Gawron, V.J. (2000)

Human Performance Measures Handbook.
Mahwaw, NJ: Lawrence Erlbaum Associates

3.2.29 Subjective Workload Assessment Technique

General description - The Subjective Workload Assessment Technique (SWAT) combines ratings of three different scales (see Table 25) to produce an interval scale of mental workload. These scales are: (1) time load, which reflects the amount of spare time available in planning, executing, and monitoring a task; (2) mental effort load, which assesses how much conscious mental effort and planning are required to perform a

task; and (3) psychological stress load, which measures the amounts of risk, confusion, frustration, and anxiety associated with task performance. A more complete description is given in Reid and Nygren (1988). A description of the initial conjoint measurement model for SWAT is described in Nygren (1982, 1983).

TABLE 25.
SWAT Scales

Time Load

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3. Almost never have spare time. Interruptions or overlap among activities are frequent or occur all the time.

Mental Effort Load

1. Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.
2. Moderate conscious mental effort or concentration required. Complexity of activity is moderatelyhigh due to uncertainty, unpredictability, or unfamiliarity. Considerable attention required.
3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

Psychological Stress Load

1. Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2. Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required. (Potter and Bressler, 1989, pp. 12-14).

Strengths and limitations - SWAT has been found to be a valid (Albery, Repperger, Reid, Goodyear, and Roe, 1987; Haworth, Bivens, and Shively, 1986; Masline, 1986; Reid, Shingledecker, and Eggemeier, 1981; Reid, Shingledecker, Nygren, and Eggemeier, 1981; Vidulich and Tsang, 1987; Warr, Colle, and Reid, M.G, 1986), sensitive (Eggemeier, Crabtree, Zingg, Reid, and Shingledecker, 1982), reliable (Corwin, Sandry-Garza, Biferno, Boucek, Logan, Jonsson, and Metalis, 1989; Gidcomb, 1985), and relatively unobtrusive (Crabtree, Bateman, and Acton, 1984; Courtright and Kuperman, 1984; Eggemeier, 1988) measure of workload. Further, SWAT ratings are not affected by delays of up to 30 minutes (Eggemeier, Crabtree, and LaPoint, 1983), nor by intervening tasks of all but difficult tasks (Eggemeier, Melville, and Crabtree, 1984; Lutmer and Eggemeier, 1990). Moroney, Biers, and Eggemeier (1995) concur. Also, Eggleston (1984) found a significant correlation between projected SWAT ratings made during system concept evaluation and those made during ground-based simulation of the same system.

Warr (1986) reported that SWAT ratings were less variable than modified Cooper-Harper ratings. Kilmer, et al. (1988) reported that SWAT was more sensitive to changes in difficulty of a tracking task than the modified Cooper-Harper Rating Scale was. Finally, Nygren (1991) stated that SWAT provides a good cognitive model of workload, sensitive to individual differences.

SWAT has been used in diverse environments, for example, test aircraft (Papa and Stoliker, 1988), a high-G centrifuge (Albery, Ward, and Gill, 1985), command, control, and communications centers (Crabtree, Bateman, and Acton, 1984), nuclear power plants (Beare and Dorris, 1984), domed flight simulators (Reid, Eggemeier, and Shingledecker, 1982; Skelly and Simons, 1983), tank simulators (Whitaker, Peters, and Garinther, 1989); and the benign laboratory setting (Graham and Cook, 1984; Kilmer, Knapp, Burdsal, Borresen, Bateman, and Malzahn (1988)). In the laboratory, SWAT has been used to

assess the workload associated with critical tracking and communication tasks (Reid, Shingledecker, and Eggemeier, 1981), memory tasks (Eggemeier, Crabtree, Zingg, Reid, and Shingledecker, 1982; Eggemeier and Stadler, 1984; Potter and Acton, 1985), and monitoring tasks (Notestine, 1984). Hancock and Caird (1993) reported significant increases in SWAT rating as the shrink rate of the target decreased and as the number of steps from the cursor to the target increased.

