Measuring Resilience in Children: From Theory to Practice*

14

Jack A. Naglieri, Paul A. LeBuffe, and Katherine M. Ross

Introduction

The concept of resilience, like all psychological constructs, must have certain characteristics in order to be subjected to experimental testing so as to be effectively applied to benefit our constituency. A primary characteristic is that resilience must be operationally defined in a way that is reliable across time, subjects, and researchers. Once a concept is operationalized in a reliable manner, then its validity can be examined. When we have sufficiently operationalized the concept of resilience, and there is evidence that it can be measured in a reliable and valid way, then application in clinical and educational settings becomes possible. This is an ideal sequence for the development tools for testing new concepts, but it is not how many concepts and tests used in education and psychology have been promulgated.

In practice, there is great emphasis on helping clients and pressure to implement new approaches even if they have only been minimally tested. If an idea appears logical and appears to help clients then it seems reasonable to believe that the construct possesses validity, however ill-defined that may be. Unfortunately, what seems logical and consistent with clinical experience may not be true. As noted by Garb (2003, p. 32), "Results from empirical studies reveal that it can be surprisingly difficult for mental health professionals to learn from clinical experience." This sobering point suggests that we should weigh empirical findings more heavily than clinical experience not vice versa. Science should temper enthusiasm. This is especially true when a new approach to treatment or a new concept is introduced.

There is a natural and desirable interplay between scientific research and applied practice in psychology because of the very nature of the field. We can assume that ultimately the field will advance because of the mutual respect and collaboration of those that emphasize science more than practice, and practice more than research. The need for the balanced contribution of science and practice is well illustrated by the study of factors related to resilience. Clearly, this area of study has benefited from the outstanding contributions

*We write this chapter in order to provide essential information about measurement of resilience and the tools that are currently available for that purpose. It is important for the reader to recognize that the first two authors of this chapter are authors of several of the scales included here. In order to provide as complete a view as possible of all the scales currently available for measuring protective factors, we also included scales developed by other authors. We have, therefore, limited any evaluative comments about these scales but do provide a factual presentation of their characteristics. It is our expectation that this information will provide readers sufficient information to arrive at their own conclusions regarding the relative advantages and disadvantages of these tools.

J.A. Naglieri (✉)
University of Virginia,
Charlottesville, VA, USA
e-mail: jnaglieri@gmail.com

P.A. LeBuffe • K.M. Ross
Devereux Center for Resilient Children, Villanova, PA, USA
e-mail: plebuffe@Devereux.org; kross@devereux.org

S. Goldstein and R.B. Brooks (eds.), *Handbook of Resilience in Children*,
DOI 10.1007/978-1-4614-3661-4_14,

made by those professionals whose goal has been to help children and adults survive and thrive in the face of adversity and by those researchers who have studied the complex interrelationships of variables that may be predictive of good outcome. All of these individuals, however, must be able to clearly define their constructs and measure them reliably before the validity of the concept can be assessed. That is the focus of this chapter—the challenge of reliable and valid measurement of factors related to resilience.

Resilience: Measurement Issues

Defining the Concept: What Is Resilience?

Although resilience has been studied and described since the 1950s, it has been only in about the past 2 decades that some consistency has emerged in the definition of this construct. Most contemporary researchers now agree that resilience refers to positive outcomes, adaptation or the attainment of developmental milestones or competencies in the face of significant risk, adversity, or stress. As Masten (2001) points out, the claim of resilience in an individual requires two judgments. First, that the individual has been exposed to significant risk or adversity and, second, that the individual has attained at least typical or normal developmental outcomes.

The paradigm for resilience research therefore consists first of enumerating or measuring the risks and sources of adversity in individuals' lives. Two general approaches have been used to ascertain and measure risk. The *major life events* approach focuses on episodic, highly traumatic events such as the death or divorce of a parent. Typically, major life events are measured using checklists that assess a wide range of traumatic events that have occurred in the individual's lifetime. Examples include the *Sources of Stress Inventory* (Chandler, 1981) or the *Life Events Checklist* (Work, Cowen, Parker, & Wyman, 1990).

Although major life events are clearly important sources of risk and adversity, a reliance on this approach in isolation has been criticized as incomplete. To gain a more complete picture of risk and adversity, a measure of daily hassles is recommended. Daily hassles denotes sources of risk that have lower acuity, but greater chronicity when compared to major life events. Examples for young children might include frequent changes in caregivers, poor quality childcare, and inconsistent or overly harsh discipline. The *Daily Hassles Scale* (Kanner, Coyne, Schaefer, & Lazarus, 1981) is a good example of this approach.

After having ascertained the risk in an individual's life, developmental outcomes can be assessed. This may consist of the attainment of developmental milestones or the accomplishment of major developmental tasks within normal limits. Positive outcome has also been characterized as the absence of psychopathology in an at-risk population. If the individual has attained typical or superior outcomes in the presence of risk or adversity, then resilience is inferred.

Challenges in Measuring Resilience

Measurement of those variables that allow some children to cope successfully with adversities in their lives is not simple. This is especially so because resilience is assessed on an inferential basis by an examination of risk and positive adaptation factors (Luthar & Zelazo, 2003). Resilience is an outcome, rather than a psychological construct in and of itself that can be defined and, perhaps, measured. This has led to efforts to identify variables that lead to, and therefore, can be used to predict, resilience rather than measuring it directly. These factors that lead to resilient outcomes are referred to as protective factors and are defined as characteristics or processes that moderate or buffer the negative effects of stress resulting in more positive behavioral and psychological outcomes than would have been expected in their absence (Masten & Garmezy, 1985). Rather than measuring resilience per se, assessments have instead focused on measuring these protective factors that predict resilience.

Further complicating the situation is the fact that researchers in this field (e.g., Werner, 2005; Wright & Masten, 2005) have found that risk and protective factors occur at multiple levels including

the community (e.g., dangerous neighborhoods/quality after school programs), the family (e.g., domestic violence/effective parenting), and characteristics of the child (e.g., difficult temperament/good coping skills). Although resilience is a function of the complex interaction of these multiple level protective and risk factors, and therefore, most likely is a multivariate construct, most assessments have focused only on the personal characteristics, often referred to as "within-child" protective factors. Moreover, this complex interaction may differ from person to person; that is, the impact of risk factors and the protection afforded by specific protective factors may be very person-specific. As an example, being part of a faith community is widely regarded as an important protective factor, yet the impact of a faith life in moderating risk and adversity differs from person to person. Given this complexity, how can these variables be reliably measured? How can these variables be aggregated to yield a reliable predictor of resilience?

Measurement of the wide variety of variables used to study resilience in children has been accomplished using a variety of experimental methods as well as formal and informal tests, including both standardized and unstandardized methods. The list ranges from published behavior rating and self-concept scales to informal ratings based on clinical criteria; sociometric ratings to social skills rating scales; tests of achievement to yearly grades and IQ test results; parent interviews to parenting quality questionnaires; and positive and negative emotionality, to name just a few. The field is awash in variables that have been studied. It appears that measures of most of the major psychological and educational constructs have been included in one study or another as putative protective factors. It leads one to ask the question: "What has *not* been included in the study of protective and risk factors?" Is there any variable or variables that are *unique* to this line of research?

The inclusion of such a wide variety of variables used to assess the potential for resilience suggests that researches have taken a case study approach to the research question. The typical list of measures of protective factors reads like a psychological report that includes major areas such as the child's history (physical attributes); status of the home environment (socioeconomic status, parents, siblings, etc.); current academic performance (class grades, standardized achievement test scores); intelligence test scores, behavioral and emotional status (parent and teacher rating scales, interviews, measures of self-concept, clinical classifications). The goal of casting such a broad net has been to determine which of these many variables are most important. This assessment, however, is complicated by the fact that not all of these variables share equal psychometric qualities.

