Development of Specific Aptitude Test Norms for the GATB in Nothern Ireland

By: A. MOKADEM, Algier's University.
P. HENDERSON, Queen's university of Belfast.

Introduction

The General Aptitude Test Battery (GATB) developed by the U.S. Employment Service (Dvorak 1947), remains one of the best multifactor batteries available today; to an unique extent it has not only accumulated occupational normative and validity data, but also establishes its applicability to adolescents and adults. The promise of continuing to improve its validation studies in the industrial setting enhances its value to Personnel Selection.

The GATB consists of 12 tests measuring 9 aptitudes. The aptitudes are: General Ability (G); Verbal Ability (V); Numerical Ability (N); Spatial Ability (S); Form Perception (P); Clerical Ability (C); Motor Coordination (K); Finger Dexterity (F); and Manual Dexterity (M). Of the twelve tests, some are paper and pencil tests and the remaining are apparatus based representing performance measures (see sections 1 of the GATB manual, US Dept. of Labour 1970).

The purpose of this study is to develop specific Aptitude Test Battery (GATP) norms to be used for selecting new applicants for training as «Capacitor Assemblers (electronics)» DOT classification code 726.684-018 in an American Company established in Northern Ireland. The sample consisted of 70 applicants who received the GATB; after about 12 weeks of training their performance was rated. On the basis of qualitative and quantitative analysis and the use of the multiple cutoff method, the following battery of three aptitudes K-90, F-80 and M-85 was established. The study demonstrates a statistically significant and useful relationship between the norms and performance in training.

Procedure

The procedures followed in this study was the same procedures as the validation studies recommended by the U.S.E.S. (U.S Dept. of labour 1970, Section 111). These procedures involve, (1) job analysis (2) administering the entire GATB to a sample of applicants, (3) Selecting a Criterion of Performance (4) Identifying significant aptitudes and (5) establishing norms.

Job Analysis

The initial job analysis was performed by observing the capacitor assemblers' performance on the job and by interviewing the operators and supervisors. Jobs were analysed into elements that are common to many dissimilar jobs. Personnel specification were then identified and assigned to each of these elements. On the basis of the job analysis, a job description was prepared which included all the duties required by the job. Qualitative analysis based on job analysis indicated that the following aptitudes were critical for performance of the work: General Intelligence (G), Form Perception (P), Motor Co-ordination (K), Finger Dexterity (F), and Manual Dexterity (M).

Experimental Battery

Each applicant received the separate answer sheet form (B-1002) of the GATB which is composed of 12 tests measuring 9 aptitudes. Of the 12 tests 8 are paper and pencil tests and 4 are performance tests. The paper and pencil tests appear in 3 booklets. Two of the apparatus tests (parts 9 and 10) involve the USES Pegboard; the other two (parts 11 and 12) require the USES Dexterity board. The administrating time was about $2^{1/2}$ hours. Raw scores were converted into aptitude scores with a mean of 100 and a standard deviation of 20.

Sample

The sample in this study consisted of 70 applicants (69 female and 1 male) employed at A.V.X. Corporation location Coleraine, N. Ireland. All the operators who were hired were given the experimental test battery. The majority of the operators are in the 18-24 age group.

Griterion

The criterion for this study consists of supervisory ratings. The immediate supervisor rated each trainee. The descriptive rating scale that was used had 6 items covering different aspects of job performance, (a)quantity of work (b) quality of work (c) accuracy of work (d) job knowledge (e) aptitude required for work and (f) variety of work. Each item has a five -point rating scale corresponding to different degrees of proficiency. For the purpose of scoring the items, weights of 1 to 5 were assigned to the responses (outstanding, above average, over-age, below average and somewhat limited). Summing the ratings for an operator provided a criterion score.

The criterion distribution was dichotomized, so as to place as nearly as possible one-third of the operators in the low criterion group. The criterion cutting scores was in this case set at 18 which placed 33% in the low criterion group and 67% in the high criterion group.

