ASSESSING DIFFERENTIAL PREDICTION BIAS IN THE DEVELOPING COGNITIVE ABILITIES TEST ACROSS GENDER, RACE/ETHNICITY, AND SOCIOECONOMIC GROUPS

GARY L. CANIVEZ
Eastern Illinois University

TIMOTHY R. KONOLD
University of Virginia

This study investigated the differential predictive validity of the Developing Cognitive Abilities Test (DCAT) in a heterogeneous sample of 863 sixth-grade students. Bias was assessed across gender, race/ethnicity, and socioeconomic dimensions. Level H of the DCAT was administered during the students' sixth-grade year, and selected subtests from the Iowa Tests of Basic Skills (ITBS) were administered 8 months later during their seventh-grade year. Overall, results indicated a general lack of bias in the DCAT in predicting ITBS scores across gender, race/ethnicity, and socioeconomic groups.

The Developing Cognitive Abilities Test (DCAT) (Beggs & Mouw, 1980) is a group-administered test that reflects Bloom's (1956) hierarchy of cognitive development within three content areas (verbal, quantitative, and spatial). The DCAT was revised and restandardized in 1989 with items continuing to reflect Bloom's taxonomy. Basic cognitive abilities tap knowledge and comprehension, application abilities tap application, and critical thinking abilities tap analysis and synthesis abilities within each of the three content areas (verbal, quantitative, and spatial) (Wick, 1990). Designed to assess the cognitive abilities of children in Grades 1 through 12, each level (C/D

Correspondence concerning this article should be addressed to Gary L. Canivez, Ph.D., Department of Psychology, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920-3099; e-mail: cfglc@eiu.edu.

Educational and Psychological Measurement, Vol. 61 No. 1, February 2001 159-171

through L) of the DCAT is composed of 27 items for each content area for a total of 81 test items.

The DCAT is part of the Comprehensive Assessment Program that also includes the National Achievement Test and the School Attitude Measure. All three measures were conormed and nationally standardized from 1988 to 1989 using a stratified, multistage probability sample with 92,397 students. The DCAT "provides continuous measurement of student growth from grades 1 through 12" and is purported to be useful in educational decision making and student evaluation (Wick, 1990, p. 1). Other uses of the DCAT include simultaneous use with achievement batteries and attitude measures to develop profiles and identify discrepancies, strengths, and weaknesses that could be used in determining areas for intervention (Wick, 1990). The DCAT is also used as a screening measure for identifying potentially gifted students in an objective manner (Wick, 1990). Perhaps the most important distinguishing characteristic of the DCAT is the link between specific items and Bloom's (1956) cognitive taxonomy.

Criterion-related validity studies, although few, have shown the DCAT to have statistically significant positive relationships with other ability measures such as the Wechsler Intelligence Scale for Children–Revised (WISC-R) and the Slosson Intelligence Test (SIT) (Karnes & Lee, 1984; Karnes, Whorton, Currie, & Cantrall, 1986) but not with the Stanford Binet Intelligence Scale Form LM (Karnes et al., 1986). However, the reported correlations between the DCAT and the WISC-R and SIT were low (< .35) and accounted for less than 12% shared variance. These low correlations are probably the result of restricted range, as the samples used in these studies were composed of gifted students, a more homogeneous group than typically used in validity studies.

Farley and Elmore (1992) found that the three subtests of the DCAT predicted performance on the Iowa Silent Reading Test (ISRT) in a sample of 165 underachieving 1st-year college students. The DCAT Verbal subtest was a better predictor of ISRT performance than the DCAT Quantitative or Spatial subtests (the DCAT Total score was not examined). The correlations, however, were generally low with only one greater than .40 (DCAT Verbal–ISRT Vocabulary, $r = .49$). Perhaps due to restricted range, the DCAT subtests separately accounted for less than 12% of the achievement variability. However, multiple regression analyses indicated that both the DCAT Verbal and Spatial subtests added statistically significant contribution to prediction of various types of reading comprehension. Khanna and Leitner (1992) reported that the DCAT moderately predicted freshman college GPA at levels comparable to ACT scores. Khanna and Sheehan (1992) provided support for the construct validity of the DCAT and its use in assessing and identifying gifted students. They found that students scoring high on the Comprehensive

Tests of Basic Skills (CTBS) also performed better in Application and Critical Thinking items of the DCAT than students scoring low on the CTBS.

