DOCUMENT RESUME

ED 445 051 TM 031 667

AUTHOR Baron, Patricia A. Bigham
TITLE Consequential Validity for High School Grades: What Is the Meaning of Grades for Senders and Receivers?
PUB DATE 2000-04-00
NOTE 59p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Admission (School); *Admissions Officers; *College Entrance Examinations; Counselors; *Grades (Scholastic); *High School Students; High Schools; Higher Education; Parents; Scores; Secondary School Teachers; Test Results; *Validity

ABSTRACT

The two major selection criteria in the college admissions process are high school record and scores on college entrance examinations. In recent years, concerns have been raised about the validity of college entrance examination scores and high school grades. This study examined the congruity of the meaning of grades between those who determine the grades (the senders) and those who in some sense use or interpret the grades (the receivers). A questionnaire was administered to 5 distinct groups: 60 high school teachers, 48 high school students, 41 parents of high school students, 115 high school counselors, and 46 college admission staff members. The results indicate that senders and receivers, in large part, agree about what grades comprise. There is disagreement among teachers, students, and parents about the frame of reference for grading (curve versus fixed). When asked to predict Scholastic Assessment Test (SAT) I scores based on one of four versions of a high school transcript, parents and students estimated higher SAT I scores than the other groups; however, no differences related to the gender of the respondent or the gender of the student. Overall, the results indicate that while there is some disagreement regarding the relative importance of grade components, there is not a clear understanding of the reference scale for high school grades, and parents and students in particular hold some beliefs that are not in concert with the education community. By investigating beliefs held by groups of senders and receivers regarding the composition of grades, the underlying scale or meaning of grades, and the expected relationship with other measures of ability, such as aptitude test scores or college performance, some understanding of the consequential validity of high school grades has been gleaned. To the extent that messages sent and received are incongruous among these groups, the validity of grades for the purposes of the users of the grades is in question. An appendix contains the study questionnaire. (Contains 13 tables and 72 references.) (SLD)

CONSEQUENTIAL VALIDITY FOR HIGH SCHOOL GRADES: WHAT IS THE MEANING OF GRADES FOR SENDERS AND RECEIVERS?

Patricia A. Bigham Baron
Rutgers University Graduate School of Education

The author would like to thank Dr. Jeffrey K. Smith for his guidance throughout, and Educational Testing Service for support. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 2000

Abstract

The two major selection criteria in the college admissions process are high school record and scores on college entrance examinations. In recent years, the validity of entrance examination scores has been scrutinized. Concerns regarding the validity of high school grades have been raised as well. Since the introduction of the concept of consequential validity presented by Messick in 1989, any investigation of validity includes a focus on users of scores and the consequences of score use. The purpose of this study was to examine the congruity of the meaning of grades between those who determine the grades (the senders) and those who in some sense use or interpret the grades (the receivers). A questionnaire was administered to five distinct groups: high school teachers, college student teachers, parents of high school students, high school counselors and college admission staff members. The results indicated that senders and receivers in large part agree about what grades comprise. There is disagreement between teachers, students and parents about the frame of reference for grading (curve versus fixed). When asked to predict SAT I scores based on one of four versions of a high school transcript, parents and students estimated higher SAT I scores than the other groups, however no differences related to gender of the respondent or gender of the student were found. Overall, the results indicate that while there is some agreement regarding the relative importance of grade components, there is not a clear understanding of the reference scale for high school grades, and parents and students in particular hold some beliefs which are non in concert with the education community. By investigating beliefs held by groups of senders and receivers regarding the composition of grades, the underlying scale or meaning of grades, and the expected relationship with other measures of ability such as aptitude test scores or college performance, some understanding of the consequential validity of high school grades has been gleaned. To the extent that the messages sent and received are incongruous among these groups, the validity of grades for the purposes employed by these users of grades is in question.

The study of validity underwent a dramatic change with the publication of Messick's (1989) landmark article, which presents the concept of consequential validity. Consequential validity gives more weight to the use of the score and the consequences of its use than earlier concepts of validity. Thinking about grade validity in this light, it is necessary to consider the uses and the users of grades, which in the case of high school grades includes parents, students, high school counselors, college admission staff and teachers. Until recently, research efforts concentrated on the construction of grades by teachers, the grade senders, with less importance placed on what grades mean to the receivers (Cizek, Rachor, and Fitzgerald, 1995; Ekstrom, 1994; Hoge and Coladarce, 1989). The educational measurement community recommends basing grades on achievement-related variables, and most contend that a grade should reflect achievement status at a particular point in time (Brookhart, 1993, 1991; Frary, Cross and Weber, 1993; Manke and Loyd, 1990; Stiggins, et al., 1989; Gronlund, 1985). In practice, teachers include a number of nonachievement factors in grading decisions, such as impressions of ability, student effort and motivation, (Cizek, Fitzgerald, and Rachor, 1995; Cizek, Rachor and Fitzgerald, 1995; Manke and Loyd, 1990, 1991; Stiggins et al., 1989) and the interaction of gender and student effort (Manke and Loyd, 1990; Griswold, 1993). Other factors which are somewhat less objective, although still considered achievement factors, are also included, such as homework and daily assignments, (Manke and Loyd, 1990, 1991; Stiggins et al., 1989), improvement, (Manke and Loyd, 1990, 1991); and class

participation, (Cizek, Rachor and Fitzgerald, 1995). No systematic study of grade users has been done to determine what their interpretations of grades are, and how they vary one from another. This study will include teachers in their roles as providers and users of grades and compare their interpretations with receivers of high school grades in an attempt to provide insight into the various meanings and interpretations of grades, and hence, their validity.

This study will specifically ask the following questions to evaluate the validity of grades:

(1) Are there differences in the meaning of grades for the following groups: teachers, students, parents, high school counselors, and college admission staff, as indicated by what they believe are the factors that should comprise grades? Are some of these factors considered more important than others? What do these groups believe should be used as a point of reference for grading?

(2) For these groups, are there differences in what they believe are the factors that teachers actually use to comprise grades? Do they believe some of these factors are considered more important than others?

(3) Do these groups differ in a systematic fashion in how they interpret a transcript of grades from a hypothetical high school student?

The outcome of grading is that grades are interpreted and used by various groups. In Messick's model of consequential validity (1989), the four facets are linked through construct validity. For each incidence of use and interpretation of grades, the relevance, utility, value implications and social consequences must all be evaluated in order to determine the degree to which grades are valid.

High school grades are provided by high school teachers as measures of high school performance. They are interpreted and used by students, parents, high school counselors and college admission staff to assist in decision making regarding the future of the student. The degree to which the actions taken and interpretations made based on grades are appropriate is the defining characteristic of consequential validity. This study evaluates both the message sent by the teacher when assigning grades and that received by various users of grades. To the extent that these messages are incongruous, the validity of grades for the purposes employed by these users of grades is in question.

Study Design

Participants in this study consisted of five distinct groups: 60 high school teachers, 48 high school students, 41 parents of high school students, 115 high school counselors and 46 college admission staff members. The high school teachers and students were voluntary participants from a middle-class regional New Jersey high school. The parents were from three middle-class high school regions in New Jersey. High school counselors and college admission staff were attendees at one of the 1998 annual regional meetings of the College Board.

Each participant was asked to respond to a questionnaire, the construction of which was guided by previous research and the results of a pilot study. Teachers had questionnaires delivered to their school mailboxes and returned them to a central location in the school. Students received questionnaires in their courses and were asked to complete them at home and return them in class the following day. High school counselors and college admission staff were asked to fill out a questionnaire during the

1998 annual regional meeting of the College Board. Parents were requested to respond and return the questionnaire to the office of the principal or the regional education office.

The questionnaire included five measures, designed as follows: The first measure asked participants to make a determination of an ideal composition of factors that make up a grade (effort, tests and quizzes, etc.). Participants were asked to assign a percentage to each of eight factors listed (see Appendix for a copy of the instrument). The second measure was analogous to the first, but asked the participants to give their opinion as to how teachers actually make such determinations. The third measure asked participants to mark on a 1-10 scale the degree to which they feel grades should be norm-referenced or criterion-referenced. For the fourth and fifth measures, an experimental design was employed. Two hypothetical high school student transcripts were created. The students were from the same high school, but had different grades, and somewhat different courses. In one transcript the student was doing fairly well, having grades of As or Bs, and one Advanced Placement course in Biology. The other transcript had lower grades, mostly Bs or Cs. To determine the effect of the student gender, each type of transcript had a male and female version. Specifically, the name of the student was either Ellen Smith or Jacob Smith. The fourth measure asked the participants to predict an SAT-V and SAT-M score for the student represented by the transcript. The fifth measure asked an open-ended question relating the transcript to application to an average public state college. The written responses to this measure were used to elucidate some of the perceptions brought to the study by the respondent groups.

Analyses

For the first two measures involving different emphases on the factors contributing to a grade, multivariate analysis of variance (MANOVA) was used to examine the data. The independent variable was group, the five levels were admissions staff, counselors, parents, teachers and students.

