


# The Other Side of the Bell Curve: Multivariate Base Rates of High Scores on the Delis-Kaplan Executive Function System

Justin E. Karr<sup>1,2,3,4,\*</sup> , Mauricio A. Garcia-Barrera<sup>5</sup>, James A. Holdnack<sup>6</sup> and Grant L. Iverson<sup>2,3,4</sup>

<sup>1</sup>Department of Psychiatry, Harvard Medical School, Boston, MA 02215, USA

<sup>2</sup>Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, MA 02129, USA

<sup>3</sup>Spaulding Rehabilitation Hospital, Spaulding Research Institute, Charlestown, MA 02129, USA

<sup>4</sup>Home Base, A Red Sox Foundation and Massachusetts General Hospital Program, Boston, MA 02129, USA

<sup>5</sup>Department of Psychology, University of Victoria, Victoria, British Columbia V8P 5C2, Canada

<sup>6</sup>Research and Statistics Consultant, Bear, DE 19701, USA

(RECEIVED February 11, 2019; FINAL REVISION August 4, 2019; ACCEPTED September 26, 2019; FIRST PUBLISHED ONLINE November 15, 2019)

## Abstract

**Objective:** Previous researchers have examined the frequency at which healthy participants obtain one or more low scores on neuropsychological test batteries, proposing five psychometric principles of multivariate base rates: (a) low scores are common, with their frequency contingent on (b) the low score cutoff used, (c) the number of tests administered/interpreted, and (d) the demographic characteristics and (e) intelligence of participants. The current study explored whether these principles applied to *high scores* as well, using the Delis-Kaplan Executive Function System (D-KEFS). **Method:** Multivariate base rates of high scores ( $\geq 75$ th,  $\geq 84$ th,  $\geq 91$ st,  $\geq 95$ th, and  $\geq 98$ th percentiles) were derived for a three-test, four-test, and full D-KEFS battery, using the adult portion of the normative sample (aged 16–89 years;  $N = 1050$ ) stratified by education and intelligence. The full D-KEFS battery provides 16 total achievement scores (primary indicators of executive function). **Results:** High scores occurred commonly for all batteries. For the three-test battery, 24.1% and 12.4% had 1 or more scores  $\geq 95$ th percentile and  $\geq 98$ th percentile, respectively. High scores occurred more often for longer batteries: 61.6%, 72.9%, and 87.8% obtained 1 or more scores  $\geq 84$ th percentile for the three-test, four-test, and full batteries, respectively. The frequency of high scores increased with more education and higher intelligence. **Conclusions:** The principles of multivariate base rates also applied to high D-KEFS scores: high scores were common and contingent on the cutoff used, number of tests administered/interpreted, and education/intelligence of examinees. Base rates of high scores may help clinicians identify true cognitive strengths and detect cognitive deficits in high functioning people.

**Keywords:** Executive function, Neuropsychological tests, Assessment, Norms/normative studies, Psychometrics, Base rates

## INTRODUCTION

A growing body of empirical research has examined the prevalence of low scores among normative samples of healthy examinees (Binder, Iverson, & Brooks, 2009; Brooks, Holdnack, & Iverson, 2011; Brooks, Iverson, & Holdnack, 2013; Brooks, Iverson, Lanting, Horton, & Reynolds, 2012; Brooks, Iverson, & White, 2009; Brooks, Sherman, & Iverson, 2010; Holdnack et al., 2017; Ivins et al., 2015; Karr, Garcia-Barrera, Holdnack, & Iverson, 2017, 2018; Mistridis et al., 2015). The normal frequency of low scores has been quantified as an advanced psychometric resource

referred to as multivariate base rates. Research on multivariate base rates has consistently demonstrated five principles: (a) low scores occur commonly among healthy individuals, (b) the number of low scores observed increases with the number of tests administered and scores interpreted, (c) the number of low scores observed is contingent on the cutoff used to define a low score (i.e.,  $\leq 16$ th percentile vs.  $\leq 2$ nd percentile), and, lastly, individuals with (d) lower intellectual functioning and (e) fewer years of education tend to present with a greater number of low scores (Brooks et al., 2013). Although many studies have focused on the frequency of low scores, no published study, to our knowledge, has examined the other side of the bell curve, exploring the normal multivariate frequency of *high scores* on a neuropsychological test battery. An understanding of how often participants within a

\*Correspondence and reprint requests to: Justin E. Karr, 79/96 Thirteenth Street, Charlestown Navy Yard, Charlestown, MA 02129, USA. E-mail: [jkarr1@mg.harvard.edu](mailto:jkarr1@mg.harvard.edu)

normative sample obtain one or more high scores would inform two aspects of clinical practice: the assessment of cognitive strengths and the assessment of deficits in patients with high premorbid functioning.

Few psychometric resources exist to properly quantify a cognitive strength. Some methods for calculating relative cognitive strengths have been widely used in the assessment of intellectual functions (Davis, 1959), but this method is not automated for all neuropsychological test batteries. Previous researchers have also discussed the detection of cognitive strengths through a close evaluation of test scatter, but there is no universal method for defining strengths in clinical assessment practice (Oakes, Lovejoy, Tartar, & Holdnack, 2013). Many clinicians simply interpret single high test scores as indicative of an underlying cognitive strength.

In addition to the assessment of cognitive strengths, researchers have also examined cognitive decline in participants with high premorbid functioning (Chodosh, Reuben, Albert, & Seeman, 2002; Deary, Starr, & MacLennan, 1998). There has been substantial emphasis on cognitive reserve in neuropsychological research, detailing the protective effects of intelligence, education, and occupational attainment against dementia risk (Scarmeas & Stern, 2004). Despite the possible protective effects of cognitive reserve, high baseline ability may obscure reductions in cognitive functioning because reduced performances may be average scores as opposed to frankly low scores in an assessment context. Clinicians often emphasize low scores in assessment practice, but high-functioning patients may present with average or even above average scores that indicate a decline from their previous level of cognitive functioning. Data on how many scores are expected to be high average or superior among healthy individuals with high intelligence or educational attainment can help clinicians understand when a high-functioning patient has more average scores than expected.

The current study used the normative data from the Delis-Kaplan Executive Function System (D-KEFS) (Delis, Kaplan, & Kramer, 2001), which is the most commonly administered test battery of executive function in clinical practice (Rabin, Paolillo, & Barr, 2016). Previous research has provided multivariate base rates of low scores for the D-KEFS (Cook et al., 2018; Crawford, Garthwaite, Sutherland, & Borland, 2011; Karr et al., 2017, 2018). This study aimed to determine if the principles of low score multivariate base rates applied to high scores as well, hypothesizing that (a) high scores will occur commonly among the D-KEFS normative sample, (b) the frequency of high scores will increase with the numbers of tests administered and scores interpreted, (c) the frequency of high scores will be related to the cutoff used to define a high score (i.e.,  $\geq 75$ th,  $\geq 84$ th,  $\geq 91$ st,  $\geq 95$ th, and  $\geq 98$ th percentiles), and, lastly, the frequency of high scores will be greater among individuals with (d) higher intellectual functioning and (e) more years of education.

