DOCUMENT RESUME

ED 384 402 JC 950 363

AUTHOR Karpp, Edward R.
TITLE Paths to Success, Volume III. Validating the California Chemistry Diagnostic Test for Local Use.
INSTITUTION Glendale Community Coll., CA. Planning and Research Office.
PUB DATE Aug 95
NOTE 22p.; For volume I, see ED 312 021, for volume II, see ED 325 170.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Chemistry; Community Colleges; *Diagnostic Tests; Placement; *Predictive Validity; State Surveys; State Universities; Statistical Data; Statistical Significance; *Test Bias; *Test Reliability; *Test Validity; Two Year Colleges

ABSTRACT

The California Chemistry Diagnostic Test (CCDT) was developed by chemistry faculties at the University of California, California State University, and California Community College systems to address dissatisfaction with existing chemistry assessment instruments. The instrument was tested at 25 campuses of the 3 systems, after which Glendale Community College (GCC) undertook a project to validate the test locally by comparing test scores with students' final grades. The validation procedure examined the test's content-related and criterion-related validity, reliability, disproportionate impact and bias, and cut scores. For the systemwide validation, a comparison of test scores and final chemistry grades for 4,023 students showed a Pearson correlation coefficient of .42, exceeding the minimum standard of .35 established by the Office of the Chancellor. The local validation at GCC included 228 students and also resulted in a coefficient of .42, indicating that the CCDT score was a significant predictor of chemistry performance. The test was also found to have a high reliability in both the statewide and the local validations. With respect to test bias, while both the statewide and local validations found small but statistically significant differences for gender and ethnic groups; the differences were considered to be too small, however, to be attributable to the instrument. (A definition of terms and suggested cut scores are included.) (KP)

VALIDATING THE CALIFORNIA CHEMISTRY DIAGNOSTIC TEST FOR LOCAL USE

(PATHS TO SUCCESS VOLUME III)

by

Edward R. Karpp

August, 1995

SUMMARY

This report describes the local validation of the California Chemistry Diagnostic Test (CCDT) at Glendale Community College. The validation study showed that this placement test is an excellent predictor of student success in general chemistry courses at Glendale College.

The CCDT was developed by the chemistry faculties of the UC, CSU, and California Community College systems to address the inadequacy and inconsistency of existing tests. A local validation procedure was developed, using traditional and new methods, to demonstrate the test's adequacy in the areas of validity, reliability, test bias, disproportionate impact, and placement scores. These are the validation topics which must be demonstrated, according to the *Standards, Policies, and Procedures for the Evaluation of Assessment Instruments Used in the California Community Colleges*, published by the Chancellor's Office.

Using a sample of Glendale College students who took the CCDT and completed chemistry courses, the Research and Planning Unit found that the placement test demonstrates all the necessary attributes. Validity, which means that the test measures what it purports to measure (i.e., chemistry skills), was demonstrated through two methods. Content-related validity was assured by the participation of a Glendale chemistry faculty member on the test development team. Criterion-related validity was established through correlation analysis: test scores are highly correlated with final chemistry course grades. Reliability, which means test scores are stable over time, was demonstrated by analysis of test scores over time. Test bias and disproportionate impact, which refer to equity issues and whether the test disciminates against gender, ethnic, and age groups, were addressed through various statistical techniques. Placement scores, referring to the adequacy of cut scores, were also validated through statistical techniques.

Additionally, multiple measures were identified which improved the predictiveness of the test. These measures were the student's highest math course completed and last math grade obtained. Several prediction methods were devised using these measures (CCDT score, highest math class, and last math grade) which successfully predict student performance in general chemistry courses.

TABLE OF CONTENTS

GRAPHS AND TABLES

Graphs

Tables

INTRODUCTION

Test Development

The California Chemistry Diagnostic Test (CCDT) was developed to address the dissatisfaction of University of California and California State University chemistry faculty with various chemistry assessment instruments previously in use.[1] This dissatisfaction was due to the wide range of topics covered in these tests and the mismatch between the skills and knowledge tested and chemistry curricula. A committee of chemistry faculty from the UC, CSU, and California Community College systems was formed to develop a testing instrument that addressed these problems. This effort also sought to encourage intersegmental articulation, so curricula in each segment could articulate to courses in the other segments.

