Psychometric Properties 3

Reliability

The reliability of an assessment instrument like the DECA-P2 is defined as "the consistency of scores obtained by the same person when reexamined with the same test on different occasions, or with different sets of equivalent items, or under other variable examining conditions" (Anastasi, 1988, p. 102). DECA-P2 scale reliability was assessed using several methods. First, the internal reliability coefficient for each scale was computed. Second, test–retest reliability of each scale was assessed. Finally, interrater reliability (two raters evaluating the same child) for each scale was determined.

Internal Reliability

Internal reliability (or internal consistency) refers to the extent to which the items on the same scale or assessment instrument measure the same underlying construct. We determined internal consistency using Cronbach's alpha (Cronbach, 1951). The internal reliability coefficients were based on the individuals included in the DECA-P2 standardization sample. The internal consistency estimates for each scale were calculated according to rater and are presented in Table 3.1. The results indicate that the DECA-P2 scales have excellent internal reliability. The Total Protective Factors scale reliability was computed using the formula provided by Nunnally and Bernstein (1994) for the reliability of a linear combination. The Total Protective Factors coefficients for parent raters (.92) and teacher raters (.95) both exceed the .90 value for a total score suggested by Bracken (1987) and also meet the "desirable standard" described by Nunnally (1978, p. 246).

The internal reliability coefficients for the three protective factor scales range from a low of .79 (Attachment/Relationships—Parent Raters) to a high of .94 (Self-Regulation—Teacher Raters). The median reliability coefficient across these three scales was .88 for parent raters and .92 for teacher raters. These median values well exceed the .80 minimum suggested by Bracken (1987). The internal reliability coefficients for the Behavioral Concerns scale, .80 for parent raters and .86 for teacher raters, also met or exceeded Bracken's recommended minimum.

Table 3.1

Internal Reliability (Alpha) Coefficients for the DECA-P2 Scales by Rater

	Raters	
Scales	**Parents**	**Teachers**
Total Protective Factors	.92	.95
Initiative	.88	.92
Self-Regulation	.90	.94
Attachment/Relationships	.79	.85
Behavioral Concerns	.80	.86

Standard Errors of Measurement

The standard error of measurement (*SEM*) is an estimate of the amount of error in observed scores, expressed in standard score units (i.e., *T*-scores). As such, the *SEM* provides an estimate of the amount of fluctuation in DECA-P2 scores that can be expected by chance; the larger the *SEM*, the greater the amount of chance fluctuation. We obtained the *SEM* for each of the DECA-P2 scale *T*-scores directly from the internal reliability coefficients using the formula

$$SEM = SD\sqrt{1 - reliability}$$

where *SD* is the theoretical standard deviation of the *T*-score (10) and the appropriate reliability coefficient is used. The *SEM*s for each DECA-P2 scale are presented in Table 3.2 according to rater. Note that the values of the *SEM* vary with the size of the reliability coefficient found in Table 3.1—the higher the reliability, the smaller the *SEM*. Because the reliability coefficients

in Table 3.1 meet or exceed recommended standards in the field, the *SEM*s in Table 3.2 are relatively small.

Table 3.2

Standard Errors of Measurement for the DECA-P2 Scale *T*-Scores by Rater

	Raters	
Scales	**Parents**	**Teachers**
Total Protective Factors	2.83	2.23
Initiative	3.46	2.83
Self-Regulation	3.16	2.45
Attachment/Relationships	4.58	3.87
Behavioral Concerns	4.47	3.74

Test–Retest Reliability

The correlation between scores obtained for the same child on two separate occasions is another indicator of the reliability of an assessment instrument. The correlation of this pair of scores is the test–retest reliability coefficient *(r)*, and the magnitude of the obtained value informs us about the degree to which random changes influence the scores (Anastasi, 1988).

To investigate the test–retest reliability of the DECA-P2, a group of parents (n = 53) and a group of teachers (n = 37) rated the same child on two different occasions separated by an interval of 6 to 8 days. Demographic information about this diverse convenience sample is provided in Table 3.3.

The results of this study are shown in Table 3.4. All of the correlations are significant ($p < .01$) and high in magnitude ranging from r = .78 (Behavioral Concerns—Parent Raters) to r = .94 (Self-Regulation—Teacher Raters). The median test–retest reliability coefficients for the three protective factor scales are .86 and .90 for parent and teacher raters, respectively. The test–retest reliability coefficients for the Total Protective Factors scale are .88 for parent raters and .95 for teacher raters. For the Behavioral Concerns scale, the reliability coefficients are .78 and .80 for parent and teacher raters, respectively. These findings indicate that all of the DECA-P2 scales have good test–retest reliability.