## METHOD

### Participants

A total of 1050 participants (47.8% men) aged 16–89 years completed the D-KEFS as part of the standardization procedure, which recruited participants across the United States in close alignment with 2000 United States census data in terms of age, race/ethnicity, education, and geographic representation (Delis et al., 2001). The following age groups were represented with their corresponding sample sizes: 16–19 ( $n = 175$ ), 20–29 ( $n = 175$ ), 30–39 ( $n = 150$ ), 40–49 ( $n = 100$ ), 50–59 ( $n = 100$ ), 60–69 ( $n = 125$ ), 70–79 ( $n = 125$ ), and 80–89 ( $n = 100$ ). The sample was predominantly White (79.2%), but included representation from a range of racial/ethnic groups (i.e., 10.6% African American, 9.1% Hispanic, and 2.1% representative of other racial and ethnic backgrounds). Education was coded as the highest level of parental education for participants 16–19 years old and highest level of personal education for participants 20–89 years old. Education ranged from fewer than 8 years to college educated:  $\leq 8$  years (5.7%), 9–11 years (12.2%), 12 years (34.3%), 13–15 years (26.7%), and 16 or more years (21.1%). Participants were screened for self-reported medical and psychiatric conditions and removed if the condition could negatively impact cognitive performance. The data were received as age-referenced scaled scores ( $M = 10$ ,  $SD = 3$ ). A subsample of participants ( $n = 823$ ) completed the Wechsler Abbreviated Scale of Intelligence (WASI) (The Psychological Corporation, 1999), with WASI scores received as age-referenced standard scores ( $M = 100$ ,  $SD = 15$ ). The data were received anonymized from Pearson, and an institutional Human Research Ethics Board approved the secondary analyses conducted herein.

### Measures

The D-KEFS includes the nine following tests with their Total Achievement scores (i.e., scores most indicative of executive function) listed in parentheses: Trail Making Test (Number-Letter Switching: Time-to-Completion), Verbal Fluency Test (Letter Fluency: Total Correct; Category Fluency: Total Correct; Category Switching: Total Correct Responses; Category Switching: Total Switching Accuracy), Design Fluency Test (Total Correct Composite Scaled Score), Color-Word Interference Test (Inhibition: Time-to-Completion; Inhibition/Switching: Time-to-Completion), Sorting Test (Free Sorting: Confirmed Correct Sorts; Free Sorting: Description Score; Sort Recognition: Description Score), Twenty Questions Test (Total Weighted Achievement Score), Word Context Test (Total Consecutively Correct), Tower Test (Total Achievement Score; Move Accuracy Ratio), and Proverb Test (Total Achievement Score: Free Inquiry). A more detailed explanation of the D-KEFS tests is available in the Technical Manual (Delis et al., 2001) and in previously

published test reviews (Baron, 2004; Homack, Lee, & Riccio, 2005; Shunk, Davis, & Dean, 2006; Swanson, 2005).

## Statistical Analyses

Because previous D-KEFS multivariate base rate calculations did not identify age-related differences in the prevalence of low scores when comparing younger and older age spans (Crawford et al., 2011; Karr et al., 2017, 2018), the high score base rates were calculated for the full sample (i.e., ages 16–89 years) without stratification by age. The lack of differences in high score base rates between age groups was verified empirically through a series of non-parametric comparisons, identifying no significant differences in high score base rates based on age after controlling for multiple comparisons. As noted earlier, all scores were received as age-referenced scaled scores, which likely explains the lack of difference in base rates based on age. The multivariate base rates were calculated based on the frequency of high Total Achievement scores for the full battery and two brief batteries. The two brief batteries consisted of three tests (i.e., Trail Making Test, Verbal Fluency Test, and Color-Word Interference Test) and four tests (i.e., Trail Making Test, Verbal Fluency Test, Color-Word Interference Test, and Tower Test), selected based on the frequency of their administration in clinical practice (Rabin et al., 2016). The full test battery included 16 Total Achievement scores, the four-test battery included 9 Total Achievement scores, and the three-test battery included 7 Total Achievement scores. All participants completed D-KEFS tests in the context of the standardization of the full D-KEFS battery, and the same participants are represented in the base rate calculations for the full and brief batteries. Variability in sample sizes for the base rates across batteries is attributable to missing data because only cases with complete data were included in base rate calculations. The participants represented in the brief battery base rates were not administered fewer tests, rather fewer scores are being interpreted in the base rate calculation. Estimated based on the two-subtest WASI (i.e., Vocabulary, Matrix Reasoning), intelligence level was used to stratify the multivariate base rates based on the following categories:  $\leq 89$ , 90–99, 100–109, and  $\geq 110$ . The cutoffs used to define high scores were equivalent to the cutoffs used to define low scores in previous studies (Karr et al., 2017, 2018), albeit on the opposite side of the bell curve:  $\geq 75$ th,  $\geq 84$ th,  $\geq 91$ st,  $\geq 95$ th, and  $\geq 98$ th percentiles.

## RESULTS

The frequencies of high D-KEFS Total Achievement scores, based on  $\geq 75$ th,  $\geq 84$ th,  $\geq 91$ st,  $\geq 95$ th, and  $\geq 98$ th percentile cutoffs, for the three-test, four-test, and full batteries are provided in Tables 1–3, respectively, with stratifications by estimated level of intelligence and years of education. High scores occurred quite commonly among the D-KEFS normative sample. When interpreting scores from the full test battery, 87.8% obtained one or more high scores at or above

the 84th percentile (see Table 3). As expected, the prevalence of high scores decreased with the use of higher cutoffs for defining a high score. When interpreting scores from the four-test D-KEFS battery, 72.9% obtained one or more scores at or above the 84th percentile, 53.7% obtained one or more scores at or above the 91st percentile, 34.3% obtained one or more scores at or above the 95th percentile, and 18.1% obtained one or more scores at or above the 98th percentile (see Table 2).

As anticipated, the base rates of high scores were lower for the brief batteries than for the total battery. For the three-test, four-test, and full batteries, 24.1%, 34.3%, and 46.0% of the sample obtained at least one score at or above the 95th percentile, respectively (see Tables 1–3). Even when using the highest cutoff to define a high score (i.e.,  $\geq 98$ th percentile), 12.4% of the sample obtained at least one or more high scores on the three-test battery, 18.1% obtained at least one or more high scores on the four-test battery, and 24.3% obtained at least one or more high scores on the full battery (see Tables 1–3). As made clear by this example, scores that are often considered very superior based on a univariate “bell curve” interpretation occur quite commonly among healthy adults completing the D-KEFS, even when interpreting scores from just four tests (i.e., nearly 1 in 5 people; 18.1%).

Stratifications by intelligence and education had their expected relationships with the prevalence of high scores. The estimated level of intelligence from the two-test WASI was related to the base rates of high scores on the D-KEFS, where the number of high scores increased with higher WASI FSIQ. Among healthy individuals with high intellectual functioning (i.e., WASI  $\geq 110$ ), only about a third of the sample (i.e., 32.1%) obtained *no scores* at or above the 91st percentile when interpreting 7 scores from the three-test battery (see Table 1). When interpreting all Total Achievement scores from the full D-KEFS battery, all participants with this level of intellectual functioning obtained at least 1 score at or above the 75th percentile. This relationship for the three-test D-KEFS battery (i.e., 7 scores) among the total sample is displayed in Figure 1.

Years of education followed a similar positive relationship with the frequency of high scores, where a greater number of educational years were associated with a greater number of high scores. For the three-test battery, this relationship is displayed in Figure 2. Using the 84th percentile as the cutoff for a high score, when participants with 16 or more years of education were administered the three-test battery, about three quarters (i.e., 75.9%) obtained 1 or more high scores, about half (i.e., 48.6%) obtained 2 or more high scores, about a third (i.e., 37.3%) obtained 3 or more high scores, and about a quarter (i.e., 23.2%) obtained 4 or more high scores (see Table 1).

