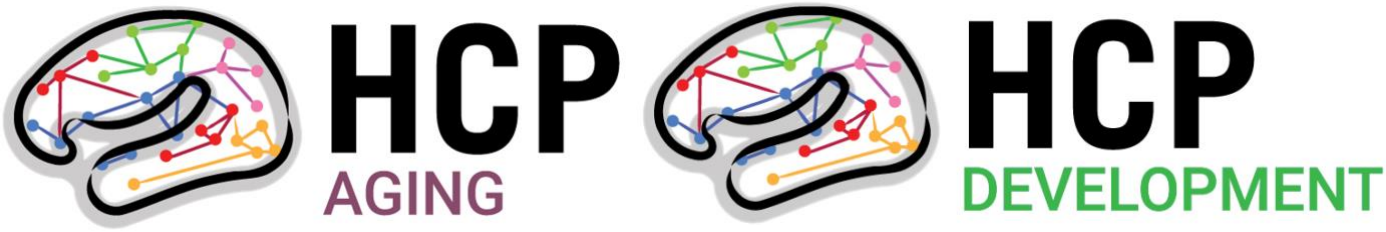

Lifespan HCP 2.0 Data Release Appendix 2:

Details and References for Behavioral & Clinical Instruments

24 February 2021

Table of Contents

Subject Information & Demographics

1.1 – Handedness

Handedness was assessed using the 11-item Edinburgh Handedness questionnaire (Oldfield 1971). An adapted 8-item version of the questionnaire was used for subjects 5- to 10-years old where the child pretended to perform activities that required using one-hand (e.g. using a hammer, using scissors). For subjects ages 11+, a combined Edinburgh handedness score (hcp_handedness_score data element) is also provided for harmonization with HCP-Young Adult and =sum(writing, throwing, scissors, toothbrush, knife_no_fork, spoon, broom, match, box, foot), where values for elements are coded left=-10, strongly prefer left (spl), usually left=-5, no preference =0, strongly prefer right (spr), usually right=5, and right=10. Missing elements are assigned the median of subject's other summands. For subjects 5- to 10-years old, a combined score of the 8-item version (hand_score data element) is provided and =sum (writing, hammer, throwing, toothbrush, hand5, spoon, scissors, hand_15_drink) where values for elements are coded left=0 and right=1. Missing elements are assigned the median of subject's other summands. Item-level responses and scores for this questionnaire can be found in the NDA data structure **edinburgh_hand01**.

Cognitive Measures

2.1 – Cognitive Status

Cognitive status was broadly measured in both studies for participants ages 14+, using the Mini Mental Status Examination (MMSE) or the Montreal Cognitive Assessment (MoCA).

Mini Mental Status Examination (MMSE)

Participants in the Development Study, age 14- to 21-years old, were administered the Mini Mental Status Exam (Folstein et al., 1975) as a broad measure of cognitive status. Participants were excluded if they scored below a 25 out 30 total points at their first visit (Crum et al., 1993). The HCP did not exclude longitudinal participants based on MMSE score on any later study visits.

Description (Folstein et al., 1975):

The first portion of the test has a maximum score of 21 and is analyzing the test takers orientation (10 points), memory (6 points), and attention (5 points). This first portion only requires verbal responses from the test taker. The second portion of MMSE requires a combination of verbal, physical, and written responses from the test taker. This portion of the test aims to analyze the test takers ability to name common items (i.e. “pencil” & “watch”), follow verbal and written commands, spontaneously write a sentence, and copy a complex polygon figure (similar to a Bender-Gestalt figure). The maximum score of this section is 9; leading to a maximum score of 30 for the entirety of the test. The questions/directions of MMSE are always administered in the same order, scored immediately, and the task itself is not timed. Item-level responses and scores for this questionnaire can be found in the NDA data structure **mmse01**.

Montreal Cognitive Assessment (MoCA)

Participants in the Aging Study, ages 36+ years old, were administered the Montreal Cognitive Assessment (Nasreddine et al., 2019) as a broad measure of cognitive status. Participants ages 36-79 years old were excluded if they scored below a 20 out 30 total points. Participants over the age of 79 years old were excluded if they scored below an 18 to account for typical cognitive decline seen in this age range.