The use of both formal and informal measures of protective factors offers a means of studying the field but the disadvantage of leading to inconsistencies within and across research investigations. For example, social status can be assessed using interviews, unstandardized questionnaires, and peer nominations but the extent to which such methods can be reliably reproduced by other researchers should also be studied. Moreover, the transition from research setting to practical application will require more refined instrumentation than is currently available to practitioners. While these methods may assist in the development of the research base for the study of resilience, well developed, reliable and valid measures are required if the important theoretical contributions made thus far can be utilized in applied settings so that children and other consumers may benefit.

In order to advance instrumentation and measurement in the field of resilience, we will present some suggestions to researchers and practitioners. In the sections that follow, we will discuss some basic measurement issues and illustrate their relevance to clinical practice. Our emphasis is on the application of concepts of resilience by child-serving professionals including both teachers and mental health professionals.

How a Test of Resilience Could Be Developed

Development of a system for measuring variables related to resilience is a task that requires important and well-established test development procedures be followed. The many methods and issues

are amply described, for example, by Crocker and Algina (1986), Nunnally and Bernstein (1994), and Thorndike (1982). Essentially, the typical test development process involves a series of steps designed to yield a defensible and usable measure of a construct or constructs. The process begins with a clear operational definition of the construct or constructs to be measured. This means that all variables of interest must be defined with such clarity that they can be evaluated via some method, be that a rating scale, observational method, or performance test. In the area of resilience, concepts such as sociability, negative affectivity, adaptability, self-referent social cognitions, which have been invoked to explain or understand resilience, would have to be defined with clarity because without a clear definition, hopes for reliable and valid measurement would be difficult at best. Definitional clarity is the sine-qua-non for the development of psychometrically sound assessment measures and approaches. This requirement is made considerably more difficult because of the evolving nature of the field of resilience.

After clearly defining the construct or constructs to be measured, the next step is the development of an initial pool of items to measure those constructs, followed by pilot testing of the items. A key consideration at this stage is adequate sampling of the various behaviors related to the construct under consideration to ensure adequate breadth of coverage, that is, content validity. The items also need to be clear, one-dimensional (that is, describe only one behavior) and, to the extent possible, free of cultural bias. The subsequent pilot tests are designed to evaluate the clarity of the items as well as the general approach to obtaining scores. At this initial stage the ways the items are presented on the page, size of the fonts, clarity of the directions, colors used on the form, position of the items on the sheet of paper, and so on, are considered. Questions like reliability and validity are not usually examined at this point because sample size typically precludes adequate examination of such questions. The goal of pilot testing is very simple—to quickly and efficiently determine if the form seems to work, if the users understand what they need to do, are we on the right track?

The next step is to conduct experiments with larger samples that allow for an examination of the psychometric qualities of the items and their correspondence to the constructs of interest. This phase is repeated until the author has sufficient confidence that the items and the scales have been adequately operationalized and the constructs adequately sampled. In each of the many iterations, experimental evidence is used to answer questions such as:

- What is the mean and standard deviation (SD) of each item?
- Do items designed to measure the same construct correlate with each other?
- Do items designed to measure the same construct correlate with other items designed to measure that same construct at higher levels than they correlate with items designed to measure different constructs?
- What is the internal reliability of those items organized to measure each construct?
- What effect does elimination of each item have on the reliability of the scale on which it is temporarily included?
- What is the factor structure of the set of items and how can item elimination be used to clarify the factor structure?
- Does the scale seem to have validity (defined in a number of different ways)?

This phase, sometimes referred to as a "tryout" stage is repeated until the scale has demonstrated at least minimally acceptable reliability and validity to warrant proceeding with standardization. The number of actual data collection efforts depends on the quality of the original concepts, the quality of the initial pool of items, the quality of the sampling used to obtain the data used to examine these questions, and the results that are found. The goal is to produce a version that is ready to be subjected to large-scale national standardization. The idea is that the cost of standardization is so great that the current status of the instrument must be of high enough quality that the risk of the final assessment failing to meet demonstrates adequate reliability and validity is greatly reduced.

The next to the last step in development of a measure for use in clinical settings is standardization and data collection to establish the reliability and validity of the final measure. This process first requires that a sample of persons who represent the population with whom the measure will be used is administered the measure so that (a) a

final group of items and scales is determined and (b) normative values can be computed. Typically, this is a nationally representative sample. Development of norms is an art as much as a science and there are several ways in which this task can be accomplished (see Crocker & Algina, 1986; Nunnally & Bernstein, 1994; Thorndike, 1982). The second task at this stage is collection of data for the purpose of establishing reliability (internal, test–retest, inter-rater, intra-rater) and validity (construct, criterion, and content, for example). Of these two, validity is clearly the more difficulty psychometric quality to assess.

There are many types of validity and, therefore, validity is not established by any single study. According to the Standards for Educational and Psychologist Testing (AERA, APA, & NCME, 1999) evidence for validity "integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses" (p. 17). It is important to note that it is not the test that is valid (as is commonly thought) but rather the interpretations and uses of test scores. In other words, the authors of the assessment have to demonstrate that the inferences about the construct (e.g., the strength of the individual's protective factors) and the decisions that are made (e.g., the individual is at risk) based on the interpretive guidelines presented in the manual are supported by evidence. That book provides 24 standards that relate to validity issues that should be addressed by test developers. This includes, for example, the need to provide evidence:

- That evidence exists to support interpretations based on the scores the instrument yields
- About the internal structure of the test
- About the organization of scales and composites within a test
- Of the relationship between the scores the instrument yields and one or more criterion variables
- For the utility of the measure across a wide variety of demographic groups or its limitations thereof
- That the measure differentiates between groups as intended

This list represents some of the issues that need to be addressed and is not intended to describe all the issues that should be examined. In the field of resilience, we believe that there are some particularly salient validity issues. For example, can variables related to resilience be operationalized into some measurable system? How effective is the measure for differentiating between children who are at risk and those who are not? How many variables need to be measured to maximally predict resilience? Is a combination of variables related to protective factors in the environment, the family, and the child, the best way to predict resilience? Do protective factors enhance outcomes only for children who are at significant risk, or all children? Can the extensive lists of child protective factors be reduced to a few key characteristics that predict which children may be resilient? The answers to these questions will help define the future of this field.

Once development of an instrument is completed then the important task of documentation begins. There is wide variation in the extent to which test authors document the development, standardization, reliability, and validity, of their measure. Some test manuals provide little if any information of the types we have described above, others provide ample descriptions. We refer the reader to examples such as the Kaufman Assessment Battery for Children—Second Edition (Kaufman & Kaufman, 2004), the Devereux Student Strength Assessment (LeBuffe, Shapiro, & Naglieri, 1999), and the Cognitive Assessment System (Naglieri & Das, 1997). We use these examples because not only do these authors provide detailed discussion of the various phases of development, but they provide extensive discussion of how the tests should be used and the scores the tests yield interpreted.

Development of a measure does not end with the writing of the sections in the manual that describe the development, standardization, and reliability/validity of the instrument. The authors have the added responsibility to inform the users about how the scores can be used to enhance practice and improve outcomes for the individual being assessed (AERA, APA, & NCME, 1999). This may include how the scores on various

scales should be compared with one another and with scores from other tests (if appropriate) to gain a better understanding of the relative strengths and needs of the individual. Increasingly important in this era of evidence-based practice is guidance on the use of scale scores from pretests and posttests to document growth, change, or response to treatment in the individual. It is essential that the authors provide the users with the values needed for determining significance when the various scores a measure provides are compared. The test manuals should provide a thorough discussion of interpretive methods to guide the practitioner. This will enable the user to interpret the scores from an instrument in a manner that is consistent with the intent of the authors and the reliability and validity evidence that was accumulated.

