas prestam informações a clientes, que são serviços que utilizam raciocínios automatizados, possuem um valor bem menor de carga de trabalho mental quando comparados a serviços de atendimentos à reclamação na mesma central de atendimento, os quais envolvem a resolução de problemas.

Essa comprovação pode ser feita por meio das ferramentas de estimação da carga de trabalho mental, as quais são o assunto do item 2.3.1.

2.3.1 Métodos de mensuração da carga de trabalho mental

Em sua dissertação, Espíndola (2013) reuniu as categorias e definições dos métodos de avaliação da carga de trabalho mental de Jorgensen et al. (1999) e Sanders e McCormick (1993), as quais estão disponíveis no Quadro 4.

Quadro 4 - Categorias e definições dos métodos de avaliação da CTM

Categoria dos métodos	Definição
Medida das tarefas primárias	Relaciona-se diretamente com o desempenho da tarefa em si.
Medidas das tarefas múltiplas	Mensura o nível da carga por meio do uso de duas tarefas, uma tarefa primária, mais sofisticada, e uma segunda tarefa menos sofisticada e com o nível de carga já conhecido.
Medidas fisiológicas	Medem as respostas fisiológicas relacionadas com as respostas às mudanças nos níveis das cargas mentais.
Medidas subjetivas	Buscam as respostas subjetivas para as experiências relacionadas com a carga mental de trabalho, frequentemente administrada por meio de questionários aplicados ao final da realização da tarefa.

Fonte: Espíndola (2013).

Utilizando essa mesma classificação, Schonblum (2004), fez uma síntese das vantagens e desvantagens dessas categorias, a qual está disponível no Quadro 5.

Quadro 5 - Vantagens e desvantagens das categorias dos métodos da CTM

Categoria dos métodos	Vantagens	Desvantagens
Medida das tarefas primárias	Prevê uma indicação de direção do sistema total ou desempenho de subsistema enquanto é respondido pelo trabalhador.	Medidas de tarefas primárias necessariamente não provêm a capacidade residual do trabalhador.
Medidas das tarefas múltiplas	Prevê uma indicação dos recursos residuais do trabalhador.	Possível impacto artificial ou intrusão no desempenho de tarefa primária, e frequentemente encontram baixa aceitação do trabalhador.
Medidas fisiológicas	Não requer respostas públicas, registra continuamente ao longo da tarefa; de natureza multidimensional.	Custo alto, requer perícia para coleta e interpretação de dados.
Medidas subjetivas	Alta validade porque eles refletem as experiências subjetivas do operador; baixo custo e facilidade de implementação, não necessita de equipamento adicional e/ou exigência de programa de computação extensos, não intrusiva, nível alto de aceitação do trabalhador.	Nível alto de variabilidade de intersubjetividade. Pode dissociar (relatórios resultados contrastantes) com medidas de desempenho de carga de trabalho.

Fonte: Schonblum (2004).

É importante destacar que ambas as categorias “medida das tarefas primárias” e “medidas das tarefas múltiplas” são referentes a medidas de performance ou rendimento, as quais muitos autores colocam na mesma categoria.

Da mesma forma que Jorgensen et al. (1999), Sanders e McCormick (1993) e Schonblum (2004), Meshkati et al. (1990) agrupam os métodos em quatro categorias: medidas de tarefa primária, medidas de tarefa secundária, medidas subjetivas e medidas fisiológicas.

Utilizando uma classificação diferente, Lean e Shan (2012) dividem os métodos em subjetivos, utilizando questionários, e objetivos e para Byrne (2011) a avaliação deve incluir no mínimo uma medida de carga de trabalho primária, uma medida objetiva e uma medida subjetiva.

Já Moray (1988) utiliza a classificação em três categorias: métodos ou medidas subjetivas, medidas comportamentais e medidas fisiológicas.

Os métodos subjetivos baseiam-se nas impressões individuais, e subjetivas, dos indivíduos, normalmente por meio da aplicação de testes psicológicos. As medidas comportamentais geralmente são relacionadas ao método da tarefa secundária, a qual durante a execução de uma determinada tarefa, apresenta-se ao operador. A partir da realização dessa tarefa secundária faz-se uma mensuração da carga mental da tarefa primária, normalmente contabilizando-se erros, tempo de execução e dificuldades apresentadas na tarefa secundária (CORRÊA, 2003). Já a estimativa por meio de métodos fisiológicos é analisada mediante a alteração de parâmetros fisiológicos do trabalhador.

De forma semelhante a Moray (1988), Galy, Cariou e Mélan (2012) separam as categorias em medidas subjetivas, medidas de desempenho e medidas psicofisiológicas.

Com base na classificação de Moray (1988) e Galy, Cariou e Mélan (2012), a classificação adotada nesse trabalho será a seguinte: medidas subjetivas, medidas fisiológicas e medidas de desempenho.

2.4 TRABALHOS DE REVISÃO DE LITERATURA

O meio científico possui uma gama de informações, as quais, em sua maioria, estão contidas em artigos científicos, monografias, dissertações, teses, livros,

anais de congressos, entre outros, os quais podem ser acessados por qualquer pessoa e em qualquer lugar do mundo (PESSOA, 2016).

A revisão de literatura é uma forma de analisar essas informações, a qual é o foco deste trabalho de dissertação. Dessa forma, em busca das metodologias utilizadas por autores, foram buscadas dissertações de revisão de literatura disponíveis eletronicamente.

Essa busca foi feita no catálogo de teses e dissertações da CAPES e no portal Domínio Público - Biblioteca digital desenvolvida em *software* livre, utilizando o período de busca entre 2009 e 2017 e a palavra-chave “revisão de literatura”. Essa busca ainda foi restrita as áreas de avaliação: Engenharias III, Interdisciplinar, Saúde Coletiva e Serviço Social, visto que o programa de pós-graduação da presente dissertação está classificado dentro das Engenharias III e como a dissertação trata da saúde mental do trabalhador pesquisas nessa área podem ser classificadas dentro das áreas Interdisciplinar, Saúde Coletiva e Serviço Social.

A seleção final contou com 13 dissertações, as quais estão contidas no Apêndice A, com o nome do seu autor e ano de defesa, o objetivo do trabalho, a metodologia empregada, as bases de dados pesquisadas, o período de publicações analisado, a forma de coleta dos dados, o tamanho da amostra e o que foi analisado.

Todas as 13 dissertações contaram com uma revisão sistemática de literatura, as quais possuíam um referencial teórico sobre o tema da pesquisa seguido de uma metodologia de busca de artigos e posteriores análises, conforme escolhas de cada autor. Apesar disso, Tinoco (2015), diferente dos demais autores, não apresentou resultados quantitativos do seu portfólio bibliográfico, por meio de uma análise bibliométrica, apenas abordou os artigos selecionados dentro de categorias definidas pelo autor.

Todas os 13 autores fizeram a busca de artigos com base em descritores ou palavras-chave, seguidos de um processo de filtragem, até chegar ao portfólio final, o qual variou de 6 a 935 publicações analisadas. O período de análise variou de 2 anos até um período ilimitado, com uma dissertação sem especificar esse intervalo. Além disso, 12 trabalhos fizeram a sua busca em bases eletrônicas de dados e 1 dos trabalhos utilizou periódicos específicos. Em relação ao objetivo do trabalho, 12 dissertações deixaram claro no objetivo geral que seria realizada uma revisão de literatura.

Apesar das análises mostrarem grande variação de conteúdo, visto que os temas abordados eram diferentes uns dos outros, pode-se dizer que de forma geral os autores analisaram o ano e local/país da realização e/ou publicação da pesquisa, periódico e fator de impacto, amostra, metodologia/ferramentas/técnicas utilizadas nos artigos para coleta de dados para posteriores análises, vantagens e desvantagens das diferentes metodologias utilizadas, principais resultados dos estudos, conclusões levantadas em cada artigo, conclusões gerais dos estudos e trechos importantes.

A metodologia empregada nesses trabalhos e as análises feitas serviram de base para definir a revisão sistemática do estudo atual, apresentada nos próximos capítulos.

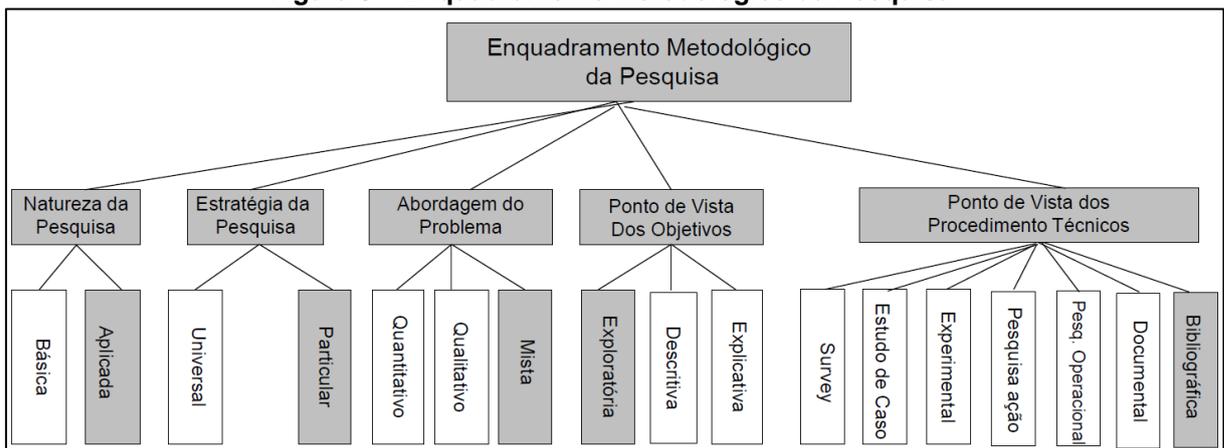
3 METODOLOGIA

A pesquisa foi dividida em duas partes: a primeira abordou uma revisão sistemática de literatura, buscando os métodos de avaliação da carga de trabalho mental, e a segunda consistiu em uma revisão de literatura de cada um dos métodos anteriormente encontrados. Essas etapas estão descritas nas seções 3.2 e 3.3. Antes, na seção 3.1, será apresentado o enquadramento metodológico da pesquisa.

3.1 ENQUADRAMENTO METODOLÓGICO

O enquadramento metodológico da pesquisa encontra-se na Figura 3.

Figura 3 – Enquadramento Metodológico da Pesquisa



Fonte: Pessoa (2016).

A natureza desta pesquisa é aplicada, pois possui interesse na aplicação, utilização e consequências práticas do conhecimento (GIL, 2008). A estratégia de pesquisa é particular, pois os resultados dependem do contexto estudado, não gerando necessariamente a repetitividade caso seja replicado o processo de pesquisa, e também as informações obtidas no estudo partem de artigos científicos coletados a partir de características previamente definidas em um contexto e período específicos (GIFFHORN, 2007; PESSOA, 2016). O problema foi abordado de forma mista, utilizando elementos quantitativos e qualitativos (MIGUEL et al., 2012). Levando em consideração os objetivos, a pesquisa é classificada como exploratória, pois pretende proporcionar visão geral acerca de determinado fato e do ponto de vista dos

procedimentos técnicos, o trabalho condiz com a pesquisa bibliográfica, desenvolvida a partir de material já elaborado (GIL, 2008).

3.2 REVISÃO SISTEMÁTICA DE LITERATURA

A metodologia empregada na primeira parte da pesquisa foi a da revisão sistemática de literatura, a qual é considerada um estudo secundário, tendo como fonte de dados os estudos primários, os artigos científicos usados como portfólio para revisão, os quais são selecionados por meio de um método pré-planejado, com procedimentos de busca, seleção e análise bem delineados e definidos, possibilitando um julgamento mais objetivo dos resultados (GALVÃO; PEREIRA, 2014; MANCINI; SAMPAIO, 2007; SCHÜTZ; SANT'ANA, 2011).

Para isso, o primeiro passo foi a definição da pergunta de pesquisa: Quais as ferramentas utilizadas para avaliação da carga de trabalho mental?

As palavras-chave escolhidas para compor a seleção foram: carga cognitiva, carga mental, ergonomia cognitiva e ergonomia mental. Nenhuma palavra-chave referente a metodologia, método ou ferramenta foi utilizada, visto que isso levaria a um número menor de artigos encontrados, sendo escolhido fazer o descarte de artigos, que não envolviam métodos ou ferramentas, de forma manual, durante as etapas de seleção do portfólio bibliográfico.

Foi realizada uma busca em bases eletrônicas de dados científicos, com o tema carga mental, em três bases de dados: *ScienceDirect*, *Scopus* e *Web Of Science*. Além disso, o período de análise foi delimitado entre os anos de 2000 e 2017.

A base *ScienceDirect* é a principal base de dados da *Elsevier* e combina autoridade, textos científicos completos e publicações técnicas e de saúde com funcionalidades inteligentes e intuitivas. A base de dados *Web Of Science* é mundialmente conhecida e dá origem ao *Journal Citation Report* (JCR), o qual lista o fator de impacto dos periódicos científicos (LACERDA; ENSSLIN; ENSSLIN, 2014). Já a base *Scopus* é o maior banco de dados científicos do mundo, se considerar o período entre 2000 e 2011 (GUERRERO-BOTE; MOYA-ANEGÓN, 2012). Além disso, as três bases de dados trazem artigos que abordam assuntos de engenharia e de saúde, conforme classificação do programa de pós-graduação e o tema da pesquisa.

As palavras-chave foram traduzidas para o inglês e assim a busca foi realizada com a seguinte combinação: "*cognitive workload*" OR "*mental workload*" OR "*cognitive ergonomics*" OR "*mental ergonomics*", todas fazendo parte do eixo único "carga mental". Além disso, as buscas limitaram-se aos campos título, resumo e palavras-chave das publicações e foram filtrados apenas artigos publicados em revistas, periódicos e anais de congressos.

Com os critérios acima estabelecidos foram encontradas 361 publicações na base de dados *ScienceDirect*, 1.983 resultados na base *Web Of Science* e 1.256 publicações na base de dados *Scopus*. Para o gerenciamento dessas publicações, foi utilizado o administrador bibliográfico gratuito *Mendeley Desktop*, o qual encontrou 1.151 arquivos duplicados, restando 2.449 documentos para serem analisados nas etapas posteriores.

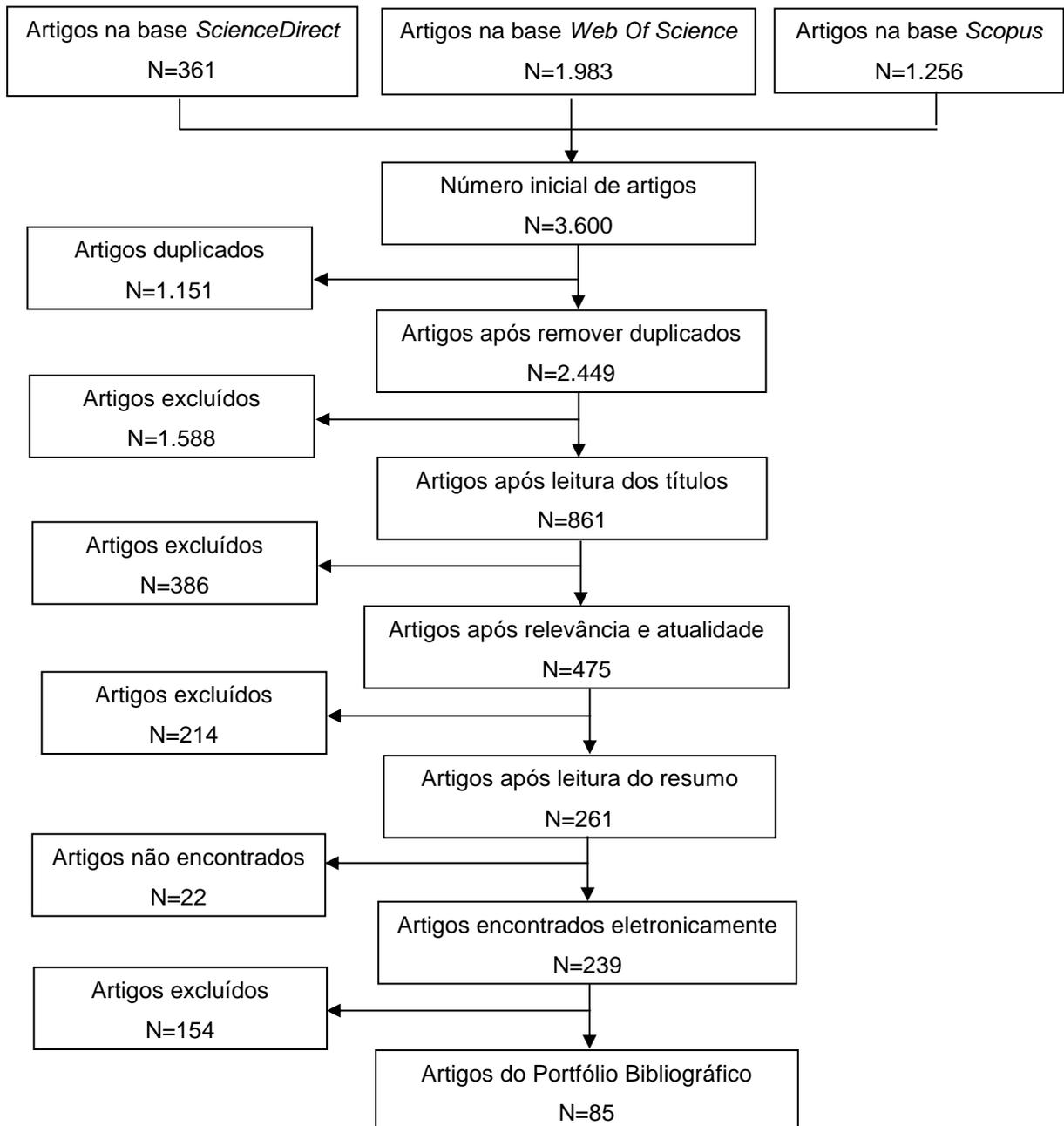
Para avaliação crítica, primeiramente foi realizada a leitura do título das 2.449 publicações, onde foram retirados os artigos em que o título não indicava a análise da carga de trabalho mental com o uso de ferramentas ou métodos e também foram eliminados os artigos de revisão de literatura, sendo excluídos 1.588 artigos, restando 861 para análises posteriores.

Na sequência, observou-se a relevância dos 861 trabalhos, por meio do número de citações no Google Acadêmico. O artigo mais citado recebeu 711 citações. Como linha de corte foi estabelecida 9 citações, de modo a representar 95% das citações totais, equivalente a 354 publicações. Outros 507 artigos foram citados menos de 9 vezes, dos quais 200 artigos não tiveram nenhuma citação. Como os artigos publicados mais recentemente possuem um menor tempo de acúmulo de citações, dos 507 artigos que receberam menos de 9 citações foram selecionadas as publicações de 2016 e 2017, as quais representaram 121 artigos, para se juntarem às 354 publicações relevantes. Dessa forma, após essa etapa restaram 475 artigos para prosseguimento da seleção do portfólio final.

A etapa posterior consistiu na leitura do resumo dos 475 artigos restantes. Excluíram-se as publicações em que o resumo explicitava tratar-se apenas de uma revisão sobre determinada ferramenta, sem nenhuma aplicação prática em trabalhadores. Assim, após leitura dos 475 resumos, restaram 261 artigos para leitura na íntegra. Dos 261 artigos desejados para leitura na íntegra, 22 não estavam disponíveis eletronicamente, restando 239 publicações para leitura integral.

A análise das 239 publicações foi realizada por meio da leitura da introdução, da metodologia e da conclusão, para ver se os artigos estavam alinhados com o objetivo e a pergunta de pesquisa. Dessa forma, foram mantidos os artigos que possuíam a aplicação de algum método ou ferramenta para avaliação da carga de trabalho mental. Assim, o portfólio bibliográfico final contou com 85 artigos, para posteriores análises. A Figura 4 apresenta os passos acima descritos.

Figura 4 - Etapas de construção do portfólio bibliográfico



Fonte: Elaborado pelo autor (2018).

As principais informações dos 85 artigos selecionados como portfólio bibliográfico (PB) foram coletadas, para posteriores análises bibliométrica e do conteúdo.

A análise bibliométrica envolveu o ano de publicação, o fator de impacto, os autores, os periódicos e as palavras-chave. Já a análise do conteúdo abordou as ferramentas de análise da carga de trabalho mental, classificando os trabalhos nas categorias: medidas subjetivas, medidas fisiológicas, medidas de desempenho ou a junção de mais de um dos tipos de métodos, e separando por ferramenta. Também foi feita a classificação dos indicadores das medidas fisiológicas utilizadas nos trabalhos. Os trabalhos ainda foram organizados conforme uma classificação estratificada dos trabalhadores e segundo o país de aplicação, buscando relações com os métodos de análise da CTM empregados.

3.3 ANÁLISE DOS MÉTODOS DE AVALIAÇÃO DA CTM

Com a revisão sistemática de literatura foram encontrados os métodos de avaliação da carga de trabalho mental mais utilizados no período entre 2000 e 2017. Assim, a segunda parte do trabalho consistiu em uma revisão e análise de cada um desses métodos: medidas fisiológicas, NASA-TLX, escala subjetiva, questionário, SWAT, tarefa secundária, ATWIT, RSME, SURG-TLX, BSPE, JDS, ESCAM, GSW, SOI, ISA, ISTAS 21, MCH, tarefa primária, IWS, RTLX, SURG-TLX + GOALS e TLI.

A ordem de apresentação das ferramentas seguiu a mesma citada acima e a mesma ordem apresentada nas tabelas dos resultados da revisão sistemática de literatura, seguindo a ordem decrescente do número de vezes em que elas foram encontradas nos artigos do portfólio bibliográfico.

Dessa forma, foram apresentados os seguintes dados de cada ferramenta:

- História e evolução da ferramenta;
- Metodologia da coleta de dados;
- Processo de análise dos dados e interpretação dos resultados;
- Principais aplicações;
- Vantagens e desvantagens;
- Análise crítica.

Como alguns métodos não possuem um procedimento específico, foram buscadas as técnicas mais utilizadas em cada caso, descrevendo as metodologias empregadas nos artigos do PB e outras relevantes encontradas na literatura.

Foi apresentada também uma linha do tempo com os métodos de avaliação da carga de trabalho mental encontrados no portfólio bibliográfico.

4 RESULTADOS E DISCUSSÕES

4.1 REVISÃO SISTEMÁTICA DE LITERATURA

4.1.1 Análise Bibliométrica

O Quadro 6 apresenta uma ordenação dos artigos que compõe o PB, a partir do ano de publicação.

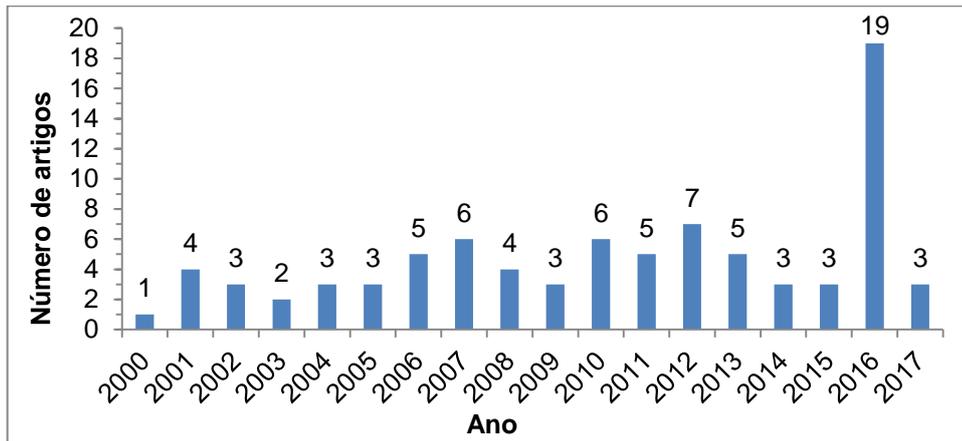
Quadro 6 - Artigos do PB organizados por ano de publicação

Ano	Artigos do PB
2000	(WEINGER et al., 2000)
2001	(BERGUER; SMITH; CHUNG, 2001), (GALSTER et al., 2001), (LING et al., 2001), (METZGER; PARASURAMAN, 2001)
2002	(EAST; BAUER JR; LANNING, 2002), (MOON et al., 2002), (WILSON, 2002)
2003	(COLLET et al., 2003), (WILSON; RUSSELL, 2003)
2004	(AVERTY et al., 2004), (MURAI et al., 2004), (KUIJER et al., 2004)
2005	(LAHELMA et al., 2005), (PICKUP et al., 2005), (KAWADA; OOYA, 2005)
2006	(AHLSTROM; FRIEDMAN-BERG, 2006), (GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006), (KERICK; ALLENDER, 2006), (PARK; JUNG, 2006), (WARD et al., 2006)
2007	(BOULTINGHOUSE et al., 2007), (DI NOCERA; CAMILLI; TERENCEZI, 2007), (LAHTINEN et al., 2007), (MURAI et al., 2007), (SALEEM et al., 2007), (WILSON; CALDWELL; RUSSELL, 2007)
2008	(DE RIVECOURT et al., 2008), (HANNULA et al., 2008), (MURAI et al., 2008), (YAO et al., 2008)
2009	(GOULD et al., 2009), (HENELIUS et al., 2009), (LIANG et al., 2009)
2010	(BYRNE et al., 2010), (DEY; MANN, 2010), (LEHRER et al., 2010), (LOPEZ et al., 2010), (RUBIO VALDEHITA et al., 2010), (ZHENG et al., 2010)
2011	(AYAZ et al., 2011), (HUTTUNEN et al., 2011), (KATAOKA; SASAKI; KANDA, 2011), (SCHULZ et al., 2011), (ZOER et al., 2011)
2012	(ARELLANO et al., 2012), (AYAZ et al., 2012), (ELFERING; GREBNER; HALLER, 2012), (MITROPOULOS; MEMARIAN, 2012), (YANG et al., 2012), (ZHENG et al., 2012a), (ZHENG et al., 2012b)
2013	(BYRNE et al., 2013), (KARHULA et al., 2013), (MOUZÉ-AMADY et al., 2013), (PEKKARINEN et al., 2013), (SAFARI et al., 2013)
2014	(DADI et al., 2014), (HARRISON et al., 2014), (LEE et al., 2014)
2015	(BERG et al., 2015), (KOCA et al., 2015), (WHEELLOCK et al., 2015)
2016	(DARVISHI et al., 2016), (ERASLAN; CAN; ATALAY, 2016), (FALLAHI et al., 2016), (LIN; LU, 2016), (MANNARU et al., 2016a), (MANNARU et al., 2016b), (MANSIKKA et al., 2016a), (MANSIKKA et al., 2016b), (MARTIN et al., 2016), (MAZLOUMI et al., 2016), (MURAI et al., 2016), (OTHTMAN; ROMLI, 2016), (SÖNMEZ et al., 2016), (VÁSQUEZ et al., 2016), (WEIGL et al., 2016), (WIDYANTI; LARUTAMA, 2016), (YU et al., 2016), (ZARE et al., 2016), (ZAWIAH; DAWAL, 2016)
2017	(RUBIO-VALDEHITA; RODRIGO-TAPIAS, 2017), (SCERBO; BRITT; STEFANIDIS, 2017), (WILSON et al., 2017)

Fonte: Elaborado pelo autor (2018).

Conforme a Figura 5, é possível observar o número de artigos do portfólio bibliográfico por ano de publicação. A maior concentração de artigos compreendeu o intervalo de 2010 a 2017, com 60% das publicações, com o ano de 2016 possuindo 22% dos artigos do PB.

Figura 5 - Distribuição das publicações por ano



Fonte: Elaborado pelo autor (2018).

Além disso, é possível dispor os artigos em relação ao fator de impacto do periódico, o qual classifica os periódicos por relevância científica, analisando o número de citações (LACERDA; ENSSLIN; ENSSLIN, 2014). Dessa forma, o FI de um determinado ano de um periódico é determinado pela divisão do número de citações de artigos ou de comunicados científicos do periódico nos dois anos anteriores pelo número de artigos publicados pelo periódico nesse mesmo período (RUIZ; GRECO; BRAILE, 2009). Sendo assim, o fator de impacto do ano de 2015 refere-se as publicações do periódico de 2013 e 2014.

A Tabela 1 apresenta os artigos, estratificados por ano, por meio de informações estatísticas relacionadas ao FI do seu periódico. Foi utilizado o fator de impacto de 2015 para todas as situações, visto que no momento da pesquisa era o mais recente, publicado pela JRC.

Tabela 1 - Estatística das variações do FI dos periódicos por ano de publicação dos artigos

Ano	Média FI	Mdn. FI	Desv. Pad. FI	Faixas do Fator de Impacto									Q.	%
				0 - 0.5	0.5 - 1.0	1.0 - 2.0	2.0 - 3.0	3.0 - 4.0	4.0 - 5.0	5.0 - 6.0	6.0 - 9.0			
2000	1.284	1.284	0.000			1						1	1,18	
2001	1.037	0.888	1.060	2		1	1					4	4,71	
2002	1.256	0.405	1.836	2				1				3	3,53	
2003	1.370	1.370	0.000			2						2	2,35	
2004	0.708	0.755	0.686	1	1	1						3	3,53	
2005	2.228	2.219	0.519			1	2					3	3,53	
2006	0.912	1.099	0.635	1		3						4	4,71	
2007	1.013	0.715	1.226	3	2	1		1				7	8,24	
2008	1.113	1.465	0.742	1		3						4	4,71	
2009	0.725	0.462	0.886	2		1						3	3,53	
2010	2.816	3.012	1.894	1		1	1	2		1		6	7,06	
2011	1.908	1.449	2.161	2		2				1		5	5,88	
2012	2.162	2.495	1.871		2	1		2		1		6	7,06	
2013	1.641	1.449	1.247	1		3	1	1				6	7,06	
2014	2.164	1.800	1.235			2		1				3	3,53	
2015	4.412	3.309	3.730			1		1			1	3	3,53	
2016	0.998	0.250	1.498	11	1	4	1	1		1		19	22,35	
2017	1.119	1.384	1.012	1		2						3	3,53	
TOTAL				28	6	30	6	10	0	4	1	85	100,00	
PERCENTUAL (%)				33	7	35	7	12	0	5	1	100		

Fonte: Elaborado pelo autor (2018).

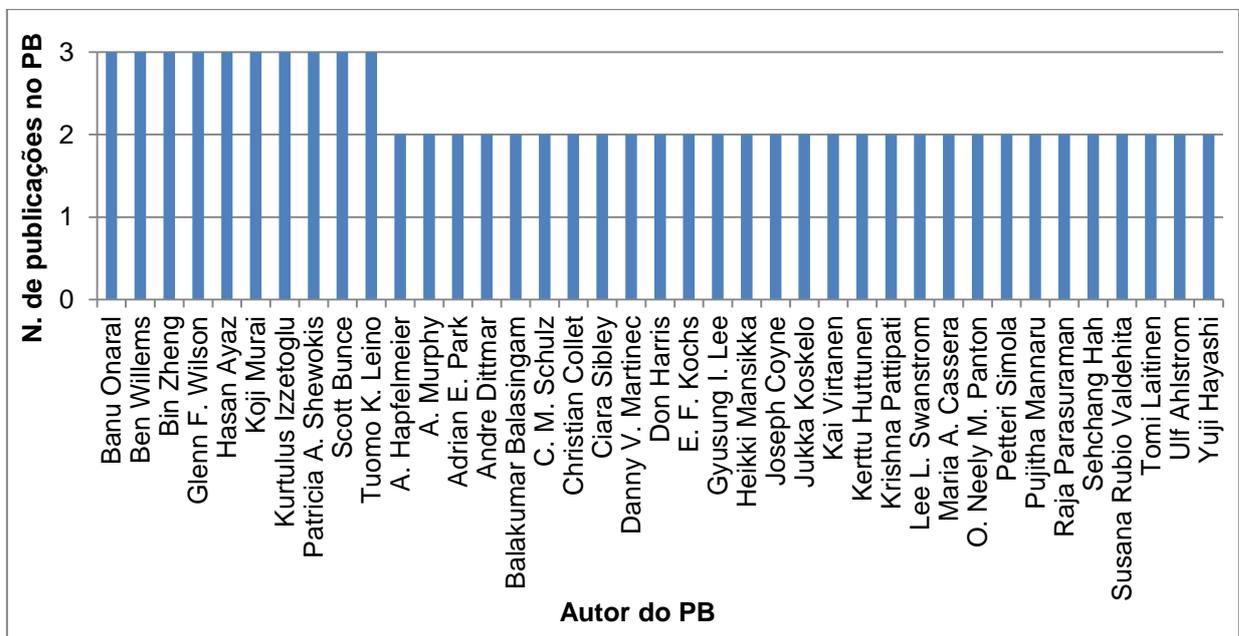
O fator de impacto das publicações variou de 0, em periódicos que não possuíam classificação, a 8.569, tendo média geral de 1.554, mediana de 1.370 e desvio padrão de 1.621, mostrando relevância científica geral da amostra, quando utilizado o FI como critério de avaliação das publicações. Além disso, apenas 4 anos apresentaram média do fator de impacto dos periódicos dos artigos menor que 1.000 e 60% dos artigos individualmente apresentaram FI dos periódicos maior ou igual a 1.000. Ressalta-se a relevância científica, visto que esse foi um dos critérios para seleção do portfólio.

Por meio do número de citações dos artigos no Google Acadêmico, 3 artigos apresentaram maior destaque: (WILSON, 2002) com 280 citações, (BERGUER; SMITH; CHUNG, 2001) com 247 citações e (AYAZ et al., 2012) com 227 citações. Além disso, 4 outros artigos tiveram entre 100 e 200 citações (AHLSTROM;

FRIEDMAN-BERG, 2006; LAHELMA et al., 2005; METZGER; PARASURAMAN, 2001; WILSON; RUSSELL, 2003).

Foram encontrados 357 autores diferentes nas publicações do PB. Dentre eles, 10 autores se destacaram, com 3 publicações cada: Banu Onaral, Ben Willems, Bin Zheng, Glenn F. Wilson, Hasan Ayaz, Koji Murai, Kurtulus Izzetoglu, Patricia A. Shewokis, Scott Bunce e Tuomo K. Leino. Outros 29 autores tiveram duas publicações no PB. A Figura 6 apresenta a relação dos autores com o número de artigos no PB, com o mínimo de 2 artigos.

Figura 6 - Autores com duas ou mais publicações no PB



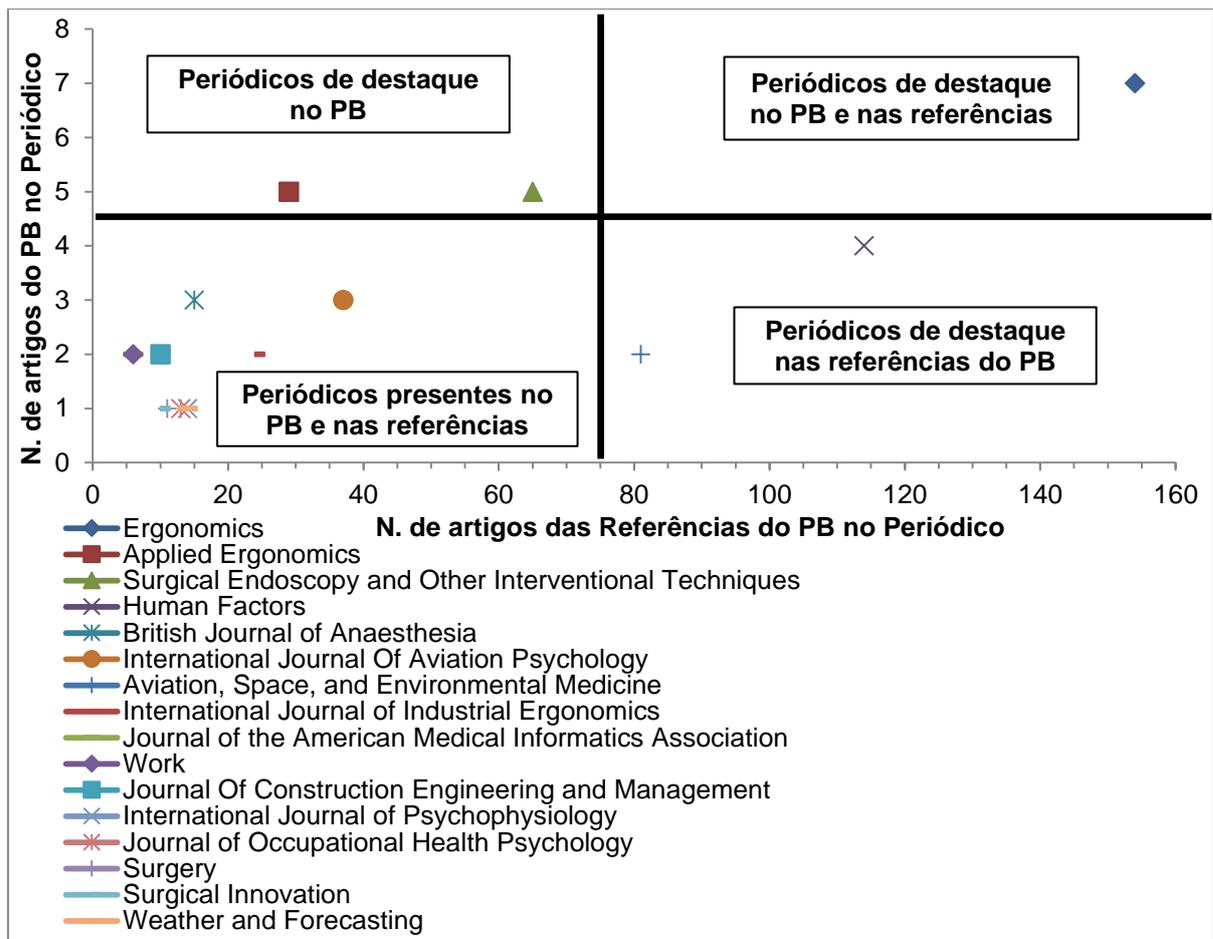
Fonte: Elaborado pelo autor (2018).

Buscando identificar os periódicos do portfólio com mais publicações do tema carga mental, realizou-se o cruzamento do número de vezes que o periódico aparecia no portfólio bibliográfico com o número de vezes que o periódico aparecia nas referências do PB, conforme Figura 7. Neste gráfico estão presentes os periódicos que possuíam pelo menos 2 artigos no PB ou apareciam no mínimo 10 vezes nas referências dos artigos do PB.

Foram encontrados 58 periódicos diferentes no PB, com maior destaque para o periódico "*Ergonomics*", com sete artigos no portfólio bibliográfico e 154 publicações nas referências do PB. Os periódicos "*Applied Ergonomics*" e "*Surgical Endoscopy and Other Interventional Techniques*" destacaram-se dentro do PB, com

cinco artigos cada. Já dentro das referências destacaram-se os periódicos "*Human Factors*" e "*Aviation, Space and Environmental Medicine*", com 114 e 81 artigos nas referências do PB, respectivamente. Nenhum periódico brasileiro foi encontrado no PB.

Figura 7 - Publicações das referências e dos artigos do PB por periódico

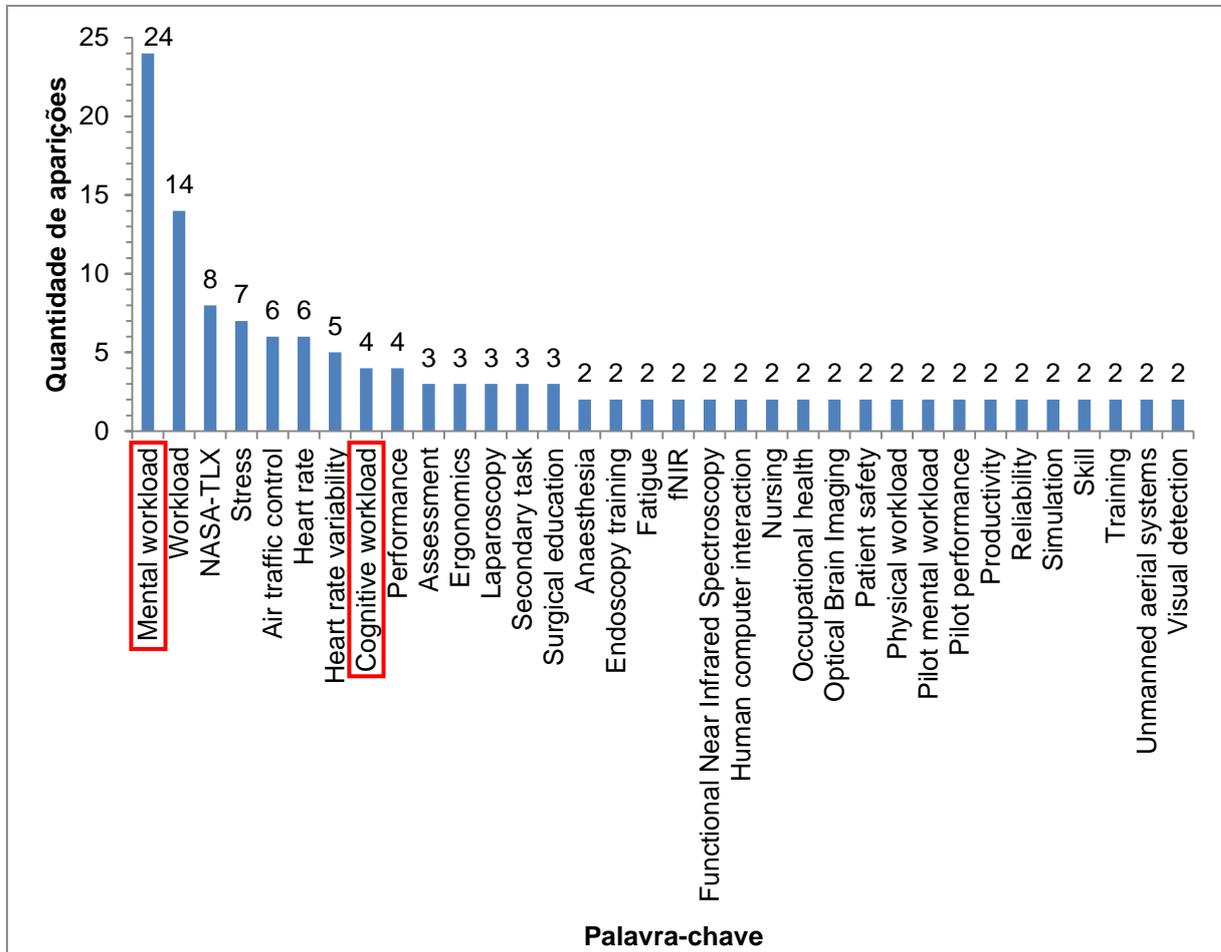


Fonte: Elaborado pelo autor (2018).

No portfólio bibliográfico foram encontradas 207 palavras-chave diferentes. 34 delas foram usadas em mais de um artigo, representando 43% das palavras-chaves, as quais estão dispostas na Figura 8. Dessas 34 palavras-chave, duas foram empregadas na atual pesquisa: "*mental workload*" e "*cognitive workload*", as quais encontram-se destacadas na Figura 8. O principal destaque foi a palavra-chave "*mental workload*", aparecendo em 24 trabalhos, seguida das palavras "*workload*" e "*NASA-TLX*", citadas 14 e 8 vezes, respectivamente. Ainda a palavra-chave "*stress*" foi utilizada 7 vezes, as palavras-chave "*air traffic control*" e "*heart rate*" apareceram 6

vezes e a palavra "heart rate variability" foi empregada 5 vezes. Todas as demais palavras-chave foram usadas em menos de 5 trabalhos.

Figura 8 - Principais palavras-chave do PB



Fonte: Elaborado pelo autor (2018).

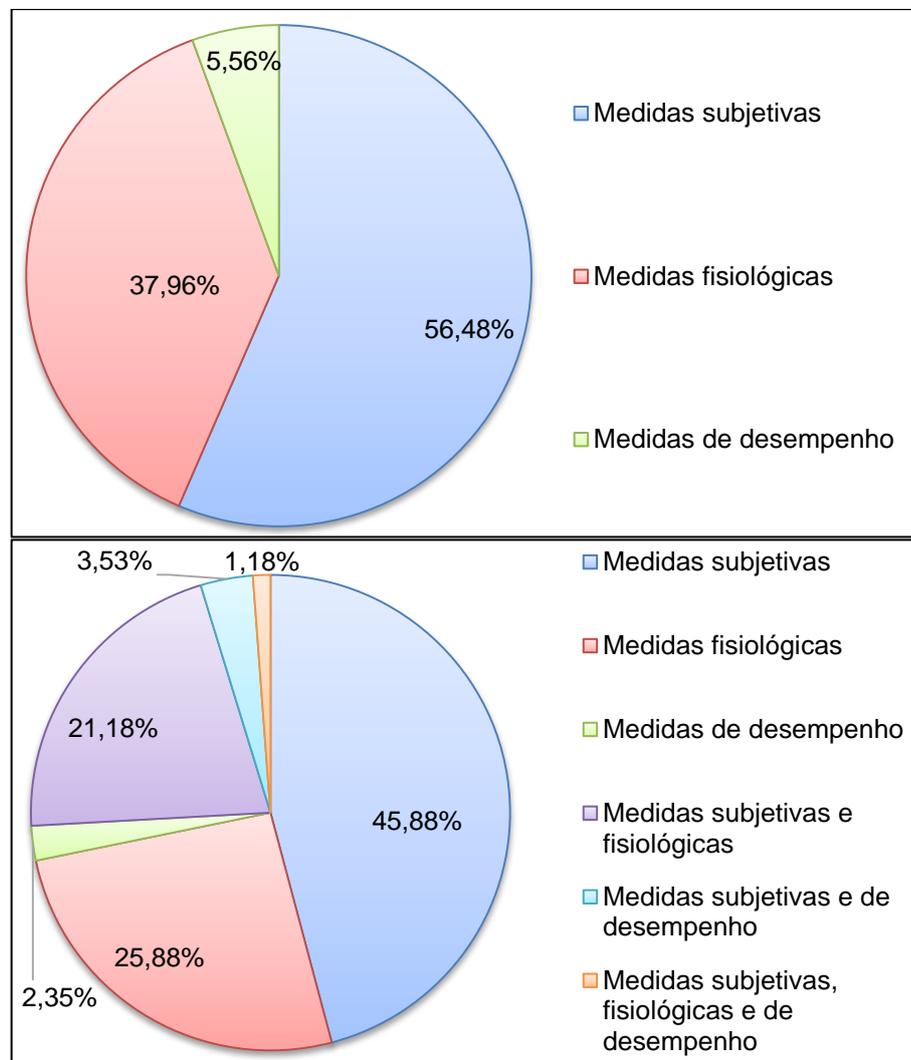
4.1.2 Análise do Conteúdo

A análise do conteúdo buscou categorizar os métodos de avaliação da carga de trabalho mental aplicados em cada artigo, descobrir as ferramentas mais utilizadas, os trabalhadores analisados e os países que estudam a CTM, buscando relações entre todos.

As categorias das ferramentas utilizadas para avaliar a carga de trabalho mental podem ser divididas em medidas subjetivas, medidas fisiológicas e medidas de desempenho (MORAY, 1998). Assim, o uso de medidas subjetivas foi encontrado

em 61 artigos, o uso de medidas fisiológicas foi visto em 41 artigos e o uso de medidas de desempenho foi encontrado em 6 artigos do portfólio bibliográfico. Partindo de outro ponto de vista, os trabalhos foram agrupados do seguinte modo: 39 artigos utilizaram apenas medidas subjetivas, 22 utilizaram apenas medidas fisiológicas, 18 utilizaram medidas subjetivas e fisiológicas, 3 utilizaram medidas subjetivas e de desempenho, 2 utilizaram apenas medidas de desempenho e 1 utilizou as três categorias. A classificação dos artigos, conforme os dois métodos acima descritos, está representada na Figura 9, por meio do percentual.

Figura 9 - Classificação dos trabalhos do PB



Fonte: Elaborado pelo autor (2018).

Foram buscados todos os métodos utilizados nos trabalhos do PB. A Tabela 2 apresenta todas as ferramentas encontradas, conforme nome exibido no(s) artigo(s), contendo a quantidade de vezes em que apareceu e o percentual

correspondente. A Tabela 2 também apresenta uma análise de cada ferramenta por fator de impacto dos periódicos, contendo a média, a mediana e o desvio padrão do FI, além do número de artigos que utilizou a ferramenta por faixa de fator de impacto.

Tabela 2 – Ferramentas de análise da carga de trabalho mental analisadas por meio de dados estatísticos do FI dos seus periódicos

Ferramenta	Méd. FI	Mdn. FI	Desv. Pad. FI	Faixas do Fator de Impacto								Q.	%
				0 – 0.5	0.5 – 1.0	1.0 – 2.0	2.0 – 3.0	3.0 – 4.0	4.0 – 5.0	5.0 – 6.0	6.0 – 9.0		
Medidas fisiológicas	1.383	1.370	1.559	17	3	13	2	3		3		41	35,7
NASA-TLX	1.661	1.264	1.873	10	5	13	1	4		2	1	36	31,3
Escala subjetiva	2.058	1.410	1.919	1	1	2	1			1		6	5,1
Questionário	0.885	0.462	0.878	3		1	1					5	4,3
SWAT	0.794	0.732	0.736	2	1	1						4	3,4
Tarefa secundária	3.600	3.667	1.706			1		2		1		4	3,4
ATWIT	1.400	1.400	0.566		1	1						2	1,7
RSME	1.449	1.449	0.000			2						2	1,7
SURG-TLX	3.425	3.425	0.163					2				2	1,7
BSPE	5.616	5.616								1		1	0,9
JDS	2.000	2.000				1						1	0,9
ESCAM	0.226	0.226		1								1	0,9
GSW	1.449	1.449				1						1	0,9
SOI	0.755	0.755			1							1	0,9
ISA	1.972	1.972				1						1	0,9
ISTAS 21	0.715	0.715			1							1	0,9
MCH	0.464	0.464		1								1	0,9
Tarefa primária	0.000	0.000		1								1	0,9
IWS	1.713	1.713				1						1	0,9
RTLX	0.000	0.000		1								1	0,9
SURG-TLX + GOALS	2.523	2.523					1					1	0,9
TLI	0.755	0.755			1							1	0,9
TOTAL				37	14	38	6	11	0	8	1	115	100
PERCENTUAL (%)				32	12	33	5	10	0	7	1	100	

Fonte: Elaborado pelo autor (2018).

Foram encontradas 22 ferramentas, destacando-se o método de análise de medidas fisiológicas e a ferramenta de análise subjetiva NASA-TLX, os quais juntos representaram praticamente 67% das aparições. Ainda foram encontradas em mais

de um trabalho as ferramentas: escala subjetiva, questionário, SWAT, tarefa secundária, ATWIT, RSME e SURG-TLX. Em contrapartida, 13 ferramentas foram encontradas em apenas 1 artigo do PB: BSPE, JDS, ESCAM, GSW, SOI, ISA, ISTAS 21, MCH, tarefa primária, IWS, RTLX, SURG-TLX + GOALS e TLI. O Apêndice B apresenta os artigos do PB organizados conforme as ferramentas de análise da CTM.

Os periódicos que publicaram artigos contendo análises das 9 principais ferramentas, as quais foram encontradas em pelo menos uma publicação, apresentaram de forma geral uma boa média de fator de impacto, com valor acima de 1.000, com exceção das ferramentas “questionário” e “SWAT”, com valor levemente inferior. Em relação a mediana, as mesmas ferramentas também foram as únicas a apresentarem valor de FI menor que 1.000. Essas 9 ferramentas representam, juntas, 88% das aplicações.

A maior concentração de artigos do PB apresentou fator de impacto dos periódicos entre 0 e 0.5 e entre 1.0 e 2.0, representando 68% das publicações. Em seguida, 14% das publicações apresentaram FI dos periódicos entre 0.5 e 1.0 ou entre 3.0 e 4.0. A faixa entre 4.0 e 5.0 não apresentou publicações e a faixa entre 6.0 e 9.0 apresentou apenas um artigo.

Outra análise que pode ser feita é a relação entre as ferramentas de análise da carga de trabalho mental com o ano de publicação dos artigos que possuíam suas aplicações, observando a evolução, ou o declínio, do uso de cada ferramenta. Essa relação está presente na Tabela 3, com o nome da ferramenta e o número de vezes que ela foi utilizada no PB em cada ano do intervalo pesquisado.

É possível observar que a utilização das ferramentas está bem espalhada no intervalo de tempo de 2000 a 2017. Como as ferramentas mais utilizadas são as medidas fisiológicas e o NASA-TLX, é possível fazer uma melhor análise do seu uso no domínio de tempo utilizado. Ambas as ferramentas apresentaram uma constância de publicações por ano, variando entre 0 e 4 publicações, com exceção de 2016, visto que 23% dos artigos do PB foram publicados nesse ano, número muito mais elevado que qualquer outro ano. Por exemplo, o ano de 2016 possuiu exame de 19 artigos, enquanto o ano de 2012, segundo ano com mais publicações dentro do PB, possuiu 7 artigos.

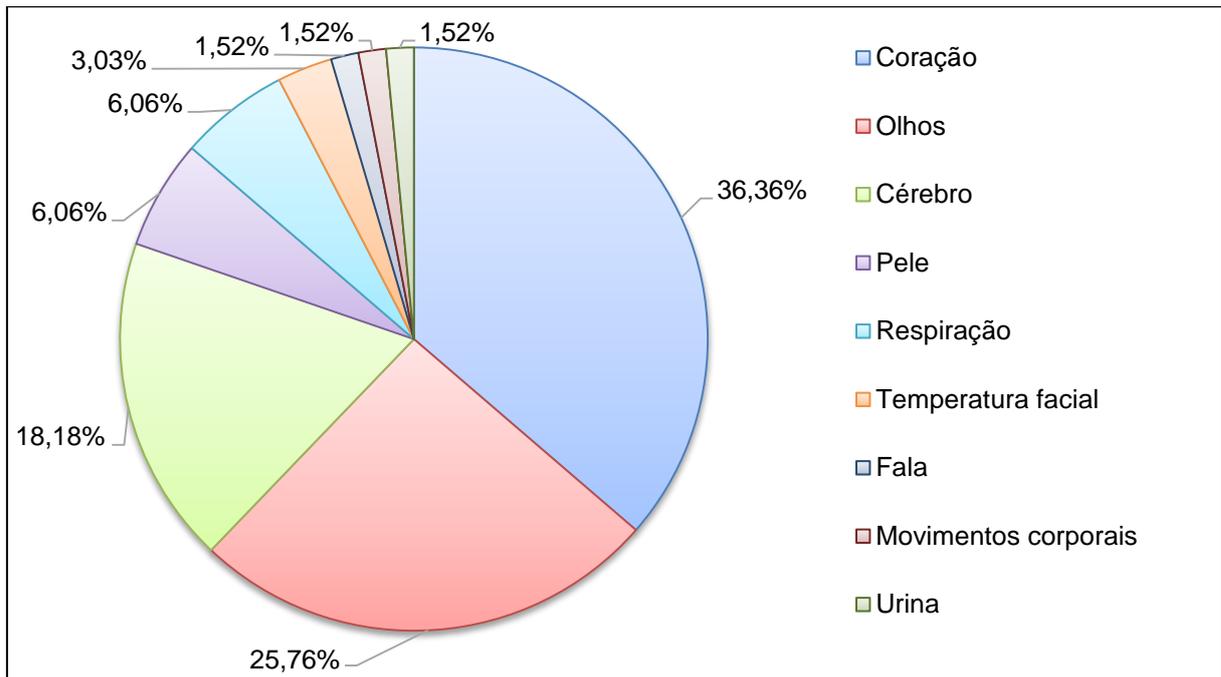
Tabela 3 – Aplicação das ferramentas da carga de trabalho mental por ano

Ferramenta	Ano de Publicação																	Total	
	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16		17
Medidas fisiológicas		2	3	2	2		2	4	4	3	2	4	2	1	1		9		41
NASA-TLX		2		1	1		1	4		1	4	1	4	3	2	2	8	2	36
Escala subjetiva			1		1	1		1				1		1					5
Questionário	1					1				1		1	1						5
SWAT						1	1										2		4
Tarefa secundária											2		1	1					4
ATWIT							1								1				2
RSME							1		1										2
SURG-TLX																1	1		2
BSPE																	1		1
JDS														1					1
ESCAM																	1		1
GSW														1					1
SOI					1														1
ISA																		1	1
ISTAS 21													1						1
MCH																	1		1
Tarefa primária																	1		1
IWS						1													1
RTLX																	1		1
SURG-TLX + GOALS																	1		1
TLI					1														1
TOTAL	1	4	4	3	6	4	6	9	5	5	8	7	9	8	4	3	26	3	115

Fonte: Elaborado pelo autor (2018).

Para a análise das medidas fisiológicas foram encontrados, nos trabalhos do PB, 9 indicadores diferentes: coração, olhos, cérebro, respiração, pele, temperatura facial, fala, movimentos corporais e urina, os quais estão apresentados na Figura 10.

Figura 10 - Indicadores de medidas fisiológicas utilizados para analisar a CTM



Fonte: Elaborado pelo autor (2018).

O coração foi analisado 24 vezes, os olhos 17 vezes, o cérebro 12 vezes, a pele 4 vezes, a respiração 4 vezes, a temperatura facial 2 vezes e a fala, os movimentos do corpo e a urina 1 vez cada. Dessa forma, os três principais indicadores corresponderam, juntos, a 80% das análises. O Apêndice C apresenta os artigos do PB organizados conforme os indicadores fisiológicos.

Quando o indicador utilizado foi o coração, a maior parte dos trabalhos, 13 artigos, analisou dados retirados do eletrocardiograma. Além disso, 8 trabalhos analisaram a frequência cardíaca, mas não especificaram como foi feita a medição. Ainda, 2 artigos analisaram a pulsação e 1 publicação não especificou quais dados do coração foram analisados.

Em relação aos olhos, 6 trabalhos analisaram o eletroculograma. Os outros 11 não especificaram o método utilizado, mas analisaram o comportamento visual, por meio dos movimentos oculares, medidas pupilares e taxa e duração das piscadas.

O cérebro foi analisado por meio de 3 técnicas diferentes: 8 trabalhos utilizaram o eletroencefalograma, 3 utilizaram a espectroscopia funcional de infravermelho próximo e 1 analisou as perturbações espectrais relacionadas ao evento.