No measure of criterion reliability was obtained as there was only one rating available, but it is indicated in the USED Test guide (1967, Vol. 1), that the usual reliability rating for this scale is about 0.90.

Design

The first attempt to establish the predictive validity of the tests was frustrated. Testing of applicants was carried out at the beginning of training between December 1978 and March 1979, and Criterion data was collected in June 1979. As applicants received tests at different times (according to the recrutment) but later were rated at approximately the same time information was not used. Thus the study considered another sample of 70 applicants who received tests between September and November 1979 and were rated sequentially between 15th January and 26th March 1980. Criterion data was obteined after about 12 to 16 weeks of training.

Analysis

Selection of significant aptitudes.

The initial step in the analysis is to identify those aptitudes to which show evidence of validity and job relatedness. Both qualitative and quantitative factors are considered in establishing test norms. The statistical results are analysed in conjunction with job analysis. The following are recommended by the compilers of the GATB as indications that a particular aptitude is important for success on the job (section 111 of the GATB manual).

(a) A high mean score relative to the other mean scores of the experimental sample.

(b) A relatively low standard deviation.

(c) A rating of «important» or «critical» on the basis of the job analysis.

(d) A correlation with the cirterion that is significant at least at the 0.05 level. The decision rule for inclusion in trial norms is to satisfy any two of a, b, and, c, or do alone.

Table (1) summarises the statistical results from which the norms are derived, and Table (2) summarises the qualitative and statistical consideration. Most factors that were thought to be important to the job, proved to be highly significant. The aptitude «M» was expected to be highly correlated with the criterion, because the job requires manual dexterity, but the validity coefficient against the criterion was low (r = 0.159). Nontheless M was included in the final test norms as the relatively high.

Table 1

Means, Standard Deviations (SD), Ranges and Correlation with the Criterion for the Aptitude of the GATB

Aptitudes	Mean	SD	Range	r
G-General learning ability	90.18	14.96	58-123	0.332**
V-Verbal Ability	93.44	12.43	59-143	0.181
N-Numerical Ability	91.21	16.71	66-123	0.447**
S-Spatial Ability	93.48	15.81	63-125	0.113
P-Form Perception	105.58	19.27	51-141	0.320**
Q-Clerical Perception	110.01	14.97	75-147	0.312**
M-Motor Coordination	107.71	14.41	77-135	0.331**
F-Finger Dexterity	100.11	19.47	42.135	0.294*
M-Manual Dexterity	98.00	15.81	64-129	0.159

* Significant at 0.05 level.

** Significant at 0.01 level.

mean of this aptitude was thought to imply that applicants having poor dexterity had already been removed from the employed group. Although Numerical Ability Score (N) correlated highly with the Criterion (V=0.447), it was not included in the final test norms because (a) it was rated as irrelevant to the job and (b) it has a relatively low mean and high standard deviation. The aptitudes to be considered in the final test norms are; General Ability (G); Verbal Ability (V); Form Perception (P); Clerical Perception (Q); Motor Coordination (K); Finger Dexterity (F); and Manual Dexterity (M). Establishing of final Test Norms :

Table 2
Summary of qualitative and quantitative data

		G	V	N	S	P	Q	K	F	M
Job analysis	Critical/ important					x		x	x	x
	Irrelevant			x						
Aptitude scores	Relatively high mean		x			x	x	x		
	Relatively low standard deviation		x				x	x		
	Significant r with criterion	x		x		x	x	x	x	
Aptitude considered for inclusion		G	V	N	S	P	Q	K	F	M

These aptitudes are now left to be used to develop a battery of 2, 3 or 4 aptitudes with cutting scores at the point (a) where about the same percentage will meet the cutting scores as the percentage place in the high criterion group and (b) which will maximize the relationship between the battery and the criterion.

Scores for these aptitudes are analysed to determine the extent to which each of the aptitudes and the appropriate cutting scores discriminate between

individuals in the high and low criterion groups. The GATB procedure recommends that the final norms should be set to eliminate a proportion of applicants in the range equal to the number in the low criterion group - 10%.

