



Creyos Health Report Interpretation Guide

Quickly gain validated and powerful brain health insights.



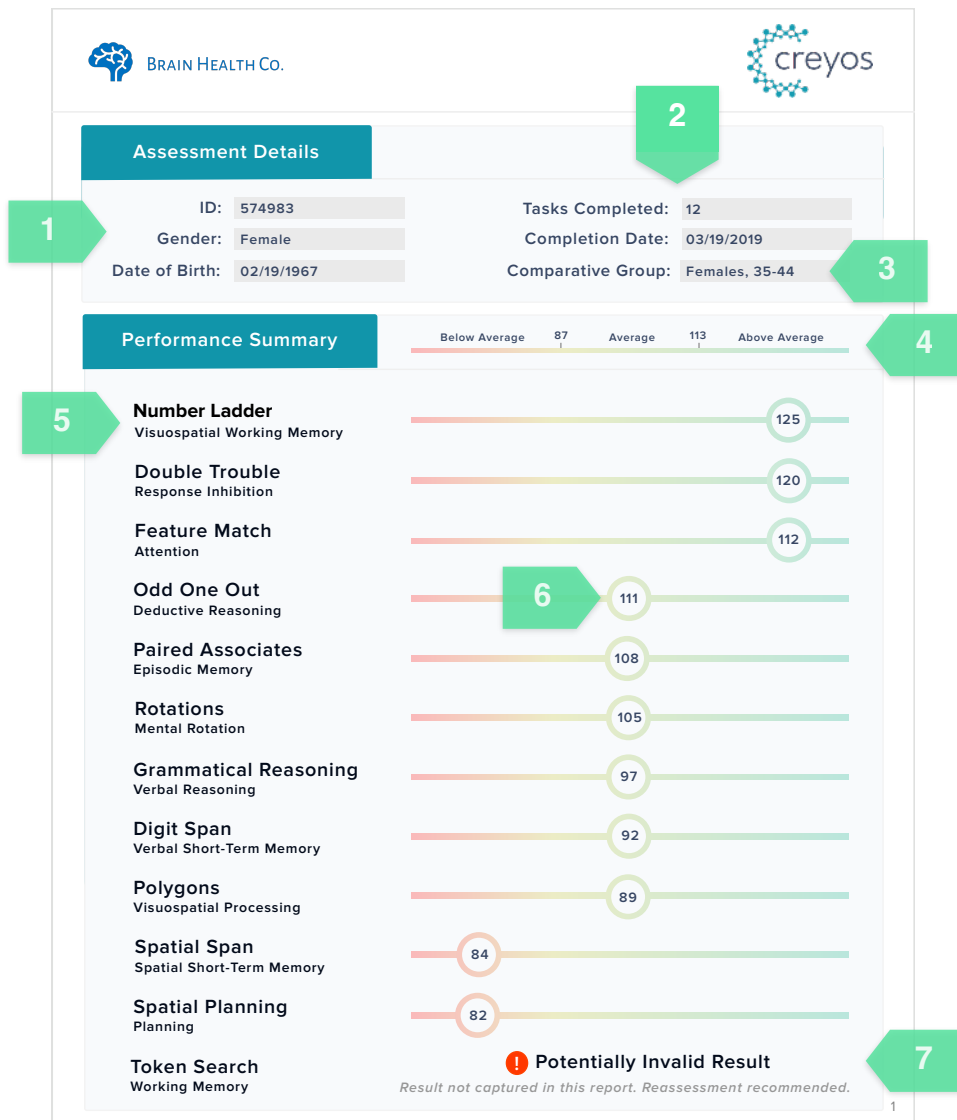
- A. Core Components of the Creyos Health Report
- B. Score Calculations
- C. Task Validity Overview and Conditions for Validity
- D. Meaningful Change Indicator Overview
- E. Answers to Common Interpretation Questions
- Appendix: Raw Score Calculations

Important Information Before You Read This Document

Creyos (formerly Cambridge Brain Sciences) provides a scientifically-validated and objective measure of an individual's cognition, however, it is not a diagnostic tool. Creyos Health should be used in conjunction with other information and clinical judgement to reach conclusions regarding an individual's health. Ultimately, Creyos Health does not replace the judgement of a practitioner and Creyos does not assume responsibility for the outcome of decisions made based on Creyos Health data.

