

Reliability

The reliability of assessments like the DECA-I/T 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). We assessed the reliability of both the DECA-I and the DECA-T using several methods. First, we computed the internal reliability coefficients for each scale. Second, we assessed standard error of measurement (SE_M). Third, we assessed the test-retest reliability of each scale. Finally, we determined the interrater reliability for each scale.

Internal Reliability

Internal reliability (also known as internal consistency) refers to the extent to which the items on the same scale or assessment instrument measure the same underlying construct. High internal reliability, which is desirable, indicates that the items assess the same characteristic of the child (i.e., construct) and, therefore, truly comprise a single scale. In contrast, low internal reliability indicates that the items measure a variety of different child characteristics and, therefore, do not comprise a single scale.

We determined the internal reliability of each scale and for each form using Cronbach’s alpha (Cronbach, 1951). In practice, this statistic can vary from .00 (low) to .99 (high). The internal reliability coefficients (alphas) were based on the DECA-I/T standardization sample and estimates for each were calculated separately for each Rater (parent/family member or early care and education professional) and are presented in Table 2.1a (Infants) and Table 2.1b (Toddlers).

The results in these tables indicate that both the DECA-I and the DECA-T have high internal reliability. For the infant form (DECA-I) the Total Protective Factors Scale alpha for both Parent Raters (.90 to .94) and Teacher Raters (.93 to .94) met or exceeded the .90 minimum for a total score suggested by Bracken (1987) in each age group. In addition, these values met the “desirable standard” described by Nunnally (1978, p.246). The same was true for the toddler form (DECA-T) with Parent Raters at .94 and Teacher Raters at .95.

Table 2.1a

Internal Reliability (Alpha) Estimates for DECA-I Scales by Rater

Scale	Raters	
	Parents	Teachers
1-3 Months		
Initiative	.87	.87
Attachment/Relationships	.80	.93
Total Protective Factors	.90	.93
3-6 Months		
Initiative	.86	.91
Attachment/Relationships	.87	.91
Total Protective Factors	.90	.94
6-9 Months		
Initiative	.90	.89
Attachment/Relationships	.89	.89
Total Protective Factors	.94	.93
9-18 Months		
Initiative	.87	.90
Attachment/Relationships	.92	.91
Total Protective Factors	.93	.94

The internal reliability coefficients for the DECA-I scales (Initiative and Attachment/Relationships) were also high. These ranged from a low of .80 (1 to 3 Months Attachment/Relationships Parent Rater) to a high of .93 (1 to 3 Months Attachment/Relationships Teacher Rater). The median reliability coefficient across both scales was .87 for Parent Raters and .90 for Teacher Raters. These median values met or exceeded the .80 minimum for scale scores suggested by Bracken (1987).

Table 2.1b

Internal Reliability (Alpha) Estimates for DECA-T Scales by Rater

Scale	Raters	
	Parents	Teachers
Attachment/Relationships	.87	.90
Initiative	.92	.94
Self-Regulation	.79	.83
Total Protective Factors	.94	.95

The internal reliability coefficients for the DECA-T remaining scales (Attachment/Relationships, Initiative and Self-Regulation) were high as well. These ranged from a low of .79 (Self-Regulation Parent Rater) to a high of .94 (Initiative Teacher Rater). The median reliability coefficient across these three scales was .87 for Parent Raters and .90 for Teacher Raters. These median values also met or exceeded the .80 minimum for scale scores suggested by Bracken (1987).

Standard Errors of Measurement

The standard error of measurement (SE_M) is another index of the reliability of test scores. It is an estimate of the amount of error in the observed score, expressed in standard score units (i.e., T scores). We obtained the SE_M for each of the DECA-I/T Scale T scores directly from the internal reliability coefficient (r) using the formula,

$$SE_M = \sigma \sqrt{1-r}$$

where σ is the theoretical standard deviation of the T score (10) and the appropriate reliability coefficient (r) is used (Atkinson, 1991). The SE_M for each DECA-I and DECA-T scale according to Rater are presented in Table 2.2a and Table 2.2b. The SE_M s varied with the size of the internal reliability coefficient reported in Tables 2.1a and 2.1b—the higher the reliability, the smaller the standard error of measurement.

Table 2.2a

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

Scale	Raters	
	Parents	Teachers
1-3 Months		
Initiative	3.61	3.61
Attachment/Relationships	4.47	2.65
Total Protective Factors	3.16	2.65
3-6 Months		
Initiative	3.74	3.00
Attachment/Relationships	3.61	3.00
Total Protective Factors	3.16	2.45
6-9 Months		
Initiative	3.16	3.32
Attachment/Relationships	3.32	3.32
Total Protective Factors	2.45	2.65
9-18 Months		
Initiative	3.61	3.16
Attachment/Relationships	2.83	3.00
Total Protective Factors	2.65	2.45

Test-Retest Reliability

The correlation between scores obtained for the same child by the same Rater 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).