In the present study

$$\frac{\text{low criterion}}{\text{whole sample}} = \frac{23}{70} = 0.328.$$ The maximum

proportion and the minimum proportion to be excluded by the battery should then be 70(.328 ± .100) that is respectively 30 and 16.

Trial D Cutting Scores are set at one standard deviation below the mean aptitude scores of the sample with deviations at five- and ten- point intervals above and below these values, rounded to the nearest whole number. The numbers in the high and the low criterion groups who score below these trials cutting scores and then noted in a work table of aptitudes x trial cutting scores.

The resulting table, known as «criterion splits» is then reduced by applying a series of eight principles to «key scores» for each aptitude. These are the scores which both reject at least 50% of the minimum number, high and low together, to be screened out by the final norms, and at least 25% of the low criterion group.

Each key score is then tried against each of the other trial cutting scores in further work tables of «nor-splits» of two, three, and four aptitude norms. The recommended nine-step procedure is intended to settle on final norms which satisfy three conditions; 1) they should screen out a proportion of applicants that is within ± 10% of the proportion of the sample in the low criterion group; 2) the proportion of the low criterion group which is screened out should be at least 50%; 3) at least 0.05 significance of correlation between final norms and the criterion (phicoefficient).

The trial norms consisted of the various combinations of 2,3 and 4 aptitudes G,V,P,Q,K,F and M with appropriate cutting scores (full instructions and

recommended work sheets for the analysis of the cutt off process can be found in the U.S Dept. of Labour 1967. Test Development Guide Vol. 2). These combinations resulted in a battery of three aptitudes and with cutting scores K90, F80, and M85.

Validity of the Battery

Table (3) illustrates the significant relationships between the job performance criterion and the SATB (K90, F80 and M85) for the validation sample. The validity co-efficient which was based on the fourfold combination of «pass» and «fail» on these tests of the battery and «high» and «low» on the dichotomized criterion indicate that the Phi Co-efficient of 0.030 is statistically significant (P 0.01).

Multiple regression was conducted between aptitude K, F and M. A multiple correlation of 0.382 (significant at 0.01 level) was abtained.

Table 3

Predictive validity of the Battery K-90, F-80, and M-85

Sample N°	High criterion Group		low criterion Group		Chi square	Significant level	Phi coefficient
	Below cutting scores	Above cutting scores	Below cutting scores	Above cutting scores			
70	9	38	12	11	6.52	0.01	0.30

Effectiveness of the Battery

Table (4) illustrates the effectiveness of the battery in the total validation sample. 67% of the non-test selected workers used for this study were good workers. If the workers had been test selected with the above norms, 78% would have been good workers. 33% of non-test selected workers used for this study, were poor workers. If the workers had been test-selected with the above norms, 25% would have been poor workers. The overal efficacy of the cutoff pattern KFM is indicated by the tetrachoric correlation of 0.53 between predicted status and criterion ratings.

Table 4
Effectiveness of Norms K-90, F-80 and M-85

	Without tests	With tests
Good trainess	67%	78%
Poor trainess	33%	22%

Discussion and conclusion

The results demonstrated in this study are generally satisfactory and indicate that, the GATB norms of K-90, F-80 and M-85 established for the occupation of capacitor, assemblers can predict success in training and early apparent success on the job. This is indicated by (a) the significant correlation with the criterion (b) the validity coefficient which was based on the fourfold combination (= 0.030, P 0.001), (C) the effectiveness of the battery which was indicated by the tetrachoric correlation of 0.53 and (d) when incorporated into the occupational Aptitude Pattern (OAP-62) included in section II of the GATB manual (U.S Dept. of Labour 1970). These norms met the OAP norms with the same aptitudes and within 10 points.

The validity coefficient of this research was somewhat affected by the type of the design in that the training period in which trainees were rated varied for different trainee. An adjustment was made to shorten this period of time to reduce the effects of practive by considering only those who were closer to the rating time (12 to 16 weeks). The effects of practice are very well known in the psychology of learning (e.g. Fleishman 1972). This also leads to the consistency of the criterion over time (Du Bois 1966) Ghiselli and Haire 1960).