The Importance of Psychometric Characteristics

Why Reliability Matters

Good reliability is essential for all measurements used for research as well as in applied settings to ensure accuracy. Reliability is important to the practitioner because it reflects the amount of error in the measurement. Recall that any obtained score is comprised of the true score plus error (Crocker & Algina, 1986). Because we can never directly determine the true score, we describe it on the basis of a range of values within which the person's score likely falls with a particular level of probability. The size of the range is determined by the reliability of the measurement with higher reliability resulting in smaller ranges. This is why in practice we say, for example, that a child earned an IQ of 105 (±5); meaning that there is a 90% likelihood that the child's true IQ score falls within the range of 100–110 (105 ± 5). The range of scores (called the confidence interval) is computed by first obtaining the standard error of measurement (SEM) from the reliability coefficient and the standard deviation (SD) of the score in the following formula (Crocker & Algina, 1986):

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

The SEM is considered the average standard deviation (68% of the normal curve is in this range) of the theoretical distribution of a person's scores around the true score. Thus, if we add and subtract 1 SEM from an obtained score, we can say that there is a 68% chance (the percentage of scores contained within ± 1 SD) that the person's true score is contained within that range. Recall that 68% of cases in a normal distribution fall within +1 and −1 standard deviation. Second, the SEM is multiplied by a *z* value of, for example 1.64 or 1.96, to obtain a confidence interval at the 90 or 95% levels, respectively. The resulting value is added to and subtracted from the obtained score to yield the confidence interval. For example, the 95% confidence range for a test score with a reliability of 0.95 and an obtained score of 100 is 93 (100 − 7) to 107 (100 + 7). It is important to note that the higher the reliability the smaller the interval of scores that can be expected to include the child's true score. The smaller the range, the more precise practitioners can be in their interpretation of the results, resulting in more accurate decisions regarding the child. The relationships between reliability and confidence intervals are provided in Fig. 14.1 for *T*-scores ($M = 50$; SD = 10) and IQ scores ($M = 100$; SD = 15).

The SEM is, of course, most important when individual decisions are made because the larger the SEM the more likely scores will differ as a function of low reliability. The lower the reliability, the more likely there will be disparity among scores, for example on a variety of measures of protective factors. These inconsistent results can complicate the interpretation of findings and make a clear understanding of a child's strengths and needs more difficult. Without reliable measures of strengths and needs, planning effective support strategies or interventions becomes problematic and ultimately child outcomes may be adversely impacted.

Reliability of specific scores also influences the comparisons among scores. For example, if a researcher or practitioner is concerned with determining if a particular protective factor score received by a child is significantly higher than the scores received on other protective factor scales and therefore represents a significant strength for

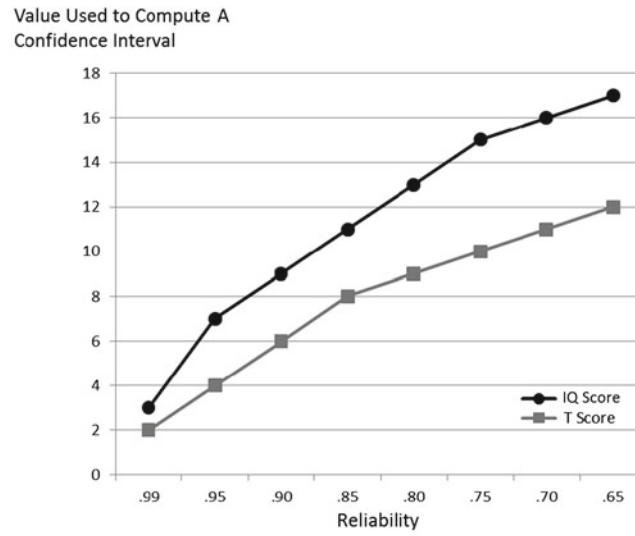

Fig. 14.1 Relationship between reliability and confidence intervals

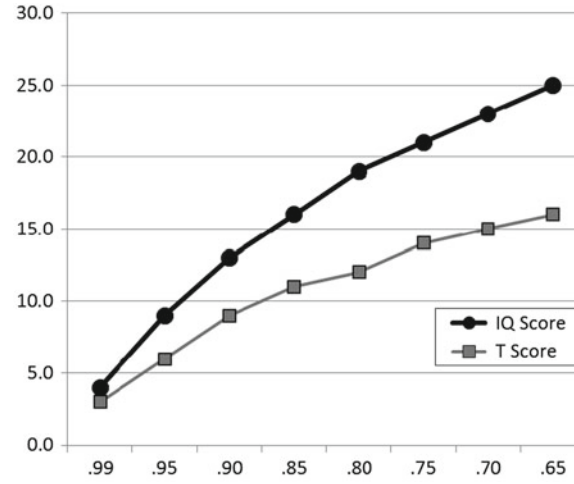

Fig. 14.2 Differences required for significance when comparing IQ or *T*-scores based on scale reliability

the child, the ability to make that determination is directly related to each factor's reliability coefficient because the calculation of the SEM is based on the reliability. In fact, the formula for the difference between two scores earned by an individual is calculated using the SEM of each score.

$$\text{Difference} = Z \times \sqrt{\text{SEM } 1^2 + \text{SEM } 2^2}$$

Applying this formula to IQ test scores and *T*-scores as shown in Fig. 14.2, we see that as the reliability goes down, the differences needed when comparing two scores increase dramatically. This means that scores from measures with reliability of 0.70 from two different teachers would have to differ by 15 points to be significant at the 95% level. This means that test scores with higher reliability reduce the influence of *measurement error* on the different scores. Clearly, in both research and clinical settings, variables with high reliability are needed.

How Much Reliability Is Needed?

Bracken (1987) provided suggested thresholds for acceptable levels of test reliability. He suggested that individual variables should have at

least an internal reliability estimate of 0.80 or greater and total scales an internal consistency of 0.90 or greater. These guidelines should be further considered in light of the decisions being made. For example, if a score is used for screening purposes where over identification is preferred to under identification, a 0.80 reliability standard for a total score may be acceptable. If, however, important decisions are made, for example, dealing with special educational placement, then a higher (e.g., 0.95) standard should be deemed more appropriate (Nunnally & Bernstein, 1994).

In summary, it is advisable that researchers and clinicians who examine scores from measures of protective factors look for scores that have internal reliability estimates of 0.80 or higher and composite scores comprised of several variables that have an internal reliability estimates of 0.90 or greater. If a rating scale's score has not been constructed to meet these requirements, then its inclusion in research and applied practice should be questioned. This is particularly important because the extent to which two variables can reliably correlate is influenced by the reliability of each variable. Clinicians are advised not to use measures that do not meet these standards because there will be too much error in the measurement to allow for confidence in the result. This is especially important because the decisions clinicians make can have significant impact on the life of a child. We therefore urge the reader to carefully examine the reliability findings of any tool they choose to use.

Why Validity Matters

Validity refers to the extent to which empirical evidence and theory supports the recommended uses and interpretations of scores derived from an assessment. Researchers who study resilience are faced with the first responsibility of carefully and clearly defining the construct they intend to evaluate. Given the inferential nature of the study of resilience, one of the greatest validity questions concerns which variables are associated with or predictive of resilience and how is the relevance of each variable demonstrated. Much of the research conducted in this area has attempted to examine these issues to varying degrees. The field has increasingly focused on identifying those variables that predict resilience in the face of adversity.

Validity of a measure of resilience is, therefore, more complicated than demonstrating the validity of an achievement test or measure of depression, for example. The number of variables that has been examined is substantial, there is considerable inconsistency in the psychometric quality of the variables studied, and the research on the relative importance of the many variables is still evolving. This makes for an exciting area of research but one that clinicians should approach with appropriate cautions.

Our view is that practitioners have a responsibility to use measures that have been developed in the manner we have briefly outlined above and that nonstandardized approaches should be avoided. We believe that the quality of the decisions made based on any assessment tool is directly related to the quality of the assessments themselves. Responsible practitioners should be aware of the psychometric attributes of any tools that are used. We will, therefore, discuss the psychometric characteristics of a number of measures available to practitioners so that the relative advantages and limitations of the tools can be understood.

