13 The Assessment of Executive Functioning Using the Delis-Kaplan Executive Functions System (D-KEFS)

Tammy L. Stephens

The *Delis-Kaplan Executive Function System* (D-KEFS) is the first nationally standardized set of tests to evaluate higher-cognitive functions (executive functions) in both children and adults (ages 8–89 years). Published in 2001, the D-KEFS includes nine stand-alone tests that measure a wide array of verbal and nonverbal executive functions.

Theoretical Approach

A single theoretical model was not utilized within the development of the D-KEFS. Instead, the authors and developers incorporated a cognitive-process approach in the design of the tests in order to ensure that both the fundamental and higher-level components of executive functions could be quantified (Delis, Kaplan, & Kramer, 2001a, 2001b).

Constructs

- *Executive Functions* (*EF*) have been defined as mental functions associated with the ability to engage in purposeful, organized, self-regulated, and goal-directed behaviors (McCloskey, Perkins, & Divner, 2009) and integrate, synthesize, and organize other cognitive processes (Wecker, Kramer, Wisniewski, Delis, & Kaplan, 2000). EF permit a person to perform certain higher-order cognitive tasks that enable academic achievement.
- *Planning* involves the setting of short- or long-term goals and the establishment of a behavioral routine (strategy) to accomplish the set goals (McCloskey et al., 2009).
- *Inhibition* has been defined as the ability to resist or suppress urges to perceive, feel, think, or act on first impulse (McCloskey et al., 2009).
- *Attention* is a basic cognitive skill that allows for successful completion of set goals. The primary aspect of attention is the ability to focus on and respond to stimuli in the environment (Dehn, 2006).
- *Perception* involves the use of sensory and perception processes that take information in from the external environment or "inner awareness" to tune into perceptions, emotions, thoughts, or actions as they are occurring (McCloskey et al., 2009).
- *Switching* refers to a change of focus or alteration of perceptions, emotions, thoughts, or actions in reaction to what is occurring in the internal and external environments (McCloskey et al., 2009).

Description of the D-KEFS

Subtest Background and Structure

The D-KEFS consists of nine tests that measure verbal and nonverbal executive functions. Each test is designed as a stand-alone instrument that

T.L. Stephens (✉)
Pearson Clinical Assessments, San Antonio, TX, USA
e-mail: Tammy.Stephens@Pearson.com

S. Goldstein and J.A. Naglieri (eds.), *Handbook of Executive Functioning*,
DOI 10.1007/978-1-4614-8106-5_13,

can be administered individually or along with other D-KEFS tests, depending on the referral question. The D-KEFS is composed of the following instruments.

D-KEFS Trail Making Test

D-KEFS Trail Making Test consists of two tasks, (1) visual cancellation and (2) a series of connect-the-circle tasks. The test includes five conditions with *Condition 4*: *Number-Letter Switching* measuring the primary executive-function task, a means of assessing flexibility of thinking on a visual-motor sequencing task. The other four conditions allow the examiner to quantify and derive normative data for several key components necessary for performing the switching task. The five conditions are as follows:

- *Condition 1*: *Visual Scanning*
  - Task: Locate and slash through all the 3s on the page
- *Condition 2*: *Number Sequencing*
  - Task: Connect numbers in numerical order (e.g., 1-2-3-4)
- *Condition 3*: *Letter Sequencing*
  - Task: Connect letters in alphabetical order (e.g., A-B-C-D)
- *Condition 4*: *Number-Letter Sequencing*
  - Task: Switch between connecting the numbers and letters alternating between numerical order and alphabetical order (e.g., 1-A-2-B)
- *Condition 5*: *Motor Speed*
  - Task: Trace dotted lines from "Start" to "End"

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Trail Making Test*: scaled scores ($M = 10$; $SD = 3$) based on completion times and one error measure, a composite scaled score, contrast scaled score, and cumulative percentile ranks for most of the error measures. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual*. A brief description of each follows:

- *Completion-Time Scores*:
  - The primary scoring measure for each of the five conditions of the D-KEFS Trail Making Test is the number of seconds that the examinee takes to complete each condition. The raw score (in seconds) for each of the Conditions 1–5 is converted to a scaled score ($M = 10$; $SD = 3$). Additional combined scores can also be derived.
- *Contrast Measures*:
  - This measure allows the examiner to determine whether an impairment exists in an underlying component skill resulting in poor performance on Condition 4: Number-Letter Switching. For this reason, performance on each of the four baseline tasks is parceled out from performance on the Number-Letter Switching condition by the computation of a series of contrast measures. The contrast measures are derived by subtracting the completion-time scaled score for each component task (Conditions 1, 2, 3, or 5) or the Number Sequencing+Letter Sequencing composite from the completion-time scaled score on the switching task (Condition 4). A new scaled score, with a mean of 10 and a standard deviation of 3, is derived for each scaled score difference (Delis et al., 2001a, 2001b, p. 47).
- *Optional Error Scores*:
  - Cumulative percentile ranks can be derived for several types of errors. An in-depth description of each type of error can be found within the D-KEFS Examiner's Manual (Delis et al., 2001a, 2001b, p. 48). A brief description of each follows:

    Omission Error occurs whenever an examinee fails to mark a target 3 for Condition 1: Visual Scanning.

