

## Measuring Outcomes with the DESSA

Jennifer L. Fleming and Paul A. LeBuffe

December 2014

The Devereux Student Strengths Assessment (DESSA; LeBuffe, Shapiro, & Naglieri, 2009, 2014) is a 72-item, standardized, norm-referenced behavior rating scale that assesses social-emotional competencies of children in kindergarten through eighth grade. It was designed to be used in a variety of settings including schools, out-of-school time programs, and child welfare settings, and can be completed by parents, teachers, and program staff. It assesses eight key social and emotional competencies: Self-Awareness, Self-Management, Social Awareness, Relationship Skills, Personal Responsibility, Decision Making, Goal-Directed Behavior and Optimistic Thinking. The DESSA yields a total score, referred to as the Social-Emotional Composite (SEC), as well as scores for each of the eight competencies.

The value of the DESSA is that it not only provides a psychometrically-sound measure of social and emotional competence in individual children and groups of children, but it can also be used for progress monitoring and program evaluation purposes.

**Progress monitoring** consists of administering an assessment such as the DESSA multiple times throughout a school year in order to examine how children are responding to a particular intervention or program. The goal of progress monitoring is to obtain feedback during the intervention or program so that the nature or intensity of the intervention or program can be adjusted to maximize the chance of a successful outcome. Progress monitoring may occur multiple times over the year. In contrast, **program evaluation** refers to examining the progress or change observed over a defined period of time such as a school year to determine overall program effectiveness.

The purpose of this guide is to aid educators in utilizing the DESSA to both monitor the acquisition of social and emotional competencies of children and to evaluate the impact of interventions, such as social and emotional learning (SEL) programs. We will first describe the importance of using a standardized measure to plan and inform programs and interventions generally. Next, specific information about the DESSA will be presented related to

measuring outcomes, including initial administration, the types of scores derived from the DESSA, and a brief description of measurement error. Following this introduction to the topic, three sections will be presented describing concrete ways the DESSA can be used to measure outcomes for 1) individual children, 2) groups of children (e.g., classrooms, grades, buildings), and 3) program evaluation and quality improvement purposes.

### IMPORTANCE OF USING A STANDARDIZED ASSESSMENT AT DIFFERENT STAGES OF IMPLEMENTATION

In order to effectively implement an SEL program or intervention, it is essential to base decisions on some reliable indicator. This can be viewed as an ongoing process, in which the indicator (e.g., social and emotional competencies as measured by the DESSA) is used prior to beginning the implementation of an SEL program, throughout the duration of a program to monitor progress, and following the implementation of a program to evaluate outcomes.

Collecting needs-assessment data is an important consideration when initially selecting an SEL program to use. This data will help to inform administrators about the type of program to select. For example, if large groups of children are found to fall within the need for instruction range on a competency such as self-management, it would be necessary to select a curriculum that targets skills related to self-management.

Once an SEL program has been selected and implementation has begun, it is important to periodically collect data to monitor how children are responding to the program. A typical structure for collecting this data would be to administer an assessment such as the DESSA at the start of the program (usually at the beginning of a school year), mid-way through the school year, and then at the end of the school year. By doing so, administrators would have information about how individual children or

groups of children are responding to the intervention. They could then, based on these results, modify the interventions prior to the end of the school year. This may include adjusting the frequency or intensity of the interventions being implemented, or moving from group interventions to more individualized supports. This approach enables practitioners to identify and modify ineffective practices while the program is being implemented thereby maximizing the likelihood of positive outcomes by the end of the program period. Ultimately, this information would allow educators to be more intentional as they deliver lessons and reinforce skills with the children to have the greatest impact.

Additionally, by assessing particular outcomes, such as social-emotional competencies, administrators would have the ability to evaluate the impact of the curriculum. For example, administrators would be able to assess whether a program or intervention led to improvements in children's social-emotional competencies. This impact could further be examined at the classroom, grade, building, or district-level and inform decisions about quality improvement, such as professional development opportunities to provide for staff.

## GENERAL INFORMATION ABOUT THE DESSA

It is recommended that a rater have approximately four weeks of exposure to a child prior to completing a rating. More specifically, it is recommended that raters have contact with the child for two or more hours for at least three days per week over the course of the four-week period, which equates to approximately 24 hours of exposure to the child. Ultimately, it is essential that a rater has sufficient opportunity to observe the child's behavior in a variety of situations and know a child well enough to complete a DESSA rating and do so accurately.