Table 2.2b

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

Scale	Raters	
	Parents	Teachers
Attachment/Relationships	3.61	3.16
Initiative	2.83	2.45
Self-Regulation	4.58	4.12
Total Protective Factors	2.45	2.24

Table 2.3a

Characteristics of DECA-I Test-Retest Reliability Sample

Characteristic	Raters	
	Parents	Teachers
Size of Sample (n)	20	23
Age (Months)		
Mean	7.9	9.3
SD	3.8	3.8
Gender		
Boys	40%	44%
Girls	60%	56%
Race		
Native American	5%	5%
Asian/Pacific Islander	5%	5%
African American		5%
Hispanic	5%	5%
Caucasian	80%	75%
Mixed Race	5%	5%
Other		

Table 2.3b

Characteristics of DECA-T Test-Retest Reliability Sample

Characteristic	Raters	
	Parents	Teachers
Size of Sample (n)	22	20
Age (Months)		
Mean	26.1	26.2
SD	4.8	4.6
Gender		
Boys	50%	50%
Girls	50%	50%
Race		
Native American		
Asian/Pacific Islander		
African American	5%	5%
Hispanic		
Caucasian	90%	95%
Mixed Race		
Other	5%	

To investigate the test-retest reliability of the DECA-I/T, a group of parents ($n=20$ for DECA-I and $n=22$ for DECA-T) and a group of teachers ($n=23$ for DECA-I and $n=20$ for DECA-T) rated the same child on two different occasions separated by a minimum of 24 hours and a maximum of 72 hours. Descriptive information on the children rated in this study is provided in Table 2.3a and Table 2.3b.

Table 2.4a presents the results of Test-Retest Reliability Study for the DECA-I. All of the correlation coefficients were statistically significant ($p < .001$), which indicates the scales have very good test-retest reliability. Overall, parents were more consistent in their evaluation of the children's behavior across time. For parents, the higher correlation was found on the Initiative Scale (.94), and the lower on the Attachment/Relationships Scale (.86). The higher correlation for early care and education professionals was found on the Attachment/Relationships and Total Protective Factor Scales (.84) and the lowest on the Initiative Scale (.83).

Table 2.4a

Test-Retest Reliability Coefficients for DECA-I Scores Obtained at a 24- to 72-Hour Interval

Scale	Parents	Teachers	Overall
Initiative	.94***	.83***	.87***
Attachment/Relationships	.86***	.84***	.83***
Total Protective Factors	.91***	.84***	.85***

*** ($p \leq .001$)

Table 2.4b

Test-Retest Reliability Coefficients for DECA-T Scores Obtained at a 24- to 72-Hour Interval

Scale	Parents	Teachers	Overall
Attachment/Relationships	.97***	.96***	.97***
Initiative	.99***	.98***	.98***
Self-Regulation	.92***	.72***	.85***
Total Protective Factors	.99***	.91***	.97***

*** ($p \leq .001$)

Table 2.4b presents the results of Test-Retest Reliability Study for the DECA-T. All of the correlation coefficients were also statistically significant ($p < .001$), which indicates the scales have very good test-retest reliability. As was the case with infants, parents were somewhat more consistent in their assessment of toddlers than teachers, although both raters are highly reliable. For parents, the highest correlation was found on the Initiative Scale (.99), and the lowest on the Self-Regulation Scale (.92). The highest correlation for early care and education professionals was found on the Initiative Scale (.98) and the lowest on the Self-Regulation Scale (.72).

Interrater Reliability

The correlation between scores obtained for the same child at the same time by two different Raters is another indicator of the reliability of an assessment instrument. The magnitude of the obtained value informs us about the degree of similarity in the different Raters' perceptions of the child's behavior.

A set of ratings included two independent ratings of the same child completed on the same day. The ratings were provided by either two early care and education professionals or two parents (or other family members).