Usage in simulated flight has also been extensive (Haworth, Bivens, and Shively, 1986; Nataupsky and Abbott, 1987; Schick and Hann, 1987; Skelly and Purvis, 1985; Skelly, Reid, and Wilson, 1983; Thiessen, Lay, and Stern, 1986; Ward and Hassoun, 1990). For example, Bateman and Thompson (1986) reported that SWAT ratings increased as task difficulty increased. Their data were collected in an aircraft simulator during a tactical mission. Vickroy (1988), also using an aircraft simulator, reported that SWAT ratings increased as the amount of air turbulence increased. Fracker and Davis (1990) reported significant increases in SWAT as the number of simulated enemy aircraft increased from 1 to 3. Hancock, Williams, Manning, and Miyake (1995) reported that SWAT was highly correlated with the difficulty of a simulated flight task. However, See and Vidulich (1997) reported significant effects of target type and threat status on SWAT scores in a combat aircraft simulator. There were no significant correlations of SWAT with overall workload but two subscales correlated with peak workload (effort, $r = +0.78$; stress, $r = +0.76$).

Usage in actual flight has been extensive. For example, Pollack (1985) used SWAT to assess differences in workload between flight segments. She reported that C-130 pilots had the highest SWAT scores during the approach segment of the mission. She also reported higher SWAT ratings during the preflight segments of tactical, rather than proficiency, missions. Haskell and Reid (1987) found significant difference in SWAT ratings between flight maneuvers and also between successfully completed maneuvers and those that were not successfully completed. Gawron, et al. (1988) analyzed SWAT ratings made by the pilot and copilot four times during each familiarization and data flight: (1) during the taxi out to the runway, (2) just prior to a simulated drop, (3) just after a simulated drop, and (4) during the taxi back to the hangar. There were significant segments effects. Specifically, SWAT ratings were highest before the drop and lowest for preflight. The ratings during postdrop and postflight were both moderate.

Experience with SWAT has not been all positive, however. For example, Boyd (1983) reported that there were significant positive correlations between the three workload scales in a text-editing task. This suggests that the three dimensions of workload are not independent. This, in turn, poses a problem for use of conjoint measurement techniques. Derrick (1983) and Hart (1986) suggest that three scales may not be adequate for assessing workload. Further, Battiste and Bortolussi (1988) reported a test/retest correlation of +0.751 but also stated that, of the 144 SWAT ratings reported during a simulated B-727 flight, 59 were zero.

Corwin (1989) reported no difference between in-flight and post-flight ratings of SWAT in only two of three flight conditions. Gidcomb (1985) reported casual card sorts and urged emphasizing the importance of the card sort to SWAT raters. A computerized version of the traditional card sort was developed at the Air Force School of Aerospace Medicine. This version eliminates the tedium and dramatically reduces the time to complete the SWAT card sort. Haworth, Bivens, and Shively (1986) reported that,

although the SWAT was able to discriminate control configuration conditions in a single-pilot configuration, it could not discriminate these same conditions in a pilot/copilot configuration. Wilson, Hughes, and Hassoun (1990) reported no significant differences in SWAT ratings among display formats, in contrast to pilot comments. van de Graaff (1987) reported considerable (60 points) intersubject variability in SWAT ratings during an in-flight approach task. Hill, Iavecchia, Byers, Bittner, Zaklad, and Christ (1992) reported that SWAT was not as sensitive to workload or as accepted by Army operators as NASA TLX and the Overall Workload Scale.

Vidulich and Tsang (1986) reported that SWAT failed to detect resource competition effects in dual-task performance of tracking and a directional transformation task. Vidulich (1991) reported test-retest reliability of +0.606 in SWAT ratings for tracking, choice RT, and Sternberg tasks. In addition, Rueb, Vidulich, and Hassoun (1992) reported that only one of three difficult simulated aerial refueling missions had SWAT scores above the 40 redline.