Table 3.3

Sample Characteristics for the DECA-P2 Test-Retest Reliability Study

	Parent Sample		Teacher Sample	
	n	%	*n*	%
Size of Sample	53		37	
Age				
Mean	4 years, 5 months		4 years, 3 months	
SD	8.7 months		9.1 months	
Gender				
Boys	22	42	20	54
Girls	31	58	17	46
Race				
American Indian/Alaskan Native	0	0	0	0
Asian	0	0	0	0
Black/African American	5	11	1	3
Native Hawaiian/Pacific Islander	0	0	0	0
White	35	74	33	89
Two or More Races	7	15	2	5
Other	0	0	1	3
Hispanic Ethnicity	8	15	2	5
Region of Residence				
Northeast	19	36	14	38
South	0	0	0	0
Midwest	30	57	20	54
West	4	7	3	8
Other/Missing	0	0	0	0

Table 3.4

Test-Retest Reliability Coefficients for Two DECA-P2 Ratings by the Same Parent or Same Teacher for the Same Child Over a 6- to 8-Day Interval

	Raters	
Scales	**Parents**	**Teachers**
Total Protective Factors	.88*	.95*
Initiative	.88*	.89*
Self-Regulation	.86*	.94*
Attachment/Relationships	.81*	.90*
Behavioral Concerns	.78*	.80*

*$p < .01$

Interrater Reliability

The correlation between scores obtained for the same child by two different raters who observe the same child in the same environment at approximately the same time is an indicator of the interrater reliability of an assessment instrument. The magnitude of the correlations between these scores tells us about the degree of similarity in the different raters' perceptions of the child's behavior. Therefore, we examined the interrater reliability of the DECA-P2 by comparing ratings obtained from two parents who live in the same household with the child (n = 31) or two teachers, or a teacher and teacher aide,

Table 3.5

Sample Characteristics for the DECA-P2 Interrater Reliability Study

	Parent Sample		Teacher Sample	
	n	%	*n*	%
Size of Sample	31		52	
Age				
Mean	4 years, 3 months		4 years, 3 months	
SD	9.8 months		8.7 months	
Gender				
Boys	15	48	19	36
Girls	16	52	33	64
Race				
American Indian/Alaskan Native	0	0	0	0
Asian	3	10	2	4
Black/African American	1	3	17	33
Native Hawaiian/Pacific Islander	0	0	0	0
White	27	87	26	50
Two or More Races	0	0	3	6
Other	0	0	2	4
Don't know	0	0	2	4
Hispanic Ethnicity	11	35	9	17
Region of Residence				
Northeast	13	42	8	15
South	8	26	34	65
Midwest	1	3	10	20
West	9	29	0	0

who work in the same classroom (n = 52). In these studies, the sample size (n) refers to the number of *pairs* of adults rating the child. Demographic information about these two samples is presented in Table 3.5.

The correlations of a set of ratings obtained for the same children by two parents or two teachers (or a teacher and a teacher aide) are provided in Table 3.6. Because the samples were relatively small, the standard deviations were not always exactly equal to the normative value of 10, so the correlation coefficients were corrected for this inconsistency in range. The corrected correlations were calculated applying the formula provided by Guilford and Fruchter (1978) to the obtained correlation between the two sets of ratings. That is, the obtained correlation was corrected first for one sample standard deviation and then for the second sample standard deviation. In some instances, the corrected correlation is lower than the original one because the standard deviation of the sample exceeded the normative value of 10. Inversely, when the standard deviation of the sample was less than 10, the corrected value was higher than the

Table 3.6

Interrater Reliability Coefficients for Two DECA-P2 Ratings by Two Parents or Two Teachers for the Same Child

Scales	Mean Rater 1	*SD* Rater 1	Mean Rater 2	*SD* Rater 2	Obtained *r*	Corrected *r*
	Parents (n = 31)					
Total Protective Factors	48.7	7.3	49.0	7.9	.33	.51*
Initiative	50.1	8.5	50.5	9.2	.49*	.59*
Self-Regulation	48.8	7.4	48.0	8.7	.26	.39
Attachment/ Relationships	48.6	4.9	48.7	7.3	.56*	.76*
Behavioral Concerns	48.9	8.7	49.5	10.2	.42	.46*
	Teachers (n = 52)					
Total Protective Factors	51.2	9.8	51.2	9.8	.71*	.72*
Initiative	50.1	8.6	49.0	8.5	.66*	.77*
Self-Regulation	52.8	10.8	52.8	10.8	.74*	.68*
Attachment/ Relationships	52.9	10.2	52.9	10.2	.37*	.36*
Behavioral Concerns	47.1	9.9	47.5	8.5	.64*	.70*

*$p < .01$

original obtained correlation. The corrected correlation coefficients provide a better estimate of the correlations that would be found in the population of preschool children and are discussed below.