A. Core Components of the Creyos Health Report: Summary Page

The Creyos Health report is designed to be easy to read and shareable with the individual assessed. A brief overview of the components of the report is shown below, starting with the one-page summary. Every Creyos Health report will first feature a summary page to enable you to quickly scan above average, average, and below average performance areas as detailed below:



A. Core Components of the Creyos Health Report: Task Results Page

The Creyos Health report is designed to be easy to read and shareable with the individuals assessed. A brief overview of the components of the individual task results is below.



B. Score Calculations: Standard Scores, Percentiles, and Classifications

Raw scores on each task are compared to the population to be presented as standard scores and percentiles. Scores and percentiles are then adjusted to better reflect the individual's true cognitive ability. To learn more about the raw scores powering these calculations, see Appendix A. To learn more about the Creyos suite of tasks, refer to the [Scientific Overview document](#).

Standard Scores

An individual's raw score on a task is compared with a population of individuals of the same gender and age group. The score is then scaled so that it is compared with a distribution of scores with a mean of 100 and a standard deviation of 15. In other words, an individual with a score of 100 is exactly at the average score for that task, and scores above or below 100 can easily be understood in relation to this average.

All tasks are presented on the same standardized scale, so results can be directly compared to one another.

Percentiles

Percentile information is shown in two ways: i) As a range below the bell curve that represents the distribution of all scores in the population; and ii) As a numerical value representing the midpoint of that range. An individual's raw score, when combined with information from the Creyos normative database, provides information about the range in which the individual's true cognitive ability can be confidently estimated to lie within. The midpoint of the range lines up with the individual's standard score, and provides a single value useful for tracking changes over time.

Score Classifications

Each score is classified as average, below average, or above average based on the standard score. If the individual scored better than 80% or more of the population (a standard score of 113 or more), the score is classified as above average. If the individual scored better than 20% or less of the population (a standard score of 87 or less), the score is classified as below average. Most scores fall in-between, and are classified as average.

B. Score Calculations: Score Adjustments and Trends

Score Adjustments

Both standard scores and percentiles are adjusted to better represent the individual's true cognitive ability. The final adjusted standard score shown on the individual's report and percentile range are based on three pieces of information:

- The individual's raw score on the task.
- The reliability of the task.
- The distribution of scores on the task in the Creyos database.

This final score is adjusted to reduce the role of chance and better estimate the individual's "true" score. In particular, a phenomenon known as *regression to the mean*—the tendency for an extreme score to be closer to the mean the next time the task is taken—is accounted for. Therefore, extremely high or low scores are less likely to be due to chance, particularly on tasks where chance occurrences could potentially result in these outliers. Because the Creyos database contains millions of task scores, these predictions about true cognitive ability are highly accurate and based on known properties of each task.

For more information on the techniques used to estimate standard score and determine the percentile range, and the reasoning behind them, consult the calculation of *standard error of estimation* in the following paper:

- McManus, I. C. (2012). [The misinterpretation of the standard error of measurement in medical education: A primer on the problems, pitfalls and peculiarities of the three different standard errors of measurement.](#) *Medical Teacher*, 34, 569-576.

Trends and the Line of Best Fit

Information about the seven most recent task administrations is available in each report where more than one score for a task is available. Trend reports show information about the standard score, percentile midpoint, score classification, and date of assessment.

A *line of best fit* is calculated using a simple regression model, showing a dotted line that best fits the data. The line does not necessarily pass through every data point, but rather shows the general trend in scores over time.

Note that data points are always matched to age norms as of the date of the task administration. This could have implications for interpreting very long-term trends that involve age—for example, a downward trend over several years would indicate decline *above and beyond* what would be expected due to normal aging.