Vidulich and Pandit (1987) concluded that SWAT was not an effective measure of individual differences. This conclusion was based on no significant correlation of SWAT with any of the scales on the Jenkins Activity Survey, Rotter's Locus of Control, the Cognitive Failures Questionnaire, the Cognitive Interference Questionnaire, the Thought Occurrence Questionnaire, the California Q-sort, and the Myers-Briggs Type Indicator.

Data requirements - SWAT requires two steps to use: scale development and event scoring. Scale development requires subjects to rank, from lowest to highest workload, 27 combinations of three levels of the three workload subscales. The levels of each subscale are presented in Table 24. Reid, Eggemeier, and Nygren (1982) describe their individual differences approach to scale development. Programs to calculate the SWAT score for every combination of ratings on the three subscales are available from the Air Force Research Laboratory at Wright-Patterson Air Force Base. A user's manual is also available from the same source.

During event scoring, the subject is asked to provide a rating (1, 2, 3) for each subscale. The experimenter then maps the set of ratings to the SWAT score (1 to 100) calculated during the scale development step. Haskell and Reid (1987) suggests that the tasks to be rated be meaningful to the subjects and, further, that the ratings not interfere with performance of the task. Acton and Colle (1984) reported that the order in which the subscale ratings are presented does not affect the SWAT score. However, it is suggested that the order remain constant to minimize confusion. Eggleston and Quinn (1984) recommended developing a detailed system and operating environment description for prospective ratings.

Thresholds - Minimum value is 0, maximum value is 100. High workload is associated with the maximum value. In addition, ratings of the time, effort, and stress scales may be individually examined as workload components (Eggemeier, McGhee, and Reid, 1983).

Sources -

Acton, W.H. and Colle, H. The effect of task type and stimulus pacing rate on subjective mental workload ratings. Proceedings of the IEEE 1984 National Aerospace and Electronics Conference (pp. 818-823). Dayton, OH: IEEE; 1984.

Albery, W., Repperger, D., Reid, G., Goodyear, C., and Roe, M. Effect of noise on a dual task: subjective and objective workload correlates. Proceedings of the National Aerospace and Electronics Conference. Dayton, OH: IEEE; 1987.

Albery, W.B., Ward, S.L., and Gill, R.T. Effect of acceleration stress on human workload (Technical Report AMRL-TR-85-039). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory; May 1985.

Bateman, R.P and Thompson, M.W. Correlation of predicted workload with actual workload using the subjective workload assessment technique. Proceedings of the SAE AeroTech Conference; 1986.

Battiste, V. and Bortolussi, M.R. Transport pilot workload: A comparison of two subjective techniques. Proceedings of the Human Factors Society 32nd Annual Meeting. 150-154; 1988.

Beare, A. and Dorris, R. The effects of supervisor experience and the presence of a shift technical advisor on the performance of two-man crews in a nuclear power plant simulator. Proceedings of the Human Factors Society 28th Annual Meeting. 242-246. Santa Monica, CA: Human Factors Society; 1984.

Boyd, S.P. Assessing the validity of SWAT as a workload measurement instrument. Proceedings of the Human Factors Society 27th Annual Meeting. 124-128; 1983.

Corwin, W.H. In-flight and post-flight assessment of pilot workload in commercial transport aircraft using SWAT. Proceedings of the Fifth Symposium on Aviation Psychology. 808-813; 1989.

Corwin, W.H., Sandry-Garza, D.L., Biferno, M.H., Boucek, G.P., Logan, A.L., Jonsson, J.E., and Metalis, S.A. Assessment of crew workload measurement methods, techniques, and procedures. Volume I - Process methods and results (WRDC-TR-89-7006). Wright-Patterson Air Force Base, OH; September 1989.

Crabtree, M.A. Bateman, R.P., and Acton, W.H. Benefits of using objective and subjective workload measures. Proceedings of the Human Factors Society 28th Annual Meeting (pp. 950-953). Santa Monica, CA: Human Factors Society; 1984.