The first MANOVA performed on the factors contributing to a grade was done using a Euclidean distance as the dependent variable. In this analysis, the Euclidean distance (E) was calculated to determine the within group difference across all 8 factors of grade composition on questions 1 and 2. A difference was calculated for each respondent between the response to the first survey question, which asked "Please attribute 100 points to indicate what YOU think each (of the factors) SHOULD be worth in determining a grade" and the response to the second question, "Please attribute 100 points to indicate what you think MOST TEACHERS use to determine a grade". The formula for E, defined as the degree to which responses to Q1 and Q2 are different, follows:

Euclidean Distance = $\sqrt{D_1^2 + D_2^2 + D_3^2 + D_4^2 + \ldots + D_8^2}$, where $D_1^2 = (Q_{11} - Q_{21})^2$, $D_2^2 = (Q_{12} - Q_{22})^2$, etc., and Q_{11} is the response on the first factor (Class Participation) when answering what should be the weight of each factor, and Q_{21} is the response on the first factor when answering what is the weight you think most teachers use. Thus, D_1 is the difference for each respondent on the first factor between what they think a factor should be worth in grading, and what they think a factor is weighted by most teachers.

Based on the results which showed high internal consistency on the part of the teachers responses, the second MANOVA used responses by the teachers to the first survey question, which asked "Please attribute 100 points to indicate what YOU think each (of the factors) SHOULD be worth in determining a grade"as the dependent variable for teachers. The intent was to obtain responses from teachers that represent teachers practice. For all other groups, responses to the second question, "Please attribute 100 points to indicate what you think MOST TEACHERS use to determine a grade" were used as the dependent variables.

For measure 3, which asked whether grades should be norm- or criterion-referenced, an ANOVA was performed with group as the independent measure. For measures 4 and 5, there was an experimental design placed on the study.
The basic design of this part of the study was a 5 x 2 x 2 (Group x Gender of target x Level of transcript) factorial design. Group membership was a fixed variable for the five groups. The analysis performed was a MANOVA with SAT I Verbal and SAT I Math as the dependent measures. Additional analyses were completed as necessary to understand the effects. Measure 5 asked for a written response regarding the predicted college performance of the student. Due to a lack of response to the open-ended question, analysis of this measure could not be completed.

Results

The survey instrument was designed to guide answers to three research questions:

1) Do senders and receivers of the message sent by grades perceive the same message?

2) Do the five groups of grade users differ in their grading preference between a criterion-referenced or fixed reference versus a sliding scale or "grading on the curve"?

3) Are grades interpreted differently by different groups? Is there a different prediction based on grades for girls and boys? If grades are interpreted differently, is it a function of group membership, the level of the grades, the gender of the student, or a combination of these factors?

Survey responses were collected on five groups: college admissions officers, high school counselors, high school students, high school parents, and high school teachers. Four versions of the survey were spiraled so that an equal number of responses to the experimental transcript could be collected. The type of transcript that each person in any of the five groups might have received was based on a 2 x 2 design, gender x level of grade. As can be seen in Table 1, the spiraling was largely successful. Close to equal numbers of responses were collected for the four versions for all but the teachers. For teachers, more high grade female transcripts were returned than for the other three versions. It is believed that this was caused either by a printing error, or distribution of the transcripts at the high school, but is not due to the number of missing responses being related to the spiraling. Also, across groups, considerably more high school counselor survey responses were returned. This is known to be due to the number of available respondents. For the overall sample of 309 cases, there were twice as many female respondents as males.

As can be seen in Table 2, the subject area in which the largest group of teachers in this sample work is English, followed by science, math, special education and history.

For the group in this study, the mean years experience teaching was equal to 17.14, with a standard deviation equal to 9.01. The number of years teaching ranged from 1 to 37 years but was highly skewed, with most of the teachers having a good deal of experience.

Senders' and Receivers' Beliefs about Factors Contributing to a Grade

The first analysis of these data was an examination of the responses to two questions posed to the senders and receivers of grades: Question 1 -- What do you think the weight for the eight factors listed below should be; and Question 2 -- What do you think most teachers actually do? The response requested was a distribution of 100 points to eight factors: (a) Class Participation, (b) Attendance, (c) Homework, (d) Improvement (from the previous year or semester's performance), (e) Tests (and quizzes), (f) Papers, (g) Effort, and (h) Growth (during this semester/year).

Across all five groups, the rank order of the eight factors' contribution to a grade quite consistent, as can be see in Tables 3 and 4. All groups reported that tests are the most important factor, followed by papers, weights ranged between 50 and 60 oints for these two factors for both questions. Of the remaining 40 to 50 points, about 20 to 25 points were applied to homework and class participation, and for all groups these were the two factors listed next in importance, although effort was a very close third for some groups. The remaining apportioning of points was done among effort, attendance, growth, and improvement. The ordering of these four factors indicates more variability across groups than for the first four factors.

In this initial review of the responses, the similarities are striking. For all groups, the emphasis on achievement variables is far above any other factors, and this is the case

when groups were asked what they think most teachers do and what they think the weighting should be. In addition to the similarities, there are some differences in responses by group within each question, as well as across the two questions. These differences will be described next, by question.

Table 3 shows the means on responses for all five groups to the first question: What do you think each factor should be worth? The order of the factors is fairly consistent, except for students. According to college admission officers, parents and teachers, the factors should be ordered, from most important to least, as Tests, Papers, Homework, Class Participation, Effort, Attendance, Growth, and Improvement. Counselors were in general agreement with these three groups, except that they ranked participation in class above homework. Students, however, ranked the last three factors differently. From highest to lowest, students ranked the first five factors the same as counselors: Tests, Papers, Homework, Class Participation, Effort, but ordered the last three factors, Improvement, Attendance and Growth, indicating that students placed improvement higher than attendance and growth.

On the second question, how do you think teachers determine grades, all groups except parents rank order the factors from highest to lowest as follows: Tests, Papers, Homework, Class Participation, Effort, Attendance, Growth, and Improvement, as shown on Table 4. Parents ranked everything the same except for Attendance and Growth, which they switched, indicating that they thought most teachers place more emphasis on growth than attendance when determining a grade.

It was expected that the responses by teachers would be about the same to both questions, given the assumption that the teachers responding believe that most teachers

practice what they believe they should do in terms of weighting these factors into grades. Looking at both tables, it can be seen that this is true for teachers, who rank ordered the factors the same way in both versions of the task.

What Groups Think Grades Should Be Versus What They Believe Occurs in Practice

The first multivariate analysis of variance procedure was performed on data from a combination of the responses to what senders and receivers said should be done and what they believe teachers actually do. A Euclidean distance measure was used to assess the difference in the responses to the two questions; What weights should be placed on components to grades (question 1) and what do most teachers actually do when they assign grades (question 2). A difference for each respondent was calculated on each of the eight factors, i.e., difference on class participation = weight provided as a response to Q1 for class participation minus weight provided as a response to Q2 for class participation. The Euclidean Distance variable, E, was defined as the degree to which overall responses to Q1 and Q2 are different.

A one way analysis of variance was performed on the dependent variable, E, to determine the group variation on how differently grades are perceived to be composed by teachers from what receivers would do if they composed grades. This distance, E, was significantly related to the independent variable, Group. As expected, teachers were the most consistent in the responses to the two questions. Table 5 illustrates the differences among the five groups thinking on what should be done in grading and what they believe most teachers do. Significant mean differences were found between counselors and teachers. This result indicates that the difference between what counselors believe grades

should comprise and what counselors believe teachers do to construct grades is significantly greater than the difference for teachers, who are more consistent in their beliefs about what should be done versus what most teachers do.

How Different are Receivers and Senders on Grade Composition

Given the consistency of responses by teachers on the questions, "What should be done?" and "What do most teachers do?" the two additional multivariate analysis of variance (MANOVA) procedures were performed using what teachers thought should be done in the construction of grades as the teacher response and receivers (admissions officers, counselors, students and parents) beliefs about what teachers do in practice. This was operationalized by using the teachers response to the question: "Please attribute 100 points to indicate what YOU think each SHOULD be worth in determining a grade", and the receivers response to: "Please attribute 100 points to indicate what you think MOST TEACHERS use." The eight dependent variables were Class Participation, Attendance, Homework, Improvement, Tests, Papers, Effort, and Growth. The independent variable was group.

The multivariate results indicated a significant group difference. For the univariate analyses, there were significant group differences on six of eight of the factors: attendance, homework, improvement, tests, papers, and effort. Responses on class participation and growth did not result in a significant difference due to group. These statistically significant results were a function of differences between senders and receivers, and differences among receivers. The results will be presented separately, first to address the issue of the message sent versus the message received. Secondly,

differences among receivers in the perception of what message is being sent will be reported.

Table 6 illustrates where significant differences between senders and receivers occur on attendance, papers, effort and class participation. Compared to teachers, students overestimate the degree to which papers count toward a grade. Parents believe that both attendance and effort count less than teachers suggest they should count. Counselors reported that teachers use effort less than teachers report it should count in grades. Students think that teachers are using class participation less than teachers report it should count. The differences between senders and receivers of grades address the issue of the whether the message being sent is being understood as intended. In addition, there is the issue of whether differences exist among receivers in interpreting grades.

