

AD P00901

Using Rating Scales to Determine Aptitude Requirements of Army Systems

Paul G. Rossmeissl and John A. Dohme
U S Army Research Institute

The United States Army is facing the possibility of a serious manpower shortage in the not to distant future. Three factors, operating concurrently, are contributing to this shortage. First, census data indicate that the quantity of individuals available for military service (18-25 year olds) will decline throughout this century and, if the birth rate remains unchanged, for the foreseeable future. Also, standardized aptitude and achievement test scores have shown a consistent decline over the past 15 years (Waters, Eitlberg, & Laurence, 1981). Taken together, these two factors indicate increased future competition among the armed forces and the civilian sector for qualified personnel, with the competition expected to be most severe for the more highly skilled individuals.

The third factor is the increasing technological sophistication of the Army's new systems. It is widely accepted that increased sophistication is increasing operator and maintainer job complexity and in turn increasing skill requirements and quantitative demand for personnel (Kerwin, Blanchard, Atzinger, & Topper, 1980), although quantitative evidence of this suspected trend is lacking (GAO, 1981). The Army, therefore, faces the possibility of increasing quantitative and qualitative personnel demands while the capability of the population to fill that demand is decreasing.

This specter of manpower shortage makes it all the more important that the Army investigate and develop techniques that will help make optimal use of the personnel that are available.

This paper reports the results of a study to assess the feasibility of using rating scales to estimate the aptitudes or abilities required to operate and maintain Army systems. If accurate aptitude estimates can be obtained in this manner, the methodology could prove to be useful in two manners. First, the scales could be used to estimate the aptitude requirements of Army systems still in the design process, (Rossmeissl, Kostyla, and Baker, 1981). Second, aptitude requirement information from systems about to be, or already, fielded could be used to develop selection and classification instruments to assist in the assignment of personnel to jobs.

The current research investigated three aspects of the utility of obtaining estimates of army aptitude requirements using rating scales. If rating scales are to be useful in the context they should show three properties: they should have high inter-rater reliability, they should reliably discriminate among the aptitudes being investigated, and they should discriminate among different Army jobs.

Method

Rating Scale Development. Army aviation was selected as a test bed for investigating the use of rating scales, so a set of scales were developed that would be directly relevant to four Army helicopter missions: Aeroscout, Attack, Cargo, and Utility.

Using task analysis procedures thirty aptitudes or abilities were identified as being possible requirements for the helicopter missions. One rating scale was then developed for each of those aptitudes. The final rating scales were very similar to those used by Fleishman (1972, 1975) in that each scale contained the Fleishman definition of the aptitude and a seven point linear rating scale. The current scales did differ from those typically used by Fleishman, however, in that the scale anchor points were directly relevant to Army aviation.

To develop these aviation-specific anchors for the 30 abilities, an ARI psychologist and an ARI Master Aviator developed as many Army aviation task statements as possible for each aptitude. The objective was to create anchor candidates that would cover the range of each aptitude from the least to the greatest amount required in performing all four Army aviation missions. In other words, to develop mission general statements that would be common to all four missions. For each ability, 152 anchor candidates were generated using the Aircrew Training Manuals (ATMs) and helicopter Operator's Manuals (-10's) as guides.

Once the anchor candidates were generated, two Standardization Instructor Pilots (SIPs) were brought in to represent each mission and a roundtable discussion was held to eliminate those candidate statements that did not apply to all four missions. Certain mission oriented candidates were also eliminated because they were not part of the training regimen for a given mission. In addition, the eight SIPs edited the working of the candidates to improve their clarity.

The remaining anchor candidates were included in a questionnaire instrument that was administered to 44 field experienced Army Warrant Officer aviators. These subjects were either current field aviators or students in the Warrant Officer Senior course (WOSC) at Fort Rucker. The subjects, who were mostly CW3 and CW4 ranks, were distributed across the four missions as follows: Aeroscout 20%, Attack 27%, Utility 23%, Cargo 30%. The anchor development questionnaire was adapted from the methodology used by Fleishman (1972-1973). Subjects assigned a value from 1-7 to each candidate corresponding to the amount of the given aptitude required to perform that task. Conceptual definitions were provided for each aptitude. The mean and standard deviation for each of the 288 anchor candidates were calculated and an attempt was made to select three anchors for each aptitude: one high, one low and one medium. In a few cases (6 of the 30 aptitudes) it wasn't possible to develop three anchors because the mean values clustered toward one end of the seven point scale, so only two were created. For each aptitude, the criterion was to select anchors that had small standard deviations, preferably 1.5 or less. The anchors were selected judgmentally to obtain the highest and lowest mean ratings having small standard deviations and also the rating closest to midscale (4.0) having a small standard deviation.

