DOCUMENT RESUME

ED 131 097 TM 005 805

AUTHOR Schroeder, Lee L.
TITLE A Note on the Probability of Errors in Decisions Based on Tests of the College Level Examination Program.
NOTE 9p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Correlation; Decision Making; *Equivalency Tests; Error Patterns; *Grades (Scholastic); *Higher Education; Probability; *Scores; *Test Interpretation; *Test Validity
IDENTIFIERS *College Level Examination Program

ABSTRACT

College Level Examination Program (CLEP) Tests were normed on a national basis, administering the test to nationwide samples of subjects. Norms appear in the booklet, CLEP Scores: Interpretation and Use, and consist of the test score means for groups of students receiving grades of A, B, C, D, and F in the relevant course, the proportion of students receiving each grade, the sample size, and the correlation coefficient between test scores and earned course grades. The validity of the tests for the purpose of selecting creditable students from the population of test takers is assessed through an examination of the systematic differences in test means across earned grades, and the correlation coefficient between earned grades and test scores. This paper supports the assertion that the correlation coefficient is a misleading statistic for the purpose of validity assessment in this context. A decision theoretic procedure is developed which focuses on the likelihood of errors in test based decisions. The decision theoretic means of validity assessment was applied to the data of all CLEP tests discussed in the norming literature. The analysis showed a dramatically incoherent value system displayed among the tests, with wildly fluctuating error likelihoods and ratios. Further, the likelihood of each error type was found to be substantially greater than would be supposed based on an examination of the correlation coefficients between earned grades and test scores. (Author/MV)

ED131097

A NOTE ON THE PROBABILITY OF ERRORS IN DECISIONS BASED ON TESTS OF THE COLLEGE LEVEL EXAMINATION PROGRAM

by

Dr. Lee L. Schroeder
Director of Measurement and Evaluation
Burlington County College
Pemberton, New Jersey 08068

A NOTE ON THE PROBABILITY OF ERRORS IN DECISIONS BASED ON TESTS OF THE COLLEGE LEVEL EXAMINATION PROGRAM

A review of the correlation between subject area tests of the College Level Examination Program and relevant course grades shows these correlations to exist to a moderate degree in the norming sample (r values range from .37 to .67).[1] While these correlation coefficients are statistically significant, and even large, when compared to the results of most validity studies, their size may contribute to a false sense of confidence in the validity of the test with regard to it's use.

Another view of the validity of the test may be developed from a decision theoretic viewpoint with closer association between the analytic procedure and test use. Let decision D_1 indicate the decision that the student has the skills associated with a given course, and decision D_2 indicate the decision that the student does not have such skills. Define groups P_1, the students for which D_1 was the appropriate decision, and P_2, the students for which D_2 was appropriate.

For an appropriate test, let T represent a chosen threshold. The rule is adopted that a score greater than T causes decision D_1 and a score less than T causes decision D_2. Two error conditions exist.

Error Type I: Decision D_1 is made for a member of P_2

Error Type II: Decision D_2 is made for a member of P_1

If x_i is the score of subject s_i, define:

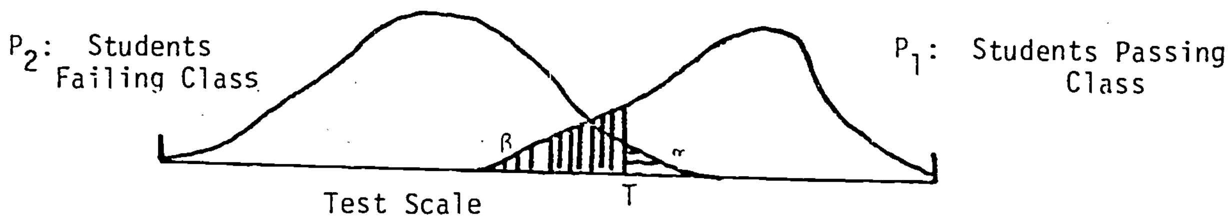
$$\alpha = \text{Probability } (x_i > T | s_i \varepsilon P_2)$$