C. Task Validity Overview and Conditions for Validity

Creyos Health reports include a Validity Indicator—a prominent marker appearing when highly unusual performance patterns are detected. Possible reasons for a potentially invalid result may be that the individual taking the task was distracted, or misunderstood the instructions. When a report indicates that a task is potentially invalid, practitioners often reassess that task after confirming that the individual completed it to the best of their ability and fully understood the rules.

Each task has a set of parameters* that must be met for the score to be considered valid. Utilizing the vast Creyos normative database, validity conditions have been computed for each task, and are shown below for reference:

Feature	Number Ladder	Spatial Span	Token Search	Paired Associates	Rotations	Polygons	Odd One Out	Spatial Planning	Grammatical Reasoning	Digit Span	Feature Match	Double Trouble**
Number of attempts	> 0	> 0	> 0	> 0	≥ 6 and ≤ 38	> 0	≥ 11 and ≤ 39	> 0	> 0	> 0	≥ 11 and ≤ 39	> 0
Number correct	≥ 4 and ≤ 12	≥ 0 and ≤ 8	≥ 1 and ≤ 13	≥ 2 and ≤ 8	≥ 4 and ≤ 32	≥ 6 and ≤ 32	≥ 7 and ≤ 34	≥ 0 and ≤ 19	≥ 5 and ≤ 46	≥ 0 and ≤ 11	≥ 9 and ≤ 33	≥ 9 and ≤ 109
Number of errors	= 3	= 3	= 3	= 3	≥ 0 and ≤ 15	≥ 0 and ≤ 15	≥ 0 and ≤ 16	≥ 0 and ≤ 4	≥ 0 and ≤ 10	= 3	≥ 0 and ≤ 9	≥ 0 and ≤ 37
Duration (seconds)	≥ 60 and ≤ 285	≥ 39 and ≤ 180	≥ 24 and ≤ 751	≥ 60 and ≤ 283	≥ 89.5 and ≤ 90.5	≥ 89.5 and ≤ 90.5	≥ 179.5 and ≤ 180.5	≥ 179.5 and ≤ 180.5	≥ 89.5 and ≤ 90.5	≥ 40 and ≤ 362	≥ 89.5 and ≤ 90.5	≥ 89.5 and ≤ 90.5
Max score	≥ 3 and ≤ 11	≥ 0 and ≤ 9	≥ 2 and ≤ 14	≥ 2 and ≤ 8	----	----	----	----	----	≥ 0 and ≤ 12	----	----
Average Score	≥ 2.7 and ≤ 7.0	≥ 0 and ≤ 6.9	≥ 2.0 and ≤ 9.4	≥ 2.0 and ≤ 5.0	----	----	----	----	----	≥ 0 and ≤ 8.4	----	----
Final Score	----	----	----	----	≥ -11 and ≤ 333	≥ -7 and ≤ 144	≥ -3 and ≤ 37	≥ 0 and ≤ 136	≥ -1 and ≤ 46	----	≥ 6 and ≤ 289	≥ -6 and ≤ 107
Correct Score***	----	----	----	----	≥ 17 and ≤ 373	----	----	----	----	----	≥ 24 and ≤ 323	----
Max Level****	----	----	----	----	≥ 4 and ≤ 19	----	≥ 8 and ≤ 20	----	----	----	≥ 4 and ≤ 17	----

* 99% of scores in the Creyos normative database fall within the bounds of these parameters.