## DISCUSSION

There is a fairly large literature on the multivariate base rates of low scores in children (Brooks et al., 2010; Cook et al., 2018), adults (Holdnack et al., 2017), and older adults

**Table 1.** Base rates of high age-referenced D-KEFS Total Achievement scores in 16–89 year-olds for the three-test battery – 7 scores: Trail Making Test (1 score), Verbal Fluency Test (4 scores), and Color-Word Interference Test (2 scores)

Number of high scores	Total sample	WASI FSIQ				Years of education				
		≤89	90–99	100–109	≥110	≤8	9–11	12	13–15	≥16
Sample size	1028	161	187	210	249	57	125	351	275	220
<b>≥75th percentile</b>										
7 high scores	2.6	–	–	1.0	6.8	–	–	.9	3.3	6.8
6 or more	8.5	–	3.7	8.1	18.9	–	1.6	5.7	8.7	18.6
5 or more	18.0	2.5	11.8	16.7	34.1	5.3	6.4	14.8	18.2	32.7
4 or more	29.5	5.6	23.0	30.5	52.2	14.0	14.4	25.1	32.0	45.9
3 or more	45.1	17.4	40.6	47.6	66.7	22.8	27.2	40.2	53.1	59.1
2 or more	64.6	40.4	61.5	67.6	83.1	47.4	48.0	62.1	69.8	75.9
1 or more	86.4	71.4	85.6	87.1	95.2	73.7	79.2	86.3	90.2	89.1
No high scores	13.6	28.6	14.4	12.9	4.8	26.3	20.8	13.7	9.8	10.9
<b>≥84th percentile</b>										
6 high scores	1.2	–	.5	1.0	2.4	–	–	0.9	1.1	2.7
5 or more	5.3	0.6	1.1	4.8	12.0	–	0.8	3.7	5.1	11.8
4 or more	11.3	0.6	4.8	10.0	23.7	1.8	4.0	8.0	11.3	23.2
3 or more	21.3	5.6	13.9	19.5	40.2	10.5	11.2	16.5	21.5	37.3
2 or more	35.3	14.9	26.7	35.7	56.6	17.5	21.6	28.8	42.9	48.6
1 or more	61.6	39.1	53.5	63.8	80.3	47.4	48.0	57.5	64.4	75.9
No high scores	38.4	60.9	46.5	36.2	19.7	52.6	52.0	42.5	35.6	24.1
<b>≥91st percentile</b>										
6 high scores	0.2	–	–	–	0.4	–	–	–	–	0.9
5 or more	0.7	–	–	0.5	1.6	–	–	0.3	0.4	2.3
4 or more	3.1	–	0.5	2.4	8.0	–	0.8	1.7	2.2	8.6
3 or more	8.9	1.2	3.7	7.1	19.3	–	4.0	6.0	9.1	18.6
2 or more	18.4	4.3	11.2	15.7	34.9	7.0	8.8	14.0	20.4	31.4
1 or more	43.2	24.8	32.1	42.9	63.1	21.1	29.6	39.0	46.2	59.5
No high scores	56.8	75.2	67.9	57.1	36.9	78.9	70.4	61.0	53.8	40.5
<b>≥95th percentile</b>										
5 high scores	0.2	–	–	–	0.4	–	–	–	–	0.9
4 or more	0.8	–	–	1.4	1.6	–	0.8	0.6	0.4	1.8
3 or more	2.3	–	0.5	1.9	6.4	–	0.8	1.1	1.8	6.4
2 or more	7.3	1.2	4.3	5.2	15.3	1.8	3.2	4.0	6.9	16.8
1 or more	24.1	8.7	13.4	24.8	39.8	12.3	16.0	17.7	26.5	39.1
No high scores	75.9	91.3	86.6	75.2	60.2	87.7	84.0	82.3	73.5	60.9
<b>≥98th percentile</b>										
5 high scores	0.1	–	–	–	–	–	–	–	–	0.5
4 or more	0.1	–	–	–	–	–	–	–	–	0.5
3 or more	0.6	–	–	–	2.0	–	–	–	0.4	2.3
2 or more	2.9	0.6	0.5	2.4	7.2	–	1.6	1.7	2.9	6.4
1 or more	12.4	3.7	4.3	11.9	22.5	7.0	5.6	7.7	13.1	24.1
No high scores	87.6	96.3	95.7	88.1	77.5	93.0	94.4	92.3	86.9	75.9

WASI FSIQ = Wechsler Abbreviated Scale of Intelligence Full Scale Intelligence Quotient.

Note. All values represent cumulative percentages except for the rows labeled “No high scores,” which provide the percentage of the normative sample with no scores falling above the high score cutoffs. Standardization data from the *Delis-Kaplan Executive Function System (D-KEFS)*. Source: © 2001 NCS Pearson, Inc. used with permission. All rights reserved.

(Brooks et al., 2009b; Palmer, Boone, Lesser, & Wohl, 1998), and on comprehensive test batteries (Brooks et al., 2011; Crawford, Garthwaite, & Gault, 2007; Schretlen, Testas, Winicki, Pearlson, & Gordon, 2008) and domain specific batteries, such as memory (Brooks, Iverson, Feldman, & Holdnack, 2009; Brooks, Iverson, Holdnack, & Feldman, 2008) or executive function (Karr et al., 2017, 2018). Very little is known, however, about the probability of healthy people obtaining high scores when a battery of neuropsychological

tests is administered. This study explored whether the principles of multivariate base rates (Brooks et al., 2013) apply to high scores in a similar way that they apply to low scores. All principles of low score base rates were applicable to high score base rates using the D-KEFS: high scores occurred commonly within the normative sample, the frequency of high scores increased with the numbers of tests administered and scores interpreted, the frequency of high scores was related to the cutoff used to define a high score (i.e., ≥75th, ≥84th,

**Table 2.** Base rates of high age-referenced D-KEFS Total Achievement scores in 16–89 year-olds for the four-test battery – 9 scores: TMT (1 score), VF (4 scores), CWIT (2 scores), and TWT (2 scores)

Number of high scores	Total sample	WASI FSIQ				Years of education				
		≤89	90–99	100–109	≥110	≤8	9–11	12	13–15	≥16
Sample size	823	121	144	171	202	39	96	275	225	188
<b>≥75th percentile</b>										
9 high scores	0.4	–	–	0.6	0.5	–	–	0.4	0.4	0.5
8 or more	1.6	–	0.7	0.6	4.5	–	–	0.7	1.8	3.7
7 or more	7.5	0.8	2.1	4.1	16.8	–	1.0	3.6	9.8	15.4
6 or more	16.3	2.5	8.3	14.6	30.7	7.7	2.1	11.3	17.3	31.4
5 or more	26.2	4.1	20.1	26.3	47.5	15.4	9.4	20.7	28.9	42.0
4 or more	39.1	12.4	32.6	43.3	60.4	25.6	20.8	32.7	43.1	55.9
3 or more	58.0	27.3	54.9	60.8	78.7	38.5	41.7	52.0	63.1	72.9
2 or more	76.5	60.3	77.8	76.6	88.1	64.1	69.8	74.2	78.2	84.0
1 or more	93.2	86.0	91.0	93.0	98.5	84.6	91.7	91.6	95.1	95.7
No high scores	6.8	14.0	9.0	7.0	1.5	15.4	8.3	8.4	4.9	4.3
<b>≥84th percentile</b>										
8 high scores	0.1	–	–	–	–	–	–	–	–	0.5
7 or more	0.5	–	–	0.6	0.5	–	–	0.4	0.4	1.1
6 or more	3.4	–	1.4	2.3	6.9	–	–	2.2	3.6	7.4
5 or more	8.3	0.8	2.1	7.0	16.3	–	3.1	5.5	7.1	18.1
4 or more	16.6	4.1	9.7	13.5	32.2	10.3	8.3	11.6	17.3	28.7
3 or more	27.8	9.1	18.8	25.7	47.5	17.9	15.6	20.7	29.3	44.7
2 or more	45.2	25.6	40.3	43.3	65.8	35.9	32.3	37.5	50.7	58.5
1 or more	72.9	61.2	70.1	69.6	85.1	66.7	64.6	68.0	72.0	86.7
No high scores	27.1	38.8	29.9	30.4	14.9	33.3	35.4	32.0	28.0	13.3
<b>≥91st percentile</b>										
7 high scores	0.1	–	–	–	–	–	–	–	–	0.5
6 or more	0.4	–	–	–	1.0	–	–	–	–	1.6
5 or more	1.3	–	–	1.2	3.0	–	1.0	0.7	–	4.3
4 or more	5.0	0.8	1.4	3.5	10.9	–	2.1	2.9	5.3	10.1
3 or more	11.8	2.5	4.2	8.8	23.8	2.6	7.3	8.4	11.6	21.3
2 or more	24.5	10.7	18.1	19.3	42.6	15.4	12.5	19.6	26.2	37.8
1 or more	53.7	42.1	45.1	50.3	68.8	41.0	40.6	47.6	54.7	70.7
No high scores	46.3	57.9	54.9	49.7	31.2	59.0	59.4	52.4	45.3	29.3
<b>≥95th percentile</b>										
6 high scores	0.2	–	–	–	0.5	–	–	–	–	1.1
5 or more	0.2	–	–	–	0.5	–	–	–	–	1.1
4 or more	1.1	–	–	1.8	2.0	–	1.0	1.1	0.4	2.1
3 or more	3.2	–	1.4	2.9	5.9	2.6	2.1	1.8	2.7	6.4
2 or more	10.6	3.3	6.3	7.0	20.3	2.6	6.3	7.3	8.9	21.3
1 or more	34.3	29.8	24.3	31.6	45.0	33.3	24.0	28.0	35.6	47.3
No high scores	65.7	70.2	75.7	68.4	55.0	66.7	76.0	72.0	64.4	52.7
<b>≥98th percentile</b>										
6 high scores	0.1	–	–	–	–	–	–	–	–	0.5
5 or more	0.1	–	–	–	–	–	–	–	–	0.5
4 or more	0.1	–	–	–	–	–	–	–	–	0.5
3 or more	0.7	–	–	–	1.5	–	–	–	0.9	2.1
2 or more	3.9	1.7	0.7	2.9	7.9	–	1.0	2.9	3.6	8.0
1 or more	18.1	19.8	9.7	14.0	23.3	17.9	10.4	12.0	18.7	30.3
No high scores	81.9	80.2	90.3	86.0	76.7	82.1	89.6	88.0	81.3	69.7