Rating Scale Evaluation. The rating scale approach to aptitude assessment was then evaluated by having Army aviators estimate the aptitude requirements of the four helicopter missions using the rating questionnaires with the aviation anchors served as points of reference on the seven point aptitude scales. The questionnaires were administered to experienced unit aviators (minimum total hours 700, minimum hours in mission 200) at Fort Campbell, Kentucky: Hunter Army Airfield, Georgia and to a few combat and combat support unit aviators at Fort Rucker who had recently been gained from field assignments. A total of 73 warrant officer aviators were sampled 19 Aeroscout, 19 Attack, 17 Cargo and 18 Utility.

Results

Inter-rates reliability. To estimate the inter-rater reliability of rating scales the aeroscout mission was chosen for detailed analysis. The data from the nineteen aeroscout aviators was factor analyzed across the thirty aptitudes. Analysis runs were conducted investigating the possibility of uncovering one through six factors in the data. However, if the interrater agreement for the scales is high a single factor should account for the data. The results of the two factor analysis for the nineteen subjects are shown in Table 1. As can be seen from the table the

Table 1
Two Factor Loadings of Aeroscout Data

<u>Factor 1</u>	.63	.60	.60	.68	.83	.47	.69	.54	.82	.74
<u>Factor 2</u>	-.47	.38	.13	.03	-.16	.17	.06	.36	-.13	-.22
<u>Factor 1</u>	.84	.67	.84	.66	.49	.69	.16	.57	.70	
<u>Factor 2</u>	.19	-.07	.17	-.43	.30	-.39	.31	-.19	-.08	

data can be captured pretty well by a single factor. Fifteen of the nineteen subjects loaded on the first factor at over .5. No loadings on the second factor were over .5 and any second factor loadings between .4 and .5 were negative. Statistically 93% of the variance in the data could be attributed to factor one. This finding of a single factor indicates that most of the aviators were performing the task in a similar manner and the inter-rater agreement of the rating scales was high.

Discrimination Among Aptitudes. To determine whether or not the rating scales were able to discriminate among the thirty aptitudes an analysis of variance was conducted on the data from the fifteen aeroscout aviators who loaded greater than .50 on factor 1 above. The results of this analysis showed that the rating scales were able to discriminate among the aptitudes ($F_{29,14}=13.6, p < .01$). Given the successful analysis of variance, a Newman-Keules test was conducted to uncover any trends in the mean scores among the thirty aptitudes. The results of this analysis indicated that the aptitude ratings tended to fall statistically into three categories: primary requirements, secondary requirements, and incidental or low requirements. The aptitudes that were classed as primary or low requirements are given in Table 2. The remaining twenty aptitudes fell into the class of secondary requirements.

Table 2

High and Low Aeroscout Aptitude Requirements

Primary Requirements

stamina
 stress tolerance
 time sharing
 divided attention
 perceptual speed

Incidental Requirements

written expression
 visualization
 number facility
 static strength
 finger dexterity

Mission/Job Discrimination. To determine if the rating scale methodology was able to discriminate among the aptitudes required for the four different helicopter missions a two-way analysis of variance was conducted. The results of this analysis showed a statistically significant ($F_{29,001}=30.07$ $p < .01$) main effect of aptitude, again indicating that the rating scales were able to show differences among the aptitudes. However, both the main effect of mission and the mission helicopter interaction did not reach statistical significance ($F_{3,69}=2.21$ $p=.095$ and $F_{87,001}=1.44$, $p=.76$ respectively). Taken together these latter two findings indicate that the rating scales were not able to uncover any differences in the aptitudes required to fly the four different helicopter missions.

Discussion

The results discussed above showed that the rating scale methodology succeeded in two of the three properties that were investigated. The methodology showed acceptable inter-rater reliabilities and was able to successfully discriminate among the levels of required aptitudes. Thus, it appears that the approach may be useful in analyzing Army jobs.

However, in this case the rating scale methodology was unable to distinguish among the aptitudes required to fly the four different missions. This finding is probably not surprising since the four different jobs are pretty similar. But it does indicate that the approach has limitations, and further research should be conducted to determine how much jobs should differ before the rating scales will uncover different aptitude requirements.

References

- Fleishman, E. A. On the relation between abilities and human performance. American Psychologist, 1972, 27, 1017-1032.
- Fleishman, E. A. Toward a taxonomy of human performance. American Psychologist, 1975, 30, 1127-1149.
- Kerwin, W. T., Blanchard, G. S., Atzinger, E. M., and Topper, P. E. Man-Machine Interface: A Growing Crisis (Among Top Problem Areas, Discussion Paper No. 2) US Army Material Systems Analysis Activity, Aberdeen Proving Ground, MD, 1980.
- Rossmessl, P. G. Kostyla, S. J., and Baker, J. Human Aptitude Assessment Techniques for System Designers. Proceedings of the 23rd Annual Conference of the Military Testing Association (Vol II), Arlington, VA 1981.
- US General Accounting Office. Effectiveness of US Forces can be increased through improved weapon system design, PSAD 81-17. Washington, DC, Author, 1981.
- Waters, B. K., Eitelberg, M. J., and Laurence, J. H. Military and civilian test score trends (1950-1980). Proceedings of the 23rd-Annual Conference of the Military Testing Association (Vol. II), Arlington, VA, 1981.