The committee developed the CCDT in two forms, Form A and Form B. These forms were pilot tested on a sample of over 6,700 college students. The most predictive and least biased test items were used in the final form of the test, Form C. Form C was validated at 25 campuses of the three systems, including 19 community colleges (one of which was Glendale Community College). The results of the Form C validation project were promising: The CCDT predicted general chemistry course grade for students in all three systems. Gender and ethnic differences were minimal.

In addition to the intersegmental validation, it was necessary for Glendale College to validate the final form of the placement test locally. Local validation is required by the Chancellor's Office; moreover, it is necessary to determine whether the test is a useful assessment instrument on campuses where it is being used. A test might be valid on a global level (i.e., when large numbers of students are pooled), but it might not be useful on a local level, where students are more diverse than in the general population. The student body of Glendale College is highly diverse in terms of ethnicity, language background, and academic preparation. It is conceivable that the CCDT, while valid when predicting the performance of UC and CSU students, could be ineffective when predicting the chemistry performance of Glendale's students.

Local Validation Procedure

The Research and Planning Unit developed a validation procedure to address the validity standards defined by the Chancellor's Office. The following is a list of these standards, along with definitions:

- **Validity**: *Validity* means that the test measures what it is supposed to measure, in this case, chemistry knowledge and skills. According to the Chancellor's Office standards, two types of validity must be addressed: content-related validity and criterion-related validity. *Content-related validity* means that the test's content is related to the curricula of the courses into which students are placed. *Criterion-related validity* means that the test predicts chemistry course performance. The CCDT shows excellent content- and criterion-related validity.
- **Reliability:** *Reliability* refers to a test's consistency. If the same student's scores are very different on different administrations of the test, then the test is not reliable. In the current study, reliability was addressed by examining the stability of test scores over time.
- **Disproportionate Impact and Test Bias:** These related topics address how the test places different gender, ethnic, and age groups. If a test shows *disproportionate impact*, then different groups are affected differentially by the test. For example, males might score lower on a test than females. *Test bias* refers to the content of the test and whether its items are biased in some manner against any gender, ethnic, or age group. The CCDT shows some very small disproportionate impact (most tests do), but its effect on student placement is inconsequential.
- **Cut Scores:** *Cut scores* define the placement system which uses the test to assign students to courses. In the current study, pass rates and other statistics were computed for a range of potential cut scores, and effective chemistry cut scores were identified.

The remaining sections of this report describe in detail the validation procedure of the CCDT. The test was found to meet all the Chancellor's Office standards; locally, it is a very good predictor of student performance in general chemistry courses.

CONTENT-RELATED VALIDITY

Intersegmental Development

The California Chemistry Diagnostic Test (CCDT) was developed to address the dissatisfaction of University of California and California State University chemistry faculty with various chemistry assessment instruments previously in use.[1] The committee responsible for the CCDT's development used the "Statement on Preparation in Natural Science Expected of Entering Freshmen," prepared by the academic senates of the UC, CSU, and California Community College systems. This document, along with the test development committee, identified eight topic areas normally taught in first-term general chemistry courses. These eight topics were (1) compounds and elements, (2) states of matter, (3) reactions of matter, (4) structure of matter, (5) periodic properties, (6) solutions, (7) qualitative concepts in thermodynamics and kinetics, and (8) laboratory skills.[2] The initial two versions of the CCDT, Forms A and B, consisted of items covering these eight topics.

The third version of the CCDT, Form C, was constructed from the test items that were both reliable and predicted general chemistry course grades. Items from all eight chemistry categories and an additional mathematics category were included in Form C.

Local Studies

The Glendale College chemistry faculty was involved with test development and pilot testing from the beginning of the project. Dr. Margaret Henley, a Glendale chemistry professor, was a member of the Chemistry Diagnostic Testing Project, the panel of chemistry experts that planned the test, developed pilot items, coordinated pilot testing, and selected the items to include on the final form. During the development phase, the local faculty reviewed the pilot tests and was aware of the test's content. The Glendale chemistry department voted to accept the CCDT test.

CRITERION-RELATED VALIDITY

Intersegmental Development

The intersegmental development team reported that "...the test is a valid predictor of student performance."[3] For a validation sample of 4,023 California college and university students, the CCDT test score and general chemistry course grade showed a Pearson correlation coefficient of 0.42.