Henry and Bardo (1987) found statistically significant relationships between the DCAT and the Achievement Series Test (*Achievement Series Technical Manual*, 1979) in a heterogeneous sample of 7,007 students in 4th through 12th grades who were part of the DCAT standardization sample. Correlations were moderate to high in magnitude and provided better estimates of the true relations than previous published studies due to examination of a large sample of "normal" youths. Correlations between the DCAT Verbal subtest and Achievement Series Reading and Language subtests were higher than correlations between the DCAT Verbal subtest and the Achievement Series Mathematics subtest. Correlations between the DCAT Quantitative subtest and Achievement Series Mathematics subtest were higher than correlations between the DCAT Quantitative subtest and Achievement Series Reading and Language subtests, supporting the DCAT's construct (convergent) validity. Henry and Bardo (1990) later found the DCAT subtests and Total score moderately predicted later performance on the Medical College Admissions Test (MCAT) in a sample of 122 nontraditional premedical students. The DCAT Total score was the single best predictor of MCAT Biology, Physics, Science Problems, and Quantitative scores based on stepwise regression analyses.

Canivez (2000) replicated the findings of Henry and Bardo (1987) as the DCAT was found to have statistically significant and substantial predictive validity coefficients with the Iowa Tests of Basic Skills (ITBS) (Hieronymus et al., 1990). Canivez also found that the DCAT Verbal subtests correlated significantly higher with the ITBS Reading, Language, and Vocabulary subtests than did the DCAT Quantitative or Spatial subtests. The DCAT Quantitative subtest correlated significantly higher with the ITBS Mathematics subtest than did the DCAT Verbal or Spatial subtests, thus providing additional evidence to support the DCAT's construct validity.

Intelligence tests or measures of mental/cognitive abilities such as the DCAT are frequently used as predictors of school achievement. When considering the predictive validity of scores on cognitive ability tests, it is critical to examine the potential bias among ethnic minorities, gender, and socioeconomic status (SES). Reynolds and Kaiser (1990) indicated that test bias related to differential predictive validity is the most serious form of test bias. In studies of the WISC, WISC-R, and WISC-III (Wechsler, 1949, 1974, 1991), results indicated that they generally predict achievement equally well across different groups, and, although rare, when differential predictive validity has been found, achievement of minority students is typically *overpredicted* (Poteat, Wuensch, & Gregg, 1988; Reschly & Reschly, 1979; Reschly & Sabers, 1979; Reynolds & Gutkin, 1980; Reynolds & Hartlage,

1979; Weiss & Prifitera, 1995; Weiss, Prifitera, & Roid, 1993). In two separate studies of predictive validity, Canivez (1997, 1998) found that coefficients of the DCAT and ITBS did not differ between Caucasians and Black/African Americans, Hispanic/Latinos, or Native Americans, although these ethnic minorities obtained significantly lower scores than Caucasians on both the DCAT and ITBS. Although this preliminary evidence indicated that regression line slopes for Caucasians, Black/African Americans, Hispanic/Latinos, and Native Americans were similar (i.e., no slope bias), it is possible that *y*-intercept bias might be present, which might suggest that the DCAT may overpredict or underpredict ITBS scores for various minority groups. To adequately study both slope and *y*-intercept bias, the present study used multivariate techniques for testing both slope and intercept bias simultaneously (Potthoff, 1966; Reynolds & Kaiser, 1990).