Description (Nasreddine et al., 2019):

The MoCA is a test that was developed to screen for mild cognitive impairment (MCI) based on the commonly seen domains of impairment. This test requires an administrator to verbally give directions to a test taker, has a maximum score of 30 points, and is scored across eight cognitive domains: visuospatial abilities, short-term memory, executive function, attention, concentration, working memory, language and orientation. During this task, visuospatial abilities are assessed using a clock-drawing task and a three-dimensional cube copy. Short-term memory recall is assessed with a task which involves two learning trials of five nouns and delayed recall after approximately 5 minutes. Multiple aspects of executive functions are assessed using an alternation task adapted from the Trail Making B task, a phonemic fluency task, and a two-item verbal abstraction task. Attention, concentration, and working memory are evaluated using a sustained attention task (target detection using tapping), a serial subtraction task, and repeating lists of digits forward and backward. Language is assessed using a three-item confrontation naming task with low-familiarity animals (i.e. “lion”, “camel”, “rhinoceros”), exact repetition of two syntactically complex sentences, and the aforementioned fluency task. Finally, orientation to time and place is evaluated via verbal responses from the test taker. Item-level responses and scores for this questionnaire can be found in the NDA data structure **moca01**.

Scoring:

- Visuospatial/Executive – 5 points
- Naming – 3 points
- Attention – 6 points
- Language – 3 points
- Abstraction – 2 points
- Delayed Recall – 5 points
- Orientation – 6 points

2.2 – Language Experience and Proficiency

Language experience and proficiency was measured in all participants using the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007). This questionnaire has been shown to be valid in collecting self-reported proficiency and experience data from bilingual and multilingual participants ages 14 to 80 years old (Kaushanskaya, Blumenfeld & Marian, 2019). This self-reported questionnaire was administered by an RA to participants, in both studies, ages 18+ years old. In the Development study, for participants ages 5 to 17 years old, LEAP-Q was administered by an RA to the parent as a parent-about-child report. Item-level responses for this instrument can be found in the NDA data structure **leap01**.

2.3 – Fluid Intelligence

Fluid intelligence describes an ability to identify the underlying rules and concepts when problem solving (Geary et al., 2019). This study assesses the fluid intelligence of participants in the Development Study (e.g., between the ages of 5 to 21 years old) using the appropriate Wechsler Intelligence Scale as outlined below.

Wechsler Preschool & Primary Scale of Intelligence, Revised (WPPSI-IV-R; age 5)

Description (Pearson Website; Wechsler, 2012):

The WPPSI-IV-R is an intelligence test designed for children ages 2 years 6 months to 7 years 7 months developed by David Wechsler in 1967. This test can be used, as an assessment of general intellectual functioning, as part of an assessment to identify intellectual giftedness, and to identify cognitive delay and learning difficulties. HCP administered WPPSI-R to assess the cognitive development of children ages 5 to 7 years old. This study only used the matrix reasoning task, which involves the participant looking at an array of images with one missing square and identifying the picture that best fits that array from five options, on an iPad, using the platform Q-Interactive on an iPad. Participants were given practice images before beginning the task. Scores for this assessment can be found in the NDA data structure **wppsiiv01**.

Wechsler Intelligence Scale for Children (WISC-V; ages 6-16)

Description (Pearson Website; Wechsler et al., 2014)

WISC-V was used to assess a child's cognitive development, with respect to the child's chronological age. Using such comparisons with other sources of data, the WISC can contribute information concerning a child's development and psychological well-being. Very high or very low scores may suggest contributing factors for adjustment difficulties in social contexts that present problems in accepting such development diversity (or that cannot accommodate more than a certain level of high cognitive functioning). This study only used the Matrix Reasoning subtest, which involves the participant looking at an array of images with one missing square and identifying the picture that best fits that array from five options. Matrix Reasoning measures higher-order executive functioning abstract problem solving, inductive reasoning, and spatial reasoning ability. This task is administered to participants ages 8 to 16 years old via an iPad using a platform called Q-Interactive. WISC-IV can be used to measure non-verbal reasoning skills, broad visual intelligence, and perceptual organization skills. Participants are given practice images before beginning the task. Scores for this task can be found in the NDA data structure **wisc_v01**.

Wechsler Adult Intelligence Scale (WAIS–IV; ages 17-21)

Description (Pearson Website; Wechsler, 2003)

The WAIS-IV is one of the most advanced tests for cognitive ability based on recent research in the cognitive neuroscience areas and is the most recent version of the Wechsler Adult Intelligence Scale, premiering in 2008. The task was administered to participants, ages 16 to 21 years old, via an iPad using a platform called Q-Interactive. HCP only used the matrix reasoning task, which involves the participant looking at an array of images with one missing square and identifying the picture that best fits that array

from five options. WAIS—IV can be used to assess non-verbal reasoning skills, broad visual intelligence, and perceptual organization skills. Participants were given practice images before beginning the task. Scores for this task can be found in the NDA data structure **wais_iv_part101**.

