DOCUMENT RESUME

ED 044 078 HE 001 792

| | |
|---|---|
| AUTHOR | Tetlow, William L., Jr. |
| TITLE | STAG: A Conceptual Approach for Comparing Grades Within and Among Colleges and Universities. |
| INSTITUTION | Cornell Univ., Ithaca, N.Y. |
| PUB DATE | 14 May 70 |
| NOTE | 9p.; Paper presented at the Tenth Annual Forum of the Association for Institutional Research, New Orleans, Louisiana, May 14, 1970 |
| EDRS PRICE | EDRS Price MF-$0.25 HC-$0.55 |
| DESCRIPTORS | *Academic Achievement, Academic Performance, Evaluation, *Evaluation Techniques, *Grade Point Average, Grades (Scholastic), *Grading, *Higher Education |
| IDENTIFIERS | STAG, *Standard Transformation of Academic Grades |

ABSTRACT

There are variables in existing grading systems that limit their effectiveness as generalized indicators of academic performance. For example: (1) the multiplicity of grading systems now in use create interpretive dilemmas for educators and students alike; (2) attempts to standardize encounter resistance, which hinders a common approach; (3) present numerical and quality-point averaging systems fail to account for substantial grading differences that exist among the various departments of an institution; and (4) rank-order academic standings of heterogeneous student groups, based upon grade-point or quality point averages, often give a misleading index of student achievement. The Standardized Transformation of Academic Grades (STAG) approach attempts to solve these difficulties by converting existing quality-point or numeric averages into a normalized and standardized scale. A conversion equation is devised for a particular reference group based upon actual grade distribution data obtained from class rank lists. The resulting "transformed" grade has the advantage of standard degrees of mean value and variability, permitting a more valid interdepartmental comparison. Since the conversion is accomplished post-facto, the necessity to enforce any pre-set grading standard is avoided. (AF)

ED0 44078

STAG: A Conceptual Approach for Comparing Grades Within and Among Colleges and Universities

William L. Tetlow, Jr.
Director, Office of Institutional Studies
Cornell University, New York



American educators have experimented with various types of evaluation and grading systems for almost two hundred years. An historical review suggests that we have almost come full-circle in attempts to symbolize a student's academic performance and achievement. Proponents for grading systems argue chiefly in terms of the necessity for simplified evaluation measures for internal and external use. They also cite benefits of competition and other pedagogical values, and a reasonable degree of measurement reliability. Opponents of grades cite lack of validity and uniformity and claim misdirected motivation, mechanization of learning, stifling of creativity, artificiality, and the protection and encouragement of inadequate teaching. However, attempts to eliminate the assignment of evaluative symbols or descriptions have historically suffered from an inability to withstand efficacious and impressive pressures favoring grading systems. Furthermore, educational research and communication of information concerning the academic performance of students has been hampered by the multiplicity and inadequacy of grading scales presently in use.

## THE PROBLEM

There are several deficiencies of existing grading systems which limit the effectiveness of them as indicators of academic performance. In addition to the problem of incomplete validity there are other deficiencies which have hampered communication and research. Some of these are:

1. *There exist many different types of grading scales currently in use thus creating problems of interpretation and communication.* According to Smallwood[1] the idea of definite grading scales using descriptive adjectives began to appear in American higher education around 1775 and by 1800 there were a variety of scales in use. By the early 1900's many colleges were using the A-to-F system or some variant of it. Miller[2] noted in 1967 that the A-to-F system, together with some special grades and a grade-point average (GPA) calculated on a 4.00 system is largely standardized in American colleges. However, there are many institutions that use quite variant schemes. For example, Rutgers University employs a "reverse" GPA system where 1.000 is the highest average (Distinguished) and 5.000 is the lowest (Failed). Usually, colleges with such atypical grading systems encounter difficulties in explaining and interpreting their system to others; in some cases, the students are at a competitive disadvantage for fellowships and graduate school admissions. Anyone who has had to interpret grades and GPA's from different colleges is aware of the difficulties of comparability.

2. *Most attempts at standardization encounter resistance which limits their universality of use.* An insistence on the use of pre-established grading "standards" or "guidelines" almost automatically encounters resistance on both philosophical and practical bases. Attempts to enforce either an absolute or a relative reference are met with resistance on the grounds of abridgment of academic freedom. Most "post-facto" compensating systems are complex, cumbersome and/or unintelligible to many educators, students, and laymen; their complexity often negates their positive aspects and restricts their utility.