Both the obtained and corrected correlations are presented in Table 3.6. These results indicate that pairs of parents or pairs of teachers who saw the children in the same environment at the same time rated the children similarly. All of the correlations for teacher raters and all but one of the correlations for parent raters are significant ($p < .01$). The Total Protective Factors scale correlations are .51 for parent raters and .72 for teacher raters. For the three individual scales, the values range from .36 (Attachment/Relationships—Teacher Raters) to .77 (Initiative—Teacher Raters). The median correlation coefficients for the three protective factor scales are .59 and .68 for parent and teacher raters, respectively. The correlation coefficients for the Behavioral Concerns scale are .46 and .70 for parent and teacher raters, respectively.

Stability of DECA-P2 Ratings

The correlation coefficients reported for the test–retest and interrater reliability studies indicate that the pairs of raters in each study ranked the children similarly. However, the coefficients do not indicate the actual similarity in the scores. Tables 3.7a and 3.7b provide the pretest and posttest mean scale scores and standard deviations received by the children in the test–retest study by parents and teachers, respectively.

Table 3.7a

Test–Retest *T*-Score Stability: Pretest and Posttest Mean Scale Scores and Standard Deviations–Parent Raters

	Pretest		Posttest	
Scales	**Mean**	***SD***	**Mean**	***SD***
Total Protective Factors	51.5	9.9	51.7	11.2
Initiative	51.1	9.5	51.8	10.1
Self-Regulation	50.9	9.7	50.9	10.9
Attachment/Relationships	52.1	8.8	51.5	10.7
Behavioral Concerns	49.0	11.0	48.9	10.6

Table 3.7b

Test–Retest *T*-Score Stability: Pretest and Posttest Mean Scale Scores and Standard Deviations–Teacher Raters

	Pretest		Posttest	
Scales	**Mean**	***SD***	**Mean**	***SD***
Total Protective Factors	48.5	9.0	48.5	8.2
Initiative	47.8	8.7	48.6	7.7
Self-Regulation	47.4	9.7	46.8	9.1
Attachment/Relationships	50.6	9.0	50.6	8.7
Behavioral Concerns	53.4	9.2	53.3	9.0

For parent raters, on average, the absolute value of the pretest–posttest mean score difference was less than one *T*-score point on the three protective factor scales (0.43), on the Total Protective Factors scale (0.2), and on the Behavioral Concerns scale (0.1). The results for teacher raters were very similar, with differences of less than one *T*-score point on the three protective factor scales (0.47), on the Total Protective Factors scale (0.0), and on the Behavioral Concerns scale (0.1). These results demonstrate that the DECA-P2 ratings are very stable across a 6- to 8-day interval for both parent and teacher raters.

Tables 3.8a and 3.8b present the interrater reliability study mean scale scores and standard deviations for parent and teacher raters, respectively. The absolute value of the difference between the mean scale scores provided by pairs of raters was calculated based on the figures in Tables 3.8a and 3.8b. Pairs of parents differed, on average, by less than one *T*-score point across the three protective factor scales (0.43), on the Total Protective Factors scale (0.3), and on the Behavioral Concerns scale (0.7). Pairs of teachers differed by an

Table 3.8a

Interrater *T*-score Stability: Mean Scale Scores and Standard Deviations–Parent Raters

	Pretest		Posttest	
Scales	**Mean**	***SD***	**Mean**	***SD***
Total Protective Factors	48.7	7.3	49.0	7.9
Initiative	50.1	8.5	50.5	9.2
Self-Regulation	49.1	7.1	48.8	7.4
Attachment/Relationships	48.0	8.7	48.6	7.9
Behavioral Concerns	48.9	8.7	49.6	10.2

Table 3.8b

Interrater *T*-Score Stability: Mean Scale Scores and Standard Deviations–Teacher Raters

	Pretest		Posttest	
Scales	**Mean**	***SD***	**Mean**	***SD***
Total Protective Factors	50.8	9.7	52.3	10.6
Initiative	50.5	9.1	50.3	9.9
Self-Regulation	51.0	9.7	52.8	10.8
Attachment/Relationships	50.6	10.0	52.8	10.2
Behavioral Concerns	47.5	9.7	47.8	8.2

average of 1.4 *T*-score points on the three protective factor scales, by 1.5 *T*-score points on the Total Protective Factors scale, and by about one-third of a *T*-score point (0.3) on the Behavioral Concerns scale. The largest of these differences, 1.5 *T*-score points on the Total Protective Factor scale, is still less than the corresponding standard error of measurement (*SEM*) of 2.23, again indicating that the DECA-P2 ratings are very stable across raters.