Courtright J.F. and Kuperman, G. Use of SWAT in USAF system T&E. Proceedings of the Human Factors Society 28th Annual Meeting (pp. 700-703). Santa Monica, CA: Human Factors Society; 1984.

Derrick W.L. Examination of workload measures with subjective task clusters. Proceedings of the Human Factors Society 27th Annual Meeting (pp. 134-138). Santa Monica, CA: Human Factors Society; 1983.

Eggemeier, F.T. Properties of workload assessment techniques. In P.A. Hancock and N. Meshtaki (Eds.) Human mental workload (pp. 41-62). Amsterdam: North-Holland; 1988.

Eggemeier, F.T., Crabtree, M.S., and LaPoint, P. The effect of delayed report on subjective ratings of mental workload. Proceedings of the Human Factors Society 27th Annual Meeting (pp. 139-143). Santa Monica, CA: Human Factors Society; 1983.

Eggemeier, F.T., Crabtree, M.S., Zingg, J.J., Reid, G.B., and Shingledecker, C.A. Subjective workload assessment in a memory update task. Proceedings of the Human Factors Society 26th Annual Meeting, Santa Monica, CA: Human Factors Society. 643-647; 1982.

Eggemeier, F.T., McGhee, J.Z., and Reid, G.B. The effects of variations in task loading on subjective workload scales. Proceedings of the IEEE 1983 National Aerospace and Electronics Conference (pp. 1099-1106). Dayton, OH: IEEE; 1983.

Eggemeier, F.T., Melville, B., and Crabtree, M. The effect of intervening task performance on subjective workload ratings. Proceedings of the Human Factors

Society 28th Annual Meetings (pp. 954-958). Santa Monica, CA: Human Factors Society; 1984.

Eggemeier, F.T. and Stadler, M. Subjective workload assessment in a spatial memory task. Proceedings of the Human Factors Society 28th Annual Meeting (pp. 680-684). Santa Monica, CA: Human Factors Society; 1984.

Eggleston, R.G. A comparison of projected and measured workload ratings using the subjective workload assessment technique (SWAT). Proceedings of the National Aerospace and Electronics Conference, Volume 2, 827-831; 1984.

Eggleston, R.G. and Quinn, T.J. A preliminary evaluation of a projective workload assessment procedure. Proceedings of the Human Factors Society 28th Annual Meeting (pp. 695-699). Santa Monica, CA: Human Factors Society; 1984.

Fracker, M.L. and Davis, S.A. Measuring operator situation awareness and mental workload. Proceedings of the Fifth Mid-Central Ergonomics/Human Factors Conference, Dayton, OH, 23-25 May 1990.

Gawron, V.J., Schiflett, S., Miller, J., Ball, J., Slater, T., Parker, F., Lloyd, M., Travale, D., and Spicuzza, R.J. The effect of pyridostigmine bromide on inflight aircrew performance (USAFSAM-TR-87-24). Brooks Air Force Base, TX: School of Aerospace Medicine; January 1988.

Gidcomb, C. Survey of SWAT use in flight test (BDM/A-85-0630-7R.) Albuquerque, NM: BDM Corporation; 1985.

Graham, C.H. and Cook, M.R. Effects of pyridostigmine on psychomotor and visual performance (TR-84-052); September 1984.

Hancock, P.A. and Caird, J.K. Experimental evaluation of a model of mental workload. Human Factors. 35(3), 413-419; 1993.

Hancock, P.A., Williams, G., Manning, C.M., and Miyake, S. Influence of task demand characteristics on workload and performance. International Journal of Aviation Psychology. 5(1), 63-86, 1995.

Hart, S.G. Theory and measurement of human workload. In J. Seidner (Ed.) Human productivity enhancement, Vol. 1 (pp. 396-455). New York: Praeger; 1986.

Haskell, B.E., and Reid, G.B.The subjective perception of workload in low-time private pilots: A preliminary study. Aviation, Space, and Environmental Medicine. 58, 1230-1232; 1987.