TMT = Trail Making Test; VF = Verbal Fluency Test; CWIT = Color-Word Interference Test; TWT = Tower Test.

Note. All values represent cumulative percentages except for the rows labeled “No high scores,” which provide the percentage of the normative sample with no scores falling above the high score cutoffs. Only participants without missing data were included in the base rates calculation. Source: Standardization data from the *Delis-Kaplan Executive Function System (D-KEFS)*, © 2001 NCS Pearson, Inc. used with permission. All rights reserved.

**Table 3.** Base rates of high age-referenced D-KEFS Total Achievement scores for full nine-test battery in 16–89 year-olds – 16 scores: TMT (1 score), VF (4 scores), DF (1 score), CWIT (2 scores), ST (3 scores), 20Q (1 score), WC (1 score), TWT (2 scores), and PT (1 score)

Number of high scores	Total sample	WASI FSIQ				Years of education				
		≤89	90–99	100–109	≥110	≤8	9–11	12	13–15	≥16
Sample size	789	118	138	165	192	36	92	267	217	177
<b>≥75th percentile</b>										
15 high scores	0.4	–	–	–	1.0	–	–	–	0.5	1.1
14 or more	1.8	–	–	–	5.2	–	–	0.4	1.4	5.6
13 or more	3.5	–	0.7	–	10.9	–	–	0.7	4.6	9.0
12 or more	6.7	–	0.7	3.6	16.1	–	1.1	3.7	7.4	14.7
11 or more	10.3	0.8	0.7	7.3	24.5	–	1.1	6.0	12.0	21.5
10 or more	14.3	0.8	3.6	9.1	34.9	–	2.2	8.2	16.6	29.9
9 or more	20.2	1.7	9.4	16.4	44.8	–	3.3	14.2	22.6	39.0
8 or more	28.4	1.7	16.7	24.2	58.9	13.9	7.6	19.9	30.4	52.5
7 or more	37.3	4.2	23.9	35.8	71.4	19.4	12.0	29.2	42.4	59.9
6 or more	46.3	8.5	31.9	47.3	81.8	27.8	25.0	36.3	53.5	67.2
5 or more	60.3	19.5	47.8	67.3	92.2	44.4	35.9	49.8	69.1	81.4
4 or more	69.8	30.5	63.0	75.2	95.8	52.8	50.0	60.7	77.0	88.7
3 or more	80.5	49.2	78.3	86.7	97.4	63.9	68.5	74.9	86.2	91.5
2 or more	91.1	77.1	91.3	92.7	98.4	77.8	89.1	88.4	93.5	96.0
1 or more	98.0	93.2	97.8	98.8	100	91.7	98.9	96.3	99.5	99.4
No high scores	2.0	6.8	2.2	1.2	0	8.3	1.1	3.7	0.5	0.6
<b>≥84th percentile</b>										
14 high scores	0.1	–	–	–	–	–	–	–	–	0.6
13 or more	0.3	–	–	–	0.5	–	–	–	–	1.1
12 or more	0.6	–	–	–	1.6	–	–	–	–	2.8
11 or more	1.9	0.8	–	0.6	2.6	–	1.1	0.4	1.8	5.1
10 or more	4.4	0.8	0.7	1.2	9.9	–	1.1	1.9	5.1	10.2
9 or more	6.8	0.8	0.7	2.4	17.7	–	1.1	3.7	6.9	15.8
8 or more	9.5	0.8	0.7	5.5	24.5	–	2.2	4.9	10.1	21.5
7 or more	13.8	0.8	1.4	9.1	33.3	5.6	2.2	9.4	13.8	28.2
6 or more	20.0	1.7	7.2	12.1	46.9	5.6	7.6	13.5	21.2	37.9
5 or more	27.1	3.4	13.8	18.8	60.4	8.3	10.9	19.5	29.0	48.6
4 or more	37.6	7.6	23.9	32.7	73.4	16.7	16.3	27.0	43.8	61.6
3 or more	52.3	18.6	40.6	54.5	80.7	38.9	35.9	42.7	56.7	72.9
2 or more	69.3	39.0	60.9	72.7	92.7	61.1	52.2	62.2	73.7	85.3
1 or more	87.8	73.7	84.1	89.7	97.9	77.8	77.2	83.9	91.7	96.6
No high scores	12.2	26.3	15.9	10.3	2.1	22.2	22.8	16.1	8.3	3.4
<b>≥91st percentile</b>										
13 high scores	0.1	–	–	–	–	–	–	–	–	0.6
12 or more	0.1	–	–	–	–	–	–	–	–	0.6
11 or more	0.1	–	–	–	–	–	–	–	–	0.6
10 or more	0.4	–	–	–	1.0	–	–	–	–	1.7
9 or more	1.0	–	–	0.6	2.1	–	1.1	–	0.5	3.4
8 or more	2.3	–	–	0.6	6.3	–	1.1	0.4	1.4	7.3
7 or more	3.7	–	0.7	1.2	8.9	–	1.1	1.1	3.2	10.2
6 or more	6.0	0.8	0.7	1.8	14.1	2.8	2.2	2.2	5.5	14.7
5 or more	10.8	1.7	2.2	5.5	26.6	2.8	4.3	7.5	9.2	22.6
4 or more	16.0	1.7	2.9	12.7	39.1	2.8	7.6	12.4	13.8	31.1
3 or more	27.1	6.8	14.5	21.8	53.6	5.6	12.0	21.7	27.6	46.9
2 or more	43.2	17.8	35.5	37.6	72.9	33.3	25.0	34.8	45.6	64.4
1 or more	68.9	54.2	59.4	66.7	87.0	61.1	58.7	61.0	72.4	83.6
No high scores	31.1	45.8	40.6	33.3	13.0	38.9	41.3	39.0	27.6	16.4
<b>≥95th percentile</b>										
11 high scores	0.1	–	–	–	–	–	–	–	–	0.6
10 or more	0.1	–	–	–	–	–	–	–	–	0.6
9 or more	0.3	–	–	–	0.5	–	–	–	–	1.1
8 or more	0.4	–	–	–	1.0	–	–	–	–	1.7