$$\beta = \text{Probability } (x_i < T | s_i \varepsilon P_1)$$

---

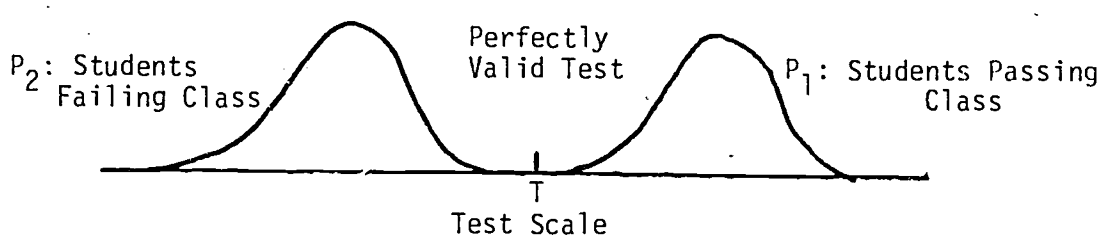
[1]College Entrance Examination Board "CLEP Scores: Interpretation and Use" (1973) p. 22.

These error probabilities are illustrated below:

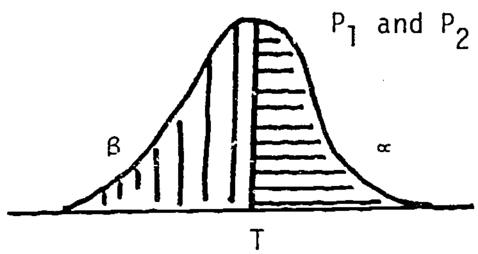


Within the framework, let P_1 be the students receiving a grade of A, B or C in a course relevant to a given CLEP subject area examination. P_2 consists of those students receiving a D or F in such courses. D_1 and D_2 are decisions to grant or not grant credit for the course via the CLEP examination. T is the recommended threshold, the mean scaled score of the student receiving a grade of C in the course. Further, α is the proportion of students receiving grades of D or F with scores above T. β is the proportion of students with grades of A, B or C who have scores below T.

It may be noted that it would be possible to define a <u>Perfectly Valid</u> test as one where α and β are zero or extremely near zero. This occurs when the distributions are disjoint.



A <u>Completely Invalid</u> test could be defined as one where α and β equal .5. This occurs when the distributions are coincidental.



Test Scale

Having developed this approach to validity analysis, it is possible to compute the magnitude of α and β for the CLEP subject area examinations. Information for the computation is provided by Educational Testing Service.[2] In this table, the means of each within final grade group, the percent of the total norming sample represented by each within final grade group and the total number of subjects is given for each test. From this the number of subjects receiving each grade may be computed. This information may be symbolized as follows:

Mean and Percentages for
Within Grade Groups

Grade	A	B	C	D	F
Mean	M_a	M_b	M_c	M_d	M_f
Number	n_a	n_b	n_c	n_d	n_f

The mean CLEP score for students with grades of C or better is computed as:

$$\mu_p = \frac{n_a M_a + n_b M_b + n_c M_c}{n_a + n_b + n_c}$$

and the mean score for students with grades less than C is computed as:

$$\mu_p = \frac{n_d M_d + n_d M_f}{n_d + n_f}$$

[2] Ibid., p. 22.

To compute the standard deviation of the C or better and less than C student group, the standard deviation of the within grade groups is required. Since this information is not readily available from Educational Testing Service, an estimate is needed. A conversation with Walter Shea of ETS lead to the use of the value 5 as a standard deviation for all within grade groups, Accordingly,

$$\sigma_j^2 = 25 = \sum_{i=1}^{n_j} (x_{ij} - M_j)^2$$ where j = A,B,C,D, or F and n_j refers

to the number of subjects receiving a grade of j.