In this study as there was only one rating available, there was no measure of criterion reliability. It is important to have at least two ratings made by different supervisors, so as to provide evidence for the validity of the test and reduce bias effects (Mier 1973), Thorndike (1949) noted that bias in a criterion measure may operate to reduce either its relevance or its reliability. Smith and Kendal (1963) reasoned that rating errors might be reduced

substantially by involving job incumbents in choosing and defining job performance rating dimensions via critical incidents procedures. Rating errors can be reduced when special efforts are taken ro help raters do a better job (Dunnette and Barman 1979). Errors of measurement are also reduced by training the raters (e.g, Latham, Wexley and Purssel 1975, Borman 1975 and Bernardin 1978).

Although the utility of the GATB is abvious in saving testing time and skills and in providing several tests measuring several aptitudes in one battery, certain limitations have been indicated by several authors. One of these limitations is that all tests are highly speeded (Anastasi 1976) which may reduce their validity for some purposes (Cronbach 1970). The major limitation of the GATB is the consistent reliance on a multiple cut off method rather than on regression equations. Weiss (1976), indicates; «this method tends to ignore errors of measurement in the predictor variables by specifying exact points on the predictor variable distribution». On the other hand, the GATB with its work sheets which are illustrated in the detailed manual offers a conversion selection procedure for the busy personnel officer who does not have access to sophisticated multivariate techniques.

References:

Anastasi, A. (1976) «Psychological Testing» MacMillan.

Borman, W.C. (1975) «Effects of instruction to avoid halo effect on reliability and validity of performance evaluation rating» J. Appl. Psychol 60, PP. 556-560.

Bernardin, H.J. (1978) "The effects of rate training on leniency and halo errors in Students rating of instructors for rating scales". J.Appl. Psychol. **63**, pp. 301-8.

Cronbach, C.J. (1970) "Aspects of the criterion problem on personnel research in manpower planning" NATO Conference, English University Press.

Dunnette, M.D. and Borman, W.C. (1979) "Personnel Selection and classification systems". Ann. Rev. Psychol. 30. pp. 447-525.

Dvorak, B.J. The new USES General Aptitude Test Battery. J. Appl. Psychol. **31**, pp. 372-376.

Fleishman, E.A. (1972) "On the relation between abilities, learning, and human performance". Amer. Psychologist. 27, pp 1018-1032.

Ghiselli, E.E. (1956) "Differentiation of individuals in terms of their predictability". J. Appl. Psychol **40** pp 374-377.

Ghiselli, E.E. (1960) "The validation of selection tests in the light of dynamic character criteria". Pers. Psychol **30**.

Ghiselli, E.E. (1966) "The Validity of Occupational Aptitude Tests" Wiley.

Lathan, G.P. Wexley, K.M. and Pursell, E.D. (1975) "Training managers to minismize rating errors in the observation of behaviour "J. Appl. Psychol. **60**, pp. 550-555.

Maier, N.R.F. (1973) "Psychology in industrial organizations" Boston, Houghton.

Mosier, C.I. (1951) "Problems and designs of cross-validation" Education and Psychol. measure **11**, pp. 5-11.

Saunders, D.R. (1956) "Moderator variables in prediction "Education and Psychol. measure **16** pp. 209-222.

Smith, Pac and Kendal, K.M. (1963) "Re-translation of expectation: an approach to the construction of unambiguons anchors of rating scales." J. Appl. Psychol. 47, pp. 149-155.

Thorndike, R.C. (1949) "Personnel Selection" New-York. Wiley.

U.S. Department of Labour (1967) Test Development Guide. Vol. 1 and 3.

U.S. Department of Labour (1970) "Manual for the USES General Aptitude Test Battery, section 111. Washington, D.C.

U.S. Department of Labor (1970): Manual for the USES General Aptitude Test Battery" Section I administration.

Weiss, D.J. (1976) "Multivariate Procedure" in **M.D. Dunnette** (Ed.) Handbook of industrial and organizational psychology. Rand McNally College Publishing Company.