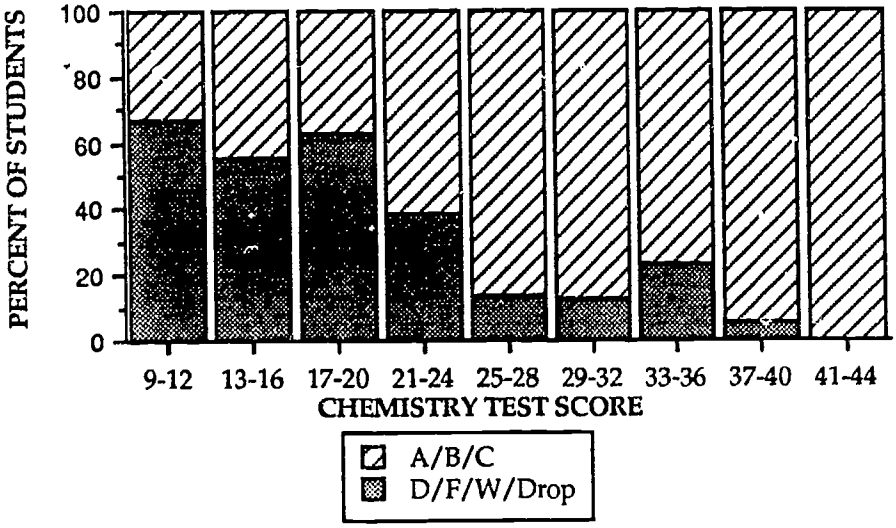
Local Studies

The relationship between CCDT scores and course performance was also studied locally. The test was shown to be significantly correlated with chemistry course performance at Glendale. Moreover, regression analysis identified multiple measures which serve as additional predictors of performance. The following section describes these results.

Correlation Analysis. The correlation between CCDT score and Chemistry 101 grade was 0.42, based on data from 228 students. This correlation coefficient is statistically significant ($p < 0.0005$). It also exceeds the minimum requirement of 0.35 established by the Chancellor's Office's standards.

The relationship between CCDT score and chemistry course performance is shown in Graph 1 (next page). The proportion of students passing and failing Chemistry 101 is shown for different ranges of CCDT scores. As CCDT score increases, the proportion of students passing tends to increase. Passing Chemistry 101 is thus associated with higher CCDT scores.

Graph 1. CCDT Score by Chemistry 101 Success



Regression Analysis. Multiple regression showed that CCDT score is a significant predictor of chemistry performance, even when other relevant variables are considered. For this analysis, the dependent variable was Chemistry 101 grade, because the assessment test's purpose is to predict success in Chemistry 101. If a student's chance of succeeding in Chemistry 101 is questionable, then he or she should be placed in the introductory Chemistry 110 course. Chemistry 101 grade was converted to a 4-point scale (with A's coded as four). Withdrawals were included as non-successes (coded as zero).

The goal of multiple regression analysis is to identify a group of student variables which optimally predict course performance. Potential predictor variables were those which were likely to be correlated with chemistry grade: High School GPA, Years Since Taking a Math Course, Years of English Courses in High School, Last Math Grade, Highest Math Course Completed, and CCDT score.

Analysis was performed using the stepwise regression procedure of SPSS^{X}. CCDT scores entered the regression equation first, indicating that it was the strongest available predictor of chemistry grade. Highest Math Course Completed and Last Math Grade were the only other variables entering the regression equation.

The regression equation consisting of these three variables showed a Multiple R of 0.55 and an adjusted R^2 of 0.297, indicating that approximately 30% of the variance in chemistry grades is predicted by these three variables. The resulting regression equation explained a highly significant proportion of the variance, $F = 31.62$ ($p < 0.00005$). Additionally, as required by the Chancellor's Office standards, the three measures of students' ability are not highly intercorrelated (the CCDT-Highest Math Course Completed correlation is $r = 0.26$, and the CCDT-Last Math Grade correlation is $r = 0.29$; the Highest Math Course Completed-Last Math Grade correlation is only 0.07).