(Continued)

**Table 3.** (Continued)

Number of high scores	Total sample	WASI FSIQ				Years of education				
		≤89	90–99	100–109	≥110	≤8	9–11	12	13–15	≥16
7 or more	1.1	–	–	1.2	2.1	–	1.1	–	0.9	3.4
6 or more	1.9	–	–	1.2	4.7	–	1.1	–	1.8	5.6
5 or more	3.3	–	–	1.2	7.3	–	1.1	1.1	1.8	10.2
4 or more	5.7	0.8	0.7	2.4	14.1	–	1.1	4.5	4.1	13.0
3 or more	9.8	0.8	2.9	4.8	23.4	–	3.3	6.0	9.7	20.9
2 or more	21.3	5.9	9.4	18.8	41.1	11.1	10.9	16.5	21.2	36.2
1 or more	46.0	33.9	34.8	43.6	65.1	36.1	35.9	38.6	49.8	59.9
No high scores	54.0	66.1	65.2	56.4	34.9	63.9	64.1	61.4	50.2	40.1
≥98th percentile										
10 high scores	0.1	–	–	–	–	–	–	–	–	0.6
9 or more	0.1	–	–	–	–	–	–	–	–	0.6
8 or more	0.1	–	–	–	–	–	–	–	–	0.6
7 or more	0.1	–	–	–	–	–	–	–	–	0.6
6 or more	0.1	–	–	–	–	–	–	–	–	0.6
5 or more	0.6	–	–	1.2	1.0	–	1.1	–	–	2.3
4 or more	1.9	–	–	1.2	3.6	–	1.1	0.4	0.5	6.8
3 or more	3.7	–	–	1.8	8.3	–	1.1	2.2	1.4	10.7
2 or more	8.5	1.7	2.2	5.5	19.3	–	3.3	6.0	6.0	19.8
1 or more	24.3	20.3	13.8	20.6	37.0	16.7	15.2	17.2	27.6	37.3
No high scores	75.7	79.7	86.2	79.4	63.0	83.3	84.8	82.8	72.4	62.7

TMT = Trail Making Test; VF = Verbal Fluency Test; DF = Design Fluency Test; CWIT = Color-Word Interference Test; ST = Sorting Test; 20Q = Twenty Questions Test; WC = Word Context Test; TWT = Tower Test; PT = Proverb Test.

Note. All values represent cumulative percentages except for the rows labeled “No high scores,” which provide the percentage of the normative sample with no scores falling above the high score cutoffs. Only participants without missing data were included in the base rates calculation. Source: Standardization data from the *Delis-Kaplan Executive Function System* (D-KEFS), © 2001 NCS Pearson, Inc. used with permission. All rights reserved.

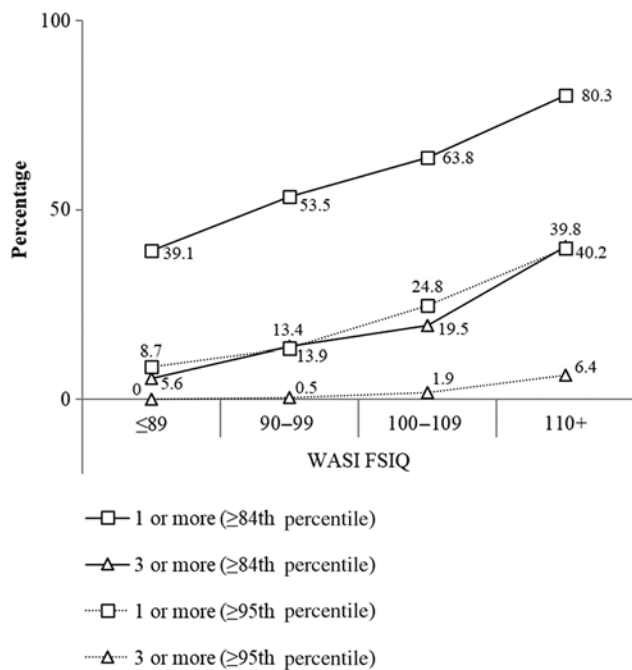
≥91st, ≥95th, ≥98th percentile), and, lastly, the frequency of high scores was greater among individuals with higher intellectual functioning and more years of education.

It is possible to use the base rates of high scores to identify a *pattern* of executive function test performance that is above average. For example, as seen in Table 1, when considering 7 scores derived from three tests, obtaining 5 or more scores at or above the 75th percentile [scaled score (SS) ≥12] occurs in only 18.0% of the standardization sample. Obtaining 4 or more scores at or above the 84th percentile (SS ≥13) occurs in only 11.3%, and obtaining 2 or more scores at or above the 91st percentile (SS ≥14) occurs in 18.4%. All of those patterns of performances would be considered above average for adults in the standardization sample. Having a single high score on tests of executive function is common, but having multiple high test scores reflects a pattern of performance that is more likely to represent a strength in executive function.

Previous research has applied multivariate base rates of low scores to the diagnosis of mild cognitive impairment to reduce the risk for misdiagnosis based on over-interpreting a single low score (Brooks et al., 2009b; Brooks et al., 2008; Brooks, Iverson, & White, 2007; Mistradis et al., 2015). In a similar manner, the high score base rates could be applied to assist with the detection of reduced cognitive functioning among high-functioning patients. For example, consider a 68-year-old man with a successful career as a manager and Master’s degree level of education, who presents to a clinic

with concerns of cognitive decline, reporting issues with disorganization and memory (and brain imaging revealed mild small vessel ischemic disease). As part of the evaluation, a neuropsychologist administers three common D-KEFS tests: the Trail Making, Verbal Fluency, and Color-Word Interference Tests, providing seven Total Achievement scores for interpretation. If this patient obtained all scores between the 16th (SS = 7) and 63rd (SS = 11) percentiles, with only one score at the 16th percentile, the neuropsychologist might conclude that this represents intact executive function ability. After all, having one or more low D-KEFS scores (16th percentile) out of 7 occurs in 45.9% of people with university degrees (Karr et al., 2017). However, as seen in Table 1, 89.1% of healthy adults with 16 or more years of education would have obtained one or more scores at or above the 75th percentile and this patient obtained no scores above this threshold. Therefore, it is not the presence of low scores that reflects a possible cognitive deficit; it is the absence of high scores. His performance is consistent with just 10.9% of the normative sample.

As a second example, imagine a high-functioning, university educated business woman is being evaluated because her multiple sclerosis has worsened. Her estimated longstanding intelligence is in the high average to superior classification range. On the three-test D-KEFS battery, she obtains the following scaled scores: 7, 8, 9, 10, 10, 11, and 11. According to the base rates of low scores, presented elsewhere (Karr et al., 2017), having 2 or more scores at or below the 25th percentile



**Fig. 1.** The data included in this figure were based on interpretation of the three-test D-KEFS battery (seven Total Achievement scores) for the sub-sample of participants aged 16–89 years old administered the WASI FSIQ. Standardization data from the Delis-Kaplan Executive Function System (D-KEFS). Source: © 2001 NCS Pearson, Inc. used with permission. All rights reserved. WASI FSIQ = Wechsler Abbreviated Scale of Intelligence Full Scale Intelligence Quotient.

occurs in 27.8% of those with above average intelligence and having one or more low scores at or below the 16th percentile occurs in 38.6% of high-functioning people. As such, the clinician might be inclined to interpret that pattern of performance as broadly normal. However, as seen in Table 1, only 4.8% of people with high average or superior intellectual functioning obtain *no* scores at or above the 75th percentile when considering those 7 scores. Again, the absence of a high score might reflect an acquired deficit in her case.