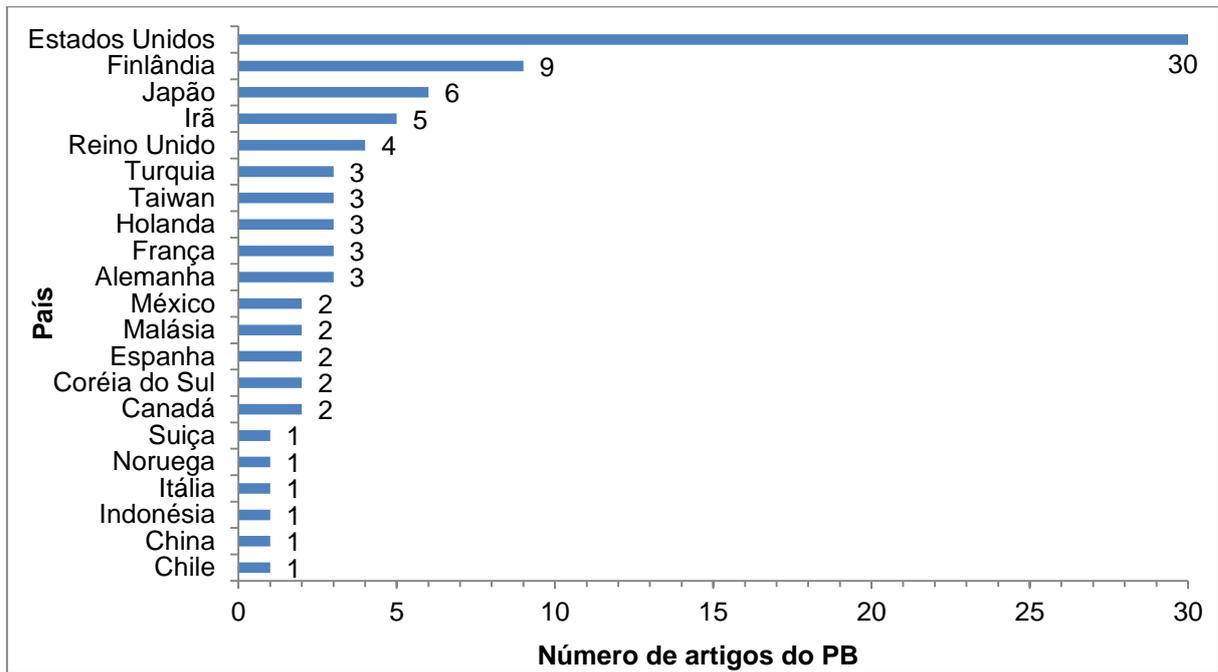
A pele foi estudada por 2 métodos: 2 trabalhos estudaram o nível de condutância cutânea tônica e 2 utilizaram a atividade eletrodérmica. A respiração foi analisada por meio das medições de ondas respiratórias, observando a frequência respiratória. A temperatura facial utilizou medidas da temperatura do rosto, comparando principalmente a temperatura nasal, nos 2 artigos do PB.

Os movimentos do corpo utilizaram movimentos horizontais, movimentos da cabeça e dados eletromiográficos. O trabalho que analisou a urina utilizou a taxa de excreção urinária de adrenalina e o artigo que analisou a fala utilizou dados da voz.

Os artigos do portfólio bibliográfico foram desenvolvidos em 21 países diferentes, com destaque para os Estados Unidos com 30 trabalhos, a Finlândia com 9 trabalhos, o Japão com 6, o Irã com 5 e o Reino Unido com 4 trabalhos, com a soma desses países representando 63% das publicações do PB. Além disso, foram realizados trabalhos na Turquia, em Taiwan, na Holanda, na França, na Alemanha, no México, na Malásia, na Espanha, na Coreia do Sul, no Canadá, na Suíça, na Noruega, na Itália, na Indonésia, na China e no Chile. A Figura 11 apresenta todos os países onde foram realizados os artigos do PB, com o número de vezes. O Apêndice D apresenta os artigos do PB organizados conforme os países de aplicação das pesquisas.

Em relação aos países de aplicação e as ferramentas utilizadas, nos Estados Unidos o destaque foram as medidas fisiológicas que apareceram 14 vezes e o questionário NASA-TLX que foi utilizado 12 vezes; na Finlândia as medidas fisiológicas foram utilizadas 6 vezes e as escalas subjetivas 4 vezes; no Japão as medidas fisiológicas foram utilizadas 5 vezes; no Irã o questionário NASA-TLX foi utilizado 4 vezes e no Reino Unido o questionário NASA-TLX e a análise de tarefa secundária foram utilizados 2 vezes.

Figura 11 - Países de aplicação da pesquisa



Fonte: Elaborado pelo autor (2018).

Para melhor visualização, a Figura 12 contém um mapa mundial, com a indicação dos países onde foram realizados os artigos do PB juntamente com a porcentagem de trabalhos. Além disso, os cinco principais países estão destacados.

Figura 12 – Mapa com países onde foram desenvolvidos os trabalhos do PB



Fonte: Elaborado pelo autor (2018).

A última análise envolveu os trabalhadores que foram avaliados. Os trabalhadores foram estratificados em 12 grupos: trabalhadores de saúde, pilotos, controladores de tráfego aéreo, operadores de máquinas e/ou equipamentos, trabalhadores navais, motoristas, trabalhadores industriais, operadores de sistemas aéreos não tripulados (UAS), controladores rodoviários, trabalhadores não identificados nos artigos, vários profissionais em um mesmo artigo e outros trabalhadores. A Tabela 4 apresenta cada uma dessas estratificações de trabalhadores com o número de trabalhos que eles apareceram e o percentual. O Apêndice E apresenta os artigos do PB organizados conforme os grupos estratificados de trabalhadores.

Tabela 4 - Trabalhadores analisados nos artigos do PB

Grupo estratificado de trabalhadores	Quantidade	Percentual (%)
Trabalhadores de saúde	25	29,41%
Pilotos	14	16,47%
Controladores de tráfego aéreo	10	11,76%
Operadores	6	7,06%
Trabalhadores navais	5	5,88%
Motoristas	4	4,71%
Trabalhadores industriais	3	3,53%
Operadores de UAS	2	2,35%
Controladores rodoviários	2	2,35%
Não identificado	2	2,35%
Vários profissionais	6	7,06%
Outros	6	7,06%
TOTAL	85	100,00%

Fonte: Elaborado pelo autor (2018).

Os artigos que envolveram os trabalhadores de saúde foram realizados principalmente nos Estados Unidos, 11 vezes, e no Reino Unido e na Alemanha, 3 vezes em cada. As principais ferramentas utilizadas foram o NASA-TLX, aplicado 14 vezes, as medidas fisiológicas, em 5 artigos, e a análise da tarefa secundária, em 4 trabalhos.

Os pilotos foram analisados 5 vezes na Finlândia e 4 vezes nos Estados Unidos, utilizando 13 vezes as medidas fisiológicas, 4 vezes o NASA-TLX e 3 vezes as escalas subjetivas. Os controladores de tráfego aéreo foram analisados

principalmente nos Estados Unidos (7 vezes) e no Irã (2 vezes), por meio principalmente das medidas fisiológicas (6 vezes) e do NASA-TLX (5 vezes).

Os operadores de máquinas e/ou equipamentos tiveram como país de destaque a Coréia do Sul, com 2 artigos e como ferramenta principal de análise o NASA-TLX, em 4 trabalhos. As publicações que envolveram trabalhadores navais foram concentradas no Japão, com 4 trabalhos, e tiveram como principal método de análise as medidas fisiológicas, aparecendo nos 5 trabalhos. Trabalhos com motoristas foram realizados em 4 países distintos e as análises de medidas fisiológicas foram encontradas em 3 trabalhos. Já os artigos que envolveram operadores de sistemas aéreos não tripulados foram desenvolvidos ambos nos Estados Unidos, utilizando medidas fisiológicas.

De forma geral, os procedimentos que possuíam várias tarefas e/ou etapas, permitindo alguma parada ou intervalo entre elas, utilizaram principalmente avaliações por medidas subjetivas multidimensionais, principalmente utilizando a ferramenta NASA-TLX, e as atividades que não permitiam parada, principalmente envolvendo atividades de aviação, de navegação e rodoviárias, optaram preferencialmente pelo uso de medidas fisiológicas, muitas vezes utilizando o NASA-TLX para comparação e validação, e pelo uso de medidas subjetivas unidimensionais, podendo utilizá-las para avaliar a carga de trabalho mental durante o procedimento.

4.2 MÉTODOS DE AVALIAÇÃO DA CARGA DE TRABALHO MENTAL

4.2.1 Medidas Fisiológicas

As medidas fisiológicas buscam relacionar mudanças fisiológicas com mudanças nos níveis de carga mental, o que acontece em estudos desde 1957, segundo Wierwille (1979). Espíndola (2013) acredita que quando comparadas com as medidas subjetivas e medidas de desempenho, elas apresentam a vantagem de ser medidas continuamente, não exigir resposta pública e possuir natureza multidimensional. Em contrapartida, como desvantagem, requerem altos custos e necessitam cuidados na avaliação e interpretação dos resultados.

Vários ramos de atividades têm utilizado como recurso de avaliação da carga mental as medidas fisiológicas, com grande destaque para a aviação, devido à

complexidade das tarefas e a busca por medições em tempo real das atividades realizadas, pois existem grandes flutuações de carga de trabalho mental durante a operação de voo, devido às demandas de tarefas que variam de muito altas a muito baixas (DI NOCERA; CAMILLI; TERENCEZI, 2007; YAO et al., 2008; WILSON, 2002). Assim, torna-se importante medir a CTM para utilizar em projetos de aeronaves, análise de missões ou simplesmente para avaliar o desempenho do piloto (YAO et al., 2008).

Dessa forma, devido as complexidades do voo, a determinação da carga de trabalho mental do piloto requer mais de uma medida fisiológica, buscando-se examinar também as relações entre as múltiplas medidas (WILSON, 2002). Isso não se restringe as atividades da aviação: a importância de avaliar as respostas de medidas fisiológicas diferentes é evidente para os demais ramos, como a navegação e os transportes terrestres.

Assim sendo, são utilizadas usualmente algumas medidas fisiológicas para avaliar a CTM: frequência cardíaca (HR) e variabilidade da frequência cardíaca (HRV), as quais podem ser retiradas do eletrocardiograma (ECG); medidas retiradas do eletroencefalograma (EEG), as quais monitoram a atividade cerebral; movimentos oculares, por meio do eletrooculograma (EOG), utilizando medidas pupilares (PD), número de piscadas, tempo médio de permanência e a duração da fixação; medidas da respiração e medidas eletrodérmicas, por meio da condutância da pele (SC) (COLLET et al., 2003; DE RIVECOURT et al., 2008; EAST; BAUER JR; LANNING, 2002; GOULD et al., 2009; YAO et al., 2008; WILSON, 2002).

As taxas de intermitência dos olhos podem conter informações valiosas relacionadas as demandas visuais da tarefa, a frequência cardíaca pode ser útil para determinar a resposta global quanto às demandas de tarefas e o eletroencefalograma pode fornecer informações sobre alta carga de trabalho e desatenção (LIANG et al., 2009).

Dessa forma, as análises fisiológicas serão divididas em nove indicadores diferentes: coração, olhos, cérebro, respiração, pele, temperatura facial, fala, movimentos corporais e urina, os quais serão descritos subsequentemente.

4.2.1.1 Coração

A maioria dos estudos que utiliza dados do coração como medida fisiológica retira esses dados do eletrocardiograma (ECG), o qual é coletado por meio de eletrodos, que podem ser de estanho ou Ag/AgCl (prata/cloreto de prata) e podem ser descartáveis ou reutilizáveis, colocados sobre o esterno, a clavícula ou o tórax. Para o contato entre os eletrodos e o corpo humano ocorrer da melhor forma, a pele é limpa com álcool. A gravação dos dados é feita por meio de um gravador e uma placa de armazenamento, os quais posteriormente são transferidos para um computador para análises (WILSON, 2002; WILSON; CALDWELL; RUSSELL, 2007; YAO et al., 2008).

Em vez dos eletrodos separados, pode-se utilizar uma cinta peitoral e um relógio de pulso. A cinta peitoral possui sensores que medem os intervalos entre os batimentos, os quais são enviados para o relógio de pulso que possui uma memória (MURAI et al., 2008).

Outro método que pode ser utilizado é o uso de uma camisa de náilon leve com sensores fisiológicos embutidos. Esses sensores de ECG são posicionados no ponto médio do lado da caixa torácica. Assim, os dados de ECG brutos também são coletados e depois exportados para um programa de *software* para análise (LEHRER et al., 2010).

Diversos dados podem ser coletados do ECG. Dentre eles, as ondas R são destaque, porque a amplitude do pico R é notável em comparação com outras ondas e assim os picos são mais facilmente encontrados (MURAI et al., 2004). Com isso, podem ser calculados os intervalos entre batimentos (IBIs) entre ondas R sucessivas. Com os dados IBI e por meio de *software*, pode ser calculada a variabilidade da frequência cardíaca (HRV) nas bandas média e alta. Dados da frequência cardíaca média podem ser apresentados em batimentos por minuto (bpm) e dados da variabilidade da frequência cardíaca em hertz (Hz) (MARTIN et al., 2016; MURAI et al., 2004; YAO et al., 2008; WILSON, 2002).

Podem ser utilizados dados medidos no domínio do tempo e da frequência. Utilizando o domínio do tempo, pode-se calcular a média, o desvio padrão e a razão entre média e desvio padrão da frequência cardíaca. Já a análise baseada no domínio da frequência calcula o valor de baixa frequência (LF), alta frequência (HF) e SNS (SNS = LF/HF), usando uma Transformada Rápida de Fourier (FFT) e o Método de

Entropia Máxima (MEM) (DE RIVECOURT et al., 2008; GOULD et al., 2009; WILSON, 2002).

Os intervalos dos batimentos cardíacos também podem ser extraídos do ECG por meio de um sistema Fuzzy, utilizando para isso um algoritmo de suavização para gerar curva de variação de largura de pulso. Com isso, pode-se desenvolver uma função para representar a carga de trabalho mental como uma função das variáveis retiradas do ECG (MOON et al., 2002).

Com os dados do ECG também podem ser desenvolvidos modelos de redes neurais artificiais (ANN) e regressão multilinear (MLR), por meio de um *software*. O modelo ANN é treinado com um algoritmo de retropropagação e o modelo MLR é construído por meio de cálculos matriciais apropriados (HANNULA et al., 2008; LIANG et al., 2009).

Várias métricas de HRV linear e não linear também podem ser utilizadas. As medidas do domínio do tempo das métricas da HRV podem ser o SDNN (desvio padrão do intervalo normal ao normal (NN)), a RMSSD (raiz quadrada das diferenças médias quadráticas de intervalos NN sucessivos) e a pNN50 (porcentagem de intervalos NN sucessivos maiores que 50 ms). O SDNN é uma estimativa da HRV geral e a RMSSD e o pNN50 são estimativas de componentes de curto prazo da HRV. Já os métodos não lineares podem ser os desvios padrão da plotagem de Poincaré (SD1, SD2), a entropia aproximada (ApEn), a entropia amostral (SampEn), a entropia de Shannon da distribuição de probabilidade dos comprimentos de linha diagonais (ShanEn), a dimensão de correlação (D2) e a entropia de permutação (PeEn). Ainda, em busca de uma conclusão mais robusta sobre as métricas de frequência cardíaca linear ou não linear, uma abordagem multiparamétrica baseada em modelos de regressão logística pode ser realizada (MARTIN et al., 2016).

Embora os resultados variem, dependendo do ramo de trabalho estudado, de forma geral os estudos mostram que a variação do tamanho do pulso ou a HR tende a aumentar quando se espera que a carga de trabalho mental aumente (DE RIVECOURT et al., 2008; MOON et al., 2002). A frequência cardíaca também mostra aumento conforme aumento da responsabilidade do trabalhador e aumento de HR conforme menor a experiência (LAHTINEN et al., 2007; YAO et al., 2008). Além disso, aumentos bruscos da HR podem ocorrer durante incidentes particularmente difíceis (LEHRER et al., 2010).

Apesar disso, em alguns casos a frequência cardíaca não permite comparações gerais de carga de trabalho entre indivíduos ou sessões com trabalhos diferentes (SCHULZ et al., 2011).

Em relação as medidas mais sensíveis, a HRV pode ser menos sensível a mudanças graduais no esforço mental do que a HR, pois embora a HRV possa ser adequada para distinguir flutuações graduais nas demandas de tarefas em níveis de carga de trabalho baixos a intermediários, a HR pode ser mais adequada para distinguir essas flutuações em níveis altos de carga de trabalho (DE RIVECOURT et al., 2008).

Outro estudo apresentou que o valor de SNS é melhor que a frequência cardíaca na efetividade em encontrar as características da carga mental para cada condição de navegação. Os resultados mostraram que os valores de SNS aumentam enquanto o navegador humano precisa julgar o manuseio do navio e sentir a carga de trabalho emocional (MURAI et al., 2004).

Para Henelius et al. (2009) a medida IBI apresenta o melhor desempenho de classificação, com efeitos mais visíveis e consistentes, usando medições de curto prazo, podendo ser usada para distinguir CM baixa e alta durante um teste multitarefa computadorizado.

Ainda pode ser utilizada a informação combinada da HRV em ambas as bandas de frequência, buscando melhores resultados para determinar os níveis intermediários de esforço mental (DE RIVECOURT et al., 2008).

Para testar as hipóteses, comparando os trabalhadores e as diferentes tarefas e/ou etapas das tarefas executadas, muitos métodos estatísticos podem ser utilizados, com destaque para a análise de variância (ANOVA) e testes t (KATAOKA; SASAKI; KANDA, 2011; YAO et al., 2008; WILSON; CALDWELL; RUSSELL, 2007).

Em relação aos trabalhadores estudados, foram encontrados estudos com pilotos, controladores de tráfego aéreo, operadores de usinas nucleares, navegadores, coordenadores de portos, operadores de veículos aéreos não tripulados, trabalhadores de saúde e operadores agrícolas (DEY; MANN, 2010; HENELIUS et al., 2009; MOON et al., 2002; MURAI et al., 2004; MURAI et al., 2016; MURAI et al., 2007; SCHULZ et al., 2011; WILSON, 2002; WILSON; CALDWELL; RUSSELL, 2007; WILSON; RUSSELL, 2003).

4.2.1.2 Olhos

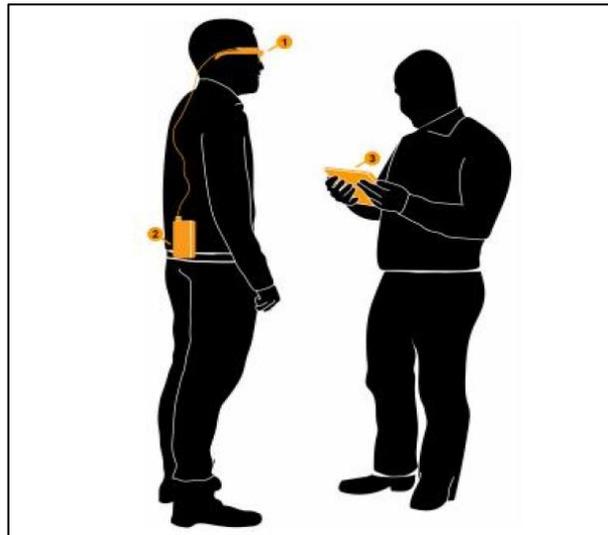
Quando são utilizados os olhos para avaliar a carga de trabalho mental, usualmente é utilizado o eletrooculograma (EOG) para coleta dos dados. Para isso, pode-se utilizar eletrodos autoadesivos, de prata/cloreto de prata, colocados acima e/ou abaixo do olho direito e nas laterais de ambos os olhos, além de utilizar um eletrodo de referência adicional sobre o osso zigomático direito. Para melhorar a aderência entre a pele e os eletrodos, pode-se limpar a região com álcool (BERGUER; SMITH; CHUNG, 2001; WILSON, 2002).

Para registrar a atividade dos movimentos oculares, pode-se utilizar também um oculômetro que consiste em um sistema de rastreamento dos olhos e cabeça, registrando a atividade de cada movimento dos olhos em grandes quantidades por segundo. Este sistema permite a coleta de dados oculares sem o uso de instrumentos invasivos e/ou desconfortáveis na cabeça. Ele usa diodos infravermelhos próximos para gerar padrões de reflexão na córnea dos olhos. Uma câmera coleta esses padrões de reflexão, juntamente com outras informações visuais e algoritmos de processamento de imagens que identificam recursos relevantes, incluindo os padrões de reflexão dos olhos e da córnea, os quais são analisadas posteriormente por meio de algoritmo computadorizado (AHLSTROM; FRIEDMANBERG, 2006; DI NOCERA; CAMILLI; TERENCEZI, 2007; KATAOKA; SASAKI; KANDA, 2011; ZHENG et al., 2012a).

O rastreamento ocular pode consistir em uma unidade principal, uma unidade de gravação e um *software*, conforme apresentado na Figura 13, respectivamente pelos números 1, 2 e 3. A unidade principal é muitas vezes referida como um vidro usável (rastreador ocular).

A unidade de gravação retém a bateria e mantém os dados gravados em um cartão de memória. Ela é conectada à unidade principal via cabo HDMI e é controlada a partir de um tablet ou de um computador que esteja executando o *software*. O *software* é autorizado a gerenciar e controlar o rastreador visual e considerar dados de rastreamento ocular registrados em tempo real (OTHMAN; ROMLI, 2016).

Figura 13 – Sistema de rastreamento ocular



Fonte: Othman e Romli (2016).

Os dados digitalizados, armazenados em um disco de computador e passados para um programa de *software*, podem ser corrigidos e submetidos a uma transformação rápida de Fourier (FFT) (WILSON; CALDWELL; RUSSELL, 2007).

Os dados visuais também podem ser coletados por meio de uma câmera de vídeo digital situada à frente do rosto do trabalhador. Assim, enquanto o participante realiza suas tarefas, filmagens do seu rosto são gravadas, as quais podem ser analisados manualmente com o uso de um programa de computador para sequenciar o tempo dos movimentos oculares (DEY; MANN, 2010).

Por meio do EOG, pode-se medir a taxa de piscar dos olhos ou o número de piscadas, as durações de intermitência, a frequência e distância sacádica, a duração da fixação, o diâmetro da pupila, a análise de padrões visuais, a entropia e o indicador estatístico índice de vizinho mais próximo (NNI) (AHLSTROM; FRIEDMAN-BERG, 2006; BERGUER; SMITH; CHUNG, 2001; DI NOCERA; CAMILLI; TERENCEZI, 2007; WILSON, 2002; WILSON; RUSSELL, 2003).

Utilizando os dados coletados que apresentam significância estatística, pode-se desenvolver modelos de regressão linear múltipla individual, buscando relacionar a CTM a essas medidas (AHLSTROM; FRIEDMAN-BERG, 2006).

Com o diâmetro da pupila podem ser realizadas medidas da média, desvio padrão e energia total do diâmetro da pupila, a qual utiliza uma transformada de Fourier. Com essas medidas, pode-se calcular a razão do diâmetro da pupila (PDR), o índice de variabilidade da pupila (PVR) e a relação harmônica da pupila (PHR), as

quais utilizam medidas dos valores acima citados, respectivamente, divididos por valores de referência (MANNARU et al., 2016b).

Análises estatísticas podem ser utilizadas para determinar diferenças estatísticas entre os dados. Os destaques são as análises de variância, teste U de Mann-Whitney, teste t e correlação de Pearson (BERGUER; SMITH; CHUNG, 2001; DI NOCERA; CAMILLI; TERENCEZI, 2007; KATAOKA; SASAKI; KANDA, 2011; WILSON; CALDWELL; RUSSELL, 2007; ZHENG et al., 2012a).

Como resultados de estudos, encontra-se que a taxa de intermitência dos olhos diminui com o aumento do nível de concentração mental, mas também aumenta com o nível de demanda de tarefas (BERGUER; SMITH; CHUNG, 2001).

Constatou-se que o aumento da carga de trabalho ou aumento na dificuldade da tarefa corresponde a uma redução linear na média de duração das piscadas e na média de distância sacádica, além de um aumento linear no diâmetro médio da pupila e no conteúdo energético (AHLSTROM; FRIEDMAN-BERG, 2006; MANNARU et al., 2016b; WILSON; CALDWELL; RUSSELL, 2007).

Em relação ao NNI, aplicado às fixações oculares e sensível à variação na carga mental, encontra-se que valores mais elevados de NNI são associados a fases de alta carga de trabalho (DI NOCERA; CAMILLI; TERENCEZI, 2007).

O tempo de permanência e a duração da fixação diminuem com o aumento das demandas de tarefas, pois a duração mais longa de fixação está associada à menor carga de trabalho mental (DE RIVECOURT et al., 2008; DEY; MANN, 2010; SCHULZ et al., 2011).

Para Schulz et al. (2011) o diâmetro médio ponderado da pupila não mostra uma diferença significativa entre sessões, sugerindo que o diâmetro da pupila não pode servir para comparação geral da carga de trabalho entre sessões ou indivíduos. Para os mesmos autores, a amplitude da sacada também não reflete as flutuações da carga de trabalho (SCHULZ et al., 2011).

Em relação aos trabalhadores estudados, foram encontrados pilotos, controladores de tráfego aéreo, operadores de sistemas aéreos não tripulados, operadores agrícolas e trabalhadores de saúde, como anestesiologistas e cirurgiões (DEY; MANN, 2010; SCHULZ et al., 2011; WILSON, 2002; WILSON; RUSSELL, 2003; WILSON; CALDWELL; RUSSELL, 2007; ZHENG et al., 2012a).

4.2.1.3 Cérebro

O cérebro é uma parte do corpo humano muito investigada na busca de medidas fisiológicas para determinar a carga de trabalho mental. Isso é feito normalmente com base nos dados do eletroencefalograma (EEG), retirados por meio de eletrodos, que podem ser de estanho ou banhados a ouro e reutilizáveis, os quais podem ser colocados sobre o couro cabeludo e com número variado de canais. Podem ser utilizados eletrodos nos mastoides (orelhas) como referência e como terra. Cada local de colocação de EEG e mastoide deve ser limpo com acetona e os eletrodos podem ser fixados com colódio e, em seguida, preenchidos com gel eletrolítico (LING et al., 2001; WILSON, 2002; WILSON; CALDWELL; RUSSELL, 2007). Os eletrodos podem ser colocados em locais baseados na montagem internacional 10-20, com seis locais de eletrodos (FZ, PZ, O1, O2, P3 e P4) com uma referência de lóbulo eletronicamente conectada (FALLAHI et al., 2016; ZAWIAH; DAWAL, 2016). Os dados do EEG também podem ser obtidos usando uma tampa especial com eletrodos rotulados usando o mesmo sistema internacional 10-20 (EAST; BAUER JR; LANNING, 2002).

Modelos de redes neurais podem ser desenvolvidos para analisar os dados do eletroencefalograma e com os resultados obtidos pode-se classificar o nível de carga mental da tarefa. Esses dados podem ser captados como uma transformada rápida de Fourier, os quais geram gráficos após a leitura feita por um *software*. Esses dados podem ser segmentados em bandas de frequência, buscando a banda de frequência que desempenha um papel vital na análise do EEG. Essas bandas podem ser cinco: delta (1-3 Hz), teta (4-7 Hz), alfa (8-12 Hz), beta (13-30 Hz) e ultrabeta (31 a 42 Hz). Com base nessas bandas e nos canais de retirada dos dados (os eletrodos), são escolhidas as características, como entradas a serem usadas na rede neural, podendo utilizar uma função de energia para escolher os melhores dados. Um modelo de rede neural que pode ser utilizada é uma rede multicamadas (MLN), a qual opera com um algoritmo (LING et al., 2001).

Caso não sejam usadas redes neurais, pode-se analisar os dados de EEG utilizando as mesmas cinco bandas para analisar a potência, utilizando correção dos dados e suavização, que pode ser feita por meio de operador laplaciano ou gaussiano, utilizando *software*. Para extração das frequências, pode-se utilizar uma transformada rápida de Fourier. Pode-se utilizar ainda sistemas Fuzzy para gerar uma função que

representa a carga de trabalho mental com base nos dados coletados (MOON et al., 2002; WILSON, 2002).

Os dados também podem ser passados por meio de uma transformada rápida de Fourier, realizada em cada sinal EEG para cada segundo de dados brutos. A FFT faz a leitura das potências das cinco bandas de frequência, podendo também apresentar um resultado da soma dessas potências. Com esses dados, e também com outros dados retirados de outros indicadores fisiológicos, como coração e olhos, é possível elaborar um modelo para prever a CTM ou para comparar os resultados com outros métodos que preveem as condições de carga mental (EAST; BAUER JR; LANNING, 2002; WILSON; RUSSELL, 2003).

Pode ser utilizada análise de perturbação espectral relacionada ao evento (ERSP), por meio das bandas de frequências teta (4-7 Hz) e alfa II (11-13 Hz) (KERICK; ALLENDER, 2006). Podem ser calculadas a média e o desvio padrão dos dados (WILSON; CALDWELL; RUSSELL, 2007).

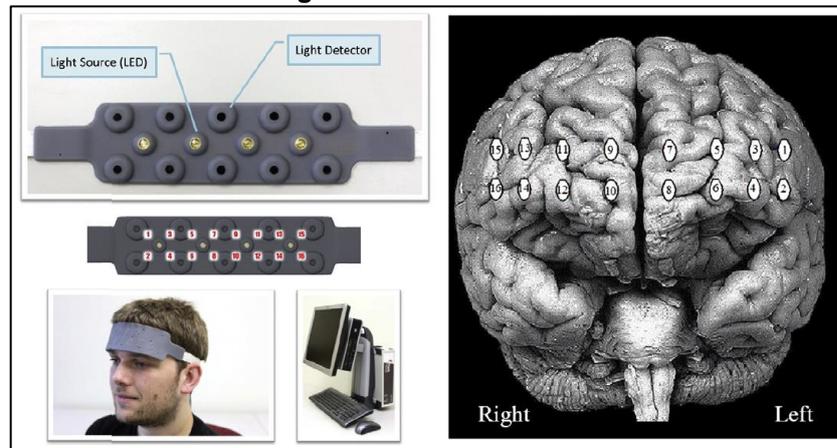
Os dados do EEG podem ser utilizados para o cálculo dos ERPs (dados eletrocardiográficos e potenciais relacionados a eventos). A amplitude do componente P300 ERP é calculada a partir do canal Pz do EEG e é utilizada para verificar a CTM nos diferentes blocos de tarefas. Os ERPs são calculados a partir de tons alvo corretamente reconhecidos (HENELIUS et al., 2009).

Uma outra técnica não invasiva utilizada para monitorar alterações no córtex é a espectroscopia funcional de infravermelho próximo (fNIR). Ela monitora as alterações de concentração de hemoglobina oxigenada (oxi-Hb) e hemoglobina desoxigenada (deoxi-Hb), ou seja, monitora as alterações hemodinâmicas relacionadas à atividade cognitiva, por meio de imagem cerebral óptica, utilizando luz infravermelha próxima. A tecnologia fNIR permite o design de sistemas de monitoramento seguros e acessíveis, permitindo que os dispositivos sejam miniaturizados, portáteis e operados por bateria, tornando assim possível o estudo em campo. Para isso, o córtex pré-frontal pode ser monitorado usando um sistema de onda contínua fNIR, o qual é composto de três módulos: uma antena flexível, que contém fontes de luz e detectores para permitir uma rápida colocação de todos os sensores, uma caixa de controle para gerenciamento de *hardware* e um computador que executa a aquisição de dados, conforme Figura 14.

Para aquisição e visualização de dados utiliza-se um *software*, os quais são dados brutos da fNIR e precisam ser filtrados para atenuar o ruído de alta frequência.

As alterações da oxigenação do sangue no córtex pré-frontal podem ser calculadas usando a Lei Modificada de Beer Lambert (MBLL) para períodos de tarefa com respeito a períodos de descanso no início de cada tarefa. Um modelo de regressão polinomial de primeira ordem separado pode ser estabelecido para cada participante e com isso os parâmetros do modelo são estimados usando dados mínimo e máximos dos respectivos dados, buscando mapear a oxigenação (AYAZ et al., 2011, 2012; HARRISON et al., 2014).

Figura 14 – Sistema fNIR



Fonte: Ayaz et al. (2012).

A análise de variância, comparações pareadas, testes t, correção de Geisser-Greenhouse com os testes post hoc de Tukey, análise de variância bayesiana em parcelas subdivididas (BANOVA) e teste de Friedman podem ser utilizadas para avaliar estatisticamente os dados (AYAZ et al., 2011; HARRISON et al., 2014; WILSON; CALDWELL; RUSSELL, 2007; ZAWIAH; DAWAL, 2016).

Estudos mostraram que estressores podem levar a mudanças significativas nas bandas teta e alfa do EEG (FALLAHI et al., 2016). Além disso, para Zawiah e Dawal (2016), uma redução nas ondas beta podem ser um sinal claro de cansaço e sonolência entre as pessoas.

Em relação aos segmentos com maior demanda, são observados aumentos topográficos da diminuição da potência da banda alfa ou aumento da potência da banda delta quando comparado com aqueles encontrados durante os níveis mais baixos de carga de trabalho (WILSON, 2002). Já Moon et al. (2002) observaram que as potências alfa e teta aumentam quando se espera que a carga

mental aumente. Ainda, aumentos no EEG de ondas lentas podem ser associados à diminuição do estado de alerta (WILSON; CALDWELL; RUSSELL, 2007).

Para os dados do ERSP, foi identificado que o pico de potência teta difere em função do estresse no tempo e o pico de potência alfa difere em função da carga da tarefa e da carga de decisão. No geral, os resultados mostram que o estresse no tempo tem os efeitos mais profundos e generalizados sobre a percepção e o desempenho da carga de trabalho (KERICK; ALLENDER, 2006).

Quanto aos dados do ERP, Henelius et al. (2009) encontraram diminuição da amplitude do P300 com o número crescente de tarefas simultâneas.

Quando analisadas as mudanças na oxigenação do sangue no córtex pré-frontal dorsolateral, medidas pela fNIR, foi encontrada associação com o aumento da carga cognitiva (AYAZ et al., 2011). Além disso, o nível de especialização parece influenciar a resposta hemodinâmica no córtex pré-frontal dorsolateral, pelo menos para algumas tarefas complexas (AYAZ et al., 2012).

No que se refere aos trabalhadores estudados, foram encontrados pilotos, controladores de tráfego aéreo, operadores de turbina de usinas nucleares, fuzileiros, operadores de veículos aéreos não tripulados, trabalhadores de saúde, maquinistas e controladores de trânsito (FALLAHI et al., 2016; HENELIUS et al., 2009; KERICK; ALLENDER, 2006; MOON et al., 2002; WILSON, 2002; WILSON; RUSSELL, 2003; WILSON; CALDWELL; RUSSELL, 2007; ZAWIAH; DAWAL, 2016).

4.2.1.4 Pele

Quando a pele é utilizada como medida fisiológica para avaliar a CTM, pode-se analisar o nível de condutância cutânea tônica (SC). Para isso, utiliza-se um par de eletrodos de prata/cloreto de prata autoadesivos, os quais podem ser colocados próximos ao nervo ulnar da palma da mão ou nos dedos do indivíduo após limpeza, a qual pode ser feita com água destilada (BERGUER; SMITH; CHUNG, 2001; GOULD et al., 2009). Em relação a SC, espera-se um aumento estatisticamente significativo no nível de condutância de pele com o aumento do nível de dificuldade da tarefa (BERGUER; SMITH; CHUNG, 2001).

Para avaliar a resistência da pele pode-se definir um índice, porque a amplitude da resposta depende de um valor pré-estimado. O tempo durante o qual o

participante “responde” aos estímulos sem se referir ao valor inicial (ou nível tônico) define a duração da perturbação ôhmica e revela o processamento da informação pelos participantes (COLLET et al., 2013). Segundo os resultados de Collet et al. (2013), quanto maior a carga de trabalho, maior a duração de perturbação ôhmica.

Pode-se utilizar também a atividade eletrodérmica (EDA), registrada a partir de eletrodos reutilizáveis de Ag/AgCl colocados no arco do pé, os quais podem ser aplicados sem limpeza da pele. Podem ser analisados o número de respostas EDA, amplitudes, tempos de subida, tempos de recuperação e níveis tónicos, por meio de *software* (WILSON, 2002).

Buscando descobrir se os movimentos das pernas influenciam as respostas da EDA do pé, pode-se registrar dados eletromiográficos (EMG) da panturrilha da perna, os quais são analisados espectralmente e os níveis de potência são submetidos às mesmas análises estatísticas dos dados da EDA. Segundo os resultados de Wilson (2002), os dados EMG não possuem relação com as mudanças na EDA (WILSON, 2002).

As diferenças estatísticas entre os dados podem ser avaliadas por meio do teste ANOVA e do teste U de Mann-Whitney (BERGUER; SMITH; CHUNG, 2001; GOULD et al., 2009).

Como trabalhadores avaliados por medidas da pele, foram encontrados estudos com profissionais da saúde, pilotos, motoristas e cadetes marítimos (BERGUER; SMITH; CHUNG, 2001; COLLET et al., 2003; GOULD et al., 2009; WILSON, 2002).

4.2.1.5 Respiração

Medidas da respiração também são utilizadas para avaliar os diferentes níveis de esforço mental. Com esse propósito, pode-se investigar a taxa de respiração (RR). As ondas de respiração podem ser gravadas continuamente com um pequeno dispositivo de gravação preso ao peito, o qual inclui um gravador, alça e uma placa de armazenamento de registros onde o gravador pode ser instalado. Esse dispositivo de gravação possui eletrodos, para captação dos dados. Após ser amarrado ao peito, acontece a gravação automática das ondas de respiração, que podem ser medidas

por pletismografia indutiva. Depois das gravações, os dados são transferidos para um computador para análises (YAO et al., 2008).

O RR pode ser calculado automaticamente a partir dos intervalos de pico das ondas respiratórias. Os dados são apresentados como médias \pm desvio padrão. ANOVA pode ser utilizada para testar as diferenças entre os grupos (YAO et al., 2008).

Com base nos resultados de YAO et al. (2008), buscando analisar as respostas de pilotos conforme as diferentes fases de voo, os pilotos menos experientes podem possuir ocasionalmente RR mais elevados do que os especialistas em voo, porém durante as fases de voo mais carregadas, como a aproximação final e a aterrissagem manual, não foram encontradas diferenças significativas na RR (YAO et al., 2008).

4.2.1.6 Temperatura facial

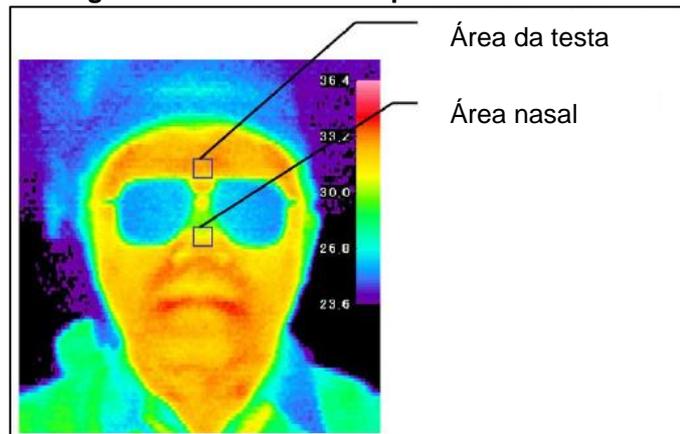
A temperatura facial também é um indicador fisiológico que pode ser utilizado para avaliar a carga de trabalho mental. Para isso, pode-se medir a temperatura nasal. Um testador mede a temperatura facial incluindo a temperatura nasal por termografia. Além disso, ele registra a situação experimental, o comportamento do sujeito, a conversação e o intervalo R-R. Os dados do intervalo R-R são os dados de referência para confirmar a característica da temperatura nasal (MURAI et al., 2007, 2008).

Os registros da temperatura são feitos por meio de um aparelho de termografia que fica à frente do sujeito, o qual está em pé realizando as suas tarefas normalmente. O cálculo da temperatura utiliza a diferença entre a temperatura nasal e a temperatura da testa para avaliar a CTM do operador, conforme mostrado na Figura 15. São calculados os valores médio, máximo e mínimo da medição nasal e frontal, sendo o valor médio utilizado como índice (MURAI et al., 2007).

A relação entre a temperatura nasal (T_n) e temperatura da testa (T_f) com a carga mental é a seguinte:

- Caso 1) Temperatura $T_n - T_f < 0$: emoções ruins (tensão/estresse);
- Caso 2) Temperatura $T_n - T_f = 0$: emoções normais;
- Caso 3) Temperatura $T_n - T_f > 0$: boas emoções (sem estresse) (MURAI et al., 2007, 2008).

Figura 15 – Local das temperaturas utilizadas



Fonte: Murai et al. (2007).

Segundos os estudos de Murai et al. (2007, 2008), realizados em navegadores, a temperatura nasal diminui quando o navegador inicia sua carga de trabalho mental, ou seja, quando são tomadas de decisões para manuseios do navio e julgamentos para evitar colisões.

4.2.1.7 Fala

Muitos parâmetros acústicos como frequência fundamental (F0), intensidade, ataque de energia e gradiente de decaimento, jitter, microtremor fisiológico, duração de pausas e segmentos de fala, velocidade de fala, disfluências de fala, localização de formantes e inclinação espectral já foram estudados como carga vocal ou características vocais induzidas por estresse psicológico ou carga de trabalho mental. Dentre esses indicadores, a F0 e a intensidade da voz têm sido os parâmetros de fala mais frequentemente associados à carga cognitiva (HUTTUNEN et al., 2011).

As gravações da fala são obtidas de forma não invasiva, sendo assim viáveis para monitorar a carga cognitiva. A fala pode apresentar mudanças devido a diferentes demandas de CTM, as quais podem ser vistas na prosódia (ritmo, estresse e entonação da fala), visto que ela transmite muitas características importantes das mensagens faladas. Dessa forma, torna-se importante reconhecer mudanças na voz causadas por mudanças na carga de trabalho, sendo muito importantes na questão de segurança, como na aviação (HUTTUNEN et al., 2011).

As medições sonoras podem ser feitas por meio de um dosímetro de ruído, o qual deve ser colocado no local em que se deseja gravar a fala, buscando captar os ruídos de fundo. Para a captação da fala, podem ser utilizados microfones com pré-amplificadores, os quais possuem uma resposta de frequência plana por meio das frequências típicas da fala humana e são colocados próximos da boca dos trabalhadores, presos em um capacete. Após as gravações, os dados são transferidos para um computador e subsequentemente são filtrados para eliminar quaisquer artefatos de baixa frequência causados por interferências da rede elétrica (HUTTUNEN et al., 2011).

Analisando-se a F0 e a intensidade da voz, emprega-se um algoritmo de extração da voz no domínio do tempo, utilizando-se a correspondência de formas de onda, para análises de forma automática. Os dados das medidas da fala são fundidos com os dados de ruído de fundo e carga cognitiva com base no tempo sincronizado entre as diferentes fontes de dados. No estudo de Huttunen et al. (2001), que envolveu pilotos, tanto a média da F0 quanto a intensidade vocal média dos pilotos aumentaram quando as cargas cognitivas aumentaram.

4.2.1.8 Movimentos corporais

Os movimentos corporais podem ser medidas na busca de relações com mudanças de outras variáveis fisiológicas, como por exemplo na frequência cardíaca, visto que a atividade física aumenta necessariamente a atividade cardíaca. Dessa forma, pode-se utilizar um medidor no pulso para registrar os movimentos do sujeito. Assim, se as alterações na frequência cardíaca são causadas pela atividade do corpo, espera-se que os movimentos do pulso estejam altamente correlacionados com a frequência cardíaca, a qual pode ser relacionada com a carga de trabalho mental (WILSON, 2002).

4.2.1.9 Urina

A urina também pode ser utilizada para verificar a carga de trabalho mental, principalmente por meio da taxa de excreção urinária de adrenalina. Para isso, amostras de urina são recolhidas antes, durante e após a realização das atividades,

conforme critérios estabelecidos, em frascos contendo ácido cítrico, para garantir a vida útil da amostra. Após coleta, os frascos devem ser mantidos o mais frio possível até o início da preparação (KUIJER et al., 2004).

As concentrações urinárias (nanogramas/mililitro) são multiplicadas pelo volume da amostra de urina correspondendo (mililitros). Essa quantidade (nanogramas) é dividida pelo período de tempo (minutos) entre a amostra analisada e a amostra anterior, obtendo-se assim a taxa de excreção urinária média para esse período (nanogramas/minuto). Para testar as diferenças na CTM podem ser utilizadas análise de variância multivariada (MANOVA) e teste t (KUIJER et al., 2004).

Kuijer et al. (2004) buscaram analisar as diferenças na CTM analisando a tarefa de somente coletar lixo, somente dirigir o caminhão de lixo e alternar entre essas duas tarefas durante um dia de trabalho. Por meio das análises da urina eles descobriram que a taxa de excreção de adrenalina durante apenas a condução foi maior do que durante a coleta apenas e rotação em geral e a taxa de excreção de noradrenalina foi menor durante a condução apenas quando comparada com a coleta apenas e a rotação em geral. Dessa forma, descobriu-se que no caso estudado a rotação de trabalho resultou em diminuição da carga de trabalho mental, em termos da taxa de excreção de adrenalina, em comparação com a condução apenas. Rotação de trabalho não aumentou a carga de trabalho mental em comparação com coleta de lixo apenas (KUIJER et al., 2004).

4.2.1.10 Visão dos autores sobre as medidas fisiológicas

São muitas as vantagens encontradas na utilização de medidas fisiológicas para avaliar a carga de trabalho mental. Para Wilson e Russell (2003), os sinais psicofisiológicos são atraentes por estarem continuamente disponíveis e sua coleta não interferir no desempenho do trabalho dos indivíduos analisados (WILSON; RUSSELL, 2003).

Wilson (2002) e Lehrer et al. (2010) acreditam que a frequência cardíaca, o piscar dos olhos e dados encefalográficos fornecem informações adicionais que vão além das encontradas em medidas subjetivas, permitindo o exame muito mais significativo dos efeitos, visto que os dados fisiológicos melhoram a detecção de tarefas, incluindo diferenciação sensível entre cargas de tarefas moderadas e altas.

Em relação as medidas do coração, os dados de HR e HRV são muito úteis na avaliação de voos altamente exigentes (LAHTINEN et al., 2007). Além disso, Lehrer et al. (2010) acreditam que quando comparados com dados dos olhos e do encéfalo, os dados do eletrocardiograma são mais facilmente usáveis e analisáveis, podendo utilizar dispositivos comercialmente disponíveis, além de possibilitarem a detecção de alta carga, particularmente onde a diferenciação pode não ser intuitivamente óbvia, inclusive não sendo possível medir pelo NASA-TLX.

Henelius et al. (2009), Lehrer et al. (2010) e Mansikka et al. (2016a) colocam o monitoramento por ECG como um método relativamente barato e simples para avaliar a HR e a HRV, com alta validade e que não gera intrusos, além de não ser invasivo.

Em relação aos olhos, um potencial benefício do uso da atividade dos movimentos oculares como um correlato da carga de trabalho cognitiva é que ela fornece a possibilidade de capturar flutuações na carga de trabalho que ocorrem em intervalos de tempo curtos, ou seja, podem fornecer medidas de carga de trabalho mais contínuas, momento a momento, as quais não são refletidas nas classificações subjetivas de carga de trabalho (AHLSTROM; FRIEDMAN-BERG, 2006; ZHENG et al., 2012a). Além disso, os movimentos oculares são relativamente fáceis de coletar em ambientes operacionais reais, a tecnologia recente de rastreamento ocular não é intrusiva e os dados são relativamente fáceis de analisar (DI NOCERA; CAMILLI; TERENCEZI, 2007).

Ao mesmo tempo, os movimentos oculares, diferentemente de potenciais relacionados a eventos e a variabilidade da frequência cardíaca, são insensíveis aos movimentos dos membros, não necessitam muito treinamento para montar o equipamento e o procedimento de calibração pode ser realizado em pouco tempo.

O NNI, por exemplo, computado das fixações oculares, é sensível às variações na carga mental, evidenciando-se como uma medida em tempo real da carga mental e como um gatilho para sistemas automatizados. Outro benefício adicional desse índice é que ele não precisa necessariamente de extrema precisão e alta resolução temporal, fornecendo uma medida independente do domínio que pode eventualmente ser usada em ambientes operacionais para coletar informações em tempo real sobre a carga do operador, tornando-se muito interessante no domínio da aviação (DI NOCERA; CAMILLI; TERENCEZI, 2007).

De forma adicional, com as medidas de movimento ocular em tempo real, também é possível identificar o local ao qual o trabalhador está olhando durante os acontecimentos (AHLSTROM; FRIEDMAN-BERG, 2006).

Medidas objetivas baseadas no cérebro podem ajudar a prevenir erros do operador e permitir a intervenção oportuna por meio da previsão de um declínio no desempenho que pode surgir tanto da sobrecarga de trabalho quanto da subcarga. O uso do EEG e dos potenciais cerebrais relacionados a eventos (ERPs) têm se mostrado fortes candidatos a medidas objetivas precisas da carga cognitiva do operador devido à sua capacidade de fornecer medidas diretas da atividade do sistema nervoso central. O EEG possui muitos atributos superiores para avaliação da carga de trabalho mental, incluindo excelente resolução temporal e uso fora de ambientes laboratoriais, devido à compactação da tecnologia associada (FALLAHI et al., 2016; HARRISON et al., 2014).

Já a espectroscopia funcional de infravermelho próximo vem ganhando reconhecimento por sua diversidade de aplicações em imagens cerebrais. Essa tecnologia utiliza imagem cerebral não invasiva implantável em campo, por possuir dispositivos miniaturizados, que mede a hemodinâmica cerebral em resposta a tarefas sensoriais, motoras ou cognitivas. Quando comparado com o EEG, o fNIR proporciona maior localização espacial e é facilmente integrado com EEG/ERPs para uma análise mais robusta (AYAZ et al., 2011; HARRISON et al., 2014).

Além das vantagens apresentadas por medidas do coração, dos olhos e do cérebro, as gravações da atividade eletrodérmica mostraram-se ferramentas confiáveis na avaliação da carga mental em campo (COLLET et al., 2003).

Apesar de todas essas vantagens, as medidas fisiológicas também possuem desvantagens. Uma delas é que muitas vezes é necessário utilizar outras formas de avaliação, seja por meio de medidas subjetivas, seja por meio de medidas de desempenho, buscando relacionar os resultados da CTM das medições fisiológicas por meio dessas outras medidas (BERGUER; SMITH; CHUNG, 2001; GOULD et al., 2009; LEHRER et al., 2010; MARTIN et al., 2016).

Outra limitação é que nunca é possível saber se as mudanças fisiológicas, por exemplo na frequência cardíaca, são derivadas da carga cognitiva, das exigências físicas ou da excitação emocional, porque o desenho experimental não controla isso. Assim, a experiência subjetiva e a avaliação podem ser a reflexão mais precisa da

carga de trabalho. Dessa forma, uma avaliação subjetiva seria útil para combinar com as análises das mudanças fisiológicas buscando avaliar a CTM (YAO et al., 2008).

Porém, quando são utilizadas outras medidas para comparar os resultados de CTM, quando já se sabe previamente qual atividade exige um maior conteúdo mental do trabalhador, as medidas fisiológicas podem apresentar resultados opostos, o que muitas vezes pode ser explicado pelo fato que tarefas que podem exigir menor carga mental são classificadas pelos indivíduos como mais difíceis, por não serem rotineiras (WILSON, 2002).

Além de algumas medidas fisiológicas possuírem um alto custo com equipamentos e até mesmo com análises, muitas vezes são recomendadas, e até mesmo necessárias, o uso de medidas de múltiplos indicadores, buscando maior confiança e significância de resultados (COLLET et al., 2003; OTHMAN; ROMLI, 2016).

Outra desvantagem que pode ser vista na utilização das medidas fisiológicas é que a experiência do trabalhador pode influenciar os resultados e também o efeito da fadiga, apresentando uma variação nas medidas fisiológicas que pode vir a ser interpretada como um aumento errôneo na carga de trabalho mental, efeito esse que não é necessariamente real (DE RIVECOURT et al., 2008). Do mesmo modo, as medidas fisiológicas podem falhar em explicar completamente a relação de tarefas mentais, carga de trabalho mental e desempenho, já que existem outros construtos que podem influenciar a CTM, sendo eles fatores constitucionais, ambientais e particulares de cada indivíduo (LEHRER et al., 2010; MANSIKKA et al., 2016a).

Alguns desafios também são encontrados: a metodologia de coleta de dados pode ser um impedimento em certas atividades, pois pode afetar a segurança e o desempenho profissional, devendo ser buscados dispositivos pequenos, fáceis de usar, não intrusivos e até mesmo que possam ser manuseados pelos próprios participantes da pesquisa; as condições utilizadas em alguns estudos são artificiais, o que acaba mascarando resultados reais no ambiente de trabalho e também muitas vezes as amostras são limitadas, não sendo possível generalizar os resultados e nem alcançar significância estatística (DEY; MANN, 2010; GOULD et al., 2009; HANNULA et al., 2008, 2009; KATAOKA; SASAKI; KANDA, 2011; LEHRER et al., 2010; LIANG et al., 2009; YAO et al., 2008).

Em relação aos dispositivos fisiológicos populares são encontradas algumas desvantagens, como a interferência nos movimentos naturais do corpo, a falta de conforto, a impraticabilidade dos dispositivos após um longo período devido ao desgaste e a interferência de fluidos corporais, como o suor (FALLAHI et al., 2016).

Outro fator desfavorável é que dependendo dos indicadores, da situação analisada e do nível de precisão almejado, são necessários longos intervalos de gravações das medidas fisiológicas, *softwares* sofisticados e grande número de iterações para alcançar os modelos estatísticos desejáveis, fato que pode vir a encarecer as análises e aumentar a duração das pesquisas (WILSON; RUSSELL, 2003).

Quando são utilizadas medidas dos olhos, alguns problemas podem ser encontrados: a luminância do local pode afetar as medidas; são recomendados registros de dados antes da realização da tarefa, para possuir dados de referência e comparação; outras variáveis podem influenciar os resultados, como movimentos manuais e coleta de informações; alocação espacial de informações; dependendo da tarefa estudada o comportamento ocular é diferente, não sendo possível generalizar as medições para diferentes aplicações e também essa técnica pode superestimar a proporção do tempo que são visualizados os diferentes locais do campo externo, visto que nem sempre o operador está buscando algo que envolva a tarefa (DE RIVECOURT et al., 2008; DEY; MANN, 2010; MANNARU et al., 2016a; OTHMAN; ROMLI, 2016; SCHULZ et al., 2011; ZHENG et al., 2012a).

Já nas análises de EEG e componentes dos potenciais relacionados a eventos, desvantagens encontradas referem-se às dificuldades para coletar e analisar esses índices em tempo real, e, nas análises da fala, ruídos do ambiente devem ser estimados e considerados para as análises da CTM, visto que o ruído pode causar o efeito Lombard, o qual é uma tendência involuntária de um falante aumentar a intensidade vocal em meio a existência de um ruído de fundo (DI NOCERA; CAMILLI; TERENCEZI, 2007; HUTTUNEN et al., 2011).

4.2.2 National Aeronautics and Space Administration Task Load Index (NASA-TLX)

4.2.2.1 História e evolução

O estudo com medidas subjetivas foi iniciado por Cooper e Harper, em 1969, por meio da Escala de Cooper Harper, a qual é uma escala unidimensional que foi elaborada para mensurar a carga mental proveniente de tarefas de voo. A partir dela, duas vertentes surgiram, as quais deram origem aos métodos hoje conhecidos como NASA-TLX e SWAT (CARDOSO, 2010).

O Índice de Carga de Tarefa NASA foi desenvolvido pela *Human Performance Group*, no Centro de Pesquisas Ames da NASA, e descrito por Hart e Staveland (1988), sendo esse um procedimento de taxa multidimensional que provê uma pontuação global da carga de trabalho baseado em uma média ponderada de seis escalas avaliadas: demanda mental, demanda física, demanda temporal, desempenho próprio, esforço e frustração (CARDOSO, 2010).

Hart e Staveland (1988) acreditavam que a carga de trabalho mental não é uma propriedade inerente, mas depende dos requisitos da tarefa, das circunstâncias do momento da realização da tarefa e das habilidades e percepções do trabalhador. Diante disso, esses autores propuseram um *framework* conceitual com base em diferentes fontes modificadoras de carga de trabalho, as quais foram enumeradas e relacionadas. Esse *framework*, chamado de Escala de Avaliação Bipolar da NASA, possuía como escalas: dificuldade da tarefa, pressão temporal, tipo de atividade, esforço físico, esforço mental, desempenho próprio, frustração, estresse e fadiga.

Por meio do *framework* foram analisados os dados dentro de cada experimento para determinar a sensibilidade das escalas individuais, classificações gerais da carga de trabalho e pontuações ponderadas da carga de trabalho. Em seguida, os dados de experimentos semelhantes foram mesclados em seis categorias. Análises correlacionais e de regressão foram realizadas com base nesses dados, bem como em toda a base de dados, para determinar (1) a associação estatística entre as classificações e (2) o grau em que essas escalas, tomadas como um grupo, prediziam as classificações gerais da carga de trabalho. Os resultados dessas análises foram então usados para selecionar um conjunto limitado de escalas

e o procedimento de ponderação para uma nova técnica multidimensional de classificação da carga de trabalho (HART; STAVELAND, 1988).

Por meio dessas análises foi proposto o NASA-TLX, que consiste em seis escalas de componentes. Três dimensões dizem respeito às demandas impostas ao sujeito (demandas mental, física e temporal) e três à interação do sujeito com a tarefa (esforço, frustração e performance). Uma média ponderada dessas seis escalas, para refletir a contribuição de cada fator para a carga de trabalho de uma atividade específica, foi proposta como uma medida integrada da carga de trabalho global (HART, 1986).

Cada uma das seis escalas ou dimensões está descrita no Quadro 7.

Quadro 7 – Definição das seis dimensões do NASA-TLX

Dimensão	Ponto inicial/final	Descrição
Demanda mental	Baixa/Alta	Quanta atividade mental e perceptiva foi necessária (por exemplo, pensando, decidindo, calculando, lembrando, olhando, pesquisando, etc.)? A tarefa era fácil ou exigente, simples ou complexa, tolerante ou exigente?
Demanda física	Baixa/Alta	Quanta atividade física foi necessária (por exemplo, empurrando, puxando, girando, controlando, ativando, etc.)? A tarefa era fácil ou exigente, lenta ou rápida, folgada ou energética, tranquila ou fatigante?
Demanda temporal	Baixa/Alta	Quanta pressão de tempo você sentiu devido à taxa ou ao ritmo em que as tarefas ou elementos da tarefa ocorreram? O ritmo era lento e vagaroso ou rápido e frenético?
Performance/ desempenho	Baixo/Alto	Quão bem-sucedido você acredita que foi em realizar os objetivos da tarefa definida pelo experimentador (ou por você mesmo)? Quão satisfeito você está com o seu desempenho no cumprimento desses objetivos?
Esforço	Baixo/Alto	Quão duro você teve que trabalhar (mentalmente e fisicamente) para atingir seu nível de desempenho?
Nível de frustração	Baixo/Alto	Quão inseguro, desanimado, aborrecido, estressado e irritado vs. seguro, gratificado, satisfeito, relaxado e sereno você se sentiu durante a tarefa?

Fonte: Adaptado de Hart (1986).

4.2.2.2 Análise dos dados e interpretação dos resultados

A estimativa da carga de trabalho por meio do NASA-TLX depende da avaliação de cada uma das seis escalas, combinadas com o peso de cada uma delas, os quais variam conforme as tarefas realizadas e os sujeitos (HART, 1986).

A primeira parte da avaliação é feita por meio da definição dos pesos das seis dimensões, os quais mostram a contribuição das escalas para a definição da carga de trabalho para a tarefa analisada e foi descrito por Hart (1986). Para isso, as seis dimensões são colocadas aos pares, gerando quinze combinações diferentes. Os trabalhadores circulam o membro de cada par que eles acreditam contribuir mais para a carga de trabalho naquela tarefa. Com isso, cada fator assume um valor, ou ponderação, de 0 (não é relevante) a 5 (mais importante que qualquer outro fator), o qual é equivalente ao número de vezes em que o fator foi circulado. Um conjunto de pesos é obtido após cada tarefa. O mesmo conjunto de pesos pode ser utilizado para atividades diferentes, ou variações da tarefa, desde que as contribuições dos seis fatores para sua carga de trabalho sejam muito semelhantes (HART, 1986).