Method

Participants

Participants from the Canivez (2000) study were used in the present study to investigate bias in the DCAT and were previously described in detail. The sample ($N = 863$) was composed of approximately equal numbers of males (48%) and females (52%). These students were heterogeneous with respect to both race/ethnicity and SES. Racial/ethnic classifications included Caucasians ($n = 566$), Black/African Americans ($n = 77$), Hispanic/Latinos ($n = 170$), Native Americans ($n = 28$), and Asian Americans ($n = 22$). SES classifications were made by considering the student's eligibility for the free or reduced school lunch program within the school district that is based on family size and income. Accordingly, four categories were identified: no free or reduced lunch ($n = 558$), reduced lunch ($n = 70$), free lunch ($n = 130$), and free lunch with Aid to Families With Dependent Children (AFDC) ($n = 105$). Of the total sample, 778 (90.2%) students' primary home language was English, 62 (7.2%) students' primary home language was Spanish, and the remaining 23 (2.6%) students' primary home languages was some type of Asian or Native American language. All students in the present sample were determined to be English proficient by school personnel. Students in special education were not specifically excluded from administration of the DCAT or ITBS, but data on special education participation were not available and thus unknown in this sample.

Instruments

Predictor. The DCAT (Wick, Beggs, & Mouw, 1989) is a group-administered test of mental abilities for students in Grades 1 through 12 (Level C/D

through Level L), which includes Verbal, Quantitative, and Spatial subtests that combine to provide a Total score. Internal consistency (KR-20) coefficients for each grade, level, content area, and thinking skills area are presented in the technical manual and are acceptable, ranging from .70 to .96. Most of the internal consistency estimates were in the mid-.80s (Wick, 1990). As expected, internal consistency estimates were highest for the DCAT Total score, a global composite score, ranging from .88 (Grade 1, Level C/D) to .96 (Grade 4, Level F).

A review of the original DCAT by Fox (1985) indicated that score reliability was good but that validity research was greatly needed. Aylward's (1992) review of the current edition of the DCAT was positive, indicating that it was well designed, psychometrically sound, and provided a broad range of assessment requiring recall, recognition, application, transformation, and integration skills. Aylward also noted the lack of validity data in the DCAT manual.

Criterion. The ITBS (Hieronymus et al., 1992) is a popular group-administered test of academic achievement composed of Vocabulary, Reading, Language Usage, Work Study, and Mathematics subtests. Test-retest stability coefficients over a 1-year time interval were mostly in the .70 to .90 range, and internal consistency and alternate forms reliability coefficients were in the .80s and .90s (Gregory, 1996). Reviews by Lane (1992) and Raju (1992) were positive, noting sound measurement practices, high technical standards (i.e., internal consistency and alternate forms reliability), and good content validity.

Procedure

Students were administered Level H of the DCAT in March 1993 by their sixth-grade classroom teachers as part of the school district's gifted education screening and evaluation process. Test answer forms were sent to the test publisher for scoring, and results were returned to the school district on a data disk. The DCAT data set included raw scores and percentile ranks. Selected subtests (Vocabulary, Reading, Language Usage, Mathematics Problem Solving) from the ITBS were administered in October 1993 during the students' seventh-grade year as part of the state-mandated academic achievement testing program. This provided a natural opportunity to investigate the short-term predictive validity of DCAT scores. ITBS results were also provided to the district on data disks. The ITBS data set included raw scores, grade equivalent scores, percentile ranks, and normal curve equivalent scores. DCAT and ITBS data sets were merged using the common student identification number for analyses. All data analyses used raw score data due to the absence of standard scores for the DCAT.

Data Analyses

A thorough investigation of differential criterion-related validity (predictive validity) involves joint consideration of both *y*-intercept and slope differences among groups (Potthoff, 1966; Reynolds & Kaiser, 1990). Potthoff's (1966) procedure is often used for addressing issues of differential prediction (Bossard, Reynolds, & Gutkin, 1980; Glutting, 1986; Glutting, Oakland, Konold, 1994; Weiss & Prifitera, 1995) because it provides a *simultaneous F* test for both intercept and slope differences, thereby controlling Type I error rates. Following the identification of a statistically significant omnibus *F*, the procedure provides follow-up comparisons to detect whether differences exist between *y* intercepts, slopes, or both. When either *y*-intercept or slope differences are found, the use of a common regression equation generally results in underpredicting criterion performance for the group with the higher mean criterion score. In the former case, errors of prediction are constant across all points of the predictor. However, slope differences suggest nonparallel regression lines and nonconstant errors of prediction, wherein the size of the errors vary across different points of the predictor scale. Nonconstant errors of prediction are also problematic when both *y*-intercept and slope differences are observed; here, however, interpretations become more challenging because the direction of bias may change if regression lines cross (see Reynolds & Kaiser, 1990).