    Commission Error occurs whenever an examinee marks a letter or a number that is not a 3.

    Sequencing Error occurs when an examinee makes a connection within the correct set of symbols for the condition being administered (numbers or letters) but connects the wrong item within that set.

Set-Loss Error occurs when an examinee draws a line connecting an item that belongs to the wrong set of symbols (numbers or letters) for the condition being administered.

Time-Discontinue Error occurs when an examinee failed to connect one or more items because the time limit for that condition had elapsed (Delis et al., 2001a, 2001b).

D-KEFS Verbal Fluency Test

The *D-KEFS Verbal Fluency Test* evaluates an individual's ability to generate words fluently in a phonemic format (Letter Fluency), from overlearned concepts (Category Fluency), and while simultaneously shifting between overlearned concepts (Category Switching). The *D-KEFS Verbal Fluency Test* consists of a standard form and an alternate form. The three testing conditions for the *D-KEFS Verbal Fluency Test* consist of the following:

- *Condition 1*: *Letter Fluency* consists of three 60-s timed trials.
  - Trial 1: Examiner names as many words as possible that begin with letter F (B).
  - Trial 2: Examiner names as many words as possible that begin with letter A (H).
  - Trial 3: Examiner names as many words as possible that begin with letter S (R).
- *Category Fluency* consists of two 60-s timed trials
  - Trial 1: Examiner names as many animals (clothing) as possible in a 60-s time limit.
  - Trial 2: Examiner names as many boys' names (girls' names) as possible in a 60-s time limit.
- *Category Switching* consists of one 60-s timed trial.
  - Trial: Examiner names as many vegetables and musical instruments while switching back and forth.

A 60-s time limit for each trial of each condition is allowed (Delis et al., 2001a, 2001b).

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Verbal Fluency Test*: Total Correct Scores, Primary Contrast Scores, and Optional Error Scores. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual* (Delis et al., 2001a, 2001b). A brief description of each follows:

- *Total Correct Scores*:
  - This measure represents the number of correct words generated within each 60-s trial for each condition. Consult the D-KEFS Examiner's Manual for guidelines for scoring correct responses (Delis et al., 2001a, 2001b).
- *Primary Contrast Scores*:
  - These measures allow the examiner to compare the examinee's performance on one task in comparison (e.g., Letter Fluency) to another task (e.g., Category Fluency) to determine whether the examinee exhibits a disproportionate impairment in one relative to the other.
- *Optional Error Scores*:
  - Numerous optional measures can be derived for the *D-KEFS Verbal Fluency Test* (e.g., three letter-fluency trials, two category-fluency trials, and one witching trial). Refer to the D-KEFS Examiner's Manual for in-depth descriptions (Delis et al., 2001a, 2001b).

D-KEFS Design Fluency Test

The *D-KEFS Design Fluency Test* evaluates an examinee's ability to draw as many different designs as possible in a 60-s time limit. The examinee is presented with a record form containing rows of boxes, with each containing an array of dots and instructed to draw a different design in each box using only four lines to connect the dots. The *D-KEFS Design Fluency Test* consists of the following three conditions:

- *Condition 1*: *Filled Dots*
  - Task: Examinee is asked to draw the designs connecting filled dots.
  - Assesses basic design fluency.

- *Condition 2*: *Empty Dots Only*
- *Condition 3*: *Switching*
  - Task: Examinee is asked to draw designs by alternately connecting filled and empty dots.
  - Assesses design fluency and cognitive flexibility.

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Design Fluency Test*: Primary Scores, Contrast Scores, and Optional Error Scores. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual*.

- *Primary Scores*:
  - This measure represents the total number of correct designs generated in each of the three 60-s individual conditions. Consult the D-KEFS Examiner's Manual for guidelines for scoring correct responses (Delis et al., 2001a, 2001b).
- *Contrast Scores*:
  - Some examinees exhibit adequate design fluency skills, except when they must simultaneously engage in cognitive shifting to generate the designs in the switching condition. Consequently, a contrast score is derived for directly comparing the examinee's ability to generate designs with switching (Condition 3) relative to his or her ability to generate designs without switching (composite scaled score for Conditions 1 and 2). Refer to the D-KEFS examiner's Manual for additional information regarding Contrast Scores (Delis et al., 2001a, 2001b).
- *Optional Error Scores*:
  - Numerous process and error measures can be derived for the *D-KEFS Design Fluency Test*. Refer to the D-KEFS Examiner's Manual for in-depth descriptions (Delis et al., 2001a, 2001b).

D-KEFS Color-Word Interference Test

The *D-KEFS Color-Word Interference Test* evaluates the examinee's ability to inhibit an overlearned verbal response (i.e., reading the printed words) in an attempt to generate a conflicting response naming the dissonant ink colors in which the words are printed. The D-KEFS Color-Word Interference Test is based on the procedure. The test consists of the following three conditions:

- Condition 1: Basic naming of color patches.
- Condition 2: Basic reading of color words printed in black ink.
- Condition 3: (Inhibition); examinee must inhibit reading the words in order to name the dissonant ink colors in which those words are printed.
- Condition 4: (Inhibition/Switching/Cognitive Flexibility); examinee must switch back and forth between naming the dissonant ink colors and reading the words.