Além da ponderação é necessária avaliar a magnitude de cada um dos fatores. Para isso, os trabalhadores devem marcar a escala na posição desejada, a qual varia de 0 a 100. Cada escala é representada por uma linha de 12 cm de comprimento, a qual é dividida em 20 partes iguais, possuindo descrições bipolares em seus extremos, conforme mostrado no Quadro 7, e marcados de 5 em 5 pontos. Caso o indivíduo marque entre duas marcas, o valor considerado é o valor correspondente ao marco da direita. Essas classificações podem ser feitas durante uma tarefa, depois de segmentos da tarefa ou após a realização completa da tarefa (HART, 1986).

Por fim, para pontuar a carga de trabalho global de cada sujeito, a magnitude de cada fator é multiplicada pelo peso, ambos dados pelo próprio trabalhador. A soma dos resultados dessas multiplicações é então dividida por 15 e chega-se ao valor da carga de trabalho global individual do sujeito (HART, 1986).

Na escala TLX, o bom desempenho está associado a um número baixo, já que uma carga de trabalho menor geralmente é acompanhada por um melhor desempenho (HART; STAVELAND, 1988).

4.2.2.3 Procedimento experimental

Hart (1986) divide em cinco os eventos para aplicação do NASA-TLX:

- i. Instruções: são explicadas as escalas e as instruções necessárias para preenchimento do NASA-TLX aos participantes;

ii. Familiarização: para garantir que foi entendido o procedimento, os sujeitos utilizam as escalas após praticar algumas atividades;

iii. Avaliações: os participantes realizam as atividades nas condições específicas da pesquisa e realizam o preenchimento das seis escalas. O número de avaliações deve ser igual ao número de tarefas vezes o número de sujeitos;

iv. Pesos: os participantes preenchem o formulário correspondente aos pesos, por meio dos círculos em cada par de fatores comparados, conforme cada tarefa e/ou cada segmento da tarefa, conforme a pesquisa que está sendo realizada;

v. Resumo: o pesquisador realiza a conferência das avaliações de cada sujeito e calcula o índice da carga de trabalho global individual do sujeito para cada segmento da tarefa.

4.2.2.4 Coleta de dados

De forma geral, os pesquisadores que utilizaram o NASA-TLX seguiram o modelo indicado por Hart e Staveland (1988) e realizaram a aplicação do método após os trabalhadores completarem a tarefa que seria mensurada, ou seja, após cenários e condições diferentes de aviação, entre tarefas com níveis de dificuldades diferentes, após completar um turno de trabalho ou um dia de trabalho, após realização do procedimento cirúrgico ou ao final de cada ensaio (AVERTY et al., 2004; BOULTINGHOUSE et al., 2007; BYRNE et al., 2010; GALSTER et al., 2001; GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006; METZGER; PARASURAMAN, 2001; SAFARI et al., 2013; WHEELOCK et al., 2015; WILSON; RUSSELL, 2003; ZHENG et al., 2012b).

4.2.2.5 Adaptações do NASA-TLX

A aplicação do NASA-TLX é feita tanto por meio físico, em papel, como em meio eletrônico, utilizando computador e internet (BOULTINGHOUSE et al., 2007; DARVISHI et al., 2016; SALEEM et al., 2007). Além disso, os trabalhadores podem ser abordados no seu local de trabalho ou podem ser levados a uma sala separada (GOULD et al., 2009).

O método inicial propõe a avaliação das escalas em um escore entre 0 e 100 pontos. Pesquisadores fizeram adaptações e estudos apresentaram escalas abrangendo valores entre 0 e 10 pontos, (MITROPOULOS; MEMARIAN, 2012), 0 e 120 pontos (SCERBO; BRITT; STEFANIDIS, 2017) e 6 e 126 pontos (WHEELLOCK et al., 2015). Além disso, Averty et al. (2014) não utilizaram todas as 6 escalas, mas apenas 5 delas: demanda mental, demanda temporal, performance, esforço e frustração.

O procedimento completo e inicial desenvolvido por Hart (1986) possui a avaliação das 6 escalas, além da sua ponderação, por meio de comparação, para estimar um valor geral de carga de trabalho, o qual é seguido pela maioria das aplicações (SAFARI et al., 2013; RUBIO VALDEHITA et al., 2010). Porém, muitos estudos não utilizam o procedimento inteiro: alguns utilizam análise apenas das escalas separadamente (LOPEZ et al., 2010; MARTIN et al., 2016; RUBIO-VALDEHITA; RODRIGO-TAPIAS, 2017; SALEEM et al., 2007), outros utilizam a soma das 6 dimensões (MITROPOULOS; MEMARIAN, 2012) e outros utilizaram a média aritmética simples das 6 escalas para determinar um índice geral (DADI et al., 2014; SCERBO; BRITT; STEFANIDIS, 2017; ZARE et al., 2016). Isso acontece, pois, alguns estudiosos consideram o procedimento de ponderação dimensional com benefício limitado (SALEEM et al., 2007).

Um método diferente proposto para a ponderar as escalas foi o uso de um sistema Fuzzy-TLX, baseado em critérios múltiplos de agregações gerais, denominado integrais Sugeno. Para isso, Mouzé-Amady et al. (2013) desenvolveram um método onde pode ser utilizado um critério externo (subjetivo ou objetivo) para determinar os pesos das classificações, desde que esse critério global esteja no mesmo intervalo que o utilizado pelas escalas (por exemplo entre 0 e 100 pontos), sendo utilizada a frequência cardíaca e a GSW nos casos estudados. Esse modelo proposto por Mouzé-Amady et al. (2013) permite a coleta de pesos de maneira mais rápida e mais compatível com os ambientes reais de trabalho, sendo melhor fundamentado matematicamente que o NASA-TLX, pois evita uma série de falhas, como violações da independência da classificação de peso. Além disso, admite um tratamento genuinamente qualitativo de medidas subjetivas de carga de trabalho.

Eraslan, Can e Atalay (2016) desenvolveram um método similar, utilizando tomada de decisões multicritério difusa, por meio dos métodos Fuzzy e AHP (*Analytical Hierarchy Process*), para calcular a carga de trabalho mental utilizando o

NASA-TLX, o SWAT e a MCH. A avaliação difusa do AHP foi utilizada para avaliar os resultados obtidos na prática, evitando resultados indesejáveis, visto que as pessoas que respondem aos questionários estão em diferentes situações mentais. Dessa forma, é uma abordagem adequada para múltiplos critérios de decisão conflitantes que complicam a tomada de decisão e é útil para identificar relações complexas entre esses critérios para avaliar alternativas de decisão (ERASLAN; CAN; ATALAY, 2016).

4.2.2.6 NASA-TLX relacionado com outros métodos

Muitos pesquisadores utilizaram a comparação do NASA-TLX com outras formas de avaliar a carga de trabalho mental, tanto com métodos de desempenho (GOULD et al., 2009), por meio de tarefa primária (AVERTY et al., 2004; OTHMAN; ROMLI, 2016) ou secundária (BYRNE et al., 2010; SCERBO; BRITT; STEFANIDIS, 2017; ZHENG et al., 2012b), como por métodos objetivos, utilizando medidas fisiológicas (WILSON; RUSSELL, 2003), como medidas dos olhos (DI NOCERA; CAMILLI; TERENCE, 2007; ZHENG et al., 2012a) e medidas do coração (KATAOKA; SASAKI; KANDA, 2011; LEHRER et al., 2010; MARTIN et al., 2016), e por outros métodos subjetivos (ARELLANO et al., 2012; LEHRER et al., 2010).

Além disso, a carga de trabalho mental foi relacionada a fatores como medidas antropométricas, características sociodemográficas e informações sobre o trabalho (SAFARI et al., 2013; SÖNMEZ et al., 2016), desempenho (KOCA et al., 2015), complexidade da tarefa (PARK; JUNG, 2006), estresse e condições ergonômicas (GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006), fadiga (ARELLANO et al., 2012), ansiedade (WHEELLOCK et al., 2015), satisfação no trabalho (BOULTINGHOUSE et al., 2007), percepção sobre segurança do paciente, erro médico e riscos do paciente (LOPEZ et al., 2010), capacidade para o trabalho (SAFARI et al., 2013), consciência da situação (LIN; LU, 2016) e carga e distúrbios físicos (DARVISHI et al., 2016; LEE et al., 2014; ZARE et al., 2016).

4.2.2.7 Análises estatísticas

Muitas análises estatísticas são utilizadas, tanto para validação dos modelos, como para comparações entre tarefas e/ou amostras, buscando também generalizações para populações.

Para validação do NASA-TLX podem ser utilizados o coeficiente de correlação intraclassa (ICC) (PARK; JUNG, 2006), os índices Keizer Meyer Olkin (KMO) e o alfa de Cronbach para testar a adequação da amostra e a confiabilidade dos dados, respectivamente (ARELLANO et al., 2012; SÖNMEZ et al., 2016).

Para comparações das amostras e das diferentes adaptações ou fases da tarefa, podem ser utilizados testes t (DARVISHI et al., 2016; RUBIO-VALDEHITA; RODRIGO-TAPIAS, 2017; SALEEM et al., 2007), teste U de Mann-Whitney (KOCA et al., 2015), Qui-quadrado, teste de Fischer (ZARE et al., 2016) e análise de variância (DADI et al., 2014; LEE et al., 2014; OTHMAN; ROMLI, 2016; SCERBO; BRITT; STEFANIDIS, 2017; YANG et al., 2012). Além disso, buscando relações entre escalas e/ou fatores e os resultados do NASA-TLX, podem ser utilizadas correlação de Pearson (AVERTY et al., 2004; LIN; LU, 2016; WHEELLOCK et al., 2015; ZHENG et al., 2012a) ou Spearman (BOULTINGHOUSE et al., 2007) e podem ser desenvolvidos modelos de regressão logística (DARVISHI et al., 2016; GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006; LEHRER et al., 2010) ou regressão múltipla (SAFARI et al., 2013).

4.2.2.8 Benefícios do NASA-TLX

O NASA-TLX é um método validado e amplamente utilizado. No experimento de Hart e Staveland (1988), as classificações do NASA-TLX foram obtidas rapidamente, levando menos de um minuto para obter a avaliação das seis escalas e não mais que dois minutos para obter os pesos para cada tipo diferente de tarefa, o que sugere que a escala de classificação multidimensional proposta é uma ferramenta prática para aplicar em ambientes operacionais, além da análise de dados ser substancialmente mais fácil de realizar quando comparada com o SWAT, que requer um programa de análise conjunta especializado.

Hart e Staveland (1988) acreditam que o NASA-TLX é mais sensível a manipulações experimentais de carga de trabalho. Cada fator foi construído para contribuir com informações independentes sobre a estrutura de diferentes tarefas. Assim, os autores acreditam que o NASA-TLX fornece informações adicionais sobre as tarefas que não estão disponíveis no SWAT ou na escala original de nove fatores.

A combinação ponderada de fatores fornece um indicador sensível da carga de trabalho geral entre diferentes tarefas e entre os diferentes níveis de cada tarefa, enquanto os pesos e a magnitude das classificações das escalas individuais fornecem informações de diagnóstico importantes sobre a fonte específica de carregamento da tarefa (HART; STAVELAND, 1988).

A escala de classificação NASA-TLX tem se mostrado uma ferramenta apropriada na autoavaliação da carga de trabalho (AVERTY et al., 2004). Além disso, Park e Jung (2006), confirmam a repetitividade do NASA-TLX por meio do teste ANOVA.

Além do mais, muitos pesquisadores mostraram-se satisfeitos com a ferramenta para validar as hipóteses e determinar o que foi pretendido no estudo (LEE et al., 2014; LIN; LU, 2016; MITROPOULOS; MEMARIAN, 2012; RUBIO VALDEHITA et al., 2010; SÖNMEZ et al., 2016; YANG et al., 2012 ZARE et al., 2016).

Algumas dicas levantadas por pesquisadores são aplicar o método logo após a conclusão da tarefa, para reduzir o viés de memória, e utilizar medidas comparativas de carga de trabalho, sejam elas de desempenho ou medidas fisiológicas (BYRNE et al., 2010; WHEELLOCK et al., 2015).

De forma geral, o NASA-TLX foi traduzido para uma série de idiomas e modificado de várias maneiras; a ferramenta é aplicada tanto por instituições governamentais, por universidades e por indústrias; estudos podem ser feitos em ambientes reais de trabalho, em laboratórios ou ambientes simulados e, além disso, o seu uso deixou de ser apenas no ramo da aviação e passou a ser aplicado em outras áreas (HART, 2006).

Controladores de tráfego aéreo, operadores de reatores nucleares, pilotos, trabalhadores de saúde (enfermeiros, médicos, anestesistas, psicólogos, cirurgiões), trabalhadores da indústria de eletrônicos, cadetes navais, professores, pessoal administrativo, policiais e vigilantes municipais, pessoal da manutenção e serviços, trabalhadores de usina nuclear, operadores de máquinas, trabalhadores têxteis, trabalhadores de construção, funcionários de um banco e soldadores são exemplos

de trabalhadores analisados pelo NASA-TLX (ARELLANO et al., 2012; BOULTINGHOUSE et al., 2007; BYRNE et al., 2010; DARVISHI et al., 2016; DI NOCERA; CAMILLI; TEREZI, 2007; GOULD et al., 2009; GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006; METZGER; PARASURAMAN, 2001; MITROPOULOS; MEMARIAN, 2012; PARK; JUNG, 2006; RUBIO VALDEHITA et al., 2010; SAFARI et al., 2013; SALEEM et al., 2007; YANG et al., 2012; ZARE et al., 2016; ZHENG et al., 2012a).

4.2.2.9 Limitações do NASA-TLX

Alguns problemas, ou limitações, foram encontrados na utilização do método NASA-TLX. Darvishi et al. (2016) e Zheng et al. (2012a) acreditam que as respostas ao NASA-TLX são afetadas pela memória do trabalhador e a capacidade de julgar a dificuldade da tarefa, algo individual de cada pessoa, o que pode ser uma limitação do método quando buscadas comparações de resultados.

O nível de experiência dos participantes e os instrumentos utilizados para realizar as tarefas também são apontados como fatores que podem limitar o uso do método (SCERBO; BRITT; STEFANIDIS, 2017).

Park e Jung (2006), não encontraram confiabilidade necessária para o estudo, utilizando o coeficiente de correlação intraclasse. Sönmez et al. (2016) também não encontraram um valor de consistência interna ideal para o questionário (SÖNMEZ et al., 2016).

O NASA-TLX também não permite a aplicação entre tarefas de um mesmo procedimento em alguns casos específicos, pois não existe nenhuma pausa entre as tarefas antes de completar todo o protocolo do procedimento (DI NOCERA; CAMILLI; TEREZI, 2007).

O uso de simulação, de condições artificiais e o número limitado de participantes, não conseguindo generalizar os resultados para uma população, também são apontados como fatores limitantes (BOULTINGHOUSE et al., 2007; KATAOKA; SASAKI; KANDA, 2011; LEE et al., 2014; SALEEM et al., 2007).

Gould et al. (2009) não conseguiram comparar os resultados do NASA-TLX para avaliar as tarefas, pois além do método não possui uma linha de referência para

as análises, as tarefas não foram estressantes o suficiente para mostrar algum resultado (GOULD et al., 2009).

Ainda, a ponderação das escalas pode ser um fator limitante, pois para Mouzé-Amady et al. (2013) ela não deve ser feita pela mesma pessoa que avalia as escalas separadamente. Outro problema é que caso seja colocada uma ponderação para cada escala em determinada tarefa, isso pode ser diferente, na prática, para cada trabalhador que executar essa tarefa. Além disso, muitas vezes é difícil para a pessoa ponderar as escalas, visto que elas podem ser muito abstratas. Por último, o cálculo pode se tornar muito demorado (MOUZÉ-AMADY et al., 2013).

4.2.3 Escala Subjetiva

No presente trabalho foram definidas como escalas subjetivas os métodos subjetivos para avaliar a carga de trabalho mental que não possuíam uma metodologia validade e empregada de forma geral, mas somente uma escala, a qual apresenta um resultado de CTM.

Uma dessas escalas é a VAS (*Visual Analogue Scales* – Escala de Analogia Visual), a qual foi difundida principalmente a partir dos estudos de Aitken e Zealley de 1969 e 1970 para mensurar experiências subjetivas, mesmo já possuindo aplicações na década de 1920. McCormack et al. (1988) descreveram o método tipicamente como consistindo de uma linha de dez centímetros ancorada em ambas as extremidades com palavras descritivas dos máximos e mínimos da dimensão que está sendo medida, possuindo pontuação entre 0 e 100 pontos, porém já foram utilizados intervalos de 5, 10, 13, 20, 75, 120 e 200 pontos. Com a escala definida, os sujeitos são solicitados a indicar sua avaliação marcando a linha no ponto apropriado entre os dois extremos. Com isso, as escalas são pontuadas medindo-se a distância do ponto inicial da escala até a marca.

Para a construção da VAS, Scott e Huskisson (1976) descreveram algumas etapas:

- i. Definir a resposta a ser observada;
- ii. Definir os extremos de resposta da escala, escolhendo frases descritivas que sejam curtas, prontamente compreendidas e não tão extremas, pois assim elas nunca serão empregadas;

iii. Determinar uma pergunta apropriada, padronizada antes do experimento, para introduzir a escala;

iv. Fazer pontos de corte definitivos para a linha, que deve ter um comprimento que possa ser compreendido como uma unidade.

Embora os rótulos verbais definam os pontos inicial e final da VAS, recomenda-se que nem números nem rótulos verbais sejam usados para definir pontos intermediários, pois isso pode causar um agrupamento de escores em torno de um dígito preferido. Também é preferível apresentar a VAS em um formato horizontal em vez de vertical (MCCORMACK et al., 1988).

Wewers e Low (1990) descreveram a VAS como uma estratégia de medição conveniente, fácil, rápida e útil em uma ampla variedade de cenários clínicos e de pesquisas para medir vários fenômenos subjetivos.

Uma evolução da VAS foi que em vez de utilizar apenas escalas de item único, existem estudos que utilizam multi itens, abordando determinada variável por meio de múltiplas escalas VAS. Além disso, apesar da maioria das VAS descritas na literatura serem projetadas para autoavaliação, vários estudos utilizam elas por meio da classificação de observadores, os quais relatam níveis significativos de confiabilidade entre avaliadores quando usadas dessa maneira (MCCORMACK et al., 1988).

Huttunen et al. (2011) em seu estudo adotaram a VAS para avaliar a carga de trabalho mental de pilotos, por meio do seu preenchimento pelo instrutor de voo. Nesse estudo foi utilizada uma linha horizontal de 10 cm na qual as pontuações das variáveis de carga variavam de 0 (nenhuma carga cognitiva) a 100 (mais alta taxa imaginável de nível de carga cognitiva). As variáveis utilizadas foram: consciência da situação, informação (quantidade, complexidade e criticidade da carga de informação) e carga de decisão. Além disso, foram utilizadas medidas fisiológicas da fala para avaliar a CTM. As medidas subjetivas das escalas VAS foram relacionadas com dados da fala por meio da correlação de Pearson.

Wilson (2002), apesar de não definir a sua escala como VAS, utilizou uma escala de 0 a 100 pontos, com 100 indicando carga de trabalho extremamente alta, para definir a CTM de pilotos. As estimativas foram dadas ao final de cada um dos segmentos de voo. Além dessa medida subjetiva, foram utilizadas medidas da frequência cardíaca, da variabilidade da frequência cardíaca, do piscar dos olhos, da atividade eletrodérmica e da atividade elétrica cerebral registrada topograficamente.

Para todas as medidas, comparações pareadas dos segmentos foram realizadas por meio de teste t pareado.

Kuijer et al. (2004), buscando comparar as demandas de trabalho, carga de trabalho e recuperação entre a condução de caminhões, coleta de lixo e rotação entre esses dois trabalhos, entre dias e durante o dia, quantificaram a carga de trabalho por meio da frequência cardíaca, do consumo de oxigênio, de avaliações subjetivas e das taxas de excreção urinária de catecolaminas. Durante a jornada de trabalho, os participantes preencheram as escalas para esforço percebido, fadiga percebida, atividade percebida e tensão percebida. A escala de esforço percebido variou de 0 (não muito esforço) a 120 (esforço extremo), a escala de fadiga percebida variou de 0 (nada cansado) a 10 (extremamente cansado) e as escalas para a percepção da atividade e para a tensão percebida variaram de 1 (ativo ou tranquilo, respectivamente) a 5 (exausto ou tenso, respectivamente).

A escala de esforço percebido foi avaliada após cada período de coleta e antes de ir ao depósito de lixo e seu escore médio foi calculado somando-se todos os escores durante o dia e dividindo o resultado pelo número de períodos de coleta. A escala de percepção de fadiga foi preenchida no início do dia e após cada período de coleta e seu escore médio foi calculado somando-se todos os escores durante o dia, dividindo o resultado pelo número de períodos de coleta e subtraindo o escore do início do dia. Já as escalas para a percepção da atividade e tensão foram preenchidas no início do dia e após cada tempo de amostra urinária durante o dia de trabalho, sendo o valor médio calculado da mesma forma que o escore médio de fadiga. Teste ANOVA foi utilizado para testar as diferenças entre as combinações de trabalho e as escalas subjetivas utilizadas (KUIJER et al., 2004).

Karhula et al. (2013) avaliaram tanto a carga de trabalho mental como física por meio de uma escala Likert de cinco pontos, variando de muito baixa a muito alta. Em seu trabalho foi explorada a relação de tensão no trabalho com as horas de trabalho, carga de trabalho percebida dependendo do turno, sonolência e recuperação de enfermeiros. Alguns testes estatísticos, como teste de Pearson, teste t, teste exato de Fisher, ANOVA, teste U de Mann-Whitney e análise de regressão logística foram utilizados para comparar as diferenças entre grupos de trabalho e características do trabalho, dependendo da variável e normalidade da distribuição da variável (KARHULA et al., 2013).

Já Lahelma et al. (2005) utilizaram uma questão perguntando quão pesado era o trabalho para avaliar a carga de trabalho mental. Caso as respostas fossem “muito pesado” ou “bastante pesado” o autor classificava como alta carga mental.

Com base nesses trabalhos, apesar das vantagens, algumas desvantagens podem ser levantadas no uso de escalas subjetivas para avaliar a carga de trabalho mental: por ser um método simples e direto, podem ser encontradas dificuldades no entendimento, principalmente em algumas populações específicas; existe uma tendência dos participantes em marcar um valor central ou próximo do valor central, podendo mascarar os resultados reais e também existem dificuldades por parte do pesquisador e aplicador da pesquisa em definir os extremos da escala, a pontuação da escala e a pergunta, de forma a tornar o procedimento de fácil compreensão e rápido preenchimento.

4.2.4 Questionário

Os questionários também são uma forma de avaliar a carga de trabalho mental de forma subjetiva, por meio de autoavaliação. Com esse propósito, os pesquisadores costumam utilizar um único questionário dividido em seções com temas diferentes, normalmente utilizando perguntas de caráter sociodemográfico e do trabalho, perguntas voltadas a carga de trabalho e, dependendo do estudo, possuem ainda seções abrangendo outras áreas de interesse da pesquisa, buscando relações entre todas elas.

Weinger et al. (2000) fizeram uma pesquisa com anestesistas, buscando medir a carga de trabalho mental associada a tarefas nas vias respiratórias. Para isso, elaboraram um questionário listando dez tarefas clínicas comuns utilizadas em procedimentos de vias aéreas, baseadas em revisão de literatura e opinião de um médico especialista, as quais englobavam a grande maioria das atividades de anestesia. Cada tarefa foi pontuada em três categorias de carga de trabalho: esforço físico, esforço mental e estresse psicológico, utilizando uma escala de analogia visual de sete pontos, sendo o valor 1 correspondente a praticamente nenhum esforço/estresse e 7 indicando o maior esforço/estresse possível.

Os sujeitos foram solicitados a classificar cada tarefa com base em um desempenho uniforme de 1 minuto de tarefa, considerando uma classificação média

para todas as vezes que realizaram essa tarefa. Ainda, foram instruídos a não classificar as tarefas que nunca haviam realizado. Além das perguntas relacionadas a carga de trabalho, o questionário também foi composto por perguntas demográficas, envolvendo a instituição de trabalho, idade, gênero e tempo de experiência na função (WEINGER et al., 2000).

Kawada e Ooya (2005) estudaram trabalhadores de uma fabricante de automóveis, buscando relações entre a carga de trabalho e reclamações de saúde no trabalho, envolvendo horas extras. Para isso, utilizaram um questionário que consistiu em treze itens para reclamações de saúde e sete itens para carga de trabalho (referência de carga de trabalho, trabalho extra, trabalho irregular, trabalho noturno, descanso, carga de trabalho mental e carga de trabalho física) com três escolhas em cada item: 0, 1 ou 3, sendo que a pontuação ficava mais alta quando os sintomas ou a carga aumentavam.

Após isso, as pontuações foram somadas e categorizadas em grupos. A carga de trabalho foi categorizada em quatro grupos: 0 a 2 como grupo A, 3 a 5 como grupo B, 6 a 8 como grupo C e acima de 8 como grupo D. No caso de queixas de saúde: escore de 0 a 3 como grupo I, de 4 a 7 como grupo II, de 8 a 14 como grupo III e acima de 14 como grupo IV (KAWADA; OOYA, 2005)

Liang et al. (2009) analisaram os fatores que afetavam a carga mental dos engenheiros de manutenção de uma central nuclear, utilizando um questionário e posteriormente um modelo mental de carga de trabalho baseado na técnica de rede neural. Para preparação do questionário foram agrupados os fatores em cinco grupos: carga de trabalho humana, máquina, materiais, método e ambiente, sendo elaborado um diagrama de causa e efeito, o qual deu origem as perguntas.

O questionário consistiu em vinte e cinco questões divididas em cinco partes: seis perguntas sobre a carga de trabalho humana, cinco perguntas sobre tarefas de manutenção, quatro perguntas sobre materiais, seis perguntas sobre métodos para realizar tarefas e quatro questões sobre o ambiente. Como resposta foi utilizada uma escala de 6 pontos, de discordo totalmente a concordo fortemente.

Para elaboração do modelo para estimar a carga de trabalho mental, foram utilizadas ainda medidas da taxa de pulso e medidas baseadas no desempenho da tarefa. Assim, o número de erros de manutenção, a complexidade, a familiarização, os materiais, os meios para realizar a tarefa e o ambiente referiram-se à carga mental

subjetiva, o tempo foi uma medida baseada no desempenho e a taxa de pulso foi um medidor fisiológico (LIANG et al.,2009)

Zoer et al. (2011) estudaram trabalhadores de uma companhia ferroviária, por meio de seis aspectos da carga de trabalho psicossocial e três resultados de saúde mental. Com isso, os pesquisadores buscaram a associação entre os aspectos da carga de trabalho e as queixas de saúde por meio de análise de regressão logística em quatro grupos etários. Para isso, foi aplicado um questionário envolvendo primeiramente perguntas sobre as características sociodemográficas e do trabalho, os seis aspectos da carga de trabalho psicossocial (pressão no trabalho, carga mental, carga emocional, autonomia no trabalho, apoio social dos colegas e apoio social dos supervisores) e os três resultados de saúde mental (fadiga relacionada ao trabalho, estresse e *burnout*).

A escala de pressão de trabalho concentrou-se exclusivamente em aspectos quantitativos da carga de trabalho em termos de velocidade e quantidade de trabalho e consistiu em onze itens, a carga mental consistiu em sete itens, a carga emocional consistiu em sete itens, a autonomia do trabalho foi avaliada com uma escala composta por onze itens sobre tempo e controle do método que um funcionário podia ou não ter em sua situação atual de trabalho, o suporte social dos colegas foi avaliado com nove itens e o apoio social dos supervisores consistiu em nove itens. Todas as escalas usaram quatro categorias de resposta: nunca, às vezes, frequentemente ou sempre. Um escore total entre 0 e 100 foi calculado para cada escala, onde uma pontuação maior indicava maior ou pior carga psicossocial (ZOER et al., 2011).

Elfering, Grebner e Haller (2012) estudaram controladores ferroviários, buscando utilizar falhas cognitivas como um mediador da influência potencial de demandas de trabalho mental e conscienciosidade em viagens arriscadas. Para isso, foi utilizado um questionário dividido em cinco partes, o qual os participantes foram instruídos a preencher durante o horário de trabalho.

A primeira parte tratava do comportamento arriscado durante o deslocamento, possuindo três itens que abordavam a pressão do tempo durante o deslocamento, negligência acidental de um sinal de parada e violação não intencional do direito de passagem dos outros. O segundo item envolvia acidentes de trajeto, com um item perguntando se os participantes estavam envolvidos em acidentes durante o deslocamento para o trabalho e para casa. A terceira parte era sobre falha cognitiva

e foi usada a Escala WCF, solicitando a frequência de falhas cognitivas no trabalho. A quarta parte tratava de interrupções, pressão de tempo e demanda de concentração no trabalho. A quinta e última parte era referente a consciência, utilizando um modelo de cinco fatores de personalidade.

A carga de trabalho foi medida na quarta parte do questionário, utilizando a versão abreviada do Instrumento de Análise de Tarefas Estressadas, com as subescalas interrupções, pressão de tempo e demanda de concentração como indicadores. A escala de interrupções no trabalho compreendeu quatro itens, por exemplo: "Quantas vezes você é interrompido por outros colegas durante a atividade de trabalho?". A escala de pressão do tempo também consistiu em quatro itens, por exemplo, "Quantas vezes você está sob pressão de tempo?". As demandas de concentração também compreenderam quatro itens, por exemplo, "Com que frequência você precisa reter informações difíceis de lembrar? (Por exemplo, quantidades, nomes, números de telefone, endereços, nomes de arquivos, registros, etc.)". O formato de resposta de todos os itens compreendia uma escala de 1 (muito raramente/nunca) a 5 (muito frequentemente/constantemente) (ELFERING; GREBNER; HALLER, 2012).

De forma geral, os questionários são enviados via correio, para a casa dos participantes ou para o local de trabalho, ou podem ser aplicados diretamente no local de trabalho (WEINGER et al., 2000).

Quando utilizados os questionários, os pesquisadores buscam relacionar a carga de trabalho mental com características sociodemográficas e do trabalho e com outros dados de saúde dos trabalhadores. Para isso, métodos estatísticos são utilizados, como análises de regressão logística ou multivariada (KAWADA; OOYA, 2005; ZOER et al., 2011).

Vantagens da utilização dos questionários consistem na sua facilidade e baixo custo para aplicação, além de serem menos invasivos, não influenciando na realização das atividades pelos trabalhadores. Além disso, muitos questionários são validados, possuem alta reprodutibilidade e conseguem abordar diversas questões importantes para a saúde do trabalhador (WEINGER et al., 2000).

Entretanto, algumas limitações são encontradas nesse tipo de metodologia. Por ser uma técnica subjetiva, existe uma influência individual muito grande, da mesma forma que nas escalas subjetivas. Além disso, a disponibilidade (maior impacto de experiências mais frequentes, recentes ou relevantes), o enquadramento

(forma que a pergunta é formulada e ordem das perguntas) e o uso de formatos de respostas diferentes podem ser fatores que influenciam os resultados. Ainda é necessário um cuidado com o tamanho do questionário, pois questionários longos possuem taxas de respostas menores e causam cansaço nos respondentes, podendo apresentar resultados errôneos (WEINGER et al., 2000).

Sendo assim, é fundamental a definição clara das tarefas a serem avaliadas, das escalas de medidas e do significado dessa avaliação, para garantir o entendimento e conseqüentemente a validade das respostas.

4.2.5 Subjective Workload Assessment Technique (SWAT)

Como foi abordado na seção 4.2.2, que trata sobre o NASA-TLX, o estudo da carga de trabalho mental por meio de medidas subjetivas foi iniciado por Cooper e Harper, em 1969, por meio da Escala de Cooper Harper, a qual é a uma escala unidimensional, e deu origem aos métodos hoje conhecidos como SWAT e NASA-TLX, ambos métodos multidimensionais (CARDOSO, 2010).

Inicialmente, a Escala de Cooper Harper deu origem a Escala de Sheridan – Simpson, a qual utilizava três dimensões subjetivas de avaliação da carga. Essa escala sofreu novas alterações, até originar o método de Reid, Shingledecker e Eggemeier (1981), o qual indica a carga mental de uma tarefa ou atividade por meio da mensuração de três dimensões: tempo, esforço mental e estresse, chamado SWAT ou *Subjective Workload Assessment Technique* (CARDOSO, 2010).

Três princípios primordiais orientaram o desenvolvimento do SWAT: (i) desenvolver uma medida tão precisa quanto possível, minimizando a intromissão do procedimento de coleta de dados sobre a situação operacional, (ii) colocar restrições mínimas de medição na complexidade da tarefa de julgamento que é exigida dos operadores que fazem avaliações de carga de trabalho, e (iii) fornecer um mecanismo para testar a validade do modelo formal de medição (REID; NYGREN, 1988).

Esse método foi desenvolvido como procedimento generalizado para avaliar a carga de trabalho mental do piloto. O modelo consiste em um procedimento de duas etapas, em que cada trabalhador, ou grupo de trabalhadores, desenvolve uma escala e uma fase de pontuação das escalas. Durante a fase de desenvolvimento da escala, os dados necessários para desenvolver uma escala de carga de trabalho

são obtidos de um grupo de sujeitos. Durante a fase de pontuação de eventos, os sujeitos avaliam a carga de trabalho associada a uma tarefa específica e/ou segmento de tarefa. Esses dois eventos são distintos e ocorrem em momentos diferentes (REID; SHINGLEDECKER; EGGEMEIER, 1981).

O SWAT utiliza três fatores para determinar a carga mental de uma atividade: tempo, esforço mental e estresse. Além disso, cada uma das dimensões é avaliada por uma escala de três pontos, composta dos níveis descritos no Quadro 8.

A aplicação desse método consiste em duas fases: uma fase de obtenção da escala da carga de trabalho mental e outra da avaliação dos níveis de carga mental.

Quadro 8 - Definições dos níveis das dimensões do SWAT

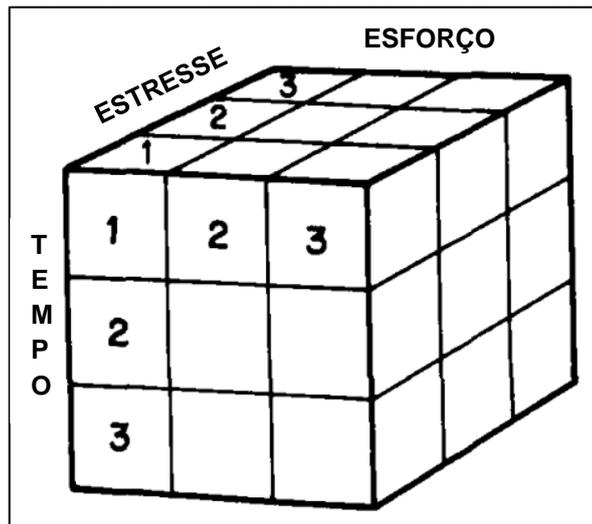
Dimensão	Níveis
Tempo	1. Normalmente sobra tempo: possibilidade de pausas durante a realização do trabalho.
	2. Às vezes sobra tempo: há possibilidades de realizar pausas, porém com uma frequência não muito definida.
	3. Raramente sobra tempo: nunca ou quase nunca sobra tempo para o operador fazer pausas.
Esforço mental	1. Pouca exigência mental: o trabalho é fácil de realizar, não exigindo muita capacidade mental (atenção, concentração, memória e percepção). Tarefas são muitas vezes executadas automaticamente.
	2. Moderada exigência mental: o trabalho exige moderada capacidade de concentração, atenção, memória e percepção. Moderada/alta complexidade de tarefas devido à incerteza, imprevisibilidade ou falta de familiaridade.
	3. Alta/elevada exigência mental: o trabalho requer muito das capacidades mentais dos trabalhadores (atenção, concentração, percepção e memória). Tarefas muito complexas, exigindo total atenção.
Estresse	1. Baixo nível de estresse: durante a execução dos trabalhos, o ambiente motiva para o trabalho e proporciona que o trabalhador se mantenha em equilíbrio. Pouco risco, confusão, frustração ou ansiedade e podem ser facilmente adaptados.
	2. Moderado nível de estresse: quando ocorrências do trabalho podem impactar no equilíbrio do trabalhador. O grau de risco, confusão, frustração ou ansiedade visivelmente aumentam a carga de trabalho e exigem uma compensação significativa para manter um desempenho adequado.
	3. Elevado nível de estresse: quando ocorrências do trabalho sempre impactam no equilíbrio do trabalhador. O nível de risco, confusão, frustração ou ansiedade aumentam grandemente a carga de trabalho e requerem que as tarefas sejam executadas apenas com o mais alto nível de determinação e autocontrole.

Fonte: Adaptado de Reid, Shingledecker e Eggemeier (1981) e Cardoso (2010).

Na primeira fase é definida a escala, segundo a importância que os avaliadores dão a cada uma das dimensões como fonte de carga mental. Esse procedimento é feito antes da tarefa e para obter os dados para o desenvolvimento da escala é feita a combinação entre os três níveis das três dimensões, totalizando 27

possíveis combinações (3x3x3), as quais são definidas por uma matriz tridimensional, mostrada na Figura 16. Assim, os trabalhadores irão ordenar as 27 combinações em ordem crescente, conforme classificação dos níveis de carga de trabalho representados por cada combinação (CARDOSO, 2010).

Figura 16 - Matriz tridimensional das dimensões do SWAT



Fonte: Reid e Nygren (1988).

Reid, Shingledecker e Eggemeier (1981) descreveram que esse procedimento pode ser feito por meio da organização de um barulho com as 27 combinações, sendo que cada carta tem um descritor para cada uma das dimensões. O indivíduo então coloca as cartas em ordem da menor carga de trabalho para a maior carga de trabalho, sendo que ele é instruído a basear seu ranking na sua experiência geral e não em qualquer situação particular da tarefa. Na verdade, a ordenação acontece com apenas 25 combinações, visto que a combinação inicial, e com menor nível de carga de trabalho mental, é definida pela escala 1 para as três dimensões (1, 1, 1) e a combinação final, com maior nível de CTM, é definida pela escala 3 para as três dimensões (3, 3, 3).

A ordem das combinações resultantes deste procedimento de classificação de cartões é então usada como dado de entrada para a análise de medição conjunta. Para ajudar na ordenação, os sujeitos são encorajados a pensar em situações de suas próprias experiências que podem ser adequadamente descritas por uma combinação das cartas. Com isso, eles podem comparar essas situações e fazer um julgamento

sobre qual das situações representa a maior carga de trabalho percebida (REID; NYGREN, 1988).

Dada a ordem de classificação que os sujeitos derivaram para as combinações dos níveis do construto tridimensional, o algoritmo é usado para procurar um conjunto de valores de escala aditiva que descreve a ordem dos níveis das três dimensões compostas. Essa análise pode ser realizada na ordenação de cada sujeito ou em uma ordem de consenso obtida pela média de um grupo de sujeitos (REID; NYGREN, 1988).

O primeiro passo para analisar a ordem de classificação dos cartões consiste na determinação do nível de concordância entre um grupo específico de indivíduos, por meio do Coeficiente de Concordância de Kendall (W). Uma "regra geral" que foi estabelecida é que se o W for igual ou superior a 0.75, há concordância suficiente para fazer uma única escala que represente todos os sujeitos, com poucas chances de corromper qualquer dimensão. Porém, uma exceção a essa prática é uma situação em que o foco da investigação pertence a uma variável de diferenças individuais (REID; NYGREN, 1988).

No caso do Coeficiente de Concordância de Kendall ser menor do que o corte proposto, um procedimento chamado prototipagem SWAT é utilizado. Para o propósito da prototipagem SWAT, seis ordenamentos hipotéticos foram desenvolvidos, baseados em uma regra que define a importância relativa de cada uma das três dimensões.

As ordenações dos avaliadores estão correlacionadas com cada uma dessas ordenações de protótipos usando o coeficiente de correlação de Spearman para determinar a importância relativa que cada sujeito coloca em cada uma das três dimensões. Normalmente, a prototipagem permite que um grupo de sujeitos seja dividido em dois ou três subgrupos homogêneos. Uma vez que tenha sido determinado quantos grupos são necessários para refletir as ponderações para um grupo particular de sujeitos, a análise conjunta é realizada para cada subgrupo separadamente. A saída da análise conjunta fornece uma escala, variando de 0 a 100, que lista um valor de escala para cada uma das células da matriz tridimensional representando combinações de níveis de tempo, esforço mental e estresse. Essa escala pode, então, ser utilizada para atribuir valores da escala a novas situações por meio da parte do procedimento SWAT chamada de pontuação de eventos (REID; NYGREN, 1988).

Já a segunda fase, realizada após as tarefas, consiste na determinação dos valores de carga mental para cada uma das dimensões, atribuindo um valor de 1 a 3, para cada tarefa, ou segmento de tarefa, desejada. Essa combinação vai ser equivalente ao valor de um dos 27 cartões organizados na fase anterior. Dessa forma, esses valores são convertidos em uma pontuação geral da carga de trabalho, encontrando o valor da escala associado à combinação durante a fase de desenvolvimento da escala, conforme o cartão gerado. Posteriormente, essas pontuações podem ser usadas em uma análise da diferença na carga de trabalho associada ao desempenho das tarefas (CARDOSO, 2010; REID; NYGREN, 1988).

Pesquisadores podem utilizar apenas o SWAT para medir a carga de trabalho mental ou podem utilizar outros métodos concomitantemente. Pickup et al. (2005), por exemplo, desenvolveram e validaram uma escala integrada de carga de trabalho, e para isso utilizaram o SWAT para comparar e relacionar os resultados de CTM, por meio da correlação de Spearman.

Kerick e Allender (2006) utilizaram, além do SWAT, medidas fisiológicas do cérebro e medidas de desempenho, por meio de tarefas primária e secundária, para avaliar a CTM de soldados. Eles ainda analisaram o estresse, por meio de classificações subjetivas de eventos, buscando relações com a carga de trabalho mental.

Fallahi et al. (2016), utilizaram o SWAT e medidas fisiológicas do cérebro e dos olhos para avaliar controladores de tráfego de uma cidade. Os resultados de CTM que envolviam o SWAT foram apresentados por meio do resultado de cada dimensão (1 a 3) e também pelo score final global do SWAT (0 a 100). Os resultados de carga mental foram classificados de 0 a 40 como baixa, 41 a 60 como moderada e 61 a 100 como alta.

Já Eraslan, Can e Atalay (2016) utilizaram o SWAT, juntamente com o NASA-TLX e a Escala de Cooper Harper, para desenvolver um modelo de tomada de decisões multicritério difusa, utilizando o método Fuzzy e o método AHP (*Analytical Hierarchy Process*), para calcular a carga de trabalho mental. A avaliação difusa foi utilizada para avaliar os resultados obtidos na prática, evitando resultados indesejáveis. Essa estrutura hierárquica é uma abordagem adequada para múltiplos critérios de decisão conflitantes que complicam a tomada de decisão e é útil para identificar relações complexas entre esses critérios para avaliar alternativas de decisão.

Reid e Nygren (1988) apresentaram diversas vantagens do método SWAT: (1) ele é baseada em propriedades formais da teoria de medição conjunta, (2) pode ser utilizado para dados individuais e de grupo, (3) é apenas ordinal (ordem de classificação), (4) a tarefa de ordenação de classificação das 27 combinações de carga de trabalho tem validade, (5) o algoritmo de escala SWAT produz simultaneamente estimativas escalonadas de intervalos dos níveis das três dimensões de carga de trabalho, bem como estimativas de seus efeitos combinados, (6) estimativas de diferenças individuais da importância de cada dimensão para avaliar a carga de trabalho podem ser obtidas e (7) uma vez que a escala tenha sido obtida, várias tarefas ou subtarefas podem ser facilmente marcadas por meio de um procedimento não intrusivo.

Em relação as limitações, como qualquer medida subjetiva, o SWAT apresenta informações relativas dos indivíduos, podendo gerar resultados equivocados e não possíveis de generalizações em muitos casos. Além disso, o procedimento recebe críticas por muitas vezes ser demorado para formulação das escalas, aplicação e análise dos dados e por ter baixa sensibilidade devido a possuir uma escala de apenas três pontos para as suas dimensões (FALLAHI et al., 2016).

Ainda, quando comparado com o NASA-TLX, apresenta a desvantagem de normalmente apresentar maior tempo de aplicação, visto que envolve 27 combinações para expressar o peso das dimensões de cada tarefa, e de apresentar apenas três dimensões, enquanto o NASA-TLX considera seis dimensões na análise da carga mental (CARDOSO, 2010).

4.2.6 Tarefa secundária

A análise de tarefa secundária, ou, em inglês, *Secondary Task Technique*, é uma das formas de analisar a carga de trabalho mental por meio de medidas objetivas, também chamadas de medidas de desempenho ou de performance, ou seja, baseada no desempenho do trabalhador em realizar uma tarefa enquanto realiza a sua atividade principal de trabalho.

Estudos envolvendo tarefa secundária para analisar a carga de trabalho, e não apenas o desempenho, são encontrados desde Knowles (1963), o qual considerava que uma das melhores maneiras de medir a carga do operador era

fazendo com o que ele executasse uma tarefa auxiliar, ou secundária, ao mesmo tempo em que executava sua tarefa primária, expressando a carga de trabalho em função das pontuações de desempenho da tarefa secundária.

Knowles (1963) acreditava que a tarefa secundária não deveria interferir fisicamente e nem interromper o desempenho da tarefa principal, devendo essa ser simples e adaptada a cada situação de trabalho. Alguns exemplos citados pelo autor são o uso de um botão de resposta, aritmética mental de resposta verbal, monitoramento visual de uma atividade e relato pressionando um botão.

Segundo Byrne et al. (2010) esse método consiste em adicionar uma segunda tarefa minimamente invasiva, cujo desempenho é facilmente medido, à tarefa principal em estudo. Com isso, supõe-se que se a carga de trabalho associada à tarefa principal se tornar excessiva, haverá certa limitação dessa carga, a qual irá prejudicar o desempenho da tarefa secundária. Dessa forma, monitorando-se continuamente o desempenho da tarefa secundária será possível determinar a CTM demandada para realizar a tarefa principal. Os autores citam como exemplo de aplicação da tarefa secundária em anestesistas a utilização de erros em problemas matemáticos simples, a latência aumentada na identificação de uma luz de alerta e maiores imprecisões na manutenção de registros para avaliar a carga mental do trabalhador.

Norman e Bobrow (1975) descreveram uma abordagem psicológica desenvolvida com base no modelo de processamento de informações humanas para avaliar a carga de trabalho mental. Segundo esse modelo, o sistema mental tem capacidade limitada, ou seja, ao realizar uma tarefa que requer mais recursos mentais, uma pessoa ficará com uma quantidade menor de recursos mentais adicionais para realizar uma tarefa secundária. Como os recursos mentais são reduzidos, o desempenho começa a se deteriorar e essa degradação é inicialmente direcionada para a tarefa designada como secundária. Levando isso em consideração, é possível estimar a carga de trabalho requerida por uma tarefa principal por meio do desempenho da tarefa secundária.

Para selecionar o tipo de tarefa secundária, pode-se utilizar a teoria de múltiplos recursos, descrita por Wickens (2002). Essa teoria acredita que o ser humano possui diferentes modalidades sensoriais, as quais podem provocar vários tipos de informações simultaneamente. Além disso, cada modalidade possui propriedades distintas e pode processar informações até certo limite. Assim, o excesso de carga de trabalho em uma modalidade sensorial pode causar problemas

e resultar em desempenho mais lento ou erros na tarefa. Isso explica o fato de normalmente o motorista, que está utilizando a visão, possuir maior compreensão de uma instrução ouvindo-a ao invés de lê-la, enquanto está dirigindo, pois nesse caso ele estaria utilizando órgãos de sentidos diferentes e não estaria sobrecarregando a mesma modalidade sensorial. Assim sendo, o efeito de uma tarefa secundária será mais significativo quando for utilizada a mesma modalidade sensorial da tarefa principal (WICKENS, 2002).

Seguindo esses preceitos, Zheng et al. (2010, 2012b) realizaram dois estudos com cirurgiões, buscando avaliar a carga de trabalho mental por meio de uma tarefa secundária, a qual utilizava as mesmas dimensões sensoriais das tarefas primárias. Em ambos os casos a tarefa principal consistia em tarefas cirúrgicas, as quais exigem grande utilização da visão, e foi utilizada uma tarefa secundária de detecção visual.

No primeiro estudo, Zheng et al. (2010) buscaram testar a hipótese que cirurgiões experientes possuem mais recursos mentais para realizar uma tarefa secundária, em comparação com cirurgiões menos experientes, enquanto realizam suturas laparoscópicas. Antes de iniciar o teste, os participantes preencheram um questionário para fornecer dados demográficos, nível de treinamento e experiência laparoscópica. Além disso, eles foram instruídos a concentrar-se na tarefa de sutura e a completar o máximo possível de suturas, sendo que a tarefa de detecção visual, que era a tarefa secundária, deveria ser realizada somente quando se sentiam à vontade para fazê-lo.

Nesse caso, como tarefa secundária, foi pedido ao cirurgião para detectar sinais visuais verdadeiros de uma lista de sinais falsos apresentados ao campo visual periférico, enquanto o cirurgião estava realizando a tarefa de sutura laparoscópica.

A qualidade da sutura foi avaliada pelo sistema de pontuação de fundamentos da cirurgia laparoscópica (FLS), sendo que a pontuação da FLS foi submetida a rigorosas avaliações de validação e confiabilidade e o desempenho da tarefa foi pontuado pela velocidade e precisão dos movimentos. Penalizações foram aplicadas quando a sutura foi conduzida de forma imprecisa (medida pelo desvio das posições de pontos predeterminados) ou a qualidade do nó foi baixa (medida pelo aperto do nó e pelo espaço entre os tecidos). O desempenho foi relatado por meio de uma escala de 100 pontos, sendo que a velocidade de movimento mais rápida, maior precisão de movimento e maior número de suturas completadas resultaram em

maiores escores de desempenho. Já a tarefa secundária foi avaliada calculando-se a taxa de perda de sinais verdadeiros ou a detecção de sinais falsos.

Já no segundo estudo, Zheng et al. (2012b) testaram a hipótese que os cirurgiões precisam de mais recursos mentais na realização de procedimentos de cirurgia endoscópica transluminal por orifício natural do que na realização de procedimentos laparoscópicos, deixando assim menos recursos mentais para o desempenho eficaz de uma tarefa de detecção visual secundária.

A tarefa secundária consistiu na detecção de um sinal visual verdadeiro, o qual era representado por um quadrado de 30 mm com canto redondo e com uma borda de 2 mm, aparecendo no monitor de um computador laptop a esquerda do computador que aparecia a realização da tarefa primária, durante um período de 6 minutos. Foram emitidos 60 sinais verdadeiros e 300 sinais falsos, de forma aleatória, sendo os sinais falsos representados por quadrados retangulares, octógonos ou quadrados de canto redondo com a borda mais fina. Cada um dos formatos foi exibido no monitor durante 0,5 segundos, utilizando um intervalo de 0,5 segundos. Para contabilizar o desempenho, os participantes deveriam dizer a palavra “*check*” para os sinais verdadeiros, sendo que a falta de um sinal verdadeiro e a resposta verbal incorreta a um sinal falso foram coletados como detecção de erros.

Nesse estudo, após completarem as tarefas, os participantes responderam ao NASA-TLX em um computador, para possuir uma medida subjetiva da carga de trabalho mental.

Ambos os casos utilizaram estatística ANOVA para comparar as médias entre as tarefas e os grupos analisados. Além disso, testes de Pearson foram realizados entre as experiências cirúrgicas e o desempenho das suturas, bem como entre as experiências cirúrgicas e o desempenho de detecção visual para investigar quais desses fatores se correlacionavam melhor com o treinamento cirúrgico (ZHENG et al., 2010, 2012b).

Byrne et al. (2010, 2013) também realizaram dois estudos com profissionais de saúde, utilizando tarefa secundária para avaliar a carga de trabalho mental. Porém, diferente dos estudos de Zheng et al. (2010, 2012b), foram utilizadas tarefas secundárias com dimensões sensoriais diferentes das tarefas primárias.

Em ambos os estudos foram feitas análises durante casos cirúrgicos, o qual exige grande utilização dos sensores visuais, utilizando simultaneamente medidas secundárias de vibração, as quais demandam dimensões táteis.

Nesses casos, antes do estudo iniciar, um dispositivo vibrotátil controlado por computador foi anexado ao braço de cada sujeito utilizando uma cinta de material macio e coldre. Um sinal foi enviado a este dispositivo em intervalos aleatórios, de 10 a 90 segundos, fazendo com que ele vibrasse enquanto eram realizadas as tarefas primárias, as quais consistiam em tarefas cirúrgicas de atendimento ao paciente.

A tarefa secundária funcionava da seguinte forma: era emitido um sinal de vibração e ele era terminado apenas quando o sujeito pressionasse um botão no dispositivo ou pressionando todo o dispositivo ao lado do corpo. Seguindo os mesmos procedimentos anteriores, os indivíduos foram solicitados a pressionar o botão sempre que o dispositivo vibrasse durante a sessão, mas foram instruídos a priorizar sua tarefa principal, sendo que o computador fazia o registro automaticamente da hora em que o sinal foi gerado e do tempo gasto pelo participante para responder ao estímulo em milissegundos.

Como medida de segurança, foi confirmado com especialistas que o uso do sistema vibrotátil não interferiria na função de qualquer equipamento da sala de cirurgia. Além disso, embora o anestesista em estudo tenha sido instruído a cuidar do paciente, um segundo anestesista estava sempre presente para monitorar a situação e garantir a segurança do paciente. Essa tarefa secundária foi escolhida por requerer mínima carga mental e não interferir na mobilidade do sujeito ou no trabalho de outros funcionários.

Além da medida de desempenho, o NASA-TLX foi preenchido após as cirurgias, para gerar uma medida subjetiva da carga mental. Em cada caso, também foram registrados detalhes do grau de anestesia, dificuldade do caso (simples, intermediário ou complexo), hora do dia e quaisquer problemas durante a realização dos procedimentos, sem registrar nenhum dado pessoal do participante.

Para avaliação dos resultados das tarefas secundárias, os tempos de resposta foram considerados como constituídos por duas populações diferentes: uma população de respostas normais, com distribuição normal sobre um média, e outra população de respostas anormais e atrasadas. Além disso, todas as vezes em que os sujeitos declararam que eram fisicamente incapazes de pressionar o botão, porque eles tinham as mãos cheias, era feita uma anotação pelo observador, sendo a decisão sobre o anestesista estar de "mãos cheias" tomada pelo próprio participante em todos os casos, pois o observador não estava ciente do momento de cada estímulo.

Nesses estudos, também foram utilizados testes estatísticos: testes de Kolmogorov-Smirnov para verificar a distribuição normal dos dados, testes de Wilcoxon e teste U de Mann - Whitney para comparar grupos e correlação de Spearman para buscar as correlações entre os fatores mensurados.

A detecção visual mostrou-se uma maneira simples e confiável de avaliar a carga de trabalho mental, utilizando avaliação do desempenho da tarefa secundária para refletir a quantidade de carga mental usada na execução da tarefa principal e, conseqüentemente, deduzir a habilidade do trabalhador para realizar suas atividades. Além disso, é uma maneira rápida e fácil de medir competências do trabalhador (ZHENG et al., 2010).

Em relação a utilização de dispositivo vibrotátil como tarefa secundária, ele se mostrou discreto, permitindo livre movimentação dos participantes e fácil uso em diferentes tipos de ambientes, podendo ser utilizado durante prática clínica. Outra vantagem é que ele pode ser útil para monitorar o desenvolvimento de um trabalhador, por exemplo, identificando não apenas quando ele pode completar uma tarefa, mas também quando sua carga mental diminui até o ponto em que ele pode completar a tarefa como um processo quase automático, ou seja, pode ser usado para rastrear o desenvolvimento de expertise ao invés de competência (BYRNE et al., 2013).

Entretanto, também são encontradas limitações. O estudo de Zheng et al. (2010) utilizou simulação e pode, portanto, não representar situações clínicas em que as tarefas cirúrgicas são mais complicadas e as distrações são apresentadas de várias maneiras. Outra limitação consiste na orientação inicial ao teste, a qual norteia os participantes a se concentrarem na tarefa primária em vez da tarefa secundária, a qual pode não ser suficiente para conduzir os participantes durante todo o curso de teste, levando alguns participantes a diminuírem o ritmo na tarefa primária para observar a tarefa secundária, desviando a atenção e podendo introduzir vieses nos resultados.

Ainda, o aumento da atividade física durante a realização da tarefa primária pode dificultar o desempenho da tarefa secundária e também podem ser encontrados problemas na determinação de limites para respostas normais das atividades secundárias, as quais teriam mais validade se fosse possível determinar a faixa normal de tempos de resposta para cada participante antes de sua avaliação no local de trabalho (BYRNE et al., 2010; ZHENG et al., 2010).

De forma geral, a tarefa secundária pode ainda impor problemas práticos significativos, como por exemplo, a aritmética mental com repetidas verbalizações pode distrair outros clínicos, o pressionamento de um botão limita o movimento físico do sujeito, o monitorando e observação requerem um observador independente e o movimento ocular requer um monitor montado na cabeça (BYRNE et al., 2013). Ela também pode ser reconhecida como uma distração para a tarefa principal, sendo irracional usá-la na sala de cirurgia real devido a preocupações com a segurança do paciente (ZHENG et al., 2012a).

Além disso, uma dica é a utilização de outros tipos de medidas validadas para analisar a CTM, sejam elas fisiológicas ou subjetivas, as quais podem encarecer e aumentar o tempo de coleta e avaliação dos dados, porém permite comparações úteis e resultados mais confiáveis.

4.2.7 Air Traffic Workload Input Technique (ATWIT)

O método *Air Traffic Workload Input Technique* (ATWIT), ou, em português, Técnica de Entrada de Carga de Trabalho de Tráfego Aéreo é uma medida subjetiva para analisar a carga de trabalho em tempo real, por meio de uma escala unidimensional, focada em controladores de tráfego aéreo (ATC), pois eles realizam procedimentos complexos envolvendo a interação homem-máquina de múltiplos procedimentos (STEIN, 1985).

Ela foi desenvolvida em 1985, por Stein, nos Estados Unidos, pois ele observou que os métodos subjetivos que existiam até aquele momento só podiam ser aplicados após a conclusão das tarefas, o que dificultava a avaliação dos níveis variáveis de carga de trabalho ao longo de uma tarefa.

Assim, Stein (1985) desenvolveu a ATWIT a fim de coletar avaliações em tempo real da carga de trabalho, solicitando classificações de um participante em intervalos regulares durante a realização de uma tarefa. Dessa forma, essa técnica é baseada em um modelo holístico que pressupõe a carga de trabalho do operador como um todo e pode fornecer estimativas precisas.

A ATWIT consiste, originalmente, de uma estimativa da carga de trabalho em intervalos regulares, durante a realização de tarefas de voo, por meio do apertado de um botão em uma escala com 10 botões, os quais representam a carga de trabalho,

de 1 (muito baixa) a 10 (muito alta), quando um sinal visual aparece. A numeração dos botões acontece da esquerda para a direita com os números de 1 a 10 (STEIN, 1985). Essa escala de 10 pontos sugere quatro níveis de erros:

1. Todas as tarefas concluídas (pontos 1 e 2);
2. Pouca chance de erro (pontos 3 a 5);
3. Alguma chance de erro (pontos 6 a 8);
4. Provavelmente algumas tarefas serão perdidas (pontos 9 e 10)

(RANTANEN; MAYNARD; ÖZHAN, 2005).