** Double Trouble includes additional measures of validity related to performance on different types of problems, such as 1) Congruent / Congruent (CC) problems, 2) Congruent / Incongruent (CI) problems, 3) Incongruent / Congruent problems (IC) and 4) Incongruent / Incongruent (II) problems. Additional validity metrics for the Double Trouble Task include:

- % of CC problems answered correctly: ≥ 33% and ≤ 100%
- Average reaction time in seconds when attempting CC problems: ≥ 0.73 and ≤ 5.76
- Average reaction time in seconds when attempting CI problems: ≥ 0.75 and ≤ 7.16
- Average reaction time in seconds when attempting IC problems: ≥ 0.74 and ≤ 6.79
- Average reaction time in seconds when attempting II problems: ≥ 0.76 and ≤ 7.18

*** Correct score refers to the sum of all points obtained from successfully completed problems.

**** Max level refers to the difficulty level of the hardest problem successfully completed.

D. Meaningful Change Indicator Overview

Creyos Health reports include a Meaningful Change Indicator, which objectively determines whether the current assessment result represents a potentially significant change relative to two time points: the individual's previous assessment results and the individual's baseline (i.e., initial) assessment results. Conditions for meaningful change for each of these cases are shown below. Meaningful change will only be calculated if a task score is likely valid according to the Validity Indicator. Note that meaningful change is based on raw scores calculated in the background for each task, rather than the adjusted standard scores shown on reports (see Appendix for more information about raw scores).

Conditions for Meaningful Change Relative to Baseline Assessment Results

The top row of this chart indicates the attempt number, and the +/- indicates the required increase or decrease in the raw score for it to be considered potentially meaningful. For instance, if it is the individual's 6th time completing Polygons, they will need to see a raw score increase of 41.83 relative to their baseline assessment result in order for this change to be considered potentially meaningful.

Task	2nd		3rd		4th		5th		6th		7th		8th		9th		10th	
	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
Number Ladder	-0.77	0.77	-0.69	0.90	-0.65	0.91	-0.70	1.00	-0.57	0.92	-0.55	1.00	-0.54	0.97	-0.57	1.13	-0.44	1.07
Spatial Span	-0.84	0.84	-0.81	1.04	-0.65	1.01	-0.75	1.12	-0.62	1.12	-0.84	0.84	-0.49	1.14	-0.96	0.96	-0.62	1.24
Token Search	-1.44	2.14	-1.45	2.26	-1.54	2.22	-1.23	2.27	-1.21	2.45	-0.96	2.35	-1.13	2.76	-0.99	2.78	-0.81	2.60
Paired Associates	-0.76	0.76	-0.64	0.93	-0.70	0.92	-0.64	0.87	-0.63	0.98	-0.63	0.95	-0.55	0.92	-0.78	0.78	-0.71	0.71
Rotations	-49.58	49.58	-45.70	62.34	-32.97	71.80	-31.37	70.18	-30.55	80.19	-25.48	83.30	-30.24	87.76	-33.23	94.94	-29.07	84.34
Polygons	-32.71	32.71	-26.85	37.40	-25.96	38.46	-27.42	42.96	-23.07	41.83	-26.27	44.58	-26.47	49.15	-25.23	53.86	-39.10	39.10
Odd One Out	-6.55	6.55	-6.41	6.41	-6.42	6.42	-5.87	5.87	-5.16	7.29	-4.86	8.09	-4.16	7.50	-4.20	6.93	-6.22	6.22
Spatial Planning	-10.45	19.94	-11.11	24.45	-6.96	25.54	-6.52	29.95	-4.02	31.65	-5.07	32.57	-5.42	34.90	-3.64	36.59	-5.99	37.33
Grammatical Reasoning	-6.58	6.58	-4.70	8.68	-5.13	9.59	-4.96	10.60	-4.90	10.18	-4.61	9.77	-4.33	10.20	-2.39	11.66	-3.88	12.19
Digit Span	-1.24	1.24	-0.90	1.29	-0.91	1.17	-0.79	1.32	-0.74	1.27	-0.82	1.44	-0.76	1.53	-0.74	1.39	-0.77	1.45
Feature Match	-47.44	47.44	-34.23	49.98	-32.73	52.30	-31.81	55.83	-30.48	57.25	-28.51	59.55	-29.79	66.09	-32.31	58.58	-24.44	54.01
Double Trouble	-12.25	22.48	-9.84	26.27	-9.20	29.54	-6.91	33.06	-3.50	33.99	-3.89	35.33	-6.71	39.85	-5.54	40.79	-2.81	45.10

Statistical thresholds used are based on a two-tailed analysis, $p < 0.2$.