Similarly, when examining changes in a child's scores over time, a period of at least four weeks between ratings should elapse. This allows each rating to reflect a new sample of behaviors and provide time for changes in behavior to occur. Ideally, the same rater should be used for both the pretest and posttest ratings. However if this is not possible, it is essential

that the same type of rater (parent or teacher/staff) be used at all administrations.

Two standard scores are provided by the DESSA: percentile scores and *T*-scores. **Percentile scores** describe the child's relative position compared to other children who have been assessed on the DESSA. Although percentile scores are useful for communicating a child's relative standing on the assessment, they should never be used when comparing scores across scales or in statistical computations. Only *T*-scores should be used for this purpose. ***T*-scores** are standard scores that have a set mean of 50 and a standard deviation of 10, with equal units along the scale. This ensures that differences in *T*-scores have the same meaning throughout their range, and therefore can be easily compared across children and ratings. All statistical analyses discussed in this guide will use *T*-scores unless otherwise stated. For a more thorough discussion of the uses of different types of scores, see the DESSA manual (Chapter 5).

*T*-scores on the DESSA are categorized as follows: *T*-Scores of 60 and above indicate that the child has social and emotional "*strengths*," *T*-scores of 41-59 inclusive indicate that the child is showing a "*typical*" amount of positive behaviors related to social and emotional competence, and *T*-scores of 40 and below indicate that the child has a "*need for instruction*" to further develop social and emotional skills as compared to the national standardization sample. These descriptive ranges can be used to better understand the scores received on the eight DESSA scales and the SEC.

When interpreting results from any assessment, it is important to keep in mind that a given score will have some amount of **measurement error**, or random fluctuations in scores. Random error could result from such things as the rater's mood on the day of a rating. Therefore, it is expected that an assessment administered more than once may not provide the same score each time, even if behavior remains stable. Rather, scores will sometimes increase or decrease slightly due to random error alone. Importantly, random error will not have a consistent effect across an entire group of children, some of whom may show an increase and others a decrease in their scores. Because of these random, minor fluctuations in scores, you must determine whether the differences in scores between time points exceed what you would expect from chance variation. This concept is referred to as

**statistical significance.** By establishing statistical significance, you can confidently determine that the change in children's scores is not the result of chance and instead reflects an improvement or decline of the behavior or competency you are assessing. The following statistical analyses will help you to accomplish this for individual children, groups of children, and for program evaluation/quality improvement purposes.

## MEASURING OUTCOMES FOR INDIVIDUAL CHILDREN

One of the benefits of using an assessment such as the DESSA is that it allows you to monitor how an individual child is progressing in the acquisition of social and emotional skills throughout a period of intervention. One approach that can be used for this purpose is the **standard error of prediction**. This approach enables you to evaluate whether the changes in a child's *T*-scores from pretest to posttest are statistically significant. This will then inform your decision to continue with the intervention strategies already in place or alter the type or frequency of the strategies being used.

The standard error of prediction, described by Atkinson (1991), is an approach in which the user can compare the child's posttest score with a predetermined range of scores reflecting measurement error based on the pretest score. If the posttest score falls within the range of scores provided, then the difference can be explained by measurement error and you cannot conclude that a significant change occurred between pretest and posttest. However, if the posttest score falls outside the range, then you can conclude that a statistically reliable change occurred between pretest and posttest. A posttest score falling below the range would indicate the child's skills showed significant decline, while a posttest score above the range would indicate significant improvement.

The standard error of prediction has been calculated and posttest confidence ranges for each DESSA scale and the SEC are located in the DESSA manual, Appendix B, Table 1 (for parents) and 2 (for teachers). The following steps will allow you to use this approach.

**Step 1:** Using the appropriate table based on the rater (found in Appendix B of the manual),

find the child's pretest DESSA *T*-score in the first column labeled "Pretest Obtained Score."

**Step 2:** Read across the table to the column that corresponds to the DESSA scale being evaluated.

**Step 3:** If the child's posttest DESSA *T*-score falls within the posttest range provided in the table, there has been no significant change in the child's score. If, however, the posttest score falls above the posttest range, we can conclude that the child's score has shown significant improvement. If the posttest score falls below the range provided, then we conclude that the score has shown significant decline. This approach can be used to gauge the progress shown by each child on each of the eight DESSA scales and the SEC.