Potthoff's (1966) procedure was used to investigate criterion-related (predictive validity) bias with the DCAT and ITBS. Students' DCAT Total raw scores were used to predict the ITBS Reading, Language, Mathematics, and Vocabulary raw scores. Equality of slopes and *y* intercepts was investigated between classifications of race/ethnicity (Caucasians vs. African Americans, Caucasians vs. Hispanics), gender (males vs. females), and SES (no free/reduced lunch vs. reduced lunch, no free/reduced lunch vs. free lunch, no free/reduced lunch vs. free lunch and AFDC) for each of the four ITBS achievement scores. Type I error rates among the simultaneous tests were controlled through Bonferroni adjustments within each achievement criterion under consideration ($n = 6$). Groups are said to be underpredicted when the use of a common regression equation would result in a lower predicted score. Native Americans and Asian Americans were not included in the analyses because of their low prevalence in the present sample. The small sample sizes of these two groups were believed to preclude the ability to obtain stable parameter estimates and thereby render subsequent results inconclusive.

Groups demonstrating statistically significant simultaneous tests were subsequently evaluated for *y*-intercept and slope differences. Because statistically significant slope and/or intercept differences may be small from a practical perspective (Weiss & Prifitera, 1995), effect sizes ( ) are provided in absolute value for all statistically significant contrasts. The effect size for statistically significant *y*-intercept differences were calculated by comparing

the average predicted score obtained from the separate group regression equation to the average predicted score that would be obtained from a common equation. Effect sizes for statistically significant slope differences were obtained by comparing predicted scores between the separate group equations and the common regression equation at various points across the predictor scale. Cohen's (1988) recommendations for small (.2), medium (.5), and large (.8) effect sizes served as a rough guideline for interpretation.

Results

Table 1 provides the sample means and standard deviations for the ITBS achievement raw scores and DCAT Total raw scores. These are presented for the total sample as well as by race/ethnicity, gender, and SES. Predictive validity coefficients for the DCAT Total raw score and ITBS achievement raw scores are presented in Table 2. Coefficients for Reading, Language, Mathematics, and Vocabulary were all moderately high and statistically significant when the total sample was considered (all ps < .05). A similar trend was observed when validity coefficients were investigated separately within each of the nine demographic groups under consideration. These moderately high coefficients were all found to be statistically significant (ps < .05).

The more important question in test bias investigations, however, is not whether validity coefficients exceed zero but whether differential validity exists between groups (Flaughter, 1978; Humphreys, 1973). Differential validity was addressed by testing for constant errors in prediction (i.e., differences in regression slopes and criterion intercepts) through a set of 24 analyses using Potthoff's (1966) procedure. No statistically significant differences were observed for 20 comparisons when Type I error rates were controlled as reported in Table 3. The 4 statistically significant comparisons occurred when predicting Reading (males vs. females and no free/reduced lunch vs. free lunch), Language (males vs. females), and Vocabulary (Caucasians vs. Hispanics). No constant errors in prediction were observed among the six demographic contrasts when predicting Mathematics achievement.

Follow-up tests of slope and y-intercept differences were conducted among the four demographic contrasts where bias was indicated. These results revealed only y-intercept differences among three of the four contrasts. In these instances, the bias operated to underpredict achievements for groups displaying higher criterion scores: Caucasians were underpredicted when compared to Hispanics on the ITBS Vocabulary subtest, $F(1, 733) = 9.03$, $p < .01$, $_{\text{Caucasian}} = .05$, $_{\text{Hispanic}} = .13$, and females were underpredicted when compared to males on both the ITBS Reading, $F(1, 860) = 41.19$, $p < .001$, $_{\text{females}} = .14$, $_{\text{males}} = .15$, and Language subtests, $F(1, 860) = 97.55$, $p < .001$, $_{\text{females}} = .22$, $_{\text{males}} = .24$.