Por meio da ATWIT é possível avaliar a carga de trabalho tanto por meio do valor respondido como por meio do tempo de resposta. Além disso, existe um tempo máximo padrão para a resposta, após o qual não é possível mais responder, indicando que o ATC está completamente ocupado, não conseguindo apertar o botão (STEIN, 1985).

Em seu estudo de desenvolvimento, Stein (1985) utilizou três níveis de carga de tarefa (baixa, média e alta) em termos de densidade de tráfego, restrições planejadas e restrições não planejadas, para analisar a carga de trabalho de controladores de tráfego aéreo.

A carga de trabalho foi avaliada por meio da ATWIT, visto que o objetivo do trabalho era o seu desenvolvimento e sua descrição, uma vez a cada minuto, por meio do apertado de um botão, sendo que os ATCs foram instruídos a apertar o botão o quanto antes, em uma caixa especial de resposta construída para esse fim, quando o bloco na parte inferior do seu radar começou a piscar (STEIN, 1985).

Esses números foram ancorados verbalmente durante as instruções, de modo que a resposta dos controladores indicaria uma carga de trabalho que variava de muito fácil, escore 1, a muito difícil, escore 10. Para evitar erros ou dificuldades na mensuração da resposta da ATWIT, todos os participantes foram treinados e habituados ao uso da ferramenta (STEIN, 1985).

No estudo piloto, Stein (1985) observou que a ATWIT apresentou uma relação positiva entre carga de tarefa e carga de trabalho percebida, correspondendo ao objetivo do seu desenvolvimento.

Rantanen, Maynard e Özhan (2005) utilizaram a ATWIT modificada para avaliar a carga de trabalho. Assim, em vez da usual e original escala com 10 pontos, utilizaram uma versão modificada com apenas 7 pontos, utilizando como argumento

a padronização dos pontos da escala numérica entre as medidas, minimizando a confusão e o erro dos participantes.

Ahlstrom e Friedman-Berg (2006) utilizaram a ATWIT, como medida subjetiva, e medidas fisiológicas do movimento dos olhos para medir a carga de trabalho cognitiva. Para isso, utilizaram a escala de 10 pontos da ATWIT, enquanto os ATCs executavam operações de controle. A avaliação da ATWIT aconteceu a cada cinco minutos pelo sistema, o qual emitia vários bipes e acendia os botões do teclado. Os controladores tinham 20 segundos para responder pressionando um dos dez botões e caso não houvesse resposta dentro desse tempo, a ATWIT padronizava um dígito indicando que não houve resposta.

Ligda et al. (2010) modificaram a ATWIT, porque um dos objetivos era medir a consciência da situação utilizando consultas em tempo real. Assim, utilizaram uma escala ATWIT que continha questões envolvendo consultas de conscientização da situação, mas algumas vezes simplesmente pediam aos pilotos para classificar a sua carga de trabalho. Em todos os casos, a escala consistia em apenas 5 pontos.

Com esse propósito, a cada três minutos um alerta era apresentado ao operador em uma tela separada dos visores de voo, juntamente com um aviso auditivo. Após pressionar o botão “pronto”, uma pergunta de conscientização da situação ou da carga de trabalho aparecia na tela e o participante fazia a escolha de 1 a 5, conforme sua opinião. Se após um minuto nenhuma resposta fosse apresentada, a consulta era removida da tela e classificada como tempo limite atingido. Além disso, a simulação não parou enquanto os participantes respondiam às perguntas (LIGDA et al., 2010).

Utilizando esse método, foram retiradas três medidas: o intervalo entre o aparecimento do alerta “pronto” e o apertado do botão “pronto”, o intervalo entre a apresentação da pergunta e a sua resposta e o escore da resposta, sendo que Ligda et al. (2010) consideraram que o tempo para responder ao sinal “pronto” e o escore de resposta eram quantificadores da carga de trabalho, porém o intervalo para responder à pergunta foi considerado apenas um tempo de reflexão sobre a percepção da situação.

Para os autores, a ATWIT modificada mostrou ser uma medida de carga de trabalho válida e confiável, fornecendo informações mais detalhadas que as métricas subjetivas de carga de trabalho pós-execução das tarefas.

Kupfer et al. (2011) utilizaram a ATWIT e também o NASA-TLX para medir a carga de trabalho durante e após a execução das tarefas de controle do tráfego aéreo, respectivamente. A ATWIT foi solicitada a cada cinco minutos e utilizou uma escala entre 1 e 6 pontos. Segundo os autores, a escala de classificação da ATWIT não foi sensível o suficiente para detectar pequenas flutuações na carga de trabalho do controlador, devido à baixa complexidade dos cenários testados.

Já Fincannon e Ahlstrom (2016) realizaram um estudo de revisão de literatura buscando fornecer uma recomendação para o tamanho ideal da ATWIT, pois apesar de originalmente possuir uma escala de 10 pontos, muitos pesquisadores utilizavam uma variação de 7 pontos dessa escala. Para isso, utilizaram o conceito de confiabilidade da avaliação de impacto, pois escalas mais confiáveis produzem tamanhos de efeito mais fortes e reduzem os custos associados à experimentação.

Por meio desse estudo, Fincannon e Ahlstrom (2016) forneceram uma avaliação preliminar do tamanho da escala da ATWIT e os resultados suportaram o uso da versão de 10 pontos, pois foi constatado que os controladores de tráfego aéreo especializados, usados nos estudos de pesquisa do ATC, tendem a relatar níveis mais baixos de carga de trabalho. Assim, a versão de 10 pontos fornece mais pontos na extremidade inferior da escala, podendo fornecer uma avaliação mais adequada e confiável da carga de trabalho do controlador. Além disso, os resultados indicaram que a força da correlação entre os escores da ATWIT e a contagem de aeronaves foi maior para a versão de 10 pontos do que para a versão de 7 pontos.

Alguns autores utilizam uma nomenclatura diferente para a ATWIT, a qual é chamada de *Workload Assessment Keypad* (WAK), conhecida em português como Teclado de Avaliação de Carga de Trabalho.

Na verdade, Fincannon e Ahlstrom (2016) colocam que o WAK é um teclado físico com botões utilizado para a ATWIT, visto que também pode ser utilizado um teclado *touch screen*, o qual será apertado conforme a classificação de CTM escolhida pelo ATC. Da mesma forma que a ATWIT, além da escala usual com 10 pontos, e nesse caso um WAK com 10 botões, existem adaptações.

Hah, Willems e Schulz (2010) e Hat et al. (2016) utilizaram um WAK original com 10 botões, o qual foi aplicado de dois em dois minutos, durante a realização das tarefas. Truitt et al. (2015) e Hale e Stanley (2017), da mesma forma, também utilizaram um WAK com 10 botões. Já Prevot et al. (2015) e Sturrea et al. (2015)

utilizaram um WAK com 6 botões, com uma escala de 1 a 6 pontos, com coletas a cada três minutos.

De forma geral, a Técnica de Entrada de Carga de Trabalho de Tráfego Aéreo apresenta diversas vantagens: é uma medida real e momentânea da carga de trabalho, sendo possível realizar comparações durante a execução de uma tarefa, não apresentando apenas uma estimativa final; não depende da memória dos participantes, visto que acontece em tempo real, durante a realização da atividade; é de fácil aplicação; não apresenta maiores complexidades para análise dos dados e pode ser adaptada dependendo do contexto da pesquisa. Como desvantagem, é preciso um cuidado por ser uma técnica que é de certa forma invasiva, pois os participantes precisam respondê-la durante a realização da sua atividade primária e depende de um *software* para gerenciamento dos intervalos.

4.2.8 Rating Scale of Mental Effort (RSME)

O método *Rating Scale of Mental Effort* (RSME), ou, em português, Escala de Avaliação do Esforço Mental, é mais uma forma de avaliar a carga de trabalho mental por meio de medidas subjetivas, ou seja, utilizando autoavaliação dos trabalhadores. Porém, diferente dos métodos NASA-TLX e SWAT, que utilizam escalas múltiplas, a RSME utiliza uma metodologia com apenas uma escala para avaliação do esforço, por isso classificado como uma medida unidimensional.

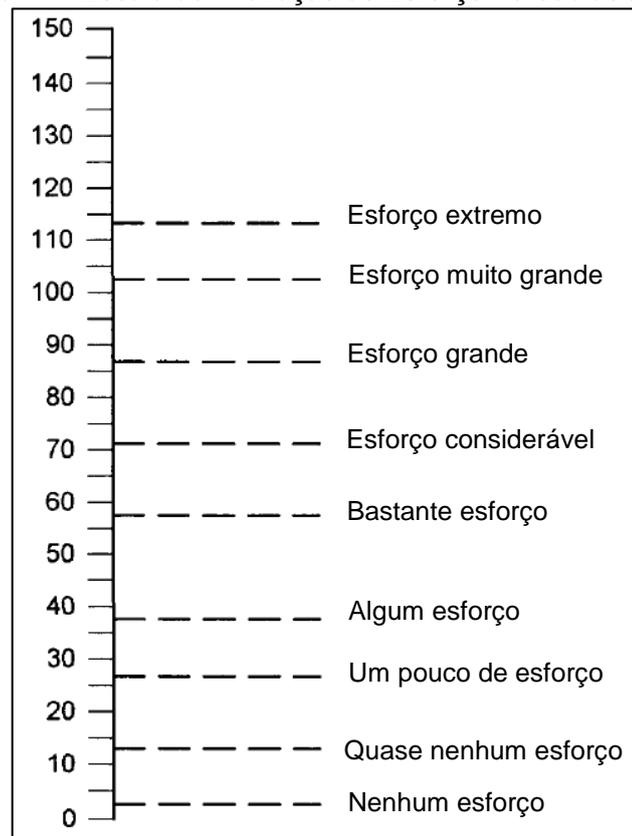
Zijlstra (1993) inspirou-se no sucesso e na facilidade de aplicação das escalas de Cooper Harper e de Borg para desenvolver a escala RSME, para medir o esforço. Sendo assim, em 1993, na Holanda, ele desenvolveu, testou e validou o método, o qual foi desenvolvido com base em três requisitos: prioridade para a tecnologia, divisão ideal de tarefas entre homem e máquina e facilidade na execução do teste.

Para desenvolver a escala, dois grupos, com diferentes educação e formação, foram convocados e cada um deles desenvolveu um modelo, sendo que um utilizava o conceito de “*inspanning*” e o outro de “*moeite*”, duas palavras holandesas para “esforço”. O começo e o final da escala foram definidos por meio de palavras estratégicas e intervalos realísticos entre esses dois pontos, por meio de estudos. A intenção da escala era que ela fosse usada durante a execução da tarefa.

Mais tarde, a classificação das palavras foi traduzida para números, para facilitar a análise (ZIJLSTRA, 1993).

Para isso, primeiramente o esforço foi mensurado em 17 palavras, recebendo uma letra de A a Q, as quais, cada uma, representavam um significado qualitativo diferente para o esforço. Na segunda fase, a letra Q foi tomada como referência e as demais letras, ou conceitos qualitativos, receberam um valor para quantificar o esforço conforme os conceitos anteriormente utilizados. Já na terceira e última fase, as escalas foram comparadas e com isso foi definida a escala final, utilizando apenas 9 palavras, e relacionando-as com um valor numérico. Esta escala numérica vertical, com escore entre 0 e 150 pontos e originalmente com 30 divisões, com cada divisão representando 5 pontos na escala, deu origem a RSME, sendo sua escala apresentada na Figura 17.

Figura 17 – Escala de Avaliação do Esforço Percebido (RSME)



Fonte: Adaptado de Zijlstra (1993).

Para validar o modelo, Zijlstra (1993) testou a ferramenta com estudantes e com motoristas de ônibus e chegou à conclusão que realmente era possível

mensurar o esforço mental, com confiança, utilizando a RSME. Além disso, ela se mostrou de fácil aplicação tanto em laboratório quanto em situações reais de trabalho.

Ward et al. (2006) aplicaram a RSME para analisar a carga mental de motoristas de ônibus em situações de trânsito rápido de ônibus, utilizando uma espécie de muro para separar a pista dos ônibus com as demais pistas, e com ou sem o uso de um sistema de suporte de pista de protótipo (LSS) para apoiar o controle do veículo dentro dos limites estreitos da pista em situações de baixa visibilidade. Esse estudo foi feito visto que se acreditava que a largura da pista e a necessidade de antecipar ou interagir com outro tráfego na faixa adjacente poderiam ser fatores estressantes para os motoristas de ônibus. Assim, foram utilizadas três variáveis independentes como possíveis estressores: densidade de tráfego, largura da pista e uso do sistema LSS. Já as variáveis dependentes foram a carga de trabalho mental e o comportamento de condução em termos de segurança e mobilidade.

Com esse propósito os autores utilizaram a Escala de Avaliação do Esforço Mental para estimar o esforço mental subjetivo ao final de cada teste, a qual teve seu resultado apresentado graficamente por meio do valor médio dos participantes. Além disso, os motoristas também foram encorajados a comentar sobre quaisquer estressores percebidos ou eventos estressantes durante cada teste, os quais foram registrados pelo pesquisador, buscando relações com o esforço mental relatado.

De Rivecourt et al. (2008) estudaram pilotos, avaliando sua carga mental por meio de medidas fisiológicas, utilizando a atividade cardiovascular e ocular, e da medida subjetiva RSME. A apresentação da RSME aconteceu após o voo, para evitar interferência com a execução da tarefa, sendo dependente da memória do piloto. Assim, não foi possível obter uma classificação confiável de cada um dos 37 elementos do voo, porém foi possível avaliar os 6 segmentos de tarefa, apresentando o valor médio da Escala de Avaliação do Esforço. Como a RSME foi utilizada apenas para estimar o esforço investido durante os seis segmentos de voo, não foi possível verificar se as classificações subjetivas também aumentaram com o aumento do nível de demanda de tarefas (DE RIVECOURT et al., 2008).

Ghanbary Sartang et al. (2016) buscaram avaliar a efetividade da RSME para mensurar a carga de trabalho mental de enfermeiros. Para isso, o trabalho consistiu em três questionários: o primeiro referente a informações demográficas, incluindo idade, gênero e experiências de trabalho; o segundo foi a aplicação do NASA-TLX e o terceiro a aplicação da RSME, sendo que esses dois métodos foram

utilizados para avaliar a CTM. Com os dados coletados, análises estatísticas foram realizadas, como teste t, análise de variância e teste de correlação de Pearson.

Por meio desse estudo, Ghanbary Sartang et al. (2016) mostraram que a RSME é um bom método para avaliar a carga mental, visto que mostrou correlações aceitáveis com os índices do NASA-TLX. Além disso, essa escala não requer nenhum dispositivo especial, é simples e fácil de usar, barata, rápida e possui aplicabilidade no local de trabalho sem interferir no trabalho dos indivíduos.

Inicialmente desenvolvida em holandês, a RSME foi traduzida posteriormente para o inglês e a partir do inglês traduzida para outros idiomas. Porém, nem sempre uma simples tradução é o suficiente para empregar o método de forma confiável. Dessa forma, Widyanti, Johnson e De Waard (2013) adaptaram a RSME para utilizar na Indonésia, pois os autores acreditavam que o fator cultural era diferente do europeu e assim a escala não se aplicaria da mesma forma na Indonésia.

Para isso, primeiramente os rótulos descritivos dos pontos de ancoragem da RSME em inglês foram traduzidos para o indonésio usando retrotradução. Um grupo de 80 estudantes indonésios foi então convidado a combinar os rótulos descritivos com os nove pontos de ancoragem originais da RSME, acabando por transformar as 9 palavras de ancoragem em 7. Depois, um grupo separado de 80 estudantes indonésios foi solicitado a colocar os rótulos descritivos em uma versão da RSME com apenas a escala de 150 pontos marcada e nenhum ponto de ancoragem separado. Após essas três fases, os pontos de ancoragem derivados desses dados foram utilizados para criar uma versão adaptada da RSME para o indonésio (WIDYANTI; JOHNSON; DE WAARD, 2013).

Para testar a validade dessa nova escala, as sensibilidades da RSME original e da RSME adaptada foram comparadas em um experimento no qual os participantes realizaram uma tarefa de busca e memorização com diferentes níveis de dificuldade que mostraram induzir de forma confiável diferentes níveis de carga mental. As classificações da carga de trabalho mental subjetiva foram comparadas usando ANOVA mista com o grupo (RSME original vs. adaptada) como um fator entre indivíduos e condição de dificuldade como um fator dentro dos sujeitos.

Segundo Widyanti, Johnson e De Waard (2013) a RSME adaptada com um número reduzido de rótulos descritivos pareceu ser usada de forma diferente da RSME original. Como um intervalo mais amplo de valores é usado, o instrumento adaptado pode ser mais sensível a mudanças na demanda do que o original. No

entanto, os participantes holandeses em um estudo anterior utilizaram um intervalo ainda maior da RSME do que os indonésios que usam a RSME original em indonésio ou adaptada, o que parece apontar para diferenças fundamentais em como holandeses e indonésios avaliam a carga de trabalho, fato que sugere que não apenas as propriedades da escala, mas também a cultura, influencia na medição da carga mental subjetiva.

Como limitação do estudo, Widyanti, Johnson e De Waard (2013) indicaram o fato do método adaptado ter sido testado apenas com alunos, sendo que uma condição mais real de resultados seria experimentada caso fossem utilizadas populações de trabalhadores e uma gama maior de tarefas.

De forma geral, a RSME, uma escala de avaliação unidimensional, pode ser tão útil e sensível a mudanças na carga mental quanto as escalas multidimensionais, e isso tem sido mostrado repetidamente, por meio da sensibilidade a mudanças na carga de trabalho mental em ambientes laboratoriais e aplicados (WIDYANTI; JOHNSON; DE WAARD, 2013). Além disso, a simplicidade do método e a rapidez para a aplicação tornam-se fatores favoráveis ao emprego em ambientes reais de trabalho, visto que não interferem na rotina de trabalho.

Porém, algumas desvantagens são encontradas, principalmente referentes a adaptação linguística da escala, pois algumas palavras utilizadas na escala original muitas vezes não possuem tradução em outros idiomas, além do fato de existir o fator cultural, que faz com que trabalhadores de nacionalidades diferentes interpretem e realizem a avaliação da CTM de forma diferente, mesmo utilizando o mesmo instrumento.

4.2.9 Surgical Task Load Index (SURG-TLX)

O método *Surgical Task Load Index* (SURG-TLX), ou, em português, Índice de Carga de Tarefa Cirúrgica, é uma forma de determinar a carga de trabalho mental de cirurgiões por meio de uma medida subjetiva, a qual utiliza uma escala multidimensional.

Ela foi desenvolvida e validada por Wilson et al. (2011) para fornecer informações diagnósticas da carga de trabalho multidimensional específica de cirurgias e determinar sua utilidade em fornecer informações sobre o impacto de

várias fontes de estresse nas demandas percebidas. Além disso, os autores buscaram esse desenvolvimento pois a sala cirúrgica é um ambiente multifacetado que expõe os cirurgiões e suas equipes a condições consideráveis de indução de estresse intraoperatório, tais como a complexidade de procedimentos, pressão de tempo, avaliação de pares, multitarefa e distrações.

Como nessa época o NASA-TLX já era um instrumento validade e bastante utilizado, inclusive na área da saúde, a intenção do SURG-TLX era de manter a estrutura geral do NASA-TLX, mas torná-lo mais relevante para as demandas específicas da cirurgia.

Assim, para o seu desenvolvimento, Wilson et al. (2011) consideraram o procedimento adotado pelo método DALI, uma variante do NASA-TLX utilizada para motoristas de carros. Para o DALI, primeiramente foram determinadas suas seis dimensões (esforço de atenção, demanda visual, demanda auditiva, demanda temporal, distração e estresse situacional) pela discussão com um número de especialistas e depois um estudo foi projetado para testar a sensibilidade e o diagnóstico do instrumento para tarefas típicas de direção, sendo confirmada a sensibilidade das seis dimensões do DALI.

Dessa forma, na busca do desenvolvimento do SURG-TLX, foi feita uma pesquisa qualitativa para identificar os principais estressores intraoperatórios e foram consideradas quais dimensões do NASA-TLX e do DALI melhor se aproximavam das demandas enfrentadas pelos operadores cirúrgicos. De tal modo, chegou-se à conclusão que as três dimensões de demandas de tarefas do NASA-TLX (demanda mental, demanda física e demanda temporal) e as duas dimensões de demanda ambiental do DALI (distração e estresse situacional) eram úteis ao SURG-TLX. Já a sexta dimensão escolhida deveria ser relacionada a complexidade da tarefa, em vez de relacionada ao esforço e/ou à frustração. Assim, as seis dimensões do SURG-TLX foram definidas e formuladas por Wilson et al. (2011) da seguinte forma:

1. Demandas mentais: Quão mentalmente fatigante foi o procedimento?
2. Demandas físicas: Quão fisicamente fatigante foi o procedimento?
3. Demandas temporais: Quão rápido foi o ritmo do procedimento?
4. Complexidade de tarefas: Quão complexo foi o procedimento?
5. Estresse situacional: Quão ansioso você se sentiu durante a realização do procedimento?
6. Distrações: Quão distrativo foi o ambiente operacional?

Após essa formulação, oito cirurgiões experientes foram consultados quanto as dimensões escolhidas para o SURG-TLX. Apesar de serem levantados diversos fatores (por exemplo, a negatividade de outros na sala de cirurgia, a indisponibilidade do equipamento preferido e as expectativas do paciente), houve um consenso geral que as dimensões refletiam as demandas típicas experimentadas na cirurgia.

Assim, foram confirmadas as seis dimensões e o procedimento utilizado foi definido, de forma semelhante ao NASA-TLX, por meio de duas etapas: a primeira parte envolvendo o cálculo de pesos das seis dimensões seguindo um conjunto de 15 comparações pareadas, dois a dois, entre as seis escalas, sendo que a dimensão com maior peso é o fator contribuinte mais importante para a carga de trabalho percebida, com os escores variando de 0 a 5, e a segunda parte envolvendo a classificação das seis escalas bipolares, que refletem as dimensões separadas ancoradas entre extremos, conforme mostrado na Figura 18.

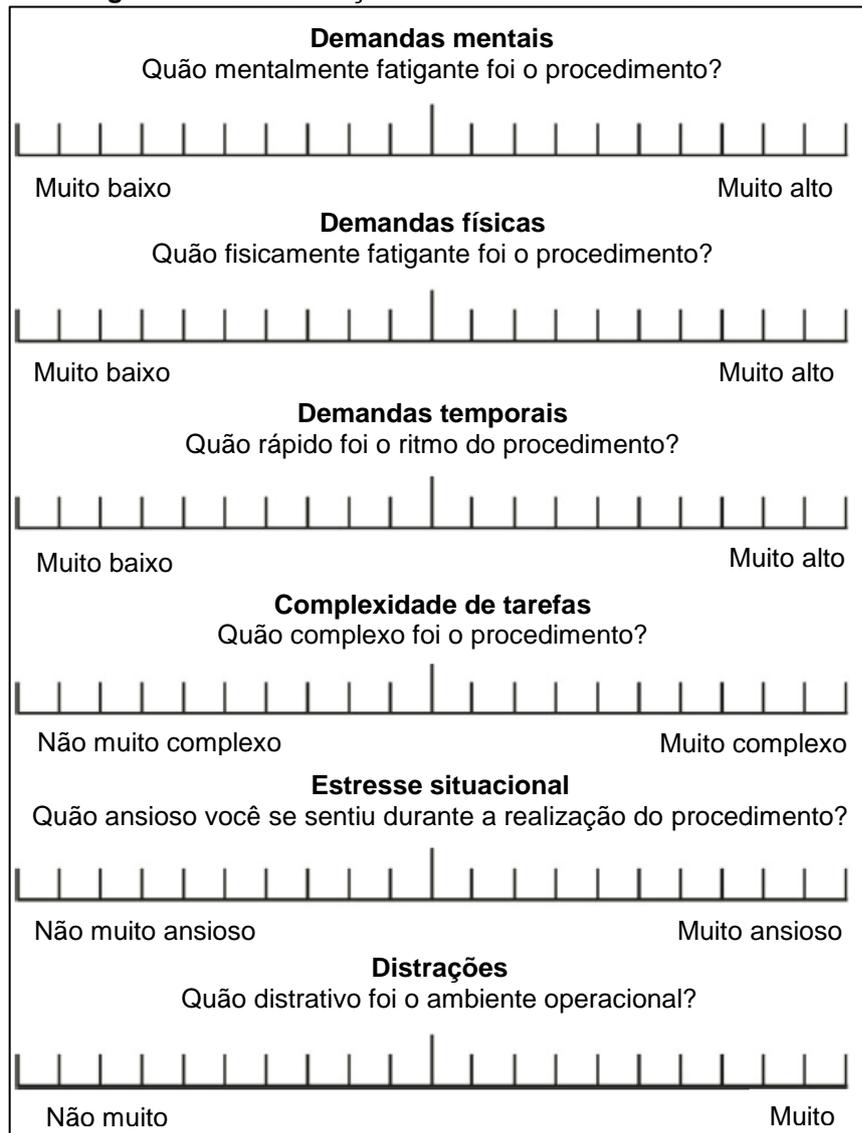
Cada escala, que representa uma dimensão, recebe uma pontuação de 0 a 100. A pontuação dada por um participante para cada domínio na escala bipolar é então multiplicada pela classificação atribuída a esse domínio, produzindo uma pontuação final para cada dimensão cognitiva. Por fim, a pontuação total da carga de trabalho é determinada pela soma das pontuações das seis dimensões.

Com o SURG-TLX desenvolvido, Wilson et al. (2011) precisavam validá-lo. Para isso, operadores cirúrgicos estagiários, expostos a vários estressores intraoperatórios, uma vez que realizaram uma tarefa laparoscópica bem validada, foram analisados, tanto por meio do seu desempenho na realização do procedimento como por meio do SURG-TLX.

Uma média de carga de trabalho para cada dimensão e para a carga de trabalho total foi calculada para cada uma das condições de interesse e submetida a análise unidirecional de variância. Ainda, cinco hipóteses relacionadas as dimensões do SURG-TLX foram levantadas, utilizando relações entre as dimensões e as condições utilizadas para a realização das tarefas.

Por meio desse estudo, o SURG-TLX foi validado e se mostrou uma medida multidimensional da carga de trabalho cirúrgica sensível a uma variedade de diferentes estressores cirúrgicos, incluindo fadiga física, pressão de tempo, multitarefa e aumento da complexidade.

Figura 18 - Classificação das seis escalas do SURG-TLX



Fonte: Adaptado de Wilson (2011).

Além disso, Wilson et al. (2011) citaram diversos pontos fortes do SURG-TLX: primeiro, a capacidade de avaliar o motivo de um procedimento ser difícil, especialmente quando realizado sob várias condições exigentes ou estressantes; segundo, o SURG-TLX pode ajudar os cirurgiões a tomarem melhores decisões sobre as prováveis demandas associadas à introdução de novas técnicas ou tecnologias e terceiro, a combinação de intervenções de treinamentos apropriados às necessidades do operador só pode ser auxiliada por informações diagnósticas sobre as fontes de sobrecarga ou estresse, o que é conseguido por meio desse procedimento. Assim, as soluções de treinamento devem, portanto, ser direcionadas para aumentar os recursos de enfrentamento para as demandas particulares.

Uma dica dos autores é que se deve buscar sempre a conclusão da comparação das dimensões e a avaliação da escala Likert de cada dimensão, para seguir o procedimento completo. Apesar disso, a escala Likert, por si só, pode fornecer um análogo visual informativo das demandas dos procedimentos. Além disso, utilizando apenas a avaliação individual das dimensões, o SURG-TLX apresenta maior utilidade clínica, pois pode ser administrado rapidamente para ajudar a orientar o processo de autorreflexão dos cirurgiões que acabaram de apresentar um desempenho ruim e caso a ponderação relativa entre duas dimensões permaneça obscura, as comparações pareadas poderão então ser usadas para distinguir qual delas possui maior contribuição para a carga de trabalho.

Já como alguns fatores que podem dificultar a aplicação do SURG-TLX, Wilson et al. (2001) citaram a dificuldade para categorizar certos procedimentos, a implementação da tecnologia na sala de cirurgia e a falta de treinamento, ou treinamento insuficiente, para quem está completando o procedimento de avaliação da tarefa cirúrgica (WILSON et al., 2011).

Berg et al. (2015), avaliaram o impacto do aumento da temperatura no desempenho operatório laparoscópico e no estresse cognitivo do cirurgião. Para isso, foi avaliado o desempenho do cirurgião em tarefas laparoscópicas em duas temperaturas ambientes, 19°C e 26°C, e o estresse cognitivo foi medido por meio do SURG-TLX.

Nesse estudo, Berg et al. (2015) examinaram tanto o instrumento completo quanto a escala bipolar de cada um dos seis domínios cognitivos, justificado pelo fato que a porção de comparação pareada do instrumento completo força os participantes a escolherem um dos dois tipos de estressores, não permitindo que não seja indicada uma situação que não teve nenhum dos dois tipos de estressores.

Dessa forma, no estudo de Berg et al. (2015), a resposta forçada da parte de comparações do instrumento era menos útil para abordar a questão de pesquisa, pois o interesse específico estava em comparar a quantidade de estresse, incluindo a opção sem estresse, que os participantes sentiam em cada domínio cognitivo examinado. Com esse propósito, o SURG-TLX foi administrado durante os últimos 5 minutos em cada condição ambiental específica, sendo que os participantes receberam instruções padronizadas sobre como completar o instrumento.

As variáveis contínuas foram comparadas com o uso de análise não paramétrica com o teste de Wilcoxon, as variáveis categóricas foram comparadas pelo

teste exato de Fisher e os resultados foram relatados como mediana e intervalo interquartilico (BERG et al., 2015).

Weigl et al. (2015) avaliaram o impacto de diferentes interrupções no fluxo de trabalho intraoperatório na capacidade dos cirurgiões para gerenciar sua carga de trabalho de forma segura e eficiente. Para isso, as interrupções foram avaliadas por observações de especialistas, usando uma ferramenta de observação bem validada, e os cirurgiões, os enfermeiros e os anestesistas avaliaram sua carga de trabalho intraoperatória diretamente após a conclusão do caso com base em três itens do SURG-TLX: demanda mental, estresse situacional e distração.

Os pesquisadores utilizaram apenas três itens do SURG-TLX por razões práticas, principalmente pela brevidade. Os itens foram selecionados com base em recomendações prévias de especialistas, bem como por meio de revisão de literatura. Cada dimensão foi pontuada de 0 a 100 pontos e não foi utilizada nenhuma comparação para ponderar os itens.

A apresentação dos valores aconteceu pela soma e média das variáveis em estudo, utilizando ainda estatística inferencial e análises de variância para explorar as diferenças entre os grupos e correlação parcial para examinar as associações entre as interrupções e a carga de trabalho dos cirurgiões (WEIGL et al., 2015).

WEIGL et al. (2016) realizaram outro estudo buscando investigar o efeito da interrupção no fluxo cirúrgico na carga de trabalho intraoperatória do cirurgião e no desempenho técnico.

Com esse propósito, em primeiro lugar foi examinado o efeito de interrupções intraoperatórias na carga de trabalho, em segundo lugar foram observados os efeitos das variações da carga de trabalho mental no desempenho cirúrgico e em terceiro lugar foram investigados os efeitos de diferentes fontes de ruptura na carga de trabalho mental e no desempenho técnico. Além disso, os autores buscaram contribuir para a validação do SURG-TLX como um instrumento adequado para avaliar a carga de trabalho cirúrgica.

Os cirurgiões do estudo avaliaram sua carga de trabalho intraoperatória diretamente após a conclusão da respectiva simulação, com base nas seis dimensões do SURG-TLX. Os escores totais da soma SURG-TLX foram calculados para as primeiras cinco dimensões e a sexta dimensão, distrações percebidas, foi tratada separadamente, uma vez que era uma característica específica do cenário. Além disso, não foi utilizada ponderação das escalas.

Para todos os dados, médias e desvios padrões foram calculados. Para comparações das classificações SURG-TLX entre os cenários de treinamento e de interrupção, foi aplicado teste t de amostras pareadas. Ainda, análise de regressão linear foi utilizada para obter associações entre a carga mental e o desempenho cirúrgico (WEIGL et al., 2016).

Forsyth et al. (2018), estudaram as interrupções vivenciadas pelos enfermeiros de emergência e buscaram estabelecer a validade de convergência de uma medida objetiva da carga de trabalho, relacionando as características da interrupção com as medidas objetivas e subjetivas da carga de trabalho.

As interrupções foram capturadas em tempo real em turnos de 8 ou 12 horas usando uma ferramenta de interrupções de fluxo de trabalho já validada anteriormente. Os dados coletados em cada interrupção incluíam modalidade, prioridade e local onde a interrupção ocorreu. Ao meio e ao final do turno, o Índice de Carga de Tarefa Cirúrgica e a Ferramenta de Avaliação Cognitiva Rápida (RCAT) foram administrados aos enfermeiros participantes para medir a carga de trabalho de forma subjetiva e objetiva. Nesse estudo, o SURG-TLX utilizou uma escala de 0 a 20 pontos para cada dimensão e no final utilizou a soma de todas para dar uma carga de trabalho geral SURG-TLX.

Foram realizados testes t pareados nos dados de carga de trabalho do meio e do final do turno e correlações foram realizadas para identificar associações entre interrupções, duração do turno e mudanças nos dados de carga de trabalho SURG-TLX e RCAT. Ainda, o RCAT foi validado com base nas dimensões do SURG-TLX (FORSYTH et al., 2018).

Abdelrahman et al. (2016) realizaram um estudo buscando comparar o estresse do cirurgião e a carga de trabalho entre dois métodos laparoscópicos diferentes, utilizando quatro medidas para avaliar a carga de trabalho do cirurgião: SURG-TLX, frequência cardíaca máxima, nível de cortisol salivar e usabilidade dos instrumentos.

A frequência cardíaca máxima e os níveis de cortisol salivar, medidas fisiológicas, foram amostrados pelo cirurgião antes da designação aleatória do procedimento cirúrgico, no período intraoperatório e ao final do procedimento. Já o SURG-TLX foi aplicado apenas após cada procedimento, juntamente com a pesquisa de usabilidade do instrumento. Nesse caso, foi utilizada apenas a análise das

dimensões separadamente do SURG-TLX, utilizando testes t, teste de Wilcoxon e ANOVA não paramétrico de Kruskal-Wallis nas variáveis dependentes.

Apesar de haver risco de viés no uso dos questionários subjetivos por um cirurgião, o SURG-TLX foi bem validado e utilizado nos conjuntos cirúrgicos, e o risco do viés foi minimizado pelo uso dos métodos objetivos fisiológicos (ABDELRAHMAN et al., 2016).

Já Dias et al. (2018), por meio de uma revisão sistemática de literatura que buscava os métodos utilizados para avaliar a carga cognitiva dos cirurgiões, mostraram um dado estatístico interessante e atual: o SURG-TLX foi a segunda medida de autorrelato mais utilizada, perdendo apenas para o NASA-TLX, ferramenta que o originou.

Apesar de ser um método validado e confiável, de fácil e rápida aplicação e totalmente voltado para situações cirúrgicas, conforme apresentado acima, o SURG-TLX pode apresentar algumas desvantagens, como o viés subjetivo, efeitos retrospectivos ou viés de recordação em como os cirurgiões e outros membros da equipe relatam o procedimento realizado, e também o sistema de pontuação, o qual não permite medir as variações da carga de trabalho mental durante o procedimento (WEIGL et al., 2015, 2016).

Uma observação importante feita por Dias et al. (2018) é que como um instrumento de autorrelato, o SURG-TLX é valioso quando o objetivo é avaliar a carga cognitiva global em diferentes procedimentos cirúrgicos e avaliar as curvas de aprendizagem dentro da educação cirúrgica baseada em competências, porém quando o objetivo é avaliar a carga cognitiva relacionada a estágios operatórios específicos, ferramentas em tempo real devem ser usadas, pois permitem a captura das flutuações da carga cognitiva. Nesse caso, uma combinação de métodos subjetivos e objetivos pode fornecer uma medida ideal da carga de trabalho mental dos cirurgiões, fato que pode ser visto no trabalho de Abdelrahman et al. (2016), utilizando o SURG-TLX alinhado com medidas fisiológicas.

4.2.10 Borg Scale of Perceived Exertion (BSPE)

Borg Scale of Perceived Exertion (BSPE), *Borg's Rating of Perceived Exertion (RPE)* ou, em português, Escala de Borg de Esforço Percebido é uma forma

de medir subjetivamente a carga de trabalho mental, por meio da mensuração de uma escala unidimensional.

A Escala de Borg de Esforço Percebido foi proposta em 1970, por Borg, originalmente com o objetivo de ser uma medida direta e simples do grau de esforço percebido, por meio de um método que poderia ser utilizado por praticamente todos, não apenas por indivíduos com capacidade superior para lidar com números expressos em proporções, como acontecia com os métodos que existiam até então. O esforço percebido é um indicador direto para avaliar o grau de estresse físico e também um indicador indireto da capacidade de trabalho físico (BORG; NOBLE, 1974).

Dessa forma, a escala foi desenvolvida para que comparações interindividuais pudessem ser feitas em estudos aplicados, objetivando construir uma escala capaz de medir o esforço físico subjetivo. O critério de validação utilizado foi a frequência cardíaca, pois naquela época ela já era conhecida como um bom indicador da tensão metabólica e era amplamente utilizada, visto que estudos mostravam o seu aumento linear com o aumento da carga de trabalho (BORG, 1982).

Diversas escalas foram testadas, mas a BSPE original utilizava um escore entre 6 e 20 pontos para avaliar o esforço percebido, os quais equivalem a um décimo da frequência cardíaca (60 a 200 batimentos por minuto). Essa escala está representada no Quadro 9.

Quadro 9 - Borg Scale of Perceived Exertion

6	
7	Muito leve
8	
9	Leve
10	
11	Relativamente leve
12	
13	Um pouco pesado
14	
15	Pesado
16	
17	Muito pesado
18	
19	Exaustivo
20	

Fonte: Adaptado de Borg (1970).

Como apresentado anteriormente a RPE de Borg foi construída para seguir a frequência cardíaca para o esforço em bicicleta ergométrica e para homens

saudáveis de meia-idade realizando um trabalho de moderado a difícil, de modo que a frequência cardíaca equivalia a cerca de 10 vezes o valor da RPE (BORG; NOBLE, 1974).

Essa escala tornou-se muito popular e foi traduzida para muitos idiomas. Além disso, muitos estudos mostraram altas correlações de avaliações e frequências cardíacas, além de altas correlações com outras variáveis fisiológicas (BORG, 1982).

Pouco tempo depois, Borg desenvolveu uma adaptação para a BSPE, modificando a magnitude e os escores, buscando uma maior facilidade para aplicação e principalmente melhor entendimento dos participantes para avaliação do esforço percebido. Essa escala adaptada possui escores de 0 a 10, sendo possível uma estimativa de magnitude livre acima de 10, representando um esforço percebido extremamente forte, conforme apresentado no Quadro 10.

Quadro 10 – BSPE modificada

0	Nenhum	
0.5	Extremamente fraco	Apenas perceptível
1	Muito fraco	
2	Fraco	Leve
3	Moderado	
4	Um pouco forte	
5	Forte	Pesado
6		
7	Muito forte	
8		
9		
10	Extremamente forte	Quase máximo
•	Máximo	

Fonte: Adaptado de Noble et al. (1983).

Comparativamente, a BSPE antiga é melhor para a maioria dos estudos simples aplicados de esforço percebido, para testes de esforço e para previsões e prescrições de intensidade de exercícios em esportes e reabilitação médica. Já a nova escala pode ser especialmente adequada para determinar outros sintomas subjetivos, como dificuldades respiratórias e dores (BORG, 1982).

A escala original de Borg (1970), descrita e usada primeiramente para mensurar o esforço percebido de performances físicas, pode ser usada para mensurar dificuldade percebida para performances mentais, que é o foco de pesquisa do presente estudo (BORG; NOBLE, 1974).

Weinger, Reddy e Slagle (2004) avaliaram a carga de trabalho de anestesistas por meio de técnicas fisiológicas (frequência cardíaca), subjetivas e

processuais (latência da resposta a uma luz de alarme e densidade da carga de trabalho). As medidas subjetivas foram feitas por meio de autoavaliação e pela classificação de um observador, utilizando a BSPE original, durante intervalos aleatórios de 7 a 15 minutos.

Para ajudar na ponderação da escala, os participantes foram explicitamente aconselhados de que a classificação típica da carga de trabalho no momento da intubação oral de rotina era de aproximadamente 12. A ordem de classificação solicitada pelo computador era primeiro o registro do observador e, em seguida, o do participante para sua autoavaliação. Os autores consideraram positivo o uso de múltiplas medidas de carga de trabalho, pois elas podem fornecer um perfil mais abrangente das demandas de trabalho de casos clínicos (WEINGER; REDDY; SLAGLE, 2004).

Schulz et al. (2011) analisaram a validade e a confiabilidade da RPE para um ambiente de simulação em escala real, utilizando cenários de indução de anestesia geral com e sem incidentes críticos.

Cada participante classificou sua carga de trabalho usando a BSPE após três sessões de simuladores diferentes, imediatamente após cada sessão. Para análise de validade e confiabilidade, utilizou-se análise de regressão univariada e multivariada e coeficiente de correlação de concordância. Além disso, mediu-se também a frequência cardíaca, pois supõe-se que a RPE está correlacionada com a frequência cardíaca dos participantes (SCHULZ et al., 2011).

Os resultados mostraram uma correlação significativa das pontuações da frequência cardíaca e da BSPE dos participantes. Ainda, a BSPE demonstrou refletir as diferenças esperadas na carga de trabalho e correlacionou-se estreitamente com as avaliações do observador, mostrando-se fácil de administrar e uma ferramenta válida para a avaliação subjetiva da carga de trabalho em configurações de simulador, possuindo confiabilidade moderada.

Como desvantagens, os autores colocaram o fato de ser utilizada uma versão alemã da RPE, o que pode limitar a comparabilidade dos resultados e as investigações usando a escala original em inglês. Além disso, indicaram o fato da ferramenta consistir em apenas uma escala com uma medida global de esforço percebido, não sendo capaz de distinguir os vários aspectos da carga de trabalho. Nesse caso, indicaram o NASA-TLX como sendo superior, pois é uma abordagem multidimensional e inclui seis categorias de carga de trabalho (SCHULZ et al., 2011).

Martin et al. (2016) utilizaram medidas da frequência cardíaca e avaliações subjetivas para avaliar a carga de trabalho associada à fase de indução, manutenção e emergência da anestesia. Como medidas subjetivas foram utilizadas a BSPE, com escore entre 6 e 20 pontos, e as seis escalas do NASA-TLX. Apesar do foco do trabalho ser nas medidas fisiológicas, as avaliações da escala de Borg diferiram significativamente entre os estágios da anestesia.

Já Leedal e Smith (2005), em seu estudo, relataram que a escala original de Borg geralmente se correlaciona intimamente com as taxas cardíacas reais. Além disso, citam como vantagem o fato da BSPE ser uma medida global, ao contrário de algumas escalas subjetivas que não levam em conta todas as demandas.

4.2.11 Job Demands Scale (JDS)

A *Job Demands Scale* (JDS), ou, em português, Escala de Demandas de Trabalho é uma forma subjetiva de avaliar a carga de trabalho mental, utilizando uma escala multidimensional.

Esse método foi desenvolvido inicialmente por Karasek, em 1979, buscando um modelo de gerenciamento do estresse no trabalho, prevendo a tensão mental, resultado da interação entre as demandas de trabalho e a latitude da decisão de trabalho, diferente dos modelos anteriores que as utilizavam de forma separada (PEKKARINEN et al., 2013). Assim, Karasek (1979) objetivou, com a construção dessa escala, medir os estressores psicológicos envolvidos na carga de trabalho, os quais são relacionados a tarefas inesperadas e aos conflitos pessoais relacionados ao trabalho.

Para isso, Karasek (1979) considerou duas variáveis independentes: as demandas de trabalho, que medem as fontes de estresse, e a latitude de decisão, que trata do controle do trabalho. Por meio delas, acontece a tensão de trabalho, a qual ocorre quando a demanda de trabalho é alta e a latitude de decisão é baixa. Além disso, essa mesma combinação também está associada à insatisfação no trabalho.

No seu modelo criado para os Estados Unidos, Karasek (1979) considerou dentro da variável independente “latitude de decisão”, duas subescalas: i. Descrição de habilidade e ii. Autoridade de decisão. A subescala descrição de habilidades é construída de quatro itens: alto nível de habilidade requerido, necessidade de

aprender coisas novas, trabalho não repetitivo e criatividade exigida. Já a subescala autoridade de decisão é composta de outros quatro itens: liberdade sobre como realizar o trabalho, permissão para tomar decisões, possibilidade de auxiliar na própria decisão e ter muito a dizer no trabalho.

Já a variável independente “demandas de trabalho” possui sete itens: requer trabalho rápido, requer trabalhar difícil, muito trabalho a ser feito, não há tempo suficiente, trabalho excessivo, não há tempo para finalizar o trabalho e demandas conflitantes (KARASEK, 1979).

Os itens de cada uma das variáveis independentes são transformados em perguntas, por exemplo “Seu trabalho exige que você trabalhe muito rápido?” ou “Você tem pouco tempo para realizar o seu trabalho?” e essas perguntas são respondidas em uma escala de 4 pontos, com 1 significando “discordo totalmente” e 4 significando “concordo totalmente”. Pode-se calcular o escore médio para cada escala, por meio da avaliação dada pelos participantes a cada item, e também o escore geral da JDS.

Essa escala foi testada e validada nos Estados Unidos e também foi desenvolvida e testada uma escala similar na Suécia. Com esses estudos, Karasek (1979) encontrou algumas limitações: o impacto das relações sociais no nível de grupo e organizacional não é considerado, o impacto específico de diferentes tipos de latitude de decisão e demandas de trabalho não é distinguido, o impacto das diferenças individuais na percepção das demandas de trabalho não é avaliado e os indicadores apresentam variações conforme os diferentes tipos de trabalho.

Posteriormente, Karasek et al. (1998) desenvolveram um questionário de Conteúdo de Trabalho (*Job Content Questionnaire – JCQ*), como uma ferramenta autoadministrada para avaliação das características sociais e psicológicas do trabalho, mais abrangente que a JDS, utilizando cinco escalas de avaliação: (i) latitude de decisão, (ii) demandas psicológicas, (iii) apoio social, (iv) demandas físicas e (v) insegurança no trabalho. Esse instrumento tem um comprimento recomendado de 49 perguntas, possuindo 5 partes, conforme cada uma das escalas acima nomeadas.

Como colocado anteriormente por Karasek et al. (1979), a tensão no trabalho é decorrente da latitude de decisão e das demandas psicológicas, e assim as escalas (i) e (ii) podem envolver a carga de trabalho mental, da mesma forma que a ferramenta JDS, que também utiliza essas escalas independentes.

No JCQ, a escala que trata da latitude de decisões contém seis componentes, totalizando dezenove questões. Esses componentes são: descrição de competências, autoridade de decisão, subutilização de competências, autoridade de decisão de grupo de trabalho, autoridade formal e influência sindical/representativa (KARASEK et al., 1998).

Já a escala das demandas psicológicas e carga mental possui quatro componentes, totalizando nove questões. Os quatro componentes são: demandas psicológicas gerais, ambiguidade do papel, concentração e perturbação do trabalho mental (KARASEK et al., 1998).

O JCQ foi projetado para autoadministração e tem sido muitas vezes incluído como uma seção em outros questionários, nos quais é incluída uma pequena frase introdutória sobre como responder às perguntas. Para a versão completa recomendada estima-se um tempo de conclusão de aproximadamente quinze minutos. Além disso, as questões do JCQ utilizam linguagem simples, buscando o entendimento pelos participantes em todos os níveis de ensino, e uma escala de 4 pontos para avaliação, conforme utilizado na JDS, facilitando a ponderação quantitativa das questões (KARASEK et al., 1998).

Essa ferramenta foi testada em seis estudos conduzidos em quatro países, sendo autoadministrado na maioria dos estudos ou administrada durante uma entrevista face a face, geralmente no local de trabalho, buscando comprovar a consistência das escalas do JCQ em diferentes culturas de trabalho e gênero.

Por meio desses estudos foram encontradas semelhanças substanciais nas médias, desvios padrões e correlações entre as escalas, e nas correlações entre escalas e variáveis demográficas, tanto para homens quanto para mulheres. Além disso, foi encontrada boa confiabilidade para a maioria das escalas e os resultados sugeriram que as características do trabalho psicológico são mais semelhantes entre os estudos realizados nos Estados Unidos do que nas outras fronteiras (KARASEK et al., 1998).

Apesar dos resultados positivos encontrados, a escala de demanda psicológica foi criticada por falhar em cobrir adequadamente o trabalho emocional na interação social e cargas de trabalho cognitivas. Porém, para Karasek et al. (1998), há evidências de que a versão da escala JCQ recomendada por nove perguntas é uma melhoria para as cargas de trabalho cognitivas e a correlação é maior com a descrição de habilidades, que também avalia o envolvimento cognitivo.

Ainda em relação aos resultados, Karasek et al. (1998) encontraram associações consistentes entre tensão mental e as escalas do Jcq, mas os efeitos diferenciais das características do trabalho foram observados: exaustão e *burnout* estão mais associadas a altas demandas psicológicas, enquanto depressão e ansiedade estão mais fortemente associadas à baixa latitude de decisão.

Pekkarinen et al. (2013) examinaram se os recursos do trabalho, como controle do trabalho, apoio social e justiça distributiva, moderam as associações de altas demandas de trabalho, induzidas pela carga de trabalho física e mental, com sintomas musculoesqueléticos entre enfermeiros geriátricos.

Com esse propósito, os autores examinaram a carga de trabalho como indicada pelas características estruturais e residentes (dependência física, deficiências cognitivas e problemas comportamentais) e as demandas de trabalho percebidas induzidas pela carga de trabalho física e mental em relação às queixas musculoesqueléticas.

A carga de trabalho mental, interesse do presente estudo, foi medida utilizando três itens da Escala de Demandas de Trabalho, pois, para os autores, tais escalas de demandas parciais de trabalho avaliam os mesmos conceitos subjacentes que a escala completa das demandas de trabalho psicológico, que reflete o quanto os funcionários trabalham com afinco. Essa escala de três itens indagou sobre as demandas mentais do trabalho em termos de carga de trabalho excessiva, trabalho duro e tempo insuficiente para trabalhar (PEKKARINEN et al., 2013).

A ponderação dos itens da escala aconteceu por meio de entrevista, na qual os participantes escolheram um escore entre 1 (discordo totalmente) e 5 (concordo totalmente). Foi calculado o escore médio para a escala e quanto maior o resultado, maior era a indicação de carga de trabalho mental.

Ainda foi utilizada uma escala de seis itens, retirada da JSD, para medir a descrição de habilidade, com os itens necessidade de aprender coisas novas, capacidade de desenvolver novas habilidades, alto nível de habilidade requerido, variedade de tarefas, trabalho não repetitivo e criatividade exigida. Uma escala com três itens também foi utilizada para medir a latitude de decisão: permissão para tomar decisões, liberdade sobre como realizar o trabalho e ter muito a dizer no trabalho. Ambas foram utilizadas para avaliar o controle do trabalho e utilizaram o mesmo esquema de pontuação acima, com escore entre 1 e 5 para cada item e escore médio para cada escala (PEKKARINEN et al., 2013).

Balducci, Avanzi e Fraccaroli (2016) estudaram o *workaholismo*, um vício voltado a viver para o trabalho ou trabalhar compulsivamente, buscando relações com as experiências afetivas no trabalho, demandas de trabalho e os efeitos a longo prazo da saúde mental.

Esse estudo foi conduzido em uma agência nacional de serviços de saúde italiana, por meio da coleta de questionários anônimos e autorrelatados, aplicados durante o horário de trabalho, utilizando de profissionais de saúde e funcionários administrativos e da limpeza. Para medir o *workaholismo* foi utilizada uma versão de 10 itens do *Dutch Work Addiction Scale* (DUWAS) e o sofrimento mental foi avaliado pela versão de doze itens do *General Health Questionnaire* (GHQ-12).

As demandas de trabalho foram medidas por meio de cinco itens do JCQ da escala de demandas de trabalho, os quais foram avaliados pela pontuação original do método, ou seja, score entre 1 e 4 (BALDUCCI; AVANZI; FRACCAROLI, 2016).

Tesi, Aiello e Giannetti (2018) exploraram como as demandas de trabalho e o bem-estar psicológico dos assistentes sociais estão relacionados ao envolvimento no trabalho.

Com esse propósito, utilizaram duas escalas do JCQ para medir as demandas de trabalho psicológico e físico, utilizando a pontuação original. A escala de demandas psicológicas no trabalho, com nove itens, foi utilizada para medir a carga de trabalho mental e a escala de demandas físicas no trabalho, com cinco itens, mediu a carga de trabalho física. Ainda, a carga emocional foi medida utilizando uma escala de demandas emocionais específica, composta por seis itens, em uma escala Likert de cinco pontos, variando de 1 (nunca) a 5 (sempre) (TESI; AIELLO; GIANNETTI, 2018).

Para trabalhos futuros, Tesi, Aiello e Giannetti (2018) sugeriram a consideração de outras facetas das demandas de trabalho, como o conflito de funções e a insegurança no emprego, pois nesse estudo foram consideradas apenas três tipos de demandas de trabalho: psicológicas, físicas e emocionais.

Já Vilas Boas e Cerqueira (2017) averiguaram o JCQ como uma das ferramentas mais importantes para avaliar os fatores psicossociais e organizacionais subjacentes ao estresse no trabalho, buscando também discutir a versão português do JQC.

Em relação as escalas do JCQ, Vilas Boas e Cerqueira (2017) discutiram sobre a divisão da latitude de decisão em duas dimensões e sobre a adequação dos

itens da escala exigências psicológicas do trabalho, utilizando análise fatorial exploratória para validar.

Com a permissão de Karasek, os autores traduziram 22 itens da versão completa das escalas JCQ para o português, utilizando retrotradução, medindo quatro escalas: latitude de decisão, demandas psicológicas de trabalho, suporte do supervisor e suporte dos colegas de trabalho.

Todas as escalas utilizaram a classificação original, baseada em intensidade, por meio de uma escala Likert de quatro pontos, variando de 1, ou “discordo totalmente”, a 4, ou “concordo totalmente”. Além disso, como os questionários são autoadministrados, eles foram distribuídos diretamente aos participantes ou por seus gerentes, sendo que cada questionário continha uma carta em anexo explicando seus objetivos e garantindo a confidencialidade (VILAS BOAS; CERQUEIRA, 2017).

Em relação aos resultados encontrados, a correlação entre escalas tendeu a ser fraca, mas significativa, o que indicou que são conceitos bastante independentes, mas existe uma relação entre eles, ou seja, há apenas uma mínima sobreposição no que as escalas estão medindo.

De forma geral, Vilas Boas e Cerqueira (2017) encontraram validade aceitável e comparabilidade substancial das escalas JCQ, mostrando que a versão portuguesa do JCQ é bastante fiel e válida para avaliar as condições de trabalho psicossocial dos trabalhadores portugueses, podendo ajudar a avaliar adequadamente o estresse no trabalho em Portugal.

Como proposta para melhorar a consistência interna da escala e ter uma estrutura fatorial mais forte, Vilas Boas e Cerqueira (2017) recomendaram a exclusão de três itens: trabalho repetitivo, trabalho árduo e sem trabalho excessivo. Dessa forma, como a escala de demandas de trabalho psicológico teria apenas três itens, torna-se importante descobrir novos itens que melhor avaliem as restrições psicológicas que caracterizam o local de trabalho.

Assim, com a retirada dos três itens, os autores encontraram problemas de consistência interna apenas na escala exigências psicológicas, porém os resultados sugerem que as propriedades psicométricas da versão portuguesa do JCQ são satisfatórias, embora seja necessária melhorar a escala de exigências psicológicas (VILAS BOAS; CERQUEIRA, 2017).

De forma geral, em relação aos métodos JDS e JQC, originado do JDS por meio da agregação de mais escalas ou variáveis independentes, apesar deles serem utilizados principalmente para determinar as demandas de trabalho, eles também podem ser utilizados para analisar a carga de trabalho mental dos trabalhadores, como utilizado em artigos apresentados acima, por tratarem de demandas basicamente mentais ou psicológicas. Além disso, possuem escalas de fácil ponderação e entendimento, já validadas em muitos países, e aplicadas em diferentes situações de trabalho.

4.2.12 Escala Subjetiva de Carga de Trabalho Mental (ESCAM)

A *Escala Subjetiva de Carga Mental de Trabajo (ESCAM)*, *Subjective Mental Workload Scale*, ou, em português, Escala Subjetiva de Carga de Trabalho Mental, é uma escala multidimensional e subjetiva, utilizada para avaliar a carga mental percebida do trabalho de forma geral e não especificamente em uma tarefa ou em momentos particulares no desempenho de uma atividade, como acontece com outros métodos subjetivos (ROLO; HERNÁNDEZ; DÍAZ, 2010).

A ESCAM foi desenvolvida por Rolo, Díaz e Hernández, em 2009, na Espanha, buscando suprir limitações existentes nos modelos subjetivos até então desenvolvidos, como, por exemplo, procedimento complexo, aplicação apenas em um setor específico de trabalhadores, procedimento de medida não especificado, método intrusivo ou utilização da carga mental apenas como um fator dentro de avaliações mais abrangentes de condições de trabalho e ambiente.

Para desenvolver esse instrumento, Rolo, Díaz e Hernández (2009) utilizaram revisão de literatura e análise dos métodos já existentes para avaliar a carga de trabalho mental. Além disso, os autores consideraram três grandes dimensões: demandas temporais da tarefa, recursos de processamento que a tarefa demanda e aspectos de natureza emocional.

Assim, Rolo, Díaz e Hernández (2009) primeiramente desenvolveram uma escala com cinco variáveis:

1. Quantidade e complexidade das informações utilizadas pelo trabalhador no desenvolvimento do seu trabalho;

2. Exigências cognitivas do trabalho realizado, referindo-se à atenção, concentração e memorização;
3. Organização temporária do trabalho, incluindo aspectos relacionados ao tempo requerido, tempo disponível, intervalos e ritmo de trabalho;
4. Características das tarefas executadas, em relação ao número de tarefas, complexidade das mesmas, interrupções, atrasos, recuperação de atrasos e consequências de erros;
5. Consequências para a saúde derivadas de um desequilíbrio na carga de trabalho mental.

Por último, foram consideradas as características individuais dos participantes (idade, sexo, estado civil, qualificação), dados sobre o trabalho (cargo, tempo de serviço, experiência exigida), condições ambientais (iluminação, ruído, temperatura) e condições organizacionais em que o trabalhador está inserido (ROLO; DÍAZ; HERNÁNDEZ, 2009).

Assim, foi desenvolvida a ESCAM, contendo 31 itens, os quais deveriam ser analisados por meio de uma escala Likert de cinco pontos, contendo as cinco dimensões citadas acima e informações referentes às características sociodemográficas e dados sobre o trabalho e as condições ambientais.

Buscando validar o instrumento, Rolo, Díaz e Hernández (2009) realizaram a aplicação da ESCAM em uma amostra contendo trabalhadores e estudantes universitários e após realizar análises da consistência interna chegaram em um instrumento com cinco fatores, totalizando 20 itens, conforme mostrado no Quadro 11, dando origem a Escala Subjetiva de Carga de Trabalho Mental.

Com a ESCAM final, a avaliação de cada item continuou consistindo em um escore de 1 a 5 pontos, com 1 indicando o menor nível de carga mental possível e 5 o máximo valor possível.