### Example

Suppose a child receives a *T*-score of 37 on the DESSA Self-Management scale at the beginning of the school year. This indicates that the child is in need of instruction in this area. Based on this information, a teacher decides to implement a variety of strategies designed to improve the child's social and emotional skills related to Self-Management. The teacher implemented the interventions for the first half of the school year, and in January, re-administered the DESSA for this child. At this administration, the child receives a *T*-score of 42 on the Self-Management scale, which falls just within the typical range of scores.

Although this appears to be an improvement, the teacher wants to check whether the increase in scores was significant (i.e., not due to measurement error). Therefore, the teacher examines the standard error of prediction by looking up the child's pretest score on the Pretest-Posttest Comparison Table in Appendix B of the manual. In order for the change to be significant, the posttest *T*-score would need to fall outside the range of 31-45. In this case, the child received a score of 42, indicating that no significant change has occurred. Based on this finding, the teacher chooses to continue using the intervention strategies designed, but decides to increase how often the strategies are used with the child.

At the end of the school year, the teacher administers the DESSA for a third time. This time, the child receives a *T*-score of 52 on the Self-Management scale, which is well within the typical range of scores.

Following the same process as last time, the teacher finds the pretest score (now 42) on the table and finds the posttest confidence range of scores to be between 35 and 50. This time, the teacher can confidently say that the child's Self-Management *T*-score significantly improved since mid-year.

### *Summary*

The standard error of prediction is one approach to determining whether a significantly reliable change has occurred in *T*-scores from pretest to posttest. This approach offers the advantage of providing specific feedback for each scale and can be used with both individual children and with groups of children (classroom, school, etc.) to monitor progress during the school year. Results derived from this approach can help inform the decision to continue an intervention as is or to modify an intervention (such as increasing the type or frequency of an intervention). However, the major limitation of the standard error of prediction is that it requires a demanding criterion to show reliable change. A large gain in the posttest *T*-score is needed to reach significance. In the example just described, even though the child showed a 5-point improvement between the first and second administration, the improvement was not deemed significant using this approach. This can lead to frustration when a child's *T*-score falls just shy of a required value.

## MEASURING OUTCOMES FOR GROUPS OF CHILDREN

Groups of children may refer to a classroom, an entire grade level, all children within a school or out-of-school time program, or children within a district or community. Following the assessment administration, the scores for the grouping(s) of interest need to be examined and interpreted. As a first step, descriptive statistics could be used to better understand and describe the data. This includes techniques such as examining the mean, median, mode or range of scores the children received. The [Classroom Profile](#) is also a useful tool to examine a group of children's *T*-scores at one time. This tool provides teacher/staff with a way to see which children and how many children are

receiving scores in the strength, typical, or need for instruction ranges. Multiple Classroom Profiles, obtained at different points in time, can be compared and changes noted.

However, it's often important to move beyond describing the data to drawing conclusions about the effectiveness of the intervention for the group as a whole. In order to accomplish this inferential statistics can be used. Although there are a number of statistical tests that can be used, this guide will focus on the simplest and easiest to use inferential statistic, the paired-samples *t*-test. It is important to note that conducting this statistical test requires the use of a statistical software program, such as SPSS.

### *Paired Samples t-test*

Suppose a teacher assessed her class of students using the DESSA at the beginning of the school year (pretest) and the end of the school year (posttest). During the year, she implemented an SEL program with her class, and is now interested in determining whether the children as a group improved in their social and emotional skills throughout the year. The Social-Emotional Composite (SEC) is the best scale to use for this purpose because it is the best overall indicator of the students' social and emotional competence. The teacher first calculates the mean (or average) SEC *T*-score for the class at both pretest and posttest and finds that the mean score increased from 41 at pretest to 46 at posttest. However, the teacher knows that the improvement in scores could just be due to chance, so she conducts a **paired samples *t*-test** on the scores. A paired samples *t*-test is selected because it compares the mean scores of the *same* group of students at two time points to determine whether a statistically significant change between scores occurred.