Summary

The results of the several reliability studies of the DECA-P2 indicate that the instrument is reliable for assessing children's social and emotional competencies. The results of the internal consistency data demonstrate that the DECA-P2 meets standards suggested by Bracken (1987). The test–retest study shows that raters rank the children's scores on the DECA-P2 similarly over time. This finding is important because it shows that the scores do not tend to change in the absence of a specific program or intervention. This makes more credible the assertion that, when changes are observed over time, these changes can reasonably be attributed to a program, intervention, or some other important event in the child's life. The results of the interrater reliability study show that different parents and teachers also rank children's scores similarly. This is important because it underscores the potential significance of differences in ratings of the same child by different adults when they do occur. The stability studies further indicate that not only the rankings but also the actual mean scale scores obtained at different points in time or from different raters are quite similar. These findings support the interpretive uses of the DECA-P2 described in Chapter 5.

Validity

The validity of a test "concerns what the test measures and how well it does so" (Anastasi, 1988, p. 139). More specifically, validity "is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed purpose" (APA, 1999, p. 11). According to the *Standards for Educational and Psychological Testing* (APA, 1999), the sources of validity evidence can be conceptualized in various ways. We investigated the validity of the DECA-P2 in regard to *test content* (content validity), *internal structure* (presented in Chapter 2 on the development of the DECA-P2), *relations to other variables* (construct validity), and *test–criterion relationships* (criterion validity).

Content-Related Validity

This type of validity assesses the degree to which the domain measured by the test is represented by the test items. With respect to the DECA-P2, content-related validity addresses how well the 27 protective factor items represent the domain of behavioral characteristics related to social and emotional competence and resilience, and how well the 11 items on the Behavioral Concerns scale represent the domain of emotional and behavioral problems of preschool children.

As detailed in Chapter 2, we based the items on the DECA-P2 on a review of the literature on social and emotional competence and resilience in young children. We also conducted a number of focus groups with early care and education professionals to obtain feedback on the original DECA items as well as suggestions for new items to include. Preliminary drafts of the DECA-P2 were also carefully reviewed by our National Advisory Committee. This process resulted in systematic consideration of information regarding social and emotional competence and resilience from researchers and academic leaders, practitioners in the field, and highly respected national leaders in the early care and education field.

Criterion-Related Validity

This type of validity measures the degree to which the scores on the assessment predict either an individual's performance on an outcome or criterion

measure or the status or group membership of the individual. As a measure of behaviors related to social and emotional competence and resilience, scores on the DECA-P2 should predict the social and emotional functioning of young children. To test this hypothesis, we obtained DECA-P2 ratings on two samples of children. First, we obtained a sample of children who were reported by their parent or teacher to have been diagnosed or identified by a professional as having an "emotional or behavioral disturbance (including AD/HD)."[1] The children in this EBD sample (n = 125) were matched to a comparison group (referred to as the Non–Identified Sample (NI) sample (n = 126)) selected from the national standardization sample. Matching variables included: type of rater and the child's age, gender, race, Hispanic ethnicity, and region of residence. Table 3.9 provides descriptive information about both samples and shows that the two groups had very similar demographic characteristics.

We compared the EBD and NI groups using analysis of variance (ANOVA) procedures to contrast the three social–emotional competence scale scores. An independent samples t-test was used to compare both the Total Protective Factors scale and Behavioral Concerns scale scores for the two groups. The results of these analyses are provided in Table 3.10, which documents that there were large and significant differences between the mean scores of the EBD and NI samples on all DECA-P2 scales. The mean standard score differences and other results reported in Table 3.10 clearly show that the ratings of the two groups differed significantly, despite the similarity in demographic characteristics. All scale comparisons were significant ($p < .01$).

In addition to being statistically significant, the means of the two groups on each scale differed by approximately 60% of a standard deviation or more (d-ratios range from .58 to 1.09). The d-ratio is a measure of the size of the difference between the mean scores of two groups, expressed in standard deviation units. According to commonly accepted guidelines for interpreting d-ratios (Cohen, 1988), d-ratios of .2, .5, and .8 are interpreted as small, medium, and large, respectively. Therefore, the effect sizes reported in Table 3.10

1 We chose to use the term *emotional and behavioral disturbance* because it is widely used in mental health and special education settings and it does not label the child as would *emotionally and behaviorally disturbed*.

Table 3.9

Sample Characteristics for the DECA-P2 Criterion Validity Study

	EBD Sample		NI Sample	
	n	%	*n*	%
Size of Sample	125		126	
Rater				
Parent	35	28	35	28
Teacher	90	72	91	72
Age				
3-year-olds	34	27	34	27
4-year-olds	49	39	51	40
5-year-olds	42	34	41	33
Gender				
Boys	92	74	90	71
Girls	33	26	36	29
Race				
American Indian/Alaskan Native	4	3	2	2
Asian	0	0	0	0
Black/African American	23	19	23	19
Native Hawaiian/Pacific Islander	0	0	0	0
White	81	67	86	69
Two or More Races	11	9	12	10
Don't know	2	2	1	1
Hispanic Ethnicity	20	18	14	12
Region of Residence				
Northeast	20	16	30	24
South	42	34	39	31
Midwest	19	15	24	19
West	42	34	33	26

for the Initiative and Attachment/Relationships scales would be characterized as medium and the effect sizes for the other three scales as large. These results provide strong evidence of the validity of the DECA-P2 scales in discriminating between groups of children with emotional or behavioral disturbance and their nonidentified peers.