The remaining contrast that demonstrated bias revealed both slope, $F(1, 684) = 7.18$, $p < .01$, and y-intercept differences, $F(1, 685) = 3.87$, $p < .05$,

Table 1
Raw Score Means and Standard Deviations of the Iowa Tests of Basic Skills (ITBS) and Developing Cognitive Abilities Test (DCAT) Scales for the Total Sample and Demographic Groups

	ITBS				DCAT
Group	Reading	Language	Mathematics	Vocabulary	Total
Total sample					
M	28.8	27.0	15.5	23.4	45.9
SD	10.1	7.3	5.6	6.4	10.8
Race					
Caucasian					
M	30.5	28.3	16.5	24.8	48.4
SD	10.1	7.3	5.6	6.4	10.8
Black/African American					
M	24.2	23.1	12.3	19.8	36.8
SD	7.9	6.6	4.2	5.6	9.2
Hispanic/Latino					
M	26.0	24.8	13.5	20.8	41.7
SD	8.9	6.7	5.4	6.0	10.4
Gender					
Male					
M	27.6	25.4	15.9	23.3	46.3
SD	10.0	7.4	5.9	6.9	11.7
Female					
M	30.0	28.4	15.1	23.5	45.4
SD	9.7	6.9	5.4	6.2	11.0
Lunch status (socioeconomic status)					
No free/reduced lunch					
M	30.5	28.1	16.2	24.4	47.6
SD	10.1	7.1	5.7	6.4	11.3
Reduced lunch					
M	27.6	25.9	15.2	22.9	46.0
SD	9.5	7.6	5.5	6.4	10.5
Free lunch					
M	26.6	25.5	14.3	21.7	43.5
SD	9.3	7.0	5.5	6.6	11.3
Free lunch and AFDC					
M	23.9	23.7	13.1	20.5	39.6
SD	7.6	6.8	4.8	6.0	9.4

Note. AFDC = Aid to Families With Dependent Children.

between children receiving free lunch and children not receiving free/reduced lunch on the ITBS Reading subtest. When predicted scores from these two groups were compared to a common regression equation, children not receiving free/reduced lunch and obtaining very low DCAT scores (i.e., three standard deviations below the mean) were slightly overpredicted, = .02. The same group of students was underpredicted in comparison to a com-

Table 2
Pearson Product-Moment Correlation Coefficients Between the Developing Cognitive Abilities Test (DCAT) and the Iowa Tests of Basic Skills (ITBS) for Total Sample and Nine Demographic Subgroups

	ITBS			
DCAT Total by Group	Reading	Language	Mathematics	Vocabulary
Total sample	.72	.70	.74	.73
Caucasian	.72	.67	.74	.72
Black/African American	.55	.64	.62	.74
Hispanic/Latino	.68	.68	.69	.66
Male	.74	.71	.76	.71
Female	.72	.72	.72	.76
No free/reduced lunch	.73	.70	.75	.74
Reduced lunch	.68	.64	.67	.69
Free lunch	.60	.62	.68	.70
Free lunch and AFDC	.69	.66	.71	.63

Note. AFDC = Aid to Families With Dependent Children. All correlation coefficients were statistically significant, $p < .05$, when Type I error rates were controlled through Bonferroni adjustments within the total sample ($n = 4$) and each demographic category (i.e., ethnicity, $n = 12$; gender, $n = 8$; and SES, $n = 16$).