Therefore:

$$25 = \sum_{i=1}^{n_j} \frac{(x_{ij}^2 - 2\, M_j\, x_{ij} + M_j^2}{n_j}$$

$$25 n_j = \sum_{i=1}^{n_j} x_{ij}^2 - 2M_j \sum_{i=1}^{n_j} x_{ij} + M_j^2\, n_j$$

$$= \sum_{i=1}^{n_j} x_{ij}^2 - 2M_j^2\, n_j + M_j^2\, n_j$$

$$25 n_j = \sum_{i=1}^{n_j} x_{ij}^2 - M_j^2\, n_j.$$ If $\sum_{i=1}^{n_j} x_{ij}$ is denoted SS_j, then

$25n_j \doteq SS_j - M_j^2\, n_j$ or $SS_j = n_j\, (M_j^2 + 25)$. The standard deviation for those receiving a grade of C or better may be computed as follows:

$$\sigma_p^2 = \frac{\sum_{j=a}^{c} \sum_{i=1}^{n_j} x_{ij}^2 - \mu_p^2 N_p}{N_p}$$ where $N_p = n_a + n_b + n_c$ and is the

mean of the subjects with a grade of C or better. Thus,

$$\sigma_p^2 = \frac{SS_a + SS_b + SS_c - (n_a + n_b + n_c)\ \mu_p^2}{n_a + n_b + n_c}$$

$$= \frac{n_a\ (n_a^2 + 25) + n_b\ (n_b^2 + 25) + n_c\ (n_c^2 + 25) - (n_a + n_b + n_c)\ \mu_p^2}{n_a + n_b + n_c}$$

Similarly:

$$\sigma_f^2 = \frac{n_d\ (25 + M_d^2) + n_f\ (25 + M_f^2) - (n_a + n_b)\ \mu_f^2}{n_d + n_f}$$

If the threshold T is defined as M_c, the mean score of C group (as recommended by CEEB in the booklet mentioned above),

then $$\alpha = \int_{\frac{T-\mu_f}{\sigma_f}}^{\infty} \phi(x)dx \text{ and } \beta = \int_{-\infty}^{\frac{T-\mu_p}{\sigma_p}} \phi(x)dx$$

where $\phi(x)dx$ is the standard normal distribution.

A computer program was developed to provide the computation as outlined above for each CLEP subject examination.

The examinations and the associated error probabilities computed as indicated above, are listed below. In computing α and β, it is assumed that the test scores are normally distributed in each population.

CLEP SUBJECT AREA EXAMINATION
ERROR PROBAELITIES

Examination	α	β
Accounting, Introductory	.23	.26
American Education, History of	.21	.21
American Government	.22	.26
American History	.23	.24
American Literature	.16	.20
Biology	.27	.28
Business Law, Introductory	.07	.28
Business Management, Introductory	.13	.31
Calculus, Introductory	.19	.30
Chemistry,General	.30	.30
College Algebra	.13	.29
College Algebra-Trigonometry	.15	.28
Computers and Data Processing	.30	.29
Computer Programming, Elementary Fortran IV	.13	.22
Economics, Introductory	.20	.31
E cational Psychology	.22	.28
English Composition	.13	.29
English Literature	.37	.28
Geology	.21	.30
Human Growth and Development	.12	.27
Literature, Analysis and Interpretation of	.05	.31
Marketing, Introductory	.14	.31
Medical Technology Clincal Chemistry	.12	.27
Hematology	.12	.21
Immunohematology	.16	.27
Microbiology	.21	.27
Money and Banking	.15	.29
Psychology, General	.10	.30
Sociology, Introductory	.13	.24
Statistics	.14	.28
Tests and Measurements	.21	.25
Trigonometry	.12	.26
Western Civilization	.05	.30

A review of the values of α and β is somewhat disturbing. While the assumptions of normality and that the within grade group standard deviation is 5 may be somewhat inappropriate, and could account for the large size of α and β to some degree, it seems reasonable that this is not the entire explanation. From this decision theoretic viewpoint, therefore, it appears that the validity of the subject area examinations is of concern in many subject areas.