Table 3.10

Mean *T*-Scores, Standard Deviations, and Difference Statistics for the DECA-P2 Criterion Validity Study

	EBD Sample (*n* = 125)		NI Sample (*n* = 126)
Initiative			
Mean	42.1		47.4
SD	9.1		9.2
F Value		21.25*	
d-ratio		0.58	
Self-Regulation			
Mean	38.2		47.0
SD	8.2		9.6
F Value		60.66*	
d-ratio		0.99	
Attachment/ Relationships			
Mean	42.5		49.2
SD	10.4		9.1
F Value		29.94*	
d-ratio		0.69	
Total Protective Factors			
Mean	40.0		47.6
SD	8.9		9.6
t Value		6.57*	
d-ratio		0.82	
Behavioral Concerns			
Mean	62.7		53.2
SD	8.0		9.4
t Value		8.57*	
d-ratio		1.09	

*$p < .01$

Examination of Race and Ethnic Group Differences

The contrasted group approach can also be used to examine group differences on a variable thought to be *irrelevant* to the construct being assessed. Messick (1995) calls this construct irrelevant variance. To evaluate the appropriateness of the DECA-P2 for use with minority children, we compared the mean scores of the Black and White children and of the Hispanic and Non–Hispanic

children included in the standardization sample. The goal was to determine if these groups of children received similar ratings on the DECA-P2.

To assess the differences in the DECA-P2 ratings, we compared the means using the *d*-ratio statistic. Tables 3.11a and 3.11b present the results of these analyses. The results in these tables indicate that the DECA-P2 scores earned by Black, White, and Hispanic children were quite similar. The differences between Black and White children (Table 3.11a) for parent raters were negligible on three of the five comparisons, and the remaining two were small according to Cohen's interpretive guidelines. The highest *d*-ratio for parent raters was .38. When rated by teachers, the differences were negligible on the Initiative scale and small for the remaining four comparisons. The largest *d*-ratio was .34. For each scale, the mean score differences between Black and White children were smaller than the standard errors of measurement for the same scale and rater (see Table 3.2).

Table 3.11a

DECA-P2 Scale Scores: *d*-Ratios Comparing Black/African American and White Children

	Black/African American (*n* = 169)		Black vs White	White (*n* = 915)	
Parent Raters	**Mean**	***SD***	***d*-ratio**	**Mean**	***SD***
Initiative	49.47	10.25	−0.09	50.34	9.76
Self-Regulation	50.06	9.73	0.01	50.00	9.88
Attachment/ Relationships	47.23	10.64	−0.38	50.91	9.50
Total Protective Factors	48.74	10.33	−0.17	50.47	9.90
Behavioral Concerns	48.20	10.31	−0.22	50.41	9.89
Teacher Raters	**(*n* = 220)**			**(*n* = 1,351)**	
Initiative	48.74	10.64	−0.16	50.33	9.67
Self-Regulation	47.05	10.10	−0.34	50.42	9.83
Attachment/ Relationships	48.39	10.71	−0.24	50.74	9.67
Total Protective Factors	47.84	10.56	−0.28	50.65	9.75
Behavioral Concerns	52.82	10.57	0.33	49.50	9.94

Table 3.11b presents the findings for Hispanic versus Non–Hispanic children. For parent raters, four of the five comparisons yielded negligible differences in mean scale scores. Only one scale, Attachment/Relationships, yielded a *d*-ratio in the small range. The same pattern of results was found for teachers.

When all raters are considered together, the median effect size for Black compared to White children was .23. When Hispanic and Non–Hispanic children are compared, the median effect size was .095. These results indicate that these groups of children receive very similar mean scale scores on the DECA-P2.