A avaliação de cada um dos cinco fatores é feita de forma individual, por meio da média aritmética simples dos itens que os compõe. Ainda, pode-se ter uma escala geral de carga de trabalho mental, por meio da média dos resultados das cinco dimensões. Valores mais baixos indicam menor percepção de carga mental (ROLO; DÍAZ; HERNÁNDEZ, 2009).

Quadro 11 - Fatores e itens contidos na ESCAM

FATORES E ITENS	ESCALA LIKERT				
FATOR 1. DEMANAS COGNITIVAS E COMPLEXIDADE DAS TAREFAS					
1. O nível de esforço ou concentração mental que meu trabalho requer é	1	2	3	4	5
2. A quantidade de memorização de informações e materiais que meu trabalho requer é	1	2	3	4	5
3. O grau de complexidade da informação que devo usar no meu trabalho é	1	2	3	4	5
4. O nível de esforço mental necessário para evitar erros no meu trabalho é	1	2	3	4	5
5. O nível de ambiguidade das decisões a tomar no meu trabalho é	1	2	3	4	5
6. Normalmente, no meu local de trabalho, o número de decisões que devo tomar é	1	2	3	4	5
FATOR 2. CONSEQUÊNCIAS PARA A SAÚDE					
7. No final do dia de trabalho me sinto exausto	1	2	3	4	5
8. Eu me sinto exausto quando me levanto de manhã e tenho que enfrentar outro dia de trabalho	1	2	3	4	5
9. O cansaço que meu trabalho produz é	1	2	3	4	5
10. Eu tenho dificuldades para relaxar depois do trabalho	1	2	3	4	5
FATOR 3. CARACTERÍSTICAS DA TAREFA					
11. O número de interrupções (telefonemas, atendimento ao público, outros colegas solicitando informações, etc.) durante a execução do meu trabalho é	1	2	3	4	5
12. A quantidade de dificuldades que ocorrem quando novos procedimentos de trabalho ou programas de computador são introduzidos é	1	2	3	4	5
13. No meu trabalho, eu tenho que fazer mais de uma tarefa de cada vez	1	2	3	4	5
14. As tarefas que realizo no meu trabalho exigem uma alta concentração devido à quantidade de distração ou ruído de fundo	1	2	3	4	5
FATOR 4. ORGANIZAÇÃO TEMPORAL					
15. O tempo atribuído a cada uma das tarefas que faço é	1	2	3	4	5
16. O tempo que tenho para fazer o meu trabalho é	1	2	3	4	5
17. O tempo que tenho para tomar as decisões exigidas pelo meu trabalho é	1	2	3	4	5
FATOR 5. RITMO DE TRABALHO					
18. É possível variar o meu ritmo de trabalho sem atrapalhar o desempenho do meu departamento	1	2	3	4	5
19. Além dos intervalos regulares, o trabalho permite que eu faça pausas quando precisar	1	2	3	4	5
20. No meu trabalho, posso cometer algum erro sem ter um impacto crítico nos resultados do trabalho	1	2	3	4	5

Fonte: Adaptado de Rolo, Díaz e Hernández (2009).

Com esse trabalho, Rolo, Díaz e Hernández (2009) validaram a ESCAM por meio da comparação dos perfis de carga mental de diferentes postos de trabalho, mostrando não existirem praticamente diferenças significativas.

Em outro trabalho, Rolo, Hernández e Díaz (2010) exploraram a relação entre as condições físico-ambientais do posto de trabalho e a carga mental percebida em trabalhadores de escritório.

Para isso, os autores utilizaram cinco variáveis independentes para as condições ambientais percebidas (adequação da temperatura, iluminação, ruído, distribuição espacial e condições de higiene) e os cinco fatores da ESCAM para avaliar a carga mental subjetiva (demandas cognitivas e complexidade da tarefa,

consequências para a saúde, características da tarefa, organização temporal e ritmo do trabalho), utilizando todos os 20 itens e a pontuação original por meio da escala Likert de cinco pontos, a qual também foi utilizada para avaliar as condições ambientais. Além disso, foram solicitados dados sociodemográficos e informações sobre o trabalho.

Os participantes voluntários responderam aos questionários em seu horário de trabalho, individualmente e anonimamente, demorando cerca de 15 minutos. Os resultados da ESCAM foram apresentados por meio da média de cada um dos cinco fatores e também por meio de um escore total da carga mental, conforme o instrumento original, dividindo os participantes em dois grupos: baixa e alta carga mental percebida (ROLO; HERNÁNDEZ; DÍAZ, 2010).

Ceballos Vásquez et al. (2014) avaliaram as características psicométricas da ESCAM em funcionários universitários, em sua maioria da área da saúde, buscando validade a partir do Questionário de Avaliação de Risco Psicossocial (SUSES-ISTAS 21), instrumento já validado.

Para avaliar a confiabilidade e a validade da ESCAM foram utilizados o alfa de Cronbach, a consultoria a especialistas na área de psicologia ocupacional, ergonomia e enfermagem do trabalho, a análise fatorial exploratória dos principais componentes com rotação ortogonal e a análise de correlação entre as dimensões da ESCAM e do SUSES-ISTAS 21.

A aplicação da ESCAM e a apresentação dos resultados seguiram a escala original, apresentando uma escala confiável e válida, com alta correlação com o SUSES-ISTAS 21, mostrando a possibilidade de empregar a ESCAM para avaliar a CTM. Como vantagens do método, Ceballos Vásquez et al. (2014) indicaram o pequeno tempo para preenchimento, a facilidade de entendimento pelos participantes e a escala multidimensional e, como desvantagem, o fato de não possuir muitos resultados para comparar, devido ao pouco tempo de existência do instrumento.

Em outro estudo, Ceballos Vásquez et al. (2016) também estudaram a confiabilidade das características psicométricas da ESCAM, dessa vez com trabalhadores de saúde de hospitais chilenos, utilizando novamente o SUSES-ISTAS 21 para validação. Os instrumentos foram aplicados por meio de entrevista, que duraram entre 15 e 20 minutos.

Os resultados apresentaram a ESCAM adequada para avaliar a carga mental, com confiabilidade e validade utilizando uma amostra ampla e representativa

de diferentes grupos profissionais e a comparação com um instrumento previamente validado (CEBALLOS VÁSQUEZ et al., 2016).

Como limitações, Ceballos Vásquez et al. (2016) colocaram que seria aconselhável usar a versão média do SUSESO-ISTAS 21 e que o tamanho dos grupos profissionais não era equivalente, o que pode ter influenciado os resultados obtidos.

Ainda, como sugestões para pesquisas futuras, Ceballos Vásquez et al. (2016) indicaram uma revisão dos itens ou a inclusão de outros em fatores como a carga emocional, buscando melhor consistência interna dessas dimensões, e também seria interessante explorar profissionais de saúde pertencentes a diferentes unidades hospitalares.

Sáenz et al. (2016) estudaram a CTM em profissionais de enfermagem, por meio da ESCAM. A aplicação do questionário foi feita 15 minutos antes do final do turno correspondente por três dias consecutivos. Os resultados mostraram a validade do instrumento e a satisfação nos resultados apresentados (SÁENZ et al., 2016).

Já Rivera e Romero (2016) avaliaram a relação existente entre a carga de trabalho mental percebida pelos funcionários de uma indústria automobilística e o trabalho realizado, por meio da ESCAM aplicada no local de trabalho.

Os resultados foram apresentados por meio do escore das cinco dimensões e o instrumento apresentou como vantagens a análise rápida na identificação das áreas a serem melhoradas e o fato da ESCAM ser respondida no local de trabalho, rompendo a lacuna entre a parte teórica e prática, obtendo informações em contextos reais (RIVERA; ROMERO, 2016).

De forma geral, as vantagens encontradas na Escala Subjetiva de Carga de Trabalho Mental são muitas: baixa invasividade, fácil implementação, boa aceitação dos avaliadores, baixo custo, pouco tempo de avaliação, fácil compreensão, itens conectados com situações típicas de trabalho, solidez do instrumento e uma perspectiva holística da carga de trabalho mental abordando características das tarefas (ROLO; DÍAZ; HERNÁNDEZ, 2009; SÁENZ et al., 2016).

Porém, uma desvantagem que pode ser associada a este tipo de escala é a possibilidade de vieses de resposta devido à desejabilidade social, ou seja, respostas socialmente aceitas que não representam necessariamente a realidade do participante (ROLO; DÍAZ; HERNÁNDEZ, 2009).

4.2.13 Global Subjective Workload Scale (GSW)

A *Global Subjective Workload Scale* (GSW), ou, em português, Escala Global de Carga de Trabalho Subjetiva, é um método subjetivo utilizado para avaliar a carga de trabalho mental, aplicado por meio de uma escala unidimensional.

Mouzé-Amady et al. (2013) utilizaram ela, juntamente com o NASA-TLX, como critério de agregação para ponderação das seis escalas do NASA-TLX. A GSW consiste em uma pergunta: Quanto foi a sua carga de trabalho global para executar a tarefa? Ela é respondida por meio de um escore entre 0 (baixo) e 10 (alto).

Mouzé-Amady et al. (2013) acreditavam existir limitações na ponderação das escalas do NASA-TLX e por isso propuseram um modelo Fuzzy-TLX para essa ponderação, sem utilizar o método completo desenvolvido por Hart e Staveland (1988), baseado em critérios múltiplos de agregação, o qual necessita de pelo menos uma medida global, que pode ser uma classificação subjetiva ou objetiva global, desde que esteja no mesmo intervalo que as dimensões do NASA-TLX.

Em seu estudo, os autores estudaram os dois modelos de agregação: as medidas da frequência cardíaca, como medida objetiva, e os valores da GSW, como medida subjetiva, buscando a comparação entre ambos e também a comparação com o método NASA-TLX padrão. Para isso, o NASA-TLX utilizou a escala para as suas seis dimensões com escore entre 0 e 10 e as escalas foram preenchidas após cada tarefa, por meio de um computador ou em papel.

Comparando-se o método NASA-TLX padrão e os modelos Fuzzy-TLX, independente da forma de agregação utilizada, com o modelo Fuzzy-TLX é possível coletar informações valiosas sobre a carga de trabalho mental, sendo melhor fundamentado matematicamente e tratando qualitativamente as medidas subjetivas de carga de trabalho, utilizando apenas um critério externo para determinar os pesos das dimensões do NASA-TLX e permitindo uma coleta de dados menos demorada e compatível com os ambientes reais de trabalho (MOUZÉ-AMADY et al., 2013).

Comparando-se os modelos Fuzzy, utilizando medidas cardíacas e utilizando a GSW, o modelo com as medidas cardíacas forneceu um efeito significativo em relação as tarefas e fatores de idade, diferente da GSW, mostrando-se mais poderoso para destacar efeitos variáveis em uma amostra de dados que uma agregação subjetiva.

De forma geral, Mouzé-Amady et al. (2013) demonstraram que o modelo Fuzzy-TLX, independentemente do critério de agregação utilizado, possibilita coletar informações valiosas sobre a carga de trabalho subjetiva em vários cenários com confiabilidade, avaliando a CTM em locais onde não se pode utilizar o NASA-TLX original, sendo possível utilizar a GSW como critério de agregação.

4.2.14 Simple Objective Index (SOI)

O *Simple Objective Index* (SOI), ou, em português, Índice Objetivo Simples, é uma forma de avaliar a carga de trabalho mental por meio de um fator objetivo, ou seja, não dependente da mensuração por uma pessoa, utilizando apenas uma resposta direta.

Essa metodologia é muito utilizada para controladores de tráfego aéreo, visto que esse é um trabalho que possui muitas demandas cognitivas e o uso de um índice objetivo não atrapalha o desenvolvimento das atividades, sendo não intrusivo. Para isso, muitos estudos apresentam o número de aeronaves (N) controladas pelo ATC como um fator para mensurar a CTM.

Sperandio (1970), em seu estudo na França, analisou a influência do aumento do tráfego aéreo nos processos de dados realizados pelos ATCs, os quais podem impactar na carga de trabalho. Dessa forma, analisou a carga de trabalho conforme a densidade de tráfego aéreo, a qual foi mensurada pelo índice N.

Costa (1993) examinou controladores de tráfego aéreo por três turnos de trabalho sucessivos, utilizando o registro do número de aeronaves sob controle por hora como índice de carga de trabalho. Além do indicador N, utilizado como um SOI, a CT também foi avaliada por meio da classificação de um observador experiente em uma escala de 1 a 16 em cada hora de trabalho do ATC e subjetivamente pelo próprio ATC em uma escala de 0 a 100 ao final do turno de trabalho.

Manning et al. (2001) utilizaram o indicador N para calcular a carga de tarefa dos controladores de tráfego aéreo, por meio do método *Performance and Objective Workload Evaluation Research* (POWER), o qual possui como um dos seus fatores o número de aeronaves controladas pelo ATC. Esse método foi correlacionado com as medidas NASA-TLX e ATWIT, ferramentas já validadas, para avaliar a carga de trabalho mental, buscando a validação do POWER. Assim, os autores

encontraram relações entre N e a CT do NASA-TLX voltada para a frustração e também entre N e a ATWIT, sinalizando que um aumento no número de aeronaves controladas aumentava o valor da ATWIT, confirmando a hipótese que o fator N, no método POWER, está relacionado com a carga de trabalho mental.

Pfeiffer et al. (2015) acreditam que existem inúmeros fatores envolvidos nos aspectos emocionais e estressores, os quais fazem parte da CTM, de um controlador de tráfego aéreo e um deles é a alta carga de tráfego, a qual tem como fator crucial o número de aeronaves sob comando. Além disso, no momento do estudo, as decisões dos supervisores sobre dividir um setor eram feitas consultando as previsões de carga de trabalho, principalmente com base no número esperado de aeronaves e nas demandas dos controladores de tráfego aéreo.

Averty et al. (2004), propuseram um novo método que utiliza o número de aeronaves como um fator para avaliar a carga de trabalho dos ATCs, que leva o nome de TLI e está explicitado no tópico 4.2.22, e utilizaram comparações com o índice objetivo simples N e com o NASA-TLX. Assim, os resultados dos autores mostraram que o TLI possui uma maior correlação com o NASA-TLX do que o N, mostrando que o índice N isoladamente não reflete exatamente a carga de trabalho percebida, pois precisa levar em consideração uma carga adicional, dependente das trajetórias e condições de cada aeronave (AVERTY et al., 2004).

De forma semelhante, Athènes et al. (2002) colocaram que o número de aeronaves controladas era o dado objetivo mais utilizado para estimar a carga de trabalho do ATC, devido a facilidade de gravação, porém acreditam que existem problemas relacionados ao uso desse indicador, visto que como as aeronaves estão espalhadas pelo espaço e tempo, o índice pode ser fortemente distorcido, pois a posição e trajetória das aeronaves também devem ser levadas em consideração. Assim, acreditam que do ponto de vista real não é possível classificar cada aeronave com o mesmo peso, o que é feito com o uso do índice objetivo N.

Seguinte a mesma linha de pensamento, Pfeiffer et al. (2015) acreditam que a carga de trabalho vai além do número de aeronaves controladas, dependendo da estrutura do espaço aéreo e rotas padrão, bem como das direções das aeronaves.

4.2.15 Instantaneous Self-Assessment (ISA)

O método *Instantaneous Self-Assessment* (ISA), ou, em português, Autoavaliação Instantânea, é uma forma subjetiva de avaliar a carga de trabalho mental, por meio de uma escala unidimensional, durante a realização das tarefas.

Esse método foi desenvolvido pela Autoridade de Aviação Civil do Reino Unido (NATS) para fornecer avaliações subjetivas da carga de trabalho durante tarefas de controle de tráfego aéreo e utilizado primeiramente nos estudos de Brennan (1992) e Jordan (1992) (TATTERSALL; FOORD, 1996).

O método ISA original consiste na reposta dos controladores quando um sinal visual é apresentado em intervalos durante a tarefa, a qual é respondida utilizando um conjunto de cinco teclas para representar diferentes níveis de carga de trabalho percebida, conforme apresentado no Quadro 12. O aviso visual é apresentado no console do radar e o teclado de resposta é colocado em um local conveniente ao lado do teclado do controlador (TATTERSALL; FOORD, 1996).

Quadro 12 - Descrição da escala ISA

Nível	Carga de trabalho	Capacidade restante	Descrição
1	Subutilizada	Muitíssima	Nada para fazer. Entediante.
2	Relaxada	Ampla	Mais do que suficiente o tempo para todas as tarefas. Atividades das tarefas do ATC ocupam menos de 50% do tempo.
3	Confortável	Alguma	O controlador tem trabalho suficiente para se manter estimulado; todas as tarefas estão sob controle. Ritmo ocupado, mas estimulante. Poderia ficar continuamente neste nível.
4	Alta	Muito pouca	Certas tarefas não essenciais são adiadas; o ACT não consegue trabalhar nesse nível por muito tempo; o controlador está funcionando "no limite"; o tempo passa rápido.
5	Excessiva	Nenhuma	Algumas tarefas não estão concluídas; o controlador está sobrecarregado e não se sente no controle.

Fonte: Adaptado de Kirwan et al. (1997) e Wilson et al. (2017).

Tattersall e Foord (1996) avaliaram a sensibilidade da ISA às variações na carga de trabalho de uma tarefa de rastreamento comparando com avaliações subjetivas retrospectivas (NASA-TLX), medidas fisiológicas (frequência cardíaca) e desempenho da tarefa sob diferentes níveis de dificuldade primária e também

avaliaram os efeitos no desempenho de rastreamento de duas modalidades diferentes de reposta ISA: resposta falada ou manual.

Os sinais visuais da ISA foram apresentados em média a cada dois minutos, como um sinal aleatório, e as respostas foram feitas usando um teclado com cinco botões, cada um rotulado para representar um nível de carga de trabalho diferente (5 = excessiva, 4 = alta, 3 = confortável, 2 = relaxada, 1 = subutilizada), ou dando respostas numéricas faladas equivalentes. Os botões possuíam doze milímetros quadrados.

Tattersall e Foord (1996) utilizaram correlação de Pearson entre as diferentes medidas de carga de trabalho. Como resultados, os autores encontraram que as classificações da ISA estavam correlacionadas significativamente com as classificações do NASA-TLX, variabilidade da frequência cardíaca e desempenho da tarefa, mostrando consistência. Além disso, encontraram relação entre o desempenho e as classificações da ISA, mostrando que um aumento nas classificações de carga de trabalho estava associado ao aumento do erro de rastreamento.

Os autores consideraram a ISA sensível ao nível de dificuldade da tarefa, mostrando que a ferramenta é útil na avaliação da carga de trabalho, mesmo com evidências de alguma interferência na tarefa principal de rastreamento, pois pode existir uma competição por recursos de atenção entre a tarefa principal e a tarefa de responder a ISA.

Kirwan et al. (1997) acreditam que a medida de carga mental mais utilizada em simulações em tempo real é a técnica ISA, a qual foi projetada para ser rápida e discreta, buscando a menor intervenção possível no desempenho da tarefa principal. Para eles, os registros obtidos a partir dessa medição permitem avaliar novos equipamentos, procedimentos e atributos do sistema, por meio dos níveis de carga de trabalho percebidos. Em seu estudo, a ISA era apresentada a cada dois minutos durante a realização das tarefas, por meio de um piscar de luzes nos botões do teclado de avaliação, as quais paravam de piscar quando o participante pressionava o botão correspondente ao nível de carga de trabalho que estava experimentando naquele momento específico.

Para eles, é possível o acompanhamento em tempo real da carga de trabalho percebida, permitindo que o analista dos dados detecte alguma situação diferente. Isso também permite que os controladores sejam interrogados após o

evento, enquanto as informações relevantes ainda estão atualizadas (KIRWAN et al., 1997).

Edwards et al. (2017) utilizaram uma simulação de controle de tráfego aéreo em rota de alta visibilidade, em trânsito, para investigar a relação entre carga de trabalho, percepção da situação e desempenho do controlador.

Para avaliar a CTM os participantes responderam a ISA a cada três minutos, a qual foi apresentada no topo do escopo do radar. Essa escala sofreu alteração da sua escala original, possuindo score entre 1 e 6, buscando reduzir o viés de resposta em torno do ponto médio da escala. Além disso, a medida foi escolhida por ser breve e menos intrusiva que outras medidas, sendo possível sua aplicação sem interromper o experimento.

Edwards et al. (2017) acreditam que apesar da ISA ser breve e não intrusiva, ela pode apresentar respostas imprecisas dos participantes, como qualquer método subjetivo. Assim, podem ser utilizadas medidas objetivas ou psicológicas para relacionar a ISA, buscando uma exploração adicional.

Rodríguez et al. (2015) desenvolveram um novo modelo para avaliar a CTM e utilizaram a ISA, e também medidas pupilares, para validar esse novo método. Nesse caso, as respostas dos ATCs em relação a ISA foram registradas a cada dois minutos e meio.

Lemercier et al. (2014) estudaram o comportamento de dirigir, induzindo pensamentos distrativos aos participantes enquanto eles dirigiam. Para avaliar a carga mental utilizaram a classificação original da ISA, analisando três situações: (i) somente a tarefa de dirigir, (ii) apenas a tarefa de memória e (iii) as condições de condução e pensamentos distrativos.

Jansen et al. (2016) buscaram examinar se as transições nas demandas das tarefas se manifestam na carga de trabalho mental e no desempenho em uma configuração de tarefa dupla, utilizando uma tarefa de memória enquanto o participante dirigia um carro.

Para isso, os autores fizeram duas adaptações: em vez de utilizar um sinal visual solicitando a avaliação da ISA por meio de um teclado, utilizaram um sinal auditivo e respostas numéricas verbais, e ainda utilizaram uma escala de 7 pontos, com 1 indicando uma tarefa muito fácil e 7 uma tarefa muito difícil, pois acreditavam que a escala de 5 pontos era insuficientemente sensível para discernir entre os níveis alto e baixo de dificuldade na tarefa de dirigir.

Além da ISA, Jansen et al. (2016) utilizaram o NASA-TLX após as tarefas, buscando interpretar as classificações da ISA coletadas durante as condições experimentais. Como resultados, os autores colocam que a ISA tem baixa obstrução, por ser unidimensional, e o NASA-TLX pode ser demorada para o preenchimento, mesmo apresentando informações valiosas sobre a natureza da carga de trabalho. Assim, o desafio é desenvolver uma escala de classificação que atinja um equilíbrio entre esses dois métodos. Além disso, também sugerem que medidas fisiológicas podem ajudar a identificar mudanças na carga de trabalho em um instante anterior às classificações da ISA (JANSEN et al., 2016).

Já Wilson et al. (2017) estudaram o uso de uma nova tecnologia de radar para previsões do tempo por meteorologistas, por meio de medidas da carga de trabalho mental.

Para isso, utilizaram a ISA a cada cinco minutos, pois, diferentemente do NASA-TLX e do SWAT, poderiam obter diversas classificações da CTM durante a tarefa, além de ser um método mais rápido, menos intrusivo e mais adequado a situação da pesquisa e não possuir uma classificação com base em diversos fatores, os quais não se encaixavam com o estudo. Nesse caso, os participantes responderam a escala original da ISA e, juntamente com as classificações, forneceram um raciocínio para o nível de carga de trabalho mental escolhido.

Um resultado encontrado por Wilson et al. (2017) foi que a carga de trabalho cognitiva pode estar relacionada à personalidade dos previsores, sendo interessante e necessário investigar essa hipótese.

No que se refere as vantagens da ISA, Tattersall e Foord (1996) citam como principal a avaliação da carga de trabalho durante o desempenho da tarefa principal. Assim, a ISA pode ser muito útil para avaliar a carga de trabalho de tarefas mais demoradas, com vários elementos ou fases, pois coleta avaliações subjetivas continuamente, baseadas no tempo, e durante a realização das atividades, não sendo preciso utilizar a memória para uma classificação posterior, como outros métodos subjetivos, e também para avaliar elementos específicos da tarefa.

Em relação as desvantagens, Tattersall e Foord (1996) acreditam que pode existir certa interferência no desempenho primário da tarefa observada, o que pode ser crítico para coletar dados em situações reais de trabalho, principalmente quando ambas as tarefas exibem modalidades sensoriais comuns. Assim, deve-se buscar uma menor interferência utilizando-se de requisitos de processamento de informações

da tarefa e da técnica de medição de carga de trabalho que envolvem atividades sensoriais diferentes.

De forma geral, a ISA é uma técnica muito simples e rápida de ser aplicada, permitindo que seja construído um perfil de carga de trabalho para a tarefa em análise. Além disso, é possivelmente o método menos intrusivo das técnicas de avaliação em tempo real da CTM e possui um baixo custo. Em contrapartida, é uma técnica muito simplista, oferecendo apenas uma classificação geral da carga de trabalho, e também pode apresentar vieses relacionados a erros de relatos pelos participantes.

4.2.16 ISTAS 21

O ISTAS 21 é uma ferramenta em forma de questionário individual utilizada para avaliar os riscos ocupacionais de natureza psicossocial, os quais estão ligadas com a carga mental de trabalho, e fundamentam uma metodologia de prevenção.

O ISTAS 21 é uma adaptação em espanhol do Questionário Psicossocial de Copenhague (CoPsoQ), que foi desenvolvido pelo Instituto Nacional de Saúde Ocupacional da Dinamarca em 2000, e foi descrito por Moncada, Llorens e Kristensen (2002).

Apesar do ISTAS 21 ser um questionário individual, ele avalia a exposição a fatores de risco à saúde de natureza psicossocial de um grupo de empregados na unidade (MONCADA; LLORENS; KRISTENSEN, 2002).

Segundo os autores, a necessidade dessa adaptação se deu pelo fato do questionário precisar ser ajustado a realidade específica de cada população, empresa, ou instituição que será aplicado, levando em consideração os objetivos de análise e prevenção perseguidos e a garantia do anonimato.

Em sua versão original dinamarquesa, o CoPsoQ possui três versões: curta, com 8 dimensões e 44 questões, com o propósito de autoavaliação para os trabalhadores da empresa; média, com 26 dimensões e 95 questões, destinado ao uso profissional na avaliação de risco, planejamento e inspeção; e longa, com 30 dimensões e 141 questões, possuindo fins de pesquisa. As três versões possuem identificação e mensuração de indicadores de exposição (dimensões psicossociais) e efeito (dimensões de saúde, estresse e satisfação), as quais, em sua maioria, são

baseadas em instrumentos conhecidos e previamente validados (MONCADA; LLORENS; KRISTENSEN, 2002).

Em comparação com os instrumentos existentes até 2000, Moncada, Llorens e Kristensen (2002) acreditam que o CoPsoQ representou um avanço em termos da mais ampla conceituação de dimensões psicossociais, possuindo maior homogeneidade e gerenciamento, e com dimensões antes não avaliadas, como demandas psicológicas emocionais e o fato de esconder emoções, além de unificar a linguagem e permitir comparações entre diferentes áreas.

A adaptação do questionário CoPsoQ passou pelo processo de retrotradução e também por oficinas de autores dinamarqueses e espanhóis para discutir o conteúdo do questionário em relação as diferenças existentes entre o mercado de trabalho dinamarquês e espanhol.

Seguindo a metodologia do CoPsoQ, o ISTAS 21 também possui as versões longa, média e curta, porém as versões longa e média incluem duas dimensões adicionais: estima e dupla presença. Ainda, a versão curta não contém as duas dimensões da saúde incluídas na versão dinamarquesa.

As principais características do ISTAS 21 estão apresentadas no Quadro 13.

Quadro 13 - Principais características do ISTAS 21

Características do ISTAS 21
Questionário individual e anônimo.
Três versões: longa (para investigação), média (para avaliação de riscos em empresas com mais de 30 trabalhadores) e curta (para empresas com menos de 30 trabalhadores e autoavaliação).
Aplicável a qualquer trabalho, ocupação ou atividade econômica.
Marco conceitual baseado no modelo demanda - controle - apoio social e esforço - compensação, incluindo somente dimensões para as quais existe validação científica razoável de que afetam a saúde.
Orientado para a participação e ação.
Padronização de todas as medidas de 0 a 100 pontos.
Compara as médias obtidas com valores de referência, obtidos mediante uma amostra representativa da população ocupada.
Apresenta os resultados em forma gráfica compreendendo todos os agentes sociais.
Validade e confiabilidade constatadas.
Internacional: Dinamarca, Espanha, Reino Unido, Bélgica, Alemanha, Brasil, Holanda e Suécia.

Fonte: Adaptado de Moncada, Llorens e Kristensen (2002).

Em relação as três versões do ISTAS 21, a versão longa é utilizada para pesquisas, a versão média avalia os riscos psicossociais em empresas com mais de 30 trabalhadores e a versão curta é utilizada em empresas com menos de 30

trabalhadores e também para autoavaliação (MONCADA; LLORENS; KRISTENSEN, 2002).

O manual escrito por Moncada, Llorens e Kristensen (2002) é correspondente a versão média, a qual, juntamente com a versão curta, são as mais aplicadas em trabalhos de campo, principalmente devido ao menor custo e maior disponibilidade e acessibilidade.

Conforme descrito pelos autores, o questionário inclui 20 dimensões psicossociais ocupacionais, cobrindo uma grande diversidade de riscos de exposição no trabalho, além da dimensão de dupla presença, relacionada ao duplo horário de trabalho, profissional e doméstico, totalizando 21 dimensões.

Todos os resultados obtidos no questionário são comparados com valores de referência, os quais foram obtidos por meio de uma população representativa de trabalhadores da Comunidade Autônoma de Navarra (MONCADA; LLORENS; KRISTENSEN, 2002).

A estrutura da versão média do questionário ISTAS 21 consiste em 124 questões, das quais 6 são suprimíveis, contendo 28 escalas, conforme Tabela 5.

Tabela 5 - Divisões do ISTAS 21 em seções, escalas e perguntas

Seção	Nº. Escalas	Nº. Perguntas	Nº. Perguntas suprimidas
Primeira			
Contexto social	-	2	2
Dupla presença	1	4	-
Segunda			
Condições de emprego e trabalho	-	15	4
Terceira			
Saúde, estresse e satisfação	7	30	-
Quarta			
Dimensões psicossociais	20	73	-
Total	28	124	6

Fonte: Adaptado de Moncada, Llorens e Kristensen (2002).

As duas primeiras seções incluem a caracterização do contexto social, incluindo demandas de trabalho doméstico e familiar, e de trabalho, como ocupação, vínculo empregatício, jornada e salário. Buscando o anonimato, algumas questões podem ser suprimidas e outras devem ser adaptadas conforme a realidade de cada instituição.

Já as outras duas seções tratam da saúde, estresse e satisfação no trabalho e as dimensões psicossociais, as quais são questões universais para todos os tipos de atividades e não podem ser modificadas e nem suprimidas (MONCADA; LLORENS; KRISTENSEN, 2002).

Especificamente, a seção de saúde, estresse e satisfação no trabalho é dividida em 7 escalas, descritas por Moncada, Llorens e Kristensen (2002):

i. Satisfação com o trabalho, a qual consiste em quatro perguntas: “Você está satisfeito com suas perspectivas de emprego?”, “Você está satisfeito com as condições ambientais de trabalho (ruído, espaço, ventilação, temperatura, iluminação, ...)?”, “Você está satisfeito com o grau em que suas habilidades são usadas?” e “Você está satisfeito com o seu trabalho, levando tudo em consideração?”. São cinco as opções de resposta: muito satisfeito, bastante satisfeito, mais ou menos satisfeito, pouco satisfeito ou nada satisfeito;

ii. Saúde geral, consistindo em cinco questões: “Em geral, você diria que sua saúde é: Excelente/ muito boa/ boa/ razoável/ ruim”, e também consiste na classificação de quatro frases “Eu fico doente mais facilmente do que outras pessoas”, “Eu sou tão saudável quanto qualquer um”, “Eu acho que minha saúde vai piorar” e “Minha saúde é excelente” com uma das opções: totalmente certa, bastante certa, não sei, bastante falsa, totalmente falsa;

iii. Saúde mental, com cinco questões referentes as quatro últimas semanas: “Você esteve muito nervoso?”, “Você se sentiu tão pobre em moral que nada poderia animá-lo?”, “Você se sentiu calmo?”, “Você se sentiu desanimado e triste?” e “Você se sentiu feliz?”, utilizando a classificação: sempre, quase sempre, muitas vezes, algumas vezes, somente alguma vez, nunca;

iv. Vitalidade, com quatro questões referentes as quatro últimas semanas: “Você se sentiu cheio de vitalidade?” “Você se sentiu com muita energia?”, “Você já se sentiu exausto?” e “Você já se sentiu cansado?”, utilizando a mesma classificação de iii.;

v. Sintomas comportamentais de estresse, com quatro afirmações referentes as quatro últimas semanas: “Eu não tive coragem de estar com as pessoas”, “Eu não fui capaz de dormir bem”, “Eu tenho estado irritado” e “Eu me senti sobrecarregado”, utilizando a mesma classificação de iii.;

vi. Sintomas somáticos de estresse, com quatro afirmações referentes as últimas quatro semanas e com a mesma classificação utilizada em iii.: “Você sentiu

aperto ou dor no peito”, “Você não teve ar”, “Você sentiu a tensão nos músculos”, “Você teve dor de cabeça”;

vii. Sintomas cognitivos de estresse, com mais quatro afirmações sobre as quatro últimas semanas e a mesma classificação utilizada em iii.: “Você teve dificuldade em se concentrar”, “Custou tomar decisões”, “Você teve dificuldade em lembrar as coisas” e “Você teve dificuldade em pensar claramente”.

As dimensões psicossociais são reunidas em cinco grandes dimensões, descritas por Moncada, Llorens e Kristensen (2002): demandas psicológicas do trabalho, trabalho ativo e desenvolvimento de habilidades, apoio social na empresa e qualidade de liderança, compensação e dupla presença. Cada uma dessas dimensões possui subdimensões, as quais estão dispostas no Quadro 14, juntamente com o número de questões de cada uma.

Quadro 14 - Dimensões psicossociais incluídas na versão média do ISTAS 21

Grupo dimensional	Dimensões psicossociais	Nº. Questões
Demandas Psicológicas	Demandas qualitativas.	4
	Demandas cognitivas.	4
	Demandas emocionais.	3
	Demandas de esconder emoções.	2
	Demandas sensoriais.	4
Trabalho ativo e desenvolvimento de habilidades	Influência no trabalho.	4
	Possibilidades de desenvolvimento no trabalho.	4
	Controle sobre os tempos de trabalho.	4
	Sentido do trabalho.	3
	Integração na empresa.	4
Apoio social na empresa e qualidade de liderança	Previsibilidade.	2
	Clareza do papel.	4
	Conflito de papéis.	4
	Qualidade de liderança.	4
	Reforço.	2
	Apoio social no trabalho.	4
	Possibilidades de relacionamento social.	2
	Sentimento de grupo.	3
Compensação	Insegurança no trabalho.	4
	Estima.	4
Dupla presença	Carga e preocupação por tarefas domésticas.	4

Fonte: Adaptado de Moncada, Llorens e Kristensen (2002) e Flores et al. (2014).

Moncada, Llorens e Kristensen (2002) apresentaram as questões correspondentes a cada uma das 21 dimensões:

i. Demandas psicológicas quantitativas: “Você tem que trabalhar muito rápido?”, “A distribuição de tarefas é irregular e faz com que seu trabalho se acumule?”, “Você tem tempo para manter seu trabalho atualizado?” e “Você tem tempo suficiente para fazer o seu trabalho?”;

ii. Demandas psicológicas cognitivas: “O seu trabalho exige memorizar muitas coisas?”, “O seu trabalho exige que você tome decisões rapidamente?”, “O seu trabalho exige que você tome decisões difíceis?” e “O seu trabalho exige um grande conhecimento?”;

iii. Demandas psicológicas emocionais: “Há momentos ou situações emocionalmente desgastantes ocorrendo em seu trabalho?”, “Seu trabalho, em geral, é emocionalmente desgastante?” e “Você tem problemas para esquecer problemas de trabalho?”;

iv. Demandas psicológicas de esconder emoções: “O seu trabalho exige que você fique em silêncio sobre sua opinião?” e “O seu trabalho exige que você esconda suas emoções?”;

v. Demandas psicológicas sensoriais: “O seu trabalho exige muita concentração?”, “O seu trabalho exige uma visão detalhada?”, “O seu trabalho exige atenção constante?” e “O seu trabalho requer um alto nível de precisão?”;

vi. Influência no trabalho: “Você tem muita influência nas decisões que afetam seu trabalho?”, “Você tem alguma influência na quantidade de trabalho que lhe é atribuída?”, “A sua opinião é levada em consideração quando suas atribuições são colocadas?” e “Você tem influência na ordem em que executa as suas tarefas?”;

vii. Possibilidades de desenvolvimento no trabalho: “O seu trabalho exige que você tenha iniciativa?”, “O seu trabalho permite que você aprenda coisas novas?”, “A conclusão do seu trabalho permite que você aplique suas habilidades e conhecimentos?” e “O seu trabalho é variado?”;

viii. Controle sobre os tempos de trabalho: “Você pode decidir quando você faz uma pausa?”, “Você pode tirar férias mais ou menos quando quiser?”, “Você pode deixar seu trabalho para conversar com um parceiro?” e “Se você tem um problema pessoal ou familiar, pode deixar seu trabalho pelo menos uma hora, sem ter que pedir uma autorização especial?”;

ix. Sentido do trabalho: “Suas tarefas fazem sentido?”, “As tarefas que você faz parecem importantes para você?” e “Você se sente comprometido com sua profissão?”;

x. Integração na empresa: “Você gostaria de permanecer na empresa em que você está pelo resto de sua vida profissional?”, “Você fala com entusiasmo sobre sua empresa para outras pessoas?”, “Você acha que os problemas da sua empresa também são seus?” e “Você acha que sua empresa é de grande importância para você?”;

xi. Previsibilidade: “Na sua empresa, você é informado com bastante antecedência das mudanças que podem afetar seu futuro?” e “Você recebe todas as informações que precisa para fazer bem o seu trabalho?”;

xii. Clareza do papel: “Você sabe exatamente que margem de autonomia você tem em seu trabalho?”, “O seu trabalho tem objetivos claros?”, “Você sabe exatamente quais tarefas são de sua responsabilidade?” e “Você sabe exatamente o que se espera de você no trabalho?”

xiii. Conflito de papéis: “Você faz coisas no trabalho que são aceitas por algumas pessoas e não por outras?”, “Você é obrigado a fazer coisas contraditórias no trabalho?”, “Você tem que fazer tarefas que acha que deveriam ser feitas de maneira diferente?” e “Você tem que executar tarefas que parecem desnecessárias?”;

xiv. Qualidade de liderança: Seus chefes imediatos: “Eles se certificam de que cada trabalhador tenha boas oportunidades de desenvolvimento profissional?”, “Eles planejam bem o trabalho?”, “Eles resolvem bem os conflitos?” e “Eles se comunicam bem com os trabalhadores?”;

xv. Reforço: Com que frequência... “Você fala com seu superior sobre como você executa seu trabalho?” e “Você fala com seus colegas sobre como você realiza seu trabalho?”;

xvi. Apoio social no trabalho: Com que frequência... “Você recebe ajuda e apoio de seus parceiros?”, “Seus colegas estão dispostos a ouvir seus problemas no trabalho?”, “Você recebe ajuda e apoio do seu superior imediato?” e “Seu superior imediato está disposto a ouvir seus problemas no trabalho?”;

xvii. Possibilidades de relacionamento social: “O seu trabalho está isolado dos seus colegas?” e “É possível conversar com seus colegas durante o trabalho?”;

xviii. Sentimento de grupo: “Existe um clima bom de trabalho entre você e seus colegas?”, “Você ajuda seus companheiros no trabalho?”, e “Em seu trabalho, você acha que faz parte de um grupo?”;

xvix. Insegurança no trabalho: “Você está preocupado com o quão difícil seria encontrar outro emprego no caso de você ficar desempregado?”, “Você está preocupado em mudarem sua tarefa contra sua vontade?”, “Você está preocupado em mudarem o horário (turno, dias da semana, horas de entrada e saída) contra a sua vontade?” e “Você está preocupado se variam o seu salário (não atualizam, diminuem, introduzem um salário variável, pagam em espécie, etc.)?”;

xx. Estima: “Meus superiores me dão o reconhecimento que eu mereço”, “Em situações difíceis no trabalho, recebo o apoio necessário”, “No meu trabalho eles me tratam injustamente” e “Se eu pensar em todo o trabalho e esforço que fiz, o reconhecimento que recebo no meu trabalho parece apropriado”;

xxi) Dupla Presença: “Que parte da família e do trabalho doméstico você faz? Eu sou a principal pessoa responsável e faço a maioria das tarefas familiares e domésticas/ Faço aproximadamente metade das tarefas familiares e domésticas/ Faço aproximadamente um quarto das tarefas familiares e domésticas/ Eu só faço tarefas muito específicas/ Eu não faço nenhuma ou quase nenhuma dessas tarefas”, “Se você perder um dia de casa, as tarefas domésticas que você faz são deixadas por fazer?”, “Quando você está na empresa, você pensa em tarefas domésticas e familiares?” e “Há momentos em que você precisaria estar na empresa e em casa ao mesmo tempo?”.

Com exceção da dimensão “insegurança no trabalho”, a qual é respondida por meio das opções “muito preocupado/ bastante preocupado/ mais ou menos preocupado/ pouco preocupado/ nada preocupado”, as demais dimensões são respondidas com as opções “sempre/ quase sempre/ muitas vezes/ algumas vezes/ somente alguma vez/ nunca”.

O processo de intervenção do ISTAS 21 é sugerido por Moncada, Llorens e Kristensen (2002) como um *checklist*, o qual deve conter a designação do grupo avaliado, a preparação para o trabalho de campo, o desenvolvimento do trabalho de campo, as análises, as prioridades de intervenção, o informe final da avaliação dos riscos psicossociais e o plano de ação preventiva, a aplicação e seguimento das medidas preventivas e a avaliação das medidas preventivas.

A pontuação de cada uma das dimensões psicossociais é padronizada entre 0 e 100 pontos. Essas dimensões podem ser positivas, nas quais a situação mais favorável para a saúde ocorre em escores altos, e negativas, aquelas para as quais as situações mais favoráveis para a saúde ocorrem em pontuação baixas. A distância entre o escore do participante e o escore ideal (0 ou 100) é observada, assim como a distância ao escore da população de referência (MONCADA; LLORENS; KRISTENSEN, 2002).

Com esses resultados os trabalhadores são classificados em três tercís: verde, o qual é o mais favorável para a saúde; amarelo, intermediário para a saúde; e vermelho, o mais desfavorável para a saúde. Os pontos de corte para os tercís são estabelecidos calculando-se os percentis exatos de 33,33 e 66,66, estabelecendo assim os valores de referência da população para todas as dimensões, os quais estão dispostos no manual do ISTAS 21 de Moncada, Llorens e Kristensen (2002).

Assim, os resultados de cada uma das 21 dimensões são apresentados em um gráfico de colunas por meio da porcentagem de trabalhadores em cada tercíl, favorecendo a compreensão e definição das áreas que precisam ser melhoradas no trabalho, possibilitando estabelecer prioridades e orientações para as ações preventivas (MONCADA; LLORENS; KRISTENSEN, 2002).

O método ISTAS 21 apresenta excelente nível de validade, confiabilidade e concordância com a versão original CoPsoQ, com as três versões (longa, média e curta) podendo ser utilizadas para fins de pesquisa e de avaliação de riscos em pequenas, médias e grandes empresas. Suas escalas apresentam coerência interna e a versão intermediária é um bom reflexo da versão longa, apresentando concordância (MONCADA et al., 2005).

Martínez et al. (2010) buscaram encontrar as causas de estresse nos trabalhadores de um centro de reabilitação de saúde mental, e comparar os resultados entre os profissionais que realizam o atendimento direto a pacientes com transtornos psiquiátricos e os trabalhadores administrativos. Para isso, utilizaram o NASA-TLX para avaliar a carga de trabalho e o ISTAS 21, versão curta, para avaliar os fatores de risco psicossocial no trabalho, além de um questionário sobre os dados gerais dos trabalhadores, como gênero, idade, tempo de empresa, jornada de trabalho e escolaridade.

Os autores apresentaram os resultados do ISTAS 21 por meio dos tercís separados das cinco grandes dimensões: demandas psicológicas, trabalho ativo e

desenvolvimento de habilidades, apoio social na empresa e qualidade de liderança, compensação e dupla presença. Nesse caso, foi possível aplicar a versão curta do questionário devido a amostra possuir menos de 20 trabalhadores.

Arellano et al. (2012) avaliaram subjetivamente a carga de trabalho e a fadiga entre operadores de tornos CNC. Para isso, utilizaram o NASA-TLX e a versão curta do ISTAS 21 para avaliar a carga de trabalho e os efeitos de fatores psicossociais na carga de trabalho, respectivamente. Os questionários foram entregues aos supervisores da equipe e eles ficaram responsáveis pela distribuição aos trabalhadores. Com os resultados, a confiabilidade do ISTAS 21 foi consistente com os valores de adequação da amostra, porém torna-se necessário aumentar a amostra para melhorar a confiabilidade.

Oñate e Meyer (2012) estudaram ergonomicamente um operador de uma empresa de mineração, buscando identificar as condições de trabalho nas quais as cargas de trabalho física e mental poderiam estar acima das capacidades do operador. Para isso, utilizaram registros eletrônicos, entrevistas, pesquisas, revisão dos padrões da empresa, um estudo de tempo e análise da carga de trabalho física e mental.

A carga de trabalho mental foi avaliada por meio do NASA-TLX, juntamente com a versão curta do ISTAS 21, buscando encontrar os riscos psicossociais no trabalho do operador. Por meio dos resultados encontrados, foram propostas soluções de melhorias, as quais, após um tempo, foram testadas e mostraram um resultado positivo para a saúde e carga mental do trabalhador estudado (OÑATE; MEYER, 2012).

Flores et al. (2014) estudaram uma equipe de cuidados paliativos, que atenta adultos com câncer, buscando descrever características do trabalho associados à síndrome de *burnout*. Para isso, avaliaram a síndrome de *Burnout*, fatores de carga mental e riscos psicossociais, por meio do *Maslach Burnout Inventory* (MBI), NASA-TLX e a versão curta do ISTAS 21, respectivamente, bem como utilizaram um questionário de dados pessoais.

A partir dos resultados, Flores et al. (2014) encontraram a necessidade e importância de intervir na equipe de cuidados paliativos, pois os trabalhadores são expostos a altas demandas que podem prejudicar a qualidade de vida e de trabalho (FLORES et al., 2014).

Já Ceballos-Vásquez et al. (2015) analisaram a percepção de fatores psicossociais e a carga de trabalho mental de enfermeiros que trabalham em unidades de terapia intensiva, utilizando três instrumentos: questionário sociodemográfico, no qual foram incluídas variáveis como idade, instituição, unidade, funções desempenhadas e turno de trabalho, ISTAS 21 (versão curta), orientado à avaliação de fatores psicossociais, incluindo cinco dimensões, e a Escala Subjetiva de Carga Mental de Trabalho (ESCAM). Os resultados apresentaram relações entre o ISTAS 21 e a ESCAM, mesmo utilizando a versão curta do ISTAS 21, a qual não é recomendada para organizações de médio ou grande porte, mesmo mostrando confiabilidade e utilizando pouco tempo para implementação.

Apesar de não ser diretamente um método para avaliar a carga de trabalho mental, o ISTAS 21 avalia os riscos psicossociais no trabalho, os quais podem ser relacionadas com a CTM, conforme apresentado nos trabalhos acima. Dessa forma, o questionário normalmente é aplicado juntamente com algum outro método subjetivo de avaliação da carga de trabalho mental, sendo usualmente utilizado o NASA-TLX.

Com as versões longa, média e curta é possível aplicar o ISTAS 21 em todos os tamanhos de empresas e nos mais diversos ramos. Os artigos relatados acima possuem aplicações da versão curta por possuírem uma amostra pequena e pela sua facilidade e agilidade de coleta, em contraste com as versões média e longa que possuem mais questões, e, conseqüentemente, demandam um maior tempo para preenchimento, podendo apresentar certa resistência por meio dos participantes e até mesmo liberação da empresa para realização de trabalhos. Apesar das pequenas diferenciações, as três versões possuem validade e confiabilidade e cabe a cada profissional enquadrá-las ao contexto de trabalho avaliado.

4.2.17 Modified Cooper Harper Scale (MCH)

A *Modified Cooper Harper Scale* (MCH), ou, em português, escala modificada de Cooper Harper é uma das formas mais conhecidas de avaliar a carga de trabalho mental de forma subjetiva, por meio de uma escala unidimensional, a qual é uma adaptação da escala original de Cooper Harper (CH).

A escala original de Cooper Harper foi fundamentada por Cooper e Harper, em 1969, para mensurar a carga mental proveniente de tarefas de voo. Dessa forma,

como foi um dos pioneiros no assunto, o método tornou-se muito estudado e modificações foram sugeridas com o passar do tempo (COOPER; HARPER, 1969).

Cooper e Harper, após um estudo de revisão nas escalas existentes para avaliação de qualidades de manuseio de pilotos, em 1966, utilizaram alguns conceitos e estruturas já existentes e criaram uma nova escala de avaliação, buscando padronização e redução das limitações até então existentes, buscando principalmente clareza nos termos e palavras utilizadas, melhorando a comunicação e o entendimento (COOPER; HARPER, 1969).

Dessa forma, Cooper e Harper (1969) buscaram uma nova definição de qualidade de manuseio, abordando não apenas características de estabilidade e controle, mas também levando em consideração a interface de comando, o ambiente de voo e o estresse.

Assim, baseando-se na escala original de Cooper, de 1957, e na escala CAL, as quais consistem em tomadas de decisão por meio de adjetivos como resposta, os quais chegam em um escore final entre 1 e 10, Cooper e Harper desenvolveram a escala CH.

A escala de Cooper Harper consiste em uma árvore de decisões, com perguntas sucessivas aos pilotos, chegando, ao final, em um escore entre 1 e 10, o qual é referente ao nível de manuseio da aeronave e, conseqüentemente, da carga de trabalho mental. A árvore de decisões está representada na Figura 19.

Dessa forma, conforme pode ser observado na Figura 19, Cooper e Harper (1969) utilizam três decisões para levar o piloto ao escore final, são elas:

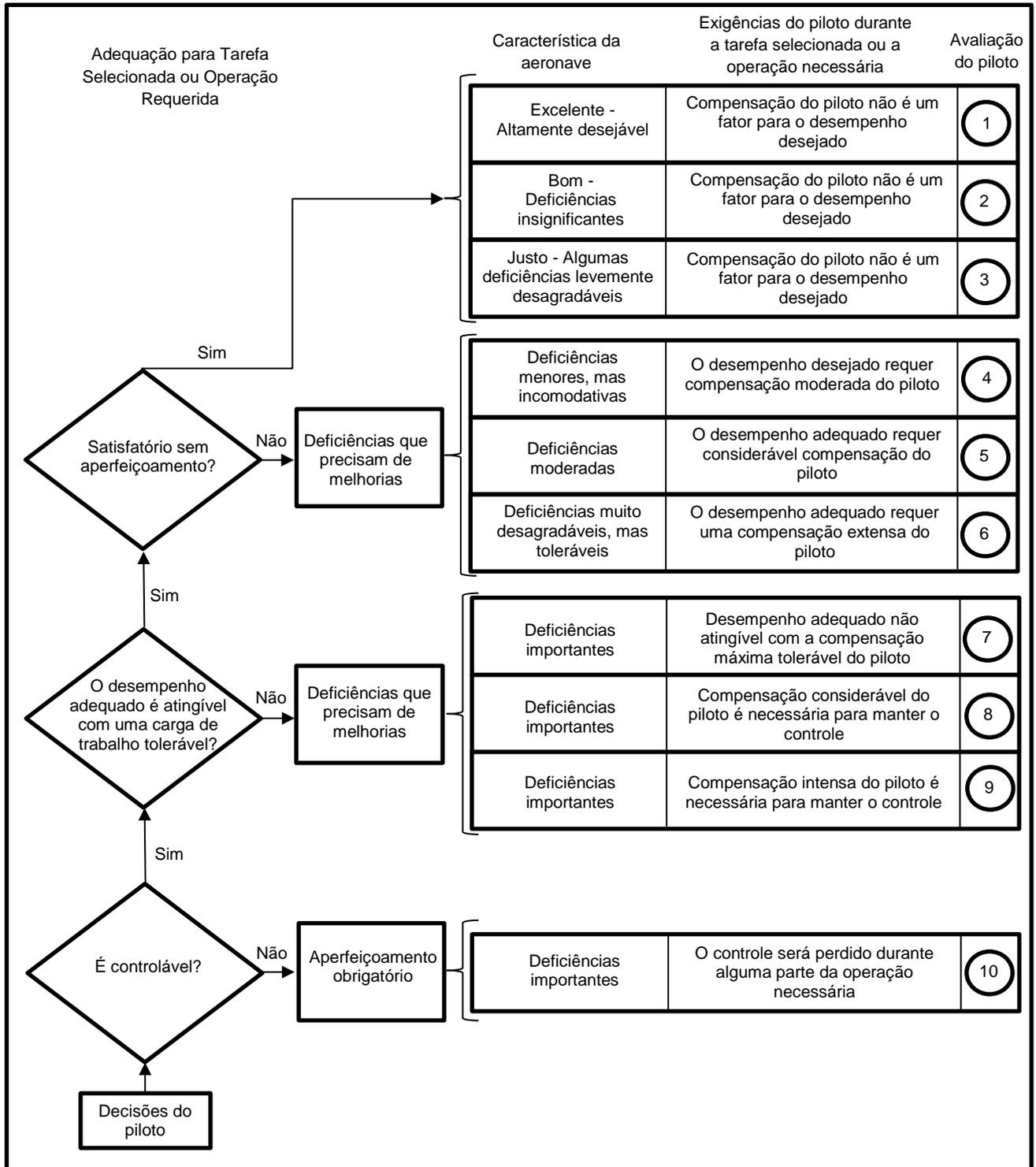
- i. O veículo é controlável?
- ii. O desempenho adequado é atingível com uma carga de trabalho tolerável?
- iii. O veículo é satisfatório sem aperfeiçoamento?

Ainda, os autores conseguem classificar qualitativamente o escore final em quatro categorias:

- i. Satisfatório – bom o suficiente sem aperfeiçoamento e, portanto, da melhor categoria;
- ii. Insatisfatório, mas tolerante - apenas bom o suficiente, adequado para o propósito, mas desejável aperfeiçoamento;
- iii. Inaceitável para o piloto - não é adequado para o propósito, mas ainda controlável;

iv. Incontrolável - inaceitável para o propósito e da mais pobre qualidade.

Figura 19 - Árvore de decisões da Escala de Cooper Harper



Fonte: Adaptado de Cooper e Harper (1969).

Assim, a escala desenvolvida por Cooper e Harper (1969) é dependente de palavras, com o escore apenas representando um resultado final. Essa escala pode

ser utilizada para avaliar voos completos, uma fase do voo, uma tarefa especificamente definida ou uma subfase, a qual, como regra geral, deve ser aplicada no local de trabalho, quando as características, o desempenho e a carga de trabalho estão recentes na memória (COOPER; HARPER, 1969).

As classificações de cada piloto sobre as situações do voo são comparadas, indicando a qualidade do voo, sendo essas informações repassadas aos engenheiros, os quais podem trabalhar em cima dos resultados, buscando atingir o objetivo proposto para a aeronave (BAPTISTA, 2015).

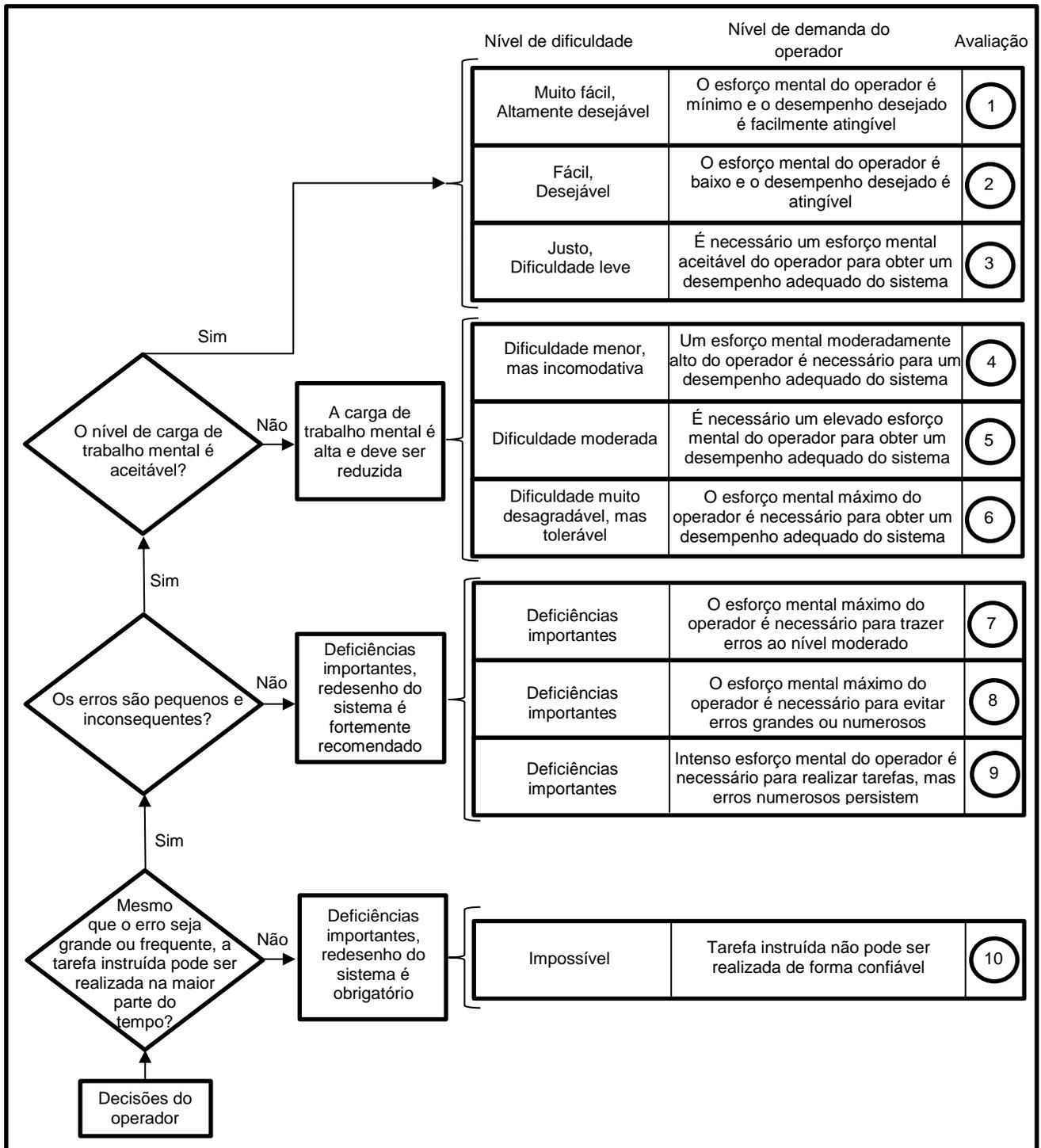
Várias versões modificadas da escala foram desenvolvidas, mantendo a árvore de decisões e a escala de 10 pontos, mas substituindo os termos que tratam da carga de trabalho (HART; WICKENS, 1990).