This analysis can be conducted with any statistical software package. Once conducted, the analysis will need to be interpreted. To interpret it, look at the significance level (called a *p* value) in the analysis output. If the *p* value is less than the significance level set (typically .05), then you can say that the change in scores between pretest and posttest is significant and not due to chance. In this example, the teacher can state that the children in her class showed significant improvements in their overall social and emotional competencies from the beginning to the end of the school year.



This analysis can also be used to examine the change in scores for each scale of the DESSA. To do this, simply conduct multiple paired samples *t*-tests. However, it is important to keep in mind that whenever multiple comparisons are made, there is a greater likelihood that one of the findings could be significant due to chance alone. To account for this, the Bonferroni correction is used. This is a statistical technique that alters the significance level set. To use it, divide .05 (the standard significance level) by the number of comparisons being made and then use that value to determine if the analyses are significant. For example, if the teacher is conducting 8 comparisons (one for each DESSA scale), she would look for *p* values less than .006 ( $.05/8 = .006$ ). In other words, you are simply being more conservative in your acceptance of what is a significant finding.

### *Summary*

Inferential statistics are useful for determining whether significant changes have occurred across multiple administrations of an assessment for a group of students. This can be useful when determining whether a classroom, grade, building, district, or community significantly improved as a result of an intervention or program. However, there are two major limitations to this approach. First, inferential statistic analyses require a statistical software package that can be costly. They also require someone with statistics experience to conduct and interpret the analyses. For assistance conducting or interpreting these analyses, contact a school psychologist or a local college or organization in your community who may be able to help. Additionally, you can contact the Devereux Center for Resilient Children to discuss opportunities for data analysis support.

The second major limitation of this approach is that, unlike the standard error of prediction described above, it does not provide information about individual students. In other words, you cannot determine whether one student significantly improved from one administration to another. It is possible that some students are responding to the program while others are not benefiting from it. The next approach described in this section will begin to address each of these limitations.

### *Examining the magnitude of change for groups of children*

In a group of children, it is possible that not all children are responding in the same way to the SEL program or intervention being used. While some children may show improvement in their social and emotional skills, other children may display no changes in their behavior or may actually display declines in competencies due to a number of factors. For example, children are constantly experiencing change and potential risk factors in their lives such as moving to a new neighborhood or school, parental divorce, or substance abuse in the home, which all may impact their social and emotional competencies. Other factors such as school or program attendance may also influence a child's response to an SEL program or intervention. Therefore, it is often useful to delve a little deeper into group data to explore whether some children are showing more change compared to other children. This can better inform and guide interventions throughout a school year and help to explain overall program outcomes.

The approach described in this section is based on guidelines for Cohen's *d*-ratio (1988), which can be used to determine the magnitude or size of the difference between two successive assessment scores. In this case, it can be used to examine scores on the DESSA SEC or on the eight DESSA scales. According to Cohen's guidelines, a change of less than 2 *T*-score points indicates no change, a change of 2-4 inclusive *T*-score points indicates a small change, a change of 5-7 inclusive indicates a medium change, and a change of 8 or more *T*-score points indicates a large change. These guidelines can be used for both positive and negative changes, so for example a change of 6 *T*-score points between a pre and posttest would indicate a medium positive change, while a -6 *T*-score point change between pre and posttest would indicate a medium negative change. It should be noted that this approach does not reflect statistical significance; the standard error of prediction, a paired-samples *t*-test, or similar statistical technique would need to be used to determine significant change. Rather, this approach is used to describe and better understand change in scores either in the absence of, or following, a statistical analysis.

To use this approach, first calculate the *T*-score point difference between pre and posttest scores for each child within the group of interest (e.g., classroom, school, etc.). Then, apply Cohen's guidelines to label

the size of change for each child. Finally, count the number of children falling within each change category (i.e., no change, small positive change, etc.) to determine how many children within the group are showing positive, negative, or no change. By taking it a step further, you can also explore how many children are displaying small, medium, or large change.

In order to assist users in using this approach, the Devereux Center for Resilient Children has created an Excel template that will automatically calculate change scores and provide the results in a table and graphical display. This template and instructions for using the template are available for download from our website at this [link](#).

### *Summary*

The approach based on Cohen's *d*-ratio is useful to explore the size of pre-post changes in groups of children to determine whether specific children (or smaller groups of children) are responding differently to an SEL program or intervention. Although this approach does not allow an administrator to determine whether statistically significance change occurred, it does provide a way to examine outcomes in the absence of statistical software. It can also provide a way to dive deeper into findings following a paired-samples *t*-test.