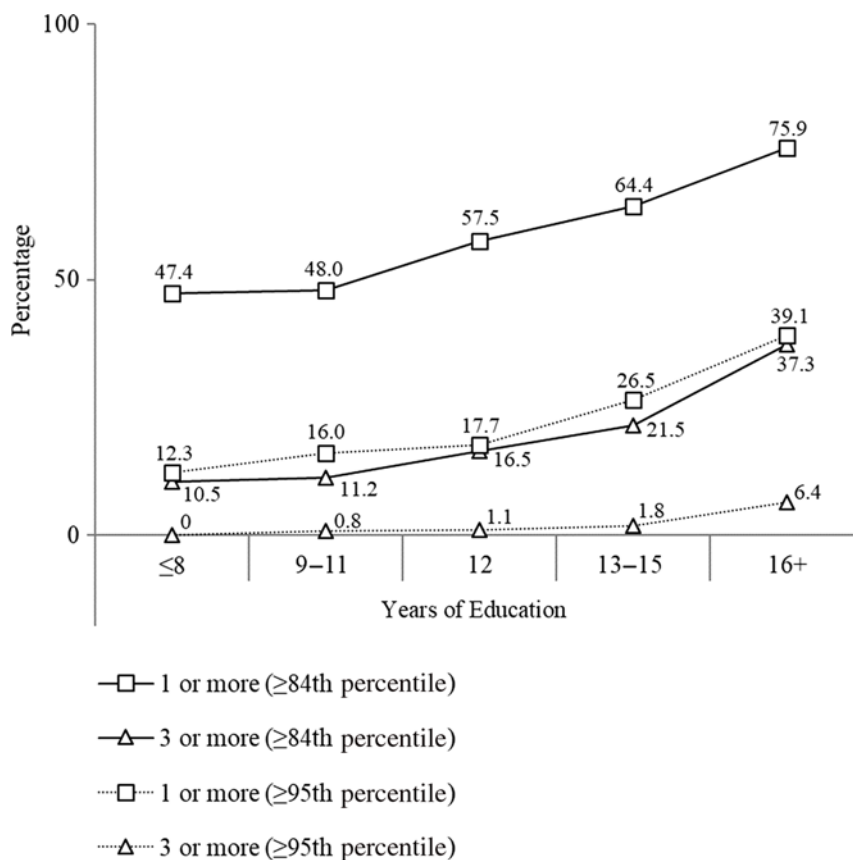
As with high-functioning individuals with many years of education, these high score base rates may have clinical utility with individuals of lower intellectual ability or educational attainment, where declines in functioning can be equally as difficult to identify. As a case example, consider a 45-year-old man presenting for an outpatient neuropsychological assessment after a workplace accident resulting in a penetrative traumatic brain injury to his ventral medial prefrontal cortex. He described himself as a low-average student throughout his life, and he left high school after 9th grade. He wants to return to work, but his employer is concerned about an increase in inappropriate comments and a short temper by the patient, both of which are new since the brain injury. His two-subtest WASI FSIQ is 88. He also completed the four-test D-KEFS battery and attained the following nine scores: 6, 7, 8, 8, 9, 9, 9, 10, and 11. His 2 scores at or below the 16th percentile ( $SS = 7$ ) were attained on the Inhibition and Inhibition/Switching conditions of the Color-Word Interference Test.

Based on low score base rates published elsewhere (Karr et al., 2018), 90.2% of individuals with an FSIQ at or below 89 would attain 4 or more scores at or below the 25th percentile and 92.0% of these individuals would attain 2 or more scores at or below the 16th percentile. Based on the education stratification, 75.0% of individuals with 9–11 years of education would attain 4 or more scores at or below the 25th percentile and 81.5% would attain 2 or more scores at or below the 16th percentile. Based on an interpretation of low scores alone, this individual appears to fall in alignment with his expected level of premorbid functioning. However, he presents with no high scores, although 86.0% of individuals with a similar WASI FSIQ and 91.7% of individuals with a similar level of education obtain at least one high score on this four-test D-KEFS battery. Although his performances are not indicative of an acquired cognitive impairment based on his profile of low scores, his lack of a single high score may correspond to a decline in cognitive functioning, possibly associated with the injury.

Each of these examples makes a case conceptualization based on the frequency of high scores, but it is important to emphasize that base rates are best used in combination with relevant case information to inform clinical judgment. A patient presenting with a specific condition may have an expected executive function deficit that occurs in isolation of other low scores within a battery. In the context of a patient with high pre-morbid functioning, this may be an average score in isolation, with this individual attaining an expected number of high scores based on normative data. Researchers using the D-KEFS have found reductions in executive function attributable to specific neurological conditions, including acquired conditions, such as traumatic brain injury (Heled, Hoofien, Margalit, Natovich, & Agranov, 2012; Strong, Tiesma, & Donders, 2011), and neurodegenerative conditions, such as frontotemporal dementia (Huey et al., 2009) and Parkinson's disease (McKinlay, Grace, Dalrymple-Alford, & Roger, 2010). In clinical practice, an interpretation of a specific profile of scores may be more useful than using multivariate base rates in isolation. Multivariate base rates are best used in combination with knowledge of the presenting condition, individual case history, and any available clinical information to assist with formulation (e.g., neuroimaging).

All three of these examples above considered education in the case conceptualization. It is important to note that age stratifications are related to the age of the examinee. The base rates were calculated using the full sample, ages 16–89 years; however, education was coded as parental education for participants 16–19 years old, and although education was coded as personal education for participants 20–29 years old, some of these participants may have fewer years of education due to age and progress through schooling (i.e., in active pursuit of a higher education, but yet to achieve a degree) as opposed to limited opportunity or intention to pursue higher education. The correlations between education and high scores did not meaningfully change with the exclusion of 16–19 or 16–29 year olds from the sample, but it is important for the clinician to consider (a) parental education as opposed to





**Fig. 2.** The data included in this figure were based on interpretation of the three-test D-KEFS battery (seven Total Achievement scores) for the sub-sample of participants aged 16–89 years old. Standardization data from the Delis-Kaplan Executive Function System (D-KEFS). *Source:* © 2001 NCS Pearson, Inc. used with permission. All rights reserved.

personal education for an examinee between the ages of 16 and 19 years, and (b) a 20–29 year-old in active pursuit of higher education may not be comparable to a middle-aged or older adult with 12 or 13–15 years of education. The WASI stratifications may have greater utility for young adults who are actively pursuing a college or university degree.

This study has several limitations. First, the standardization sample was used to calculate both the norm-referenced scores and the multivariate base rates. If an independent sample of participants was administered the D-KEFS and their scores were standardized using the published D-KEFS norms, base rates of high scores for this sample might differ from the standardization sample, especially if this sample was drawn from a different population (e.g., all college-educated adults). A replication of the base rates with a new group of participants would be informative to determine the comparability of the frequencies of high scores to the standardization sample.

Another limitation pertains to the derivation of base rates for different permutations of the D-KEFS battery. There were no separate samples administered briefer or longer batteries of D-KEFS tests, rather the base rates for briefer batteries were calculated based on scores from a subset of tests that the participants were administered. Participants exposed to more tests could have greater familiarity with testing or reduced test anxiety, strengthening performance, or alternatively, some of these

participants could experience increased fatigue or inattention due to lengthier test administration.

The study is also limited by intelligence stratifications based on the first edition of the WASI, which was the available edition at the time the D-KEFS was standardized. The base rates cannot be considered precisely accurate if alternatives are used to estimate intellectual (or premorbid intellectual) functioning. Instead, the clinician should consider the general principles set out in our tables, not the actual base rates. It is also important to note that the intelligence-stratified base rates are based on the currently obtained WASI FSIQ, not estimated premorbid intelligence. The clinician needs to determine whether to interpret performances based on current or estimated premorbid intelligence. For example, if a person with a known neurological condition obtained a WASI FSIQ of 109, it would be reasonable for the clinician to assume that the person's premorbid FSIQ was at least 110 and use the base rate tables for those with high average intelligence.