CCDT Score, Highest Math Course Completed, and Last Math Grade, therefore, reliably predict Chemistry 101 grade. It was determined that a simple method of calculating the combined predictor variable was desirable. A simple formula (as opposed to, say, a regression formula with its decimal coefficients) has the advantage that individuals can readily calculate it—this is advantageous for both students and counselors. Of several simple formulas studied, the most effective was shown to be:

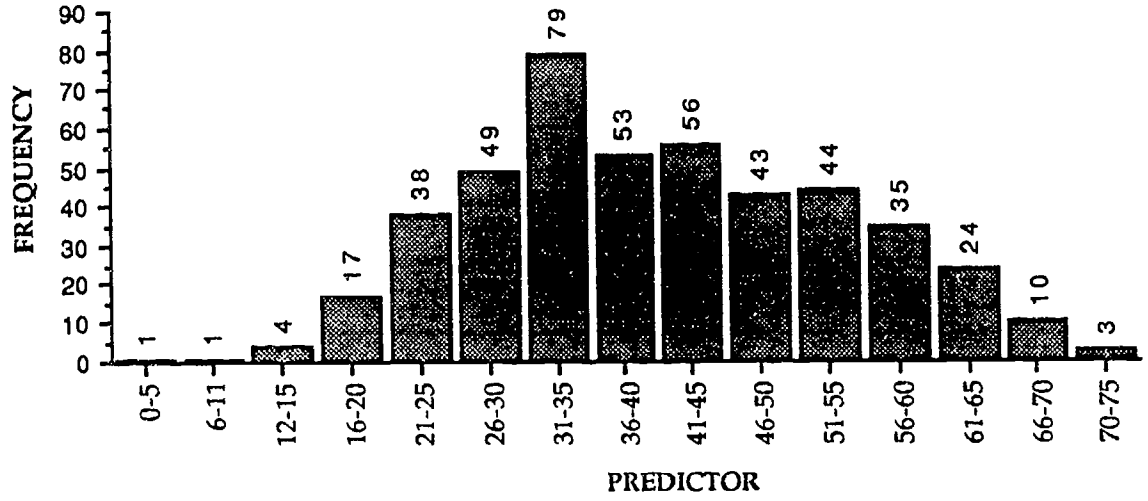
PREDICTOR = CCDT + (HIGHEST MATH COURSE x LAST MATH GRADE).

In this formula, CCDT is the raw assessment test score, with a possible range of zero to 44. Highest Math Course is converted to a 7-point numerical scale (None = 0, Basic Math = 1, Algebra I = 2, Geometry = 3, Algebra II = 4, Trigonometry = 5, College Algebra/Precalculus = 6, and Calculus = 7). Last Math Grade is represented by a 4-point numerical scale (0=F, 1=D, 2=C, 3=B, 4=A). Both Highest Math Course and Last Math Grade are self-reported by students.

The prediction equation predicts chemistry grade with a regression coefficient of 0.54, which is nearly identical to that of the full regression equation, 0.55. The predictor also has a relatively normal distribution (see Graph 2, next page), with a mean of 40.56 (for the full sample, N=457, that took the CCDT between Fall 1991 and Fall 1993).

The CCDT therefore shows acceptable criterion-related validity. The derived prediction equation, with its incorporation of multiple measures, also shows acceptable criterion-related validity.

Graph 2. Chemistry Predictor Distribution



PREDICTOR = CHEMISTRY SCORE + (LAST MATH GRADE x HIGHEST MATH COURSE)

RELIABILITY

Intersegmental Development

Based on a sample of 4,023 students, the intersegmental validation reports the CCDT's standard error was 0.12; based on the 1,789-student community college subsample, the standard error was 0.18. These values show that the amount of error in the test is very small. Such a narrow standard error suggests the CCDT is a reliable instrument.

Coefficient alphas of 0.87 for the full sample and 0.86 for the community college subsample were also reported. These high values indicate the test is internally consistent.

Local Studies

Both CCDT scores and the derived prediction equation show stability over time. Table 1 (next page) shows the means for both scores for the years in which data were

available. Although the means of both scores increased somewhat, the differences are not statistically significant.

Table 1. CCDT and Predictor Means Over Time

	CCDT		*Predictor*	
Calendar Year	*Mean*	*N*	*Mean*	*N*
1991	23.08	139	38.55	128
1992	24.52	200	40.52	1
1993	25.51	151	41.29	129

DISPROPORTIONATE IMPACT AND TEST BIAS

Intersegmental Development

The intersegmental development project addressed the potential for gender and ethnic differences in CCDT scores. Their conclusion was that, whereas some differences exist, those differences are extremely small.

Pilot testing of the CCDT examined possible gender and ethnic differences for each test item. Two forms of the test (A and B) were pilot tested, and only the best test items were included in the final version of the test, Form C. According to the intersegmental development team, "Among the criteria used for selecting items were chemistry content, item-total correlation, item difficulty, item characteristic curve, distractor pattern, sex differences, and ethnic differences."[4] The intersegmental developers thus selected test items for the final form of the CCDT to minimize gender and ethnic bias.