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Color-Word Interference Test*: Completion-Time Scores, Contrast Scores, and Error Scores. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual*.

- *Completion-Time Scores*:
  - This measure is the primary method used to analyze performance on the D-KEFS Color-Word Interference Test and is based on the number of seconds that the examinee takes to complete each of the four conditions (Delis et al., 2001a, 2001b).
- *Contrast Scores*:
  - This measure allows the examiner to distinguish between a deficit in higher-level abilities of inhibition or cognitive flexibility

and impairments in the fundamental skills of basic naming and reading. Refer to the D-KEFS Examiner's Manual for in-depth descriptions (Delis et al., 2001a, 2001b).

- *Error Scores*:
  - This measure categorizes errors for each of the four conditions of the D-KEFS Color-Word Interference Test. Naming Errors, Reading Errors, Inhibition Errors, and Inhibition/Switching Errors can be analyzed. Refer to the D-KEFS Examiner's Manual for in-depth descriptions (Delis et al., 2001a, 2001b).

D-KEFS Sorting Test

The *D-KEFS Sorting Test* evaluates the individual's ability to initiate problem-solving behavior in verbal and nonverbal modalities. The D-KEFS Sorting Test consists of two conditions: Free Sorting and Sort Recognition.

For Condition 1: Free Sorting—the examinee is presented six mixed-up cards that display both perceptual features and printed words. The examinee is asked to sort the cards into two groups, with three cards per group. The sorts are made according to many different concepts or rules as possible, and upon completion of the sort, the examinee must describe the concepts employed to generate each sort. Each of the two card sets has a maximum of eight target sorts: three sorts based on verbal-semantic information from the printed words and five based on visual-spatial features or patterns on each card (Delis et al., 2001a, 2001b).

For Condition 2: Sort Recognition—the examiner sorts the same sets of cards into two groups (three cards in each group) according to the eight target sorts. After the examiner completes is each sort, the examinee instructed to identify and describe the correct rules or concepts used to generate the sort. Corrective feedback is never given during administration of this test, to minimize possible adverse effects of repetitive negative feedback. Additionally, the examinee's problem-solving performance is scored in terms of both accuracy of the sorting responses and the descriptions of the sorting concepts. Finally, the formal assessment of examinees' descriptions of the sorting rules provides information about their conceptual reasoning skills.

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Sorting Test*: Sorting and Descriptive Measures, Composite Scores and Contrast Scaled Scores. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual*.

- *Sorting and Descriptive Measures*, raw scores are converted into scaled scores or cumulative percentile ranks.
- *Composite Scores* for combined description measures (e.g., Condition 1: Free Sorting and Condition 2: Sort Recognition) can also be derived.
- *Contrast Scaled Scores*, a percent accuracy score, and a percent description accuracy score can be determined (Delis et al., 2001a, 2001b, p. 115).

D-KEFS Twenty Questions Test

The *D-KEFS Twenty Questions Test* evaluates the examinee's ability to identify the various categories and subcategories represented in the 30 objects and to formulate abstract, yes/no questions that eliminate the maximum number of objects regardless of the examiner's answer. The task involves the examinee being presented with a stimulus page depicting pictures of 30 common objects. The examinee is instructed to ask the fewest number of questions that result in the elimination of the most objects (Delis et al., 2001a, 2001b).

Executive functions assessed by the *D-KEFS Twenty Questions Test* include:

- The ability to perceive the various categories and subcategories represented by the 30 objects

- The ability to formulate abstract, yes/no questions that eliminate the maximum number of objects regardless of the examiner's answer
- The ability to incorporate the examiner's feedback in order to formulate more efficient yes/no questions

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Twenty Questions Test*: an Initial Abstraction Score, Total Questions Asked Score, and Total Weighted Achievement Score can be derived for the Primary Measure. Additionally, three optional process and error scores can also be derived: spatial questions, set-loss questions, and repeated questions. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual* (Delis et al., 2001a, 2001b). A brief summary of each score follows:

- *Initial Abstraction Score*:
  - Quantifies the level of abstract thinking represented by the first question asked by an examinee on each item. The minimum number of objects eliminated by the first question is summed across items 1–4 to obtain a raw score. The raw score is then converted to a scaled score.
- *Total Questions Asked Score*:
  - The fewer yes/no questions an examinee asks, the better is his/her performance on the *D-KEFS Twenty Questions Test*. The Total Questions Asked Score quantifies this aspect of performance and serves as a global achievement measure. Further, this variable is based on the number of yes/no questions asked until the target object is identified for each item, and these scores are summed across the four items on the test to obtain a raw score. The raw score is then converted to a scaled score.
- *Total Weighted Achievement Score*:
  - This measure was developed specifically to account for those individuals who fortuitously arrive at the correct answer after asking only one or two highly concrete questions. A total weighted achievement raw score is obtained by summing the raw scores for the four items. The raw score is converted to a scaled score.
- *Spatial Questions*:
  - This measure reflects the number of yes/no questions asked that attempt to eliminate objects based on their location on the stimulus page (e.g., "Is it in the top left side of the page?"). Only 2.7% of the normative sample asked spatial questions. The total raw score is summed across all four items and is transformed into a cumulative percentile rank.
- *Repeated and Set-Loss Questions*:
  - These two error measures reflect the number of repetitive and set-loss questions, respectively, summed across all four items of the test. The total raw scores for repetition errors and set-loss questions are transformed into cumulative percentile ranks.