Uma dessas adaptações foi feita em 1981, por Wierwille, que propôs uma escala que ficou conhecida como a escala Modificada de Cooper Harper (MCH), pois a escala original avaliava apenas as qualidades de manuseio de aeronaves e a carga mental associado ao piloto. Dessa forma, Wierwille buscou criar uma escala para avaliar a carga de trabalho em sistemas diferentes daqueles em que o operador humano executava tarefas essencialmente motoras, abrangendo atividades perceptivas, cognitivas e de comunicação, tarefas que são consequências de uma maior automatização (WIERWILLE; CASALI, 1983).

Assim, a escala MCH busca medir situações em sistemas mais complexos e automatizados, onde os operadores humanos são usados com menos frequência como elementos ativos do sistema de controle e mais frequentemente com as suas habilidades de percepção, monitoramento, avaliação, comunicação e resolução de problemas (WIERWILLE; CASALI, 1983).

A escala MCH está representada na Figura 20 e mantém a árvore de decisões encontrada na escala original. Porém, conforme descrito por Wierwille e Casali (1983), os termos foram substituídos de forma mais apropriada a outras atividades humanas, envolvendo termos como realização da tarefa, erros, dificuldade, desempenho e carga mental, buscando aplicações em sistemas mais modernos.

Figura 20 - Árvore de decisões da MCH



Fonte: Adaptado de Wierwille e Casali (1983).

Wierwille e Casali (1983) validaram a MCH em três experimentos diferentes, cada um deles abrangendo uma situação de trabalho diferente, as quais não envolviam tarefas motoras e abrangiam alguma forma de automatização, visto que a MCH objetiva avaliar a carga mental nessas modalidades de tarefas. Dessa

forma, o primeiro experimento enfatizou o aspecto perceptivo da atividade humana, o segundo experimento foi projetado para testar a sensibilidade da carga cognitiva da escala MCH e o terceiro experimento foi projetado para enfatizar as comunicações verbais, incluindo detecção, reconhecimento, abrangência e resposta. A classificação da escala MCH foi fornecida pelos participantes imediatamente após cada voo.

Com esses experimentos, os autores encontraram que as classificações da escala MCH são válidas, estatisticamente confiáveis e fornecem uma boa avaliação geral da carga de trabalho. Dessa forma, Wierwille e Casali (1983) recomendaram que a escala MCH seja incluída em experimentos nos quais a CTM deve ser avaliada, devendo manter um bom design experimental, instruindo os participantes adequadamente, por meio dos termos importantes e procedimentos que devem ser seguidos para conseguir a classificação, e executando as avaliações imediatamente após cada situação.

Bevilacqua et al. (2013) estudaram a carga de trabalho mental de trabalhadores de um grupo automotivo, utilizando para isso a MCH, após implementação de um sistema visual. Assim, a escala foi administrada antes e depois das modificações introduzidas.

Mansikka, Virtanen e Harris (2018) analisaram a sensibilidade do NASA-TLX, da escala MCH e da média do intervalo entre batimentos cardíacos sucessivos, em um dispositivo de treinamento de voo. Para ambas as escalas, os participantes foram treinados e forneceram a sua ponderação, relativa as tarefas realizadas, após cada teste.

Os autores encontraram que todas as medidas diferenciaram a maioria das condições de tarefa e existe uma forte correlação entre o NASA-TLX e a MCH, sugerindo que as classificações forneceram resultados semelhantes. Esse resultado é útil no sentido que a MCH pode ser uma opção mais apropriada para avaliar a carga de trabalho mental em um ambiente que possui limitações de tempo, visto que a classificação da MCH demora menos tempo que a do NASA-TLX (MANSIKKA; VIRTANEN; HARRIS, 2018).

Já Eraslan, Can e Atalay (2016) utilizaram a MCH, juntamente com o NASA-TLX e o SWAT para desenvolver um modelo de tomada de decisões multicritério difusa para calcular a carga de trabalho mental.

De forma geral, a escala de CH é utilizada em situações com aplicações motoras e a escala MCH é utilizada em questões perceptivas, cognitivas e de

comunicação, a qual tornou possível aumentar as aplicações da escala de Cooper Harper. Ambas as escalas foram testadas em diversos estudos e apresentaram sensibilidade e confiabilidade necessárias para avaliar a carga de trabalho mental, mostrando-se um procedimento de rápida aplicação e fácil entendimento, após explicação aos participantes.

Porém, como ambas as escalas são unidimensionais, a sua sensibilidade geral pode ser questionada, pois ignora a natureza multidimensional do processamento de informações humanas, não distinguindo as demandas de processamento de tarefas. Além disso, como são subjetivas e restringidas ao participante, o qual expressa sua opinião, podem existir dificuldades de quantificação pelo participante e também como são necessárias várias avaliações de participantes diferentes, buscando resolver as variações de opinião, ocorre um aumento da duração do programa de testes, gerando maior gasto econômico.

4.2.18 Tarefa primária

A medição de tarefa primária, ou medição direta do desempenho, é uma forma de avaliar a carga de trabalho mental por meio do desempenho do trabalhador em relação a sua própria atividade.

As medidas de tarefas primárias como meio de avaliar a carga de trabalho começaram a ser utilizadas nos anos de 1960, porém foi após 1970 que as aplicações ganharam um uso mais notório, principalmente em questões relacionadas a aviação (WILLIGES; WIERWILLE, 1979).

Williges e Wierwille (1979) supõem que a medida que a carga de trabalho mental aumenta, o desempenho tende a alguma degradação, pois os recursos necessários para executar as tarefas estão disponíveis em quantidades finitas. Dessa forma, conforme a dificuldade é aumentava, mais recursos são necessários para manter um desempenho consistente. Assim, a hipótese fundamenta-se no fato que uma diminuição do desempenho pode indicar um aumento de CTM.

Dessa forma, os autores citam algumas medidas de desempenho que podem ser utilizadas na aviação, como o desempenho de aterrissagem, desempenho de mira e número de erros processuais. A conclusão bem-sucedida da missão deve ser definida em termos das medidas.

Williges e Wierwille (1979) diferenciam três maneiras de utilizar a tarefa primária como forma de avaliar a carga de trabalho mental: medidas isoladas ou individuais, medidas múltiplas e modelos matemáticos.

As medidas individuais são o uso de tarefas primárias analisadas individualmente. Muitas vezes essas medidas podem ser utilizadas como uma verificação do desempenho, utilizando um outro método principal para avaliar a carga de trabalho. Williges e Wierwille (1979) acreditam que as medidas isoladas só conseguem avaliar situações com alta carga de trabalho mental, necessitam de um equipamento para detecção, condicionamento e gravação e são específicas de cada tarefa, sendo necessárias novas medidas conforme a situação é modificada.

As medidas múltiplas envolvem o uso de mais de uma medida primária para avaliar a mesma tarefa, fornecendo uma visão mais completa do comportamento do trabalhador e do desempenho operacional do sistema. Nesse caso, Williges e Wierwille (1979) colocam que cada natureza específica de tarefa exige um novo conjunto de medidas. Além disso, utilizando-se muitas medidas pode-se ter algumas que não sejam alteradas de maneira confiável como uma função da carga de trabalho.

Já os modelos matemáticos envolvem equações matemáticas do desempenho do trabalhador. Para Williges e Wierwille (1979), a estimativa da carga de trabalho por modelagem matemática é geralmente confinada a tarefas específicas, bem restritas e talvez repetitivas, como o controle manual e o monitoramento de instrumentos. Tanto os modelos matemáticos, como os outros dois grupos, não são aconselháveis para avaliar a CTM de forma separada.

Hicks e Wierwille (1979) citam como aspectos positivos do uso de medidas primárias de desempenho o fato da medição não ser intrusiva, uma vez que o comportamento envolvido já está presente, e a possibilidade de utilizar diversas medidas, conseguindo expandir a análise. Como desvantagem, os autores citam a dificuldade em generalizar as medidas, pois tarefas diferentes devem ser analisadas de maneiras diferentes.

Hart e Wickens (1990) descrevem um relacionamento mais complexo entre carga de trabalho e desempenho: para tarefas relativamente simples, o desempenho consistente é mantida ao longo de um intervalo de níveis de dificuldade, mesmo com o aumento do volume de trabalho; para tarefas moderadamente difíceis, o desempenho deteriora-se de forma linear com o aumento da exigência da tarefa; e para tarefas muito difíceis, algumas tarefas podem ser ignoradas, mantendo um nível

consistente de esforço em outras. Assim, o desempenho das tarefas primárias correlaciona-se com a carga de trabalho apenas para tarefas moderadamente difíceis.

Hart e Wickens (1990) consideram difícil avaliar a carga de trabalho a partir de medidas de desempenho, pois se duas tarefas não competem pelos mesmos recursos, o desempenho pode ser mantido, mas a carga de trabalho será aumentada; em contrapartida, se duas tarefas competem pelos mesmos recursos, o desempenho de uma ou de ambas pode ser reduzido, caso os recursos sejam insuficientes, mas a carga de trabalho não será necessariamente aumentada. Ainda, os autores colocam o tempo para executar a tarefa como um fator muito importante nas medidas de desempenho, pois acreditam que praticamente qualquer tarefa pode ser executada de forma correta se houver tempo necessário para isso.

Os autores acreditam que medidas diretas de desempenho de tarefas são geralmente indicadores ruins de carga mental, pois não diagnosticam a fonte de carga e não refletem os investimentos em recursos causados por mudanças nas demandas de tarefas, apesar delas indicarem se o trabalhador foi ou não capaz de realizar a tarefa.

Assim, Hart e Wickens (1990) colocam que a medida que as demandas de tarefa são aumentadas, os trabalhadores geralmente conseguem manter um nível consistente de desempenho, embora a carga de trabalho possa ser maior ou menor, dependendo da estratégia adotada. Por outro lado, decréscimos de desempenho que acompanham um aumento nas demandas de tarefas podem refletir um nível de esforço constante, ou reduzido, ou o limite das capacidades do trabalhador, dependendo novamente das prioridades e estratégias escolhidas e da capacidade individual de cada um.

Para Cain (2007), as medidas de tarefa primária precisam ter contexto, relevância, representação e treinamento em tarefa apropriada para serem confiáveis, e podem ser representadas pela velocidade, precisão, reação ou tempos de resposta e taxas de erro.

Porém, Cain (2007) coloca algumas situações em que as tarefas primárias podem não ser suficientes ou adequadas para avaliar a carga de trabalho: quando a variabilidade nas demandas das tarefas é insuficiente para resultar em mudanças observáveis no desempenho da tarefa principal, quando as tarefas primárias utilizam modalidades sensoriais diferentes, quando a tarefa principal causa efeitos incidentes,

como fadiga ou estresse, e quando outros fatores, como estratégia, afetam o desempenho e a carga de trabalho de forma diferente.

Como desvantagem, Cain (2007) acredita que medidas de desempenho de tarefas primárias são menos práticas do que medidas subjetivas, não possuem generalizações e padronizações e não são suficientes por conta própria para avaliar a CTM.

Othman e Romli (2016) avaliaram a carga de trabalho mental de pilotos por meio de medidas de tarefa primária, do NASA-TLX e de medidas oculares. A avaliação do desempenho foi registrada em tempo real e foram utilizados o tempo de resposta do piloto e a taxa de operação precisa durante o intervalo de tempo de ocorrência de informações anormais e a conclusão correta da tarefa de reação.

Os autores categorizaram o tempo de resposta preciso e a taxa de conclusão da tarefa em relação à tarefa de voo de operação de anormalidade em três níveis de carga mental: alta, moderada e baixa. Utilizando o teste ANOVA, encontraram que há uma indicação de que existe um efeito significativo da carga de trabalho mental tanto para o tempo de resposta quanto para a porcentagem de precisão da tarefa de voo, pois à medida que o nível de carga de trabalho mental aumentou, a porcentagem de precisão diminuiu e o tempo de resposta aumentou.

Butmee e Lansdown (2017) determinaram os efeitos das mudanças de estado de controle, de automatizado para manual e vice-versa, na carga de trabalho mental de motoristas, utilizando medidas subjetivas, rastreamento ocular e medidas de desempenho de tarefa primária e secundária. As tarefas primárias envolveram análises de velocidade, distância, posição, precisão e tempo.

Já Longo (2016) revisou estudos da carga de trabalho mental na medicina, e encontrou diversos trabalhos que utilizaram o desempenho primário das tarefas como indicador da CTM, utilizando análises de tempo, precisão, velocidade e taxa de erro.

Segundo Longo (2016), em contextos clínicos, as medidas de tarefas primárias são mais úteis para estimar a capacidade ociosa do clínico para lidar com tarefas adicionais e são adequadas para tarefas e períodos curtos de trabalho mental. Porém, além de possuírem baixa capacidade de distinguir o desempenho de várias tarefas paralelas, são sensíveis apenas a grandes mudanças na carga de trabalho mental.

4.2.19 Integrated Workload Scale (IWS)

A *Integrated Workload Scale* (IWS) ou Escala Integrada de Carga de Trabalho, é um método subjetivo de avaliação da carga de trabalho mental que utiliza uma escala unidimensional, a qual pode ser utilizada para avaliar a CT em tempo real, por meio de aplicações em intervalos predefinidos.

Pickup et al. (2005) desenvolveram essa escala pois a Network Rail, uma proprietária da infraestrutura ferroviária do Reino Unido, estava interessada na avaliação da carga de trabalho mental dos sinalizadores ferroviários e até aquele momento não existia um método desenvolvido diretamente para o setor ferroviário, apesar da grande variedade de ferramentas existentes.

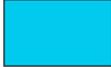
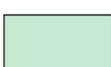
Além dos requisitos de validade, confiabilidade, aceitabilidade e sensibilidade, a Network Rail solicitou outros requisitos, como ser um método qualitativo e/ ou quantitativo, permitir medição instantânea e/ ou contínua, permitir o rastreamento de picos e vales de carga, utilização online ou off-line e escala com administração rápida e minimamente invasiva, para uso em campo. Assim, Pickup et al. (2005) consideraram adequada uma escala de classificação unidimensional e utilizaram a ISA como inspiração, modificando descritores, os quais deveriam refletir como os próprios sinalizadores se sentem em relação aos seus níveis de carga de trabalho.

Para isso, Pickup et al. (2005) realizaram entrevistas com sinalizadores, buscando encontrar as fontes de cargas percebidas, as causas do esforço e da demanda e os tipos de efeitos decorrentes do seu trabalho. Assim, os autores conseguiram obter as principais dimensões da carga de trabalho na terminologia própria dos sinalizadores.

Com as entrevistas e consenso com outros pesquisadores, Pickup et al. (2005) elaboraram onze descritores de carga de trabalho e solicitaram que os sinalizadores os colocassem em ordem, em uma linha de alta a baixa carga de trabalho. Além disso, eles poderiam classificar como confuso os descritores que possuíam alguma dificuldade de interpretação.

Com base nas respostas dessas classificações, Pickup et al. (2005) organizaram a ferramenta IWS, configurando-a com nove descritores, representada na Figura 21.

Figura 21 - Escala Integrada de Carga de Trabalho (IWS)

	Não é exigente	O trabalho não é nada exigente
	Esforço mínimo	Esforço mínimo necessário para se manter no topo da situação
	Algum tempo livre	Ativo com algum tempo livre para concluir trabalhos menos essenciais
	Esforço moderado	Trabalho exigente, mas gerenciável com moderado esforço
	Pressão moderada	Pressão moderada, o trabalho é administrável
	Muito ocupado	Muito ocupado, mas ainda capaz de fazer o trabalho
	Esforço extremo	Esforço e concentração extremos são necessários para garantir que tudo seja feito
	Lutando para acompanhar	Muito alto nível de esforço e demanda, lutando para acompanhar tudo
	Trabalho muito exigente	Trabalho muito exigente, com problemas complexos ou múltiplos para lidar e, mesmo com níveis muito elevados de esforço, é incontrolável

Fonte: Adaptado de Pickup et al. (2005).

Como mostrado na Figura 21, os descritores são colocados da menor carga de trabalho para a maior, sendo representados pelo descritor principal, além de uma explicação a direita e uma cor representativa, a qual demonstra esse aumento de CT, buscando maior usabilidade.

Como resultado da IWS, é gerado um gráfico da carga de trabalho experimentada pelo operador durante um cenário, uma tarefa, um evento ou um período específico desejado. Além disso, apesar de permitir o cálculo de um valor médio de carga de trabalho ao longo do tempo estudado, a IWS foi desenvolvida para rastrear picos altos e baixos da carga de trabalho, então a apresentação dos dados é feita por meio do score, de 1 a 9, para cada medição da escala, conforme intervalo definido no estudo, mostrando todo o comportamento do operador durante o procedimento (PICKUP et al., 2005).

A aplicação da ferramenta pode acontecer de forma manual, com o investigador utilizando um cronômetro e solicitando as respostas nos intervalos definidos no estudo, por meio de um registro em papel. Essa técnica apresenta como vantagem a sua simplicidade e nenhuma possibilidade de erro tecnológico. Porém,

como desvantagem, os dados precisam ser inseridos em algum *software* para gerar um gráfico com os dados, ocupando um maior tempo do investigador (PICKUP et al., 2005).

Já a aplicação por meio de dispositivos eletrônicos agiliza a coleta, processamento e análise dos dados, diminuindo a invasividade no ambiente de trabalho do sinalizador. Essa aplicação acontece por meio de um computador portátil, podendo possuir uma tela sensível ao toque ou respostas por meio de mouse e/ou teclado, com um *software* que solicita a resposta da IWS em intervalos predefinidos, usualmente utilizando um alarme sonoro. Como o *software* registra o escore, a hora em que a solicitação foi feita e a hora em que a classificação foi realmente recebida, pode-se gerar dois gráficos, com o tempo de espera e a classificação IWS (PICKUP et al., 2005).

Para validação do instrumento, Pickup et al. (2005) realizaram dois estudos experimentais. O primeiro utilizou simuladores próximos as condições reais de trabalho e solicitou a IWS a cada dois minutos, a qual foi respondida por meio de um teclado. Ainda, ao final, os participantes preencheram um questionário sobre a usabilidade da IWS e também preencheram o SWAT com uma pequena adaptação.

Já o segundo experimento utilizou um cenário que gerou maior carga de trabalho para os operadores, para testar a extremidade superior da escala de avaliação IWS, a qual foi respondida por meio do toque em uma tela, a cada dois minutos. O mesmo questionário para testar a usabilidade da ferramenta também foi aplicado ao final do experimento (PICKUP et al., 2005).

Com esses estudos, Pickup et al. (2005) encontraram que o primeiro experimento apresentou relação entre a IWS e o SWAT, mostrando que ambos mediam a mesma coisa. Já o segundo experimento encontrou uma boa sensibilidade da ferramenta. Além disso, em ambos os experimentos, os participantes relataram que a IWS refletia a intensidade com que os operadores se sentiam trabalhando, representando de forma precisa a carga de trabalho. Ainda, a IWS se mostrou não intrusiva.

Com esses resultados experimentais, Pickup et al. (2005) realizaram um estudo de campo: o sinalizador ferroviário foi analisado em seu ambiente de trabalho, sendo solicitado a cada cinco minutos a responder a IWS, por meio de dois métodos de captura, em folha de registro e pelo pressionamento de um botão várias vezes até mostrar o número apropriado. Para facilitar o pressionamento do botão, a escala foi

invertida, exigindo menos pressionamentos em situações de maior carga. Com os resultados dessa pesquisa de campo, os autores encontraram que os escores da IWS apresentam uma boa indicação da demanda e do esforço relativo às tarefas registradas na linha de tempo.

Dessa forma, Pickup et al. (2005) comprovaram a validade e a sensibilidade da IWS na indústria ferroviária, identificando os altos e baixos na carga de trabalho, podendo integrar a IWS com a linha de tempo e observações dos avaliadores e/ou especialistas, em busca das razões para as flutuações na carga de trabalho.

Siegel e Schraagen (2014) realizaram um estudo com sinalizadores ferroviários para analisar o estudo de resiliência do sistema homem-máquina, utilizando para isso medidas da carga de trabalho. A carga de trabalho foi analisada de forma objetiva, subjetiva e fisiológica, ou seja, por meio da Carga de Tarefa Cognitiva Externa (XTL), da IWS e da HRV, sendo a HRV utilizada para validação.

A XTL utiliza três dimensões: complexidade da tarefa, duração da tarefa e troca de tarefas, utilizando parâmetros disponíveis em tempo real. Já a IWS foi medida de duas formas: em um computador separado, próximo a cada estação de trabalho, e integrada dentro do sistema operacional (SIEGEL; SCHRAAGEN, 2014).

Siegel e Schraagen (2017) estudaram novamente a resiliência do sistema ferroviário em relação aos sinalizadores ferroviários, conforme diferentes situações poderiam ocorrer durante o turno de trabalho. Para isso, utilizaram, dessa vez, apenas medidas de carga mental objetiva e subjetiva, utilizando a IWS como método subjetivo e em tempo real. Nesse estudo, os autores assumiram que uma classificação da IWS maior ou igual a 5 acontecia como consequência de um evento externo sério.

Com a IWS, além de avaliar os altos e baixos da carga de trabalho durante todo o intervalo do procedimento, Siegel e Schraagen (2017) também identificaram que o método foi capaz de revelar um incidente, por meio de um valor mais alto no escore da IWS.

Broekhoven (2016) também investigou os sinais de resiliência em uma sala de controle de tráfego ferroviário, utilizando para isso a carga de trabalho. Nesse caso, o autor utilizou três medidas de carga de trabalho: a IWS, uma medida fisiológica da atividade eletrodérmica e observação comportamental. Além disso, essas três medidas foram comparadas com a XTL, para analisar se esse algoritmo é um bom indicador para prever a carga de trabalho em sinalizadores ferroviários. Assim, o autor comparou as quatro formas de medir a carga de trabalho mental.

A IWS era solicitada a cada cinco minutos, com 30 segundos de duração, na estação de trabalho do sinalizador, sendo possível acompanhar de forma automática e contínua os escores. Além disso, para manter uma alta taxa de resposta, foi possível que os sinalizadores abrissem e indicassem o escore na ferramenta durante os cinco minutos, utilizando mouse ou teclado, e também foi permitido ajuste no valor dos cinco minutos anteriores, oportunizando o operador a se concentrar principalmente em lidar com a situação. Caso nenhuma resposta fosse computada, o último valor era copiado, sob suposição de que não houve alteração da carga de trabalho. Ainda, a IWS foi agrupada: escore 1 ou 2 foi classificado com baixa carga de trabalho e escores entre 3 e 9 foram classificados como alta carga de trabalho (BROEKHOVEN, 2016).

Para comparar os métodos de avaliação da carga de trabalho, Broekhoven (2016) utilizou a IWS como medida de referência, utilizando a classificação baixa ou alta dos escores da IWS. De forma geral, os resultados mostraram que as medidas são viáveis para avaliar a CTM, porém uma melhor correspondência com a IWS acontece quando a carga de trabalho é muito alta. O grande foco do trabalho foi na avaliação da XTL para utilização como indicador para prever a carga de trabalho, o qual foi validado, utilizando comparação com a IWS.

Já Balfe, Sharples e Wilson (2015) compararam três níveis diferentes de automação na sinalização ferroviária: um nível alto, onde um agente automatizado definia rotas para trens usando informações de horários; um nível médio, onde os trens eram roteados por caminhos predefinidos, e um nível baixo, onde o operador era responsável pelo movimento de todos os trens. Para isso, foram utilizadas condições operacionais normais e anormais para os três níveis e foi avaliado o desempenho, o comportamento e a carga de trabalho subjetiva dos sinalizadores ferroviários, utilizando um simulador, o qual é utilizado para treinamento de operadores iniciantes.

Dessa forma, os autores utilizaram a escala IWS para medir a carga de trabalho percebida, a qual foi fornecida em papel aos sinalizadores, e a cada dois minutos ao longo do experimento eles faziam a avaliação do escore de forma verbal, o qual era anotado pelo pesquisador que acompanhava o experimento.

Balfe, Sharples e Wilson (2015) conseguiram encontrar relações entre o desempenho e as medidas de carga de trabalho da IWS para cada nível de automação, atingindo o objetivo do experimento.

Sendo assim, a IWS é uma ferramenta para avaliação da carga de trabalho mental muito útil para trabalhadores ferroviários, visto que foi desenvolvida especialmente para eles, apresentando como vantagens o fato de ser uma medida em tempo real, capaz de perceber os altos e baixos da carga de trabalho, sendo possível relacionar com os motivos que levaram a essas variações. Ainda é um método que foi classificado como não intrusivo e de fácil implementação.

4.2.20 Raw-NASA-TLX (RTLX)

O Raw-NASA-TLX (RTLX) é uma adaptação da ferramenta NASA-TLX, a qual é uma medida subjetiva, por meio de autorrelato, de avaliação da carga de trabalho mental. Todo o método NASA-TLX foi anteriormente descrito e analisado no tópico 4.2.2.

A adaptação feita pelo Raw-NASA-TLX consiste em uma simplificação do método original, buscando maior brevidade e rapidez nas respostas. Dessa forma, relembrando o NASA-TLX, a sua avaliação consiste em duas partes, descritas por Hart (1986). A primeira delas consiste na atribuição de pesos, entre 0 e 5, para cada uma das seis dimensões (demanda mental, demanda física, demanda temporal, desempenho, esforço e nível de frustração). Para isso, as seis dimensões são comparadas aos pares, gerando quinze pares de combinação. Dessa forma, os participantes circulam a dimensão, em cada par de combinação, que possui maior contribuição na carga de trabalho da tarefa realizada, conforme opinião própria. O peso vai ser igual à quantidade de vezes que a dimensão foi escolhida entre as combinações.

Já a segunda parte consiste na avaliação da magnitude de cada uma das seis dimensões, por meio da atribuição de um valor de 0 a 100, marcado em uma escala horizontal, originalmente de 12 cm e com 20 divisões (HART, 1986).

Por fim, a pontuação da carga de trabalho global individual para cada participante é calculada por meio da multiplicação do peso da dimensão pela sua magnitude, utilizando posteriormente a soma desses seis valores resultantes, a qual será dividida por 15, chegando ao escore final de CTM global.

Segundo Hart (2006) a modificação mais comum feita no NASA-TLX consiste na eliminação do processo de ponderação das dimensões, ou seja, é retirada

a primeira parte da avaliação do método. Essa modificação ficou conhecida como Raw-NASA-TLX e ganhou popularidade por ser mais simples e fácil de aplicar, visto que as avaliações são simplesmente calculadas para cada uma das seis dimensões e elas são adicionadas para criar uma estimativa global da carga de trabalho, sem levar em consideração qualquer ponderação.

Em 1989, três anos após o desenvolvimento do NASA-TLX, já possuíam estudos utilizando a ferramenta sem levar em consideração a etapa de ponderação das dimensões, mesmo que alguns autores não utilizassem a denominação Raw-NASA-TLX.

Em seu estudo, Byers (1989) comparou o NASA-TLX e o RTLX, buscando verificar se realmente era necessária a ponderação das dimensões. Como resultado, o autor encontrou empiricamente que os escores tradicionais do NASA-TLX e os escores considerando todas as dimensões com o mesmo peso estavam quase perfeitamente correlacionados. Assim, Byers (1989) encontrou uma escala igualmente sensível, mostrando que ficava a critério do pesquisador qual das versões utilizar.

Nygren (1991) buscou examinar questões fundamentais da medição, em sua grande maioria esquecidas durante o dimensionamento das técnicas NASA-TLX e SWAT, para avaliar a carga de trabalho mental percebida. Com base nesse estudo, o autor fez duras críticas ao modelo de ponderação do NASA-TLX, defendendo o não uso de ponderação das dimensões, por ser ineficiente, devido a restrições severas que são colocadas no modelo, com base em argumentos concretos. Para defender o seu ponto de vista, Nygren (1991) utilizou quatro principais argumentos.

O primeiro consiste no fato de que a ponderação produz apenas uma ordem relativa de importância associada a cada fator, porém, no modelo original, ela acaba se tornando uma equação de previsão linear, o que não deveria acontecer. No segundo argumento, Nygren (1991) aborda a questão do valor dos pesos: como eles variam de 0 a 5 e no final de toda a soma é utilizada a divisão por 15, devido ao número de comparações dos pares, o peso máximo que uma dimensão pode receber, utilizando uma escala de 0 a 1, é de 0,33, não refletindo a realidade de muitos casos, os quais podem ter alguma dimensão que represente mais que 33% da carga mental.

O terceiro argumento também fala dos valores dos pesos, pois, segundo Nygren (1991), como os pesos são derivados de comparações relativas entre pares, caso uma das dimensões nunca seja escolhida, ela vai receber um peso zero e, conseqüentemente, não terá impacto nenhum na definição da carga global de

trabalho, fato que muitas vezes não é real. O que essas comparações entre os pares de fatores podem indiciar é que determinada dimensão apenas é a menos importante no conjunto, mas não que ela seja necessariamente irrelevante para o cálculo da carga de trabalho.

Ao mesmo tempo, só é possível ponderar uma das dimensões com o valor zero e conseqüentemente as outras cinco dimensões serão utilizadas no cálculo da carga global de trabalho, mesmo que na prática elas não sejam relevantes, pois por esse método não é possível retirar mais que uma dimensão do cálculo final. Apesar de em algumas aplicações não se esperar que a ponderação cause grandes erros, visto que se espera que as dimensões menos importantes tenham pouca contribuição, não deixa de ser uma representação distorcida dos verdadeiros julgamentos de importância de um indivíduo para dimensões da carga de trabalho (NYGREN, 1991).

Já no quarto argumento, Nygren (1991) fala que os pesos do NASA-TLX estão condenados a ter pouco poder discriminatório real em um modelo de previsão, o que é agravado pelo fato dos pesos possuírem intervalo de 0 a 0,33 e assim possuírem pequena diferenciação intervalar.

Dessa forma, Nygren (1991) acredita que o sistema de ponderação do NASA-TLX é ineficiente e, por isso, precisa ser modificado e/ou combinado com algum outro tipo de procedimento que permite uma faixa de pesos irrestritos. Para ele, o sistema de ponderação é apenas uma ordenação da importância dimensional, não devendo ser usado para calcular um valor global de carga de trabalho.

Ainda sobre o sistema de ponderação, Nygren (1991) observou em seus experimentos que os desvios padrões dos escores utilizando ponderação são próximos aos valores que seriam esperados de um modelo simples de ponderação com pesos iguais para as dimensões, indicando que os pesos reais do NASA-TLX não acrescentaram nenhuma melhoria significativa na redução da variabilidade entre os sujeitos. Assim, muitas vezes é melhor utilizar uma ponderação igual e simples, somando-se todas as escalas e dividindo pelo número de dimensões, pois, dessa forma, o escore final será a média de todos os seis componentes de carga de trabalho, ou, caso aplicado, será uma média das dimensões que se encaixam com a tarefa executada (NYGREN, 1991).

Hendy, Hamilton e Landry (1993) apoiam a ideia de Nygren (1991) e acreditam que a ponderação do NASA-TLX pouco tem a oferecer na geração de uma estimativa de carga geral de trabalho. Ainda, os resultados do estudo deles suportam

a conclusão que uma classificação univariada da carga de trabalho pode ser mais sensível do que uma escala composta linearmente de vários fatores relacionados à carga de trabalho, que é o caso do NASA-TLX original. Além disso, uma boa estimativa pode ser obtida a partir da média aritmética simples dos fatores individuais, que é o objetivo do RTLX.

Liu e Wickens (1994) examinaram a estrutura de medidas para avaliar a carga mental na execução de tarefas de decisão e monitoramento, sendo uma dessas medidas o NASA-TLX, por meio da escala original e do Raw-NASA-TLX. Com isso, os autores sugerem que a análise das dimensões separadamente pode fornecer informações detalhadas sobre a carga de trabalho, as quais podem não ser obtidas utilizando o escore geral de carga de trabalho do método original. Porém, utilizando comparação dos dois métodos, os autores encontraram maior sensibilidade nos dados utilizando o NASA-TLX original, com ponderação das escalas.

Widyanti e Larutama (2016) observaram a relação entre o desempenho do *lean manufacturing* e a carga de trabalho mental experimentada por funcionários de uma companhia de aviões, a qual foi avaliada por meio do RTLX. Nesse caso, os autores utilizaram escalas com escore entre 0 e 10 pontos para as dimensões e encontraram valores igualmente sensíveis ao método original.

Já Hu et al. (2016) buscaram a relação entre a carga de trabalho e o desempenho de novatos em tarefas laparoscópicas, antes e após receberem treinamento. Para avaliar a carga de trabalho foi utilizado o Raw-NASA-TLX, aplicado após a realização das tarefas, analisando apenas as dimensões, por meio de um escore entre 0 e 20 pontos. Apesar dos autores considerarem o uso do NASA-TLX sem a ponderação uma limitação do estudo, apresentaram como vantagens do RTLX a maior facilidade de aplicação, principalmente em um ambiente clínico, e uma sensibilidade tão boa quanto a do NASA-TLX original.

Sendo assim, o Raw-NASA-TLX apresenta-se como um método simples e rápido para avaliar a carga de trabalho mental, o qual é apoiado em um método muito utilizado e validado, o NASA-TLX. Além disso, com base nos autores apresentados, o RTLX, comparado com a versão original, verificou-se ser mais sensível (HENDY; HAMILTON; LANDRY, 1993), menos sensível (LIU; WICKENS, 1994) ou igualmente sensível (BYERS, 1989), o que mostra que fica a critério do pesquisador qual metodologia utilizar, levando em consideração o contexto da pesquisa e os trabalhadores analisados.

4.2.21 Surgical Task Load Index (SURG-TLX) + Global Operative Assessment of Laparoscopic Skills (GOALS)

O método SURG-TLX + GOALS, utiliza conceitos dos métodos SURG-TLX e GOALS. O método SURG-TLX foi descrito e analisado anteriormente no tópico 4.2.9, então primeiramente será descrita e avaliada a metodologia da GOALS, para posteriormente discutir ambos os métodos compilados.

O método *Global Operative Assessment of Laparoscopic Skills (GOALS)*, ou, em português, Avaliação Operacional Global de Habilidades Laparoscópicas, uma medida subjetiva que pode ser utilizada para avaliar a carga de trabalho mental em cirurgias minimamente invasivas, foi desenvolvido em Montreal, no Canadá, por Vassiliou e colegas, em 2005 (GUMBS; HOGLE; FOWLER, 2007).

Vassiliou et al. (2005) desenvolveram a GOALS, pois existia a necessidade de uma ferramenta de avaliação intraoperatória, que atendesse a altos padrões de confiabilidade e validade, para as habilidades laparoscópicas.

Em seu estudo, Vassiliou et al. (2005) desenvolveram a GOALS, consistindo em uma escala global de cinco itens, uma lista de verificação com 10 itens e duas escalas de analogias visuais (VAS), para competência e dificuldade da tarefa, e utilizaram para avaliar os participantes durante uma colecistectomia laparoscópica, sendo a avaliação feita pelo cirurgião assistente, por dois observadores treinados e por autoavaliação, enquanto dissecavam a vesícula biliar.

Com o desenvolvimento da GOALS, Vassiliou et al. (2005) esperavam possuir uma ferramenta para avaliar o desempenho na sala cirúrgica, fornecendo apoio baseado em evidências para criação de treinamentos. Assim, esperava-se um impacto significativo na prática e treinamento em cirurgia minimamente invasiva, contribuindo para a definição de competência técnica e proporcionando aos residentes uma experiência educacional formativa e eficiente.

O primeiro passo no desenvolvimento da GOALS foi a identificação de aspectos importantes das habilidades técnicas exclusivas da laparoscopia. Para isso, foram utilizadas análises de vídeos de uma variedade de procedimentos laparoscópicos e foi utilizada a opinião de laparoscopistas especializados. A GOALS foi desenvolvida com base na ferramenta Avaliação Estruturada Objetiva de

Habilidades Técnicas (OSATS), criada por Martin et al. (1997) para avaliar cirurgias abertas.

Os itens na GOALS são pontuados usando uma escala Likert de 5 pontos com âncoras em 1, 3 e 5, onde “1” representa o nível mais baixo de desempenho e “5” é considerado o desempenho ideal, conforme descrição do Quadro 15. A pontuação total para a escala de classificação global é a soma das pontuações para cada um dos cinco itens, podendo chegar a uma pontuação total de 25 pontos. Os cinco itens da GOALS são descritos por Vassiliou et al. (2005):

i. Percepção de profundidade: Este item avalia quão confortável o operador está trabalhando com um sistema ótico monocular, que fornece uma imagem bidimensional em um monitor, em comparação com a visão tridimensional binocular oferecida pela cirurgia aberta. Ele descreve habilidades que variam de instrumentos amplamente oscilantes e frequentes até a direção precisa dos instrumentos no plano correto. Uma dica para uma percepção de profundidade ruim pode incluir a ativação do cautério antes do contato com os tecidos (sem perceber que o plano correto ainda não foi atingido);

ii. Destreza bimanual: O item mede quão bem o cirurgião é capaz de otimizar o uso de ambas as mãos. O operador ignora completamente a mão não dominante ou ele proporciona uma exposição ideal usando ambas as mãos de forma complementar? Se o avaliador está constantemente perguntando ao participante o que sua outra mão está fazendo, então o residente provavelmente não domina a destreza bimanual;

iii. Eficiência: Este item mede a fluidez e o progresso do procedimento. O operador muda constantemente de uma área de dissecação para outra, ou ele faz o maior progresso possível em uma área de exposição?;

Quadro 15 - Avaliação Operacional Global de Habilidades (GOALS)

Domínio	Descrições das âncoras				
	1	2	3	4	5
Percepção de profundidade	Constantes erros de alvo, grandes oscilações, lento para corrigir.		Algum erro ou falta de alvo, mas rápido para corrigir.		Direciona com precisão os instrumentos no plano correto para o alvo.
Destreza bimanual	Usa apenas uma mão, ignora a mão não dominante, falta de coordenação entre as mãos.		Usa ambas as mãos, mas não otimiza a interação entre as mãos.		Habilmente usa ambas as mãos de maneira complementar para fornecer exposição ideal.
Eficiência	Esforços incertos e ineficientes, muitos movimentos experimentais, constantes mudanças no foco ou persistência sem progresso.		Movimentos lentos, mas planejados, razoavelmente organizados.		Conduta confiante, eficiente e segura, mantém o foco na tarefa até que seja melhor realizada por meio de uma abordagem alternativa.
Manipulação de tecidos	Movimentos bruscos, rasga o tecido, fere estruturas adjacentes, controle insatisfatório da garra, pinça frequentemente escorrega.		Manipula os tecidos razoavelmente bem, traumas menores no tecido adjacente (ex. sangramento desnecessário ocasional ou escorregamento da pinça).		Lida bem com os tecidos, aplica tração apropriada, danos insignificantes nas estruturas adjacentes.
Autonomia	Incapaz de completar toda a tarefa, mesmo com orientação verbal.		Capaz de concluir a tarefa com segurança, com orientação moderada.		Capaz de concluir a tarefa de forma independente, sem solicitar orientação.

Fonte: Adaptado de Vassiliou et al. (2005).

iv. Manipulação de tecidos: O quarto item mede o manuseio adequado dos tecidos, o que também inclui o uso apropriado dos instrumentos laparoscópicos. Este item tenta verificar quão bem o operador é capaz de se adaptar ao efeito de fulcro de instrumentos longos por meio de trocartes e à perda relativa de *feedback* tátil. Os instrumentos são usados com cautela e são usados apropriadamente, dependendo da qualidade dos tecidos no caso em questão?;

v. Autonomia: O item final avalia a independência técnica, medindo a quantidade de orientação que o operador precisa (se houver) para executar a tarefa de maneira segura e apropriada. Se o cirurgião responsável assumir a operação, porque o operador não conseguiu progredir ou não pôde realizar o procedimento com segurança, isso seria uma evidência para atribuir uma pontuação baixa para a autonomia.

Além da GOALS, foi desenvolvida uma lista de verificação específica de 10 itens para a atividade de dissecação da vesícula biliar do fígado. Essa lista de verificação consiste em 10 itens para verificar se eles já estão prontos ou não. Ela foi desenvolvida por meio de consulta com especialistas em laparoscopia e inclui os itens disponíveis no Quadro 16. Os itens recebem 1 ponto se foram feitos corretamente e nenhum ponto se eles não foram feitos, chegando em um total máximo de 10 pontos.

Quadro 16 – Lista de verificação para dissecação da vesícula biliar

Tarefas	Feito (1 ponto)	Não feito (0 pontos)
1. Utiliza cautério apenas quando todas as áreas condutoras estão no campo de visão.		
2. Tem bom controle do instrumento, minimiza o recuo.		
3. Segura a vesícula biliar perto de cliques para iniciar a dissecação.		
4. Reajusta a tensão na vesícula biliar para otimizar a exposição.		
5. Evita dissecação no fígado causando sangramento excessivo.		
6. Evita a perfuração da vesícula biliar.		
7. Evita o derramamento de cálculos biliares.		
8. Maximiza a dissecação útil em 1 área antes de mudar a abordagem.		
9. Realiza dissecação em plano apropriado a maior parte do tempo.		
10. Evita a necessidade de aquisição de cirurgião.		
Total		

Fonte: Adaptado de Vassiliou et al. (2005).

Por fim, ainda foram desenvolvidas duas escalas de analogias visuais (VAS) para competência geral e dificuldade do caso. Cada VAS consiste em uma linha de 10 cm, com âncoras descritivas em cada extremidade. O avaliador é instruído a

colocar um “X” na linha para indicar a pontuação que ele atribui para a construção específica que a escala está medindo.

A primeira VAS mede quão tecnicamente difícil é o procedimento por razões de anatomia, patologia, planos teciduais ou equipamentos, possuindo na extremidade esquerda a palavra “extremamente fácil” e na extremidade direita a palavra “extremamente difícil”. A segunda VAS mede a competência técnica geral, com a extremidade esquerda possuindo “não foi possível concluir a tarefa com a orientação máxima” e a extremidade direita com “poderia executar a tarefa de forma segura e independente (totalmente competente)”. As pontuações são determinadas numericamente, medindo a distância ao longo da linha de 10 cm onde a marca foi colocada.

A validação da GOALS, da lista de verificação e das escalas VAS foi feita durante a dissecação do leito hepático da vesícula biliar como parte de uma colecistectomia laparoscópica eletiva. A colecistectomia laparoscópica foi escolhida porque é um procedimento comum realizado por muitos cirurgiões gerais. Para isso, os participantes foram avaliados pelo cirurgião responsável, por dois observadores treinados e por autoavaliação.

Para avaliar a consistência interna dos itens da GOALS foi utilizado o alfa de Cronbach, que foi calculado para os conjuntos de dados correspondentes a cada avaliador. Os coeficientes de correlação intraclasse (ICCs) avaliaram a confiabilidade interavaliadores da GOALS, a VAS para a competência, a VAS para a dificuldade e a lista de verificação da tarefa, sendo que os ICCs foram calculados para cada um dos cinco itens da GOALS, bem como para o score total. Para cada variável, as médias de grupos iniciantes e experientes foram comparadas usando os testes t e teste U de Mann-Whitney.

Como resultados do estudo de Vassiliou et al. (2005) a consistência interna da GOALS foi excelente para todos os quatro grupos de avaliadores e a lista de verificação demonstrou confiabilidade, entre avaliadores, inferior em comparação com a GOALS. Ainda, diferente da GOALS, a lista de verificação não identificou uma diferença significativa entre operadores novatos e experientes, sugerindo que ela não é uma medida válida de habilidade operacional. Já a VAS para competência geral, embora fornecesse ao avaliador uma orientação menos estruturada do que a GOALS, demonstrou ter uma confiabilidade interavaliador e demonstrou evidências de validade de construto.

Assim, a GOALS se apresentou como uma “combinação correta” de estrutura e objetividade, ao mesmo tempo em que permite que o avaliador incorpore sua experiência para uma avaliação mais abrangente e, possivelmente, mais útil. Além disso, os dados sugerem que a lista de verificação não aumenta a eficácia da avaliação intraoperatória e não suporta sua inclusão continuada (VASSILIOU et al., 2005).

A GOALS também é superior a uma lista de verificação ou a uma VAS, pois fornece aos residentes um *feedback* formativo e somativo. O *feedback* relacionado a componentes específicos das habilidades técnicas é muito atraente quando utilizado em programas de treinamento. Além disso, é interessante o *feedback*, pois os residentes poderão direcionar sua prática e melhorar seu desempenho de acordo com as informações concretas recebidas.

Ainda a alta correlação entre os escores do cirurgião e do observador implica que o cirurgião responsável pode concluir com segurança a avaliação sem a necessidade de um observador na sala de operações, fato que é menos dispendioso e demorado, favorecendo assim o uso prático da GOALS. Além disso, os cinco itens que compõem a GOALS são específicos para habilidades laparoscópicas, mas eles não se concentram em um procedimento específico, o que está em contraste com a lista de verificação, que contém itens relacionados especificamente à remoção da vesícula biliar do leito hepático. Assim sendo, a GOALS poderia ser utilizada para outros procedimentos, diferente da lista de verificação, que precisaria, no mínimo, de adaptações.

Com esse estudo, Vassiliou et al. (2005) concluíram que a GOALS é viável, confiável e válida, sendo superior a lista de tarefas e a VAS para avaliação de habilidade técnica por avaliadores experientes. Assim, os resultados apoiam o uso da GOALS no treinamento e avaliação de habilidades laparoscópicas, podendo ser um meio eficaz de fornecer *feedback* formativo e focado sobre as habilidades técnicas em cirurgia minimamente invasiva. Ainda, a GOALS pode ser usada para medir a eficácia do treinamento em simuladores para melhorar o desempenho operacional (VASSILIOU et al., 2005).

Vassilio et al. (2007) investigaram a confiabilidade e a validade da GOALS aplicada em análises por vídeo, pois a observação direta pode muitas vezes apresentar vieses relacionados a fatores pessoais e profissionais, no sentido de desencorajar os avaliadores a dar um *feedback* negativo, e de ser intensivo em termos

de tempo e trabalho. Além disso, como é utilizado vídeo para realizar o procedimento laparoscópico, a gravação acontece de forma prática e inerente ao equipamento, tornando esse método atraente e aparentemente menos tendencioso na avaliação da habilidade intraoperatória.

Para isso, foram estudados cinco cirurgiões iniciantes e cinco cirurgiões experientes, durante uma colecistectomia laparoscópica. Para avaliação, foram utilizados dois observadores diretos e, posteriormente, quatro laparoscopistas avaliaram os procedimentos gravados, usando os cinco itens da GOALS e sua pontuação total. Para avaliar a confiabilidade entre avaliadores diretos, entre avaliadores de vídeo e entre avaliadores diretos e por vídeo, foi calculado o coeficiente de correlação intraclasse. Além disso, a validade de construto foi avaliada pela análise de variância.

Por meio dos dados do estudo, a GOALS pode ser usada para avaliar a habilidade laparoscópica com base em performances gravadas em vídeo, mas o treinamento do avaliador desempenha um papel importante na garantia da confiabilidade e validade do instrumento. Além disso, a experiência com a ferramenta na sala de operação pode melhorar a confiabilidade da classificação do vídeo e pode ser útil para avaliadores de treinamentos. Assim, a GOALS pode ser utilizada tanto por vídeo como por classificação direta, o que aumenta sua versatilidade, preservando sua confiabilidade e validade. Uma vantagem demonstrada é que se um observador externo precisa avaliar o procedimento, o uso do vídeo pode não ser apenas mais eficiente em termos de tempo, mas também mais prático e flexível (VASSILIO et al., 2007).

Chang et al. (2007) realizaram um estudo com o mesmo objetivo do estudo de Vassilio et al. (2007), buscando validar a GOALS para procedimentos analisados por vídeo, utilizando também cirurgiões novos e experientes. Nesse estudo, além de utilizar as cinco escalas da GOALS, foi avaliado o grau de dificuldade da operação levando em conta fatores como inflamações, aderências, hábitos corporais e variações anatômicas, conforme escala VAS de Vassiliou et al. (2005). As pontuações para os cirurgiões experientes e novatos foram comparadas e analisadas estatisticamente usando análise de variância de fator único. Além disso, a confiabilidade interobservadores foi calculada utilizando o alfa de Cronbach.

Por meio desse estudo foi documentada a consistência dentro do grupo de revisores para cada fita de vídeo e demonstrado que a GOALS pode ser utilizada para

quantificar e diferenciar a habilidade cirúrgica laparoscópica de um cirurgião novato de um experiente, utilizando o desempenho gravado por vídeo (CHANG et al., 2007).

Gumbs, Hogle e Fowler (2007) buscaram validar a GOALS para todo o procedimento de colecistectomia laparoscópica e também para apendicectomia laparoscópica, utilizando, além dos cinco domínios da GOALS, a dificuldade do procedimento, em uma escala Likert de 5 pontos, da mesma forma que Chang et al. (2007). Para comprovar os resultados e comparar os escores médios dos participantes, em ambos os casos foi utilizado teste t.

Esse estudo também mostrou que foi possível diferenciar os grupos de cirurgiões residentes e experientes na avaliação de procedimentos laparoscópicos completos pela GOALS. Como vantagens do método, o baixo custo, o pouco tempo para completar as escalas e a facilidade para obter os dados ao final de cada procedimento foram citados pelos autores. Além disso, a GOALS inclui padronização, exposição de deficiências e avaliação de padrões técnicos mínimos em programas de residência (GUMBS; HOGLE; FOWLER, 2007).

Já Kramp et al. (2015) estimaram a validade e confiabilidade da GOALS para cirurgiões residentes com pouca experiência, durante cirurgia laparoscópica. Nesse estudo, a GOALS consistiu em seis itens, sendo os cinco itens normais da GOALS (percepção de profundidade, destreza bimanual, eficiência, manipulação de tecidos e autonomia) e um item referente ao nível de dificuldade, da mesma forma que os dois estudos anteriores. Para poder comparar com a OSATS, utilizada para cirurgias abertas, os itens da GOALS foram convertidos para uma escala de 10 pontos, em vez da habitual escala Likert de 5 pontos.

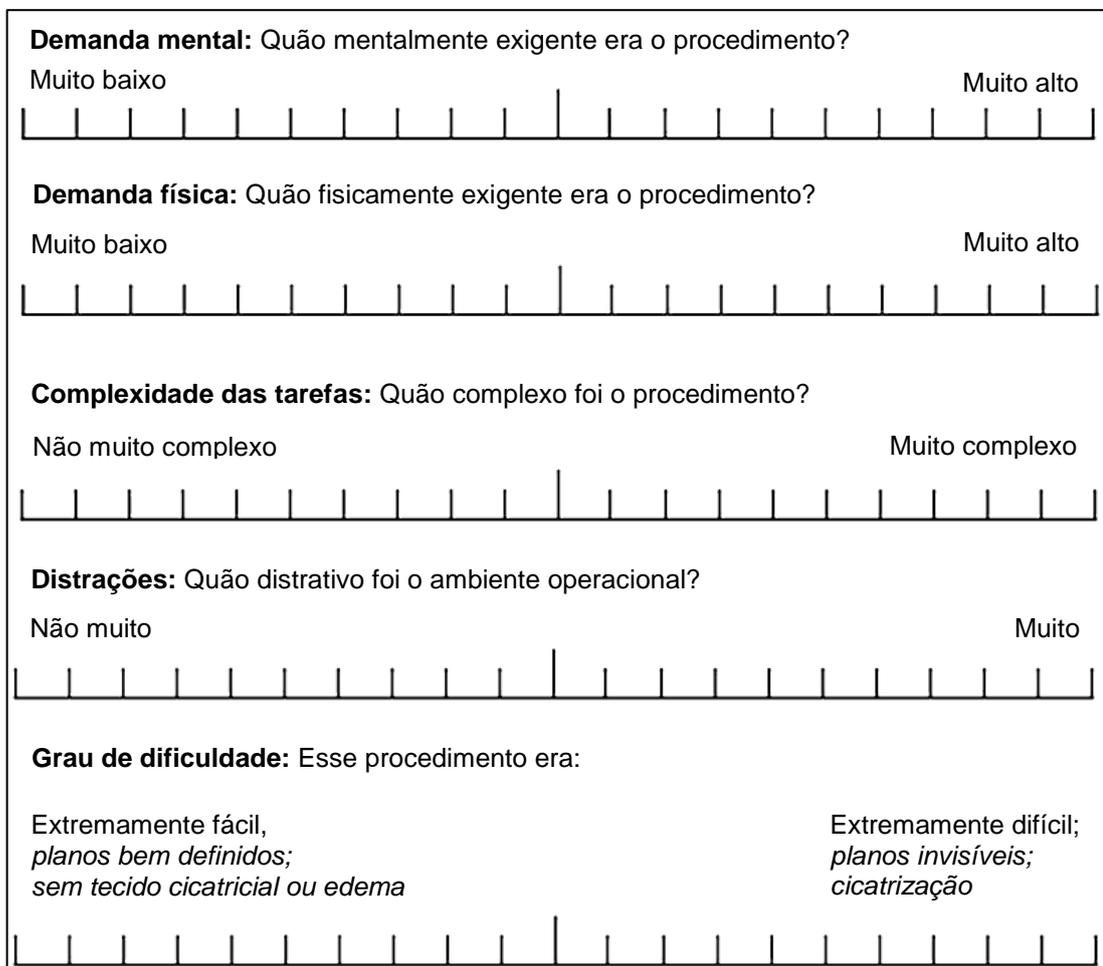
Os autores calcularam a correlação dos escores GOALS com os da Avaliação Estruturada Objetiva de Habilidades Técnicas. Além disso, a validade de construto foi estimada, juntamente com a confiabilidade intraclasse do modelo.

Por meio desses testes, a GOALS mostrou constructos significantes e válidos para avaliação de aprendizes ou novatos. Ainda, quando comparada com a OSATS, a GOALS contém critérios para avaliar as cirurgias minimamente invasivas (KRAMP et al., 2015).

Como tanto o SURG-TLX como a GOALS avaliam cirurgiões, Yu et al. (2016) e Hallbeck et al. (2017) resolveram utilizar em seus trabalhos a junção desses dois métodos, porém utilizando apenas uma parte de cada um deles.

Assim, uma versão modificada do questionário validado SURG-TLX foi utilizada com a adição da pergunta da GOALS para aumentar sua relevância na mensuração da carga de trabalho intraoperatória. Esse questionário contém quatro dimensões do SURG-TLX (demanda mental, demanda física, complexidade e distração) e uma pergunta da GOALS, em relação ao grau de dificuldade, de extremamente fácil até extremamente difícil, conforme Figura 22. A classificação do participante para cada escala é relatada de 0 a 100 pontos, sendo que a escala possui 20 divisões.

Figura 22 - Questionário SURG-TLX + GOALS



Fonte: Adaptado de Yu et al. (2016).

Yu et al. (2016) administraram o questionário mostrado na Figura 22 a cada membro da equipe cirúrgica imediatamente após cada procedimento cirúrgico. Os resultados foram apresentados por meio de um gráfico para cada uma das cinco dimensões, considerando cada uma das funções dos membros da equipe cirúrgica.

A comparação do tempo entre as técnicas cirúrgicas foi realizada com o Teste U de Mann-Whitney. Para cada variável dependente do questionário, foi utilizado um modelo de efeitos mistos com combinações fatoriais completas de fatores do caso. Além disso, o tempo de procedimento foi incluído como uma covariável para ajustar a duração do procedimento. Testes post hoc para múltiplas comparações foram realizados com correção de Bonferroni.

O estudo demonstrou que esse questionário é uma ferramenta que permite detectar diferenças na carga de trabalho intraoperatória para variações na complexidade do caso e entre diferentes funções dos membros da equipe cirúrgica, fornecendo métricas quantitativas para identificar limitações da carga de trabalho e problemas de desempenho na sala de cirurgia. Ainda, esta ferramenta de questionário pode ser usada para monitorar aumentos potenciais ou desequilíbrio na carga de trabalho como resultado de inovações em tecnologias robóticas ou laparoscópicas (YU et al., 2016).

Já no estudo de Hallbeck et al. (2017), foi estudada a incorporação de pequenos intervalos de 5 minutos, após cada período de 30 minutos de trabalho, durante os procedimentos cirúrgicos, buscando reduzir os riscos ergonômicos na equipe cirúrgica.

Antes dos dias operatórios estudados, cada cirurgião teve que preencher um questionário preliminar, o qual cobriu o desconforto corporal autorrelatado relacionado ao trabalho, o que o cirurgião fez para mitigar esse desconforto, se eles sentiram que a dor afetaria a capacidade de realizar seu trabalho, o nível de dor nas últimas 24h após a cirurgia e como essa dor impactou o sono, juntamente com informações demográficas.

Nesse estudo, o SURG-TLX e a pergunta da GOALS também possuíam uma escala de avaliação de 0 a 100 pontos e por meio desse questionário foi possível responder as hipóteses da pesquisa (HALLBECK et al., 2017).

De forma geral, esse questionário mostrou-se uma ferramenta eficaz para quantificar a carga de trabalho intraoperatória em toda a equipe cirúrgica, garantindo que as demandas físicas e mentais não excedam os limites, a partir do qual o desempenho pode diminuir e o risco de lesões pode aumentar. Além disso, essa ferramenta tem o potencial de medir a segurança dos procedimentos atuais e orientar o projeto de intervenções de carga de trabalho (YU et al., 2016).

4.2.22 Traffic Load Index (TLI)

O *Traffic Load Index* (TLI), ou, em português, Índice de Carga de Tráfego, é um instrumento que combina medidas objetivas e medidas subjetivas para avaliar a carga de trabalho mental.

Os primeiros estudos que utilizaram esse indicador foram os de Athènes et al. (2002) e os de Averty et al. (2002, 2004), todos realizados na França, porém a melhor descrição está contida no trabalho de Averty et al. (2004).

Como o trabalho dos controladores de tráfego aéreo exige certo esforço cognitivo, muitos estudos foram realizados com esses trabalhadores, porém a avaliação da carga de trabalho mental acontecia somente por aspectos objetivos, como pelo número e distribuição de aeronaves controladas por cada ATC por meio de variáveis do tráfego, ou somente por fatores subjetivos, como avaliação do desempenho por terceiros ou métodos já validados, como o NASA-TLX e o SWAT.

Assim, Averty et al. (2004) desenvolveram um método que abrangesse tanto medidas objetivas quanto medidas subjetivas, utilizando variáveis objetivas do tráfego e uma carga adicional imposta por efeitos subjetivos, incluindo gravidade dos conflitos e pressão do tempo para a sua resolução, o qual foi denominado TLI.

A justificativa para desenvolver um novo método foi devido ao fato de o indicador número de aeronaves (N), segundo Averty et al. (2004), apesar de facilmente gravado, não ser suficiente, pois se deve levar em consideração a situação do tráfego, ou seja, a maneira como as aeronaves se espalham pelo espaço e pelo tempo.