Additional limitations pertain to the D-KEFS standardization sample, including the age of the dataset, small sample sizes for some education stratifications, and lack of co-norming with other commonly administered test batteries. The D-KEFS standardization sample was stratified to align with the 2000 United States census based on regional distribution and

multiple demographic characteristics, including sex, education, and race/ethnicity. Considering the test remains in popular use roughly two decades later, the sample has likely become less consistent with the population over time, which limits the utility of the base rate tables in the assessment of individuals from groups underrepresented in the normative data (e.g., individuals from racial and ethnic minority backgrounds). Education stratifications were based on proportional representation in the United States population, which resulted in some groups having very small sample sizes (e.g., for the three-test battery, only 57 participants had 8 or fewer years of education). These small sample sizes reduce the generalizability of these normative groups in clinical practice. The normative sample also lacks co-normative data with other commonly administered test batteries (e.g., Wechsler Adult Intelligence Scale, Fourth Edition; Wechsler Memory Scale, Fourth Edition). When administered in conjunction with tests of other domains, the normal frequency of high scores would likely be greater due to an increased number of tests administered and interpreted; however, without co-normative data, multivariate base rates of high scores for a more comprehensive battery cannot be derived. A final limitation is that base rates are provided for only three permutations of the D-KEFS battery, and neuropsychologists in clinical practice may use a more flexible approach, where fewer, more, or a different combination of D-KEFS tests are administered.

Despite these limitations, these findings provide the first multivariate base rates calculated for high scores in neuropsychological assessment, complementing the abundant research examining the normal occurrence of low scores among healthy adults. These findings support the principles of multivariate base rates as applied to *the other side of the bell curve*, and future research can evaluate if these principles hold for other test batteries commonly administered in clinical practice. Although it has yet to be empirically evaluated, the normal prevalence of high scores on the D-KEFS may be similar to other batteries. Multivariate base rate research has consistently demonstrated that obtaining one or more low scores is fairly common across diverse test batteries. The observed base rates vary modestly across batteries, and principles of low score base rates have been established as opposed to specific rules-of-thumb for expected frequencies of low scores. After future studies replicate the current findings on high scores base rates with different test batteries, an omnibus study aggregating base rates across all test batteries evaluated would be of great clinical utility for neuropsychologists seeking to understand the normal frequency of both low and high scores on a test battery. The preparation of additional multivariate base rates and methods for using these high score base rates in conjunction with low score base rates will improve the interpretation of cognitive strengths and weaknesses in clinical practice. Further, high score base rates may improve the detection of mild cognitive impairment or the early stages of dementia among individuals with high premorbid ability, and future research should evaluate their sensitivity at detecting cognitive decline in high-functioning patients.

## ACKNOWLEDGEMENTS

Data used for the analyses reported within this manuscript were provided by Pearson, Inc. (2001). Standardization data from the Delis-Kaplan Executive Function System (D-KEFS). Copyright© 2001 NCS Pearson, Inc. used with permission. All rights reserved. San Antonio: Pearson, Inc.

Mauricio A. Garcia-Barrera notes that this work is related, in part, to his grant entitled Characteristics and Correlates of Intraindividual Variability in Executive Control Processes (Natural Sciences and Engineering Research Council of Canada; 418676-2012). Grant L. Iverson notes that this work is related, in part, to the TBI Endpoints Development (TED) Initiative and a grant entitled Development and Validation of a Cognition Endpoint for Traumatic Brain Injury Clinical Trials (subaward from W81XWH-14-2-0176).

Mauricio A. Garcia-Barrera has served in the past as a consultant for Pearson. James A. Holdnack has been previously employed by Pearson, the company that owns and distributes the D-KEFS. Grant L. Iverson has received research support from test publishing companies in the past, including PAR, Inc., ImPACT Applications, Inc., and CNS Vital Signs. He acknowledges current unrestricted philanthropic support from ImPACT Applications, Inc. He receives royalties for one neuropsychological test (Wisconsin Card Sorting Test-64 Card Version).

## CONFLICTS OF INTEREST

The authors have nothing to disclose.

## REFERENCES

- Baron, I.S. (2004). Delis-Kaplan Executive Function System. *Child Neuropsychology*, 10(2), 147–152. doi: [10.1080/09297040490911140](https://doi.org/10.1080/09297040490911140)
- Binder, L.M., Iverson, G.L., & Brooks, B.L. (2009). To err is human: “Abnormal” neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, 24(1), 31–46. doi: [10.1093/arclin/acn001](https://doi.org/10.1093/arclin/acn001)
- Brooks, B.L., Holdnack, J.A., & Iverson, G.L. (2011). Advanced clinical interpretation of the WAIS-IV and WMS-IV: Prevalence of low scores varies by level of intelligence and years of education. *Assessment*, 18(2), 156–167. doi: [10.1177/1073191110385316](https://doi.org/10.1177/1073191110385316)
- Brooks, B.L., Iverson, G.L., Feldman, H.H., & Holdnack, J.A. (2009a). Minimizing misdiagnosis: Psychometric criteria for possible or probable memory impairment. *Dementia and Geriatric Cognitive Disorders*, 27(5), 439–450. doi: [10.1159/000215390](https://doi.org/10.1159/000215390)
- Brooks, B.L., Iverson, G.L., & Holdnack, J.A. (2013). Understanding and using multivariate base rates with the WAIS-IV/WMS-IV. In J.A. Holdnack, L.W. Drozdick, L.G. Weiss, & G.L. Iverson (Eds.), *WAIS-IV, WMS-IV, and ACS: Advanced Clinical Interpretation* (pp. 75–102). San Diego, CA: Elsevier Science. doi: [10.1016/B978-0-12-386934-0.00002-X](https://doi.org/10.1016/B978-0-12-386934-0.00002-X)
- Brooks, B.L., Iverson, G.L., Holdnack, J.A., & Feldman, H.H. (2008). Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International*

