Proceedings of the Iowa Academy of Science

Volume 57 | Annual Issue Article 57

1950

Development of a Test of Aptitude for Veterinary Medicine

William A. Owens
Iowa State College

Development of a Test of Aptitude for Veterinary Medicine

By William A. Owens

Particularly since the war, schools of veterinary medicine have been able to enroll for training only a relatively small percentage of their applicants. An increased emphasis upon the importance of identifying the best qualified candidates has resulted in a careful reexamination of selection procedures and in recognition of the possible utilization of some sort of veterinary aptitude test.

Problem

Accordingly, the problem of the present investigation was to discover or to develop an efficient predictor, or predictors, of scholastic success during the first professional year of veterinary training.

Populations

Preliminary findings are based upon the records of all (N = 133) freshmen and sophomores who were enrolled in the School of Veterinary Medicine at The Iowa State College during the academic year 1947-'48.

Validational findings are based upon the academic records of 150 pre-veterinarians tested and *subsequently* enrolled in Veterinary Medicine at either Cornell University (N = 25), Michigan State College (N = 41), Kansas State College (N = 49), or Iowa State College (N = 35) during the academic year 1948-'49.

Method

Indices already available were examined as to their predictive utility. These included: high school academic average, pre-veterinary college average, grade in certain specific pre-veterinary courses, scores on the A. C. E. psychological examination, and subtest scores on Form 20 of the Moss Aptitude Test for Medical Professions (3).

Materials

Since none of the above proved to be a highly satisfactory predictor of academic success, it was decided to construct four special purpose tests. Two of these were 50-item achievement tests, constructed with the assistance of the departments concerned, and over the content of the two most predictive pre-veterinary courses —

The writer wishes to acknowledge the invaluable counsel and assistance of Dr. James E. Wert.

chemistry and zoology. The remaining two were, respectively, 60- and 50-item aptitude tests designed to measure the same abilities as the most predictive pair of sub-tests in the Moss A. T. M. P. The first, called "Paragraph Comprehension," is a reading test; and the second, designated as "verbal Memory," involves the timed study of standard selections with a subsequent objective test upon accuracy of recall. Both are entirely new, and their content was judged by members of *that* staff at Iowa State to be representative veterinary content. The usual psychometric procedures relative to establishment of time limits, analysis of items, and general revision, were employed following a cross-sectional experimental study of the four in 1947-'48 and prior to a longitudinal study of their validity in 1948-'49. The chemistry and zoology achievement tests were combined subsequent to the 1947-'48 study, since their intercorrelation approached their reliabilities; the composite, shortened to 80 items, was simply called "Pre-veterinary Achievement." It may be noted that none of *these* test results have been employed in selection and that the data are, in this regard, uncontaminated.

Criterion

In the validational study, the criterion adopted for the evaluation of veterinary aptitude was that of academic success during the first semester or first two quarters of professional training. A total grade-point average for this period was employed for three reasons: (1) such an average is highly correlated with ultimate scholastic standing; (2) academic mortality is heaviest at this time; and (3) it seemed to be the longest waiting interval in a longitudinal study compatible with the urgency of the case. As these data were received from the several cooperating schools, each subject's grade-point average was assigned a standard score value in a distribution for the institution from which it had been obtained.

Statistical Treatment

Statistical analysis of these data follows the common correlational form with the exception that a discriminant function style of analysis has been employed in obtaining estimates of the weights to be assigned to each of the several new tests in order to produce composite scores which maximize the differences between certain performance groups. In addition, multiple biserial correlations have been employed at one point, after a method described by Wert (4), to estimate the combined effects of the tests in predicting the dichotomous criterion of performance.

Results

The primary results of this investigation have been summarized in 4 tables. Prior to some systematic comment upon them, it seems appropriate to observe that the "corrected" odd-even reliability of the composite or total test scores derived from the several weighted sub-test scores was 0.88. This estimate is based upon results obtained from the entire tested population of 424 candidates for admission into veterinary training at the previously named cooperating institutions. Total test reliability thus seems reasonably satisfactory.[1]

To proceed, in Table 1 are shown the correlations between various predictors and freshman veterinary average. Under "Existing Indices" it is interesting to note that pre-veterinary chemistry average appears to be the best single predictor. A possible artifact involved is the relatively large variance in chemistry grades as compared with that typical of other pre-veterinary subjects.