HE001792

Paper presented at the Tenth Annual Forum of the Association for Institutional Research, Roosevelt Hotel, New Orleans, Louisiana, May 14, 1970

3. *Present numerical and quality-point averaging systems fail to take into account substantial grading differentials existent between the various sub-divisions of an institution.* In fact, present averaging and comparison practices are based upon an implicit (and almost always unrecognized) assumption that such differentials do not exist! In a study involving 38 Minnesota colleges Hood[3] concluded that, "This study has shown that the distribution of ability levels and the distribution of academic grades differ considerably among different colleges and types of colleges. Therefore, a particular grade-point-average represents differing levels of academic achievement at different colleges." What should be added to his statement is that this is indeed true even when the colleges are divisions of a single university. Data I have collected from a broadly representative sample of twenty colleges and universities[4] during the last four years revealed substantial divisional differences in grading distributions within a single institution. For example, quality point averages of 2.70 in the Liberal Arts College and 2.20 in the College of Architecture at one institution were the median averages for those divisions and thus represented comparable academic achievement relative to each curriculum. However, performance comparisons are made and statistical aggregations (e.g. fraternity chapter averages) are computed in a fashion which ignores these differences and treats equally all grades of the same numerical value. The "university average" in the example cited was computed as 2.50 and the inference is thus fostered that the Architecture student is *below* average in performance while his compatriot in Liberal Arts is *above* average. The actual situation is that both individuals are performing *equally relative to their respective curricula.*

4. *Rank-order academic standings of heterogeneous student groups when based upon grade-point or quality point averages often give an incomplete and incorrect index of student achievement.* The relationship between GPA's and subsequent class rank do not include the assets of standard degrees of mean value or variability. Thus in knowing only any given student's rank or GPA one does not have a complete picture of the relative performance of the student. When the rankings are based upon group averages which do not take into account the divisional differences mentioned previously they can actually be substantially in error. In a continuation of the previously mentioned example, a fraternity comprised principally of architects will be ranked considerably below its appropriate standing while another fraternity consisting of liberal arts students will be ranked higher than it should be on the basis of the curricular performance of its members (because of the differences in grading distributions of the two colleges).

## THE STAG APPROACH

### Overview

The Standardized Transformation of Academic Grades (STAG) approach is an attempt to resolve the difficulties of educational research and effective communication caused by the aforementioned deficiencies of present grading systems. The STAG approach is used to convert existing quality-point or numeric averages into a normalized[5] and standardized scale. Each conversion (regression) equation is based upon actual grade distribution data for the particular reference group in question and obtained from class ranking lists. The resultant transformed grade has the advantage of standard degrees of mean value and variability thus permitting more valid comparisons and a greater variety of statistical manipulations. Since the conversion is accomplished post-facto the necessity to enforce any pre-set grading standards is avoided. The institution need not change its traditional and familiar grading system but can use the STAG grades whenever inter or intra-institutional comparisons are deemed desirable or necessary. The standardized grade has the added feature of providing information about the relative academic performance of each individual; relative to his curriculum and relative to other individuals.

Detail of STAG Approach

The STAG approach is a relatively simple and straightforward one consisting of the following four operations:

1. *Determine the appropriate reference sub-groups* for comparison or analysis. In most universities each school, college or major division should comprise a separate sub-group. Further sub-divisions may be deemed desirable; for example, the Liberal Arts College students may be sub-divided into Humanities, Social Sciences, Physical Sciences, and Biological Sciences divisions. Sub-divisions on the basis of sex are not relevant for most applications.[5]

   The decision on whether to use one standardized distribution for the entire division (e.g. Agriculture) or to determine the distribution for each class within that division will depend upon the application. Separate distributions for each class will be more accurate than one for the entire division but the latter, if based upon the junior year distribution[6], has the advantage of being a goal referent.[7]

2. *Determine the mean GPA and standard deviation (S.D.) for the reference groups.* For example, if your reference group is junior agriculture students determine the mean GPA and S.D. for all junior agriculture students on the basis of the one year's grades.