### *Examining outcomes for children who began in the need for instruction range*

The final approach in this section will focus on the subset of children within a group who began the school year with social and emotional needs. These children, who received *T*-scores of 40 or below on one or more of the DESSA scales or SEC, often receive more intentional social and emotional interventions or supports in addition to a universal SEL program. Often, schools and programs will follow these children with the progress monitoring method described earlier in this guide in order to track their progress and inform the interventions they are receiving. In addition to this progress monitoring, it is often useful to examine the overall outcomes for this subgroup of children at the end of a school year.

The most straightforward way to examine DESSA outcomes for children who began the school year in the need for instruction range would be to simply add up

the number of children who were moved out of the need range and into either the typical or strength range across the program year. It is often a powerful statement to be able to say, for example, that 200 of the 250 children within a program (or 80%) who started the year in the need range were moved out of that range and into the typical or strength ranges. Given the relationship between low social and emotional competencies and school and life outcomes, this would be an important and meaningful outcome to highlight.

### *Summary*

In this section, we've highlighted three approaches for examining outcomes for groups of children. Each of these approaches has strengths and limitations, and although each can be used on its own, the combination of the three often lead to the most complete, useful, and powerful message.

## **PROGRAM EVALUATION AND QUALITY IMPROVEMENT**

In this final section, the use of the DESSA for evaluating programs and quality improvement purposes will be discussed. As noted above, the DESSA can be used to examine the effectiveness of SEL programs or intervention strategies. This can be done by examining the changes in children's DESSA scores before and after intervention.

However, two important points need to be considered when evaluating interventions, as recommended by Jacobsen and Truax (1991). These include determining both that a statistically reliable change *and* a clinically (or educationally) meaningful change has occurred. Statistically reliable change is addressed through the pretest-posttest comparison described in the "Measuring Outcomes for Individual Children" section. This will allow you to confidently state that the obtained differences are not simply chance findings.

Once a statistically reliable change is found, you can then determine the meaningfulness of that change. This is important, because just knowing a significant change exists does not tell you about whether that change actually makes a difference to the children, teachers, parents, or school. Ultimately, the best possible outcome for a child would be to have strengths within all eight DESSA scales. The worst outcome for a child

would be if all DESSA scores fell within the need for instruction range. By examining this in combination with statistically reliable change, you can better understand the impact an intervention or program had on the children.

After determining that the change is significant, determine whether the change is meaningful, by examining the value of the posttest *T*-scores. With the DESSA, three clinically meaningful changes could occur. An *optimal* outcome would occur when a child's posttest *T*-score is within the strength range. A *favorable* outcome would occur when a child's posttest *T*-score is within the typical range. Finally, a *negative* outcome would occur when a child's posttest *T*-score is within the need for instruction range.

The advantage of this approach is that you can look at the effectiveness of interventions for each DESSA scale on a child-by-child basis. This allows you to determine which children benefited from which interventions in which areas. On a broader scale, you could also aggregate findings across children, classrooms, schools, etc. for program evaluation or quality improvement. For example, suppose this approach leads to the conclusion that district-wide, students significantly improved in Social Awareness, but not Relationship Skills. An administrator might then make the decision to provide teachers with additional professional development in Relationship Skills strategies. It could also lead to the decision for a school or program-wide focus on Relationship Skills for the following year.

## CONCLUSIONS

It is our hope that this guide provides both broad recommendations and specific techniques for evaluating outcomes for individual children and groups of children using the DESSA. Although only a few of the many approaches for examining outcomes were discussed, we have found these approaches to be particularly useful for better understanding children's outcomes and using that information to better inform and guide the SEL programming and interventions being used in school and out-of-school time settings. For more information about the approaches discussed, please contact Jennifer Fleming at [jfleming2@devereux.org](mailto:jfleming2@devereux.org).

## REFERENCES

- Atkinson, L. (1991). Three standard errors of measurement and the Wechsler Memory Scale-Revised. *Psychological Assessment*, 3, 136-138
- Cohen, J. (1988). Statistical power analysis in the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.
- Jacobsen, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- LeBuffe, P. A., Shapiro, V. B., & Naglieri, J. A. (2009, 2014). The Devereux Student Strengths Assessment (DESSA). Charlotte, NC: Apperson.