Haworth, L.A., Bivens, C.C., and Shively, R.J. An investigation of single-piloted advanced cockpit and control configuration for nap-of-the-earth helicopter mission tasks. Proceedings of the 42nd Annual Forum of the American Helicopter Society. 657-671; 1986.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklad, A.L., and Christ, R.E. Comparison of four subjective workload rating scales. Human Factors. 34: 429-439; 1992.

Kilmer, K.J., Knapp, R., Burdsal, C., Borresen, R., Bateman, R.P., and Malzahn, D. A comparison of the SWAT and modified Cooper-Harper scales. Proceedings of the Human Factors 32nd Annual Meeting. 155-159; 1988.

Lutmer, P.A. and Eggemeier, F.T. The effect of intervening task performance and multiple ratings on subjective ratings of mental workload. Paper presented at the 5th Mid-central Ergonomics Conference, University of Dayton, Dayton, OH; 1990.

Masline, P.J. A comparison of the sensitivity of interval scale psychometric techniques in the assessment of subjective mental workload. Unpublished masters thesis, University of Dayton, Dayton, OH; 1986.

Moroney, W.F., Biers, D.W., and Eggemeier, F.T. Some measurement and methodological considerations in the application of subjective workload measurement techniques. International Journal of Aviation Psychology. 5(1), 87-106, 1995.

Nataupsky, M. and Abbott, T.S. Comparison of workload measures on computer-generated primary flight displays. Proceedings of the Human Factors Society 31st Annual Meeting (pp. 548-552). Santa Monica, CA: Human Factors Society; 1987.

Notestine, J. Subjective workload assessment and effect of delayed ratings in a probability monitoring task. Proceedings of the Human Factors Society 28th Annual Meeting (pp. 685-690). Santa Monica, CA: Human Factors Society; 1984.

Nygren, T.E. Conjoint measurement and conjoint scaling: A users guide (AFAMRL-TR-82-22). Wright-Patterson Air Force Base, OH: Aerospace Medical Research Laboratory; April 1982.

Nygren, T.E. Investigation of an error theory for conjoint measurement methodology (763025/714404). Columbus, OH: Ohio State University Research Foundation, May 1983.

Nygren, T.E. Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. Human Factors. 33, 17-33; 1991.

Papa, R.M. and Stoliker, J.R. Pilot workload assessment: A flight test approach. Washington, DC: American Institute of Aeronautics and Astronautics, 88-2105, 1988.

Pollack, J. Project report: an investigation of Air Force reserve pilots' workload. Dayton, OH: Systems Research Laboratory; November 1985.

Potter, S.S. and Acton, W.H. Relative contributions of SWAT dimensions to overall subjective workload ratings. Proceedings of Third Symposium on Aviation Psychology. Columbus, OH: Ohio State University; 1985.

Potter, S.S. and Bressler, J.R. Subjective workload assessment technique (SWAT): A user's guide. Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory; July 1989.

Reid, G.B., Eggemeier, F., and Nygren, T. An individual differences approach to SWAT scale development. Proceedings of the Human Factors Society 26th Annual Meeting (pp. 639-642). Santa Monica, CA: Human Factors Society; 1982.

Reid, G.B., Eggemeier, F.T., and Shingledecker, C.A. In M.L. Frazier and R.B. Crombie (Eds.) Proceedings of the workshop on flight testing to identify pilot workload and pilot dynamics, AFFTC-TR-82-5 (pp. 281-288). Edwards AFB, CA: May 1982.

Reid, G.B. and Nygren, T.E. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock and N. Mehtaki (Eds.) Human mental workload (pp. 185-218). Amsterdam: North Holland; 1988.

Reid, G.B., Shingledecker, C.A., and Eggemeier, F.T. Application of conjoint measurement to workload scale development. Proceedings of the Human Factors Society 25th Annual Meeting. 522-526; 1981.

Reid, G.B., Shingledecker, C.A., Nygren, T.E., and Eggemeier, F.T. Development of multidimensional subjective measures of workload. Proceedings of the IEEE International Conference on Cybernetics and Society. 403-406; 1981.