When a grade or grades are received, are all receivers getting the same message? In the analysis comparing what the receiver groups perceive to be the practice of teachers, there is some disagreement among the non-teacher groups, as illustrated in Figure 1. On the other hand, all groups believe tests should carry the most weight in grade development, and place growth and improvement near the bottom of the list. As Table 7 illustrates, significant differences among receivers were found for five of the eight factors: Attendance, Homework, Improvement, Papers and Effort. Students think that teachers use homework less in grading than other groups. This was significantly different from what admissions officers and counselors reported. In contrast, students believe that papers are contributing more to a grade than other groups. Admissions officers and counselors report they think teachers are using papers less than students believe the factor contributes. On the other hand, admissions officers believe that attendance and effort are

being used significantly more than parents perceive these factors contribute to grades. While no group attributed a large percentage of weight to improvement, admissions officers believe teachers place a higher weight on improvement, and effort, relative to other groups. This difference was found to be significantly higher statistically than counselors and parents.

By and large the groups in this study were similar in their understanding of what teachers were using in the composition of grades, but these results indicate that there were some differences in the way receivers perceive grades to be constructed. Taken in combination with the results that compared teachers ideal grading construction with what is perceived to be the message sent, group differences can be seen on most of the factors.

Overall, the analysis of factors in grading indicate that some aspects of what is communicated by grades are unclear, but by and large there is great agreement about what constitutes a grade. Where differences exist, the difference may be based on an underlying belief of some groups about the meaning of grades, or there may be some other basis for the misunderstanding, but why the differences exist is not addressed in this research. What can be evaluated from these data is the difference between what these groups believe should be the order and degree of importance of these components in grading and how they believe teachers are using these elements.

Grading on a Curve or Grading using a Fixed Reference

In addition to looking at the composite factors in grades, this study of grading practice included the reference scale definition as part of the criteria used in grades. To examine beliefs about the criteria of a grading scale, all groups were asked to respond to a 10-point scale, where they had to circle the number which best represented their feeling

about how grading should be approached. On this scale, one was defined as equal to grading on the curve, ten was defined as equal to a criterion-referenced or fixed standard, and five and a half was labeled as a mix of the two. The analysis concerned the issue of whether groups differed in their response regarding grading on a curve versus a fixed standard. A one-way analysis of variance (ANOVA) was performed using the five groups of grade users as the independent variable and the response to the curve versus fixed standard question as the dependent variable.

The results of the ANOVA indicated a significant group difference in means between teachers and parents, teachers and students, admissions officers and students, and counselors and students. As shown in Table 9, teachers response was the highest mean score, indicating that teachers believed more than other groups that a fixed standard such as A = 90 to 100, B = 80 to 89, etc. should be applied in grading. Students were the group who believed most strongly that grading should include a distribution of grades, such that approximately the same number of As, Bs, Cs, etc. would be produced. On the scale used, one was equal to "grade on the curve"; the students mean response is approximately equal to what was labeled "mix of the two".

Previous research (Frary et al., 1993) indicated that teachers may differ by subject area on what point of reference or standard should be used in grading. Specifically, the group identified as "largely math and science teachers" was reported to stand out as different in their beliefs about the point of reference for grading. The 1993 study found this group of 48 teachers to be largely in agreement with a ranking approach. The data in the current study were grouped so that two of the subject areas could be analyzed; sample sizes were too small to do any further analyses. Using 13 English teachers and 17 math and science

teachers to determine if the response to grading standards differed by subject area, a one-way ANOVA was performed. A significant difference on grading on a curve vs. a fixed reference was not found.

Analysis of the Experimental Design on the High School Transcript

The last set of results is based on predictions of SAT I Verbal and Math scores for a hypothetical student's transcript. This transcript was one of four possible varieties. The student is either named Ellen or Jacob Smith, and had either fairly good grades in a somewhat challenging course load, including one course in Advanced Placement Biology, or somewhat poor grades with a less challenging curriculum. The first analysis was conducted using the full sample; MANOVA was used to evaluate group differences, where group had five levels, admissions officers, counselors, parents, students and teachers. A second analysis was performed on the education professionals only. The sample for the second analysis consisted of teachers, college admission officers and high school counselors. MANOVA was used to study the effect of gender of the respondent on prediction of SAT I Verbal and Math scores for the educational professional sample.

Prediction of SAT I V and SAT I M by Five Groups based on a High School Transcript

The first MANOVA was a 2 x 2 x 5 (Transcript level x Gender x Group) between-subjects design, performed on the two dependent variables: SAT I Verbal (SAT I V) and SAT I Math (SAT I M). The independent variables were level of transcript (high grades and low grades) and gender of transcript (male and female) and group (admissions officers, counselors, parents, students and teachers). The results of the analysis revealed two significant main effects. A significant transcript level effect was found, which is not

surprising. It was expected that the high grade transcript would produce higher predicted SAT scores, and vice versa.

The second significant effect was a group difference. This difference was across level and gender of transcript, which indicates that some group or groups were predicting SAT scores higher or lower than another group or groups, even given the difference in the grades on the transcript. It is important to note that overall the predictions made based on the transcripts were quite reasonable. Respondents to the questionnaire were not found to predict SAT scores for the high grade transcript in the 700s, and for the low grade transcript in the 300s. Rather, the high grade transcript prompted higher SAT M scores than SAT V scores, and the low grade transcript lower SAT M scores than SAT V. This seems reasonable given the grades in various courses on the two types of transcripts.

The results of two ANOVA procedures revealed three main effects. On SAT I V, the group effect and the transcript level effect were significant. On SAT I M, the effect for transcript level was significant. As expected, the transcripts with high grades produced higher estimated SAT I V and SAT I M scores than the transcripts with low grades, as can be seen in Tables 10 and 11. In addition, significant group differences were found between college admission officers and both the student and parent groups on estimating SAT I V across transcripts. This finding indicates that there was a tendency on the part of parents and students to estimate higher SAT I V scores than college admission officers. The lack of an interaction effect, however, indicates that there was not a tendency to give higher or lower scores to the same level transcript when gender was varied, nor was there a tendency for some groups to predict higher scores for either gender on either SAT I V or SAT I M.

Beyond predicting SAT scores, respondents were asked to respond to an open-ended question about the student represented by the transcript. The question asked was whether it would be a good choice for this student to apply to the middle level college in the state college system. Most respondents left this question blank, disallowing a quantitative analysis of these data. Descriptions of some responses, and trends for some groups are included in the discussion section.

The Relationship of Gender of the Respondent to Prediction of SAT I Scores

The next analysis looked at the effect of the gender of the respondent on the prediction of SAT I scores. Similar to the previous analysis, this MANOVA used the predicted SAT I scores as the dependent variables, but this inquiry did not include a group variable. Only the three educational professional groups (teachers, admissions officers and counselors) were included in the sample, and these were collapsed. This was done in order to produce a large enough sample to include gender of the respondent as an independent variable. The sample of education professionals was comprised of 228 respondents, 63% female and 37% male. As can be seen in Table 12, the male/female distribution for admissions officers was close to 50/50, but for counselors and teachers the number of female respondents was larger, with 64 and 72 % accordingly.

A multivariate analysis of variance was performed to examine the prediction of SAT I V and M scores by this group. The analysis revealed only one main effect, grade level of the transcript, was significant. The stepdown analysis of variance for SAT I V on the educational professional group indicates a significant difference in the prediction of the score. The factors considered in this analysis were level of the transcript, gender on the transcript, and gender of the respondent. None of the interaction terms indicated a

statistically significant difference. Table 13 illustrates the differences in the mean SAT I scores predicted by this group for each variant of the transcript. ($p > .10$). Of the main effects, neither gender of respondent nor the gender on the transcript were found to be contributing to the overall group difference ($p > .10$). Only the main effect of the level of the transcript was statistically significant, $F(1, 212) = 56.93$, $p < .0001$.

Findings for SAT I M on the education professional group are similar to results for SAT I V. The effect of level of the transcript is significant, $F(7, 212) = 40.94$, $p < .0001$, but no other main effects, nor any interactions, were significant ($p > .05$). Based on the analyses done on the educational professionals in this study, there is no evidence that males and females predict statistically different SAT I scores for the students provided in this questionnaire.

Discussion

This study looked at the consequential validity of grades, evaluating the perspectives of the senders of grades, teachers, and the receivers of grades, students, parents, high school counselors and college admissions officers. Three aspects of consequential validity of grades were evaluated: the composition of grades, the scale on which grades are based, and how grades relate to perceptions about students' performance on standardized achievement measures. Grade composition was analyzed using eight factors, selected based on earlier work as the most essential factors in grading, and groups were asked to weight each in importance. Group differences in beliefs about grading standards were evaluated by allowing participants to choose between grading on a curve and using a fixed standard. Finally, the use of grades to make judgements about students was considered in two ways: groups were asked to estimate scores on the SAT I Verbal

and Math tests based on a student's transcript, and were asked to consider the college application process for the same student. Group differences in these three components of grading were evaluated to assess the consequential validity of grades.