Table 2.5a

Characteristics of DECA-I Interrater Reliability Sample

Characteristic	Raters	
	Parents	Teachers
Size of Sample (n)	23	20
Age (Months)		
Mean	10.4	11.0
SD	4.1	3.3
Gender		
Boys	43.5%	55.0%
Girls	56.5%	45.0%
Race		
Native American		
Asian/Pacific Islander		
African American	8.7%	45.0%
Hispanic	4.0%	
Caucasian	83.3%	51.0%
Mixed Race	4.0%	4.0%
Other		

Table 2.5b

Characteristics of DECA-T Interrater Reliability Sample

Characteristic	Raters	
	Parents	Teachers
Size of Sample (n)	23	30
Age (Months)		
Mean	26.3	27.5
SD	4.5	5.5
Gender		
Boys	62.5%	45.6%
Girls	37.5%	54.4%
Race		
Native American		
Asian/Pacific Islander		10.0%
African American	4.0%	40.0%
Hispanic		3.0%
Caucasian	92.0%	44.0%
Mixed Race	4.0%	3.0%
Other		

Two different comparisons were made: 1) Teacher Rater-Teacher Rater and 2) Parent Rater-Parent Rater. We collected Teacher Rater-Teacher Rater pairs of ratings on 20 infants and 30 toddlers. We also collected Parent Rater-Parent Rater pairs of ratings on 23 infants and 23 toddlers. Demographic information on the children rated is provided in Table 2.5a and Table 2.5b.

Table 2.6a presents the results of this study for the infants. The interrater reliability coefficients for Parent Rater-Parent Rater who saw the child in the same environment were high and statistically significant ($p < .01$). The coefficients ranged from a high of .58 for parent pairs on Total Protective Factors to a low of .53 for parent pairs on the other scales. This indicates that different parents or family members rate the same child very similarly on the DECA-I when observing the child in the same environment. The Teacher Rater-Teacher Rater coefficients, while mildly high, were not statistically significant. This could very well mean that teachers have not observed the infants as much as parents.

Table 2.6a

Interrater Reliability Coefficients for DECA-I Scores

Scale	Parent-Parent	Teacher-Teacher
Initiative	.53**	.33
Attachment/Relationships	.53**	.29
Total Protective Factors	.58**	.32

** ($p \leq .01$)

Table 2.6b presents the results of this study for the toddlers. Unlike the infants, the interrater reliability coefficients for both pairs (Teacher Rater-Teacher Rater and Parent Rater-Parent Rater) who saw the child in the same environment were high and statistically significant. The coefficients ranged from a high of .64 for Teacher pairs on Self-Regulation to a low of .47 for teacher pairs on Attachment/Relationships. This indicates that different early care and education professionals and different parents or family members rate the same child very similarly on the DECA-T when observing the child in the same environment.

Table 2.6b

Interrater Reliability Coefficients for DECA-T Scores

Scale	Parent-Parent	Teacher-Teacher
Attachment/Relationships	.49*	.47*
Initiative	.61***	.49**
Self-Regulation	.63***	.64***
Total Protective Factors	.58**	.52**

* ($p \leq .05$)

** ($p \leq .01$)

*** ($p \leq .001$)



Summary

The results of the internal consistency, test-retest, and interrater reliability studies indicate that the DECA-I/T is a reliable tool for assessing infant and toddler protective factors. The results of the internal consistency study demonstrated that the DECA-I/T meets the desirable standards that measurement and testing professionals have recommended. The test-retest study showed that Raters give very similar ratings on the same child across relatively short periods. This indicates that the DECA-I/T is not easily impacted by random changes, but tends to provide a consistent assessment of the child within a single setting, and stability of multiple assessments over time. The results of the interrater reliability studies demonstrate that pairs of parents have similar perceptions of both infants and toddlers, and that teachers show strong agreement in their ratings of toddlers. These results should assure parents and early care and educational professionals alike that the DECA-I/T is a reliable assessment package that can be used with confidence.

Validity


The validity of a test “concerns what the test measures and how well it does so” (Anastasi, 1988, p. 139). More specifically, validity studies investigate the evidence that supports the conclusions or inferences that are made based on test results and the interpretive guidelines presented in the test manual. According to the Standards for Educational and Psychological Testing (APA, 1985), validity evidence can be conceptualized as related to content, prediction (criterion), and construct. We investigated the validity of the DECA-I/T in regard to each of these three areas, and convergence in the case of the DECA and DECA-I/T.

Content Validity

Content validity assesses the degree to which the domain measured by the test is represented by the test items. With respect to the DECA-I/T, content validity addresses how well the protective factor items represent the entire domain of within-child behavioral characteristics related to resilience in infants and toddlers.

As detailed in Chapter 1 of this manual, the content of the DECA-I/T was based on a thorough review of the resilience literature related to young children; results of national focus groups conducted with parents, teachers and infant and early childhood mental health professionals; and careful review of other infant and toddler social and emotional instruments. This resulted in a large initial pool of 112 distinct strength-based behaviors. The authors and DECI (Devereux Early Childhood Initiative) staff critically reviewed this set of potential items. Specifically, they were asked if they thought any content pertaining to within-child protective factors for infants and toddlers was missing. The consensus was that there was ample coverage of content with no skills/topics missing.

