CODAP: A Comparison of Single versus Multi-Factor Task Inventories

Johnny J. Weissmuller & Michael R. Staley
of Sensible Systems, Inc.

Sandra West
Personnel Department, City of Fort Worth

BACKGROUND

Within many military and civilian agencies, there is a fundamental technology which supports both the operational and research occupational analysis programs. The core of this technology is called CODAP, an acronym for the Comprehensive Occupational Data Analysis Programs. The CODAP "system" is a set of analysis tools and procedures which use, as raw material, information provided by the members of the occupational field being studied. This system may be used to revise classification structures, assess job related skills, verify the relevance of training courses and a host of other applications in which an accurate knowledge of job content at the task level is desirable (Thew and Weissmuller, 1979; Morsh and Christal, 1966; Christal, 1974; Levine, Ash, Hall and Sistrunk, 1983, Christal and Weissmuller, 1988).

Several years ago the City of Fort Worth first used the Comprehensive Occupational Data Analysis Programs (CODAP) system to analyze their clerical families with the help of the Air Force Human Resources Laboratory. More recently, the Fort Worth Personnel Department undertook a second CODAP project to study patrol officers. Concerned about cost factors, a proposal was made to collect three Subject-Matter-Expert (SME) ratings in a side-by-side format. As Sensible Systems was mastering booklets and processing the data, they became involved in this discussion and strongly recommended against such a format. When asked for specific evidence of problems encountered in using this technique in a job survey setting, Sensible was unable to point to a documented study. From word-of-mouth reports by people who have used multi-factor booklets, it seemed that the operational practitioners who found the degree of contamination too great simply disposed of their data (and never documented the problem) while other practitioners simply claimed the loss of reliability was within acceptable limits given the cost savings. In order to more formally assess the impact of side-by-side ratings of independent factors, the City of Fort Worth and Sensible Systems, Inc. cooperated in collecting and analyzing subject matter expert (SME) ratings in both the single-factor and triple-factor rating booklet formats. This paper documents the results of that study.

METHOD

Five survey booklets were prepared. The first booklet contained rater background information along with a 52-item physical activity inventory (relative importance), and a 91-item knowledge, skills, and ability (KSA) inventory. The KSA inventory was rated on an "importance to success" scale and a "value at entry" scale. To preclude the kind of induced correlation being examined in the rest of this study, the KSA inventory was listed TWICE, once with the importance scale and, in the following section, with the "value at entry" scale. To provide comprehensive tracking and the widest rater base for the KSA items, this booklet was administered to all raters.

The second rater booklet contained the 468-item task inventory with three side-by-side rating columns. The three rating columns were always ordered (from left to right) as "Importance," "Physical Demands," and "Learning Difficulty." It was believed that "physical demands" represented a significantly different kind of factor than the other two. It was hoped that placing "physical demands" between the other two would help reduce the correlation between "Importance" and "Learning Difficulty." The instructions directed the SME to rate importance first. Immediately following the inventory, further instructions told the SME to return to the beginning of the booklet and then rate physical demands. Final instructions directed the SME to again return to the beginning of the booklet and rate learning difficulty.

The third, fourth, and fifth rater booklets also contained the 468-item task inventory with a single rating column. The third booklet contained the "Importance" factor while the fourth booklet contained the "Physical Demands" factor and the fifth booklet contained "Learning Difficulty."

All raters were given the first, background/KSA booklet. The sample was divided in approximately a 1-to-2 ratio to receive either the 3-factor booklet (n=23) or TWO of the three single factor booklets (n=49). Unfortunately, no records were generated to record which of the two booklets raters completed first. Upon return of the booklets, all data were automated and entered into the atCODAP system.

ANALYSIS

Each rater sample by factor combination subjected to interrater reliability analyses in the atCODAP system. Data were processed as raw responses (no adjustments) and no deletions were allowed. Vectors of mean values were retained and correlated to compare treatment effects.

The primary hypothesis being tested was that the three factor booklet would show increasing contamination (induced correlation) from left to right. This contamination should have effects on several statistics associated with the 3-factor booklet as compared to the 1-factor booklets. One would expect higher interrater agreement (r11) in single factor booklets due to fewer distractions. This pattern was shown (.222/.142, .252/.159, and .267/.099). One would also expect the intercorrelations between the different factors to be higher in the 3-factor booklet sample. These data are shown in Table 1.

As mentioned earlier, mean rating vectors (task factors) were saved at the end of each interrater reliability analysis. Correlating these vectors allows one to compare how well the results of a 3-factor booklet agree with a 1-factor booklet. If the contamination spreads as predicted, one would expect a drop in the inter-treatment correlations from importance to physical demands to learning difficulty between the 3-factor booklet and the 1-factor booklet vectors. This pattern is shown (.921, .912, .752). The relationship is not expected to be symmetric. For example, the correlation between a 3-factor importance vector (little contamination) with 1-factor difficulty vector (little contamination) should be much lower than the correlation between a 1-factor importance vector (little contamination) with a 3-factor difficulty

vector (high contamination) and the difference is apparent (0.167 vs 0.594). The pattern repeats for importance vs physical demands (0.389 vs 0.539) and for physical demands vs difficulty (0.366 vs 0.554). These relationships are shown in Table 2.