3. *Transform the grade to any standardized scale score* by means of the equation:

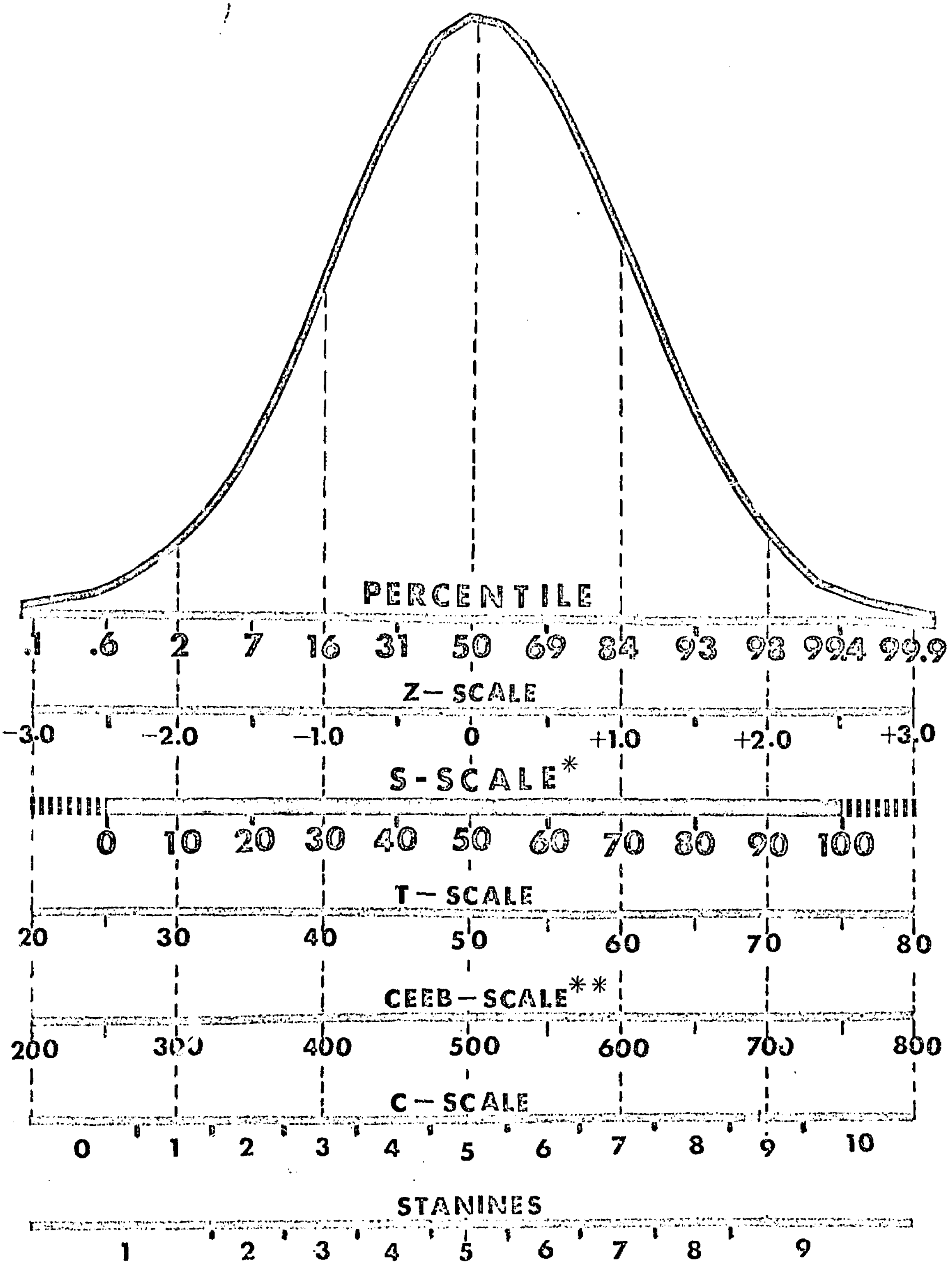
$$\text{Standard Score} = \frac{\text{S.D. of New Scale}}{\text{S.D. of Referent Group}} \times \left(\text{Student's GPA} - \text{Mean GPA of Referent Group}\right) + \text{Mean Value of New Scale}$$

   For example, using the S-scale[8] (see following chart) and the mean GPA of 2.40 and a standard deviation of .60 for junior agriculture students, a student with a 3.00 GPA in that division has an S-score of 70:

$$\text{Standard Score} = \frac{20}{.60} \times (3.00 - 2.40) + 50 = 70$$

   The choice of standard scale will depend upon the particular application in mind. The "T" and "CEEB" scales are familiar to many researchers but the scale limits (practically speaking 20-80 or 200-800) are confusing to many faculty members and students who are not mathematically inclined. The stanine scale suffers from the same problem but has the practical advantage of occupying only one column on computer punched-card records. The S-scale uses a 100 point scale and its limits are more understandable to many people.[9] For guidance and counseling students on grade performance, low scores that result from the S-scale may have valuable "shock" effects.