Rueb, J., Vidulich, M., and Hassoun, J.A. Establishing workload acceptability: an evaluation of a proposed KC-135 cockpit redesign. Proceedings of the Human Factors Society 36th Annual Meeting. 17-21; 1992.

Schick, F.V. and Hann, R.L. The use of subjective workload assessment technique in a complex flight task. In A.H. Roscoe (Ed.) The practical assessment of pilot workload, AGARDograph No. 282 (pp. 37-41). Neuilly-sur-Seine, France: AGARD; 1987.

See, J.E. and Vidulich, M.A. Assessment of computer modeling of operator mental workload during target acquisition. Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting. 1303-1307, 1997.

Skelly, J.J. and Purvis, B. B-52 wartime mission simulation: Scientific precision in workload assessment. Paper presented at the 1985 Air Force Conference on Technology in Training and Education. Colorado Springs, CO; April 1985.

Skelly, J.J., Reid, G.B., and Wilson, G.R. B-52 full mission simulation: Subjective and physiological workload applications. Paper presented at the Second Aerospace Behavioral Engineering Technology Conference, 1983.

Skelly, J.J. and Simons, J.C. Selecting performance and workload measures for full-mission simulation. Proceedings of the IEEE 198 National Aerospace and Electronics Conference. 1082-1085; 1983.

Thiessen, M.S., Lay, J.E., and Stern, J.A. Neuropsy-Chological workload test battery validation study (FZM 7446), Fort Worth, TX: General Dynamics; 1986.

van de Graaff, R.C. An in-flight investigation of workload assessment techniques for civil aircraft operations (NLR-TR-87119 U). Amsterdam, the Netherlands: National Aerospace Laboratory; 1987.

Vickroy, S.C. Workload prediction validation study: The verification of CRAWL predictions. Wichita, KS: Boeing Military Airplane Company; 1988.

Vidulich, M.A. The Bedford Scale: Does it measure spare capacity? Proceedings of the 6th International Symposium on Aviation Psychology. 2, 1136-1141; 1991.

Vidulich, M.A. and Pandit, P. Individual differences and subjective workload assessment: Comparing pilots to nonpilots. Proceedings of the International Symposium on Aviation Psychology. 630-636, 1987.

Vidulich, M.A. and Tsang, P.S. Techniques of subjective workload assessment: a comparison of SWAT and NASA-Bipolar methods. Ergonomics. 29(11), 1385-1398; 1986.

Vidulich, M.A. and Tsang, P.S. Absolute magnitude estimation and relative judgment approaches to subjective workload assessment. Proceedings of the Human Factors Society 31st Annual Meeting. 1057-1061; 1987.

Ward, G.F. and Hassoun, J.A. The effects of head-up display (HUD) pitch ladder articulation, pitch number location and horizon line length on unusual altitude recoveries for the F-16 (ASD-TR-90-5008). Wright-Patterson Air Force Base, OH: Crew Station Evaluation Facility; July 1990.

Warr, D.T. A comparative evaluation of two subjective workload university measures: the subjective assessment technique and the modified Cooper-Harper Rating. Masters thesis, Dayton, OH: Wright State, 1986.

Warr, D., Colle, H., and Reid, G.B. A comparative evaluation of two subjective workload measures: The subjective workload assessment technique and the modified Cooper-

Harper Scale. Paper presented at the Symposium on Psychology in the Department of Defense. USAFA, Colorado Springs, CO.; 1986.

Whitaker, L.A., Peters, L., and Garinther, G. Tank crew performance: Effects of speech intelligibility on target acquisition and subjective workload assessment. Proceedings of the Human Factors Society 33rd Annual Meeting. 1411-1413; 1989.

Wilson, G.F., Hughes, E., and Hassoun, J.A Physiological and subjective evaluation of a new aircraft display. Proceedings of the Human Factors Society 34th Annual Meeting. 1441-1443; 1990.