Results of this study

The results of the analysis on grade composition done in this study contribute to the research in an area in which a good deal of work has been done. Prior to this study, work in this area concentrated primarily on teachers, students and parents. The present research considered counselors and admissions officers as well. Among the five groups in this study, four groups of grade receivers, and the teachers who send the grades, there is a good deal of agreement regarding the importance of achievement-based factors vs. non-achievement based factors in the construction of grades. The message teachers send with grades and the message received by students, parents, and high school counselors and college admissions staff is similar. The rank ordering of eight factors was close to the same across all five groups. Senders and receivers rank achievement factors highest in grade composition. Responses for all groups indicated that tests and papers should contribute more than half of the weight in grade composition. All five groups believe that most teachers use tests and papers in this way when determining grades, and that homework and class participation are the factors that should be and are believed to be the next biggest contributors. Affective factors received the least weight.

There are some differences in the message sent versus the message users are receiving. Teachers report that tests and papers, followed by homework and class participation, are considered in the construction of grades. The next factor considered by teachers is effort. Students also said that the contribution of tests and papers are most

important, but give significantly more weight to papers than teachers do. This difference is made up for by class participation; students believe that participation carries less weight than do teachers. Parents report that they believe effort and attendance to be less a part of grades than is reported by teachers. Students are in agreement with high school counselors on the degree to which effort contributes to a grade, but both of these groups places less emphasis on effort than teachers say it is worth. The difference between the message sent and received appears to be found among students and parents, and to some degree high school counselors. Students are interpreting grades to be based more on papers and less on class participation than the intended weighting reported by teachers, and parents underestimate how much attendance and effort is a part of the grade. Counselors are also underestimating the contribution of effort. It is possible that the higher weight placed on papers by high school students is due to the fact that the questionnaire asked respondents to consider a course in U.S. History. Students placing higher weight on papers for a course in U.S. History, rather than a course in math or science, may be an artifact of the questionnaire design. In addition, it may also be the case that students focused more on this aspect of the question than teachers, who may have been more likely to think of the courses they teach.

There is some disagreement on the emphasis on achievement versus non-achievement factors. On the average, teachers reported that 16% of a grade should be based on the four non-achievement factors; improvement, growth, attendance, and effort. Among receivers, the belief reported was that teachers are developing grades that use these factors for between 7% and 21% of the grades. College admissions officers placed more weight on the non-achievement factors than every other group.

In addition to the differences between the senders and the users of grades, there is some disparity among the receivers on how teachers grade. These differences are not restricted to non-achievement factors; five of the eight factors in this study were found to be significantly different for the users. On all five of these factors, the college admissions group differed from at least one other receiver group; parents and counselors differed from the admissions group on the amount of emphasis on affective factors, students on homework and papers.

In addition to perceptions on message sent and received, respondents indicated what they believed grades should comprise. High school counselors reported that improvement, growth and class participation should contribute more than they believe teachers weight these factors. Correspondingly, they believe the contribution of homework and tests should be less.

The second aspect of grading that this research considered was the degree to which groups are in agreement regarding the standard upon which grades are based. Findings regarding the use of a curve versus a fixed standard in grading indicate that there are group differences among users. Teachers believe most strongly that there should be a fixed standard for grading; students were closest to the idea that grades should be on a curve. No group indicated that grading on a curve should be used extensively. Significant differences were found between teachers and both parents and students; parents and students responded that there should be a mix of the two methods. Also, college admissions and high school counselor groups response were significantly different from students.

Finally, this study analyzed responses to a question which asked senders and receivers of grades to evaluate a transcript of a girl or boy student and estimate what this student was likely to achieve on the SAT I Verbal and Math tests. A follow-up question was also asked, where an assessment of the student's plans for college was asked for, based on the same student's transcript.

When asked to predict SAT I Verbal and Math scores based on the transcript of a student, there was a group difference in predicting SAT I Verbal. Students and parents estimated higher scores than college admission officers. No difference in prediction was found between senders and receivers on the basis of gender of the student represented by the transcript. Responses to the open-ended question regarding a college choice for the student described in the transcript suggested that there are some differences in how these groups view the process. Teachers stayed closest to the actual information provided in the transcript, i.e., grades and courses, when providing a rationale for the decision. Other groups were more likely to include opinions about other factors entering into the decision, such as whether or not the level of motivation of the student was sufficient, and how extracurricular activities play into the college admission process. The degree to which these considerations were differentially present for the receivers could not be evaluated quantitatively given the question posed, but is an area for future research.

Examining Current Findings in the Context of Related Literature

The findings in the present study generally indicate that grades are based primarily on achievement factors. However, findings also confirm earlier research that teachers to some degree include non-achievement factors such as effort, improvement, growth and attendance in grade composition (Cross and Frary, 1999; Cizek, Fitzgerald, and Rachor,

1995; Cizek, Rachor and Fitzgerald, 1995; Manke and Loyd, 1990, 1991; Stiggins et al., 1989). The comparisons made in this study indicate that teachers and grade receivers largely agree that a mixture of achievement and non-achievement factors is being used in determining grades, confirming recent findings that students and parents see grading as comprising a mixture of achievement and non-achievement information (Cross and Frary, 1999). In this study, the proportional contribution of achievement and non-achievement factors was evaluated, which had not been done previously, and this clearly indicated that achievement factors are being given the predominance of weight. It was also found that most receivers report the non-achievement factors to be contributing less to grades than teachers report. Viewed in the context of consequential validity (Messick, 1989) these findings provide evidence that grades are being interpreted similarly by senders and the receivers in this study. College admission officers and high school counselors, parents and for the most part students weight similarly the factors in determining grades.

Another somewhat less firm finding also provides some evidence of construct validity. Teachers are the most consistent group in responding to two questions, "What should be the weights on grading factors?" and "What to most teachers use as the weights on grading factors?" There is some disparity between what teachers say is done and what teachers say they should be doing; teachers reported that non-achievement factors should be worth more than they think teachers are currently weighting non-achievement factors. Earlier work did not address the issue of the proportion of a grade attributable to various factors, nor were comparisons made among all of these user groups.

The results on what should be included in a grade suggest that the receivers would like something other than achievement factored into the grade, which may have some

influence on teachers' actual practice. Teachers may be including these non-achievement factors into grading to reach the goal of fairness (Manke and Loyd, 1991; Pilcher, 1989), which is high on the list of criteria for a grading policy for both students (Loyd, Nava, and Hearn, 1991) and teachers (Loyd & Manke, 1990, 1991).

Another finding worth noting is the difference for students in how much weight papers are given; students reported they believe this factor is weighted higher than teachers report is the case. Students may perceive the amount of work associated with papers as higher than what teachers perceive, and therefore students place more weight on this factor. Previous research lends partial support for this theory, finding that both teachers (Brookhart, 1993) and students (Burton, 1983) perceive grades as equal to pay for work performed by students. In the open-ended responses in the present study, students indicated awareness that the process of college admission is dependent on a number of factors, but continually mention effort or hard work. One student wrote about the transcript with high grades, "Yes, she should do well because she did well in HS, and she is probably a studious student so she will work hard." Another student said of the transcript with low grades, "She won't fit in well. She'll have to work fairly hard to keep up and do well." And of the same student transcript, "(She will do well), because she has an A-B average, but it also depends on ambition, extra curricular, etc." In these open-ended responses, the inclusion of effort in the construction of grades is being implied.

Previous research (Pilcher, 1994; Griswold, 1993; Loyd, Nava and Hearn, 1991) indicates that teachers, parents and students all believe that effort should be included when developing grades; however no direct comparison as to how much effort should be worth was made. In addition, earlier studies (Cizek et al., 1995; Manke and Loyd, 1990,

1991; Stiggins et al., 1989) indicated that students and teachers agreed that improvement should be included in grades. The results from the present study confirm these findings, but indicate that for all groups, the relative contribution in determination of grades of both effort and improvement is perceived to be small.

The other critical component in determining grades is the scale on which grades are placed. The results of earlier work are mixed (Robinson and Craver, 1994; Frary et al., 1993; Stiggins et al., 1989); mostly it appears that teachers are sometimes using a criterion-reference, and neither parents nor students were found to be clear on the difference. The results of this study are similar to previously reported results, with teachers closest to preferring a straight criterion reference approach and students leaning most toward norm referencing. Earlier work (Frary, et al, 1993) suggests that a difference in use of a fixed standard by subject area exists, such that math and science teachers were more likely to adhere to a set criteria, however the data in this study could not confirm a difference by subject area. When reviewing these results within the framework of consequential validity, the findings are somewhat troublesome. Value implications, according to Messick (1989), are that part of the score or grade to which judgements or emphasis, positive or negative, are placed. In the case of grading on a curve versus a more strictly defined criterion-referenced grade, the grade of B takes on different meanings. Specifically, a B is either the grade that a predefined proportion of the class should receive, or it is the grade received by students who have mastered some amount of material taught in the course. Given the results indicating that teachers are defining grades differently than students and parents, the validity of the grade is in question.