2.4 – Episodic Memory

Episodic memory is a past oriented memory system that is useful for recalling events as they were happening (e.g., like a movie; Wheeler, 2001). The Aging study measures episodic memory using Rey Auditory Verbal Learning Test and the NIH Toolbox Picture Sequence Memory Task (outlined below). The Development study only measures episodic memory using the NIH Toolbox Picture Sequence Memory Task.

Rey Auditory Verbal Learning Test (RAVLT; ages 36+)

Description (Rey 1941):

The Rey Auditory Verbal Learning Test is an easily administered measure of verbal learning and memory, with components related to short-term memory, learning, immediate and delayed recall and recognition memory. Designed to gauge indices such as immediate memory span, learning and learning strategies, retroactive and proactive interferences, and comparison types of errors (e.g. intrusions), the RAVLT provides qualitative measures of learning and memory that are not easily discernible from other tests that approach memory as a unitary entity. The RAVLT consists of a semantically unrelated word list (List A) and a similar interference list (List B). For the first five trials, the subject is prompted with words from List A at the beginning of each trials then asked to recall as many as they can. Novel words not on List A and repetitions are also tracked and scored. On the sixth trial, words from an interference list (List B) are read out loud and then the subject is asked to recall as many words as they can. For the final trial, the subject is asked to recall as many words as they can from List A without hearing the list of words. The interfering list (List A for trial 6, List B for trial 7) is also tracked and scored if the subject recalls any of the interfering words. Many clinicians and researchers regard the RAVLT as a purer measure of verbal memory functioning. This test has been widely used in clinical and research contexts in several countries, being considered as a valid and effective measurement of episodic memory and sensitive to memory deficits found in several clinical conditions such as Alzheimer's disease, Mild Cognitive Impairment and major depression, demonstrating strong criterion validity. This study administered RAVLT using an iPad via the platform of Q-Interactive. Scores for this task can be found in the NDA data structure **ravlt01**.

Interpretation:

1. A low score on immediate free recall of List B words as compared to Trial 1 immediate free recall of List A words may be related to a high degree of proactive interference. Such deficits are often related to retrieval problems and associated with anterior lesions. However, in some cases this may reflect cognitive fatigue, as the repeated trials are very demanding on arousal and attention skills (Rey 1941; Schoenberg et al., 2006; Zhao et al., 2012).

2. A low score on Trial VI for List A words relative to Trial V may be indicative of either high degrees of forgetting during the short delay or retroactive interference or a combination of these

problems. These may reflect subcortical lesions in cases where information cannot be stored over time. A recognition test is useful to see if there is true inability to retain new memories or reflects a retrieval problem as well (Rey 1941; Schoenberg et al., 2006; Zhao et al., 2012).

Picture Sequence Memory Test (ages 5+)

Picture Sequence Memory Test Ages 5-6 Form A v2.0, Picture Sequence Memory Test Age 7 Form A v2.0, Picture Sequence Memory Test Ages 8+ Form A v2.0, Picture Sequence Memory Test Ages 8+ Form B v2.0, Picture Sequence Memory Test Ages 8+ Form C v2.0

Description (NIH Toolbox Website):

The Picture Sequence Memory Test is a measure developed for the assessment of episodic memory for ages 3-85 years. It involves recalling increasingly lengthy series of illustrated objects and activities that are presented in a particular order on the iPad screen, with corresponding audio-recorded phrases played. The participants are asked to recall the sequence of pictures demonstrated over two learning trials; sequence length varies from 6-18 pictures, depending on age. Participants are given credit for each adjacent pair of pictures they correctly place (i.e., if pictures in locations 7 and 8 are placed in that order and adjacent to each other anywhere, such as slots 1 and 2, one point is awarded), up to the maximum value for the sequence, which is one less than the sequence length. (That is, if 18 pictures are in the sequence, the maximum score on that trial is 17 -the number of adjacent pairs of pictures that exist). The test takes approximately seven minutes to administer. HCP administered this task to participants in both studies who were ages 5+, as well as to the parents of participants in the development study. For longitudinal subjects, Form A was used in visit 1 (V1), Form B was used in visit 2 (V2), and Form C was used in visit 3 (V3).