D. Meaningful Change Indicator Overview

Conditions for Meaningful Change Relative to Previous Assessment Results

The top row of this chart indicates which attempts are being compared, and the +/- indicates the required increase or decrease in the raw score for the change to be considered potentially meaningful. For instance, if it is the individual's 7th time completing Spatial Span, a raw score increase of 0.75 relative to their previous (i.e., 6th) attempt will be considered potentially meaningful.

Task	1st to 2nd		2nd to 3rd		3rd to 4th		4th to 5th		5th to 6th		6th to 7th		7th to 8th		8th to 9th		9th to 10th	
	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
Number Ladder	-0.77	0.77	-0.76	0.76	-0.68	0.68	-0.75	0.75	-0.78	0.78	-0.63	0.63	-0.79	0.79	-0.79	0.79	-0.62	0.62
Spatial Span	-0.84	0.84	-0.84	0.84	-0.88	0.88	-0.86	0.86	-0.81	0.81	-0.75	0.75	-0.91	0.91	-0.69	0.69	-0.95	0.95
Token Search	-1.44	2.14	-1.47	1.47	-1.48	1.48	-1.31	1.31	-1.23	1.23	-1.21	1.21	-1.18	1.18	-1.38	1.38	-1.03	1.03
Paired Associates	-0.76	0.76	-0.73	0.73	-0.74	0.74	-0.72	0.72	-0.69	0.69	-0.63	0.63	-0.74	0.74	-0.64	0.64	-0.75	0.75
Rotations	-49.58	49.58	-52.09	52.09	-48.17	48.17	-55.22	55.22	-46.38	46.38	-48.31	48.31	-45.05	45.05	-58.69	58.69	-49.69	49.69
Polygons	-32.71	32.71	-32.05	32.05	-32.39	32.39	-30.03	30.03	-29.63	29.63	-30.97	30.97	-34.10	34.10	-33.09	33.09	-31.21	31.21
Odd One Out	-6.55	6.55	-5.99	5.99	-5.59	5.59	-5.46	5.46	-5.36	5.36	-6.37	6.37	-5.36	5.36	-4.55	4.55	-4.47	4.47
Spatial Planning	-10.45	19.94	-14.90	14.90	-13.97	13.97	-15.42	15.42	-14.83	14.83	-14.89	14.89	-15.07	15.07	-13.22	13.22	-16.20	16.20
Grammatical Reasoning	-6.58	6.58	-6.42	6.42	-6.40	6.40	-6.36	6.36	-6.75	6.75	-6.94	6.94	-6.28	6.28	-5.79	5.79	-6.09	6.09
Digit Span	-1.24	1.24	-1.02	1.02	-1.07	1.07	-0.94	0.94	-0.96	0.96	-1.04	1.04	-1.01	1.01	-0.87	0.87	-0.93	0.93
Feature Match	-47.44	47.44	-45.85	45.85	-40.91	40.91	-45.46	45.46	-42.78	42.78	-43.26	43.26	-41.41	41.41	-38.79	38.79	-43.19	43.19
Double Trouble	-12.25	22.48	-16.23	16.23	-14.33	14.33	-15.25	15.25	-13.48	13.48	-14.87	14.87	-13.29	13.29	-12.40	12.40	-14.79	14.79

Statistical thresholds used are based on a two-tailed analysis, $p < 0.2$.

D. Meaningful Change Indicator Overview

How Were Conditions for Meaningful Change Calculated?

The Meaningful Change Indicator compares the difference in an individual's performance on any given task between two time points to the variability in repeated measurements that would occur in the *absence* of meaningful change. The latter is estimated from a sample of healthy control subjects. The reliable change index uses the test-retest reliability and the standard deviation of scores (measured in a control sample) of a task to describe the range of possible differences that occur in repeat task completions. If an individual's change in performance from one time point to another is much larger than what is expected due to chance, then one can conclude that there was meaningful change.