Table 3.11b

DECA-P2 Scale Scores: *d*-Ratios Comparing Hispanic and Non-Hispanic Children

	Hispanic (*n* = 379)		Hispanic vs Non-Hispanic	Non-Hispanic (*n* = 888)	
Parent Raters	**Mean**	***SD***	***d*-ratio**	**Mean**	***SD***
Initiative	49.58	9.82	−0.09	50.42	9.79
Self-Regulation	49.20	9.64	−0.11	50.25	9.88
Attachment/ Relationships	48.17	10.22	−0.27	50.81	9.63
Total Protective Factors	48.75	9.68	−0.18	50.57	10.06
Behavioral Concerns	49.60	10.16	−0.05	50.11	10.05
Teacher Raters	**(*n* = 473)**			**(*n* = 1,353)**	
Initiative	49.37	9.29	−0.09	50.22	10.04
Self-Regulation	50.17	9.68	0.03	49.86	10.11
Attachment/ Relationships	48.53	9.24	−0.22	50.65	10.04
Total Protective Factors	49.34	9.54	−0.10	50.34	10.09
Behavioral Concerns	49.38	9.98	−0.07	50.11	10.14

Individual Prediction

The criterion validity of an assessment can also be determined by examining the ability of a scale's scores to accurately predict group membership. The extent to which the Total Protective Factors and Behavioral Concerns scale scores accurately predicted membership in either the EBD or the NI samples was, therefore, examined.

For the Total Protective Factors scale, we predicted that individuals with a *T*-score of less than or equal to 40 would be members of the EBD sample and those with scores above 40 would be members of the NI sample. (As explained in Chapter 5, *T*-scores of 40 and below on DECA-P2 protective factor scales indicate areas of need.) We then compared these predictions with actual group membership. Table 3.12a presents the results of this study.

As shown in Table 3.12a, low Total Protective Factors scale scores correctly predicted group membership for 58% of the EBD sample. Similarly, Total Protective Factors scale *T*-scores that fell in the Typical and Strength ranges (i.e., > 40) correctly predicted 75% of the NI sample. Overall, the Total Protective Factors scale *T*-scores correctly predicted group membership for 67% of the 249 children in this study. Significant chi-square analysis results ($\chi^2 = 28.23$, $df = 1$, $p < .001$, *phi* coefficient = .34) indicate that the Total Protective Factors scale *T*-scores were significantly related to group membership.

Table 3.12a

Actual and Predicted Group Membership Using Total Protective Factors for the DECA-P2 Criterion Validity Study

	EBD Sample		NI Sample	
	n	%	*n*	%
Actual Group Membership	123		126	
Predicted Group Membership Total Protective Factors				
TPF ≤ 40	71	58	31	25
TPF > 40	52	42	95	75

Individual prediction results based on the Behavioral Concerns scale were similar. For this scale, we predicted that individuals with a *T*-score of greater than or equal to 60 would be members of the EBD sample and those with scores below 60 would be members of the NI sample. (As explained in Chapter 5, *T*-scores of 60 and above on the Behavioral Concerns scale indicate areas of need.) We then compared these predictions with actual group membership. Table 3.12b presents the results of this study.

As shown in Table 3.12b, high Behavioral Concerns scale *T*-scores correctly predicted group membership for 70% of the EBD sample. Similarly, Behavioral Concerns scale *T*-scores that fell in the Typical range (i.e., < 60)

Table 3.12b

Actual and Predicted Group Membership Using Behavioral Concerns for the DECA-P2 Criterion Validity Study

	EBD Sample		NI Sample	
	n	%	*n*	%
Actual Group Membership	125		126	
Predicted Group Membership Behavioral Concerns				
BC ≥ 60	87	70	40	32
BC < 60	38	30	86	68

correctly predicted 68% of the NI sample. Overall, the Behavioral Concerns scale *T*-scores correctly predicted group membership for 69% of the 251 children in this study. Significant chi-square analysis results (χ^2 = 35.97, *df* = 1, $p < .001$, *phi* coefficient = .38) indicate that the Behavioral Concerns scale *T*-scores were significantly related to group membership.

It should be noted that the classification accuracy of any assessment is determined both by the psychometric properties of the assessment and the decision rules (i.e., cut scores) used to make these decisions. A less stringent decision rule will result in more children being identified as having significant social and emotional concerns. A more stringent decision rule will result in fewer children being identified. In the case of the DECA-P2, we have chosen a relatively stringent decision rule to minimize the chances of children being overidentified as having social and emotional concerns.

Construct-Related Validity

This type of validity examines the degree to which the assessment instrument measures the theoretical construct of interest. One common approach to establishing the construct validity of an assessment is to demonstrate that scores on the assessment in question correlate positively with scores of similar constructs on other well-developed measures. This is referred to as convergent validity. To examine convergent validity, we correlated *T*-scores on the DECA-P2 with standard scores from the Preschool Behavioral and Emotional Rating Scale (PreBERS; Epstein & Synhorst, 2009) and the Conners Early Childhood Scale (Conners EC; Conners, 2009). Parents ($n = 45$) and teachers ($n = 56$) completed the DECA-P2, the PreBERS, and the Conners EC in a

counterbalanced order in one session. The demographic characteristics of the children involved in this study are presented in Table 3.13 and indicate that this was a varied sample.