mon regression equation when their obtained DCAT score was at or above two standard deviations below the mean. Although the effect sizes for this group of children were small, the values demonstrated a steady increase across the DCAT scale, from a low of = .002 at two standard deviations below the mean to a high of = .12 at three standard deviations above the mean. The direction of bias for children receiving free lunch was also found to be dependent on their DCAT score. Children receiving free lunch and scoring low on the DCAT (i.e., at or below one standard deviation below the mean) would be underpredicted on the ITBS Reading subtest if a common regression equation were employed. The largest effect size for this group of children was obtained for those obtaining DCAT scores three standard deviations below the mean, = .34. Thereafter, effect sizes declined as DCAT scores increased from two standard deviations below the mean, = .19, to one standard deviation below the mean, = .04. Conversely, children receiving free lunch and obtaining DCAT scores at or above the mean were overpredicted on the ITBS Reading subtest when a common regression equation was used. Effect sizes for this group of children demonstrated a steady increase for children obtaining DCAT scores at the mean to those scoring three standard deviations above the mean, $_{\text{mean}} = .11$, $_{\text{+ one standard deviation}} = .26$, $_{\text{+ two standard deviations}} = .41$, $_{\text{+ three standard deviations}} = .56$.

Discussion

Results of the present study are quite similar to results obtained with other measures of intellectual or cognitive abilities (Poteat et al., 1988; Reschly &

Table 3
Effect Sizes and F *Values (and degrees of freedom) for Simultaneous Slope and* y*-Intercept Comparisons Between Developing Cognitive Abilities Test (DCAT) Total Score and Iowa Tests of Basic Skills (ITBS) Scales*

	ITBS			
Demographic Comparison	Reading	Language	Mathematics	Vocabulary
Caucasian versus Black/African American	0.62	0.71	0.76	0.78
	3.64	0.01	2.10	0.08
	(2,639)	(2,639)	(2,639)	(2,639)
Caucasian versus Hispanic/Latino	0.45	0.48	0.54	0.62
	1.45	0.52	1.47	5.33*
	(2,732)	(2,732)	(2,732)	(2,732)
Male versus female	0.23	0.40	0.15	0.03
	20.60**	48.72**	2.59	1.77
	(2,859)	(2,859)	(2,859)	(2,859)
No free/reduced lunch versus reduced lunch	0.29	0.30	0.18	0.24
	2.22	2.55	0.44	1.16
	(2,624)	(2,624)	(2,624)	(2,624)
No free/reduced lunch versus free lunch	0.40	0.35	0.34	0.43
	5.54*	2.05	1.23	2.50
	(2,684)	(2,684)	(2,684)	(2,684)
No free/reduced lunch versus free lunch and AFDC	0.66	0.60	0.56	0.62
	2.81	1.41	0.12	0.86
	(2,659)	(2,659)	(2,659)	(2,659)

Note. AFDC = Aid to Families With Dependent Children. Tabled effect sizes (reported above *F* values) were calculated by comparing the average predicted scores obtained from the separate group regression equations. Type I error rates were controlled through Bonferroni adjustments applied within each matrix to the number of simultaneous-test comparisons ($n = 6$).
*$p < .05$. **$p < .01$.

Reschly, 1979; Reschly & Sabers, 1979; Reynolds & Gutkin, 1980; Reynolds & Hartlage, 1979; Weiss et al., 1993; Weiss & Prifitera, 1995). The observation of no differential prediction in ITBS Reading, Language, Mathematics, or Vocabulary performance for Black/African American students in the present study is consistent with results found with the WISC-III and WIAT (Weiss & Prifitera, 1995). Among Hispanic/Latino students, only the ITBS Vocabulary subtest was differentially predicted in the present study and was restricted to *y*-intercept bias, indicating that ITBS Vocabulary scores of Hispanic/Latino students were overpredicted by the DCAT. However, the practical significance of this overprediction was found to be trivial, = .13.

Contrasts related to gender found no evidence of slope bias but yielded differential *y* intercepts for ITBS Reading and Language in which male ITBS

performance was overpredicted. Weiss and Prifitera (1995) also found the WISC-III to overpredict reading performance on the WIAT for males, but no differential prediction was observed in the language area. Here again, the overprediction observed for males in the present study was found to be of little practical significance for both Reading, = .15, and Language, = .24.