Interpretation (Scoring and Interpretation Guide):

The PSMT is a measure of episodic memory, which involves the acquisition, storage and effortful recall of new information. It is considered a strong "fluid ability" measure, with performance reaching a peak in early adulthood and declining across the life span. Measures of episodic memory such as PSMT can be extremely useful in evaluating performance of those with potential neurological impairments or other health-related problems in which memory is implicated or at risk. One can evaluate all three types of standard scores to interpret individual performance, with higher scores representing better episodic memory within the normative standard being applied (i.e., in relation to the general child or adult population, or in relation to age peers, or in relation to overall demographically comparable peers). Scores for this task can be found in the NDA data structure **psm01**. Scores derived from all NIH Toolbox Cognition tasks can be found in the NDA data structure **cogcomp01**.

Key Points:

- Measures episodic memory using sequences of pictured objects and activities in a particular order
- Participants put the pictures back into the sequence that was shown
- Examiner's may need to remind participants to watch the screen

- Examiner's should not allow participants to touch the screen while the pictures are being presented
- 3-7 have 2 practice sequences, older participants have only one practice sequence

2.5 – Executive Function/Cognitive Flexibility

Executive function is thought as the ability to think before acting, focus attention, remember instructions, and resisting temptations (Diamond, 2013). Executive function relies partially on **cognitive flexibility** which has been thought of as the ability to thinking outside the box, seeing things from another perspective, and the ability to apply different rules to responses when required (Diamond, 2013). Both the Aging and Development studies assessed cognitive flexibility by using the NIH Toolbox – Dimensional Change Card Sort Task (outlined below).

Dimensional Change Card Sort Test

Dimensional Change Card Sort Test Ages 3-7 v2.0, Dimensional Change Card Sort Test Ages 8-11 v2.0, Dimensional Change Card Sort Test Age 12+ v2.0

Description (NIH Toolbox Website):

DCCS is a measure of cognitive flexibility. Two target pictures are presented that vary along two dimensions (e.g., shape and color). Participants are asked to match a series of bivalent test pictures (e.g. yellow balls and blue trucks) to the target pictures, first according to one dimension (e.g., color) and then, after a number of trials, according to the other dimension (e.g., shape). "Switch" trials are also employed, in which the participant must change the dimension being matched. For example, after four straight trials matching on shape, the participant may be asked to match on color on the next trial and then go back to shape, thus requiring the cognitive flexibility to quickly choose the correct stimulus. This test takes approximately four minutes to administer and is recommended for ages 3-85. HCP administered this task to participants in both studies who were ages 5+, as well as to the parents of participants in the development study.

Interpretation (Scoring and Interpretation Guide):

The DCCS is a measure of executive function, specifically tapping cognitive flexibility. It is considered a "fluid ability" measure, like Flanker, with performance generally increasing through childhood and then declining across the adult age span. To interpret individual performance, one can evaluate all three types of normative scores, where higher scores indicate higher levels of cognitive. Scores for this task can be found in the NDA data structure **dccs01**. Scores derived from all NIH Toolbox Cognition tasks can be found in the NDA data structure **cogcomp01**.

Key Points:

- Measure of executive function that assesses cognitive flexibility and attention
- 2 target pictures that vary along 2 dimensions: shape and color
- Participants match sets of 2 test pictures that differ in color to the target pictures – the relevant dimension for sorting is indicated by a cue word – "shape" or "color"

- Practice items: White and brown/rabbits and sailboats
- Test items: yellow and blow/balls and trucks
- All participants use their index finger to choose which of the two pictures matches the relevant aspect of the stimulus
- Participants return their index finger back to home base between test trials
- Instructions and test trials differ for ages 3-7 and 8-85
- 8+ are presented with a mixed block of 30 shape or color items and learn about home base before learning about the test
- 3-7 learn about home base before the mixed trials test, because they are scored for reaction time

2.6 – Executive Function/Inhibition

Executive function is generally thought of as the ability to think before acting, focus attention, remember instructions, and resisting temptations (Diamond, 2013). Executive function relies partially on **inhibition** which has been thought of as the ability for self-control, resisting temptation, and resisting acting impulsively (Diamond, 2013). Both the Aging and Development studies assessed inhibition by using the NIH Toolbox – Flanker Inhibitory Control and Attention Task (outlined below).

