

2 人员工作负荷

工作负荷已被定义为一系列任务需求、努力、活动或目标实现（Gartner 和 Murphy, 1979）。任务需求（任务负荷）是要实现的目标、执行任务所允许的时间以及完成任务的绩效水平，影响付出努力的因素包括所提供的信息和设备、任务环境、被试者的技能和经验、所采取的策略以及对情境的情绪反应。

这些定义在任务负荷与工作负荷之间提供了可测量的连接。例如，直升机飞行员在保持持续悬停的状态下，其工作负荷值可为 70（分数范围是 0 ~ 100）。这个例子源自加利福尼亚州伍德兰希尔斯 Perceptronics 公司副总裁阿扎德·马德尼对一项陆军的研究。对于保持持续悬停并同时坦克瞄准任务，其工作负荷也可能是 70。这种差异是由于飞行员在仅完成悬停任务（即没有水平或垂直移动）时，对自身赋予了严格的绩效要求。但在瞄准坦克时，为了将工作负荷保持在可控水平，就放松了对悬停的绩效要求，即可以出现几英尺的移动。这些定义使任务负荷与工作负荷能够在实际工作情况下得到解释。这些定义还可以对任务负荷与工作负荷进行逻辑上的正确分析。实际上，工作负荷永远不会超过 100%（一个人不能做不可能的事情）。任何允许工作负荷超过 100% 的理论或报告结果都是不现实的。但是，根据定义，任务负荷可能超过 100%。一个例子就是测量“所需时间 / 可用时间”。根据拟定的定义，如果绩效要求设置得太高（由此增加了所需的时间）或可用时间设置得太低，则此任务负荷的测量值就可能超过 100%。总之，即使任务负荷确实超过 100%，工作负荷也不能超过 100%。

工作负荷可通过独立任务绩效（见 2.1）或次任务绩效（见 2.2）进行测量，也可通过主观测量（见 2.3）或数字模拟测量（见 2.4）。工作负荷的生理测量也已得到应用，此处不进行讨论。Caldwell 等（1994）提供了一份有关采用生理指标进行工作负荷测量的有价值的参考资料。2.5 讨论了工作负荷与绩效之间的分离。Wierville 等（1979）、O'Donnell 和 Eggemeier（1986）给出了选用工作负荷测量技术的指导原则。脑力负荷内容可见 Moray（1982）。Wierville 和 Eggemeier（1993）列出了测量方法的 4 个关键方面：诊断性、通用敏感性、可迁移性和实施要求。通用性指导如图 2.1 所示，

其中底部分支可以重复模拟。

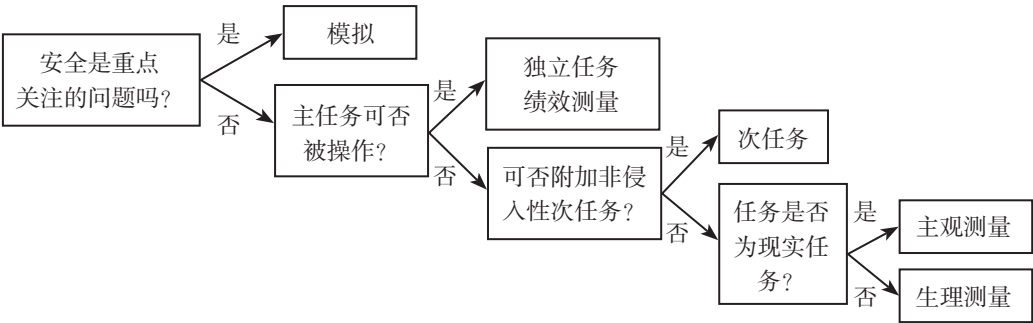


图 2.1 工作负荷测量技术的通用性指导

Stanton 等（2005）对工作负荷测量进行了类似分类，包括主任务和次任务的绩效测量、生理测量和主观测量。此外，Funke 等（2012）提出了团队工作负荷的理论以及主观、绩效、生理和策略转换测量。Sharples 和 Megaw（2015）对工作负荷测量技术进行了以下分类：

- （1）分析技术 比较分析、数学模型、专家意见、任务分析方法和模拟模型。
- （2）实验技术 主任务绩效、次任务绩效。
- （3）心理生理技术 心脏活动、大脑活动、皮肤电活动、眼功能、体液分析、肌肉和运动分析。
- （4）主观 / 操作员意见技术 单维量表、多维量表、相对判断、即时判断、访谈和观察。

最后，Matthews 和 Reinerman-Jones（2017）出版了一本关于工作负荷评估的著作。

原书参考文献

Caldwell, J.A, Wilson, G.F., and Cetinguc, M. Psychophysiological Assessment Methods (AGARD-AR-324). Neuilly-Sur-Seine, France: Advisory Group For Aerospace Research and Development, May 1994.