Table 2 (next page) shows the reported means for male versus female students and white versus nonwhite students. The scores of students from nonwhite ethnic groups (including Latinos) were combined for this analysis due to unequal sample sizes and the development project team's decision to compare the performance of minority students with that of white students.

Table 2. Statewide CCDT Means by Gender and Ethnic Groups

	Total Sample		*Community College Sample*	
Group	*CCDT Mean*	*N*	*CCDT Mean*	*N*
Males	24.74	2180	23.21	551
Females	22.31	1412	20.88	517
Whites	24.78	1387	23.52	527
Nonwhites	23.23	1203	21.80	1036
Total	23.72	4023	22.37	1789

The differences in means, though small, are statistically significant: Males achieve reliably higher scores than females, and whites achieve reliably higher scores than nonwhites. However, gender and ethnic differences account for a very small proportion of the total variance in CCDT scores, as indicated by the eta-square statistic. The development project team reported that the largest gender effect account for only 7% of the variance, and the largest ethnicity effect accounted for about 1%. Therefore, "...of all the reasons for...students to differ in their total test score, very few can be attributed to their sex or to their ethnicity".[5]

The development project team also addressed the related issue of differential prediction. Correlations between CCDT score and final course performance were calculated for gender and ethnic groups. These correlations are shown in Table 3.

Table 3. Statewide Correlations Between CCDT and Course Grade by Gender and Ethnic Groups

	Total Sample		*Community College Sample*	
Group	*Correlation*	*N*	*Correlation*	*N*
Males	0.41	1530	0.30	530
Females	0.46	1030	0.39	271
Whites	0.42	1034	0.28	279
Nonwhites	0.44	853	0.32	277
Total	0.42	2836	0.32	916

These correlation coefficients show that the CCDT is a consistently strong predictor of chemistry grade for all groups. Further, for the community college subsample, the test is a somewhat more accurate predictor of grades for females and for nonwhites. According to the development project data, therefore, the CCDT does not appear to impact gender and ethnic groups inequitably.

Local Studies

Local studies confirmed that the CCDT shows little disproportionate impact. The test shows some gender and ethnic differences locally. However, these differences are small, and evidence suggests that they are not attributable to the testing instrument.

Table 4 compares locally obtained mean CCDT scores for gender, ethnic, and age groups. Predictor scores (Predictor = CCDT + Highest Math Course x Last Math Grade) means are also compared.

Table 4. CCDT and Predictor Means by Gender, Ethnic, and Age Groups at Glendale College

Group	*CCDT Mean*	*CCDT Std. Dev.*	*N*	*Predictor Mean*	*Predictor Std. Dev.*	*N*
Males	27.44	7.09	142	43.98	11.58	138
Females	24.14	7.46	104	40.21	12.66	99
Whites	26.62	7.06	106	42.51	11.15	103
Nonwhites	25.61	7.68	140	42.33	12.93	134
Age 17 & Under	27.58	7.33	38	44.92	11.46	38
Age 18-21	26.81	6.82	124	42.95	11.76	118
Age 22 & Over	26.37	7.19	60	43.83	11.22	59
Total	26.05	7.42	246	42.41	12.16	237

For both CCDT scores and Predictor scores, the gender difference is statistically significant. Females score significantly lower than males on both measures. However, the proportion of variance accounted for by gender is very small: eta-square is 4.8% for CCDT scores and only 2.3% for Predictor scores. Although females perform somewhat

worse than males, over 95% of the variance in these scores is not explained by differences in gender.

The ethnic difference is not significant, when white students are compared with nonwhite students (including Latinos). The ethnic comparison is complicated by Glendale College's population, which consists of native and non-native English speakers in both white and nonwhite groups. The nonsignificant ethnic difference, however, suggests that there is no substantial disproportionate impact due to the test.

The CCDT and the Predictor both predict chemistry grade well for different groups of students. Table 5 shows score-grade correlations for gender, ethnic, and age groups.