Assessing meaningful change requires that these data have been obtained in a control sample. Creyos has a database of over 8 million test scores, and our normative database consists of more than 75,000 individuals. Most of these individuals have completed every task more than once, and more than 5000 people have completed every task more than 10 times. The interval between self-administered repeated assessments ranges from less than a day, to more than a month. This massive database allows us to characterize in the general population how performance on every task fluctuates naturally across a range of intervals. We are therefore able to quantify the bounds of what constitutes a meaningful change for every task.

E. Common Interpretation Questions

If an individual's scores are unusually low relative to their comparative group, is there cause for concern?

In cases in which a standard score and percentile are extremely low, or lower than expected given knowledge about the individual, a first step may be to discuss with the individual the conditions under which the assessment was completed, and then choose to reassess that individual to obtain additional data. In certain rare instances, the individual may have misunderstood the task instructions or may have been distracted during the assessment, but not enough to result in the task being flagged as potentially invalid. If the task results are *consistently* lower than expected, further investigation may be warranted.

Ultimately, it is up to the practitioner to evaluate all available sources of information to decide the appropriate course of action. This may include continued treatment of the individual within the practitioner's area of expertise, reassessing the individual, or a referral to a neuropsychologist or neurologist.

Is it normal for an individual to score high or normally on most tasks, but very poorly on one or two tasks?

It is normal to have a profile with some high scores and some low scores. Each task is designed to measure distinct cognitive abilities, and tasks within one cognitive domain (e.g., memory) are largely independent from tasks within other cognitive domains (e.g., verbal ability). Therefore, it is normal for individuals to exhibit strengths in one area of cognition, but weaknesses in another. In some cases, very low scores on a handful of tasks may indicate genuine weaknesses in one cognitive domain, and the areas of the brain associated with them. Examine patterns in the scores (e.g., was every low score a memory task?), and examine other information you may have collected to determine the appropriate next steps. In some cases, unusually low scores on a few tasks may simply be due to chance or other irrelevant factors. In these cases, you may choose to re-assess the individual to gather additional data and confirm the cognitive profile of the individual.

Is it possible for individuals to get low scores even if their true capabilities are normal or high?

Obtaining a surprisingly low score is possible, and can arise due to a number of factors, such as individuals being distracted, not understanding the task instructions, malingering (purposefully choosing wrong answers), or being unable to physically complete the assessment. In many of these rare occurrences, Creyos Health may indicate that the task result was likely invalid, but practitioners should still consider them when extremely low scores are obtained.

Is it possible for individuals to get high scores even if their true capabilities are normal or low?

Obtaining an artificially high score is highly unlikely, and only probable in cases where an individual is attempting to manipulate the results (e.g., using a pencil and paper to write down a sequence of digits in Digit Span). Taking the tests in this manner is detectable due to unusual patterns in the data, so in these cases, Creyos Health may indicate that the task result was likely invalid.

E. Common Interpretation Questions

Why do assessment results sometimes fluctuate from one session to the next?

Variability in task results is completely normal, given that cognition naturally varies over time, and a wide variety of random and non-random factors can affect the final score on any neuropsychological task. Adjustment of scores (see Section B) reduces the effect of random factors (e.g., regression to the mean), but a measure of an individual's cognition will change upon repeated task completions as result of practice effects, normal variation in performance, other random sources of error, and perhaps, meaningful change (e.g., deterioration resulting from a clinical condition). This is why the Meaningful Change Indicator is incorporated into Creyos Health. A positive indication suggests that the change in the individual's score is larger than what we would expect to happen due to chance. See Section D for more information about meaningful change.

If a meaningful change is not indicated, does that mean cognition has stayed the same?

Not necessarily. The statistical analysis for meaningful change indicates when a variation in a task result is unlikely if chance alone is operating. It does *not* follow that a lack of significance means there is not meaningful change. Practically, this means that Creyos Health reports are best interpreted as indicating when a change in a task result is unusual. If a result is not unusual, it is not definitive evidence that everything is normal, and small changes may be interpretable in the context of information specific to an individual's history or treatment.