Table 1

Correlations of Various Predictors with Freshman Veterinary Average (Preliminary Study, N = 133)

Variable	r
Existing Indices	
Total pre-veterinary average	.40
Pre-veterinary chemistry average	.47
Pre-veterinary zoology average	.41
Raw Score on A. C. E.	.02
Moss A.T.M.P. Subtests	
Visual Memory	.13
Memory for Content	.13
Comprehension and Retention	.22
General Information	—.11
Vocabulary	—.01
Understanding of Printed Material	.38
Application of Principles	.17
Logical Reasoning	.02
Four New Tests	
Chemistry Achievement Test	.27
Zoology Achievement Test	.12
Paragraph Comprehension	.47
Verbal Memory	.57
Sum of Paragraph Comprehension & Verbal Memory	.62

At 5 per cent level r = .17; at 1 per cent r = .22

[1] As given with unlimited time in the preliminary study, the reliabilities for the sub-tests are Pre-veterinary Achievement, 0.65; Paragraph Comprehension, 0.78; and Verbal Memory, 0.74.

Sub-tests three and six of the Moss A.T.M.P. suggested functions to be measured by the two new aptitude tests, and the former may be broadly considered as models for the latter.

As previously indicated, the "Four New Tests" were reduced to three following the evidence derived from the preliminary study and through the combination of the chemistry and zoology achievement tests to form the revised Pre-veterinary Achievement Test. In spite of its apparently poor validity, this content was *tentatively* retained because it was recognized that the selection process at Iowa State, involving heavy emphasis upon pre-veterinary success in chemistry and zoology, brought about a tremendous restriction in range of talent on these tests which might not be duplicated at other institutions. In the event that it were not, it was felt that they could conceivably be of substantial value.

From Table 2, then, it is apparent that each of the three new tests much better predicts academic *inability* from a relatively low cutting score than academic standing from the total range of scores

Table 2

Correlations of Test Scores With Grade-Point Average
(Validational Study, N = 150)

	P.—M Correlation	Tetrachoric r (G.P.A. —½ vs. ½)
Pre-veterinary Achievement (X_1)	0.24	0.41 (31%)
Paragraph Comprehension (X_2)	0.36 (0.47)	0.56 (25%)
Verbal Memory (X_3)	0.17	0.26 (28%)

5% r = 0.16 and 1% r = 0.21

(this admittedly *within* the successful group). As evidence, the product-moment correlations in the left-hand column are consistently and substantially smaller than the tetrachoric correlations [2] in the right-hand column. These latter were obtained by arbitrarily dichotomizing grade-point averages at their mean, and by then dichotomizing test score distributions at the points of minimum error in classification — at or slightly above the lowest quartile point. The percentage to the right of each coefficient is that below the cutting score in the truncated tail of the test distribution in question.

Since it appeared to be most efficient in each case to set cutting scores on the tests near the 25th centile, it was arbitrarily decided to break the distribution of grade-point averages at the same level and to attempt to determine how best to weight each test to maximize the differences between these two segments of the criterion.

2 Magnitudes estimated from the computing diagrams of Chesire, et. al. (1).

Thus, in Table 3, are shown the multiple, biserial correlations and discriminant-function weights based upon this proposed dichotomy in the criterion. It may be noted that the correlations for "all students" are considerably smaller than are those "within schools."

Table 3

Multiple Biserial Correlations and Discriminate Test Weights

(Validational Study, N = 150)

	Multiple Biserial r's (G.P.A. —¾ vs. ¼)	Weights in the Discriminate Function
X_2 & X_3 (all students)	0.37	
X_2 & X_3 (within schools)	0.43	$v = 2.05\,X_2 + X_3$
X_1, X_2 & X_3 (all students)	0.40	
X_1, X_2 & X_3 (within schools)	0.45 (0.57)	$v = -0.58\,X_1 + 2.13\,X_2 + X_3$

This is no doubt attributable to test-wise institutional differences so large that an individual's scores might rank him in the highest quarter at school A and in the lowest at school B. The second series of discriminate function weights in Column 2 are those which were employed to derive a single series of composite or total scores, the "V scores," for all tests.