Tools to Measure Variables Related to Resilience

The assessment of factors related to resilience in clinical practice is in its early stages. Although informal, nonstandardized tests and procedures are valuable as initial approaches to assessment, they lack the needed research and development base as well as norms calibrated on a representative national standardization sample to make them useful in research and defensible in practice. To assist educational and clinical professionals who would like to incorporate the assessment of resilience in their professional practice, we provide a review of the tools currently available for this

purpose that meet certain criteria. To be included in this listing, the evaluation tools must: (1) be published so as to be readily available to practitioners, (2) be a standardized, norm-referenced tool, (3) have a technical manual or other accessible source of psychometric information including standardization sample, reliability and validity, (4) be intended for use with children, defined as birth to 18 years. The tools that met these criteria are presented in alphabetical order.

Ages and Stages Questionnaire: Social Emotional

Purpose: The Ages and Stages Questionnaire: Social Emotional (ASQ-SE; Squires, Bricker, & Twombly, 2002) was developed for early identification and remediation of social and emotional deficits in young children. The ASQ-SE was designed for cost-effective large-scale screening of children aged 6–60 months. There are separate questionnaires for each 6-month age interval. The main purpose of the ASQ-SE is to act as a screening test, but, according to the authors, it can also be used to monitor progress, plan for intervention, and conduct research within a comprehensive community-based program.

Scale description: Each of the eight ASQ-SE questionnaires is designed for a specific age range. The number of questions ranges from 22 to 36 depending on the age. The ASQ-SE items cover seven concepts: self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interaction with people. There is also a section to identify general concerns and comments. Responses are calibrated using a multiple point format (*most of the time*, *sometimes*, or *never or rarely*). The rater can also indicate if a particular item is of particular concern. The ASQ-SE yields a total raw score, by adding the item scores; a high score is problematic. Children who receive a total score above a recommended cut-off should be referred for further evaluation. The ASQ-SE can only be completed by a parent rater. The reading level is that of a fifth to sixth grader (Squires, Bricker, & Twombly, 2003).

Psychometric characteristics: The ASQ-SE was standardized on a sample of 2,633 children with approximately 175 cases in each age group. Cronbach's alpha coefficient was reported to range from 0.67 to 0.91. The level of agreement between the total scores over two time intervals (1–3 weeks) was reported as 94%. The overall sensitivity (the ability to accurately identify children with social–emotional disabilities) was reported as 78%. The authors also measured the utility of the ASQ-SE by surveying parents. The results indicated that 97% of parents thought the assessment was "easy to understand and appropriate" (Squires et al., 2003).

Behavioral and Emotional Rating Scale

Purpose: The Behavioral and Emotional Rating Scale, Second Edition (BERS-2; Epstein, 2004) measures behavioral and emotional strengths in children aged 5–19 years using parent, teacher, and a youth self-report rating scales. The BERS-2 is intended to identify protective factors related to the child and the child's family, relying on resilience theory (King, Swerrdlik, & Schneider, 2005). Other purposes outlined in the manual are to identify children who lack strengths and who may be in need of further intervention. The BERS-2 scores can also be used to guide intervention, monitor progress, and evaluate the effectiveness of instructional programs (Epstein, 2004).

Scale description: The BERS-2 has 52–57 items, depending on the rating form. The items are divided into five scales: Interpersonal Strength, Family Involvement, Intrapersonal Strength, School Function, and Affective Strength. There is a Career Strength scale on the youth and parent form as well. The BERS-2 uses a Likert-type format where the rater is asked to reflect on the child's behavior from the last 3 months and answer "not at all like the youth" to "very much like." In addition, there are eight open-ended questions to capture additional information that may aid follow-up assessments or interventions (King et al., 2005). The results of the BERS-2 yield percentile ranks and standard scores for each scale, with a mean of 10 and standard deviation of 3. The

summation of the five scales yields the Strength Index. The rater also receives a summary form that can be used to compare results with other raters (Epstein, 2004).

Psychometric characteristics: The BERS-2 utilized the same standardization sample from the original BERS to create the norms for the teacher form. These norms were based on a sample of 2,176 normally developing children and adolescents, and 861 children and adolescents with emotional/behavioral disorders (King et al., 2005). The parent and youth forms were created and normed with the new standardization samples of 927 and 1,301 youth, respectively. The standardization sample closely matched the 2002 U.S. census data, although slightly under- or over-representing: females, Hispanics, and certain family income levels. The authors reported alpha internal consistency with coefficients ranging from 0.79 to 0.96. Test–retest reliability studies yielded correlations of 0.87–0.99 for the Strength Index. Inter-rater reliability studies indicated correlations of 0.98 for teacher–teacher and 0.54 for parent–child for the Strength Index. The subscales were slightly less reliable with correlations of 0.85–0.96 for teacher–teacher, 0.50–0.63 for parent–child, and 0.20–0.67 for parent–teacher. Validity was examined by comparing the BERS-2 to the Walker-McConnell Scale of Social Competence and School Adjustment—Adolescent Version (Walker & McConnell, 1995), the Systematic Screening for Behavior Disorders (SSBD; Walker & Severeson, 1992), the Scale for Assessing Emotional Disturbance (SAED; Epstein & Cullinan, 1998), the Social Skills Rating System (SSRS; Gresham & Elliot, 1990), and the Achenbach Teacher Report Form (TRF; Achenbach, 1991). Correlations are reported in the form of a table contained in the Examiner's manual (Epstein, 2004).

Devereux Early Childhood Assessment

Purpose: The Devereux Early Childhood Assessment (DECA: LeBuffe & Naglieri, 1999) is a nationally standardized rating scale designed to be used by preschool program directors, teachers, preschool mental health, and early childhood special educators to evaluate protective factors related to resilience in children aged 2–5 years. One of the main goals of the DECA is to help determine if children have developed adequate skills in three areas (Initiative, Self-control, and Attachment) that are related to resilience. Children who receive comparatively low scores in these three strength-based, within-child protective factors may be at risk for developing social and emotional challenges or disorders. By identifying these at-risk children early, strategies can be implemented at school and at home to help develop these protective factors, increasing the odds that the child will be able to successfully adapt to current and future risk and adversity. The rating scale also includes a brief rating of behavioral concerns.

Scale description: The DECA uses a behavior rating scale format which evaluates the frequency with which a child aged 2–5 years demonstrates specific behaviors over the past 4-week interval. A family member or early care and educational professional completes the 37 items which are scored using a 0 (Never) to 4 (Very Frequently) scale. The DECA items are organized into two dimensions: protective factors and behavioral concerns. The Protective Factors included are Initiative (11 items), Self-Control (8 items), and Attachment (8 items). A screener for behavioral concerns (10 items) is included to help identify children with emerging problem behaviors. Items on the Initiative scale assess the child's use of independent thought and action to meet his or her needs. The Self-Control scale includes items about the child's ability to experience a range of feelings and express them appropriately using words and actions. Attachment items determine if the child has developed mutual, strong, and long-lasting relationships with other children and adults. In addition, a Total Protective Factors Scale is provided. The Behavioral Concerns items measure a wide variety of problem behaviors seen in some young children. Separate norms are provided for parent and teacher raters and yield both percentile ranks and *T*-scores. Recommended descriptive terms are provided to aid in communication with parents, teachers, and

other professionals. The term "Strength" is used for protective factor *T*-scores of 60 or above. "Typical" is used to describe *T*-scores of 41–59 inclusive. "Concern" is used to describe low protective factor scores of 40 or below.