SES was the only variable to show both slope and *y*-intercept bias, which was limited to the ITBS Reading subtest comparison between the no free lunch and free lunch groups. The intersecting regression lines for these two groups make interpretation somewhat confusing in that ITBS Reading scores were underpredicted for those receiving free lunch when their DCAT scores were low, whereas ITBS Reading was overpredicted for those receiving free lunch when their DCAT scores were high. However, interpretations become clearer when effect sizes are considered with respect to Cohen's (1988) classification system. In this context, children receiving free lunch and obtaining DCAT scores two or more standard deviations above the mean were found to be overpredicted on the ITBS Reading subtest with a medium effect size. All other predicted differences for this group can be considered small in practical terms.

Limitations of the present study include examining the predictive validity of only one level of the DCAT (Level H). Thus, results may not generalize to other age groups or levels of the DCAT. Also, the time delay in administering the ITBS was approximately 8 months. Longer time intervals between DCAT and ITBS testing should be the topic of future research to examine the influence of time on the differential predictive validity of the DCAT. Finally, these data were obtained from one metropolitan school district in the Southwest, and although the sample was diverse with respect to race/ethnicity, gender, home language, and SES, it was not selected to be representative of the larger national population. Future studies should attempt to use more representative samples to generalize to the larger population. These findings, however, are certainly encouraging in that the DCAT appears to provide a generally unbiased assessment of cognitive abilities across race/ethnicity, gender, and SES.

References

Achievement series technical manual: Preliminary report no. 2. (1979). Glenview, IL: Scott, Foresman.

Aylward, G. P. (1992). Review of the Developing Cognitive Abilities Test (2nd ed.). In J. J. Conoley & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook* (pp. 279-281). Lincoln: Buros Institute of Mental Measurements, University of Nebraska.

Beggs, D. L., & Mouw, J. T. (1980). *Developing Cognitive Abilities Test*. Carbondale: Southern Illinois University Press.

Bloom, B. S. (1956). *Taxonomy of educational objectives handbook I: The cognitive domain*. New York: David McKay.

Bossard, M. D., Reynolds, C. R., & Gutkin, T. B. (1980). A regression analysis of test bias on the Stanford-Binet Intelligence Scale for Black and White children referred for psychological services. *Journal of Clinical Child Psychology*, *9*, 52-54.

Canivez, G. L. (1997, August). *Developing Cognitive Abilities Test (DCAT): Investigating psychometric characteristics and racial/ethnic bias*. Paper presented at the annual meeting of the American Psychological Association, Chicago.

Canivez, G. L. (1998, August). *Developing Cognitive Abilities Test (DCAT): Predictive validity and racial/ethnic bias*. Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Canivez, G. L. (2000). Predictive and construct validity of the Developing Cognitive Abilities Test: Relations with the Iowa Tests of Basic Skills. *Psychology in the Schools*, *37*, 1-7.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Farley, M. J., & Elmore, P. B. (1992). The relationship of reading comprehension to critical thinking skills, cognitive ability, and vocabulary for a sample of underachieving college freshmen. *Educational and Psychological Measurement*, *52*, 921-931.

Flaughter, R. L. (1978). The many definitions of test bias. *American Psychologist*, *33*, 671-679.

Fox, L. H. (1985). Review of Developing Cognitive Abilities Test. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (Vol. 1, pp. 460-461). Lincoln: Buros Institute of Mental Measurements, University of Nebraska.

Glutting, J. J. (1986). Potthoff bias analyses for K-ABC MPC and Nonverbal Scale IQs among Anglo, Black, and Puerto Rican kindergarten children. *Professional School Psychology*, *1*, 225-234.

Glutting, J. J., Oakland, T., & Konold, T. R. (1994). Criterion-related bias with the guide to the assessment of test-session behavior for the WISC-III and WIAT: Possible race/ethnicity, gender, and SES effects. *Journal of School Psychology*, *32*, 355-369.

Gregory, R. J. (1996). *Psychological testing: History, principles, and applications* (2nd ed.). Boston: Allyn & Bacon.

Henry, P., & Bardo, H. R. (1987). The predictive validity of the Developing Cognitive Abilities Test. *Educational and Psychological Measurement*, *47*, 207-214.