D-KEFS Word Context Test

The *D-KEFS Word Context Test* evaluates executive functions in the verbal modality and assesses skills such as deductive reasoning, integration of multiple bits of information, hypothesis testing, and flexibility of thinking. Specifically, the test measures the examinee's ability to discover the meaning of made-up words or mystery words based on clues given in sentences. For each mystery word, the examinee is shown five sentences (clues) that help him or her to decode the meaning of the word. With each new clue sentence for the word, previously presented sentences are also displayed. The first few sentences for each word provide vague clues about the mystery word's meaning; clues become more detailed in subsequent sentences. The examinee's task is to decode the mystery word with as few clue sentences as possible while continuing to report the correct response to the remaining clue sentences of each item (Delis et al., 2001a, 2001b).

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Word Context Test*: Total Consecutively Correct, Consistently Correct Ratio, Repeated Incorrect Responses, No/Don't Know Responses, and Correct-to-Incorrect Errors. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual* (Delis et al., 2001a, 2001b). A brief summary of each score follows:

- *Total Consecutively Correct*:
  - The primary achievement measure for the *D-KEFS Word Context Test*. A raw score is calculated by totaling the *Total Consecutively Correct Score* across items 1–10. The raw score is then converted to a scaled score.
- *Consistently Correct Ratio*:
  - This measure is computed by dividing the Total Consecutively Correct Raw Score/First Sentence Correct Raw Score and multiplying it by 100. The denominator (First Sentence Correct Raw Score) reflects the first sentence in which the examinee provides a correct response. The raw score obtained for this measure is converted to a scaled score.
- *Repeated Incorrect*:
  - This measure reflects the number of incorrect responses that are repeated within the same item, summed across the ten items of the test. The raw score is converted to a scaled score.
- *No/Don't Know Responses*:
  - This measure is the number of clue sentences to which the examinee provides either no response or "don't know" responses, after being prompted by the examiner to take a use, summed across the ten items of the test. The raw score is converted to a scaled score.
- *Correct-to-Incorrect Errors*:
  - The number of times an examinee provides a correct response on an early clue sentence and then lose set and report an incorrect response for the very next clue sentence presented
- *Repeated Incorrect Responses*:
  - This measure reflects the number of incorrect responses that are repeated within the same item, summed across the ten items of the test. A raw score is calculated and converted to a scaled score.
- *No/Don't Know Responses*:
  - This measure is the number of clue sentences to which the examinee provides either no response or "don't know" responses, after being prompted by the examiner to take a guess. This measure is summed across one item of the test to obtain a raw score. The raw score is converted to a scaled score.
- *Correct-to-Incorrect Errors*:
  - This measure reflects the number of times the examinee provides a correct response on an early clue sentence and then loses set and reports and incorrect response for the very next clue sentence presented. The raw score is transformed into a cumulative percentile rank.

D-KEFS Tower Test

The *D-KEFS Tower Test* evaluates the examinee's ability to move disks varying in size from small to large across three pegs to build a designated tower in the fewest number of moves possible (Delis et al., 2001a, 2001b). When completing the task, the examinee must follow the following rules:

1. Move only one disk at a time.
2. Never place a larger disk over a smaller disk.

The *D-KEFS Tower Test* assesses several key executive functions, including spatial planning, rule learning, inhibition of impulsive and perseverative responding, and the ability to establish and maintain instructional set.

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Tower Test*: Total Achievement Score and five optional process scores (Mean First-Move

Time, Time-Per-Move Ratio, Move Accuracy Ratio, Total Rule Violations, and Rule-Violations-Per-Item Ratio). An in-depth description of each score can be found within the *D-KEFS Examiner's Manual* (Delis et al., 2001a, 2001b). A brief summary of each score follows:

- *Total Achievement Score*:
  - This score reflects the sum of the achievement scores, including bonus points, for all items administered. The raw score is converted to a scaled score.
- *Mean First-Move Time*:
  - This score reflects the average of the examinee's first-move times. Specifically, the score is computed by taking the sum of the examinee's first-move times (in seconds) for all the items administered divided by the number of items administered. A raw score is converted into a scaled score.
- *Time-Per-Move Ratio*:
  - This measure reflects the average time the examinee takes to make each of his or her moves. The score is computed by summing the completion times for all items administered and divided by the total number of moves made for all items administered. The raw score is then converted into a scaled score.
- *Move Accuracy Ratio*:
  - This measure reflects the efficiency with which the examinee constructed the towers. The total number of moves used by the examinee across all items administered is divided by the fewest number of moves required across all items administered. The raw score is converted to a scaled score.
- *Total Rule Violations*:
  - This measure represents the total number of rule violations committed by the examinee across all items administered. The two rule violations of the test include moving more than one disk at a time and placing a larger disk on a smaller disk. The raw score is converted into a cumulative percentile rank.
- *Rule-Violations-Per-Item Ratio*:
  - This measure reflects the average number of rule violations made by the examinee relative to the number of items administered. The obtained raw score is converted to a scaled score.