Psychometric characteristics: The DECA was standardized on a national sample of 2,017 children aged 2–5 years. The Total Protective Factors Scale reliabilities for Parents and Teachers is 0.93. The average reliabilities across raters for the separate scales are as follows: Initiative (0.87), Self-Control (0.88), Attachment (0.81), and Behavioral Concerns (0.76). The validity of the DECA was studied by comparing children who varied in their social and emotional health. Two samples of children were compared: one group with known emotional/behavioral problems ($N=95$) and another that were considered typical ($N=86$). The results showed that the children with emotional/behavioral problems earned lower scores (less desirable) on the measures of Initiative (effect size (ES) of 0.78), Self-Control (ES = 1.01), Attachment (ES = 0.47), Total Protective Factors (ES = 0.89), and higher scores (also less desirable) on the measure of Behavioral Concerns (ES = 1.08). These results and others presented in the DECA Technical Manual (LeBuffe & Naglieri, 1999) indicated that the children with demonstrated emotional and behavioral problems earned scores that reflect the behavioral difficulties they have and their need for stronger factors that are associated with resilience. See Chap. 10 for more information.

It is important to note that at the time of this writing, the second edition of the DECA is in the final stages of development. The second edition has a larger standardization sample and new norms. More information about this edition can be found in the manual that will be published along with the rating scale.

Devereux Early Childhood Assessment for Infants and Toddlers

Purpose: The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T; Mackrain, LeBuffe, & Powell, 2007) was created to evaluate social–emotional skills in infants and toddlers. The DECA-I/T assesses three protective factors related to resilience: Attachment/Relationships, Initiative, and Self-Regulation. The results of this assessment can be used to identify young children's social–emotional skills and to help identify children who may be at risk or need additional assistance. The DECA-I/T can also be used as an outcome measure for early childhood programs and be used as a research tool.

Scale description: The DECA-I/T is a behavior rating scale for children aged 1 month up to 36 months. The Infant form has 33 items comprised from two protective factor scales: Initiative (18 items) and Attachment/Relationships (15 items). The Toddler form has 36 items comprised from three protective factors scales: Attachment/Relationships (18 items), Initiative (11 items), and Self-Regulation (7 items). The DECA-I/T asks family members and early care and education providers to rate the child's behavior from the past 4-week interval using a 0 (Never) to 4 (Very Frequently) scale. The Attachment/Relationship scale assesses if a mutual, strong, long-lasting relationship has developed between the infant or toddler and a significant adult. The Initiative scale determines the infant or toddler's ability to use independent thought or actions to meet his or her needs. The Self-Regulation scale assesses the toddler's ability to gain control of and manage emotions, and sustain focus and attention. A Total Protective Factors scale is provided, in addition to *T*-scores and percentile ranks for each scale.

Psychometric characteristics: The DECA-I/T was standardized on a national sample of 2,183 infants and toddlers between 4 weeks and 3 years of age. The internal reliability coefficients for the Infant form on the Total Protective Factors scale ranged from 0.90 to 0.94 for parents, and 0.93 to 0.94 for teachers. The reliabilities for the Attachment/Relationships scale ranged from 0.80 to 0.92 for parents and 0.89 to 0.93 for teachers. The reliabilities for the Initiative scale ranged from 0.86 to 0.90 for parents and 0.87 to 0.91 for teachers. The internal reliability coefficients for the Toddler form on the Total Protective Factors

scale was 0.94 for parents and 0.95 for teachers. The reliabilities for the separate scares are as follows: Attachment/Relationships (0.87 for parents and 0.90 for teachers), Initiative (0.92 for parents and 0.94 for teachers), Self-Regulation (0.79 for parents and 0.83 for teachers). The validity of the DECA-I/T was investigated by a contrasted groups approach, examining the scale scores for an identified vs. community samples. Results from both the infant and toddler forms indicate significant and meaningful differences between the identified and community samples on all scales (*d*-ratios range from 0.75 to 1.52). These results are presented in the Technical Manual (Powell, Mackrain, & LeBuffe, 2007).

Devereux Early Childhood Assessment—Clinical Form

Purpose: The Devereux Early Childhood Assessment—Clinical Form (DECA-C; LeBuffe & Naglieri, 2003) is designed to assess factors related to both resilience and emotional/behavioral problems. DECA-C is intended to be used as part of a larger assessment of emotional health and to develop intervention plans that may be needed. For this reason, the DECA-C is intended to be used by those professionals (e.g., psychologists, counselors, and those with clinical training) who have the necessary qualifications to interpret and use this clinical tool as part of child assessment. The information about both protective factors and behavior concerns provides at least three important advantages to the clinician. First, a balanced examination of the child from both positive and concern perspectives is achieved. Second, the examination of the relationships between these dimensions leads to a more complete understanding of how they individually and jointly influence the child's behavior. Third, the inclusion of both dimensions provides important information for intervention planning. See Chap. 10 for more information.

Scale description: The DECA-C uses a behavior rating scale format to evaluates the frequency with which a child aged 2–5 years demonstrated specific behaviors over the past 4-week interval. A family member or early care and educational professional completes the items which are scored using a 0 (Never) to 4 (Very Frequently) scale. The DECA-C is organized into three scales related to resilience (Initiative, Self-control, and Attachment) and four scales about behavioral concerns. These are: Attention Problems (7 items which assess difficulties with focus, distractibility, impulsivity, and hyperactivity); Aggression (7 items used to measure hostile and destructive acts); Emotional Control Problems (8 items which measure the child's difficulties in modifying the overt expression of negative emotions); and Withdrawal/Depression (9 items which address behaviors related to social isolation and lack of reciprocal interactions as well as depressed affect). Like the Total Protective Factors scale, these four Behavioral Concerns scales are combined into a Total score.

The DECA-C was standardized on a national sample of 2,017 children aged 2–5 years and normed to yield *T*-scores set at a mean of 50 and SD of 10. The Total Protective Factors Scale reliabilities for Parents and Teachers is 0.93 and the average reliabilities across raters for the separate scales are: Initiative (0.87), Self-Control (0.88), Attachment (0.81), and Behavioral Concerns (0.76). The average Behavioral Concerns scale internal reliabilities across parent and teacher raters are as follows: Withdrawal/Depression (0.73), Emotional Control Problems (0.83), Attention Problems (0.83), and Aggression (0.82) and the Total Behavioral Concerns Scale (0.91).

Psychometric characteristics: The validity of the DECA-C was examined in a series of research studies summarized in the Manual. In summary, the DECA-C effectively differentiated the groups of children who had known emotion and behavior problems with a matched comparison group of typical preschool children (see LeBuffe & Naglieri, 2003); children with known emotional and behavioral problem showed more signs of behavioral concerns and fewer signs of strong protective factor scores than the DECA-C normative sample; and that the children with documented emotional and behavioral problems in this study had needs in the Protective Factors and Behavioral Concerns Scales of the DECA-C.

The validity of the DECA-C was assessed using several other studies which are reported in the Manual by LeBuffe and Naglieri (2003) and in Chapter 15 in this volume.

Devereux Student Strengths Assessment

Purpose: The Devereux Student Strengths Assessment (DESSA; LeBuffe, Shapiro, & Naglieri, 2009) is a rating scale designed to assess social–emotional competencies that serve as protective factors for children in kindergarten through the eighth grade. The DESSA is completed by parents, teachers, or staff at schools and child-serving agencies, including after-school, social service, and mental health programs. The assessment is comprised entirely of 72 items that are described as strength-based (e.g., how well does the child get along with others). The DESSA is intended to provide a psychometrically sound, strength-based, measure of social–emotional competence in children and youth that can be used to identify individuals at risk of developing social–emotional problems before those problems emerge and identify the strengths and needs of individuals already been identified as having social, emotional, and behavioral concerns.