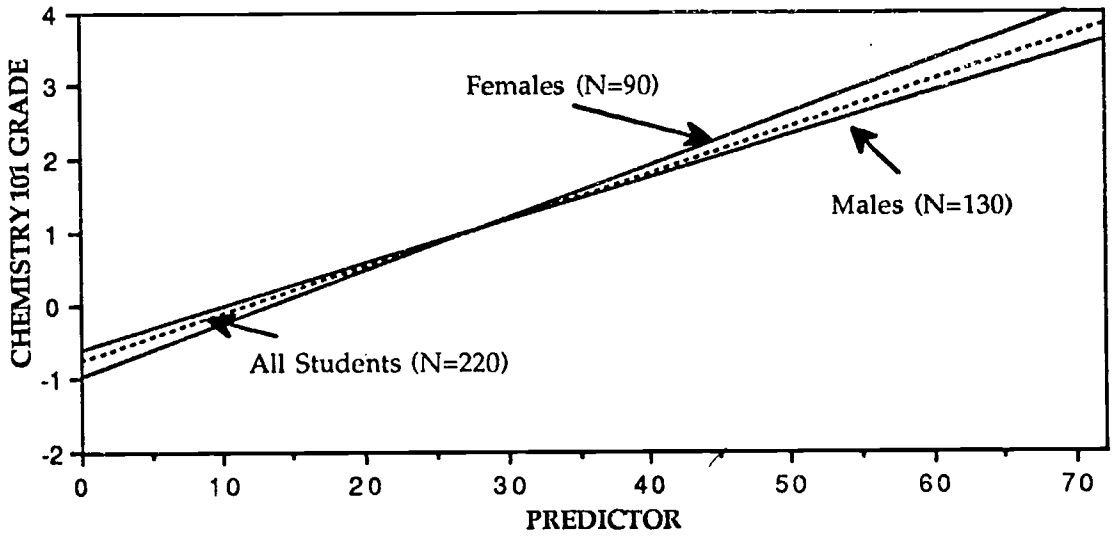
Table 5. Score-Grade Correlations for CCDT and Predictor by Gender, Ethnic, and Age Groups at Glendale College

	CCDT-Grade		*Predictor-Grade*	
Group	*Correlation*	*N*	*Correlation*	*N*
Males	0.37	133	0.48	130
Females	0.50	95	0.65	90
Whites	0.30	102	0.51	99
Nonwhites	0.50	126	0.57	121
Age 17 & Under	0.20	38	0.43	38
Age 18-21	0.44	135	0.55	129
Age 22 & Over	0.48	49	0.57	48
Total	0.42	228	0.54	220

All of these correlations (except the CCDT-Grade correlation for students age 17 and under) are statistically significant. Moreover, nearly all the correlations are comparable in value for all gender, ethnic, and age groups. Thus, both CCDT and the Predictor are significantly correlated with chemistry course performance for these gender, ethnic, and age groups.

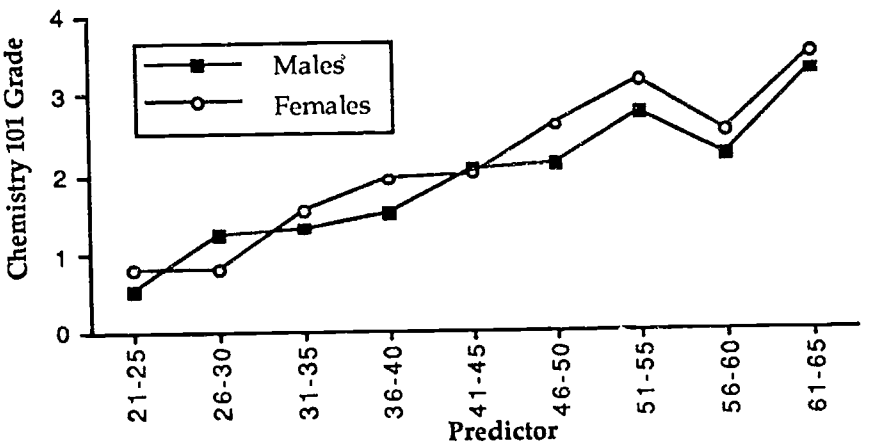
Differential prediction was also explored with regression analysis. Graph 3 (below) shows the prediction equations for gender. The x-axis represents student scores on the Predictor variable. The y-axis represents Chemistry 101 grade, coded numerically (A=4, B=3, etc.). The three lines in the graph represent regression equations for predicting Chemistry 101 grade. The dashed line is the equation for all students, and the lines for males and females are indicated in the figure. These lines are very similar, and their slopes are not significantly different, according to a t-test for slope differences.[6]

Graph 3. Regression of Predictor Equation Predicting Chemistry 101 Grade by Gender