Henry, P., & Bardo, H. R. (1990). Relationship between scores on Developing Cognitive Abilities Test and scores on Medical College Admissions Test for nontraditional premedical students. *Psychological Reports*, *67*, 55-63.

Hieronymus, A. N., Hoover, H. D., Oberley, K. R., Cantor, N. K., Frisbie, D. A., Dunbar, S. B., Lewis, J. C., & Lindquist, E. F. (1990). *Iowa Test of Basic Skills*. Chicago, IL: Riverside.

Humphreys, L. G. (1973). Statistical definitions of test validity for minority groups. *Journal of Applied Psychology*, *58*, 1-4.

Karnes, F. A., & Lee, L. A. (1984). Correlations of scores on the Developing Cognitive Abilities Test and the Wechsler Intelligence Scale for Children–Revised for intellectually gifted students. *Psychological Reports*, *54*, 373-374.

Karnes, F. A., Whorton, J. E., Currie, B. B., & Cantrall, S. W. (1986). Correlations of scores on the WISC-R, Stanford-Binet, The Slosson Intelligence Test, and the Developing Cognitive Abilities Test for intellectually gifted youth. *Psychological Reports*, *58*, 887-889.

Khanna, R. K., & Leitner, D. W. (1992, October). *An assessment of GPA prediction: DCAT as a possible solution*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago.

Khanna, R. K., & Sheehan, J. (1992, October). *The construct validity of the DCAT*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago.

Lane, S. (1992). Review of the Iowa Tests of Basic Skills. In J. J. Conoley & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook* (pp. 421-423). Lincoln: Buros Institute of Mental Measurements, University of Nebraska.

Poteat, G. M., Wuensch, K. L., & Gregg, N. B. (1988). An investigation of differential prediction with the WISC-R. *Journal of School Psychology*, *26*, 59-68.

Potthoff, R. F. (1966). *Statistical aspects of the problem of bias in psychological tests* (Institute of Statistics Mimeo Series No. 479). Chapel Hill: University of North Carolina Department of Statistics.

Raju, N. S. (1992). Review of the Iowa Tests of Basic Skills. In J. J. Conoley & J. C. Conoley (Eds.), *The eleventh mental measurements yearbook* (pp. 423-424). Lincoln: Buros Institute of Mental Measurements, University of Nebraska.

Reschly, D. J., & Reschly, J. E. (1979). Validity of WISC-R factor scores in predicting achievement and attention for four sociocultural groups. *Journal of School Psychology*, *17*, 355-361.

Reschly, D. J., & Sabers, D. (1979). Analysis of test bias in four groups with the regression definition. *Journal of Educational Measurement*, *16*, 1-9.

Reynolds, C. R., & Gutkin, T. B. (1980). A regression analysis of test bias on the WISC-R for Anglos and Chicanos referred to psychological services. *Journal of Abnormal Child Psychology*, *8*, 237-243.

Reynolds, C. R., & Hartlage, L. C. (1979). Comparison of WISC and WISC-R regression lines for academic prediction with Black and with White referred children. *Journal of Consulting and Clinical Psychology*, *47*, 589-591.

Reynolds, C. R., & Kaiser, S. M. (1990). Test bias in psychological assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed., pp. 487-525). New York: John Wiley.

Wechsler, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children–Third Edition*. San Antonio, TX: Psychological Corporation.

Weiss, L. G., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology*, *33*, 297-304.

Weiss, L. G., Prifitera, A., & Roid, G. (1993). The WISC-III and the fairness of predicting achievement across ethnic and gender groups. Advances in psychoeducational assessment; The Wechsler Intelligence Scale for Children: Third Edition. *Journal of Psychoeducational Assessment*, Monograph Series, 35-42.

Wick, J. W. (1990). *Developing Cognitive Abilities Test: Technical manual prepublication edition*. Chicago: American Testronics.

Wick, J. W., Beggs, D. L., & Mouw, J. T (1989). *Developing Cognitive Abilities Test* (2nd ed.). Chicago: American Testronics.