Scale description: The DESSA is organized into eight conceptually-derived scales that provide information about social–emotional competencies. They are: Self-Awareness (7 items), Social-Awareness (9 items), Self-Management (11 items), Goal-Directed Behavior (10 items), Relationship Skills (10 items), Personal Responsibility (10 items), Decision Making (8 items), and Optimistic Thinking (7 items). The combination of these scales is used to obtain a Social–Emotional Composite score. This composite score provides an overall indication of the strength of the child's social–emotional competence and the eight DESSA scales are used to create profiles for individuals as well as the entire classroom that describe the strengths and needs of the student and/or groups of students as compared to national norms. This information can also be used to compare ratings across raters, environments, and time to monitor progress and evaluate outcomes.

Psychometric characteristics: The DESSA was standardized on a national sample of 2,494 children in grades K through 8 by teachers and parents using both paper and pencil and online versions of the scale. The DESSA standardization sample closely approximated the K–8 population of the United States with respect to age, gender, geographic region of residence, race, ethnicity, and socioeconomic status based on the 2008 U.S. census bureau. The DESSA reliability coefficients for the Social–Emotional Composite for parent raters (0.98) and teacher raters (0.99) both exceed the 0.90 value for a total score suggested by Bracken (1987). The internal reliability coefficients for the eight social–emotional competence scales vary from 0.82 (Optimistic Thinking and Self-Awareness—Parent Raters) to 0.94 (Relationship Skills—Teacher Raters). The median reliability coefficient across these eight scales was 0.86 for parent raters and 0.92 for teacher raters. These values well exceed the 0.80 minimum suggested by Bracken (1987). The validity evidence provided in the scale's Manual suggested that DESSA scores *d* differentiate between groups of children with and without the special education designation of serious emotional disturbance, that the scales do show strong convergent validity with similar measures, and that the Social–Emotional Composite can be considered a measure of within-child protective factors. See LeBuffe, Shapiro, and Naglieri (2009) for more details or Chapter 15 in this volume.

Devereux Student Strengths Assessment—Second Step Edition

Purpose: Devereux Student Strengths Assessment—Second Step Edition (DESSA-SSE; LeBuffe, Naglieri, & Shapiro, 2011) is a 36-item, standardized, norm-referenced behavior rating scale that assesses the social–emotional competencies that serve as protective factors for children in kindergarten through the fifth grade. Developed on the basis of the social–emotional content covered in the Second Step curriculum (Committee for Children, 1997), the DESSA-SSE

can be completed by parents, teachers, or staff at schools and child-serving agencies, including after-school, social service, and mental health programs. Like all the other scales in this line from the Devereux Center for Resilient Children, the assessment uses only strength-based items. The DESSA-SSE was developed to provide a way to evaluate those specific social–emotional competencies taught in the Second Step curriculum. Specifically, the DESSA-SSE has been designed to describe the social–emotional competence of groups of children so that children's progress through the Second Step social–emotional learning program can be evaluated using a psychometrically sound, nationally normed tool.

Scale description: The DESSA-SSE is organized into five scales: Skills for Learning (9 items), Empathy (9 items), Emotional Management (9 items), Problem Solving (9 items), and a Social–Emotional Composite based on all 36 items. Raw scores on each scale are converted to *T*-scores and corresponding percentile ranks and categorical descriptions. The DESSA-Second Step Edition was standardization and normed on a sample of a total of 1,250 children in kindergarten through fifth grades who closely approximated the U.S. population with respect to age, gender, geographic region of residence, race, ethnicity, and socioeconomic status according to the 2008 U.S. census.

Devereux Student Strengths Assessment—Mini

Purpose: The Devereux Student Strengths Assessment—Mini (DESSA-mini) (Naglieri, LeBuffe, & Shapiro, 2010) is a universal screening tool developed to measure social–emotional skills that are related to mental, emotional, and behavioral disorders in order to make early intervention more possible. The DESSA-mini can be used by professionals with or without clinical training to offer a brief summary of a child's current overall social–emotional competence to determine if additional skill development should be provided. The scale can also be used for ongoing progress monitoring during the course of social–emotional interventions. The DESSA-mini is comprised entirely of strength-based items (e.g., get along with others) which are scored on a 5-point scale about how often the student engaged in each behavior over the past 4 weeks.

Scale description: The DESSA-mini is comprised of four 8-item forms which were developed to be highly correlated with the full DESSA and equal in reliability and very similar in overall mean scores. The standardization and normative sample was comprised of a total of 1,250 children and youth in kindergarten through eighth grade who closely approximated the K–8 population of the United States with respect to age, gender, geographic region of residence, race, ethnicity, and socioeconomic status according to the 2008 U.S. census. Each DESSA-mini form yields a *T*-score from the sum of the 8-item ratings.

Psychometric characteristics: The internal reliability of the four 8-item DESSA-mini forms range from 0.91 (mini 4) to 0.92 (mini 3). Each of the DESSA-mini reliability coefficients exceed the 0.90 value for a total score suggested by Bracken (1987). Validity evidence presented in the manual indicates that the DESSA-mini can be used with confidence as a screener for social–emotional competence because (a) DESSA-mini Social–Emotional Total scores are strongly correlated with the Social–Emotional Composite scores on the full DESSA; (b) there is considerable agreement between identification rates based on the DESSA and each DESSA-mini form; (c) the DESSA-mini *T*-scores differentiate groups of children with and without known social–emotional problems; and (d) the DESSA-mini and the DESSA identify children similarly regardless of race or ethnicity.

Hierarchical

Penn Interactive Peer Play Scale

Purpose: The Penn Interactive Peer Play Scale (PIPPS; Fantuzzo, Coolahan, et al., 1998; Fantuzzo et al., 1995) was developed on the idea that children's play interactions are highly indicative of their social and emotional health and predictive of future social and academic success.

This behavioral rating scale was developed with Head Start teachers and parents, assessing peer play interactions with high-risk urban youth. There is a teacher form, which is utilized in the classroom and on the playground, and there is a parent form, which is utilized in the home and neighborhood (Fantuzzo et al., 1995). The PIPPS aims to measure children's play strengths in kindergarten and is intended to be used for screening, assessment, informing curriculum, and promoting communication between parents and teachers (Fantuzzo & Hampton, 2000). The PIPPS is also only intended to be used with urban, low-income, minority children. The PIPPS was developed to identify resilient children in high-risk situations, differentiate children with positive peer interactions from those who were less successful, and to inform interventions (Fantuzzo et al., 1995).

Scale description: The PIPPS was originally standardized on a group of 312 African American high-risk children aged 38–63 months. The participants included 38 teachers from five different Head Start programs. Fantuzzo et al. utilized an exploratory factor analysis of the original items to uncover three constructs: Play Interaction, Play Disruption, and Play Disconnection. Both the teacher and the parent versions consist of 32 items. This behavior rating scale is in a Likert-type format (*never*, *seldom*, *often*, or *always*) revealing how often the teacher or parent witnessed the child displaying a certain behavior. The Play Interaction scale measures the child's play strengths, the Play Disruption scale measures antisocial behaviors that can interrupt play interactions, and the Play Disconnection scale measures withdrawal from play. The PIPPS is not intended to categorize students. If the results indicate that a child has poor play interactions, further evaluation is recommended in addition to efforts to bolster the child's skills in that area (Fantuzzo et al., 1995).

Psychometric characteristics: The PIPPS demonstrates reliability and validity in urban, low-income, African American, Kindergarten youth. Cronbach's alpha for the three scales ranges from 0.87 to 0.91. The construct validity of the PIPPS was determined using exploratory factor analysis. The PIPPS was reported to be significantly correlated with the SSRS. The PIPPS also demonstrates reliability and validity in low-income preschool children, utilizing the same comparisons as articulated above (Hampton & Fantuzzo, 2003).

Preschool Behavioral and Emotional Rating Scale

Purpose: The Preschool Behavioral and Emotional Rating Scale (PreBERS; Epstein & Synhorst, 2009) is an assessment that measures the emotional and behavioral strengths in preschool children aged 3–5 years. The preBERS can be used to identify children with low levels of emotional and behavior strengths, inform IEPs or IFSPs, guide intervention, and monitor progress. This rating scale can be completed by any adult with adequate exposure to the child and can be scored and interpreted by any professional adult who had appropriate training in tests and measurement. The preBERS is entirely strength-based and grounded in resilience research. The overarching goal of this assessment is early identification of children who may need additional support or interventions (Epstein & Synhorst, 2009).