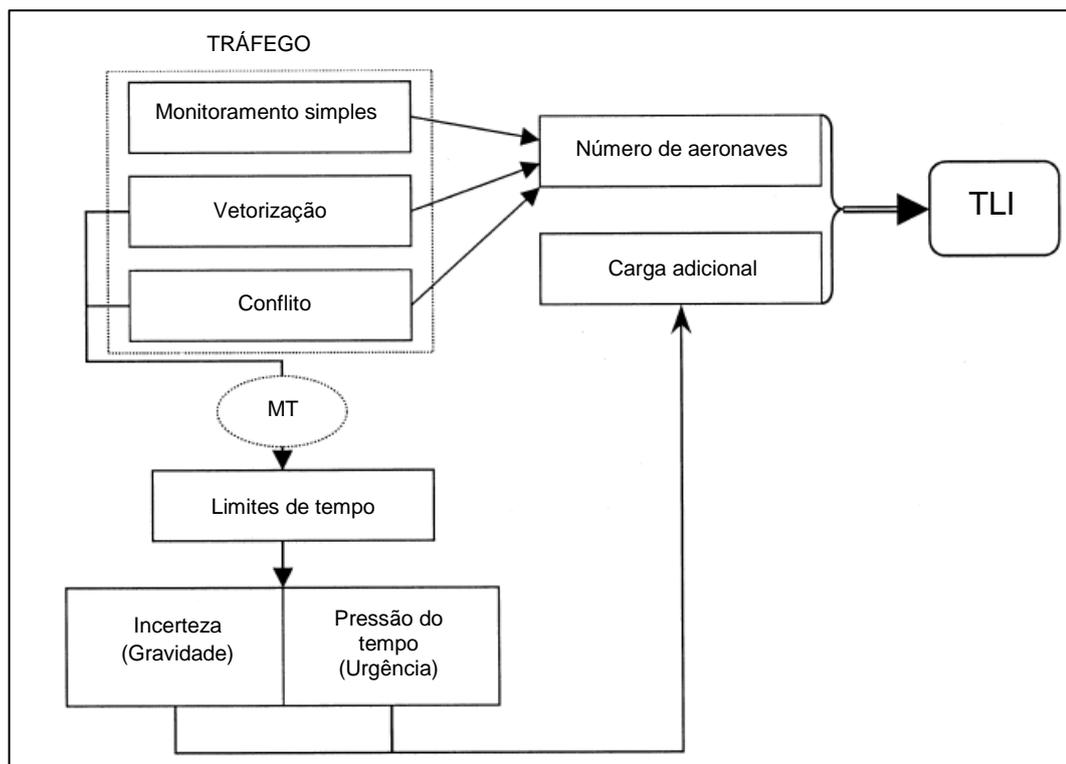
Assim, além do índice N, é importante a forma como as aeronaves se comportam e as decisões que devem ser tomadas em situações “fora do normal”, as quais envolvem o tempo de resposta (MT) a um diagnóstico de conflito e o momento em que a decisão é tomada e a incerteza da decisão, pois acredita-se que quanto maior o MT, apesar da solução ser mais precisa, maior será o estado de estresse do ATC. Dessa forma, os autores acreditam que além do número de aeronaves, deve-se levar em consideração a carga adicional resultante da ameaça de excitação excessiva, a qual é utilizada no cálculo do TLI.

Esse índice foi desenvolvido a partir da análise de 25 horas de gravações de sessões de controle, envolvendo 25 ATCs qualificados que trabalhavam de

maneira habitual. Para posterior validação e justificativa do instrumento, o TLI foi comparado com o índice objetivo simples do número de aeronaves e avaliações subjetivas da carga de trabalho, pelo NASA-TLX, por meio de cinco fatores: demanda mental, demanda temporal, desempenho, esforço e frustração (AVERTY et al., 2004).

A Figura 23 mostra um esquema de como é calculado o Índice de Carga de Tráfego, o qual é descrito por Averty et al. (2004).

Figura 23 - Diagrama esquemático para cálculo do TLI



Fonte: Adaptado de Averty et al. (2014).

A primeira etapa consiste em categorizar as aeronaves em duas classes: monitoramento simples ou envolvidas em algum problema de controle. As aeronaves envolvidas em algum problema são categorizadas de duas formas: vetorização, quando as aeronaves convergem para um mesmo aeroporto e precisam de instruções sobre vetorização para fornecer o espaçamento correto, e conflito, quando duas aeronaves estão separadas por uma distância menor do que os intervalos mínimos de segurança (AVERTY et al., 2004).

Durante o MT, intervalo de tempo durante o qual o problema de controle é diagnosticado e solucionado, acontece a carga de trabalho adicional, resultante da vetorização ou do conflito, a qual é definida pela pressão do tempo e a incerteza, as

quais são transformadas em dados numéricos para serem adicionadas ao N (AVERTY et al., 2004).

Para isso, Averty et al. (2014) calculam um índice de carga adicional para cada aeronave que está sob controle do ATC. Convencionalmente, o valor correspondendo a cada aeronave monitorada é igual a 1. Assim, quando a aeronave apresenta algum problema, ela receberá um valor maior que 1, dentro do intervalo dos limites de tempo.

A gravidade, ou seriedade, é definida pela distância entre duas aeronaves e a interseção das trajetórias de voo, a qual é classificada em quatro categorias, de A (espaçamento esperado é inadequado para garantir segurança total) a D (risco máximo de conflito). Ainda, deve-se avaliar a pressão do tempo por meio da variável urgência. Assim, utilizando-se da avaliação de um especialista, a aeronave recebe um valor entre 1 e 3.5, sendo que cada categoria possui um valor máximo ($A_{\text{máx}} = 2.0$, $B_{\text{máx}} = 2.5$, $C_{\text{máx}} = 3.0$, $D_{\text{máx}} = 3.5$), o qual é determinado com base na gravidade e urgência de cada aeronave (ATHÈNES et al., 2002; AVERTY et al., 2004).

Como a carga adicional é um fator de cada aeronave individualmente, Averty et al. (2014) adicionam o valor de carga adicional obtido para cada aeronave, em um dado momento, ao número de aeronaves, para chegar ao TLI, o qual é calculado a cada 10 segundos, visto que a carga adicional muda conforme o MT.

Assim, após cálculo do TLI, Averty et al. (2014) compararam o índice com as avaliações subjetivas do NASA-TLX e do índice objetivo N, buscando analisar a confiabilidade e a validade do TLI, utilizando condições reais de trabalho em campo.

Com os resultados, Averty et al. (2014) mostraram uma melhor correlação do TLI com o NASA-TLX do que entre o N e o NASA-TLX, mostrando que é melhor utilizar um índice que possui medidas objetivas e subjetivas, caso do TLI, em vez de apenas um índice objetivo simples.

Athènes et al. (2002) correlacionaram medidas objetivas, o TLI, o NASA-TLX e avaliações fisiológicas da carga de trabalho em situações reais de trabalho do ATC, pois acreditavam que a pressão do tempo e a incerteza, utilizadas no cálculo do TLI, influenciam no estado emocional dos controladores, agindo no nível de excitação fisiológica do organismo.

De acordo com os resultados de Athènes et al. (2002), o TLI se mostrou mais correlacionado com o NASA-TLX e com os parâmetros fisiológicas do que o

índice objetivo simples do número de aeronaves monitoradas, mostrando a vantagem de acrescentar estimativas de incerteza e pressão do tempo ao índice N.

Da mesma forma, Averty et al. (2002) utilizaram parâmetros cardíacos e a atividade eletrodérmica para avaliar o comportamento emocional do trabalho em condições reais dos ATCs, em comparação com o N, o TLI e o NASA-TLX.

Novamente os resultados mostraram uma melhor correlação do NASA-TLX com o TLI em comparação com o N. Ainda, o TLI mostrou-se mais correlacionado com a atividade fisiológica dos controladores do que o N e o NASA-TLX, independente do parâmetro fisiológico considerado (AVERTY et al., 2002).

Uma classificação interessante feita por Averty et al. (2002) foi a separação do TLI em cinco classes, correspondendo a cinco níveis diferentes de carga de trabalho global, conforme Quadro 17.

Quadro 17 - Classes da carga de trabalho de acordo com o TLI

Nível de carga de trabalho	Avaliação global da carga de tráfego	TLI
Nível 1	Subcarga (risco de hipovigilância).	$TLI \leq 2.5$
Nível 2	Carga baixa.	$2.5 < TLI \leq 5.5$
Nível 3	Carga padrão.	$5.5 < TLI \leq 9.5$
Nível 4	Carga elevada.	$9.5 < TLI \leq 14$
Nível 5	Sobrecarga (risco de estresse).	$TLI > 14$

Fonte: Adaptado de Averty et al. (2002).

Para Averty et al. (2002), o TLI é computável a partir de dados de voo de tráfego e usa uma escala de tempo de alta resolução idêntica à da tela do radar, permitindo uma conexão contínua entre aspectos subjetivos da carga de trabalho e dados objetivos. Porém, conforme mostrado em seu estudo, é vantajoso o uso juntamente com parâmetros fisiológicos.

De forma geral, Averty et al. (2004) acreditam que a principal vantagem de utilizar o TLI é que ele possui dados objetivos, por meio do número de aeronaves controladas, e dados subjetivos, definidos pela carga adicional associada a cada situação. Ainda, esse índice é baseado em regras simples, as quais, apesar das particularidades dos procedimentos, podem ser bastante similares, sendo possível generalizações.

Porém, Averty et al. (2004) acreditam que ainda precisa ser construído um modelo de trabalho geral, por meio de um algoritmo capaz de integrar os recursos de

tráfego relevantes, além de investigar dados do tráfego aéreo considerando a reatividade emocional dos conflitos, permitindo calcular o TLI em tempo real.

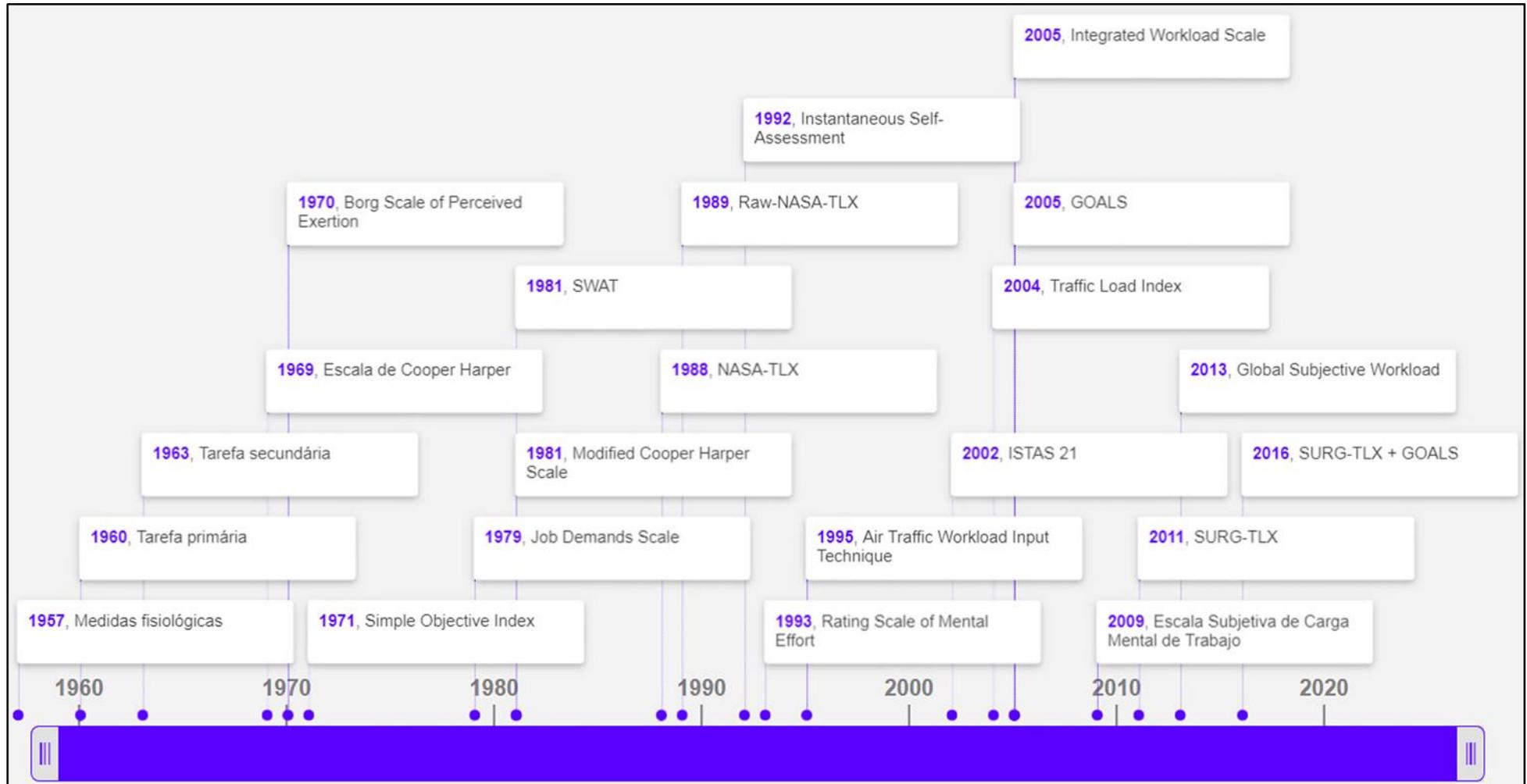
4.3 ANÁLISE GERAL DOS MÉTODOS DE AVALIAÇÃO DA CTM

Com exceção das ferramentas “questionário” e “escala subjetiva”, as quais não são propriamente únicas e normalmente não possuem validação, os demais métodos encontrados no portfólio bibliográfico, e discutidos na seção 4.2, estão traçados em uma linha do tempo na Figura 24.

A linha do tempo tem início no século XX, mais precisamente em 1957, com as medidas fisiológicas e já na década de 1960 apareceram as medidas de tarefa primária e secundária e a escala de Cooper Harper. Na década de 1970 começaram a ser utilizadas a Escala de Borg de Esforço Percebido, o Índice Objetivo Simples e a Escala de Demandas de Trabalho. Na década de 1980 a escala modificada de Cooper Harper, o SWAT, o NASA-TLX e o Raw-NASA-TLX ganharam suas versões, a partir da escala de Cooper Harper original. Na década de 1990 a Autoavaliação Instantânea, a Escala de Avaliação do Esforço Mental e a Técnica de Entrada de Carga de Trabalho de Tráfego Aéreo ganharam forma. Passando para o século XXI, na década de 2000, o ISTAS 21, o Índice de Carga de Tráfego, a Avaliação Operacional Global de Habilidades Laparoscópicas, a Escala Integrada de Carga de Trabalho e a Escala Subjetiva de Carga de Trabalho apareceram e, por último, na década atual, os métodos SURG-TLX, Carga de Trabalho Subjetiva Global e a junção do SURG-TLX com a GOALS surgiram.

Assim sendo, é possível notar que primeiramente as medidas fisiológicas começaram a ser usadas, seguidas pelas medidas de performance e por último os métodos subjetivos.

Figura 24 - Linha do tempo dos métodos de avaliação da CTM



Fonte: Elaborado pelo autor (2018).

As medidas fisiológicas são medidas continuamente, possuem respostas diretas, de natureza multidimensional, e, de forma geral, as coletas não interferem no desempenho do trabalhador. As medidas fisiológicas são utilizadas nos mais variados segmentos, podendo-se utilizar diversos indicadores, ou sentidos diferentes, buscando relações, sendo possível distinguir cargas de trabalho baixas, moderadas e altas.

Porém, requerem altos custos com equipamentos e *softwares* para avaliação, além de exigirem cuidados na avaliação e interpretação dos resultados, os quais possuem certa complexidade, muitas vezes precisando ser comparados com medidas de desempenho ou métodos subjetivos para fornecer uma estimativa da carga de trabalho mental, visto que nem sempre as flutuações nas medidas fisiológicas são relacionadas a demandas mentais do trabalho, e, em alguns casos, os equipamentos podem ser intrusivos. Outra desvantagem existente é a dificuldade em utilizar medidas fisiológicas nas empresas, devido à falta de validação por comitês de ética.

As medidas de desempenho são uma forma objetiva de avaliar a carga de trabalho mental, sendo representadas no PB pelo Índice Objetivo Simples e pelas tarefas primárias e secundárias, e consistem na análise da performance dos trabalhadores, com base no desempenho individual. O SOI e a tarefa primária consistem na mensuração da tarefa principal do operador de forma não intrusiva, sendo muito utilizados em controladores de tráfego aéreo. Para aumentar a validade dos resultados, podem-se utilizar múltiplas medidas e também o uso de medidas fisiológicas ou subjetivas concomitantemente, pois uma medida simples do desempenho não é capaz de fornecer resultados válidos de CTM. Além disso, esses métodos possuem dificuldades de generalizações das atividades, não diagnosticam a fonte de carga e não refletem os investimentos em recursos causados por mudanças nas demandas de tarefas.

Já as tarefas secundárias utilizam avaliações do desempenho de uma tarefa secundária, realizada concomitantemente com a tarefa principal. Da mesma forma que as medidas diretas de desempenho, é um método não intrusivo, rápido e fácil de ser aplicado. Porém, como limitações, algumas medidas podem ter certa intrusividade, desconcentrando a realização da tarefa principal ou até mesmo distraindo outras pessoas que trabalham no mesmo local; a tarefa principal pode ser preterida pela tarefa secundária, resultando em consequências que muitas vezes não

refletem a realidade das tarefas; e a avaliação dos resultados pode ser difícil, pois são necessárias as determinações de medidas padrões para comparação. Além disso, tanto nas tarefas primárias como secundárias, as mudanças no desempenho podem não significar mudanças na carga de trabalho mental.

As medidas subjetivas tiveram início com a escala de Cooper Harper, por meio de uma escala unidimensional que utiliza uma árvore de decisões para chegar ao escore final. Como essa escala é voltada para tarefas manuais de voo, adaptações aconteceram, dando origem a escala modificada de Cooper Harper, focada em trabalhos mais automatizados, aumentando as aplicações da CH. Por serem escalas unidimensionais, elas são adequadas para situações que possuem limitações de tempo, pois possuem aplicações rápidas, de fácil entendimento e de fácil avaliação dos resultados. Porém, ainda por serem unidimensionais, não distinguem as diferentes demandas de processamento de tarefas.

Um ano depois da escala de Cooper Harper, surgiu a Escala de Borg de Esforço Percebido, também unidimensional, porém utilizada na área de saúde. Essa escala originalmente foi construída para analisar o esforço físico e até hoje é utilizada para isso, porém existem aplicações que avaliam a carga de trabalho mental. Essa escala apresenta rápida e fácil aplicação, é simples de responder, não exigindo muitos conhecimentos dos trabalhadores, e é correlacionada com a frequência cardíaca. Porém, como a CH, não é capaz de distinguir os vários aspectos da carga de trabalho, por ser unidimensional.

Um tempo depois surgiram a Autoavaliação Instantânea, a Técnica de Entrada de Carga de Trabalho de Tráfego Aéreo, a Escala Integrada de Carga de Trabalho, a Escala de Avaliação do Esforço Mental e o Índice de Carga de Tráfego, como escalas unidimensionais.

A ISA é aplicada durante a realização das tarefas e possui várias adaptações no tamanho da sua escala. Foi desenvolvida para fornecer avaliações subjetivas da carga de trabalho durante tarefas de controle de tráfego aéreo, por meio de resposta em teclado ou falada. É uma ferramenta rápida e fácil de responder e avaliar e muito útil para situações longas, que envolvem várias tarefas ou fases diferentes, não sendo dependente da memória dos participantes. Porém, ela pode atrapalhar a execução da tarefa primária, sendo intrusiva, principalmente em situações em que a ISA utiliza os mesmos recursos sensoriais.

A ATWIT possui o mesmo princípio da ISA e o mesmo público alvo, sendo respondida por teclado físico ou *touch screen*, possuindo, assim, as mesmas vantagens e desvantagens da ISA. Por ser utilizada de forma computadorizada, necessita de um *software* para gerenciamento dos intervalos de solicitações das classificações, captação das respostas e apresentação dos resultados, os quais são medidos em tempo real e apresentam as flutuações da carga de trabalho, sendo possível a adaptação a diferentes contextos de trabalho.

A IWS também possui o mesmo princípio da ISA e da ATWIT, porém foi desenvolvida para o setor ferroviário, principalmente para sinalizadores ferroviários. Ela apresenta os picos de carga de trabalho, sendo aplicada de forma manual, por papel, ou automática, por computador, e por meio da linha de tempo de trabalho é possível relacionar os resultados com as situações. O método é classificado como não intrusivo e de fácil implementação, porém pode ser considerado intrusivo em algumas situações, dependendo da forma de coleta das informações.

A Escala de Avaliação do Esforço Mental, inspirada na ISA e no NASA-TLX, também possui uma escala unidimensional, com escores entre 0 e 150 pontos, e alguns descritores, e é focada na avaliação do esforço, sendo utilizada também para a CTM. A RSME não possui aplicação específica para uma classe de trabalhadores, não requer nenhum dispositivo especial, é simples e fácil de usar, barata, rápida, não interfere no trabalho dos participantes e apresenta resultados tão sensíveis a mudanças na carga mental quanto as escalas multidimensionais, sendo adequada para o emprego em ambientes reais de trabalho. Porém, a adaptação em países diferentes da escala desenvolvida inicialmente pode ser complicada, pois os termos qualitativos podem ser de difícil adequação linguística e a cultura interfere na interpretação do instrumento. Além disso, como ela é realizada após as tarefas, é dependente da memória.

Ainda, como escala unidimensional, existe o Índice de Carga de Tráfego. Porém, diferente das ferramentas mencionadas anteriormente, ele combina medidas objetivas e subjetivas para avaliar a carga de trabalho mental de controladores de tráfego aéreo. Como medida objetiva é utilizado o número de aeronaves, que é um dado de tarefa primária, e como medida subjetiva é utilizado um fator para cada aeronave, o qual depende da gravidade e da urgência dos problemas encontrados, conforme a situação do tráfego aéreo. A principal vantagem do método é a utilização

de dados objetivos e subjetivos, possuindo um procedimento simples e permitindo generalizações.

Além do mais, em relação aos métodos subjetivos para avaliar a carga de trabalho mental, muitos deles possuem escalas multidimensionais, levando em consideração várias dimensões para compor a carga mental, podendo avaliá-la de forma separada ou por um índice geral obtido a partir delas.

Mesmo não sendo os dois primeiros métodos a serem desenvolvidos, o NASA-TLX e o SWAT possuem certo destaque entre as ferramentas subjetivas com escala multidimensional. O NASA-TLX, com seis dimensões, possui uma rápida aplicação e fácil análise dos dados. Além disso, como é o método mais usual para avaliar a CTM, possui muitos estudos para comparação. Porém, como um método subjetivo, depende da memória do trabalhador, visto que é aplicado ao final das tarefas.

Outra desvantagem é a ponderação das escalas, que pode ser um processo que mesmo levando um tempo demasiado, pode não atingir o objetivo esperado, não fornecendo um valor real e não sendo aplicável em muitas situações. Para resolver esse problema, foi formulado o RTLX, que é uma adaptação do NASA-TLX, retirando a ponderação das escalas. Assim, esse método é mais simples e rápido que o formato original e mostra-se, em algumas situações, até mais sensível que o NASA-TLX.

Ainda, em relação a ponderação das escalas do NASA-TLX também pode ser utilizado um modelo Fuzzy, o qual pode ser agregado por meio da Escala Global de Carga de Trabalho Subjetiva, a qual é unidimensional e consiste em uma pergunta sobre a carga de trabalho global para executar a tarefa, utilizando o mesmo intervalo de pontuação do NASA-TLX. Outra forma de agregação é por meio de um indicador fisiológico.

Já o SWAT, envolvendo apenas três dimensões, é baseado em propriedades formais da teoria de medição conjunta, permite análises individuais ou em grupos e, após obter a escala pelo indivíduo, várias tarefas ou subtarefas podem ser facilmente classificadas por meio de um procedimento não intrusivo. Porém, o procedimento é mais demorado que o NASA-TLX, possui uma escala de apenas três pontos para cada dimensão e possui menos dimensões para avaliar a carga de trabalho mental.

Antes do surgimento do NASA-TLX e do SWAT, foi desenvolvida a Escala das Demandas de Trabalho, com escala multidimensional, utilizada normalmente para determinar as demandas de trabalho, porém sendo possível analisar a carga de trabalho mental. As suas escalas possuem fácil ponderação e entendimento, validação em muitos países e aplicações nas mais diversas situações de trabalho. Porém, trata-se de um questionário que pode ser considerado longo, podendo causar a fadiga e o desinteresse dos respondentes, visto que leva em torno de 15 minutos para conclusão.

No início do século XXI foi desenvolvido o questionário ISTAS 21, utilizado para avaliar os riscos ocupacionais de natureza psicossocial, os quais estão ligados a carga mental dos trabalhadores, e foca em ações preventivas com base nos resultados do grupo avaliado. Como o questionário possui as versões curta, média e longa, todos os tamanhos e ramos de empresas podem ser avaliados. Para ajudar na estimativa da CTM, normalmente aplica-se juntamente algum outro método subjetivo de avaliação, o que pode ser caracterizado como uma desvantagem, visto que o tempo de coleta, ponderação e análise dos resultados torna-se mais demorado, acarretando em custos. Normalmente é aplicada a versão curta, devido ao maior tamanho das outras duas versões, as quais, demandando um tempo mais longo de preenchimento, podem apresentar resistência dos participantes e até mesmo dificuldade de liberação da empresa.

Alguns anos depois foi desenvolvida a Avaliação Operacional Global de Habilidades Laparoscópicas, utilizada para avaliar a carga de trabalho mental em cirurgias minimamente invasivas, consistindo em uma escala global de cinco itens, uma lista de verificação com dez itens e duas escalas de analogias visuais. A GOALS é eficiente para avaliar o treinamento e as habilidades laparoscópicas, fornecendo *feedback* aos cirurgiões. Como ela pode ser utilizada por vídeo ou por classificação direta, possui grande versatilidade e flexibilidade. O baixo custo, o pouco tempo para completar as escalas, a facilidade para obter os dados ao final de cada procedimento e a possibilidade de generalização da escala global para outros procedimentos laparoscópicos também são vantagens do método. Porém, a lista de verificação é específica à remoção da vesícula biliar e os autorrelatos dependem da memória.

Da mesma forma que a GOALS, o SURG-TLX foi desenvolvido para avaliar a carga de trabalho mental de cirurgiões, mantendo a estrutura do NASA-TLX, porém adaptando para demandas específicas da cirurgia. Como o método abrange seis

dimensões para avaliar a carga de trabalho, é possível estimar qual dimensão está interferindo mais na carga de trabalho em cada procedimento ou etapa específica, possibilitando melhores decisões sobre a introdução de novas técnicas ou tecnologias, treinamentos e permitindo autoavaliação. Apesar de ser um método validado, confiável, de fácil e rápida aplicação, pode apresentar desvantagens, como o fato de ser dependente da memória, não permitindo o acompanhamento da carga de trabalho mental durante os procedimentos e não captando flutuações na CTM, e a dificuldade para categorizar certos procedimentos.

Uma nova metodologia juntou o SURG-TLX e a GOALS, a qual possui as vantagens e desvantagens de ambas as ferramentas separadas. Além disso, esse procedimento permite identificar limitações da carga de trabalho e problemas de desempenho na sala cirúrgica, orientando ações para intervenções nos procedimentos.

Ainda em relação as escalas multidimensionais, a Escala Subjetiva de Carga de Trabalho Mental avalia a carga de trabalho de forma geral, por meio de um questionário. Apesar de grande, leva em torno de 15 minutos para o preenchimento, é de fácil entendimento pelos participantes e implementação, não é invasivo, possui baixo custo e pouco tempo de avaliação. A ESCAM pode ser respondida no local de trabalho e identifica áreas a serem modificadas e melhoradas, pois os itens são conectados com situações típicas de trabalho. Porém, podem existir vieses de resposta devido à deseabilidade social e como é um método recente existem poucos resultados para comparação.

Apesar de não aparecerem na linha do tempo, a escala subjetiva e o questionário são métodos simples, normalmente unidimensional e multidimensional, respectivamente, diretos, baratos e não invasivos para avaliar a carga de trabalho mental, apresentando, de forma geral, as mesmas vantagens e desvantagens dos métodos subjetivos, porém, por possuírem diversas metodologias, acabam sendo menos empregados e vistos com menos validade e confiabilidade que os demais.

Alguns métodos são baseados em trabalhadores específicos, porém várias ferramentas permitem avaliações dos mais variados trabalhadores: medidas fisiológicas, tarefas primária e secundária, JDS, MCH, SWAT, NASA-TLX, RTLX, GSW, RSME, ISTAS 21, ESCAM.

De forma específica, o SOI, a escala de Cooper Harper, a ISA, a ATWIT e o TLI foram desenvolvidos para controladores de tráfego aéreo, a BSPE é utilizada

em trabalhadores de saúde, a GOALS, o SURG-TLX e a junção dessas duas ferramentas são utilizadas para avaliar cirurgões e a IWS é utilizada para sinalizadores ferroviários.

Mesmo com os diversos métodos e suas aplicações específicas, as vantagens e desvantagens podem ser generalizadas para as medidas fisiológicas, para as medidas de desempenho e para as medidas subjetivas.

As medidas fisiológicas são medidas de forma contínua, permitindo o acompanhamento da carga de trabalho de forma direta, multidimensional e não intrusiva. Porém, exigem procedimentos mais complexos e mais caros e mudanças fisiológicas podem não significar mudanças na carga de trabalho.

As medidas de desempenho, representadas pelas tarefas primárias e secundárias e pelo SOI, fornecem índices diretos para avaliar a carga de trabalho mental, com base no desempenho. Normalmente são obtidas de forma rápida e simples, porém, em algumas situações, podem ser intrusivas. Além disso, da mesma forma que as medidas fisiológicas, mudanças no desempenho podem não significar mudanças na carga de trabalho.

As medidas subjetivas podem ser divididas em medidas unidimensionais e multidimensionais. As medidas unidimensionais podem ser obtidas continuamente durante a tarefa ou após as tarefas e as medidas multidimensionais são obtidas após a realização das atividades.

As medidas unidimensionais obtidas durante as tarefas são a ISA, a ATWIT e a IWS. Elas apresentam a vantagem de acompanhar toda a realização do procedimento, mostrando as variações e os picos de alta e baixa CTM, são ferramentas rapidamente e facilmente respondíveis e muito úteis para situações longas, não dependendo da memória do participante. Porém, podem apresentar certa intrusividade e exigir *softwares* para gerenciamento dos intervalos de solicitações das classificações.

Já as medidas unidimensionais obtidas após as tarefas são a escala de Cooper Harper, a MCH, a BSPE, a GSW e a RSME. As escalas de Cooper Harper são baseadas em uma árvore de decisões e os demais métodos são apenas uma classificação com base nos escores e em descritores qualitativos. De forma geral, são métodos rápidos, fáceis de implementar e avaliar, de fácil entendimento e resposta pelos participantes e não intrusivos. Porém, dependem da memória dos trabalhadores e, da mesma forma que todas as medidas unidimensionais, fornecem apenas um valor

geral de carga de trabalho, não levando em consideração as diferentes demandas e as fontes das variações.

As medidas subjetivas multidimensionais são o NASA-TLX, o RTLX, o SWAT, a JDS, o ISTAS 21, a ESCAM, o SURG-TLX, a GOALS e a junção do SURG-TLX e da GOALS. Apesar de cada método possuir um foco diferente, todos utilizam diversas dimensões para analisar a carga de trabalho mental, sendo essa sua principal vantagem. De forma geral, possuem avaliações simples e de fácil entendimento pelos participantes. Porém, são métodos totalmente dependentes da memória, visto que são realizados após as tarefas, mais longos que os demais, mesmo que muitas vezes não sejam demorados, e exigem treinamentos para o preenchimento correto. Ainda, por serem dependes da memória e do julgamento individual, podem apresentar vieses de respostas, vinculados também a desejabilidade social.

Já o método TLI é diferente dos demais, pois envolve uma medida de desempenho e uma medida subjetiva, sendo essa sua principal característica, pois é um procedimento simples, que possui as vantagens das medidas objetivas e das medidas subjetivas e que permite generalizações.

O Quadro 18 apresenta as características gerais e as vantagens e desvantagens de cada ferramenta, agrupadas nas categorias descritas acima.

Como todas as categorias possuem vantagens e desvantagens, uma forma de compensar essas desvantagens é a utilização de mais de um método para avaliar a carga de trabalho mental, principalmente pertencendo a classificações diferentes, sendo o uso de ferramentas fisiológicas, de desempenho e subjetivas, de forma concomitante, o ideal para uma avaliação mais confiável, válida e segura.

Quadro 18 - Quadro resumo dos métodos de avaliação da CTM

(continua)

	Características	Vantagens	Desvantagens
MEDIDAS FISIOLÓGICAS			
Indicadores: coração, olhos, cérebro, respiração, pele, temperatura facial, fala, movimentos corporais e urina	Mudanças nos indicadores fisiológicos relacionadas a mudanças na CTM.	Medidas contínuas e em tempo real. Normalmente não intrusivas. Resposta direta.	Equipamentos mais caros e complexos. Análises mais difíceis, exigindo maiores cuidados. Alguns equipamentos podem ser intrusivos.
MEDIDAS DE DESEMPENHO			
Tarefa primária	Mudanças no desempenho da tarefa principal relacionadas a mudanças na CTM.	Direta e não intrusiva. Pode ser em tempo real. Fácil e rápida análise.	Dificuldade em generalizar as medidas. Desempenho não está necessariamente relacionado a CTM. Não são suficientes por conta própria.
Tarefa secundária	Mudanças no desempenho de uma tarefa secundária, mais simples e desenvolvida concomitantemente com a tarefa principal, relacionadas a mudanças na CTM.	Medida objetiva, simples, direta e fácil de ser aplicada.	Pode ser intrusiva. Dificuldade para limitar as respostas normais. Podem não estar relacionadas a mudanças na CTM.
SOI	Mudanças no desempenho da tarefa principal relacionadas a mudanças na CTM.	Fator objetivo. Resposta direta. Não intrusivo.	Sozinho ele não é capaz de analisar a CTM. Superficial. Apenas para ATCs.
MEDIDAS SUBJETIVAS			
Unidimensional em tempo real			
ATWIT	Escala de 1 a 10, escolhida por meio de teclado físico ou <i>touch screen</i> .	Medida direta e simples. Fácil aplicação e análise dos dados. Permite análise das variações da CTM.	Apenas para ATCs. Pode ser invasiva. Precisa ter gerenciamento de intervalos de solicitações.
ISA	Escala de 1 a 5, escolhida por meio de teclado físico ou <i>touch screen</i> .	Medida direta, simples, discreta e de baixo custo. Fácil aplicação e análise dos dados. Permite análise das variações da CTM.	Desenvolvida para ATCs. Pode ser invasiva. Precisa ter gerenciamento de intervalos de solicitações.

Quadro 18 - Quadro resumo dos métodos de avaliação da CTM

(continua)

	Características	Vantagens	Desvantagens
MEDIDAS SUBJETIVAS			
Unidimensional em tempo real			
IWS	Escala de 1 a 9, escolhida por meio de teclado físico ou <i>touch screen</i> ou verbalmente.	Medida direta e simples. Fácil aplicação e análise dos dados. Permite análise das variações da CTM, permitindo relacionar com o que aconteceu na prática.	Apenas para o setor ferroviário. Pode ser intrusiva. Precisa ter gerenciamento de intervalos de solicitações.
Unidimensional após as tarefas			
BSPE	Escala original com pontuação entre 6 e 20 pontos.	Fácil, rápida, simples e global. Não invasiva. Correlacionada com a frequência cardíaca.	Utilizada preferivelmente para medir o esforço físico percebido. Apenas para a área da saúde.
Escala de CH	Árvore de decisões. Escala de 1 a 10.	Pioneira entre as escalas subjetivas. Simples, rápida, fácil e barata. Qualitativa e quantitativa.	Apenas para trabalhos manuais de ATCs.
Escala subjetiva	Escalas desenvolvidas para contextos específicos.	Simples, rápida e fácil de aplicar a avaliar. Baixo custo.	Muitas vezes não possui validação.
GSW	Critério de agregação para as dimensões do NASA-TLX.	Utilizada juntamente com o NASA-TLX.	Menos eficiente que um critério fisiológico de agregação.
MCH	Árvore de decisões. Escala de 1 a 10.	Simples, rápida, fácil e barata. Qualitativa e quantitativa. Possui maior generalização que a Escala de CH.	Voltada para trabalhos mais automatizados.
RSME	Escala de 0 a 150, com 9 descritores qualitativos.	Pode ser tão sensível quanto as escalas multidimensionais. Rápida, simples e fácil.	Utilizada preferivelmente para medir o esforço mental. Dificuldade de adaptação linguística e cultural.
TLI	Índice objetivo do número de aeronaves e fator subjetivo para a situação do tráfego aéreo.	Combina medidas objetivas e subjetivas. Regras simples e generalizáveis. Procedimento simples.	Apenas para ATCs. Não permite o cálculo em tempo real e não possui um algoritmo direto.

Quadro 18 - Quadro resumo dos métodos de avaliação da CTM

(continua)

	Características	Vantagens	Desvantagens
MEDIDAS SUBJETIVAS			
Multidimensional			
ESCAM	Questionário com 5 fatores e 20 itens.	Análise holística do trabalho. Realizada no local de trabalho, com fácil avaliação. Itens conectados com situações típicas de trabalho.	Pode ser considerada longa. Poucos resultados para comparação.
GOALS	Escala global de 5 itens, lista de verificação com 10 itens e 2 escalas de analogias visuais.	Avalia o desempenho do cirurgião por meio de <i>feedback</i> . Avaliação pessoalmente ou por vídeo. Baixo custo, rápida e possível de generalização.	Apenas para cirurgias minimamente invasivas. Lista de verificação é específica para uma única situação.
ISTAS 21	Questionário com 3 versões: longa, média e curta.	Avalia riscos ocupacionais de natureza psicossocial. Útil para orientar ações preventivas e elaborar planos de ação.	Não avalia diretamente a CTM. Longo tempo de preenchimento.
JDS	Duas variáveis independentes: demandas de trabalho e latitude de decisão.	Escalas de fácil ponderação e entendimento.	Utilizada principalmente para determinar demandas de trabalho. Pode ser demorada.
NASA-TLX	6 dimensões: demanda mental, demanda física, demanda temporal, desempenho, esforço e nível de frustração. Escala de 0 a 100 pontos.	Método mais utilizado. Simples e rápido de aplicar e analisar.	Sistema de ponderação pode ser ineficiente.
Questionário	Questionários desenvolvidos para contextos específicos.	Facilidade e baixo custo de aplicação.	Grande tamanho do questionário e falta de validação. Dificuldade de elaboração.
RTLX	Igual ao NASA-TLX, porém sem o sistema de ponderação das escalas.	Mais rápido e simples que o NASA-TLX.	Considera todas as dimensões com o mesmo peso.
SURG-TLX	6 dimensões: demanda mental, demanda física, demanda temporal, complexidade das tarefas, estresse situacional e distrações. Escala de 0 a 100 pontos.	Baseado no NASA-TLX. Fácil e rápida aplicação.	Apenas para cirurgiões. Dificuldade para categorizar certos procedimentos.

Quadro 18 - Quadro resumo dos métodos de avaliação da CTM

(conclusão)

	Características	Vantagens	Desvantagens
MEDIDAS SUBJETIVAS			
Multidimensional			
SURG-TLX + GOALS	Compilação das ferramentas SURG-TLX e GOALS.	Vantagens do SURG-TLX, com adição de mais uma informação. Orienta ações para intervenções nos procedimentos.	Utiliza dois métodos, porém de forma incompleta.
SWAT	3 dimensões: tempo, esforço mental e estresse. Escala de 3 pontos.	Método bastante utilizado. Pode ser empregado para análise individual ou em grupo.	Procedimento pode ser demorado e ter baixa sensibilidade, por utilizar uma escala de 3 pontos.

Fonte: Elaborado pelo autor (2018).

5 CONSIDERAÇÕES FINAIS

A revisão sistemática de literatura mostrou-se uma metodologia adequada na busca das ferramentas mais utilizadas, a partir do ano de 2000, na avaliação da carga de trabalho mental, identificando os métodos mais empregados em cada grupo estratificado de trabalhadores e nos países, analisando um portfólio bibliográfico final constituído por 85 artigos, publicados entre 2000 e 2017 e alinhados com a pesquisa.

Na análise bibliométrica foram destacados os principais autores presentes nos artigos do portfólio bibliográfico, os principais periódicos, com destaques para os periódicos "*Ergonomics*", "*Applied Ergonomics*", "*Surgical Endoscopy and Other Interventional Techniques*", "*Human Factors*" e "*Aviation, Space and Environmental Medicine*", e as principais palavras-chave utilizadas, destacando-se as palavras-chave "*mental workload*", "*workload*" e "*NASA-TLX*". Além disso, foi verificado o fator de impacto dos periódicos, os quais, de forma geral, apresentaram um bom resultado.

Na análise do conteúdo foram encontradas 22 ferramentas, com destaque para o método de análises de medidas fisiológicas, principalmente analisando o coração, os olhos e o cérebro, e a ferramenta NASA-TLX. O maior número de aplicações das pesquisas do PB foi nos Estados Unidos, seguido pela Finlândia, Japão, Irã e Reino Unido e os principais trabalhadores analisados foram os trabalhadores de saúde, os pilotos e os controladores de tráfego aéreo.

Os 22 métodos encontrados no portfólio bibliográfico foram descritos, focando na sua origem, metodologia, avaliação dos resultados e aplicações usuais, e também analisados criticamente, por meio das suas vantagens e desvantagens. Assim, as ferramentas podem ser classificadas em medidas fisiológicas, medidas de desempenho e medidas subjetivas, as quais, por sua vez, são divididas em multidimensionais e unidimensionais, em tempo real ou após a realização das tarefas.

As medidas fisiológicas são medidas continuamente, da mesma forma que as medidas subjetivas unidimensionais em tempo real, podendo acompanhar o desenvolvimento da carga de trabalho de forma direta e não dependendo da memória dos participantes. As medidas de desempenho fornecem índices diretos para avaliar a CTM e são obtidos de forma rápida e simples. As medidas subjetivas multidimensionais são capazes de diagnosticar as diversas dimensões da carga de trabalho, sendo dependentes da memória, não intrusivas e realizadas após as tarefas

e as medidas unidimensionais, realizadas após os procedimentos, fornecem uma estimativa geral da CTM, sendo de fácil resposta, implementação e avaliação.

Como desvantagens, as medidas fisiológicas possuem procedimentos mais caros e complexos, as medidas de desempenho podem não estar relacionadas com a carga de trabalho mental e serem intrusivas, as medidas subjetivas multidimensionais são mais longas e dependentes da memória, as medidas subjetivas unidimensionais realizadas após as tarefas também dependem da memória e não são capazes de fornecer as diferentes demandas da CTM e as medidas unidimensionais em tempo real podem ser intrusivas e dependem do gerenciamento de intervalos para solicitações das classificações.

De forma geral, as ferramentas não são limitadas por países, visto que a maioria delas já foi traduzida e adaptada para diversos idiomas. Em relação aos trabalhadores, algumas ferramentas foram desenvolvidas especialmente para controladores de tráfego aéreo, cirurgiões, trabalhadores de saúde e sinalizadores ferroviários, porém a grande maioria pode ser utilizada nas mais variadas classes de trabalhadores.

Assim, o critério de escolha deve ser a situação de trabalho a ser verificada, os objetos da análise e os recursos disponíveis, sendo indicado, sempre que possível, o uso de pelo menos dois métodos diferentes e, idealmente, o uso de ferramentas fisiológicas, de desempenho e subjetivas, de forma concomitante, visando utilizar as vantagens e suprir as desvantagens de cada uma delas.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDELRAHMAN, Amro M. et al. Impact of single-incision laparoscopic cholecystectomy (SILC) versus conventional laparoscopic cholecystectomy (CLC) procedures on surgeon stress and workload: a randomized controlled trial. **Surgical endoscopy**, v. 30, n. 3, p. 1205-1211, 2016.

ANDRADE, J. M. V. **Diabetes Mellitus e doença periodontal: uma revisão sistemática**. 2009. 85 f. Dissertação (Mestrado em Ciências Médicas) – Universidade do Estado do Rio de Janeiro, Faculdade de Ciências Médicas, Rio de Janeiro, 2009.

AHLSTROM, Ulf; FRIEDMAN-BERG, Ferne J. Using eye movement activity as a correlate of cognitive workload. **International Journal of Industrial Ergonomics**, v. 36, n. 7, p. 623-636, 2006.

ARELLANO, Juan Luis Hernández et al. Construction of a survey to assess workload and fatigue among AMT operators in Mexico. **Work**, v. 41, n. Supplement 1, p. 1790-1796, 2012.

ATHÈNES, Sylvie et al. ATC complexity and controller workload: Trying to bridge the gap. In: **Proceedings of the International Conference on HCI in Aeronautics**. AAAI Press Cambridge, MA, 2002. p. 56-60.

AVERTY, Philippe et al. Evaluating a new index of mental workload in real ATC situation using psychophysiological measures. In: **Digital Avionics Systems Conference, 2002. Proceedings. The 21st**. IEEE, 2002. p. 7A4-7A4.

AVERTY, Philippe et al. Mental workload in air traffic control: an index constructed from field tests. **Aviation, space, and environmental medicine**, v. 75, n. 4, p. 333-341, 2004.

AYAZ, Hasan et al. Estimation of cognitive workload during simulated air traffic control using optical brain imaging sensors. In: **International Conference on Foundations of Augmented Cognition**. Springer, Berlin, Heidelberg, 2011. p. 549-558.

AYAZ, Hasan et al. Optical brain monitoring for operator training and mental workload assessment. **Neuroimage**, v. 59, n. 1, p. 36-47, 2012.

BAHIA, Secretaria da Saúde do Estado. Superintendência de Vigilância e Proteção da Saúde. Diretoria de Vigilância e Atenção à Saúde do Trabalhador. Centro Estadual de Referência em Saúde do Trabalhador. Protocolo de atenção à saúde mental e trabalho/organizado por Suerda Fortaleza de Souza/SESAB/SUVISA/DIVAST/CESAT - Salvador: **DIVAST**, 2014.

BALDUCCI, Cristian; AVANZI, Lorenzo; FRACCAROLI, Franco. The individual “costs” of workaholism: An analysis based on multisource and prospective data. **Journal of Management**, p. 0149206316658348, 2016.

BALFE, Nora; SHARPLES, Sarah; WILSON, John R. Impact of automation: measurement of performance, workload and behaviour in a complex control environment. **Applied ergonomics**, v. 47, p. 52-64, 2015.

BAPTISTA, Marta Pereira. **Controlo ótimo da atitude de um helicóptero com falha de atuadores**. 2015. Dissertação (Mestre em Engenharia Aeronáutica) - Instituto Tecnológico de Aeronáutica, São José dos Campos, São Paulo.

BERG, Regan J. et al. The impact of heat stress on operative performance and cognitive function during simulated laparoscopic operative tasks. **Surgery**, v. 157, n. 1, p. 87-95, 2015.

BERGUER, R.; SMITH, W. D.; CHUNG, Y. H. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. **Surgical endoscopy**, v. 15, n. 10, p. 1204-1207, 2001.

BEVILACQUA, Maurizio et al. Visual Management implementation and evaluation through mental workload analysis. **IFAC Proceedings Volumes**, v. 46, n. 7, p. 294-299, 2013.

BORG, Gunnar. Perceived exertion as an indicator of somatic stress. **Scand j rehabil med**, v. 2, p. 92-98, 1970.

BORG, Gunnar A. Psychophysical bases of perceived exertion. **Med sci sports exerc**, v. 14, n. 5, p. 377-381, 1982.

BORG, Gunnar. Psychophysical scaling with applications in physical work and the perception of exertion. **Scandinavian journal of work, environment & health**, p. 55-58, 1990.

BORG, Gunnar AV; NOBLE, Bruce J. Perceived exertion. **Exercise and sport sciences reviews**, v. 2, n. 1, p. 131-154, 1974.

BOULTINGHOUSE, Oscar W. et al. Assessing physician job satisfaction and mental workload. **Telemedicine and e-Health**, v. 13, n. 6, p. 715-718, 2007.

BRENNAN, S. D. An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. **Portsm. DRA Marit. Command Control Div. DRA Tech. Memo. CAD5**, v. 92017, 1992.

BROEKHOVEN, R. F. G. **Comparison of real-time relative workload measurements in rail signallers**. 2016. Dissertação de Mestrado. University of Twente.

BUTMEE, Totsapon; LANSDOWN, Terry C. Moving Between Automated and Manual Driving: Mental Workload and Performance Implications. 2017.

BYRNE, Aidan. Measurement of mental workload in clinical medicine: a review study. **Anesthesiology and pain medicine**, v. 1, n. 2, p. 90, 2011.

BYRNE, A. J. et al. The relationship between experience and mental workload in anaesthetic practice: an observational study. **Anaesthesia**, v. 68, n. 12, p. 1266-1272, 2013.

BYRNE, A. J. et al. Novel method of measuring the mental workload of anaesthetists during clinical practice. **British journal of anaesthesia**, v. 105, n. 6, p. 767-771, 2010.

BYERS, James C. Traditional and raw task load index (TLX) correlations: are paired comparisons necessary?. *Advances in Industrial Ergonomics and Safety I*: Taylor and Francis., 1989.

CAIN, Brad. **A review of the mental workload literature**. Defence Research And Development Toronto (Canada), 2007.

CARDOSO, M. S. **Avaliação da carga mental de trabalho e do desempenho de métodos de mensuração: NASA TLX e SWAT**. 2010. 128 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2010.

CASALI, John G.; WIERWILLE, Walter W. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. **Human Factors**, v. 25, n. 6, p. 623-641, 1983.

CASALI, John G.; WIERWILLE, Walter W. On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. **Ergonomics**, v. 27, n. 10, p. 1033-1050, 1984.

CEBALLOS VÁSQUEZ, Paula et al. Validación de Escala Subjetiva de Carga Mental de Trabajo en funcionarios/as universitarios. *Ciencia y enfermería*, v. 20, n. 2, p. 73-82, 2014.

CEBALLOS-VÁSQUEZ, Paula et al. Fatores psicossociais e carga mental de trabalho: uma realidade percebida pelos enfermeiros em Unidades de Terapia Intensiva 1. **Revista Latino-Americana de Enfermagem**, v. 23, n. 2, p. 315-322, 2015.

CHANG, Lily et al. Reliable assessment of laparoscopic performance in the operating room using videotape analysis. **Surgical innovation**, v. 14, n. 2, p. 122-126, 2007.

COLLET, Christian et al. Assessing workload through physiological measurements in bus drivers using an automated system during docking. **Human factors**, v. 45, n. 4, p. 539-548, 2003.

COOPER, George E.; HARPER JR, Robert P. **The use of pilot rating in the evaluation of aircraft handling qualities**. Advisory Group for aerospace research and development Neuilly-Sur-Seine (France), 1969.

CORRÊA, Fábio de Paula. **Carga Mental e Ergonomia**. 2003 148f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis, SC.

COSTA, Giovanni. Evaluation of workload in air traffic controllers. **Ergonomics**, v. 36, n. 9, p. 1111-1120, 1993.

DADI, Gabriel B. et al. Cognitive workload demands using 2D and 3D spatial engineering information formats. **Journal of Construction Engineering and Management**, v. 140, n. 5, p. 04014001, 2014.

DANIELLOU, François; BÉGUIN, Pascal. Metodologia da ação ergonômica: abordagens do trabalho real. **Ergonomia**, p. 281-301, 2007.

DARVISHI, Ebrahim et al. Subjective Mental Workload and Its Correlation With Musculoskeletal Disorders in Bank Staff. **Journal of manipulative and physiological therapeutics**, v. 39, n. 6, p. 420-426, 2016.

DE RIVECOURT, M. et al. Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. **Ergonomics**, v. 51, n. 9, p. 1295-1319, 2008.

DE SOUZA CARDOSO, Mariane; GONTIJO, Leila Amaral. Avaliação da carga mental de trabalho e do desempenho de medidas de mensuração: NASA TLX e SWAT. Evaluation of mental workload and performance measurement: NASA TLX and SWAT. **Gestão & Produção**, v. 19, n. 4, p. 873-884, 2012.

DEY, A. K.; MANN, D. D. A complete task analysis to measure the workload associated with operating an agricultural sprayer equipped with a navigation device. **Applied ergonomics**, v. 41, n. 1, p. 146-149, 2010.

DE WAARD, Dick; LEWIS-EVANS, Ben. Self-report scales alone cannot capture mental workload. **Cognition, technology & work**, v. 16, n. 3, p. 303-305, 2014.

DI NOCERA, Francesco; CAMILLI, Marco; TERENCEZI, Michela. A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. **Journal of Cognitive Engineering and Decision Making**, v. 1, n. 3, p. 271-285, 2007.

DIAS, R. D. et al. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. **British Journal of Surgery**, v. 105, n. 5, p. 491-501, 2018.

DORNELES, S. F. **Indicadores para avaliação da atenção á saúde materno infantil: uma revisão de literatura**. 2014. 62 f. Dissertação (Mestrado em Saúde Coletiva) – Instituto Sírio-Libanês em Ensino e Pesquisa, São Paulo, 2014.

DUL, Jan; WEERDMEEESTER, Bernard. Ergonomia prática. rev. ampl. **São Paulo: Edgard Blücher**, 2004.

DUMMEL, J. **Conhecimento e conduta dos profissionais de saúde diante da relação entre as doenças periodontais e as doenças cardiovasculares: uma revisão de literatura**. 2016. 77 f. Dissertação (Mestrado em Saúde Pública) -

Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2016.

EAST, Julia A.; BAUER JR, Kenneth W.; LANNING, Jeffrey W. Feature selection for predicting pilot mental workload: A feasibility study. **International Journal of Smart Engineering System Design**, v. 4, n. 3, p. 183-193, 2002.

EDWARDS, T. et al. Multifactor interactions and the air traffic controller: the interaction of situation awareness and workload in association with automation. **Cognition, Technology & Work**, v. 19, n. 4, p. 687-698, 2017.

ELFERING, Achim; GREBNER, Simone; HALLER, Martina. Railway-controller-perceived mental work load, cognitive failure and risky commuting. **Ergonomics**, v. 55, n. 12, p. 1463-1475, 2012.

ERASLAN, Ergun; CAN, Guln Feryal; ATALAY, Kumru Didem. Mental workload assessment using a fuzzy multi-criteria method/Procjena mentalnog opterecenja primjenom neizravne metode s vise kriterija. **Tehnicki Vjesnik-Technical Gazette**, v. 23, n. 3, p. 667-675, 2016.

ESPÍNDOLA, Edilene Zilma. **Avaliação da Carga Mental de Trabalho em Magistrados do Poder Judiciário de Santa Catarina**. 2013. Dissertação (Mestrado em Engenharia de Produção e Sistemas. Área de Concentração: Ergonomia) Universidade Federal de Santa Catarina (UFSC), Florianópolis-SC, 2013.

FACCHINI, L. A. Uma contribuição da epidemiologia: o modelo de determinação social aplicado à saúde do trabalhador. In BUSCHINELLI, J. T.; ROCHA, L. E.; RIGOTTO, R. M. **Isto é trabalho de gente?: vida, doença e trabalhador no Brasil**. Rio de Janeiro: Vozes, 1994.

FALLAHI, Majid et al. Analysis of the mental workload of city traffic control operators while monitoring traffic density: A field study. **International Journal of Industrial Ergonomics**, v. 54, p. 170-177, 2016.

FINCANNON, Thomas; AHLSTROM, Vicki. Scale Size of the Air Traffic Workload Input Technique (ATWIT) A Review of Research. In: **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**. Sage CA: Los Angeles, CA: SAGE Publications, 2016. p. 2078-2082.

FLORES, Daniela et al. Ocuparse del bienestar de los profesionales de la salud: un desafío pendiente. **Revista Chilena de Terapia Ocupacional**, v. 14, n. 1, p. ág. 33-44, 2014.

FORSYTH, Katherine L. et al. Interruptions Experienced by Emergency Nurses: Implications for Subjective and Objective Measures of Workload. **Journal of Emergency Nursing**, 2018.

FRUTUOSO, Joselma Tavares; CRUZ, Roberto Moraes. Mensuração da carga de trabalho e sua relação com a saúde do trabalhador. **Revista Brasileira de Medicina do Trabalho**, v. 3, n. 1, p. 29-36, 2005.

GALSTER, Scott M. et al. Air traffic controller performance and workload under mature free flight: Conflict detection and resolution of aircraft self-separation. **The International Journal of Aviation Psychology**, v. 11, n. 1, p. 71-93, 2001.

GALVÃO, Taís Freire; PEREIRA, Mauricio Gomes. Revisões sistemáticas da literatura: passos para sua elaboração. **Epidemiologia e Serviços de Saúde**, v. 23, n. 1, p. 183-184, 2014.

GALY, Edith; CARIOU, Magali; MÉLAN, Claudine. What is the relationship between mental workload factors and cognitive load types?. **International Journal of Psychophysiology**, v. 83, n. 3, p. 269-275, 2012.

GHANBARY SARTANG, A. et al. Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses. **Journal of Occupational Health and Epidemiology**, v. 5, n. 4, p. 211-217, 2016.

GIFFHORN, E. **Construção de um Modelo de Avaliação do Desempenho de Empresas terceirizadas com a utilização da metodologia MCDA-C: Um Estudo de Caso**. 2007. 258 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2007.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. Editora Atlas SA, 2008.

GOMES, T. R. **O erro médico sob o olhar do judiciário: uma análise das decisões do Tribunal de Justiça do Distrito Federal e Territórios**. 2017. 99 f. Dissertação (Mestrado em Saúde Coletiva) - Universidade de Brasília, Brasília, 2017.

GONZÁLEZ-MUÑOZ, Elvia Luz; GUTIÉRREZ-MARTÍNEZ, Rodolfo E. Contribution of mental workload to job stress in industrial workers. **Work (Reading, Mass.)**, v. 28, n. 4, p. 355-361, 2007.

GOULD, Kristian S. et al. Effects of navigation method on workload and performance in simulated high-speed ship navigation. **Applied ergonomics**, v. 40, n. 1, p. 103-114, 2009.

GRECO, R. M.; QUEIROS, V. M & GOMES, J. R. Cargas de trabalho dos técnicos operacionais da escola de enfermagem da USP. **Revista Brasileira de Saúde Ocupacional**. Vol. 25, p. 59-75, 1999.

GUERRERO-BOTE, Vicente P.; MOYA-ANEGÓN, Félix. A further step forward in measuring journals' scientific prestige: The SJR2 indicator. **Journal of Informetrics**, v. 6, n. 4, p. 674-688, 2012.

GUMBS, Andrew A.; HOGLE, Nancy J.; FOWLER, Dennis L. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. **Journal of the American College of Surgeons**, v. 204, n. 2, p. 308-313, 2007.

HAH, Sehchang; WILLEMS, Ben; SCHULZ, Kenneth. The evaluation of Data Communication for the future air traffic control system (NextGen). In: **Proceedings of**

the Human Factors and Ergonomics Society Annual Meeting. Sage CA: Los Angeles, CA: SAGE Publications, 2010. p. 99-103.

HAH, Sehchang et al. Human Factors Evaluation of Conflict Resolution Advisories for Air Traffic Control in the En Route Domain. In: **Proceedings of the Human Factors and Ergonomics Society Annual Meeting.** Sage CA: Los Angeles, CA: SAGE Publications, 2016. p. 94-98.

HALE, Mark R.; STANLEY, Raymond M. Evaluating the design and suitability of the Wildlife Surveillance Concept. In: **Integrated Communications, Navigation and Surveillance Conference (ICNS), 2017.** IEEE, 2017. p. 5A1-1-5A1-7.

HALLBECK, M. S. et al. The impact of intraoperative microbreaks with exercises on surgeons: a multi-center cohort study. **Applied ergonomics**, v. 60, p. 334-341, 2017.

HANNULA, Manne et al. Comparison between artificial neural network and multilinear regression models in an evaluation of cognitive workload in a flight simulator. **Computers in biology and medicine**, v. 38, n. 11, p. 1163-1170, 2008.

HANSSON, Gert-Åke et al. Physical workload in various types of work: Part I. Wrist and forearm. **International Journal of Industrial Ergonomics**, v. 39, n. 1, p. 221-233, 2009.

HARRISON, Joshua et al. Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. **IEEE Transactions on Human-Machine Systems**, v. 44, n. 4, p. 429-440, 2014.