# Relationship of Several Scales to a Normal Distribution




* Devised for this particular application (STAG)

** College Entrance Examination Board

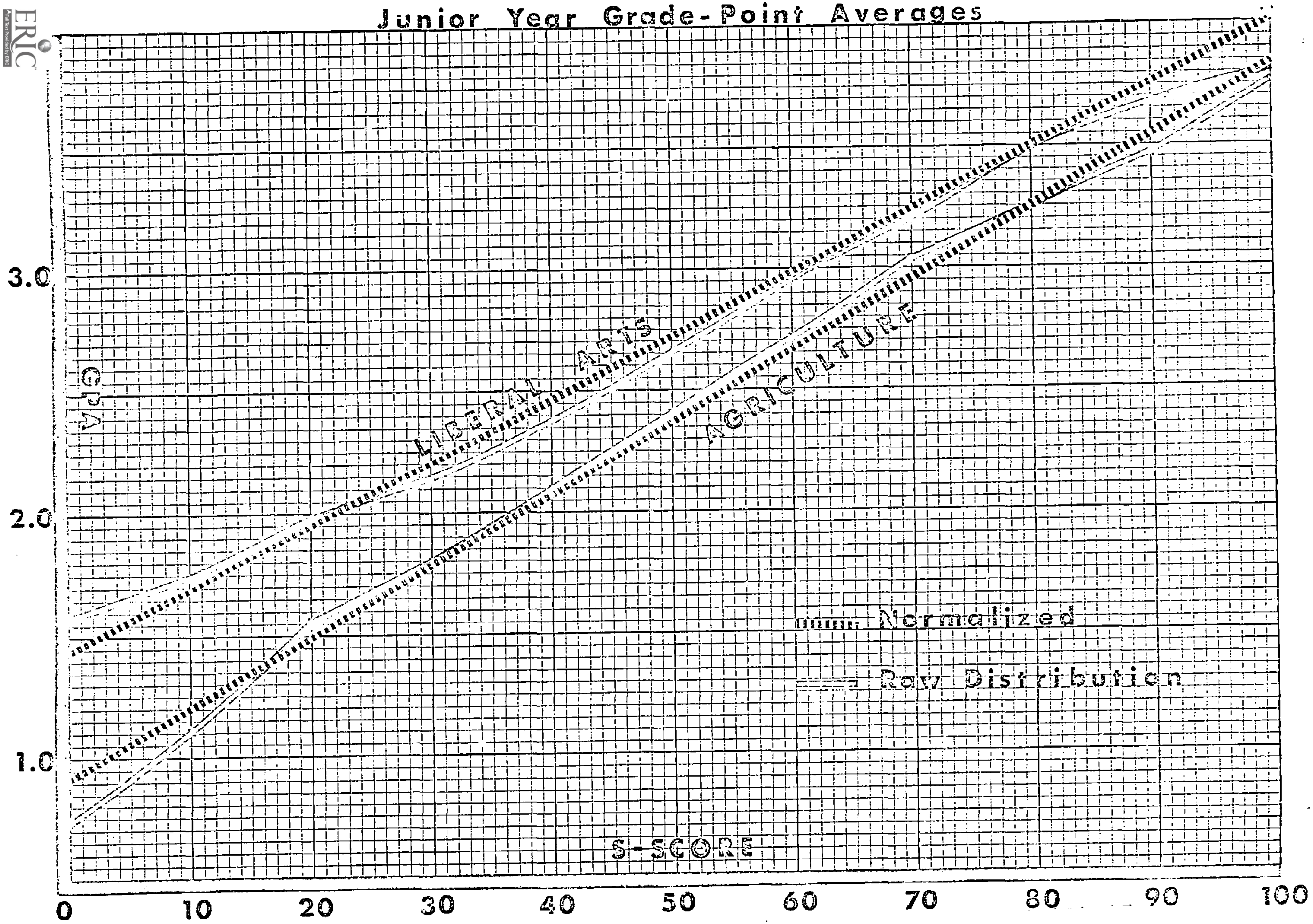
4. OPTIONAL: *Normalize the grading distributions.* Opinions differ regarding the desirability of the normalizing procedure. It is recommended here for two reasons:

   a/ It compensates for possible change variations in the original raw distribution (the curve is "smoothed") and;

   b/ The resultant distribution is more likely to remain stable over a period of years thus avoiding the necessity for frequent re-calculation.

If this option is exercised a plot of the data points should be made on either linear or arithmetic probability graph paper to determine the magnitude of the changes made by the normalizing procedure. An example of the result of the above process is shown in the following graph for two divisions of *a single* university. The graph also shows the best straight line approximation (normalized curve) for each of the two college GPA distributions. The example uses the typical 4.00 quality-point system but the STAG method can be used on any other type (e.g. 100 point, etc.) of numeric scale. The Liberal Arts student who was at the 50th percentile of his class (rank = 268/535) had a 2.68 junior year GPA. The Agriculture student who was at the 50th percentile of his class (rank = 198/395) had a 2.41 junior year GPA. Although both students ranked in the middle of their respective college classes the Liberal Arts student had a GPA that was .27 quality-points higher than the student in Agriculture. Similar differences exist at all points of the grading distribution.

Another way to highlight the variations is to see what a given GPA means in terms of relative achievement in each of the two colleges. A student obtaining a GPA of 2.00 in Liberal Arts stands at approximately the 7th percentile (S-score = 20) of his class which is one and one-half standard deviations below the mean. On the other hand an Agriculture student with a GPA of 2.00 stands at approximately the 25th percentile (S-score = 37) which is about two-thirds of a standard deviation below the mean. Thus the very same GPA represents two quite different achievement records!



Junior Year Grade-Point Averages
GPA
3.0
2.0
1.0
LIBERAL ARTS
AGRICULTURE
Normalized
Raw Distribution
S-SCORE
0
10
20
30
40
50
60
70
80
90
100

## APPLICATION OF THE STAG APPROACH

A brief example of the results of the application of the STAG approach and the different and more complete information it conveys is illustrated in the following table.

### Sample Application of STAG System Conversion

| Name | Division | Quality Point Average (0-4.30 Hi) | Standardized Grade (S-Score) (0-100 Hi) | Rank by Quality Point | Rank by S-Score |
|---|---|---|---|---|---|
| Adams | Architecture | 3.25 | 92 | 4 | 1 |
| Baker | Hotel Admin. | 3.47 | 91 | 3 | 2 |
| Carey | Liberal Arts | 3.70 | 89 | 1 | 3 |
| Damon | Engineering | 3.56 | 85 | 2 | 4 |
| Evans | Agriculture | 2.83 | 65 | 6 | 5 |
| Fisher | Commerce | 3.00 | 65 | 5 | 6 |
| Green | Architecture | 2.00 | 41 | 10 | 7 |
| Hunter | Engineering | 2.05 | 38 | 9 | 8 |
| Irving | Liberal Arts | 2.31 | 34 | 7 | 9 |
| James | Commerce | 2.10 | 27 | 8 | 10 |

In this sample application no student has lower than a 2.00 (C) average on the quality-point system and the conclusion would often be drawn by most students and other observers that all of the students are performing at or above "average". More sophisticated observers might make a mental referent of the GPA to what they believe to be "average" performance for the university. Yet when the averages are converted to the standardized scale it can easily be seen that four individuals have averages below the mean (50) and one student, James, has a score more than one standard deviation (greater than 20 points) below the mean; his average is sufficiently low to justify some concern about his future performance in his curriculum area. The standardized score immediately and readily conveys the information regarding the student's relative class standing *in his division.*

The rank order position of the students changes markedly also; the student with the fourth highest quality-point-average (Adams) has the highest S-score average. The differences are due to the different GPA distributions for each one of the university divisions. In this case, Architecture has a more suppressed distribution than Hotel Administration, Liberal Arts, and Engineering. University-wide ranking on the basis of a standardized grade is more appropriate for many internal applications such as making awards based upon scholastic performance.