Scale description: The preBERS has 42 items that are divided into four dimensions: Emotional Regulation (13), School Readiness (13), Social Confidence (9), and Family Involvement (7). There are seven open-ended questions that aim to capture any additional social, family, or community strengths. The assessment is written at a fifth-grade reading level and was created to be completed in 10 min. Each item is rated on a Likert-type scale (0 = not at all like the child, 1 = not much like the child, 2 = like the child, and 3 = very much like that child) (Drevon, 2011). The subscales each yield a raw score, a percentile rank, and scaled standard scores. The summation of the subscales yields the total scaled score or Strength Index, which is also reported in a percentile rank and a descriptive term (*Very Superior*, *Superior*, *Above Average*, *Average*, *Below Average*, *Poor*, or *Very Poor*) (Epstein & Synhorst, 2009).

Psychometric characteristics: The preBERS has a set of norms for three different standardization

samples: typical preschool children, Head Start preschool children, and Special Education preschool children. The sample size for these groups was 1,471, 962, and 1,103, respectively. Each sample was compared to the U.S. census by region, race, ethnicity, gender, parental education, family income, and disability status. The samples were mostly representative, but with some regional discrepancies in both the Head Start and Special Education groups (Drevon, 2011). The preBERS reported good internal consistency for the Strength Index, with correlations ranging from 0.96 to 0.98. Correlations were good for each subscale, as well, ranging from 0.84 to 0.97. Short-term test–retest data for the Strength Index indicated high corrected correlations, equaling 0.80 in teachers and 0.95 in parents. The subscale correlations ranged from 0.81 to 0.89 in teachers and 0.88 to 0.97 in parents. Long-term test–retest data revealed a corrected correlation of 0.79 in teachers and 0.85 in parents for the Strength Index and subscale correlations ranging from 0.72 to 0.89 in teachers and 0.83 to 0.92 in parents. The preBERS reported teacher and paraprofessional inter-rater corrected correlations between 0.71 and 0.85 for the subscales, with a 0.72 corrected correlation in the Strength Index (Epstein & Synhorst, 2009).

Resiliency Scales for Children and Adolescents

Purpose: The Resiliency Scales for Children and Adolescents (RSCA; Prince-Embury, 2008) aims to identify and measure personal qualities and vulnerabilities related to resiliency in youth aged 9–18 years. The RSCA is a screener, but can also be utilized to plan and monitor progress and outcomes. The scales are available only in a self-report format and can be administered by qualified supervisors who are professionals, knowledgeable of psychological tests and assessments (Prince-Embury & Steer, 2010). The RSCA can be used to evaluate children and adolescents' personal resiliency.

Scale description: The RSCA items are written on a third-grade reading level and use a 5-point Likert-type scale 0 (*never*) to 4 (*almost always*) to measure three global scales: Sense of Mastery (20 items), Sense of Relatedness (24 items), and Emotional Reactivity (20 items), for a total of 64 items. Each global scale consists of a group of subscales—sense of mastery: optimism, self-efficacy, and adaptability; sense of relatedness: trust, perceived social support, comfort, and tolerance; emotional reactivity: sensitivity, recovery, and impairment. The raw scores of the RSCA are converted to *T*-scores (Prince-Embury & Steer, 2010).

Psychometric characteristics: The RSCA was standardized on a group of 200 children aged 15–18 years. The sample was compared to the U.S. census on both parent education and ethnicity within each year of age and also by gender (Prince-Embury, 2008). All three global scales displayed good internal consistency scores, with alpha coefficients ranging from 0.83 to 0.95. The RSCA indicated test–retest reliability through a 12-day interval (on average), yielding correlations of 0.70–0.92. To establish validity, the RSCA was correlated with the Reynolds Bully Victimization Scale (Reynolds, 2004), the Brown ADD Scales for Children (Brown, 2001), and then Beck Youth Inventories (BYI-II; Beck, Beck, Jolly, & Steer, 2005; Sink & Mvududu, 2010). Psychometric properties for the RSCA were further explored in clinical samples of children ($n = 110$) and adolescents ($n = 178$) revealing good internal consistency among the three global scales with alpha coefficients ranging from 0.82 to 0.90 in the child population and from 0.92 to 0.94 in the adolescent population (Prince-Embury, 2010).

Conclusions

Initial conceptualizations of psychological concepts have a history of being retained across generations of psychologists. Once an idea is proposed, and especially if it is operationalized in a practical method, it can become widely used before researchers have adequately determined the ultimate value and utility of the concept. Perhaps one of the best examples is the Stanford-Binet and Wechsler IQ tests which have changed

Table 14.1 Psychometric characteristics of scales used to measure variables related to resilience

Rating scale	No. of items	Age range	Informants	Scores for scales	Comparison sample size	Sample description	Match to US population
Ages and Stages Questionnaire: Social–Emotional (ASQ-SE)	Varies	3–66 months	Parents	Raw score	2,633	National sample	No
Behavioral and Emotional Rating Scale (BERS)	52	6–9 years	Teachers, parents, self	Raw scores, percentiles, scales scores	2,176	National sample	Yes
Devereux Early Childhood Assessment (DECA)	37	2–5 years	Parents and teachers	*T*-score	2,000	National sample	Yes
Devereux Early Childhood Assessment—Clinical (DECA-C)	62	2–5 years	Parents and teachers	*T*-score	2,000	National sample	Yes
Devereux Early Childhood Assessment—Infant Toddler (DECA-IT)	33 (infant form) and 36 (toddler form)	1–36 months	Parents and teachers	*T*-score	2,183	National sample	Yes
Devereux Student Strengths Assessment (DESSA)	72	5–14 years	Parents and teachers	*T*-score	2,500	National sample	Yes
Devereux Student Strengths Assessment—Mini (DESSA-mini)	Four 8 item forms	5–14 years	Teachers	*T*-score	1,250	National sample	Yes
Devereux Student Strengths Assessment—Second Step Edition (DESSA-SSE)	36 items	5–14 years	Teachers	*T*-score	1,250	National sample	Yes
Penn Interactive Play Scale	32	preK & K	Parents and teachers	*T*-score	312	African American Head Start populations living in high-risk, low income urban populations	No
Preschool Behavioral and Emotional Rating Scale (preBERS)	42	3–6 years	Parents and teachers	Scaled scores	1,471	Typical preschool, head start, and early childhood special education	Yes
Resiliency Scales for Children and Adolescents (RSCA)	64	9–18 years	Self report	*T*-score	650	National sample	No

little since they were first published in the early 1900s. Similarly, because initial conceptualizations have such an important influence on the field, advocates of a concept such as resilience and the variables that lead to it should be mindful of the power of initial conceptualizations.

Researchers and practitioners need to be mindful that the various tools summarized in Table 14.1 of this chapter have both definitional and operational influence. Although there is a growing number of new methods for assessing the likelihood of resilience there is, as yet, much more work has to be accomplished just to adequately define the concept and the methods used in the assessment process. The use of any one of the tools described in this chapter may provide useful information about a child, but such information needs to be integrated into a larger picture. Each of the tools summarized in this chapter provides a limited examination of the child and they should be used accordingly. This is particularly important because the list of variables that influence resilience is very large and diverse, including the child's characteristics (psychological and physical), the family, both immediate and extended, as well as the community and larger societal factors. Additionally, the determination of which combination of variables best predicts resilience and the complex interactions of these variables is still evolving.