The last set of analyses done in this study were evaluations of the responses based on a high school students' transcript. The experimental design placed on the study was a fictitious student transcript which was either Jacob or Ellen, (male or female) and either mostly As and Bs or mostly Cs and Ds (high or low). What was asked of the groups of senders and receivers of grades was an estimate of the SAT I Verbal and Math score for the student represented by this transcript. There were no main effects beyond a difference due to the level of grades on the transcript. Earlier work looked at grading differences for male and female students (Manke and Loyd, 1990; Griswold, 1993), and reported gender by ability level interactions; the results here did not confirm earlier findings of this nature. In an additional analysis, responses for educational professionals were combined into one group, and evaluated using the gender of the respondent in the analysis. There was no evidence that either the student gender on the transcript or the gender of the respondent was contributing to the predicted scores, above and beyond what the grades on the transcript indicated.

Practical Implications

One aspect of examining the consequential validity of high school grades is to evaluate the message sent by teachers when reporting grades to the message received by the various grade receivers. Findings reported here indicate that largely the message sent is the message received, high school teachers, counselors, students, parents and college admission officers are in reasonable agreement about the meaning of grades. Furthermore, when these groups look at a set of grades on a transcript and consider college aptitude, there is a lot of similarity across groups.

To the extent that there may be some miscommunication, it appears that between students and teachers there is disagreement on the amount of work required and the amount of credit given for that work. The notion of grades as pay for work done may have application in trying to understand this difference.

Secondarily, students and parents are not in agreement with teachers on grading standards. Students and parents are more likely to be accepting of normative scaling in grading than are teachers. Perhaps this is related to the absolute standards to which teachers are held, such as high school graduation tests, or the difference in measurement training such that the meaning of terms about grading standards are less clear for parents and students.

Interestingly, the receivers that seem to read effort and improvement into grades and who think these ought to count are college admission officers. Perhaps for this group the role and definition of effort in grades is different than for teachers. Clearly, putting forth effort is a desirable behavior in a student at any level. However, grades may incorporate effort in various ways. If a student is trying really hard but still not showing understanding, does the exertion of effort result in a reward of higher grades? Or, is effort included in grades because by trying harder and therefore achieving better results on tests, a student earns higher grades. It is conceivable that college admissions officers and teachers are really considering effort to the same degree but define effort differently.

Areas for Further Investigation

Some of the results of this study warrant further exploration. One such result is the notion presented in earlier research of students' and teachers' perceptions of grades as pay for work done (Brookhart, 1993; Burton, 1983). The findings in this research study

indicate that this may be more the case for students than for teachers, due to results regarding the perceived weight or value placed on papers versus class participation. This is an interesting area and would be worth pursuing. An investigation which considers various kinds of work students are asked to do and the value or weight attributed to such work would be helpful both to the professional education community and to those students being asked to perform these tasks.

Another result indicated that the high school counselors in this study would give more credit for some of the non-achievement factors than teachers in the study. The reason for this is not known, but perhaps it is in the nature of student counseling that the need for information that seems more personal would be regarded highly. Similarly, the college admissions group may hold beliefs, which are different than teachers regarding grade construction due to the nature of the professional use of grades. College admissions officers include effort and improvement in their understanding of what grades mean, and include motivation among factors to consider in the admissions process. This raises the question of what the role of affective factors such as effort is in the grading process. Perhaps it is a difference in definition rather than a true difference in how much they should be included. Most college admissions staff admit that there are factors such as athletics and civic activity which contribute to the admissions equation. Results from the open-ended question in the current study indicate that students are also aware that these are components in the admissions process. An investigation into the importance of various factors in the admissions process, and how these factors are defined and included in grades by grade senders and receivers would be interesting.

In addition, this study asked about grade composition from two perspectives: What grades should comprise and what most teachers do. The results indicated that teachers believe non-achievement factors should be weighted more heavily than they believe teachers weight these factors in practice. This finding suggests an investigation into the actual practice of teachers. If the process of grading includes thinking about students' achievement and non-achievement factors, as indicated in this study, it would be informative to know how these components are incorporated into a single grade. Such an investigation would allow an evaluation of the differences between what teachers think most teachers do and what teachers say should be done.

Limitations

One limitation of this research is due to the size and nature of the sample. Interpretation of responses from the teachers and students in the study are limited by the fact that the sample are from one middle-class suburban district in the northeastern United States. The responses provided and interpretations thereof may not generalize to beliefs, attitudes and knowledge of other teachers and students in other locations. The same is true of the nature of the parent sample, which was collected from three middle-class suburban district high schools, and the size of the sample for parents limit generalizability further. The nature of the samples of high school counselors and college admissions staff is more diverse. The responses for these groups were obtained at three meetings of professional staff interested in issues surrounding college admission, however the potential for a bias due to self-selection is present, since the choice to fill out the survey at the meeting was independent for each respondent.

Some subgroup analyses could not be carried out due to sample size limitations. For college admissions staff, additional study could be done to consider whether the selectivity of the college in which the admissions officer works influences his or her perception of grading. Earlier research indicated that there may be differences among teachers based on the subject area taught. These differences could not be confirmed by this study, however the lack of confirmation could be due to sample size limitations, and should be explored further when more teachers responses have been collected.

Finally, it is important to note that in previous research (Waltman and Frisbie, 1994), parents and teachers of fourth graders were reported to have very little understanding of the concepts of growth vs. status and norm-referenced vs. criterion-referenced grading standards. This possibility was not investigated in this study.

Concluding Remarks

Consequential validity of grades takes into account the intended meaning of grades, the actual uses of grades, and the consequences of those uses. Given various receivers and users of grades, the potential exists for more than one interpretation. The findings in this study indicate that the interpretation of high school grades by parents, students, high school counselors and college admissions officers seems to be in concert with the intended meaning of teachers when sending grades. Teachers reported that most teachers weight achievement factors such as tests and papers highest in assigning grades; and although the tendency to include non-achievement factors seems prevalent, these factors are weighted as considerably less important than the achievement factors. Grade receivers are generally in agreement that this is the way teachers are operating. On the other hand, teachers indicated that the non-achievement factors should be weighted

higher, which may be an indication that teachers are not all doing what they say most teachers do. The extent to which teachers actually incorporate non-achievement factors into grades should be investigated, as well as the way in which the components are being combined into one grade.

What is less consistent across groups is the use of a fixed standard or a curve as the criteria for grading. In fact, parents and students lean more toward a mixture of norm-referenced and criterion-referenced grading system, teachers toward a criterion-referenced scale for grades. This difference allows for different interpretations of a grade. When the meaning of a grade is unclear, the communication intended by that grade is insufficient for purposes such as guidance, remediation or reward. For the consequences of grading to be valid for each purpose and grade meaning to be clear, efforts should be made so that all grade users understand the process of grading.

Finally, one fairly regular use of high school grades is as a predictor in the college admission process. Results presented here indicate that there is some disagreement in the meaning of grades as regards affective measures among senders and receivers. Whether these represent a misunderstanding in how grades are constructed, or differences in beliefs about what grades should comprise cannot be gleaned from these results. Given the prominence of high school grades in admission decisions, and the likelihood that grades will continue to be a very important component in the decision-making process, further work is needed to insure fair interpretation and valid judgements of grades by all receivers.

Table 1

Number of Responses by Group for Each Type of Experimental Transcript

	Group					
Transcript Type	Admissions	Counselors	Students	Parents	Teachers	Total
Female-high	14	29	12	11	21	87
Female-low	11	28	12	10	11	71
Male-high	11	30	13	10	15	79
Male-low	10	28	11	10	13	72

Note. See text for a description of groups.

Table 2

Percent of Teachers in Each Subject Area

Subject	Percent
Arts	3.3
Business	5.0
English	21.7
Family	3.3
Health	5.0
History	10.0
Language	8.3
Math	13.3
Science	15.0
Shop	1.7
Special	13.3
	N = 60

Table 3

Allocations for Grading Factors that Teachers Should Use To Determine Grades

	Group									
	Admissions		Counselors		Parents		Students		Teachers	
Factor	M	SD	M	SD	M	SD	M	SD	M	SD
Improvement	2.78	4.8	1.58	3.4	1.10	2.8	4.25	7.1	0.97	2.3
Growth	5.24	5.6	3.90	4.9	3.63	4.8	3.27	4.8	2.18	3.6
Attendance	6.85	5.7	5.47	6.1	3.39	4.6	4.06	4.2	5.87	5.6
Effort	9.02	5.6	6.75	5.5	6.27	5.2	9.06	9.2	6.93	5.7
Class Participation	12.50	6.4	14.14	8.1	13.37	6.9	10.92	9.0	11.88	6.7
Homework	13.43	6.5	13.08	6.4	12.63	5.0	10.69	7.4	12.83	5.1
Papers	23.54	6.8	23.48	10.1	22.49	9.2	25.35	11.9	21.67	9.0
Tests	26.85	11.5	31.75	13.5	36.29	14.2	32.08	15.0	37.58	14.9

Note. See text for a description of groups and n's. Participants allocated 100 points among the 8 factors in response to the following request: "Please attribute 100 points to indicate what YOU think each (of the factors) SHOULD be worth in determining a grade."