On the basis of the converted S-scores Evans and Fisher are shown to be equally ranked. Both students are shown to be three-quarters of a standard deviation (15 points) above the mean averages (50 points) in their respective curricular areas. When ranked by unadjusted quality points the inference is given that Fisher has achieved in a superior manner to Evans; if the criterion is performance relative to the curricular classmates the inference is incorrect. It is impossible to tell each one of them stands in relation to his classmates on the basis of the y-point-average.

## THE STAG APPROACH VS. PRESENT SYSTEMS

The STAG approach transforms quality-point or numeric averages of sub-divisions of a college or university into standardized and normalized grades. The 100-point S-scale has the attributes of a mean of 50 and a standard deviation of 20 points thus permitting ease of statistical manipulation and meaningful resultant comparisons.

An institution can use the STAG approach for inter and intra-university comparisons and since the standardized grade represents the same relative achievement level in any college or curriculum, interpretation and communication can be enhanced.

The STAG method is applied post-facto and designed to supplement existing grading systems; it is not necessary for a college or university to change its grading procedures. The system is easy to understand and apply and is adaptable to both manual or computer application.

Differentials in grading distributions of the various sub-divisions of an institution are compensated for in the STAG approach and similar standardized scores do represent similar achievement in terms of divisional performance.

A standardized score has incorporated in it information pertaining to the relationship of the individual (or group) to the mean performance level; a more complete picture of the performance of a student relative to his curricular area is thus afforded.

Class rankings of groups of students is much more appropriate for many applications when based upon standardized scores than when based on GPA's when the group is comprised of students from many different colleges or sub-divisions of a university.

The STAG method of compensating for divisional grading variations and for evaluating and comparing academic performance should be of assistance to anyone concerned with evaluating scholarship for purposes of educational research, educational-vocational counseling, financial aid, admissions or employment. Use of the STAG method should facilitate inter and intra-university comparisons on either an individual or group basis.

## FOOTNOTES

1. Smallwood, Mary L, An Historical Study of Examinations and Grading Systems in Early American Universities, (Harvard Studies in Education, Vol. 24), Cambridge, Mass: Harvard University Press, 1935.

2. Miller, Stuart, "Measure, Number, and Weight: A Polemical Statement of the College Grading Problem". (Ann Arbor: Center for Research on Learning and Teaching, The University of Michigan, 1967).

3. Hood, Albert B. "A Method of Comparing Student Achievement Levels at Defferent Colleges", Personnel and Guidance Journal, April 1967, pp. 799-803.

4. Amherst, Univ. of Cal. (Berkeley), Bowdoin, Cornell, Univ. of Georgia, Hamilton, Univ. of Illinois, Univ. of Michigan, Middlebury, Univ. of North Carolina, Northwestern, Univ. of Oregon, Rutgers, Univ. of South Carolina, Stanford, Union, Univ. of the South, Univ. of Virginia, Univ. of Washington, Wesleyan

5. The most relevant criterion is usually how the student compares with his curriculum group - not with his sex group. Graduation requirements are not different for the sexes. There may be cases, however, where sex groupings would be appropriate.

6. This choice is somewhat arbitrary but based upon the student personnel literature. Freshmen grades are considered atypical due to adjustment problems by the student; many sophomores fall victim to a "sophomore slump"; seniors attention and performance is often affected by future career plans. The junior (3rd) year is generally considered to be the most stable; the resultant distribution will thus be essentially the most general and realistic.

7. A freshman's grades would convert to a higher score on a freshman-based scale than on a junior-based scale but if the freshman *were to maintain the identical average* the latter score would show him where he *would* stand as a junior.

8. S-scale range 0-100 (Hi); Mean = 50; S.D. = 20.

9. Although it covers less than the full probability range, in this type of application the extremities (which account for only .6% of the cases on each end of the scale) are less important; all scores falling at or above the maximum are equated to 100 and likewise all falling at or below the minimum are equated to 0. The positive advantages of a familiar 100-point scale outweigh the loss of distinction at the scale extremities - a distinction that would imply an accuracy level that in some applications is not warranted and frequently mis-valued. For this same reason (implied and unwarranted accuracy) it is recommended that the S-Grade NOT be carried to any greater distinction than the nearest whole unit! If further refinement is needed some other attribute of the students should be the basis of the distinction.