Transformation of research findings into clinical practice is always tricky, and it is especially so for the concept of resilience. Application of this concept in the educational and clinical environments would benefit from greater consensus regarding the definition of resilience, the identification and measurement of protective factors, and agreement on which protective factors should be measured. Most importantly, which protective factors, especially in the within-child domain, can be strengthened, and how, and to what effect?

Clinicians should be cautious when applying the concept of resilience and they should be particularly mindful of the psychometric issues that limit application. We suggest that when given the option, measures that have documented psychometric characteristics and have norms based on a national standardization should be preferred and used within the boundaries specified by the authors. The use of well-developed, psychometrically sound assessments will greatly enhance the likelihood that we will be able to (a) obtain good information about the variables related to resilience and (b) develop and evaluate ways to improve social and emotional outcomes for children.

References

Achenbach, T. M. (1991). *Teacher Report Form*. Burlington: Department of Psychiatry, University of Vermont.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

Beck, J. S., Beck, A. T., Jolly, J. B., & Steer, R. A. (2005). *Beck Youth Inventories* (2nd ed.). San Antonio, TX: Pearson.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5*, 313–326.

Brown, T. E. (2001). *Brown Attention-Deficit Disorder Scales for Children and Adolescents*. San Antonio, TX: Pearson.

Chandler, L. A. (1981). The source of stress inventory. *Psychology in the Schools, 18*(2), 164–168.

Committee for Children. (1997). *Second step: A violence prevention curriculum* (2nd ed.). Seattle, WA: Committee for Children.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Hold, Rinehart and Winston.

Drevon, D. D. (2011). Test review [Review of the preschool behavioral and emotional rating scale, by M. H. Epstien & L. Synhorst]. *Journal of Psychoeducational Assessment, 29*(1), 84–88.

Epstein, M. H. (2004). *Behavioral and emotional rating scale, second edition: Examiner's manual*. Austin, TX: PRO-ED, Inc.

Epstein, M. H., & Cullinan, D. (1998). *Scale for Assessing Emotional Disturbance*. Austin, TX: PRO-ED.

Epstein, M. H., & Synhorst, L. (2009). *Preschool behavioral and emotional rating scale: Examiner's manual*. Austin, TX: PRO-ED, Inc.

Fantuzzo, J., Coolahan, K., Mendez, J., McDermott, P., & Sutton-Smith, B. (1998). Contextually-relevant validation of peer play constructs with African American head start children: Penn interactive peer play scale. *Early Childhood Research Quarterly, 13*, 411–431.

Fantuzzo, J., Sutton-Smith, B., Coolahan, K. C., Manz, P. H., Canning, S., & Debnam, D. (1995). Assessment of preschool play interaction behaviors in young low-income children: Penn Interactive Play Scale. *Early Childhood Research Quarterly, 10*, 105–120.

Fantuzzo, J. W., & Hampton, V. R. (2000). Penn Interactive Peer Play Scale: A parent and teacher rating system for young children. In K. Gitlin-Weiner, A. Sandgrund, & C. Schaefer (Eds.), *Play diagnosis and assessment* (2nd ed., pp. 599–620). Hoboken, NJ: Wiley.

Garb, J. N. (2003). Clinical judgement and mechanical prediction. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology, (Assessment psychology, Vol. 10,* pp. 27–43). New Your: Wiley.

Gresham, F. M., & Elliot, S. N. (1990). *Social Skills Rating System test manual*. Circle Pines, MN: American Guidance Service.

Hampton, V. R., & Fantuzzo, J. W. (2003). The validity of the Penn Interactive Play Scale with urban, low-income kindergarten children. *School Psychology Review, 10*(1), 77–91.

Kanner, A. D., Coyne, J. C., Schaefer, C., & Lazarus, R. S. (1981). Comparison of two modes of stress management: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine, 4*(1), 1–37.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (2nd ed.). San Antonio, TX: Pearson.

King, J., Swerrdlik, M. E., & Schneider, J. W. (2005). Review of behavioral and emotional rating scale. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (2nd ed.). Lincoln, NE: Buros Institute of Mental Measurements.

LeBuffe, P. A., & Naglieri, J. A. (1999). *Devereux Early Childhood Assessment (DECA) assessment, technical manual, and user's guide*. Lewisville, NC: Kaplan.

LeBuffe, P. A., & Naglieri, J. A. (2003). *Devereux early childhood assessment—clinical form*. Lewisville, NC: Kaplan.

LeBuffe, P. A., Naglieri, J. A., & Shapiro, V. B. (2011) *The Devereux Student Strengths Assessment–Second step edition (DESSA-SSE)*. Lewisville, NC: Kaplan.

LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009). *The Devereux Student Strengths Assessment (DESSA) assessment*. Lewisville, NC: Kaplan.

Luthar, S. S., & Zelazo, L. B. (2003). Research on resilience: An integrative review. In S. S. Luthar (Ed.), *Resilience and vulnerability: Adaptation in the context of childhood adversities* (pp. 510–549). Cambridge: Cambridge University Press.

Mackrain, M., LeBuffe, P. A., & Powell, G. (2007). *The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T) assessment*, Lewisville, NC: Kaplan.

Masten, A., & Garmezy, N. (1985). Risk, vulnerability, and protective factors in developmental psychopathology. In B. Lahey & A. Kazdin (Eds.), *Advances in clinical child psychology*. New York: Plenum Press.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56*, 227–238.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system interpretive handbook*. Itasca: Riverside Publishing Company.

Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2010). *The Devereux Student Strengths Assessment–Mini (DESSA-Mini) assessment*, Lewisville, NC: Kaplan.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Powell, G., Mackrain, M., & LeBuffe, P. (2007). *The Devereux Early Childhood Assessment for Infants and Toddlers (DECA-I/T), technical manual*, Lewisville, NC: Kaplan.

Prince-Embury, S. (2008). *Resiliency Scales for Children & Adolescents (RSCA)*. San Antonio, Pearson.

Prince-Embury, S. (2010). Psychometric properties of the resiliency scales for children and adolescents and use for youth with psychiatric disorders. *Journal of Psychoeducational Assessment, 28*(4), 291–302.

Prince-Embury, S., & Steer, R. A. (2010). Profiles of personal resiliency for normative and clinical samples of youth assessed by the resiliency scales for children and adolescents. *Journal of Psychoeducational Assessment, 28*(4), 303–314.

Reynolds, W. (2004). *Reynolds Bully Victimization Scale*. San Antonio, TX: Pearson.

Sink, C., & Mvududu, N. R. (2010). Review of the resiliency scales for children and adolescents. In R. A. Spies, K. F. Geisinger, & J. F. Carlson (Eds.), *The eighteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Squires, J., Bricker, D., & Twombly, E. (2002). *The ASQ:SE user's guide*. Baltimore: Paul H. Brookes Publishing.

Squires, J., Bricker, D., & Twombly, E. (2003). *The ASQ:SE User's Guide for the Ages & Stages Questionnaires, social-emotional: A parent completed, child-monitoring system for social-emotional behaviors*. Baltimore, MD: Paul H. Brookes Publishing.

Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton-Mifflin.

Walker, H. M., & McConnell, S. R. (1995). *The Walker-McConnell Scale of Social Competence and School Adjustment–Adolescent Version*. San Diego: Singular.

Walker, H. M. & Severeson, H. H. (1992). *Systematic screening for behavior disorders*. Longmont, CO: Sopris West.

Werner, E. E. (2005). What can we learn about resilience from large-scale longitudinal studies? In S. Goldstein & R. Brooks (Eds.), *Handbook of resilience in children*. New York, NY: Kluwer.

Work, W. C., Cowen, E. L., Parker, G. R., & Wyman, P. A. (1990). Stress resilient children in an urban setting. *Journal of Primary Prevention, 11*(1), 3–17.

Wright, M. O., & Masten, A. S. (2005). Resilience processes in development: Fostering positive adaptation in the context of adversity. In S. Goldstein & R. Brooks (Eds.), *Handbook of resilience in children*. New York, NY: Kluwer.