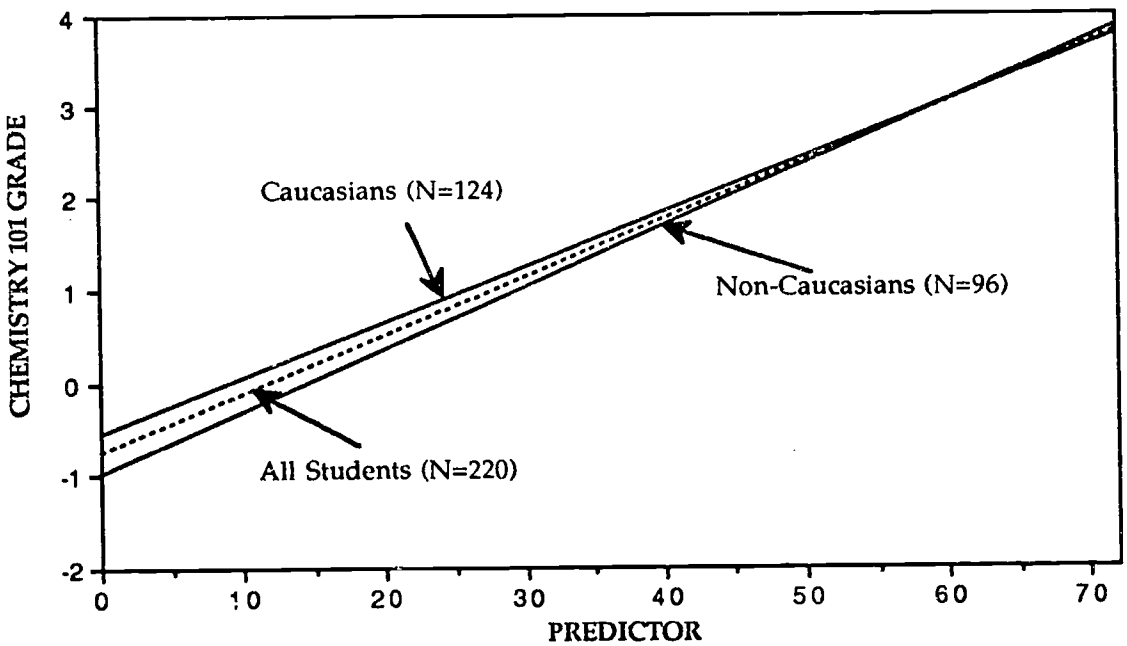
Additional analysis (see Graph 4 below) shows that males and females who score similarly on the Predictor achieve similar Chemistry 101 grades. As a predictor of general chemistry performance, this Predictor does not show marked differential prediction with regard to gender.