Funke, G.J., Knott, B.A., Salas, E., Pavlas, D., and Strang, A.J. Conceptualization and measurement of team workload: A critical need. Human Factors 54(1): 36-51, 2012.

Gartner, W.B., and Murph M.R. Concepts of workload. In B.O. Hartman and R.E. McKenzie (Eds.) Survey of Methods to Assess Workload (AGARD-AG-246). Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development, 1979.

Matthews, G., and Reinerman-Joes, L.E. Workload Assessment: How to Diagnose Workload Issues and



- Enhance Performance. Santa Monica, California: Human Factors and Ergonomics Society, 2017.
- Mora N. Subjective mental workload. *Human Factors* 24(1): 25-40, 1982.
- O' Donnell, R.D., and Eggemeier F.T Workload assessment methodology. In K.R. Boff, L. Kaufman, and Thomas (Eds.) *Handbook of Perception and Human Performance*. New York, NY: John Wiley and Sons, 1986.
- Sharples, S., and Megaw, T. Definition and measurement of human workload. In J.R. Wilson and S. Sharples (Eds.) *Evaluation of Human Work* (p. 515-548). Boca Raton: CRC Press, 2015.
- Stanton, N.A., Salmon, P.M., Walker, G.H., Barber, C., and Jenkins, *Human Factors Methods: A Practical Guide for Engineering and Design*. Aldershot, United Kingdom: Gower ebook, December 2005.
- Wierwille, W.W., and Eggemeier, F.T. Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors* 35(2): 263-281, 1993.
- Wierwille, W.W., Williges, R.C., and Schifflett, S.G. Aircrew workload assessment techniques. In B.O. Hartman and R.E. McKenzie (Eds.) *Survey of Methods to Assess Workload (AGARD-AG-246)*. Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research and Development, 1979.

2.1 工作负荷的独立绩效测量方法

绩效已被用于工作负荷测量。这些测量方法假设，随着工作负荷的增加，额外的加工处理需求将使绩效降低。O'Donnell 和 Eggemeier (1986) 定义了采用绩效进行工作负荷测量的 4 个相关问题：①负荷过轻可能会提高绩效；②负荷过重可能会导致地板效应；③可能出现信息加工策略、培训或经验的混合效应；④测量方法只适用于特定任务，而不适用于其他任务。

Meshkati 等 (1990) 指出，当任务复杂或具有多个维度时，需要采用多种任务测量方法。此外，任务测量可能具有干扰性，也可能受到工作负荷之外其他因素（如动机和学习）的影响。

独立任务绩效测量包括机组工作负荷评估（见 2.1.1）、控制动作 / 单位时间（见 2.1.2）、扫视持续时间和频率（见 2.1.3）、负荷压力（见 2.1.4）、观测工作负荷区域（见 2.1.5）、信息增益率（见 2.1.6）、相对条件效率（见 2.1.7）、速度压力（见 2.1.8）、任务难度指数（见 2.1.9）和时间裕度（见 2.1.10）。

原书参考文献

- Meshkati, N., Hancock, P.A., and Rahimi, M. Techniques in mental workload assessment. In J.R. Wilson and E.N. Corlett (Eds.) *Evaluation of Human Work. A Practical Ergonomics Methodology* (p. 605-

627). New York: Taylor & Francis Group, 1990.

O'Donnell, R.D., and Eggemeie F.T. Workload assessment methodology. In KR. Boff, L. Kaufman, and Thomas (Eds.) Handbook of Perception and Human Performance (p.42-1-42-49). New York: Wiley and Sons, 1986.

2.1.1 机组工作负荷评估系统

概述：机组工作负荷评估系统（AWAS）是英国宇航公司开发的一种时间轴分析软件，用于工作负荷预测。AWAS 需要 3 个输入条件：①飞行过程中飞行员任务的逐秒描述；②对威肯斯多资源理论加工通道中每个通道的需求；③同时性需求对单个通道的影响（Davies et al, 1995）。

优势和局限性：Davies 等（1995）报告了 AWAS 工作负荷预测与听觉辨别次任务中错误数量的相关达到 0.904。该项研究的被试者为 2 名经验丰富的飞行员，研究用的平台为一架海上勇士模拟器。

数据需求：飞行任务的逐秒时间表、对每个信息处理渠道的需求以及同时性需求的影响。

阈值：未说明。

原书参考文献

Davies, A.K., Tomosz A., Hicks, M.R., and White, J. AWAS (Aircrew Workload Assessment System): Issues of theory, implementation, and validation. In R. Fuller, N. Johnston, and N. McDonald (Eds.) Human Factors in Aviation Operations. Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP), vol. 3, Chapter 48, 1995.