D-KEFS Proverb Test

The *D-KEFS Proverb Test* qualitatively evaluates the nature of an individual's verbal abstraction skills (Delis et al., 2001a, 2001b). The D-KEFS Proverb Test consists of eight sayings that are presented in two conditions:

- *Condition 1*: *Free Inquiry*
  - Proverbs are read individually to the examinees, who attempt to interpret them orally without assistance or cues.
- *Condition 2*: *Multiple Choice*
  - The same eight proverbs are presented individually along with four alternative interpretations from which the examinee must select the best one.
  - The set of multiple-choice response alternatives for each proverb consists of:
    A correct abstract interpretation
    A correct concrete interpretation
    An incorrect, phonemically similar response
    An unrelated saying

Scales the Test Yields

Several types of scores are derived from the *D-KEFS Proverb Test*: normative data are provided for seven measures for the Free Inquiry condition and six variables for the Multiple Choice condition. An in-depth description of each score can be found within the *D-KEFS Examiner's Manual*. A brief summary of each score follows:

Scores for Free Inquiry Condition

- *Total Achievement Score*:
  - The primary measure for the Free Inquiry condition of the D-KEFS Proverb Test. The raw score is based on the sum of the individual achievement scores of all eight items. The raw score is converted to a scaled score.
- *Common Proverb Achievement Score*:
  - This measure is based on the examinee's performance on the first five items on the D-KEFS Proverb Test, which consist of high-frequency sayings with which most people are likely to be familiar. The raw score is the sum of the raw achievement scores for the first five items. The raw score is then converted to a scaled score.

- *Uncommon Proverb Achievement Score*:
  - This measure is based on the examinee's performance on the last three items on the D-KEFS Proverb Test, which consist of low-frequency sayings with which most people are less likely to have heard. The raw score is the sum of the achievement scores for the last three items. The raw score is converted to a scaled score.
- *Accuracy Only Score*:
  - This measure reflects the extent to which examinees can provide accurate interpretations of the proverbs regardless of whether their interpretations are abstract or concrete. The raw score is the sum of the accuracy scores for the eight proverbs; the raw score is converted to a scaled score.
- *Abstraction Only Score*:
  - This measure reflects the degree to which examinees provide abstract responses to the proverbs regardless of the degree of accuracy of their interpretations. The raw score is the sum of the abstraction scores for the eight proverbs. The raw score is converted to a scaled score.
- *No/Don't Know Response and Repeated Responses*:
  - This measure reflects the frequency with which an examinee makes these error types across the eight proverbs of the Free Inquiry condition. The raw score for the no/don't know and repeated responses measures is converted into cumulative percentile ranks.

Scores for Multiple-Choice Condition

- *Total Achievement Score*:
  - This measure reflects the sum of an examinee's item achievement scores on the eight items. The raw score is converted into a cumulative percentile rank.
- *Common and Uncommon Proverb Achievement Scores*:
  - The raw score for the common proverb achievement index is the sum of an examinee's item achievement scores on the first five items of the Multiple Choice condition. Raw scores are transformed into a cumulative percentile rank.
- *Endorsement Measures*:
  - This measure reflects the number of times the examinee endorses each type of alternative response (correct abstract, correct concrete, incorrect phonemic, or incorrect unrelated) summed across the items and transformed into a cumulative percentile rank (Delis,).

Administration and Scoring

Tips on Administration

The D-KEFS tests are cognitive assessment instruments, and therefore examiners must have formal training and experience in the assessment of intellectual and cognitive functions (Delis et al., 2001a, 2001b). With one exception, the D-KEFS tests were designed for use with both children and adults (ages 8–89). The D-KEFS Proverb Test was designed for adolescents and adults (ages 16–89).

Examiners should become familiar with the standardized administration procedures for each of the nine tests that make up the D-KEFS. Each of the nine D-KEFS tests stands alone; therefore, the examiner should pick the tests that will best provide the information needed to answer the referral question.

Scoring the Tests

Scaled scores, cumulative percentile ranks, contrast measures, combined scaled scores, and various optional scores can be derived from each test. For most of the measures provided by the D-KEFS tests, the raw scores are converted to scaled scores, with a mean of 10 and a standard deviation of 3. An in-depth description of scoring procedures for each test can be found within the *D-KEFS Examiner's Manual* (Delis et al., 2001a, 2001b).

Use of Scoring Software

The *D-KEFS Scoring Assistant* software automatically computes the standardized scores for both the primary and optional measures of the standard and alternate forms of the D-KEFS and prints them in a report format. Thus, the scoring software greatly enhances the efficiency with which the D-KEFS can be used in clinical practice (p. 29, manual).