The results of this study, which are presented in Table 3.14, indicate that the DECA-P2 has strong convergent validity with the total scale scores for both the PreBERS and the Conners EC. The PreBERS Strength Factor is a measure of social and emotional strengths of young children, and the Conners Global Index assesses social, emotional, and behavioral concerns. The DECA-P2 Total Protective Factors scale correlated significantly with the PreBERS Strength Index for both parent (corrected $r = .65$, $p < .01$) and teacher (corrected $r = .78$, $p < .01$) raters. As would be expected, the Total Protective Factors scale correlated negatively with Conners Global Index for both parent (corrected $r = -.37$, $p < .01$) and teacher (corrected $r = -.42$, $p < .01$) raters.

Table 3.13

Demographic Characteristics of the DECA-P2 Convergent Validity Study Sample

	Parent Sample (n = 45)		Teacher Sample (n = 57)	
	n	**%**	***n***	**%**
Age (years)	Mean = 4.02	*SD* = .783	Mean = 3.95	*SD* = .766
Mean	4 years, 5 months		4 years, 3 months	
SD	8.7 months		9.1 months	
Male	19	42.2	29	50.9
Female	26	57.8	28	49.1
Race/ethnicity				
White	18	40.0	28	49.1
Black	16	35.6	16	28.1
American Indian/ Alaskan Native	0	0.0	0	0.0
Asian	1	2.2	2	3.5
Native Hawaiian/ Pacific Islander	0	0.0	0	0.0
Two or More	4	8.9	5	8.8
Region				
Northeast	14	31.1	18	31.6
South	4	8.9	0	0.0
Midwest	18	40.0	32	56.1
West	9	20.0	6	10.5

Table 3.14

Results of the DECA-P2 Convergent Validity Study: Correlation of the DECA-P2 Total Protective Factors With PreBERS and Conners EC Summary Scales

Parents	*N*	Mean	*SD*	PreBERS Correlations Obtained *r*	PreBERS Correlations Corrected *r*
DECA-P2 Total Protective Factors	45	51.0	9.7	.55**	.63**
DECA-P2 Behavioral Concerns	45	48.9	10.8	−.29	−.32*
PreBERS Strength Factor	45	104.9	12.4	–	–
Teachers					
DECA-P2 Total Protective Factors	57	50.4	10.6	.76**	.78**
DECA-P2 Behavioral Concerns	57	52.0	10.9	−.68**	−.70**
PreBERS Strength Factor	57	99.0	13.1	–	–

Note: Correlations between the DECA-P2 and PreBERS were corrected for restriction (or expansion) in range.

DECA-P2 T-scores are set at mean of 50 and SD of 10 and PreBERS scores are set at mean of 100 and SD of 15.

p < .05
****p < .01***

Parents	*N*	Mean	*SD*	Conners Correlations Obtained *r*	Conners Correlations Corrected *r*
DECA-P2 Total Protective Factors	45	51.0	9.7	−.42**	−.34*
DECA-P2 Behavioral Concerns	45	48.9	10.9	.73**	.60**
Conners Global Index	45	56.6	13.1	–	–
Teachers					
DECA-P2 Total Protective Factors	57	50.4	10.6	−.58**	−.41**
DECA-P2 Behavioral Concerns	57	52.0	10.9	.81**	.64**
Conners Global Index	57	57.1	14.9	–	–

Note: Correlations between the DECA-P2 and Conners were corrected for restriction (or expansion) in range.

DECA-P2 and Conners T-scores are set at mean of 50 and SD of 10.

p < .05
****p < .01***

The DECA-P2 Behavioral Concerns scale showed negative correlations with the PreBERS Strength Index for both parent (corrected $r = -.36$, $p = .052$) and teacher (corrected $r = -.54$, $p < .01$) raters. Finally, the Behavioral

Concerns scale correlated positively with the Conners EC Global Index for both parents (corrected $r = .59$, $p < .01$) and teacher (corrected $r = .65$, $p < .01$). These data provide evidence that the DECA-P2 measures both positive behaviors and behavioral concerns in a way consistent with other well-developed measures. However, the moderate correlations also suggest that there are unique aspects of the DECA-P2 that provide somewhat different information than these other two scales.

Validity Study Summary

The content-related evidence provided in this chapter related the DECA-P2 items to both the research and practice on social and emotional competence in preschool children. The results of the criterion-related validity studies demonstrated that DECA-P2 scores do differentiate between groups of children with and without the special-needs designation of emotionally or behaviorally disturbed. The construct-related evidence established that the DECA-P2 scales do show strong convergent validity with similar, albeit clinically oriented, measures.