- Neuropsychological Society*, 14(3), 463–478. doi: [10.1017/S1355617708080521](https://doi.org/10.1017/S1355617708080521)
- Brooks, B.L., Iverson, G.L., Lanting, S.C., Horton, A.M., & Reynolds, C.R. (2012). Improving test interpretation for detecting executive dysfunction in adults and older adults: Prevalence of low scores on the Test of Verbal Conceptualization and Fluency. *Applied Neuropsychology: Adult*, 19(1), 61–70. doi: [10.1080/09084282.2012.651951](https://doi.org/10.1080/09084282.2012.651951)
- Brooks, B.L., Iverson, G.L., & White, T. (2007). Substantial risk of “Accidental MCI” in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 13(3), 490–500. doi: [10.1017/S1355617707070531](https://doi.org/10.1017/S1355617707070531)
- Brooks, B.L., Iverson, G.L., & White, T. (2009b). Advanced interpretation of the neuropsychological assessment battery with older adults: Base rate analyses, discrepancy scores, and interpreting change. *Archives of Clinical Neuropsychology*, 24(7), 647–657. doi: [10.1093/arclin/acp061](https://doi.org/10.1093/arclin/acp061)
- Brooks, B.L., Sherman, E.M.S., & Iverson, G.L. (2010). Healthy children get low scores too: Prevalence of low scores on the NEPSY-II in preschoolers, children, and adolescents. *Archives of Clinical Neuropsychology*, 25(3), 182–190. doi: [10.1093/arclin/acq005](https://doi.org/10.1093/arclin/acq005)
- Chodosh, J., Reuben, D.B., Albert, M.S., & Seeman, T.E. (2002). Predicting cognitive impairment in high-functioning community-dwelling older persons: MacArthur studies of successful aging. *Journal of the American Geriatrics Society*, 50(6), 1051–1060. doi: [10.1046/j.1532-5415.2002.50260.x](https://doi.org/10.1046/j.1532-5415.2002.50260.x)
- Cook, N.E., Karr, J.E., Brooks, B.L., Garcia-Barrera, M.A., Holdnack, J.A., & Iverson, G.L. (2018). Multivariate base rates for the assessment of executive functioning among children and adolescents. *Child Neuropsychology*, 25(6), 836–858. doi: [10.1080/09297049.2018.1543389](https://doi.org/10.1080/09297049.2018.1543389)
- Crawford, J.R., Garthwaite, P.H., & Gault, C.B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology*, 21(4), 419–430. doi: [10.1037/0894-4105.21.4.419](https://doi.org/10.1037/0894-4105.21.4.419)
- Crawford, J.R., Garthwaite, P.H., Sutherland, D., & Borland, N. (2011). Some supplementary methods for the analysis of the Delis-Kaplan Executive Function System. *Psychological Assessment*, 23(4), 888–898. doi: [10.1037/a0023712](https://doi.org/10.1037/a0023712)
- Davis, F.B. (1959). Interpretation of differences among averages and individual test scores. *Journal of Educational Psychology*, 50(4), 162–170. doi: [10.1037/h0044024](https://doi.org/10.1037/h0044024)
- Deary, I.J., Starr, J.M., & MacLennan, W.J. (1998). Is age kinder to the initially more able?: Differential ageing of a verbal ability in the healthy old people in Edinburgh study. *Intelligence*, 26(4), 357–375. doi: [10.1016/S0160-2896\(99\)00005-7](https://doi.org/10.1016/S0160-2896(99)00005-7)
- Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *The Delis-Kaplan Executive Function System: Technical manual*. San Antonio, TX: The Psychological Corporation.
- Heled, E., Hoofien, D., Margalit, D., Natovich, R., & Agranov, E. (2012). The Delis-Kaplan Executive Function System sorting test as an evaluative tool for executive functions after severe traumatic brain injury: A comparative study. *Journal of Clinical and Experimental Neuropsychology*, 34(2), 151–159. doi: [10.1080/13803395.2011.625351](https://doi.org/10.1080/13803395.2011.625351)
- Holdnack, J.A., Tulskey, D.S., Brooks, B.L., Slotkin, J., Gershon, R., Heinemann, A.W., & Iverson, G.L. (2017). Interpreting patterns of low scores on the NIH toolbox cognition battery. *Archives of Clinical Neuropsychology*, 32(5), 574–584. doi: [10.1093/arclin/acx032](https://doi.org/10.1093/arclin/acx032)
- Homack, S., Lee, D., & Riccio, C.A. (2005). Test review: Delis-Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology*, 27(5), 599–609. doi: [10.1080/13803390490918444](https://doi.org/10.1080/13803390490918444)
- Huey, E.D., Goveia, E.N., Paviol, S., Pardini, M., Krueger, F., Zamboni, G., Tierney, M.C., Wassermann, E.M., & Grafman, J. (2009). Executive dysfunction in frontotemporal dementia and corticobasal syndrome. *Neurology*, 72(5), 453–459. doi: [10.1212/01.wnl.0000341781.39164.26](https://doi.org/10.1212/01.wnl.0000341781.39164.26)
- Ivins, B.J., Lange, R.T., Cole, W.R., Kane, R., Schwab, K.A., & Iverson, G.L. (2015). Using base rates of low scores to interpret the ANAM4 TBI-MIL battery following mild traumatic brain injury. *Archives of Clinical Neuropsychology*, 30(1), 26–38. doi: [10.1093/arclin/acu072](https://doi.org/10.1093/arclin/acu072)
- Karr, J.E., Garcia-Barrera, M.A., Holdnack, J.A., & Iverson, G.L. (2017). Using multivariate base rates to interpret low scores on an abbreviated battery of the Delis-Kaplan Executive Function System. *Archives of Clinical Neuropsychology*, 32(3), 297–305. doi: [10.1093/arclin/acw105](https://doi.org/10.1093/arclin/acw105)
- Karr, J.E., Garcia-Barrera, M.A., Holdnack, J.A., & Iverson, G.L. (2018). Advanced clinical interpretation of the Delis-Kaplan Executive Function System: Multivariate base rates of low scores. *The Clinical Neuropsychologist*. doi: [10.1080/13854046.2017.1334828](https://doi.org/10.1080/13854046.2017.1334828)
- McKinlay, A., Grace, R.C., Dalrymple-Alford, J.C., & Roger, D. (2010). Characteristics of executive function impairment in Parkinsons disease patients without dementia. *Journal of the International Neuropsychological Society*, 16(2), 268–277. doi: [10.1017/S1355617709991299](https://doi.org/10.1017/S1355617709991299)
- Mistridis, P., Egli, S.C., Iverson, G.L., Berres, M., Willmes, K., Welsh-Bohmer, K.A., & Monsch, A.U. (2015). Considering the base rates of low performance in cognitively healthy older adults improves the accuracy to identify neurocognitive impairment with the Consortium to Establish a Registry for Alzheimer’s Disease-Neuropsychological Assessment Battery. *European Archives of Psychiatry and Clinical Neuroscience*, 265(5), 407–417. doi: [10.1007/s00406-014-0571-z](https://doi.org/10.1007/s00406-014-0571-z)
- Oakes, H., Lovejoy, D., Tartar, S., & Holdnack, J.A. (2013). Understanding index and subtest scatter in healthy adults. In J.A. Holdnack, L.W. Drozdick, L.G. Weiss, & G.L. Iverson (Eds.), *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation* (pp. 103–169). San Diego, CA: Elsevier Science. doi: [10.1016/B978-0-12-386934-0.00003-1](https://doi.org/10.1016/B978-0-12-386934-0.00003-1)
- Palmer, B.W., Boone, K.B., Lesser, I.M., & Wohl, M.A. (1998). Base rates of “impaired” neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology*, 13(6), 503–511. doi: [10.1016/S0887-6177\(97\)00037-1](https://doi.org/10.1016/S0887-6177(97)00037-1)
- Rabin, L.A., Paolillo, E., & Barr, W.B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-Year Period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, 31(3), 206–230. doi: [10.1093/arclin/acw007](https://doi.org/10.1093/arclin/acw007)
- Scarmeas, N., & Stern, Y. (2004). Cognitive reserve: Implications for diagnosis and prevention of Alzheimer’s disease. *Current Neurology and Neuroscience Reports*, 4(5), 374–380. doi: [10.1007/s11910-004-0084-7](https://doi.org/10.1007/s11910-004-0084-7)
- Schretlen, D.J., Testas, S.M., Winicki, J.M., Pearson, G.D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the*

- International Neuropsychological Society*, 14(3), 436–445. doi: [10.1017/S1355617708080387](https://doi.org/10.1017/S1355617708080387)
- Shunk, A.W., Davis, A.S., & Dean, R.S. (2006). Review of Delis-Kaplan Executive Function System (D-KEFS). *Applied Neuropsychology*, 13(4), 275–27. doi: [10.1207/s15324826an1304\\_9](https://doi.org/10.1207/s15324826an1304_9)
- Strong, C.A.H., Tiesma, D., & Donders, J. (2011). Criterion validity of the Delis-Kaplan Executive Function System (D-KEFS) fluency subtests after traumatic brain injury. *Journal of the International Neuropsychological Society*, 17(2), 230–237. doi: [10.1017/S1355617710001451](https://doi.org/10.1017/S1355617710001451)
- Swanson, J. (2005). The Delis-Kaplan Executive Function xSystem: A Review. *Canadian Journal of School Psychology*. doi: [10.1177/0829573506295469](https://doi.org/10.1177/0829573506295469)
- The Psychological Corporation. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.