Standardization, Norms, and Psychometrics

Characteristics of the Standardization Sample

The D-KEFS was standardized on a nationally representative, stratified sample of 1,750 children, adolescents, and adults, ages 8–89 years. Stratification was based on age, sex, race/ethnicity, years of education, and geographic region. The 2000 US Census figures were used as target values for the composition of the D-KEFS normative sample (Delis et al., 2001a, 2001b, p. 1).

Sixteen age groups make up the D-KEFS normative sample: 8 years, 9 years, 10 years, 11 years, 12 years, 13 years, 14 years, 15 years, 16–19 years, 20–29 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, 70–79 years, and 80–89 years. The D-KEFS sample consisted roughly of equal proportions of men and women at each age group, with the exception of the older age groups, which had more women than men, which is consistent with the census data (Delis et al., 2001a, 2001b, pp. 1–2).

The proportion of African-American, Hispanic, white, and other racial/ethnic groups sampled were stratified to approximate the 2000 US Census population estimates. Additionally, the D-KEFS sample was divided into the five major educational groups used by the US Census: less than or equal to 8 years of education, 9–11 years, 12 years, 13–15 years, and 16 years or more. Finally, the United States was divided into four major geographical areas as specified by the US Census data: northeast, north central, south, and west. All regions are well represented in the normative sample (Delis et al., 2001a, 2001b, p. 2). A more in-depth description of the normative sample can be found in the *D-KEFS Technical Manual*.

Reliability of the Scales

Reliability indicates the consistency of measurements. Consistency has several meanings, consistent within itself (internal reliability), consistent over time (test-retest reliability), and consistent with an alternate form of the measure (alternate-form reliability) (Sattler, 2008). The psychometric properties of internal consistency, stability coefficients, and alternate-form reliability were determined for the D-KEFS instruments. These measures of reliability provide a basis for deriving the standard error of measurement and confidence intervals. An in-depth synopsis of all reliability measures can be found within the *D-KEFS Technical Manual* (Delis et al., 2001a, 2001b). A brief description of each reliability scale follows:

Internal Consistency: It assumes that all items measure the same trait or construct. It is established by dividing the test into two equivalent halves (split-half reliability). This division creates two alternative forms of the test. The most common way of dividing the test is to assign odd-numbered items to one form and even-numbered items to the other (Sattler, 2008). Internal Consistency for the nine tests that make up the D-KEFS instrument is as follows:

- *D-KEFS Trail Making Test* (Combined Number and Letter Sequencing Composite Score) ranges from .60 to .81
- *D-KEFS Verbal Fluency Test* by age group and per conditions 1–3:
  - Condition 1: .68–.90
  - Condition 2: .53–.76
  - Condition 3 (Total Correct): .37–.68
  - Condition 3 (Total Switching): .51–.76
- *D-KEFS Design Fluency Test*: Not reported due to time constraints

- *D-KEFS Color-Word Interference Test* (Combined Color Naming and Word Reading Composite Score) ranges from .62 to .86
- *D-KEFS Sorting Test* by age group and per conditions 1–3:
  - Condition 1: .55–.86
  - Condition 2: .55–.84
  - Condition 3: .62–.81
- *D-KEFS Twenty Questions Test*: Internal Consistency for this test is reported based on age group and by Initial Abstraction and Total Weighted Achievement
  - Initial Abstraction: Based on age group, Internal Consistency ranges from .72 to .87
  - Total Weighted Achievement: Based on age group, Internal Consistency ranges from .10 to .55
- *D-KEFS Word Context Test*: Internal Consistency for this test is reported based on age group and Total Consecutively Correct
  - Total Consecutively Correct: Based on age group, Internal Consistency for this test ranges from .47 to .74
- *D-KEFS Tower Test*: Internal Consistency for this test is reported based on age group and Total Achievement
  - Total Achievement: Based on age group, internal consistency for this test ranges from .43 to .84
- *D-KEFS Proverb Test*: Internal Consistency on this test is reported based on age group and Total Achievement: Free Inquiry
  - Total Achievement: Free Inquiry: Based on age group, internal consistency for this test ranges from .68 to .81

Additional forms of reliability were also established for the nine D-KEFS tests; these included Alternate-Form Reliability and Test-Retest Reliability. *Alternate-Form Reliability* is the equivalent or parallel form reliability determined by administering two equivalent tests to the same group of examinees (Sattler, 2008). Specific correlation coefficients can be found within *Chapter 2 of the D-KEFS Technical Manual* (Delis et al., 2001a, 2001b). Moderate to high correlations coefficients were found.

Test-Retest is an index of stability a measure of how consistent scores are over time (Sattler, 2008). Test-Retest Reliability coefficients for the nine tests that make up the D-KEFS instrument can be found within *Chapter 2 of the D-KEFS Technical Manual* (Delis et al., 2001a, 2001b). Test-retest correlations range from moderate to high across the D-KEFS tests.