The 112-item protocol was piloted with a small national sample of 251 children prior to standardization. This protocol was further refined based on feedback from DECI staff and the National Advisory Team as well as pilot study results. The final standardization form consisted of 68 items and was sent out nationally.



The standardization data set was further reduced by ridding the sample of cases that had critical information missing, such as the date of birth of the child. The final data set was 2,183 children. Utilizing this final data set and the analytic techniques described in Chapter 1, a large number of the items were eliminated, resulting in a final 2-factor solution with 33 items for the DECA-I and a 3-factor solution with 36 items for the DECA-T. It is noteworthy that the items and scales on both the DECA-I and the DECA-T have striking similarities to the within-child protective factor scales on the DECA. All three scales include the constructs of Initiative and Attachment/Relationships. In addition, the construct of Self-Regulation on the DECA-T is quite similar to the construct of Self-Control on the DECA. The overlap and similarities also signify an important developmental trajectory that the scales follow from infancy through the preschool age. The similarity of the factor structure and scale content on the DECA and the DECA-I/T, despite the fact that they were developed with two entirely different samples, lends credence to the importance of these constructs in the social and emotional development of children from birth through age five.

Criterion Validity

Criterion validity measures the degree to which the scores on the assessment instrument predict either 1) an individual's performance on an outcome or criterion measure, or 2) the status or group membership of an individual.

Protective factors buffer children against stress and adversity, resulting in better outcomes than would have been possible in their absence. One important outcome for young children is social and emotional health. Consequently, children with high scores on the DECA-I/T Protective Factor Scales should have greater social and emotional health than children with low scores on these scales.

To test this hypothesis, we obtained DECA-I and DECA-T ratings on two samples of infants and toddlers. The "Identified" sample had known emotional and behavioral problems. These children met at least one of the two following criteria: 1) they had been referred to a mental health professional due to social and emotional challenges, or 2) they had been asked to leave a childcare setting due to their behavior.

We also obtained DECA-I and DECA-T ratings for a matched comparison group of typical infants and toddlers, the "Community" sample. Matching variables included age, gender, and race. Table 3.1a and 3.1b provide descriptive information on the samples for the DECA-I and the DECA-T showing that the two groups were demographically similar.

Table 3.1a

Characteristics of the DECA-I Validity Study Sample

Characteristic	Identified Sample		Community Sample	
		%		%
Size of Sample (n)	15		15	
Age (Months)				
Mean	10.6		11.7	
SD	5.2		4.5	
Gender				
Boys	8	53.0%	8	53.0%
Girls	7	47.0%	7	47.0%
Race*				
Native American	1	6.7%	1	6.7%
African American	3	20.0%	4	26.6%
Hispanic	4	26.6%	5	33.3%
Caucasian	6	40.0%	7	46.6%
Missing	3	20.0%		

*Totals do not add up to 100% due to multiple race

Contrasted Groups

The contrasted groups approach to assessing criterion validity examines scale score differences between groups of individuals who differ on some important variable. Multivariate Analysis of Variance (MANOVA) procedures were used to contrast scale scores for the identified and community samples. Preliminary tests of homogeneity of variance and normality were conducted, and no adverse violations of assumptions were found. Subsequently, independent *t*-tests were used to compare the Total Protective Factors scores for the two groups.

Table 3.2a presents the results of this study with the DECA-I and documents that there were significant and meaningful differences between the "Identified" and the "Community" samples on all three scales. The mean standard score differences and other results reported in Table 3.2a indicate that the ratings of the two groups differ significantly despite the similarity in demographic characteristics ($p < .01$).

Table 3.1b

Characteristics of the DECA-T Validity Study Sample

Characteristic	Identified Sample		Community Sample	
		%		%
Size of Sample (n)	69		69	
Age (Months)				
Mean	27.3		27.5	
SD	4.6		5.2	
Gender				
Boys	43	62.3%	38	55.1%
Girls	26	37.7%	31	44.9%
Race*				
Native American	8	6.7%	2	2.9%
African American	9	20.0%	7	10.1%
Hispanic	15	26.6%	19	27.5%
Caucasian	45	40.0%	43	62.3%
Asian/Pacific Islander	1	20.0%	1	1.4%
Other	3	4.3%	2	2.9%

*Totals do not add up to 100% due to multiple race

Similarly, Table 3.2b presents the results of this study with the DECA-T and documents that there were significant and meaningful differences between the "Identified" and the "Community" samples on all four scales. The mean standard score differences and other results reported in Table 3.2b strongly indicate that the ratings of the two groups differed significantly despite the similarity in demographic characteristics ($p < .01$).