Table 4

Allocations for Grading Factors Believed To Be Teachers Practice to Determine Grades

	Group									
	Admissions		Counselors		Parents		Students		Teachers	
Factor	M	SD	M	SD	M	SD	M	SD	M	SD
Improvement	2.17	4.6	0.62	1.9	0.32	1.2	0.58	2.2	0.37	1.6
Growth	3.33	5.2	1.54	5.4	1.98	7.9	1.00	2.4	1.28	3.5
Attendance	7.41	5.0	4.83	6.4	1.78	3.4	4.31	3.7	3.98	4.5
Effort	7.70	6.7	4.35	4.9	3.59	4.5	4.50	5.3	4.02	5.5
Class Participation	10.28	6.1	10.89	6.8	11.20	7.9	8.17	4.7	10.97	5.4
Homework	15.50	7.6	15.46	8.4	12.29	6.6	10.29	6.7	14.28	5.9
Papers	19.98	9.8	20.41	8.8	25.15	9.8	29.17	13.1	21.83	9.1
Tests	33.78	13.4	41.43	17.2	43.39	14.1	41.56	14.0	43.33	14.3

Note. See text for a description of groups and n's. Participants allocated 100 points among the 8 factors in response to the following request: "Please attribute 100 points to indicate what you think MOST TEACHERS use to determine a grade."

Table 5

Overall Difference Between What Groups Believe Should Be Used In Grades and What Groups Believe Teachers Use to Construct Grades

	Euclidean Distance	
Group	Mean	SD
Counselors	25.08_a	15.25
Students	24.61_{ab}	17.61
Admissions	22.80_{ab}	11.54
Parents	20.12_{ab}	12.03
Teachers	16.90_b	12.55

Note. Means that do not share subscripts differ at $p < .005$ in the Tukey's studentized range comparison.

Table 6

Mean Weights Placed by Receivers and Senders on Eight Factors in Grade Composition

	Group				
Factor	Admissions	Counselors	Parents	Students	Teachers
Class Participation	10.28	10.89	10.26	**8.17	11.88
Attendance	7.41	4.83	*1.51	4.31	5.87
Homework	15.50	15.46	12.54	10.29	12.83
Improvement	2.17	0.62	0.37	0.58	0.97
Tests	33.78	41.43	43.97	41.56	37.58
Papers	19.98	20.41	25.46	*29.17	21.67
Effort	7.70	**4.35	*3.34	4.50	6.93
Growth	3.33	1.54	2.17	1.00	2.18

Note. See text for description of groups. Means for teachers are from the response to: "How should these factors be weighted?" Means for receivers are from the response to: "What do you think most teachers do?"

* $p < .01$, ** $p < .05$ using Dunnett's significance test, criterion group = teachers.

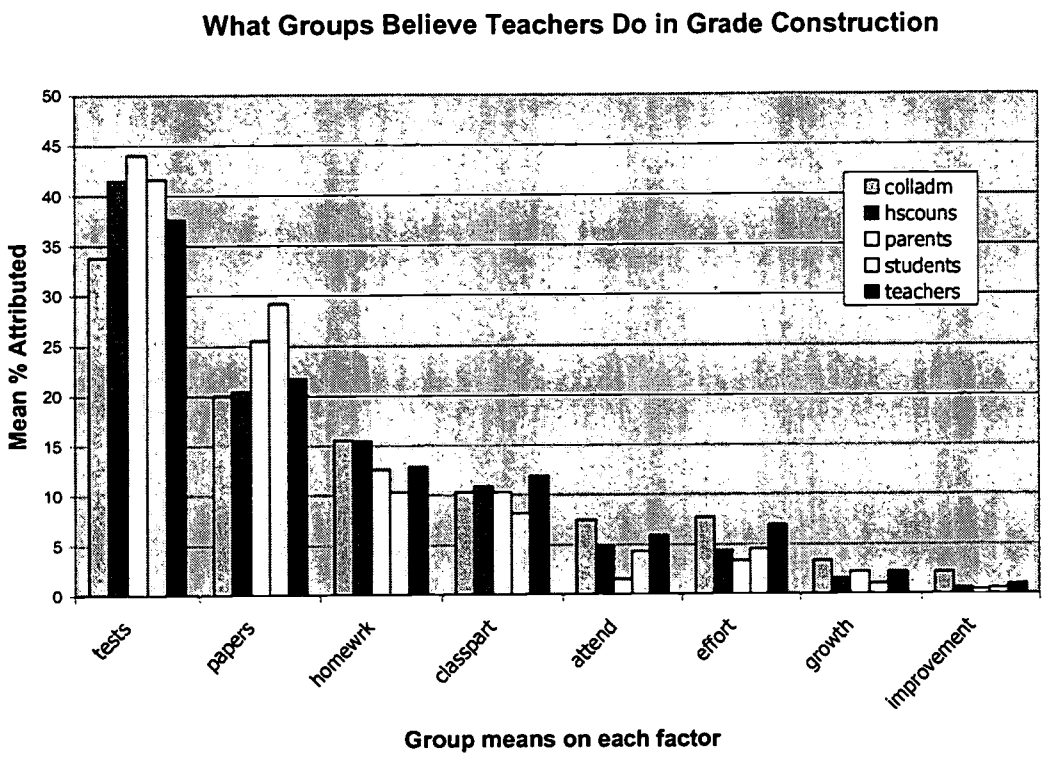

Figure 1: Receivers belief about the contribution of each factor in grades, and the senders' belief as to what each factor should contribute to a grade. Percentages are means by group.

Table 7

Differences among Receiver Groups in Beliefs about What Teachers Do

	Group			
Factor	Admissions	Counselors	Parents	Students
Improvement	2.17_a	0.62_b	0.32_b	0.58_{ab}
Attendance	7.41_a	4.83_{ab}	1.78_b	4.31_{ab}
Effort	7.70_a	4.35_b	3.59_b	4.50_{ab}
Homework	15.50_a	15.46_a	12.29_{ab}	10.29_b
Papers	19.98_a	20.41_a	25.15_{ab}	29.17_b

Note. Means in the same row that do not share subscripts differ at $p < .01$ in the Tukey's studentized range comparison.

Table 8

Difference Between Groups using Mean for Teachers as Control

	Group			
Difference Factor	Admissions	Counselors	Parents	Students
D-Attendance	3.43_a	0.85_{ab}	-2.20_b	0.33_{ab}
D-Effort	3.68_a	0.33_{ab}	-0.43_b	0.48_{ab}
D-Papers	-1.86_b	-1.42_b	3.31_{ab}	7.33_a

Note. Means in the same row that do not share subscripts differ at $p < .001$ in the Tukey's studentized range comparison.

A positive value indicates that the group believed teachers place a higher weight on this factor than the average of the teachers group response on this variable. A negative value indicates the group believes teacher in practice weight this factor lower than the average of the teachers group response to what most teachers weight it.

Table 9

Grading on a Curve versus a Fixed Standard

	Grading Scale	
Group	Mean	SD
Teachers	7.95_a	1.72
Admissions	7.40_{ab}	1.39
Counselors	7.21_{ab}	1.95
Parents	6.49_{bc}	1.99
Students	5.70_c	2.45

Note. Responses were made on a 10-point scale (1 = grade on the curve, 10 = fixed standard).

Means that do not share subscripts differ at $p < .005$ in the Tukey's studentized range comparison.

Table 10

Mean SAT I Verbal Scores Predicted by Group for Four Variants of Student Transcripts

	Experimental Transcript Type											
	High Females			High Males			Low Females			Low Males		
Group	N	Mean	s.d.	N	Mean	s.d.	N	Mean	s.d.	N	Mean	s.d.
Total	87	547	44.9	79	559	54.8	71	504	68.4	72	502	57.9
Admissions	14	552	29.9	11	555	26.9	11	472	48.5	10	477	58
Counselors	29	533	37.9	30	560	54.7	27	503	80.3	28	496	43.6
Students	12	557	53.8	13	559	76.6	12	550	52.6	11	521	64.9
Teachers	21	544	51.6	15	546	54.7	11	485	54.3	13	498	77.8
Parents	11	575	45.7	10	580	62.4	10	507	62.4	10	533	47.9

Table 11

Mean SAT I Math Scores Predicted by Group for Four Variants of Student Transcripts

	Experimental Transcript Type											
	High Females			High Males			Low Females			Low Males		
Group	N	Mean	s.d.	N	Mean	s.d.	N	Mean	s.d.	N	Mean	s.d.
Total	87	575	58.0	79	588	58.2	71	431	102.5	72	436	58.8
Admissions	14	557	37.5	11	559	49.7	11	420	53.5	10	433	52.9
Counselors	29	562	51.7	30	587	48.8	27	431	122.4	28	428	37.3
Students	12	585	63.2	13	590	64.8	12	437	150.1	11	456	89.5
Teachers	21	588	77.5	15	604	64.8	11	443	45.4	13	413	60.3
Parents	11	592	39.0	10	591	70.9	10	421	70.8	10	468	62.7

Table 12

Distribution of Education Professionals by Gender

	Group			
Gender	Admissions	Counselors	Teachers	Total
Males	23	41	16	84
Females	23	74	44	144
Total	46	115	60	221

Table 13

Mean SAT I Verbal and Math scores Predicted by Male and Female Education Professionals on Four Experimental Transcripts

			SAT I Verbal		SAT I Math	
Group	Transcript	N	Mean	Std. Dev.	Mean	Std. Dev.
Female						
	High Female	40	538	44.6	574	67.6
	High Male	35	555	54.8	583	59.7
Male						
	High Female	24	546	36.1	563	43.8
	High Male	21	555	42.0	592	48.0
Female						
	Low Female	33	483	59.0	418	54.3
	Low Male	33	496	45.7	427	42.2
Male						
	Low Female	16	511	85.2	459	148.1
	Low Male	18	486	72.4	421	55.2

REFERENCES

ACAOPU (Association of Collegiate Admissions Officers of Public Universities) (1997, January). Do you give extra quality points in calculating GPA's for ... accelerated courses in HS?