Graph 4. Course Performance by Predictor and Gender

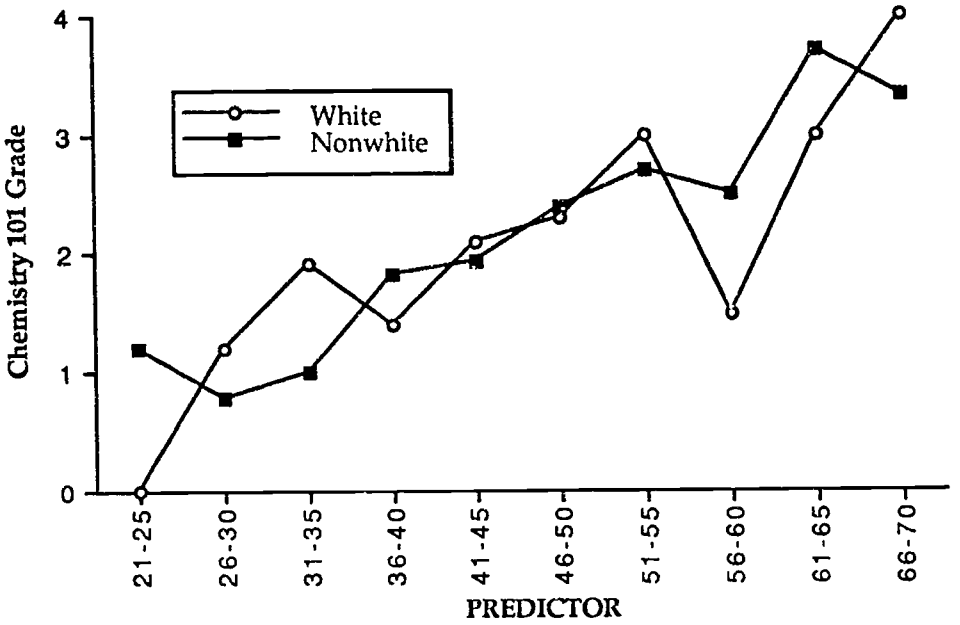


The Predictor shows little or no differential prediction with regard to ethnicity. Graph 5 (below) is a graph of the prediction equations for white students versus nonwhite students. Both regression lines are close to the dashed (total sample) line. The slopes of the white versus nonwhite lines are not significantly different. No evidence was therefore found for a differential prediction effect, comparing white versus nonwhite students. Further, white and nonwhite students who score similarly on the Predictor perform similarly in Chemistry 101 (see Graph 6, next page).

Graph 5. Regression of Predictor Equation Predicting Chemistry 101 Grade by Ethnicity



Graph 6. Chemistry Grade by Predictor and Ethnicity



CUT/PLACEMENT SCORES

Local Studies

The primary goal for establishing a preliminary cut score was to achieve a 70% success rate for students testing above the cut score and a 30% or lower success rate for students testing below the cut score. These guidelines were set by the local governance process. In addition to the guidelines, it would be advantageous to maximize the proportion of correctly placed students, compared to the baseline success rate. Table 6 shows various statistics for a range of plausible cut scores, based on the validation sample of 220 students.

Table 6. Cut Score Statistics for Predictor

Predictor Score	*% Ineligible Passing*	*% Eligible Passing*	*Correctly Placed*	*Adds to Baseline*	*Number Eligible*	*Number Eligible*	*% Eligible*
27	26.7%	75.1%	75.0%	3.2%	15	205	93.2%
28	30.4%	76.6%	75.9%	4.1%	23	197	89.5%
29	30.8%	76.4%	76.4%	4.5%	26	194	88.2%
30	*31.0%*	*78.0%*	*76.8%*	*5.0%*	*29*	*191*	*86.8%*
31	35.3%	78.5%	76.4%	4.5%	34	186	84.5%
32	*34.2%*	*79.7%*	*77.3%*	*5.5%*	*38*	*182*	*82.7%*
33	40.5%	79.2%	75.5%	3.6%	42	178	80.9%
34	*37.5%*	*81.4%*	*77.3%*	*5.5%*	*48*	*172*	*78.2%*
35	39.6%	82.0%	76.8%	5.0%	53	167	75.9%

As the second and third columns show, cut scores in the range 28-32 would result in approximately 70% of eligible students passing and 30% of ineligible students passing. These cutoffs would also place most students (about 77%) correctly.

The baseline rate of passing Chemistry 101 for this validation sample was 71.8% (i.e., without a cut score based on this predictor variable, 71.8% of these students passed Chemistry 101). Cut scores of 30, 32, and 34 increase the success rate by the largest margin. A cut score of 30 would place almost 77% of students correctly, and almost

87% of tested students would be eligible for Chemistry 101. A cut score of 32 would place almost 80% of students correctly, and about 83% of tested students would be eligible for Chemistry 101. A cut score of 34 would place over 81% of students correctly, and about 78% of tested students would be eligible for Chemistry 101.

CONCLUSION

The college is currently experimenting with cut score ranges that will optimize the placement process for students. Establishing and maintaining effective cutoffs is an ongoing process, requiring more than setting a single cutoff based on a single sample of students. Further research and tracking of students will help the college adjust cutoff ranges so that the greatest number of students is placed appropriately.

The validation procedure showed that the CCDT is an effective measure of chemistry skills at Glendale College. The test was developed with input from the chemistry faculty at Glendale, as well as other community colleges, assuring content-related validity. CCDT scores are excellent predictors of general chemistry course performance, showing high criterion-related validity. Scores are reliable as well. Additionally, the CCDT shows little or no disproportionate impact for different student groups. As an assessment instrument, the CCDT provides valid, reliable information about chemistry skills and knowledge to both students and instructors.

REFERENCES

[1]Russell, A. A. (1994). A rationally designed general chemistry diagnostic test. Journal of Chemical Education, 71 (4), 314-317.

[2]Ibid, page 315.

[3]Test Analysis: The California Chemistry Diagnostic Test Form C, California Chemistry Diagnostic Test Project, 1989, page 1.

[4]Ibid, page 3.

[5]Ibid, page 25.

[6]Howell, D. C. (1987). *Statistical methods for psychology*. Boston: PWS-Kent.