2.1.2 控制动作 / 单位时间

概述：控制动作 / 单位时间是指在测量过程中，操作员对每个控制单元实施的控制输入总数除以总时间。

优势和局限性：Wierville 和 Connor（1983）指出，这一测量技术对工作负荷具有敏感性。其具体测量值是在一台动基飞行模拟器上对控制器（副翼、升降舵和方向舵）每秒的控制输入次数。对工作负荷的操纵是通过对俯仰稳定性、气流扰动、侧风方向和速度的控制实现的。

Porterfield（1997）采用类似方法，将空中交通管制员地空通信持续时间作为工作负荷测量指标。他报告了持续时间与空中交通工作负荷输入技术（ATWIT）之间

的显著相关性 (0.88)，这是一种基于飞行员工作负荷客观 / 主观评估技术 (POSWAT) 的工作负荷评定。

Zeitlin (1995) 基于每分钟制动次数加上车辆速度的对数，建立了驾驶员工作负荷指数。该指数对道路类型 (农村、城市高速公路) 的差异具有敏感性。

数据要求：控制动作必须得到明确定义。

阈值：未说明。

原书参考文献

- Porterfield, D.H. Evaluating controller communication time as a measure of workload. *The International Journal of Aviation Psychology* 7(2): 171-182, 1997.
- Wierwille, W.W., and Connor, S.A. Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors* 25(1): 1-16, 1983.
- Zeitlin, L.R. Estimates of driver mental workload: A long term field trial of two subsidiary tasks. *Human Factors* 37(3): 611-621, 1995.

2.1.3 扫视持续时间和频率

概述：对视觉显示器的扫视持续时间和频率已被用作视觉工作负荷的测量指标。扫视的持续时间越长和 (或) 频率越高，视觉工作负荷就越高。

优势和局限性：Fairclough 等 (1993) 采用扫视持续时间计算驾驶员查看导航信息 [纸质地图与液晶显示器 (LCD) 文本显示]、前方道路、后视镜、仪表板、左后视镜、右后视镜、左窗和右窗的时间百分比。这些数据是在一辆车上采集到的，该车行驶在英国道路上，车上装有数据采集设备。作者得出的结论是，这一“测量技术被证明具有足够的敏感性，可以对纸质地图和 LCD/ 文本显示进行区分，并可检测视觉场景中其他区域的相关变化”。但作者也提醒，扫视持续时间的缩短可能反映了驾驶员对纸质地图数量和易读性的应对策略。

作者还使用扫视持续时间和频率对两种车内路线导航系统进行比较。数据采集于 23 名德国被试者，实验中驾驶的车辆上装有数据采集设备。结果显示，“随着对导航显示屏扫视频率的增加，对仪表板、后视镜和左后视镜的扫视次数均显著减少”。基于这些结果，作者得出结论，“扫视持续时间表现出对信息更新的难度更为敏感。扫视频率代表了‘视觉检测行为’的数量”。

Wierwille (1993) 基于对驾驶员视觉行为的回顾得出结论，这种行为是“相对一致的”。

数据要求：记录被试者的眼睛位置。

阈值：0 ~ 无穷大。

原书参考文献

Fairclough, S.H., Ashby, M.C., and Parkes, A. 岛 In-vehicle displays, visual workload and visibility evaluation. In A.G. Gale, I.D. Brown, C.M. Haslegrave, H.W. Kruysse, and Taylor (Eds). Vision in Vehicles IV (p.245-254). Amsterdam: North-Holland, 1993.

Wierwille, W.W. An initial model of visual sampling of in-car displays and controls. In A.G.Gale, I.D. Brown, C.M. Haslegrave, H.W. Kruysse, and S.P. Taylor (Eds.) Vision in Vehicles-IV (p. 271-280). Amsterdam: North-Holland, 1993.

2.1.4 负荷压力

概述：负荷压力指在任务过程中增加必须监测的信号源数量而产生的压力（Chiles 和 Alluisi, 1979）。

优势和局限性：负荷压力会影响任务执行中出现的错误数量。在非实验室环境下，通过增加负荷压力以测量操作员的工作负荷，可能会有困难。

数据要求：必须明确定义信号源。

阈值：未说明。

原书参考文献

Chiles, W.D., and Alluisi, E.A. On the specification of operator or occupational workload with performance-measurement methods. Human Factors 21(5): 515-528, 1979.