Use of the Test

Interpretation Methods

The D-KEFS variables measure different aspects of test performance, to include accuracy of responses, error rates, and response times. Most of the D-KEFS measures provide scaled scores. The directionality of the scaled scores is used to interpret performance; specifically, the higher the scaled score, the better the performance. This rule pertains to measures reflecting (1) accuracy scores, (2) error rates (e.g., the more errors generated, the lower the scaled score), or (3) completion times (e.g., the slower the time to solve an item or complete a condition, the lower the scaled score). Moreover, there are two types of measures in which either low- or high-scaled scores reflect different types of cognitive problems (Delis et al., 2001a, 2001b).

Contrast measures may signal cognitive difficulties if the scaled score is either too low or too high. For example, on the D-KEFS Verbal Fluency Test, one of the contrast measures is letter fluency versus category fluency. For this measure, a contrast scaled score of 7 or lower may reflect greater difficulty with letter fluency than with category fluency. In contrast, a scaled score of 13 or higher may indicate greater difficulty with category fluency than with letter fluency (Delis et al., 2001a, 2001b).

Certain time variables that measure the examinee's latency to make a response might also reflect different types of cognitive problems, depending whether the scaled score is too high or too low. For example, on the D-KEFS Tower Test, the mean first-move time variable measures the average time the examinee takes to initiate the first move for each item administered. This measure

offers an estimate of the examinee's initial planning time before engaging in problem-solving behavior. According to Delis et al. (2001a, 2001b) examinees that are too slow or too quick in generating their first moves both may be demonstrating cognitive problems for different reasons (e.g., those with activation problems may take longer for initial response, while those with impulsive tendencies may respond too fast).

The directionality of the cumulative percentile ranks is used to interpret examinee performance. Cumulative percentile ranks on the D-KEFS were scaled to reflect the percentage of the normative sample that obtained raw scores that were equal to or worse than the raw score obtained by the examinee. A cumulative percentage rank of 10 % on an error measure indicates that 10 % of the normative sample made the same or more errors on that measure.

Various levels of interpretation may be used to interpret the process-oriented tests that make up the D-KEFS instruments. Specifically, performance can be interpreted at four general levels: interpretation of achievement measures, interpretation of process measures, integration of D-KEFS findings with results from the entire cognitive and motor test battery, and inferences regarding risk factors for cognitive difficulties. A brief description of each interpretation method follows; however, more in-depth interpretation can be found within *Chapter 2 of the D-KEFS Examiner's Manual* (Delis et al., 2001a, 2001b).

Interpretation of the D-KEFS tests or conditions through the global achievement measures allow the examiner to gauge an examinee's overall level of performance on the task. Therefore, achievement measures usually provide an initial level of interpretation in addressing whether or not an examinee generally performed well or poorly on the test (Delis et al., 2001a, 2001b). Additionally, some D-KEFS tests provide achievement measures for baseline conditions that isolate more fundamental cognitive or motor skills that are needed to perform the higher-level conditions of the test.

The D-KEFS results can also be interpreted through process measures. According to Delis et al. (2001a, 2001b), the main intention of the process approach to cognitive assessment is the use of multiple measures that are designed to isolate and quantify specific aspects of test performance. Consequently, these measures provide pertinent information necessary to quantify the examinee's performance in comparison to normative data for a wide range of cognitive abilities, which include problem-solving strategies, ratio measures, error types, and temporal aspects of responding. Finally, process measures allow the examiner to pinpoint key areas of deficit and thereby guide appropriate intervention selection.

The integration of the D-KEFS findings with results from the entire battery given to the examinee is of utmost importance. Specifically, the cognitive deficit found through the administration of the D-KEFS test(s) should be validated by the use of several test measures that are also designed primarily for evaluating that ability area or are highly dependent on that ability for successful performance. The determination of whether the examinee's overall profile of cognitive strengths and weaknesses is typical of those seen in specific patient populations or atypical for most clinical disorders should also be determined (Delis et al., 2001a, 2001b).

The final level of interpretation relates to the examiner's understanding of risk factors for cognitive difficulties. When interpreting low scores on the D-KEFS, the examiner should not automatically associate the low scores to brain damage. Instead, the examiner should consider other reasons for cognitive difficulties. Possible neurostructural etiologies of cognitive deficits include prenatal exposure to alcohol, severe birth trauma, traumatic brain injuries (TBIs), or environmental neurotoxic exposure. Additionally, the examiner must consider possible nonneurostructural factors, which include depression, anxiety, obsessive thoughts, pain symptoms, or sleep deprivation and fatigue (Delis et al., 2001a, 2001b). The examiner should be familiar with characteristics of such factors and consider them within the interpretation of test results.

Identification of Special Populations

Results of the D-KEFS test(s) should be integrated and interpreted in consideration with

results obtained from other cognitive and achievement tests. When determining eligibility for special education services under one of the 13 categories listed within the Individuals with Disability Education Act (IDEA, 2004), the D-KEFS results should be used to assist in the verification of the disability criteria. Specifically, when considering a student as having a Learning Disability (LD), evaluation personnel should use a variety of assessment results (to include the D-KEFS) to investigate patterns of strengths and weaknesses in cognitive and or achievement (IDEA, 2004). Further, by investigating the examinee's performance in the areas of higher-level executive functioning, the D-KEFS results can be utilized to verify characteristics of Attention Deficit Disorder (ADD) and/or Attention Deficit Hyperactivity Disorder (ADHD) and TBI. In-depth clinical interpretation of scores can be found within the *D-KEFS Examiner's Manual* for each of the nine tests (Delis et al., 2001a, 2001b).