Besides being statistically significant, the means of the two groups on each instrument and on each scale differed by approximately one standard deviation (d -ratios range from .75 to 1.52). The d -ratio is a measure of the size of the difference between the mean scores expressed in standard deviation units. Widely accepted guidelines for interpreting d -ratios (Cohen, 1988) in comparing two groups indicate that the magnitudes of .2, .5, and .8 are interpreted as small, medium, and large, respectively. Therefore, the effect sizes in Tables 3.2a and 3.2b would all be characterized as large except Initiative on the toddler scale, which would be characterized as a medium effect size. These findings provide evidence of the validity of the DECA-I/T scales in discriminating between groups of infants and toddlers with and without social and emotional concerns.

Table 3.2a

Mean T Scores and Difference Statistics for DECA-I Validity Study

	Identified Sample (<i>n</i> =15)	Community Sample (<i>n</i> =15)
Initiative		
Mean	45.3	54.4
SD	8.6	11.8
F Value		6.20***
d-Ratio		.89
Attachment/Relationships		
Mean	41.6	54.4
SD	9.6	7.2
F Value		18.89***
d-Ratio		1.52
Total Protective Factors		
Mean	42.9	53.2
SD	9.3	9.6
t Value*		3.71**
d-Ratio		1.09

* *t*-test for independent means

** $p < .01$

*** $p < .001$

Examination of Potential Adverse Impact on Minority Children

The contrasted group approach can also be used to show that groups that differ on a variable thought to be irrelevant to the purpose of the instrument do *not* differ on scale scores. To evaluate the appropriateness of the DECA-I/T for use with minority children, we compared the mean scores of African American and Caucasian children and Hispanic and Caucasian children in the standardization sample. The goal was to determine if these groups of children received similar ratings on the DECA-I/T. To assess the difference in ratings we compared the means using the *d*-ratio statistic. It should be noted that *d*-ratios following non-significant hypothesis tests should be interpreted as being not statistically significantly different from zero (Sawilowsky & Yoon, 2002).

Table 3.2b

Mean T Scores and Difference Statistics
for DECA-T Validity Study

	Identified Sample (n=69)	Community Sample (n=69)
Attachment/Relationships		
Mean	42.6	50.3
SD	9.4	9.6
F Value		31.01***
d-Ratio		.81
Initiative		
Mean	42.7	50.4
SD	10.7	9.7
F Value		37.42***
d-Ratio		.75
Self-Regulation		
Mean	41.0	50.5
SD	9.7	9.7
F Value		35.25***
d-Ratio		.98
Total Protective Factors		
Mean	40.9	50.5
SD	8.9	9.8
t Value*		7.07**
d-Ratio		1.03

* t-test for independent means

** $p < .01$

*** $p < .001$

Table 3.3a presents the results of these analyses for the DECA-I. As shown in Table 3.3a, 19 of 12 of the mean score differences were negligible. Two of the remaining three mean score differences would be characterized as “small” and one “medium.” The average *d*-ratio when comparing scores earned by African American and Caucasian children was .09. The average *d*-ratio when comparing scores earned by Hispanic and Caucasian children was .26.

Table 3.3a

DECA-I Scale Scores: *d*-Ratios Comparing Minority and Non-Minority Children

	African-American vs. Caucasian	Hispanic vs. Caucasian
Teacher Raters		
Initiative	.08	.11
Attachment/Relationships	.01	.07
Total Protective Factors	.05	.11
Parent Raters		
Initiative	.19	.21
Attachment/Relationships	.15	.67
Total Protective Factors	.08	.41

Table 3.3b presents the results of these analyses for the DECA-T. As shown in Table 3.3b, 9 of 16 of the mean score differences were negligible. Six of the remaining mean score differences would be characterized as “small” and one as “medium.” The average *d*-ratio when comparing scores earned by African American and Caucasian children was .20. The average *d*-ratio when comparing scores earned by Hispanic and Caucasian children was .24.

Table 3.3b

DECA-T Scale Scores: *d*-Ratios Comparing Minority and Non-Minority Children

	African-American vs. Caucasian	Hispanic vs. Caucasian
Teacher Raters		
Attachment/Relationships	.07	.02
Initiative	.33	.14
Self-Regulation	.05	.10
Total Protective Factors	.11	.06
Parent Raters		
Attachment/Relationships	.30	.68
Initiative	.06	.17
Self-Regulation	.42	.29
Total Protective Factors	.27	.47