2.1.5 观测工作负荷区域

概述：Laudeman 和 Palmer（1995）开发了观测工作负荷区域，以测量飞机驾驶舱的工作负荷。这一测量技术不是基于理论，而是基于工作负荷与任务限制之间的逻辑联系。用他们的话说：“驾驶舱里的每项任务都存在一个可被客观定义的机会窗口。驾驶舱里观测到的任务工作负荷，可赋予这样一种操作性定义，即由领域专家对最大的任务重要性做出评定。在任务机会窗口期，任务重要性作为任务重要性与时间的线性函数而增大。当任务机会窗口重叠由此导致任务函数重叠时，任务函数通过相加的方式得以组合，从而产生一种复合函数，其中包括两个或多个任务函数的观测工作负荷效应。我们将这些由两个或多个任务函数组成的复合函数，称为观测工作负荷函数。



我们提出从观测工作负荷函数中提取而来的重属测量构成一个区域，即所谓的观测工作负荷区域”。

优势和局限性：Laudeman 和 Palmer（1995）报告，副机长的工作负荷管理评定与观测工作负荷区域之间存在显著相关性。这种相关性是基于 18 名双人机组人员驾驶高仿真度飞行模拟器得出的。小的观测工作负荷区域与高的工作负荷管理评定相关。错误率越高的工作人员，观测工作负荷区域越大。该项技术需要专家提供任务重要性评定，同时也需要明确定义任务的开始和结束。

阈值：未说明。

原书参考文献

Laudeman, IV., and Palmer, E.A. Quantitative measurement of observed workload in the analysis of aircrew performance. *International Journal of Aviation Psychology* 5(2): 187-197, 1995.

2.1.6 信息增益率

概述：信息增益率测量技术基于希克定律，该定律指出反应时间（RT）是传输信息量的线性函数，（Ht）： $RT=a+B(Ht)$ （Chiles 和 Alluisi，1979）。

优势和局限性：希克定律已在大量条件下得到了验证。然而，它仅限于一些分离的任务，除非该任务是正常程序的一部分，否则可能具有干扰性，尤其是在非实验室环境中。

数据要求：从反应时间估计信息增益率。通常使用机械秒表或软件时钟对时间进行采集。机械秒表需要频繁校准（如在每次试验之前），软件时钟需要稳定和恒定的电源。

阈值：未说明。

原书参考文献

Chiles, W.D., and Alluisi, E.A. On the specification of operator or occupational workload with performance-measurement methods. *Human Factors* 21(5): 515-528, 1979.

2.1.7 相对条件效率

概述：Paas 和 van Merriënboer（1993）将工作负荷评定与任务绩效指标相结合，以计算相对条件效率。评分从 1 分（非常非常低的脑力负荷）~ 9 分（非常非常高的

脑力负荷) 不等。绩效是以对测试问题正确答案的百分比进行测量的。相对条件效率被计算为“与假定效率为零的直线的垂直距离”。

优势和局限性：不同工作条件下的效率得分显著不同。

数据要求：未说明。

阈值：未说明。

原书参考文献

Paas, F.G.W.C., and van Merriënboer, J. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors* 35(4): 737-743, 1993.

2.1.8 速度压力

概述：速度压力是指通过提高一个或多个信号源的信号呈现速率而产生的压力。

优势和局限性：速度压力会影响错误的数量以及完成任务的时间（Conrad, 1956; Knowles et al, 1953）。在非实验室任务中施加速度压力可能有困难。

数据要求：该任务必须包括离散信号，其呈现速率可得到控制。

阈值：未说明。

原书参考文献

Conrad, R. The timing of signals in skill. *Journal of Experimental Psychology* 51: 365-370, 1956.

Knowles, W.B., Garvey, W.D., and Newlin, E.P. The effect of speed and load on display-control relationships. *Journal of Experimental Psychology* 46: 65-75, 1953.