Adelman, C. (1983). Devaluation, diffusion and the college connection: A study of high school transcripts, 1964-1981. U.S. Department of Education. (ERIC Document Reproduction Service No. Ed 228 244).

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1997). Standards for educational and psychological testing. Manuscript in preparation.

Anderson, J.O., Bachor, D.G. (1993, April). Assessment practices in the elementary classroom: Perspectives of stakeholders. Paper presented at the Annual meeting of the National Council on Measurement in Education, Atlanta, GA.

Beaver, W.(1996). Is it time to replace the SAT? Academe, May-June, 1-3.

Bejar, I.I. & Blew, E.O. (1981). Grade inflation and the Validity of the Scholastic Aptitude Test. American Educational Research Journal, 18(2), 143-156.

Bracey, G.W. (1984). Grade inflation? Phi Delta Kapan, 76(4), 328-329.

Briscoe, C. (1994). Making the Grade: Perspectives on a Teacher's Assessment Practices. Mid-Western Educational Researcher, 7(4), 14-25.

Briscoe, C. & Ulerick-LaMaster, S. (1991). Concept mapping for meaningful learning in a college biology class. The American Biology Teacher, 53(4), 214-219.

Brookhart, S.M.(1991). Letter: Grading practices and validity. Educational Measurement: Issues and Practice, 10(1), 35-36.

Brookhart, S.M.(1993). Teachers' grading practices: Meaning and values. Journal of Educational Measurement, 30(2), 123-142.

Burton, F. (1983, April). A study of the letter grade system and its effect on the curriculum. Paper presented at the Annual meeting of the American Educational Research Association, Quebec, Canada.

Chapman, D.W., O'Brien, C.J., and DeMasi, M.E. (1987). The effectiveness of the public school counselor in college advising. The Journal of College Admissions, 151?,(Spring 1987), 11-18.

Cizek, G.J., Rachor, R.E, & Fitzerald, S. (1985, April). Further investigation of Teachers' Assessment Practices. Paper presented at the Annual meeting of the American Educational Research Association, San Francisco, CA.

Cizek, G.J., Fitzgerald, S. & Rachor, R.E. (1995/6). Teachers' assessment practices: Preparation, Isolation, and the Kitchen Sink. Educational Assessment, 3(2), 159-179.

Cole, N.S. (1988). A realist's appraisal of the prospects for unifying instruction and assessment. In Assessment in the service of learning: Proceedings of the 1987 ETS Invitational Conference. Princeton, NJ: Educational Testing Service.

Cronbach, L.J. (1982). Designing evaluations of educational and social programs. San Francisco: Jossey-Bass.

Crooks, T.J.(1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58, 438-481.

Cross, L.H. & Frary, R. B.(1999). Hodgepodge Grading: Endorsed by students and teachers alike. Applied Measurement in Education, 12(13), 53-72.

Daugherty, R. (1995). National Curriculum Assessment: A review of policy 1987-1994. London: The Falmer press.

Dobson, J. E., Campbell, N.J., Dobson, R. (1987). Relationships among loneliness, perceptions of school, and grade point averages of high school juniors. The School Counselor, 35(2), 143-148.

Durm, M.E.(1993). An A is not an A is not an A: A history of grading. The Educational Forum, 57, 294-297.

Ekstrom, R. B. (1994). Gender differences in high school grades: an exploratory study. (College Board Rep. No. 94-3). Princeton, NJ: Educational Testing Service.

Frary, R.B., Cross, L.H., & Weber, L.J. (1993) Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. Educational Measurement: Issues and Practice, 2, 23-30.

Friedman, S.J. & Frisbie,D.A. (1995). The Influence of Report Cards on the Validity of Grades Reported to Parents. Educational and Psychological Measurement, 55(1),5-26.

Frisbie, D.A., Waltman, K.K.(1992) An NCME Instructional Module on Developing a Personal Grading Plan. Educational Measurement: Issues and Practice, 11(3),35-42.

Gose, B. (1997, July 25). Efforts to curb grade inflation get an F from many critics. Chronicle of Higher Education, A41.

Goulden, N.R. & Griffin, C.J.G. (1995) The meaning of grades based on faculty and student metaphors. Communication Education, 44, 110-125.

Griswold, P.A. Beliefs and inferences about grading elicited from student performance sketches. Educational Assessment, 1(4), 311-328.

Gronlund, N.E. (1985). Measurement and evaluation in teaching. New York: Macmillan.

Hayden, T.C. (Ed.). (1995). Peterson's handbook for college admissions (4th edition). location?

Herman, J.L. and Golan., S. (19??) The effects of standardized testing on teaching and schools. Educational Measurement: Issues and Practice, 12, 20-42.

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. Journal of Educational Psychology, 57, 469-480.

Keith, T.Z. (1982). Time spent on homework and high school grades: A large-sample path analysis. Journal of Educational Psychology, 74(2), 248-253.

Kodras, J.E. and Prather, J.E. (1978). Faculty opinions on grading standards and practices. Report 78-22 (ED 160 014). Atlanta, GA: Georgia State University.

Levy, J. and Riordan, P. (1994). Rank-in-Class, Grade Point Average, and College Admissions. National Association of Secondary School Principals, Reston, Va.

Linn, R.L. (1997). Evaluating the validity of assessments: The consequences of use. Educational Measurement: Issues and Practice, 16(2), 14-15.

Loeb, J.W. (1993). Standards for the future -- Not unscalable barriers. College Board Review, 169, 15-30.

Loyd, B.H., Nava, F.G., & Hearn D.L. (1991, April). High school students perceptions of the grading process. Paper presented at the Annual meeting of the National Council on Measurement in Education, Chicago, IL.

Maguire, T., Hattie, J., & Haig, B.(1994). Construct validity and achievement assessment. The Alberta Journal of Educational Research, 40(2), 109-126.

Manke, M.P. & Loyd, B.H. (1990, April). An investigation of non-achievement-related factors influencing teacher' grading practices. Paper presented at the Annual meeting of the National Council on Measurement in Education, Boston, MA.

Manke, M.P. & Loyd, B.H. (1991, April). A study of teachers' understanding of their grading practices. Paper presented at the Annual meeting of the National Council on Measurement in Education, Chicago, IL.

Mehrens, W.A. (1997). The consequences of consequential validity. Educational Measurement: Issues and Practice, 16(2), 16-18.

Messick, S. (1984). Evidence and ethics in the evaluation of tests. Educational Researcher, 10(9), 9-20.

Messick, S. (1984). The psychology of educational measurement. Journal of Educational Measurement, 21(3), 215-237.

Messick, S. (1989). Validity, in R. L. Linn (Ed.), Educational Measurement (3rd ed.), (pp. 13-103). Washington, DC: American Council on Education.

Messick, S. Validity of test interpretation and use. In M.C. Alkin (Ed.), Encyclopedia of Educational Research (6th ed.) (pp. 1487-1495). New York: MacMillan, 1992.

Messick, S. (1995). Validity of Psychological Assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741-749.

Mitchell, J.V., Jr. (1983). Self-assessment variables related to grade satisfaction and dissatisfaction in high school. The School Counselor, 30,(5), 368-373.

Murphy, R. and Broadfoot, P. (1995). Effective Assessment and the Improvement of Education -- A tribute to Desmond Nuttall. London: The Falmer Press.

Natriello, G. (1984). Problems in the evaluation of students and student disengagement from secondary schools. Journal of Research and Development in Education, 7(4), 14-24.

Natriello, G., and others (1985). School reform and potential dropouts. Educational Leadership, 43(1), 10-14.

Ordovensky, P. (1996, October 25). Test critics decry trend. USA Today, pp. A1-.A2.

Pilcher, J.K. (1994). The value-driven meaning of grades. Educational Assessment, 2(1), 69-88.

Popham, W.J. (1997). Consequential validity -- Right concern, wrong concept. Educational Measurement: Issues and Practice, 16(2), 9-13.

Richards, A.L. and Luong-van, S. (1992). Moderation Procedures in English and Mathematics in years 10 and 11 in the Northern Territory (CAB-R-5/1991). Northern Territory Dept. of Education, Darwin (Australia).

Shepard, L.A. (1997). The centrality of test use and consequences for test validity. Educational Measurement: Issues and Practice, 16(2), 5-8.

Shepard, L.A. (1993). Evaluating test validity. Review of Research in Education, 19,405-450.

Shepard, L.A. (1989). Why we need better assessments. Educational Leadership, 46(7),4-9.