HART, S. G. **NASA Task load Index (TLX). Volume 1.0; Paper and pencil package.** NASA Ames Research Center; Moffett Field, CA, United States. 1986.

HART, S. G.; STAVELAND, L. E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: **Advances in psychology.** North-Holland, 1988. p. 139-183.

HART, S. G. Gerenciamento da carga de trabalho da tripulação: Um Fator Crítico no Desempenho do Sistema. **Revista SIPAER**, v. 13, n. 66, p. 18-23, maio 1999.

HART, S. G. NASA-task load index (NASA-TLX); 20 years later. In: **Proceedings of the human factors and ergonomics society annual meeting.** Sage CA: Los Angeles, CA: Sage Publications, 2006. p. 904-908.

HART, Sandra G.; WICKENS, Christopher D. Workload assessment and prediction. In: **Manprint.** Springer, Dordrecht, 1990. p. 257-296.

HENDY, Keith C.; HAMILTON, Kevin M.; LANDRY, Lois N. Measuring subjective workload: when is one scale better than many?. **Human Factors**, v. 35, n. 4, p. 579-601, 1993.

HENELIUS, Andreas et al. Mental workload classification using heart rate metrics. In: **Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE**. IEEE, 2009. p. 1836-1839.

HICKS, Thomas G.; WIERWILLE, Walter W. Comparison of five mental workload assessment procedures in a moving-base driving simulator. **Human Factors**, v. 21, n. 2, p. 129-143, 1979.

HU, Jesse SL et al. Training improves laparoscopic tasks performance and decreases operator workload. **Surgical endoscopy**, v. 30, n. 5, p. 1742-1746, 2016.

HUTTUNEN, Kerttu et al. Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. **Applied ergonomics**, v. 42, n. 2, p. 348-357, 2011.

IIDA, I. ; GUIMARÃES, L. B. d. M. Ergonomia: Projeto e Produção. 3. ed. **São Paulo: Blucher**, 2016. v. 1.

JANSEN, Reinier J. et al. Hysteresis in mental workload and task performance: the influence of demand transitions and task prioritization. **Human factors**, v. 58, n. 8, p. 1143-1157, 2016.

JORDAN, C. S. Experimental study of the effects of an instantaneous self assessment workload recorder on task performance. **Report No. DRA/TM (CAD5)/92011. Farnborough: Defence Evaluation & Research Agency**, 1992.

JORGENSEN, A. H.; GARDE, A. H.; LAURSEN, B. & JENSEN, B. R. **Applying the concept of mental workload to IT-work**. Finland, 1999 – Cyberg, 1999.

KARASEK, Robert A. Job demands, job decision latitude, and mental strain: Implications for job redesign. **Administrative science quarterly**, p. 285-308, 1979.

KARASEK, Robert et al. The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. **Journal of occupational health psychology**, v. 3, n. 4, p. 322, 1998.

KARHULA, Kati et al. Association of job strain with working hours, shift-dependent perceived workload, sleepiness and recovery. **Ergonomics**, v. 56, n. 11, p. 1640-1651, 2013.

KATAOKA, Jun; SASAKI, Minako; KANDA, Katsuya. Effects of mental workload on nurses' visual behaviors during infusion pump operation. **Japan Journal of Nursing Science**, v. 8, n. 1, p. 47-56, 2011.

KAWADA, Tomoyuki; OOYA, Machiko. Workload and health complaints in overtime workers: a survey. **Archives of medical research**, v. 36, n. 5, p. 594-597, 2005.

KERICK, S. E.; ALLENDER, L. E. Effects of cognitive workload on decision accuracy, shooting performance, and cortical activity of soldiers. **SELECTED TOPICS IN ELECTRONICS AND SYSTEMS**, v. 42, p. 359, 2006.

KIRCHHOF, A. L. C. et al. Compreendendo cargas de trabalho na pesquisa em saúde ocupacional na enfermagem. **Colombia Médica**, v. 42, n. 2 Supl. 1, p. 113-119, 2011.

KIRWAN, Barry et al. Human factors in the ATM system design life cycle. In: **FAA/Eurocontrol ATM R&D Seminar, Paris, France**. 1997. p. 16-20.

KNOWLES, William B. Operator loading tasks. **Human factors**, v. 5, n. 2, p. 155-161, 1963.

KOCA, Dilek et al. Physical and mental workload in single-incision laparoscopic surgery and conventional laparoscopy. **Surgical innovation**, v. 22, n. 3, p. 294-302, 2015.

KRAMP, Kelvin H. et al. Validity and reliability of global operative assessment of laparoscopic skills (GOALS) in novice trainees performing a laparoscopic cholecystectomy. **Journal of surgical education**, v. 72, n. 2, p. 351-358, 2015.

KUIJER, P. Paul FM et al. Effect of job rotation on work demands, workload, and recovery of refuse truck drivers and collectors. **Human factors**, v. 46, n. 3, p. 437-448, 2004.

KUPFER, Michael et al. Controller support tools for schedule-based terminal-area operations. In: **Proceedings of the Ninth USA/Europe Air Traffic Management Research and Development Seminar**. 2011.

LACERDA, Rogerio Tadeu de Oliveira; ENSSLIN, Leonardo; ENSSLIN, Sandra Rolim. Research opportunities in strategic management field: a performance measurement approach. **International Journal of Business Performance Management**, v. 15, n. 2, p. 158-174, 2014.

LAHELMA, Eero et al. Occupational class inequalities across key domains of health: results from the Helsinki Health Study. **The European Journal of Public Health**, v. 15, n. 5, p. 504-510, 2005.

LAHTINEN, Taija MM et al. Heart rate and performance during combat missions in a flight simulator. **Aviation, space, and environmental medicine**, v. 78, n. 4, p. 387-391, 2007.

LAURELL, A. , NORIEGA, M. **Processo de Produção e Saúde**. São Paulo: Hucitec, 1989.

LAVILLE, A. Ergonomia. Tradução: Márcia Maria Neves Teixeira. **São Paulo: Editora Pedagógica Universitária Ltda**, 1997.

LEAN, Ying; SHAN, Fu. Brief review on physiological and biochemical evaluations of human mental workload. **Human Factors and Ergonomics in Manufacturing & Service Industries**, v. 22, n. 3, p. 177-187, 2012.

LEEDAL, J. M.; SMITH, A. F. Methodological approaches to anaesthetists' workload in the operating theatre. **British journal of anaesthesia**, v. 94, n. 6, p. 702-709, 2005.

LEE, Gysung I. et al. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. **Surgical endoscopy**, v. 28, n. 2, p. 456-465, 2014.

LEHRER, Paul et al. Cardiac data increase association between self-report and both expert ratings of task load and task performance in flight simulator tasks: An exploratory study. **International Journal of Psychophysiology**, v. 76, n. 2, p. 80-87, 2010.

LEMERCIER, Céline et al. Inattention behind the wheel: How factual internal thoughts impact attentional control while driving. **Safety science**, v. 62, p. 279-285, 2014.

LEPLAT, J., CUNY, X. **Introdução à psicologia do trabalho**. Trad. Helena Domingos. Lisboa: Fundação Calouste Gulbenkian, 1983.

LIANG, Guo-Feng et al. Evaluation and prediction of on-line maintenance workload in nuclear power plants. **Human Factors and Ergonomics in Manufacturing & Service Industries**, v. 19, n. 1, p. 64-77, 2009.

LIGDA, Sarah V. et al. Impact of conflict avoidance responsibility allocation on pilot workload in a distributed air traffic management system. In: **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**. Sage CA: Los Angeles, CA: Sage Publications, 2010. p. 55-59.

LIMA, JAA. Bases teóricas para uma Metodologia de Análise Ergonômica. In: **4º Congresso Internacional de Ergonomia e Usabilidade de Interfaces, Rio de Janeiro/RJ**. 2004.

LIN, Li-Wei; LU, Ming-Shiun. Empirical Research on the Relationship between Helicopter Pilots' Mental Workloads and Situation Awareness Levels. **Journal of the American Helicopter Society**, v. 61, n. 3, p. 1-8, 2016.

LING, C. et al. EEG signal analysis for human workload classification. In: **SoutheastCon 2001. Proceedings. IEEE**. IEEE, 2001. p. 123-130.

LIU, Yili; WICKENS, Christopher D. Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. **Ergonomics**, v. 37, n. 11, p. 1843-1854, 1994.

LONGO, Luca. Mental workload in medicine: foundations, applications, open problems, challenges and future perspectives. In: **Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium on**. IEEE, 2016. p. 106-111.

LOPEZ, Karen Dunn et al. Cognitive work analysis to evaluate the problem of patient falls in an inpatient setting. **Journal of the American Medical Informatics Association**, v. 17, n. 3, p. 313-321, 2010.

LUSTOSA, A. M. **Cuidados Paliativos e Manejo da Dor: revisão integrativa da literatura**. 2015. 91 f. Dissertação (Mestrado em Ciências da Saúde) – Universidade Cruzeiro do Sul, São Paulo, 2015.

MANCINI, M. C.; SAMPAIO, R. F. (2007). Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Rev bras fisioter**, v. 11, n. 1, p. 83-9, 2007.

MANNARU, Pujitha et al. Cognitive context detection in UAS operators using eye-gaze patterns on computer screens. In: **Next-Generation Analyst IV**. International Society for Optics and Photonics, 2016. p. 98510F.

MANNARU, Pujitha et al. Cognitive context detection using pupillary measurements. **IEEE Transactions on Human and Machine Systems**, 2016.

MANNING, Carol A. et al. **Investigating the validity of performance and objective workload evaluation research (POWER)**. FEDERAL AVIATION ADMINISTRATION OKLAHOMA CITY OK CIVIL AEROMEDICAL INST, 2001.

MANSIKKA, Heikki et al. Fighter pilots' heart rate, heart rate variation and performance during instrument approaches. **Ergonomics**, v. 59, n. 10, p. 1344-1352, 2016a.

MANSIKKA, Heikki et al. Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. **Applied ergonomics**, v. 56, p. 213-219, 2016b.

MANSIKKA, Heikki; VIRTANEN, Kai; HARRIS, Don. 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**, n. just-accepted, p. 1-22, 2018.

MARTÍNEZ, Rosario Aguirre. Evaluación ergonómica de carga mental y factores de riesgo psicosocial del trabajo, en profesionales, técnicos y administrativos de un centro de rehabilitación de salud mental privado. **Revista Chilena de Terapia Ocupacional**, n. 10, p. Pág. 85-97, 2010.

MARTINS, C. O. **Repercussão de um programa de ginástica laboral na qualidade de vida de trabalhadores de escritório**. 2005. 184 f. Tese (Doutorado em Engenharia da Produção) - Universidade Federal de Santa Catarina, Florianópolis, 2005.

MARTIN, J. et al. Linear and non-linear heart rate metrics for the assessment of anaesthetists' workload during general anaesthesia. **BJA: British Journal of Anaesthesia**, v. 117, n. 6, p. 767-774, 2016.

MARTIN, J. A. et al. Objective structured assessment of technical skill (OSATS) for surgical residents. **British journal of surgery**, v. 84, n. 2, p. 273-278, 1997.

MASCULO, Francisco Soares; VIDAL, Mario Cesar. **Ergonomia: trabalho adequado e eficiente**. Elsevier Brasil, 2013.

MAZLOUMI, A. et al. Assessment of mental workload Air Traffic Controllers based on task load factors in Air Traffic Control simulator. **Iran Occupational Health**, v. 13, n. 4, p. 39-48, 2016.

MCCORMACK, H. M.; DAVID, J. de L.; SHEATHER, S. Clinical applications of visual analogue scales: a critical review. **Psychological medicine**, v. 18, n. 4, p. 1007-1019, 1988.

MESHKATI, Najmedin et al. Techniques in mental workload assessment. 1995.

METZGER, Ulla; PARASURAMAN, Raja. The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. **Human Factors**, v. 43, n. 4, p. 519-528, 2001.

MIGUEL, Paulo Augusto Cauchick. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. Elsevier, 2012.

MINISTÉRIO DA FAZENDA. 1º Boletim Quadrimestral sobre Benefícios por Incapacidade de 2017. Adoecimento Mental e Trabalho: a concessão de benefícios por incapacidade relacionados a transtornos mentais e comportamentais entre 2012 e 2016. **Secretaria de Previdência**, 2017.

MITROPOULOS, Panagiotis; MEMARIAN, Babak. Task demands in masonry work: Sources, performance implications, and management strategies. **Journal of Construction Engineering and Management**, v. 139, n. 5, p. 581-590, 2012.

MONCADA, Salvador; LLORENS, Clara; KRISTENSEN, T. S. Manual para la evaluación de riesgos psicosociales en el trabajo. **Método Istars21 (CoPsoQ). España. Instituto Sindical de Trabajo, Ambiente y Salud (ISTAS)**, 2002.

MONCADA, S. et al. ISTAS21: Versión en lengua castellana del cuestionario psicosocial de Copenhague (COPSOQ). **Arch Prev Riesgos Labor**, v. 8, n. 1, p. 18-29, 2005.

MONTMOLLIN, Maurice de. **Vocabulaire de L' Ergonomie**. Paris: Octares (p.), 1995.

MOON, B. S. et al. Fuzzy systems to process ECG and EEG signals for quantification of the mental workload. **Information Sciences**, v. 142, n. 1, p. 23-35, 2002.

MOURA, W. **Trabalho e Doença Existencial: uma visão psicossociológica das doenças ocupacionais**. Rio de Janeiro: Laboratório de Editoração da UFRJ, 1998.

MORAY, Neville. Mental workload since 1979. **International Reviews of Ergonomics**, v. 2, p. 123-150, 1988.

MOUZÉ-AMADY, Marc et al. Fuzzy-TLX: using fuzzy integrals for evaluating human mental workload with NASA-Task Load index in laboratory and field studies. **Ergonomics**, v. 56, n. 5, p. 752-763, 2013.

MURAI, Koji et al. The mental workload of a ship's navigator using heart rate variability. **Interactive Technology and Smart Education**, v. 1, n. 2, p. 127-133, 2004.

MURAI, Koji et al. Evaluation of ship navigator's mental workload using nasal temperature and heart rate variability. In: **Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on**. IEEE, 2008. p. 1528-1533.

MURAI, Koji et al. Toward development of monitoring system of port coordinator's mental workload using adhesive plaster-type sensor. In: **World Automation Congress (WAC), 2016**. IEEE, 2016. p. 1-5.

MURAI, Koji et al. A characteristic of a navigator's mental workload based on nasal temperature. In: **Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on**. IEEE, 2007. p. 3639-3643.

NOBLE, Bruce J. et al. A category-ratio perceived exertion scale: relationship to blood and muscle lactates and heart rate. **Medicine and science in sports and exercise**, v. 15, n. 6, p. 523-528, 1983.

NORMAN, Donald A.; BOBROW, Daniel G. On data-limited and resource-limited processes. **Cognitive psychology**, v. 7, n. 1, p. 44-64, 1975.

NYGREN, Thomas E. Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. **Human Factors**, v. 33, n. 1, p. 17-33, 1991.

O'DONNELL, R.D. & EGGEMEIER, F.T. (1986). Workload assessment methodology. In K.R. BOFF, L. KAUFMAN & J.P. THOMAS (Eds.), **Handbook of perception and human performance**. Vol. II, cognitive processes and performance. p. 42-49, New York: Wiley, 1986.

OLIVEIRA, C. C. **As dificuldades encontradas no pré-natal e sua relação as gestantes sífilíticas: uma revisão integrativa**. 2015. 49 f. Dissertação (Mestrado em Saúde Coletiva) – Universidade de Fortaleza, Fortaleza, 2015.

OLIVEIRA, F. F. de. **Contribuições da pesquisa operacional para gestão da produção e operações: uma análise exploratória da literatura**. 2011. 129 f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio Grande do Norte, Natal, 2011.

OLIVEIRA, I. C. M. **Acidentes De Trânsito E A Aplicação De Medidas Preventivas: Uma Revisão Integrativa**. 2016. 76 f. Dissertação (Mestrado em Saúde Coletiva) – Universidade Federal do Acre, Rio Branco, 2016.

OLIVEIRA SECCO, Iara Aparecida de et al. Cargas psíquicas de trabalho e desgaste dos trabalhadores de enfermagem de hospital de ensino do Paraná, Brasil. **SMAD, Revista Electrónica en Salud Mental, Alcohol y Drogas**, v. 6, n. 1, 2010.

OÑATE, Esteban; MEYER, Felipe. Ergonomic study of an operator's work of a molybdenum plant. **Work**, v. 41, n. Supplement 1, p. 5950-5955, 2012.

OTHMAN, Norashiken; ROMLI, Fairuz Izzuddin. Mental Workload Evaluation of Pilots Using Pupil Dilation. **International Review of Aerospace Engineering (IREASE)**, v. 9, n. 3, p. 80-84, 2016.

PARK, Jinkyun; JUNG, Wondea. A study on the validity of a task complexity measure for emergency operating procedures of nuclear power plants—comparing with a subjective workload. **IEEE transactions on nuclear science**, v. 53, n. 5, p. 2962-2970, 2006.

PEKKARINEN, Laura et al. Job demands and musculoskeletal symptoms among female geriatric nurses: The moderating role of psychosocial resources. **Journal of occupational health psychology**, v. 18, n. 2, p. 211, 2013.

PESSOA, I. C. **Estatística das aplicações de métodos multicritério nas áreas da engenharia de produção**. 2016. 171 f. Dissertação (Mestrado em Engenharia de Produção e Sistemas) - Universidade Tecnológica Federal do Paraná, Pato Branco, 2016.

PFEIFFER, Linda et al. The mental organization of air traffic and its implications to an emotion sensitive assistance system. **Int. J. Adv. Life Sci**, v. 8, p. 164-174, 2015.

PICKUP, Laura et al. The Integrated Workload Scale (IWS): a new self-report tool to assess railway signaller workload. **Applied Ergonomics**, v. 36, n. 6, p. 681-693, 2005.

PREVOT, Thomas et al. Arrival Metering Precision Study. In: **AIAA Modeling and Simulation Technologies Conference**. 2015. p. 0653.

RANTANEN, Esa M.; MAYNARD, Philip W.; ÖZHAN, Deniz. The impact of sector characteristics and aircraft count on air traffic control communications and workload. 2005.

REID, G. B.; NYGREN, T. E. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In: **Advances in psychology**. North-Holland, 1988. p. 185-218.

REID, G. B.; SHINGLEDECKER, C. A.; EGGEMEIER, F. T. Application of conjoint measurement to workload scale development. In: **Proceedings of the Human Factors Society Annual Meeting**. Sage CA: Los Angeles, CA: Sage Publications, 1981. p. 522-526.

RICHARD, J. F. **Les Activités Mentales**. Paris: Armand Colin, 1990.

RIVERA, Miguel Macías; ROMERO, Raúl Rocha. Medición De Carga Mental De Trabajo En La Industria Automotriz En México. **European Scientific Journal, ESJ**, v. 12, n. 26, 2016.

RODRÍGUEZ, Sara et al. Pupillometry to assess Air Traffic Controller workload through the Mental Workload Model. In: **Proceedings of the 5th International Conference on Application and Theory of Automation in Command and Control Systems**. ACM, 2015. p. 95-104.

ROLO GONZÁLEZ, Gladys; DÍAZ CABRERA, Dolores; HERNÁNDEZ FERNAUD, Estefanía. Desarrollo de una escala subjetiva de carga mental de trabajo (ESCAM). **Revista de Psicología del Trabajo y de las Organizaciones**, v. 25, n. 1, p. 29-37, 2009.

ROLO, Gladys; HERNÁNDEZ-FERNAUD, Estefanía; DÍAZ-CABRERA, Dolores. Impacto de las condiciones físico-ambientales percibidas sobre la carga mental de trabajo: un estudio exploratorio en empleados de oficina. **Psycology**, v. 1, n. 3, p. 333-342, 2010.

ROSÁRIO, A. M. M. do. **Avaliação de instrumentos que investigam abuso de álcool e outras drogas em adolescentes: revisão de literatura**. 2011. 177 f. Dissertação (Mestrado em Ciências) - Universidade de São Paulo, São Paulo, 2011.

RUBIO VALDEHITA, Susana et al. LA CARGA MENTAL COMO FACTOR DE RIESGO PSICOSOCIAL. DIFERENCIAS POR BAJA LABORAL. **Ansiedad y estrés**, v. 16, 2010.

RUBIO, Susana et al. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. **Applied Psychology**, v. 53, n. 1, p. 61-86, 2004.

RUBIO-VALDEHITA, Susana; RODRIGO-TAPIAS, Isabel. Sources of Mental Workload in a Sample of Nurses and Auxiliary Nursing Technicians from Madrid (Spain). **REVISTA IBEROAMERICANA DE DIAGNOSTICO Y EVALUACION-E AVALIACAO PSICOLOGICA**, v. 1, n. 43, p. 177-185, 2017.

RUIZ, Milton Artur; GRECO, Oswaldo Tadeu; BRAILE, Domingo Marcolino. Fator de impacto: importância e influência no meio editorial, acadêmico e científico. **Rev Bras Cir Cardiovasc**, v. 24, n. 3, p. 273-8, 2009.

SÁENZ, María Pamela Bustamante et al. Evaluación de carga mental de trabajo en profesionales de enfermería del turno vespertino de un hospital público de Ciudad Juárez, Chihuahua. **CULCyT**, n. 56, 2016.

SAFARI, Shahram et al. Personnel's health surveillance at work: effect of age, body mass index, and shift work on mental workload and work ability index. **Journal of environmental and public health**, v. 2013, 2013.

SALEEM, Jason J. et al. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. **Journal of the American Medical Informatics Association**, v. 14, n. 5, p. 632-640, 2007.

SANDERS, M. K.; McCORMICK, E. J. **Human Factors in Engineering and Design**, New York: McGraw-Hill, 1993.

SANTOS, A. A. **Revisão sistemática de estudos sobre a síndrome de burnout em professores do ensino fundamental e médio**. 2010. 89 f. Dissertação (Mestrado em Saúde Coletiva) - Universidade Estadual de Feira de Santana, Feira de Santana, 2010.

SANTOS, J. M. L. **Erros de prescrição de medicamentos em pacientes hospitalizados-revisão de literatura**. 2010. 143 f. Dissertação (Mestrado em Saúde Pública) - Universidade de São Paulo, São Paulo, 2010.

SCERBO, Mark W.; BRITT, Rebecca C.; STEFANIDIS, Dimitrios. Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. **The American Journal of Surgery**, v. 213, n. 2, p. 244-248, 2017.

SCHONBLUM, Rosane. **Atributos psicométricos necessários à construção de uma medida de carga mental de trabalho**. Dissertação de Mestrado – Universidade Federal de Santa Catarina, Centro de Ciências Humanas. Programa de Pós-Graduação em Psicologia, Florianópolis, 2004.

SCHULZ, Christian M. et al. Assessment of subjective workload in an anaesthesia simulator environment: reliability and validity. **European Journal of Anaesthesiology (EJA)**, v. 28, n. 7, p. 502-505, 2011.

SCHULZ, Christian M. et al. Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment. **British journal of anaesthesia**, v. 106, n. 1, p. 44-50, 2010.

SCHÜTZ, Gustavo Ricardo; SANTOS, Saray Giovana dos; SANT'ANA, Antônio Sérgio Santos. Política de periódicos nacionais em Educação Física para estudos de revisão/sistemática. **Rev. bras. cineantropom. desempenho hum**, v. 13, n. 4, p. 313-319, 2011.

SCOTT, J.; HUSKISSON, E. C. Graphic representation of pain. **Pain**, v. 2, n. 2, p. 175-184, 1976.

SELL, I. **Projeto do trabalho humano**. Florianópolis: Editora da UFSC, 2002.

SELIGMANN-SILVA, E. **Desgaste mental no trabalho dominado**. São Paulo: Cortez, 1994.

SIEGEL, A. W.; SCHRAAGEN, Johannes Martinus Cornelis. Measuring workload weak resilience signals at a rail control post. **IIE transactions on occupational ergonomics and human factors**, v. 2, n. 3-4, p. 179-193, 2014.

SIEGEL, A. Willy; SCHRAAGEN, Johannes Martinus Cornelis. Beyond procedures: Team reflection in a rail control centre to enhance resilience. **Safety science**, v. 91, p. 181-191, 2017.

SILVA, Andrea Aparecida; DE OLIVEIRA LUCAS, Elaine Rosangela. Abordagem ergonômica do ambiente de trabalho na percepção dos trabalhadores: estudo de caso

em biblioteca universitária Approach ergonomics of the environment of work in the perception of the workers: case study in university library. **Revista ACB**, v. 14, n. 2, p. 382-406, 2009.

SÖNMEZ, Betül et al. Determination of nurses' mental workloads using subjective methods. **Journal of clinical nursing**, 2016.

SPERANDIO, J. C. Variation of operator's strategies and regulating effects on workload. **Ergonomics**, v. 14, n. 5, p. 571-577, 1971.

STEIN, E. S. Air traffic controller workload: An examination of workload probe (Report FAA/CTTN90/60). **Atlantic City, New Jersey: FAA**, 1985.

STURREA, Lindsay K. et al. Human factors assessment of disturbances to scheduled performance-based navigation arrival operations. **Procedia Manufacturing**, v. 3, p. 2442-2449, 2015.

TATTERSALL, Andrew J.; FOORD, Penelope S. An experimental evaluation of instantaneous self-assessment as a measure of workload. **Ergonomics**, v. 39, n. 5, p. 740-748, 1996.

TESI, Alessio; AIELLO, Antonio; GIANNETTI, Enrichetta. The work-related well-being of social workers: Framing job demands, psychological well-being, and work engagement. **Journal of Social Work**, p. 1468017318757397, 2018.

TINOCO, M. M.. **A relação saúde/doença no processo de trabalho dos Agentes Comunitários de Saúde: uma revisão de literatura**. 2015. 92 f. Dissertação (Mestrado em Ciências) - Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro, 2015.

TRUITT, Todd R. et al. Integrating UAS Operations in Class C Airspace. 2015.

VASSILIOU, Melina C. et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. **The American journal of surgery**, v. 190, n. 1, p. 107-113, 2005.

VASSILIOU, Melina C. et al. Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. **Surgical innovation**, v. 14, n. 3, p. 211-216, 2007.

VÁSQUEZ, P. A. C. et al. Validación de la Escala Subjetiva de Carga mental de trabajo (ESCAM) en profesionales de la salud de Chile. **Universitas Psychologica**, v. 15, n. 1, 2016.

VELÁZQUEZ, Francisco Farrer et al. **Manual de ergonomía**. Fundación MAPFRE, 1995.

VILAS-BOAS, Madalena; CERQUEIRA, Alzira. Assessing stress at work: The Portuguese version of the Job Content Questionnaire. **Avaliação Psicológica**, v. 16, n. 1, 2017.

WARD, Nicholas J. et al. An evaluation of a lane support system for bus rapid transit on narrow shoulders and the relation to bus driver mental workload. **Ergonomics**, v. 49, n. 9, p. 832-859, 2006.

WEIGL, Matthias et al. Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-scale simulated procedure. **Surgical endoscopy**, v. 30, n. 2, p. 559-566, 2016.

WEIGL, Matthias et al. The impact of intra-operative interruptions on surgeons' perceived workload: an observational study in elective general and orthopedic surgery. **Surgical endoscopy**, v. 29, n. 1, p. 145-153, 2015.

WEINGER, Matthew B. et al. Quantitative description of the workload associated with airway management procedures. **Journal of clinical anesthesia**, v. 12, n. 4, p. 273-282, 2000.

WEINGER, Matthew B.; REDDY, Swapna B.; SLAGLE, Jason M. Multiple measures of anesthesia workload during teaching and nonteaching cases. **Anesthesia & Analgesia**, v. 98, n. 5, p. 1419-1425, 2004.

WEWERS, M. E.; LOWE, N. K. A critical review of visual analogue scales in the measurement of clinical phenomena. **Research in nursing & health**, v. 13, n. 4, p. 227-236, 1990.

WHEELLOCK, Ana et al. The impact of operating room distractions on stress, workload, and teamwork. **Annals of surgery**, v. 261, n. 6, p. 1079-1084, 2015.

WICKENS, Christopher D. Multiple resources and performance prediction. **Theoretical issues in ergonomics science**, v. 3, n. 2, p. 159-177, 2002.

WIDYANTI, Ari; JOHNSON, Addie; DE WAARD, Dick. Adaptation of the rating scale mental effort (RSME) for use in Indonesia. **International Journal of Industrial Ergonomics**, v. 43, n. 1, p. 70-76, 2013.

WIDYANTI, A.; LARUTAMA, W. The relation between performance of lean Manufacturing and employee's mental workload. In: **Industrial Engineering and Engineering Management (IEEM), 2016 IEEE International Conference on**. IEEE, 2016. p. 252-256.

WIERWILLE, Walter W. Physiological measures of aircrew mental workload. **Human Factors**, v. 21, n. 5, p. 575-593, 1979.

WIERWILLE, Walter W.; CASALI, John G. A validated rating scale for global mental workload measurement applications. In: **Proceedings of the Human Factors society Annual Meeting**. Sage CA: Los Angeles, CA: Sage Publications, 1983. p. 129-133.

WIERWILLE, Walter W.; CONNOR, Sidney A. Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. **Human factors**, v. 25, n. 1, p. 1-16, 1983.

WIERWILLE, Walter W.; RAHIMI, Mansour; CASALI, John G. Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. **Human Factors**, v. 27, n. 5, p. 489-502, 1985.

WILLIGES, Robert C.; WIERWILLE, Walter W. Behavioral measures of aircrew mental workload. **Human Factors**, v. 21, n. 5, p. 549-574, 1979.

WILSON, Glenn F. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. **The International Journal of Aviation Psychology**, v. 12, n. 1, p. 3-18, 2002.

WILSON, Glenn F.; RUSSELL, Chris A. Operator functional state classification using multiple psychophysiological features in an air traffic control task. **Human Factors**, v. 45, n. 3, p. 381-389, 2003.

WILSON, Glenn F.; CALDWELL, John A.; RUSSELL, Christopher A. Performance and psychophysiological measures of fatigue effects on aviation related tasks of varying difficulty. **The international journal of aviation psychology**, v. 17, n. 2, p. 219-247, 2007.

WILSON, Katie A. et al. Forecaster Performance and Workload: Does Radar Update Time Matter?. **Weather and Forecasting**, v. 32, n. 1, p. 253-274, 2017.

WILSON, Mark R. et al. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). **World journal of surgery**, v. 35, n. 9, p. 1961, 2011.

WISNER, A. **Por dentro do trabalho: ergonomia: método & técnica**. São Paulo: FTD, 1987.

WISNER, A. **A inteligência no trabalho: textos selecionados de ergonomia**. São Paulo: Fundacentro, 1994.

YANG, Chih-Wei et al. Assessing mental workload and situation awareness in the evaluation of computerized procedures in the main control room. **Nuclear Engineering and Design**, v. 250, p. 713-719, 2012.

YAO, Yong-Jie et al. Heart rate and respiration responses to real traffic pattern flight. **Applied psychophysiology and biofeedback**, v. 33, n. 4, p. 203-209, 2008.

YU, Denny et al. Quantifying Intraoperative Workloads Across the Surgical Team Roles: Room for Better Balance?. **World journal of surgery**, v. 40, n. 7, p. 1565-1574, 2016.

ZARE, Sajad et al. The Relationship between Mental Workload and Prevalence of Musculoskeletal Disorders among Welders of Tehran Heavy Metal Structures Company in 2016. **Journal of Biology and Today's World**, v. 5, n. 12, p. 218-223, 2016.

ZAWIAH, Nurul Izzah Abd Rahman and Siti; DAWAL, Md. The mental workload and alertness levels of train drivers under simulated conditions based on electroencephalogram signals. **Malaysian Journal of Public Health Medicine**, v. 16, n. 1, p. 115-123, 2016.

ZHENG, Bin et al. Measuring mental workload during the performance of advanced laparoscopic tasks. **Surgical endoscopy**, v. 24, n. 1, p. 45, 2010.

ZHENG, Bin et al. Workload assessment of surgeons: correlation between NASA TLX and blinks. **Surgical endoscopy**, v. 26, n. 10, p. 2746-2750, 2012a.

ZHENG, Bin et al. Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. **Surgical endoscopy**, v. 26, n. 5, p. 1352-1358, 2012b.

ZIJLSTRA, Ferdinand Rudolf Hendrikus. Efficiency in work behaviour: A design approach for modern tools. 1993.

ZOER, Ilona et al. The associations between psychosocial workload and mental health complaints in different age groups. **Ergonomics**, v. 54, n. 10, p. 943-952, 2011.

APÊNDICE A - Dissertações de Revisão de Literatura

(continua)

Autor/Ano	Objetivo	Metodologia	Bases	Período	Como coletou	Amostra	O que analisou
ANDRADE, 2009	Analisar as evidências a respeito da <i>Diabetes Mellitus</i> (DM) como fator de risco para a Doença Periodontal (DP), a partir de dados da literatura científica.	Revisão sistemática de literatura	MEDLINE, SciELO, LILACS, Healthstar, SciSearch	1989 - 2007	<p>Descritores: "<i>diabetes periodontal disease case control study</i>", "<i>doença periodontal diabetes caso controle</i>", "<i>diabetes and periodontal disease</i>" e "<i>doença periodontal e diabetes</i>".</p> <p>Processo de filtragem.</p>	26 artigos	<p>a) Tipo de estudo (caso controle, transversal e longitudinal);</p> <p>b) País de publicação;</p> <p>c) Ano de publicação;</p> <p>d) Participantes: diabéticos (tipo 1 ou 2) e não diabéticos;</p> <p>e) Medidas de associação: perda de nível de inserção e/ou perda óssea;</p> <p>f) Desfecho clínico: periodontite;</p> <p>g) Significância estatística entre DM e DP;</p> <p>h) Medida estatísticas de associação utilizada nos trabalhos.</p>
SANTOS, A., 2010	Apresentar uma revisão sistemática de estudos de corte transversal sobre a prevalência da síndrome de <i>burnout</i> em professores de ensino fundamental e médio, assim como os fatores considerados de risco para seu desenvolvimento e consequências para o indivíduo e a organização.	Revisão sistemática de literatura de estudos epidemiológicos	BVS, PsycINFO, LILACS, MEDLINE (PubMed), SciELO	1989 - 2009	<p>Descritores: Cruzamento do termo "<i>Burnout</i>" com os termos (1) <i>Burnout Syndrome</i>, (2) <i>prevalence</i>, (3) <i>Risk factors</i>, (4) <i>Teacher</i>, (5) <i>crosssectional studies</i>; e (6) <i>survey</i>.</p> <p>Processo de filtragem.</p>	35 artigos	<p>a) Primeiro autor;</p> <p>b) País onde foi realizado o estudo;</p> <p>c) Ano de publicação;</p> <p>d) Periódico publicado;</p> <p>e) Número de participantes total no estudo;</p> <p>f) Características sociodemográficas e do trabalho de cada estudo;</p> <p>g) Resultado/Desfecho de cada artigo;</p> <p>h) Análise dos componentes das dimensões Exaustão Emocional (EE), Despersonalização (DE) e Realização Pessoal (RP).</p>

(continua)

Autor/Ano	Objetivo	Metodologia	Bases	Período	Como coletou	Amostra	O que analisou
SANTOS, J., 2010	Analisar a literatura sobre erros de prescrição de medicamentos ocorridos em pacientes adultos hospitalizados em unidades de clínica médica e cirúrgica.	Revisão de literatura	MEDLINE/PubMed, IPA, LILACS, Embase, Web of Science, Scopus	1999 - 2009	<p>Palavras-chave: <i>medication error (s)</i>, <i>medication order (s)</i>, <i>prescribing error (s)</i>, <i>prescription error (s)</i>, prescrição, hospital (s), <i>medication system, inpatient (s)</i>, <i>adult (s)</i>.</p> <p>Processo de filtragem.</p>	51 artigos	<p>a) Ano e local de publicação; b) Definição de erro de medicação e erro de prescrição; c) Métodos utilizados na detecção de erros de prescrição: - Tipo de estudo e métodos para coleta de dados; - Período de estudo e tamanho da amostra; - Profissional envolvido na coleta de dados; - Tipo de prescrição analisada; d) Classificação de erros de prescrição: - Quanto ao tipo; - Quanto à gravidade; e) Frequência de erros de prescrição; - Medicamentos envolvidos; f) Causas associadas a erros de prescrição; g) Estratégias de prevenção de erros de prescrição.</p>

(continua)

Autor/Ano	Objetivo	Metodologia	Bases	Período	Como coletou	Amostra	O que analisou
OLIVEIRA, F., 2011	Realizar uma análise exploratória da literatura para avaliar como as áreas da Pesquisa Operacional (PO) têm contribuído na modelagem das principais decisões em gestão da produção e operações (GPO) nos últimos 30 anos.	Revisão teórica de literatura	Periódicos: <i>European Journal of Operational Research</i> , <i>International Journal of Operation & Production Management</i> , <i>Journal of Operational Research Society</i> , <i>Production and Operations Management Society</i> , Revista Produção, Revista de Pesquisa Operacional.	1980 - 2009	Palavras-chave: estoque, programação, capacidade, layout. Processos de filtragem.	857 artigos	a) Separação dos artigos dentro de 4 subáreas: Layout, Planejamento da Capacidade, Programação da Produção e Gestão de Estoques; b) Análises dos artigos (dentro da classificação e de forma geral): - Número de artigos de cada periódico por década; - Número de artigos de cada área por periódico e por ano; - Teste de hipóteses em decisões entre as áreas dentro da subárea maior.
ROSÁRIO, 2011	Revisar estudos e descrever três importantes instrumentos validados para o Brasil e utilizados para avaliação do abuso de álcool e outras drogas em adolescentes.	Revisão de literatura	Adolec, MEDLINE, LILACS, PubMed, SciELO	1998 - 2010	Descritores: Teen ASI ou T-ASI, <i>Drug Use Screening Inventory</i> (DUSI), DUSI-R e CBCL. Processo de filtragem.	54 artigos	a) Descrição de cada um dos três instrumentos: Teen ASI, DUSI-R e CBCL; b) Classificação dos artigos quanto ao tipo de cada um dos instrumentos; - Tipo de aplicação dentro do instrumento; - Amostra; - País de aplicação; - Objetivo; - Resultados; - Conclusão; - Contexto das aplicações; c) Vantagens e limitações de cada instrumento.

(continua)

Autor/Ano	Objetivo	Metodologia	Bases	Período	Como coletou	Amostra	O que analisou
DORNELES, 2014	Identificar quais indicadores possuem atributos desejáveis para avaliar a atenção à saúde materno-infantil, subsidiando a tomada de decisão da gestão.	Revisão de literatura	Embase, LILACS, Cochrane, SciSarch, PAHO, PubMed/MEDLINE: Scopus, Emerald		Palavras-chave e descritores. Processo de filtragem.	6 artigos	a) Indicadores utilizados para analisar a saúde materno infantil; b) Local do estudo; c) Amostra; d) Classificação dos indicadores selecionados (objetivo, validade, contribuição para inovação, facilidade para coleta de dados, viabilidade/aceitabilidade, subdimensão).
LUSTOSA, 2015	Analisar a produção científica nacional e internacional referente aos cuidados paliativos e ao manejo da dor e discutir o manejo da dor em pacientes que experimentam o processo de terminalidade.	Revisão integrativa de literatura	SciELO, LILACS, MEDLINE	2009 - 2013	Palavras-chave: "dor" OR "pain" AND "cuidados paliativos" OR "palliative care". Processo de filtragem.	80 artigos	a) Ano da publicação; b) Periódico; c) País, idioma e área de publicação; d) Fator de impacto dos periódicos; e) Classificação em 6 categorias que emergiram a partir do processo de categorização; f) Trechos das publicações relevantes dentro de cada categoria.

(continua)

Autor/Ano	Objetivo	Metodologia	Bases	Período	Como coletou	Amostra	O que analisou
OLIVEIRA, C., 2015	Analisar por meio de publicados as dificuldades enfrentadas por mulheres com sífilis no pré-natal; sua relação com os serviços de saúde e os prejuízos para o feto.	Revisão integrativa de literatura	Biblioteca Virtual de Saúde	2010 - 2014	Descritores: assistência pré-natal e sífilis congênita. Processo de filtragem.	12 artigos	a) Título; b) Autores; c) Ano de publicação; d) Local de publicação; e) População estudada; f) Objetivos do estudo; g) Metodologia; h) Resultados; i) Conclusões; j) Assistência Pré-Natal e sua relação com a Sífilis Congênita; k) Qualidade do Pré-natal; l) Desfechos neonatais relacionados ao pré-natal.
TINOCO, 2015	Identificar os efeitos produzidos na saúde dos Agentes Comunitários de Saúde pelo processo de trabalho, a partir de uma detalhada revisão de literatura.	Revisão integrativa de literatura	Biblioteca Virtual em Saúde (BVS), SciELO, CAPES, LILACS, Scopus	Até 2014	Descritores, palavras-chave e operadores booleanos. Processo de filtragem.	37 artigos	a) Divisão em quatro categorias que apoiam as reflexões sobre os efeitos produzidos na saúde dos Agentes Comunitários de Saúde: Desgaste e cargas de trabalho, Estresse e Esgotamento profissional, Estratégias de enfrentamento das dificuldades, Prazer e satisfação. b) Análises dos artigos dentro de cada uma das categorias.

(continua)

Autor/Ano	Objetivo	Metodologia	Bases	Período	Como coletou	Amostra	O que analisou
DUMMEL, 2016	Investigar a literatura sobre o conhecimento e condutas dos profissionais de saúde diante da relação entre as doenças periodontais e as doenças cardiovasculares.	Revisão de literatura	Biblioteca Virtual da Saúde (BVS), Scopus, Web of Science, SocINDEX	2009 - 2015	Associação de palavras-chave. Processo de filtragem.	11 artigos	a) Ano da publicação; b) Classe do profissional; c) País de origem; d) Periódico; e) Forma como o estudo foi conduzido para que se obtivessem as informações dos profissionais; f) Conhecimento sobre a relação entre doença periodontal e doença cardiovascular; g) Particularidades do estudo.
OLIVEIRA, I., 2016	Revisar o conhecimento científico produzido acerca da efetividade das medidas de prevenção utilizadas na redução dos acidentes de trânsito no âmbito da saúde pública.	Revisão integrativa de literatura	Biblioteca Virtual em Saúde on-line (BIREME), com índices bibliográficos contidos no LILACS, MEDLINE, IBECs, Scielo	2005 - 2015	Descritores: Acidentes de Trânsito, Prevenção de Acidentes, Prevenção e Controle, Políticas Públicas, Medição de Velocidade e Saúde da População Urbana. Processo de filtragem.	14 artigos	a) Autor do artigo; b) Ano de publicação; c) Local do estudo; d) População do estudo; e) Intervenção/comparação; f) Resultados; g) Agrupamentos por semelhança e organização em categorias temáticas.

(conclusão)

Autor/Ano	Objetivo	Metodologia	Bases	Período	Como coletou	Amostra	O que analisou
PESSOA, 2016	Realizar uma pesquisa na literatura, utilizando artigos científicos de abrangência internacional a fim de apresentar a distribuição estatística das aplicações dos métodos multicritério nas áreas da Eng. de Produção.	Pesquisa bibliográfica	Scopus	2010 - 2015	ProKnow-C. Processos de filtragem.	935 artigos	a) Ano de publicação; b) Periódico; c) Métodos multicritério; d) Nacionalidade do primeiro autor; e) Fator de Impacto; f) Classificação dos artigos quanto a área da Engenharia de Produção; g) Relação entre o método multicritério usado e a área da Engenharia de Produção.
GOMES, 2017	Qualificar e quantificar os processos judiciais relacionados à saúde, que versem sobre erro médico, em 2ª instância, no Tribunal de Justiça do Distrito Federal e Territórios (TJDFT), durante o período de 2013 e 2015.	Pesquisa exploratória, documental, descritivo-analítica e quali-quantitativa	PROCESSOS: Sítio eletrônico do TJDFT. ARTIGOS: Biblioteca Virtual em Saúde – BVS, LILACS e MEDLINE.	PROC.: 2013 - 2015 ART.: 2014 - 2015	PROCESSOS: Palavra-chave: "erro médico". Apelações Cíveis e Apelações Criminais julgadas em 2ª instância. ARTIGOS: Descritores: "segurança do paciente AND direito do paciente", "segurança do paciente AND erro médico", " <i>patient safety</i> AND <i>patient rights</i> ", " <i>patient safety</i> AND <i>malpractice</i> ". Processo de filtragem.	PROC.: 202 ART.: 89	PROCESSOS: a) Ano do processo; b) Tipos de ação; c) Tipo de danos reclamados; d) Especialidade médica envolvida na ação; e) Identificação do polo passivo das ações judiciais; f) Tipos de decisão; g) Análise idiosincrática para analisar o discurso emitido nos acórdãos. ARTIGOS: a) Área em que estão relacionados; b) Natureza da pesquisa; c) Palavras-chave utilizadas nos trabalhos; d) Conceito de erro médico.

APÊNDICE B – Artigos do PB organizados conforme as ferramentas de análise da CTM

Ferramenta	Artigo
Medidas fisiológicas	(BERGUER; SMITH; CHUNG, 2001), (LING et al., 2001), (EAST; BAUER JR; LANNING, 2002), (MOON et al., 2002), (WILSON, 2002), (COLLET et al., 2003), (WILSON; RUSSELL, 2003), (KUIJER et al., 2004), (MURAI et al., 2004), (AHLSTROM; FRIEDMAN-BERG, 2006), (KERICK; ALLENDER, 2006), (DI NOCERA; CAMILLI; TEREZI, 2007), (LAHTINEN et al., 2007), (MURAI et al., 2007), (WILSON; CALDWELL; RUSSELL, 2007), (DE RIVECOURT et al., 2008), (HANNULA et al., 2008), (MURAI et al., 2008), (YAO et al., 2008), (GOULD et al., 2009), (HENELIUS et al., 2009), (LIANG et al., 2009), (DEY; MANN, 2010), (LEHRER et al., 2010), (AYAZ et al., 2011), (HUTTUNEN et al., 2011), (KATAOKA; SASAKI; KANDA, 2011), (SCHULZ et al., 2011), (AYAZ et al., 2012), (ZHENG et al., 2012a), (MOUZÉ-AMADY et al., 2013), (HARRISON et al., 2014), (FALLAHI et al., 2016), (MANNARU et al., 2016a), (MANNARU et al., 2016b), (MANSIKKA et al., 2016a), (MANSIKKA et al., 2016b), (MARTIN et al., 2016), (MURAI et al., 2016), (OTHMAN; ROMLI, 2016), (ZAWIAH; DAWAL, 2016)
NASA-TLX	(GALSTER et al., 2001), (METZGER; PARASURAMAN, 2001), (WILSON; RUSSELL, 2003), (AVERTY et al., 2004), (GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006), (PARK; JUNG, 2006), (BOULTINGHOUSE et al., 2007), (DI NOCERA; CAMILLI; TEREZI, 2007), (SALEEM et al., 2007), (GOULD et al., 2009), (BYRNE et al., 2010), (LEHRER et al., 2010), (LOPEZ et al., 2010), (RUBIO VALDEHITA et al., 2010), (KATAOKA; SASAKI; KANDA, 2011), (ARELLANO et al., 2012), (MITROPOULOS; MEMARIAN, 2012), (YANG et al., 2012), (ZHENG et al., 2012a), (ZHENG et al., 2012b), (MOUZÉ-AMADY et al., 2013), (SAFARI et al., 2013), (DADI et al., 2014), (LEE et al., 2014), (KOCA et al., 2015), (WHEELLOCK et al., 2015), (DARVISHI et al., 2016), (ERASLAN; CAN; ATALAY, 2016), (LIN; LU, 2016), (MARTIN et al., 2016), (MAZLOUMI et al., 2016), (OTHMAN; ROMLI, 2016), (SÖNMEZ et al., 2016), (ZARE et al., 2016), (RUBIO-VALDEHITA; RODRIGO-TAPIAS, 2017), (SCERBO; BRITT; STEFANIDIS, 2017)
Escala subjetiva	(WILSON, 2002), (KUIJER et al., 2004), (LAHELMA et al., 2005), (LAHTINEN et al., 2007), (HUTTUNEN et al., 2011), (KARHULA et al., 2013)
Questionário	(WEINGER et al., 2000), (KAWADA; OOYA, 2005), (LIANG et al., 2009), (ZOER et al., 2011), (ELFERING; GREBNER; HALLER, 2012)
SWAT	(PICKUP et al., 2005), (KERICK; ALLENDER, 2006), (ERASLAN; CAN; ATALAY, 2016), (FALLAHI et al., 2016)
Tarefa secundária	(BYRNE et al., 2010), (ZHENG et al., 2010), (ZHENG et al., 2012b), (BYRNE et al., 2013)
ATWIT	(AHLSTROM; FRIEDMAN-BERG, 2006), (HARRISON et al., 2014)
RSME	(WARD et al., 2006), (DE RIVECOURT et al., 2008)
SURG-TLX	(BERG et al., 2015), (WEIGL et al., 2016)
BSPE	(MARTIN et al., 2016)

Ferramenta	Artigo
JDS	(PEKKARINEN et al., 2013)
ESCAM	(VÁSQUEZ et al., 2016)
GSW	(MOUZÉ-AMADY et al., 2013)
SOI	(AVERTY et al., 2004)
ISA	(WILSON et al., 2017)
ISTAS 21	(ARELLANO et al., 2012)
MCH	(ERASLAN; CAN; ATALAY, 2016)
Tarefa primária	(OTHMAN; ROMLI, 2016)
IWS	(PICKUP et al., 2005)
Raw-NASA-TLX	(WIDYANTI; LARUTAMA, 2016)
SURG-TLX + GOALS	(YU et al., 2016)
TLI	(AVERTY et al., 2004)

APÊNDICE C – Artigos do PB organizados conforme os indicadores fisiológicos utilizados para análise da CTM

Indicador fisiológico	Artigo
Corção	(EAST; BAUER JR; LANNING, 2002), (MOON et al., 2002), (WILSON, 2002), (WILSON; RUSSELL, 2003), (MURAI et al., 2004), (LAHTINEN et al., 2007), (MURAI et al., 2007), (WILSON; CALDWELL; RUSSELL, 2007), (DE RIVECOURT et al., 2008), (HANNULA et al., 2008), (MURAI et al., 2008), (YAO et al., 2008), (GOULD et al., 2009), (HENELIUS et al., 2009), (LIANG et al., 2009), (DEY; MANN, 2010), (LEHRER et al., 2010), (KATAOKA; SASAKI; KANDA, 2011), (SCHULZ et al., 2011), (MOUZÉ-AMADY et al., 2013), (MANSIKKA et al., 2016a), (MANSIKKA et al., 2016b), (MARTIN et al., 2016), (MURAI et al., 2016)
Olhos	(BERGUER; SMITH; CHUNG, 2001), (EAST; BAUER JR; LANNING, 2002), (WILSON, 2002), (WILSON; RUSSELL, 2003), (AHLSTROM; FRIEDMANBERG, 2006), (DI NOCERA; CAMILLI; TERENCEZI, 2007), (WILSON; CALDWELL; RUSSELL, 2007), (DE RIVECOURT et al., 2008), (DEY; MANN, 2010), (KATAOKA; SASAKI; KANDA, 2011), (SCHULZ et al., 2011), (ZHENG et al., 2012a), (FALLAHI et al., 2016), (MANNARU et al., 2016a), (MANNARU et al., 2016b), (OTHMAN; ROMLI, 2016), (ZAWIAH; DAWAL, 2016)
Cérebro	(LING et al., 2001), (EAST; BAUER JR; LANNING, 2002), (MOON et al., 2002), (WILSON, 2002), (WILSON; RUSSELL, 2003), (KERICK; ALLENDER, 2006), (WILSON; CALDWELL; RUSSELL, 2007), (AYAZ et al., 2011), (AYAZ et al., 2012), (HARRISON et al., 2014), (FALLAHI et al., 2016), (ZAWIAH; DAWAL, 2016)
Pele	(BERGUER; SMITH; CHUNG, 2001), (WILSON, 2002), (COLLET et al., 2003), (GOULD et al., 2009)
Respiração	(EAST; BAUER JR; LANNING, 2002), (WILSON; RUSSELL, 2003), (YAO et al., 2008), (KATAOKA; SASAKI; KANDA, 2011)
Temperatura facial	(MURAI et al., 2007), MURAI et al., 2008)
Fala	(HUTTUNEN et al., 2011)
Movimentos do corpo	(WILSON, 2002)
Urina	(KUIJER et al., 2004)

APÊNDICE D – Artigos do PB organizados conforme os países de aplicação das pesquisas

País	Artigo
Estados Unidos	(WEINGER et al., 2000), (BERGUER; SMITH; CHUNG, 2001), (GALSTER et al., 2001), (LING et al., 2001), (METZGER; PARASURAMAN, 2001), (EAST; BAUER JR; LANNING, 2002), (WILSON, 2002), (WILSON; RUSSELL, 2003), (AHLSTROM; FRIEDMAN-BERG, 2006), (KERICK; ALLENDER, 2006), (WARD et al., 2006), (BOULTINGHOUSE et al., 2007), (SALEEM et al., 2007), (WILSON; CALDWELL; RUSSELL, 2007), (LEHRER et al., 2010), (LOPEZ et al., 2010), (ZHENG et al., 2010), (AYAZ et al., 2011), (AYAZ et al., 2012), (MITROPOULOS; MEMARIAN, 2012), (ZHENG et al., 2012b), (DADI et al., 2014), (HARRISON et al., 2014), (LEE et al., 2014), (BERG et al., 2015), (MANNARU et al., 2016a), (MANNARU et al., 2016b), (YU et al., 2016), (SCERBO; BRITT; STEFANIDIS, 2017), (WILSON et al., 2017)
Finlândia	(LAHELMA et al., 2005), (LAHTINEN et al., 2007), (HANNULA et al., 2008), (HENELIUS et al., 2009), (HUTTUNEN et al., 2011), (KARHULA et al., 2013), (PEKKARINEN et al., 2013), (MANSIKKA et al., 2016a), (MANSIKKA et al., 2016b)
Japão	(MURAI et al., 2004), (KAWADA; Ooya, 2005), (MURAI et al., 2007), (MURAI et al., 2008), (KATAOKA; SASAKI; KANDA, 2011), (MURAI et al., 2016)
Irã	(SAFARI et al., 2013), (DARVISHI et al., 2016), (FALLAHI et al., 2016), (MAZLOUMI et al., 2016), (ZARE et al., 2016)
Reino Unido	(PICKUP et al., 2005), (BYRNE et al., 2010), (BYRNE et al., 2013), (WHEELLOCK et al., 2015)
Turquia	(KOCA et al., 2015), (ERASLAN; CAN; ATALAY, 2016), (SÖNMEZ et al., 2016)
Taiwan	(LIANG et al., 2009), (LIN; LU, 2016), (YANG et al., 2012)
Holanda	(KUIJER et al., 2004), (DE RIVECOURT et al., 2008), (ZOER et al., 2011)
França	(COLLET et al., 2003), (AVERTY et al., 2004), (MOUZÉ-AMADY et al., 2013)
Alemanha	(SCHULZ et al., 2011), (MARTIN et al., 2016), (WEIGL et al., 2016)
México	(GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006), (ARELLANO et al., 2012)
Malásia	(OTHMAN; ROMLI, 2016), (ZAWIAH; DAWAL, 2016)
Espanha	(RUBIO VALDEHITA et al., 2010), (RUBIO-VALDEHITA; RODRIGO-TAPIAS, 2017)
Coréia do Sul	(MOON et al., 2002), (PARK; JUNG, 2006)
Canadá	(DEY; MANN, 2010), (ZHENG et al., 2012a)
Suíça	(ELFERING; GREBNER; HALLER, 2012)
Noruega	(GOULD et al., 2009)
Itália	(DI NOCERA; CAMILLI; TERENCEZI, 2007)
Indonésia	(WIDYANTI; LARUTAMA, 2016)
China	(YAO et al., 2008)
Chile	(VÁSQUEZ et al., 2016)

APÊNDICE E – Artigos do PB organizados conforme os grupos estratificados de trabalhadores

Trabalhadores	Artigo
Trabalhadores de saúde	(WEINGER et al., 2000), (BERGUER; SMITH; CHUNG, 2001), (BOULTINGHOUSE et al., 2007), (SALEEM et al., 2007), (BYRNE et al., 2010), (LOPEZ et al., 2010), (ZHENG et al., 2010), (KATAOKA; SASAKI; KANDA, 2011), (SCHULZ et al., 2011), (ZHENG et al., 2012a), (ZHENG et al., 2012b), (BYRNE et al., 2013), (KARHULA et al., 2013), (PEKKARINEN et al., 2013), (LEE et al., 2014), (BERG et al., 2015), (KOCA et al., 2015), (WHEELOCK et al., 2015), (MARTIN et al., 2016), (SÖNMEZ et al., 2016), (VÁSQUEZ et al., 2016), (WEIGL et al., 2016), (YU et al., 2016), (RUBIO-VALDEHITA; RODRIGO-TAPIAS, 2017), (SCERBO; BRITT; STEFANIDIS, 2017)
Pilotos	(LING et al., 2001), (EAST; BAUER JR; LANNING, 2002), (WILSON, 2002), (DI NOCERA; CAMILLI; TERENCE, 2007), (LAHTINEN et al., 2007), (DE RIVECOURT et al., 2008), (HANNULA et al., 2008), (YAO et al., 2008), (LEHRER et al., 2010), (HUTTUNEN et al., 2011), (MANSIKKA et al., 2016a), (MANSIKKA et al., 2016b), (LIN; LU, 2016), (OTHTMAN; ROMLI, 2016)
Controladores de tráfego aéreo	(GALSTER et al., 2001), (METZGER; PARASURAMAN, 2001), (WILSON; RUSSELL, 2003), (AVERTY et al., 2004), (AHLSTROM; FRIEDMAN-BERG, 2006), (AYAZ et al., 2011), (AYAZ et al., 2012), (HARRISON et al., 2014), (FALLAHI et al., 2016), (MAZLOUMI et al., 2016)
Operadores	(MOON et al., 2002), (PARK; JUNG, 2006), (DEY; MANN, 2010), (ARELLANO et al., 2012), (YANG et al., 2012), (ZARE et al., 2016),
Trabalhadores navais	(MURAI et al., 2004), (MURAI et al., 2007), (MURAI et al., 2008), (GOULD et al., 2009), (MURAI et al., 2016)
Motoristas	(COLLET et al., 2003), (KUIJER et al., 2004), (WARD et al., 2006), (ZAWIAH; DAWAL, 2016)
Trabalhadores industriais	(KAWADA; OOYA, 2005), (GONZÁLEZ-MUÑOZ; GUTIÉRREZ-MARTÍNEZ, 2006), (SAFARI et al., 2013)
Operadores de UAS	(MANNARU et al., 2016a), (MANNARU et al., 2016b)
Controladores rodoviários	(PICKUP et al., 2005), (ELFERING; GREBNER; HALLER, 2012)
Não identificado	(LAHELMA et al., 2005), (WILSON; CALDWELL; RUSSELL, 2007)
Vários profissionais	(KERICK; ALLENDER, 2006), (RUBIO VALDEHITA et al., 2010), (ZOER et al., 2011), (MOUZÉ-AMADY et al., 2013), (DADI et al., 2014), (ERASLAN; CAN; ATALAY, 2016)
Outros	(HENELIUS et al., 2009), (LIANG et al., 2009), (MITROPOULOS; MEMARIAN, 2012), (DARVISHI et al., 2016), (WIDYANTI; LARUTAMA, 2016), (WILSON et al., 2017)