Comparability of the 1999 DECA and the 2012 DECA-P2 Protective Factor Scales

The DECA-P2 protective factor scales differ from their 1999 counterparts in a number of important ways. First, about 25% of the items on the DECA-P2 are different from the original 1999 DECA. Second, the scale names have been modified to reflect these changes in content. For instance, the original Attachment scale has been broadened and reconceptualized to include Relationships. Third, new norms with a contemporary standardization sample have been developed. These changes suggest that the comparability of scale *T*-scores obtained with the original DECA versus those obtained with the new items, scales, and norms should be examined. In other words, would a child who was rated on the DECA and received a Total Protective Factors scale *T*-score of, for example, 50 expect to receive a similar score on the DECA-P2?

To address this important question, we utilized the DECA-P2 standardization sample and standardization form items. We extracted from the DECA-P2 standardization form the original DECA items and scales and scored the new standardization sample on those original scales using the 1999 norms tables. Only three items from the original DECA were not included in the DECA-P2 standardization form: one item from the Attachment scale and two items from the Behavioral Concerns scale. Utilizing the missing-item scoring rules (see Chapter 4 of this manual), we substituted the child's mean score on the other eight items for the missing item on the Attachment scale. Because two items were missing from the DECA Behavioral Concerns scale, this scale was not used and is not reported below. We then compared the results obtained with the two different forms of the Devereux Early Childhood Assessment. The results are presented in Table 3.15.

Table 3.15

Comparability of DECA and DECA-P2 *T*-Scores Derived From the Same Ratings on the Same Children

Scales	Mean DECA	*SD* DECA	Mean DECA-P2	*SD* DECA-P2	*r*	*t*	*d*-ratio
	Parents (*n* = 1,244)						
Total Protective Factors	51.8	10.7	50.1	10.0	0.97*	23.3**	0.16
Initiative	52.4	10.2	50.1	9.8	0.94*	23.2**	0.23
Self-Regulation	53.4	10.5	50.0	9.8	0.99*	73.5**	0.33
Attachment/ Relationships	50.9	11.6	50.1	9.8	0.89*	5.4**	0.07
	Teachers (*n* = 1,923)						
Total Protective Factors	51.5	9.8	49.8	9.7	0.97*	30.3**	0.17
Initiative	52.3	9.8	49.5	9.6	0.88*	26.9**	0.29
Self-Regulation	53.4	10.4	49.9	10.0	0.99*	134.7**	0.34
Attachment/ Relationships	50.0	9.2	50.0	9.9	0.95*	0.3	−0.06

p < .01
****p < .01 using Bonferroni correction***

The *T*-scores on the original DECA and the DECA-P2 correlated very highly. All of the correlation coefficients reported in Table 3.15 are significant and high in magnitude. For parent raters, the correlation coefficients ranged from

.89 (Attachment/Relationships) to .99 (Self-Regulation). For teacher raters, the correlation coefficients ranged from .88 (Initiative) to .99 (Self-Regulation).

Although the correlations were all very high, it could still be the case that the actual *T*-scores received by the children on the comparable scales differed. With the exception of Attachment/Relationships for teacher raters, the mean scale *T*-scores based on the DECA-P2 were lower than the mean *T*-scores on the DECA. On Total Protective Factors, the DECA-P2 mean scale *T*-scores were 1.7 points lower than the corresponding DECA scores for both parent and teacher raters. Also, for both parent and teacher raters the average mean scale *T*-score difference across the three protective factor scales was approximately 2.1. Paired-samples *t*-tests were used to evaluate the statistical significance of these mean score differences. Bonferroni correction procedures were used to account for inflated Type I error rates due to the multiple comparisons. The experiment-wise error rate was set at $p < .01$ with the pair-wise rate set at $p < .0025$. Using these criteria, all of the mean score comparisons were significant, with the exception of Attachment/Relationships as rated by teachers. Finally, the magnitude of these mean score differences (effect size) was evaluated using the *d*-ratio. As shown in Table 3.15 for both parent and teacher raters, the effect sizes for the Total Protective Factors and Attachment/Relationship scales were negligible. For both rater types, the effect sizes on the remaining two scales, Initiative and Self-Regulation, were small, ranging from .23 to .34.

These results indicate that scores on the 1999 DECA and the 2012 DECA-P2 are similar but different. On each scale except Attachment/Relationships for teacher raters the DECA-P2 mean scale *T*-scores were significantly lower. However, the magnitude of these differences is small. These results imply that, whenever possible, comparisons between Devereux Early Childhood Assessments should use the same form, either the 1999 DECA or the 2012 DECA-P2. In those cases where the two different forms are being compared (e.g. pretest-posttest comparisons—see Chapter 5), one can expect that, in general, the score on the DECA-P2 will be one to three points lower than the score that would have been received had the DECA been used for both ratings. DECA-P2 users should keep this in mind if they must compare results across forms. Additional guidance on comparing DECA and DECA-P2 results can be found at www.CenterForResilientChildren.org.