Smith, J.K. & Loyd, B. (199?) Principles of grading and assessment. Unpublished manuscript.

Starch, D., and Elliot, E.C. (1912). Reliability of grading high school work in English. School Review, 20, 442-57.

Starch, D., and Elliot, E.C. (1913). Reliability of grading high school work in Mathematics. School Review, 21, 254-95.

Stiggins, R.J., Frisbie, D.A., and Griswold, P.A. (1989). Inside high school grading practices: building a research agenda. Educational Measurement :Issues and Practice, 8, 5-15.

Talley, N.R. and Mohr, J.I.(1991). Weighted Averages, computer screening, and college admission in public colleges and universities. The Journal of College Admission, 132, 9-11.

Talley, N.R. and Mohr, J.I.(1993). The case for a national standard of grade weighting. The Journal of College Admission, 139, 9-13.

Thorndike, R.M., Cunningham, G.K., Thorndike, R. L., and Hagen, E. (1991). Measurement and evaluation in psychology and education (5th ed.). New York: Macmillan.

Turnbull, W.W. (1985). Student change, program change: Why SAT scores kept falling. (College Board Rep. No. 85-2). New York, NY: College Entrance Examination Board.

Traub, R.E. (1990). Assessment in the classroom: What is the role of research? The Alberta Journal of Educational Research, 35(1), 85-91.

Unger, H.G. (1990). A student's guide to college admissions. New York: Facts on file.

U.S. Congress, Office of Technology Assessment.(1992). Testing in American schools: Asking the right questions (OTA-SE7-519). Washington, DC: U.S. Government Printing Office.

U.S. Department of Education, Office of Educational Research and Improvement (1994). What do student grades mean? Differences across schools (Office of Research Rep. OR 94-3401). Washington, DC: U.S.Office of Research.

Waltman, K.K. & Frisbie, D.A. (1994).Parents' Understanding of Their Children's Report Card Grades. Applied Measurement in Education, 7(3),223-40.

Zeidner, M. (1992). Key facets of classroom grading: A comparison of teacher and student perspectives. Contemporary Educational Psychology, 17, 224-243.

APPENDIX

1. Below is a list of factors which are often used by high school teachers in determining grades. Consider a high school U.S. history course. Please attribute 100 points to indicate what YOU think each SHOULD be worth in determining a grade in such a course. If there are factors listed there that you believe should not be included in a grade, write zero (0) next to those. Please be sure when you have finished that the total adds up to 100.

_____ **participation in class**

_____ **attendance**

_____ **homework**

_____ **improvement from previous year or semester's performance**

_____ **tests, quizzes**

_____ **term papers, projects**

_____ **effort**

_____ **growth during this semester/year**

2. Below is a list of factors which are often used by high school teachers in determining grades. Consider a high school U.S. history course. Please attribute 100 points to indicate what you think MOST TEACHERS use to determine a grade If there are factors listed there that you believe should not be included in a grade, write zero (0) next to those. Please be sure when you have finished that the total adds up to 100.

_____ **participation in class**

_____ **attendance**

_____ **homework**

_____ **improvement from previous year or semester's performance**

_____ **tests, quizzes**

_____ **term papers, projects**

_____ **effort**

_____ **growth during this semester/year**

3. Some high school teachers grade "on the curve" so that there is always approximately the same number of As, Bs, Cs, etc. Some high school teachers use a fixed standard such as A= 90 to 100, B= 80 to 89, C = 70 to 79, etc. Still other high school teachers try to mix the two concepts. On the scale below, circle the number which best represents your feelings on how grading should be approached.

1 2 3 4 5 6 7 8 9 10

grade on the curve mix of the two fixed standard

Ellen/ Jacob Smith *(high- end transcript)*

Comprehensive Public High School, SES: Middle Class, School Enrollment approx. 1000

6 AP courses available: English, Biology, Calculus, Physics, Chemistry and U.S. History.

STUDENT'S ACADEMIC HISTORY

Grade 09 **Year 94-95**

Course ID	Subject	Semester	Final	Credits
310	ALGEBRA 1	Y	B	5.00
801	EART/ENVSCIENCE	Y	A	5.00
114	ENG 9	Y	B	5.00
911	PE-HEALTH 09	Y	A-	5.00
003	PER KEY/SP WRI	Y	A-	5.00
440	SPANISH 1	Y	B	5.00
211	W. HISTORY	Y	B	5.00

Grade 10 **Year 95-96**

Course ID	Subject	Semester	Final	Credits
823	BIOLOGY 1	Y	A	5.00
124	ENGLISH 10	Y	A	5.00
320	GEOMETRY	Y	A-	5.00
209	LAW & SOCIETY	Y	B-	5.00
831	MARINE SCI	Y	A-	5.00
912	PE-HEALTH 10	Y	B	5.00
450	SPANISH 2	Y	B	5.00

Grade 11 **Year 96-97**

Course ID	Subject	Semester	Final	Credits
312	ALGEBRA 2	Y	A	5.00
841	BIOLOGY AP	Y	B-	5.00
833	CHEMISTRY	Y	B	5.00
231	ECONOMICS	Y	B-	5.00
134	ENG 11	Y	B	5.00
913	PE -HEALTH 11	Y	A	5.00
440	SPANISH 3	Y	B-	5.00

Class Rank 65 Class Size 260

4. Considering the transcript for Ellen/Jacob Smith, what SAT V and M score would you predict for her/him?

SAT V _____ SAT M _____

Ellen/Jacob Smith *(low end transcript)*

Comprehensive Public High School, SES: Middle Class, School Enrollment approx. 1000

6 AP courses available: English, Biology, Calculus, Physics, Chemistry and U.S. History.

STUDENT'S ACADEMIC HISTORY

Grade 09		Year 94-95		
Course ID	Subject	Semester	Final	Credits
310	ALGEBRA 1	Y	C	5.00
711	BASIC FOODS	Y	B-	5.00
801	EART/ENVSCIENCE	Y	C	5.00
114	ENG 9	Y	C	5.00
911	PE-HEALTH 09	Y	B	5.00
440	SPANISH 1	Y	C	5.00
211	W. HISTORY	Y	A-	5.00
Grade 10		**Year 95-96**		
Course ID	Subject	Semester	Final	Credits
823	BIOLOGY 1	Y	B-	5.00
213	CONT WLD AFFRS	Y	A	5.00
124	ENGLISH 10	Y	B-	5.00
320	GEOMETRY	Y	D	5.00
912	PE-HEALTH 10	Y	A-	5.00
450	SPANISH 2	Y	D	5.00
006	WORD PROC	Y	D	5.00
Grade 11		**Year 96-97**		
Course ID	Subject	Semester	Final	Credits
013	ACCT 1	Y	C	5.00
312	ALGEBRA 2	Y	D	5.00
134	ENG 11	Y	B-	5.00
831	MARINE SCI	Y	C	5.00
913	PE -HEALTH 11	Y	A-	5.00
440	SPANISH 3	Y	B-	5.00
234	US HISTORY 1	Y	B	5.00
Class Rank 158	**Class Size 260**			

4. Considering the transcript for Ellen Smith, what SAT V and M score would you predict for her ?

SAT V _____ **SAT M** _____

5. Finally, looking again at the transcript for Ellen/Jacob Smith, consider the following. Ellen/Jacob is considering applying to a local 4-year college. The college is part of the state college system, there are 6 colleges in the state system, this one is rated in the middle. Is this a good choice for Ellen/Jacob? How do you think she/he will do? Please explain why you think this.

AERA

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

TM031667

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Consequential Validity for High School Grades: What is the Meaning of Grades for Senders and Receivers

Author(s): Patricia A. Baron

Corporate Source:

Publication Date: April, 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA - FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Patricia A. Baron

Printed Name/Position/Title: Patricia Baron Measurement Statistician2

Organization/Address: 1762 Old Trenton Rd
West Windsor NJ 08550

Telephone: 609-683-2294

FAX:

E-Mail Address: Pbaron@ets.org

Date: 7/17/00

(over)

Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 2000

Dear AERA Presenter,

Congratulations on being a presenter at AERA. The ERIC Clearinghouse on Assessment and Evaluation would like you to contribute to ERIC by providing us with a written copy of your presentation. Submitting your paper to ERIC ensures a wider audience by making it available to members of the education community who could not attend your session or this year's conference.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed, electronic, and internet versions of *RIE*. The paper will be available **full-text, on demand through the ERIC Document Reproduction Service** and through the microfiche collections housed at libraries around the world.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse and you will be notified if your paper meets ERIC's criteria. Documents are reviewed for contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at **http://ericae.net.**

To disseminate your work through ERIC, you need to sign the reproduction release form on the back of this letter and include it with **two** copies of your paper. You can drop of the copies of your paper and reproduction release form at the ERIC booth (223) or mail to our attention at the address below. **If you have not submitted your 1999 Conference paper please send today or drop it off at the booth with a Reproduction Release Form.** Please feel free to copy the form for future or additional submissions.

Mail to: AERA 2000/ERIC Acquisitions
The University of Maryland
1129 Shriver Lab
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC/AE is a project of the Department of Measurement, Statistics and Evaluation at the College of Education, University of Maryland.