2.1.9 任务难度指数

概述：任务难度指数由 Wickens 和 Yeh（1985）开发，用于对典型实验室任务的工作负荷进行分类。该指数有 4 个维度：

1. 对刺激的熟悉程度

0= 字母；

1= 空间点图案，跟踪光标。

2. 并行任务数量

0= 单任务；

1= 双重任务。



3. 任务难度

0= 记忆组块为 2;

1= 记忆组块为 4, 二阶追踪, 延迟回忆。

4. 资源竞争

0= 无竞争;

1= 刺激模式(视觉、听觉)或中央处理模式(空间、言语)的竞争(Gopher 和 Braune, 1984)。

任务难度指数是上述 4 个维度中每个维度的得分总和。

优势和局限性: Gopher 和 Braune (1984) 报告, 任务难度指数与工作负荷主观测量之间存在显著的正相关(0.93)。其数据基于 55 名男性被试者在 21 项任务中的反应, 这些任务包括斯滕伯格记忆任务、隐藏模式、卡片旋转追踪、迷宫追踪、延迟数字回忆和双耳分听。

数据要求: 需要用户在上述 4 个维度上对所要完成的任务做出描述。

阈值: 在 0 和 4 之间变化。

原书参考文献

Gopher, D., and Braune, R. On the psychophysics of workload: Why bother with subjective measures? Human Factors 26(5): 519-532, 1984.

Wickens, C.D., and Yeh, Y. POCs and performance decrements: A reply to Kantowitz and Weldon. Human Factors 27: 549-554, 1985.

2.1.10 时间裕度

概述: 在回顾了当前飞行工作负荷的测量技术后, Gawron 等(1989)指出了 5 个主要的不足之处: ①主观评定表现出很大的个体差异, 远远超出了可归因于经验和能力的差异; ②大多数测量技术并不全面, 仅评估了工作负荷的单个维度; ③许多工作负荷测量由于要求任务响应、主观评定或使用电极等而具有侵入性; ④一些测量技术在高压力状态下会让被试者感到难以理解, 如在高工作负荷环境中, 评定的含义会被遗忘, 因此飞行员给出的值会低于实际值; ⑤被试者会错误感知要执行的任务数量, 并提供错误的低工作负荷度量。Gawron 表明了工作负荷测量的目的在于识别潜在的危险状况。设计不良、程序不当、培训不足或接近灾难的状况都可能引发潜在危险。对某种情况危险状态的最客观测量指标, 是若不采取控制行动便将导致飞机被摧毁所经历的时间。这些时间包括: 直到飞机撞击的时间、直到飞机受到过度应力而断裂的

时间，以及直到飞机燃料耗尽的时间。

优势和局限性：工作负荷的时间裕度测量是定量的、客观的，与绩效直接相关，可以在任务中有针对性地使用。例如，地空导弹摧毁飞机之前的时间是空地突击任务中的一个很好的测量指标。此外，时间裕度易于从飞机性能的测量中计算出来。最后，在任何长度的间隔上均可计算出相应的时间，由此进行逐个间隔的工作负荷比较。

数据要求：只要有飞机性能数据，这种方法就很有用。

阈值：最小值为 0，最大值为无穷大。

原书参考文献

Gawron, V.J., Schiflett, S.G., and Miller, J.C. Measures of in-flight workload. In R.S. Jensen (Ed.) *Aviation Psychology* (p. 240-287). London: Gower, 1989.

2.2 工作负荷的次任务绩效测量

次任务测量法是用于衡量工作负荷的最常用技术之一。这种方法需要被试者在特定要求下完成主任务，同时利用空闲注意力或能力完成次任务。在此情况下，工作负荷的操作定义就是次任务绩效的下降程度。

优势：次任务测量法有几个优点。第一，它对被试者的能力变化较为敏感，并可区分单任务绩效所无法区分的几种设备布局（Slocum et al, 1971）；第二，它可提供因压力导致任务受损的一个敏感性指标；第三，它可提供用于比较不同任务的共同尺度。

局限性：次任务测量法的一个主要缺点是对主任务绩效产生干扰（Williges 和 Wierwille, 1979）。但是，Vidulich（1989a）从两个实验中得出结论，不干扰主任务绩效的次任务，对主任务难度并不敏感；Vidulich（1989b）认为，附加任务的敏感性与其对主任务的干扰性直接相关。

Damos（1993）分析了评估单双重任务绩效的 14 项研究的结果，得到的结论是：“单任务和多任务测量的效应量在统计学上均不为零，多任务测量的效应量高于相应的单任务测量。但相应的预测效度低”。Poulton（1965）指出，对不同敏感度的绩效测试结果进行比较是困难的。为了解决这个问题，Colle 等（1988 年）建立了双重权衡曲线以平衡不同次任务的绩效水平。在这种方法中，“两个不同的次任务与相同的主任务搭配。在每个次任务与主任务搭配时，会得到一条权衡曲线”。

另一个潜在的缺点是，被试者在完成次任务时可能会使用不同的策略。例如，