Table 4 is a summary table. In it appear the tetrachoric correlations between the dichotomized composite test scores and the dichotomized criterion. The percent to the right of each coefficient,

Table 4

Tetrachoric Correlations Between Composite Test Scores and Grade-Point Average

(Validational Study)

	Cornell University	Iowa State College	Kansas State College	Michigan State College
	N = 25	N = 35	N = 49	N = 41
High ¾ vs. Low ¼ (Criterion)	0.72 (16%)	0.70 (12%)	0.62 (18%)	0.48 (12%)

again, indicates the proportion of cases below the minimum error cutting score in the tail of the test distribution. Final results were cast in this form to make them coincident with the form of the practical question as it always arises — is a given subject above or below some critical test score, and how does this argue for his being above or below some accepted grade standard.

It is to be regretted that the number of cases for each institution

is not larger, although application of the chi-square test to these data suggests that the *least* significant tabled relationship surpasses the 2% probability level. A sort of increase in numbers may, of course, be achieved by ignoring institutional differences and combining the data. This, naturally, lowers the apparent degree of relationship; in this instance it results in a single r_t of 0.52.

Interpretation

In interpreting these results several conflicting influences must be recognized. The estimates of test-criterion relationship provided in Table 4 may be thought of as overestimates for at least two reasons. First, they are based, after the fact, upon the most efficient test cutting score; whereas, in practice it may be impractical to find or to employ such a value, which will in any case show sampling fluctuations. Second, the scoring weights established for the combination of sub-test scores were derived from a composite of four populations and then applied to this *same* composite population. A "shrinkage" in discriminative efficiency must be expected when these weights are applied to the scores obtained from a new population, although the customary effect may be minimized in this instance since the original sampling was of a broader and more heterogeneous group than could have been obtained at a single school.

Running counter to these two influences, which would tend to make the values of Table 4 appear to be overestimates, is the undoubted fact that the coefficients shown have been depressed by a marked restriction in the range of talent existing *within* the validational group. At some institutions, the standard deviation of composite test scores is 25 to 30 percent larger in the distribution of *candidates* for admission than in the distribution of *selectees*. This is a fact mainly attributable to current selection on the basis of preveterinary grade-point average — a test correlated variable. If Kelley's (2) correction for homogeneity were applied to the present data to obtain estimates of test-criterion correlations within the population of candidates, many of the relationships here reported would be substantially increased in magnitude. For example, in Table 2, the Paragraph Comprehension test would correlate 0.47 with the criterion instead of 0.36; and, in Table 3, the final multiple correlation would be 0.57 instead of 0.45.

In addition, there is evidence to suggest that within the groups admitted to veterinary training, those who *had* taken the tests, and who composed slightly more than half of the total number, were

superior in performance to those who had *not* taken the tests. At least a partial explanation is that, almost without exception, those tested had had their pre-veterinary training at one of the four cooperating institutions. By and large students with this background make better grades and fail less frequently.

In evaluating these conflicting influences the writer is disposed to judge it as not unlikely that they approximately cancel and offset each other, and that the estimates of relationship reported are, therefore, not grossly in error.

Summary and Conclusions

A veterinary aptitude test has been devised which had the following characteristics in the populations studied:

(1) It had a reliability of 0.88.
(2) It had tetrachoric validity coefficients of from 0.48 to 0.72, against a grade-point average criterion.
(3) It was a better predictor of the specified criterion than were pre-veterinary grades, singly or collectively, or the A. C. E.
(4) It predicted most efficiently, within the validational group, from a relatively low cutting score.

References

1. Chesire, L., Saffir, M., & Thurstone, L. L. Computing diagrams for the tetrachoric correlation coefficient. Chicago: The University of Chicago Bookstore, 1933.
2. Kelley, T. L. Statistical Method. New York: McMillan, 1923. pp. 223-228.
3. Moss, F. A. & Hunt, Thelma. Aptitude Test for medical professions. Prepared especially for the War Department by the authors and under a committee of the Association of American Medical Colleges.
4. Wert, James E. Unpublished manuscript available from the author. Iowa State College.

Department of Psychology
Iowa State College
Ames, Iowa