How are the age ranges used for population comparisons determined?

Age ranges were chosen based on two objectives: i) Minimizing significant differences in average performance within age ranges; and ii) Ensuring a large sample size data for that age range (and gender) to accurately compare the individual's score to the database.

Appendix: Raw Score Calculations

Raw scores are not shown on Creyos Health reports, but are shown after each task is completed, and are used in the calculation of percentile ranges, standard scores, meaningful change, and validity. Raw scores are calculated differently for every task—to learn more about the Creyos suite of tasks, refer to the [Scientific Overview document](#). The calculation of each raw score below specifies how performance is measured, before adjusting for age, gender, and other factors.

Untimed Tasks (Digit Span, Number Ladder, Paired Associates, Spatial Span, Token Search)

For untimed tasks, raw score is calculated as the sum of the number of items in all successfully completed problems, divided by the number of successfully completed problems. For example, consider a Digit Span Task attempt in which an individual progresses through the following problems: 4 digits, 5 digits, [wrong answer], 4 digits, 5 digits, 6 digits, [wrong answer], [wrong answer]. The raw score will be the sum of all correctly remembered digits ($4 + 5 + 4 + 5 + 6 = 24$) divided by the total number of correct problems solved (5), resulting in a raw score of 4.8. This methodology provides greater granularity and a truer representation of the individual's performance compared to other more basic scoring methods, such as taking the highest level achieved on the task.

Timed Tasks Without Difficulty Level Adjustments (Double Trouble, Grammatical Reasoning)

For timed tasks in which difficulty does not adjust based on performance, raw score is the number of correct answers minus the number of incorrect answers. For example, a score of 35 on Double Trouble may equate to 40 correct problems and 5 incorrect problems.

Timed Task With Difficulty Level Adjustments (Feature Match, Spatial Planning, Odd One Out, Polygons, Rotations)

Feature Match and Rotations: the raw score is the number of points for correct answers, minus points subtracted for incorrect answers. Points correspond to the number of symbols; that is, problems with more symbols are worth more points than problems with fewer symbols. For example, a correctly answered problem with 6 symbols adds 6 points to the raw score, whereas an incorrectly answered problem with 7 symbols removes 7 points from the raw score.

Spatial Planning: each problem has a minimum number of moves needed to solve it. Points for a problem are equal to twice the minimum number of moves needed, minus the number of moves actually used. The raw score is the total points collected for all solved problems. A problem will end if individuals make twice the number of moves needed, but points are not subtracted for failed problems. For example, completing a problem in 5 moves, when the minimum number of moves required is 3, results in a score of one for that problem: $(2 \times 3 \text{ minimum moves required}) - 5 \text{ moves completed} = 1$.

Odd One Out: the raw score is the number of points for correct answers, minus points subtracted for incorrect answers.

Polygons: the raw score is the number of points for correct answers, minus points subtracted for incorrect answers. More difficult problems, which have more subtle differences between the polygons, are worth more points.

About Creyos

Creyos (formerly Cambridge Brain Sciences) leads the field when it comes to accurately quantifying brain function and brain health. Our proprietary cognitive assessments have been taken millions of times and have been used in over 300 studies published in leading academic journals over the last 30 years. Owing to years of rigorous academic development, Creyos possesses one of the world's largest normative databases of cognitive function developed from 12+ million cognitive task scores. Our cognitive assessments and health questionnaires—all delivered and scored digitally—are used by healthcare practitioners treating mental health conditions, brain injuries, aging, and other patient populations throughout the world, as well as by leading researchers.

Our assessments were developed by Dr. Adrian Owen, chief scientific officer of Creyos, and one of the leading authorities on cognition. Professor Owen is the head of the Owen Lab at the Western Institute for Neuroscience (WIN), a cutting-edge cognitive neuroscience research center at Western University in Ontario, Canada.