Interventions Based on Test Results

Results of the examinee's performance on any of the nine D-KEFS tests can be used to guide intervention selection. Each of the nine individual D-KEFS tests targets low- and high-level executive functions, allowing for the identification of specific areas of weakness. Consequently, an in-depth clinical analysis of scores should be used to pinpoint specific areas of deficit resulting in the appropriate identification and selection of intervention practices.

Validity

Relationships to Other Similar Measures

The validity of the D-KEFS tests (e.g., the Stroop procedure, Trail Making Test, verbal and design fluency tests, tower tests, twenty questions procedure, and proverb interpretations) has been demonstrated in numerous neuropsychological studies conducted over the past 50 years or more. The evidence of validity has been provided in terms of sensitivity of the tests in the detection of brain damage, specifically in the frontal lobe area, and in the ability of the tests to measure areas of higher-level executive functions (Delis et al., 2001a, 2001b). Numerous studies have been conducted which investigated the validity of the nine D-KEFS tests in comparison to other cognitive measures, specifically the *California Verbal Learning Test-Second Edition* (CVLT-II; Delis, Kaplan, Kramer, & Ober, 2000) and the *Wisconsin Card Sorting Test* (WCST; Heaton, Chelune, Talley, Kay, & Curtiss, 1993).

The validity study between the CVLT-II and the D-KEFS consisted of 292 adults (33 % male and 63 % female). Standardized scores from the two instruments were used in the analyses. Correlations between the CVLT-II and the D-KEFS Sorting Test measures results found the CVLT-II immediate and delayed recall measures correlated in the low positive range with the key Sorting Test measures, including the confirmed correct sorts, free sorting description score, and sort recognition description score indices (Delis et al., 2001a, 2001b). Additionally, the vast majority of the correlations were not significant, indicating little overlap between the functions assessed by the two instruments.

A small pilot study was also conducted to investigate the relationship between the WCST and the D-KEFS tests. The sample size of the study consisted of 23 adults (65 % male and 35 % female). Correlations were run between the test scores obtained on the WCST and the D-KEFS tests. According to Delis et al. (2001a, 2001b), the number of categories completed on the WCST tended to have moderate correlations with several of the primary measures of the D-KEFS tests. Additionally, findings indicated that the perseverative responses measures of the WCST tended to correlate at somewhat lower levels with key D-KEFS measures. Complete results can be found in Table 3.59 of the D-KEFS Technical Manual (Delis et al., 2001a, 2001b).

Numerous studies have been conducted over the years investigating the use of the D-KEFS tests with individuals diagnosed with fetal alcohol syndrome (FAS), Alzheimer's disease,

and Huntington disease. Additionally, studies were conducted which analyzed the evidence of validity across the nine D-KEFS tests. Results for all the validity studies can be found within *Chapter 3 of the D-KEFS Technical Manual* (Delis et al., 2001a, 2001b).

Summary and Conclusions

The D-KEFS is the first nationally standardized set of tests to evaluate higher-cognitive functions (executive functions) in both children and adults. Published in 2001, the D-KEFS includes nine stand-alone tests that measure a wide array of verbal and nonverbal executive functions. The D-KEFS allows for the identification of areas of strengths and weaknesses in lower- and higher-level executive functions. Thus, results of each of the nine tests can be clinically interpreted and used with other assessment results in the validation of various disabilities (e.g., LD, ADD, ADHD, or TBI). Finally, an in-depth analysis of examinee performance can be used to pinpoint specific areas of weakness, making intervention selection more appropriate.

References

Dehn, M. (2006). *Essentials of processing assessment*. Hoboken, NJ: Wiley.

Delis, D.C., Kramer, J., Kaplan, E., & Ober, B. (2000). California Verbal Learning Test-Second Edition (CVLT-II). San Antonio, TX: The Psychological Corporation.

Delis, D., Kaplan, E., & Kramer, J. (2001a). *Delis Kaplan (D-KEFS) examiner's manual*. San Antonio, TX: NCS Pearson, Inc.

Delis, D., Kaplan, E., & Kramer, J. (2001b). *Delis Kaplan (D-KEFS) technical manual*. San Antonio, TX: NCS Pearson, Inc.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual—Revised and expanded*. Lutz, FL: Psychological Assessment Resource, Inc.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub. L. No. 102-446, 118 Stat. 2647 (2004).

McCloskey, G., Perkins, L., & Divner, B. (2009). *Assessment and intervention for executive function difficulties*. New York, NY: Routledge Taylor & Francis Group.

Sattler, J. M. (2008). *Assessment of children: Cognitive applications* (5th ed., pp. 102–104). La Mesa, CA: Jerome M. Sattler.

Wecker, N. S., Kramer, J. H., Wisniewski, A., Delis, D. C., & Kaplan, E. (2000). Age effects on executive ability. *Neuropsychology, 14*, 409–414.