RESULTS

It was assumed that the 1-factor booklet would define the "real" relationship between the three different factors (importance, physical demands, and learning difficulty), but the analysis showed a lower correlation (0.167) between the un-contaminated importance factor in the 3-factor booklet and the difficulty factor in the 1-factor booklet than was found between importance and difficulty in the 1-factor booklet (0.291). As each 1-factor booklet rater was given two booklets to rate, contamination effects were introduced into the 1-factor booklet data sets. Fortunately, as each rater received only two of the three booklets, it was possible to split each factor rater set into two subsets. For example, for all raters who rated task difficult, approximately half also rated importance with the remainder rating physical demands. Comparing "importance" to "difficulty" by independent subsets of raters yields a relationship of 0.166 -- much better. Similar analysis drops the "importance" to "physical demands" relationship to 0.359 (from 0.437) and the "physical demands" to "difficulty" relationship to 0.417 (from 0.445).

Although the 1-factor booklet raters rated two factors, they did this in two separate booklets, not side-by-side and hence should demonstrate minimal interference between factor ratings. The analysis described above split each rater group into half by other factor rated. Correlations between these non-random split-halves was 0.855 for importance, 0.935 for physical demands, and 0.916 for task difficulty. As no controls were imposed on the SMEs and no records were made, it is not known in which order the 1-factor booklets were completed. From the split-half analysis it appears that perhaps the importance booklet might have been done last most often and hence shows a somewhat less than expected level of stability due to contamination.

CONCLUSIONS

Even though an attempt was made to arrange the three factor booklet to minimize contamination, factor contamination did occur. The correlation between mean vectors for importance and learning difficult rose from 0.167 in the least contamination scenario to 0.594 in the highest contamination scenario. This is significant well beyond the 0.01 level. Internal measures of consistency also showed significant drops beyond the 0.01 level across the major treatment (Peters & McCormick, pp 94-95). Inspection of the z-values shows the trend towards higher contamination (4.58 for importance, 5.08 for physical demands, and 9.28 for learning difficulty). With these findings it would be easy to justify warnings against multi-factor booklets. In some projects, however, several factors may be collected simply to ensure that all the bases are covered and a minor degree of contamination is not a problem. Each CODAP study, however, must be evaluated in its own context, judging the need for "pure" factors versus project cost limitations.

TABLE 1

RELIABILITIES & CORRELATIONS
SINGLE- VS TRIPLE-FACTOR SME RATINGS
(Within Booklet Type)

	-- SINGLE FACTOR BOOKLET --			-- TRIPLE FACTOR BOOKLET --		
	IM	PD	LD	IM	PD	LD
Importance (IM)	r11=.222 rkk=.900 k=31.8			r11=.142 rkk=.792 k=22.9		
Physical Demands (PD)	0.437	r11=.252 rkk=.915 k=31.9		0.526	r11=.159 rkk=.811 k=22.8	
Learning Difficulty (LD)	0.291	0.445	r11=.267 rkk=.923 k=32.9	0.532	0.614	r11=.099* rkk=.716 k=22.8

*Rises to 0.127 with two raters deleted. Addition deletions drop rkk.

All corresponding reliabilities (r11s) and correlations are significantly different at the 0.01 level.

TABLE 2

CORRELATIONS
SINGLE- VS TRIPLE-FACTOR SME RATINGS
(Between Booklet Type)

SINGLE FACTOR BOOKLET	-- TRIPLE FACTOR BOOKLET --		
	IM	PD	LD
Importance (IM)	0.921	0.539	0.594
Physical Demands (PD)	0.389	0.912	0.554
Learning Difficulty (LD)	0.167	0.366	0.752

Corresponding off-diagonal correlations are significantly different at 0.01 level, e.g., IM-PD (0.389) vs PD-IM (0.539).

BIBLIOGRAPHY

Christal, R. E. (1974). The United States Air Force occupational research project (AFHRL-TR-73-75, AD-774 575). Lackland AFB, TX: Occupational Research Division, Air Force Human Resources Laboratory.

Christal, R. E. & Weissmuller, J. J. (1976). New CODAP programs for analyzing task factor information (AFHRL-TR-76-3, AD-A026 121). Lackland AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory.

Christal, R. E. & Weissmuller, J. J. (1988). The Job-Task Inventory. in The job analysis handbook for business, industry and government (Vol. 2). Sidney Gael (Ed). New York: John Wiley & Sons.

Goody, K. (1976). Comprehensive occupational data analysis programs (CODAP): Use of REXALL to identify divergent raters. (AFHRL-TR-76-82, AD-A034 327). Lackland AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory.

Levine, E. L., Ash, R. A., Hall, H., & Sistrunk, F. (1983). Evaluation of Job Analysis Methods by Experienced Job Analysts. Academy of Management Journal, 26(2), 339-348.

Morsh, J. E. & Archer, W. B. (1967). Procedural Guide for Conducting Occupational Analysis in the United States Air Force (PRL-TR-67-11, AD-664 037). Lackland AFB, TX: Personnel Research Laboratory.

Morsh, J. E. & Christal, R. E. (1966) Impact of the computer on job analysis in the United States Air Force (PRL-TR-66-19, AD-656 304). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Peters, D. L. & McCormick, E. J. (1966). Comparitive Reliability of Numerically Anchored versus Job-Task Anchored Rating Scales. Journal of Applied Psychology, 50(1), 92-96.

Staley, M. R. & Weissmuller, J. J. (1981). Interrater Reliability: The development of an automated analysis tool (AFHRL-TP-81-42, AD-A108 400). Brooks AFB, TX: Technical Services Division, Air Force Human Resources Laboratory.

Thew, M. C. & Weissmuller, J. J. (1979). CODAP: A current overview. Proceedings 21st Annual Conference of the Military Testing Association 462-479. San Diego CA: Military Testing Association.

Weissmuller, J. J. (1979). CODAP: Applications for state and local governments. (Exhibit). Hartford CT: Federal Technology Transfer Forum and Exposition.