Flanker Inhibitory Control and Attention Test

Flanker Inhibitory Control and Attention Test Ages 3-7 v2.0, Flanker Inhibitory Control and Attention Test Ages 8-11 v2.0, Flanker Inhibitory Control and Attention Test Age 12+ v2.0

Description (NIH Toolbox Website):
The Flanker task measures both a participant's attention and inhibitory control. The test requires the participant to focus on a given stimulus while inhibiting attention to stimuli (fish for ages 3-7 or arrows for ages 8-85) flanking it. Sometimes the middle stimulus is pointing in the same direction as the "flankers" (congruent) and sometimes in the opposite direction (incongruent). Twenty trials are conducted for ages 8-85; for ages 3-7, if a participant scores ≥90% on the fish stimuli (with no more than one congruent and one incongruent trial incorrect), 20 additional trials with arrows are presented. The test takes approximately three minutes to administer. HCP administered this task to participants in both studies who were ages 5+, as well as to the parents of participants in the development study.

Interpretation (Scoring and Interpretation Guide):
The Flanker is a measure of executive function, specifically tapping inhibitory control and attention. It is considered a "fluid ability" - the capacity for new learning and information processing in novel situations-measure, in which performance reaches a peak in early adulthood, then tends to decline across the life span (based on health and individual factors, of course). To interpret individual performance, one can evaluate all three types of normative scores, in which higher scores indicate higher levels of ability to attend to relevant stimuli and inhibit attention from irrelevant stimuli. Scores for this task can be found in the NDA data structure **flanker01**. Scores derived from all NIH Toolbox Cognition tasks can be found in the NDA data structure **cogcomp01**.

Key Points:

- Measure of executive function that measures attention and ability to inhibit automatic responses that may interfere with achieving goals
- The test requires participants to focus on a particular stimulus while inhibiting attention to the flanking stimuli
- There are two types of trials: congruent and incongruent
- These trials are mixed
- There is on block of 20 test trials for all participants

2.7 – Language/Reading Decoding

Oral Reading Recognition Test (ages 5+)

Oral Reading Recognition Test Age 3+ v2.0

Description (NIH Toolbox Website):
Separate but parallel reading tests have been developed in English and Spanish. In either language, the participant is asked to read and pronounce letters and words as accurately as possible. The test administrator scores them as right or wrong. For the youngest children, the initial items may require them to identify letters (vis-à-vis non-letter symbols) and to identify a specific letter in an array of four symbols. The test is given via CAT and requires approximately three minutes. This test is recommended for ages 7-85, but is available for use as young as age 3, if desired. Normative scores are available for ages 3-6, but the test may not be appropriate for all children, especially those who cannot yet identify any letters of the alphabet. HCP administered this task to participants in both studies who were ages 5+, as well as to the parents of participants in the development study.

Interpretation (Scoring and Interpretation Guide):
The Reading Test is a measure of reading decoding skill and, like vocabulary, is considered among the crystallized abilities; those abilities are generally more dependent upon past learning experiences and consistent across the adult life span. To interpret individual performance, one can evaluate all three types of standard scores plus the percentiles; higher scores indicate better reading ability within the normative standard being applied. Scores for this task can be found in the NDA data structure **orrt01**. Scores derived from all NIH Toolbox Cognition tasks can be found in the NDA data structure **cogcomp01**.

Key Points:

- Measures a participant's ability to read and pronounce letters and words
- For the youngest children, initial items require identification of letters in an array of 4 letters
- The test adjusts the difficulty level of items depending on the participant's performance
- The number of items presented depends on the participant's age and performance, but will approximately be 25 for most
- One set of instructions for all
- Initial items are chosen based on education and age

- Not timed

2.8 – Language/Vocabulary Comprehension

Picture Vocabulary Test (ages 5+)

Picture Vocabulary Test Age 3+ v2.0

Description (NIH Toolbox Website):

This measure of receptive vocabulary is administered in a computerized adaptive format. That is, the next question a participant receives depends on his/her response to the previous question; Computer Adaptive Testing (CAT) ensures a test that is tailored to the participant's needs. The respondent is presented with an audio recording of a word and four photographic images on the iPad screen, and is asked to select the picture that most closely matches the meaning of the word. This test takes approximately four minutes to administer and is recommended for ages 3-85. HCP administered this task to participants in both studies who were ages 5+, as well as to the parents of participants in the development study. Scores for this task can be found in the NDA data structure **tpvt01**. Scores derived from all NIH Toolbox Cognition tasks can be found in the NDA data structure **cogcomp01**.

Interpretation (Scoring and Interpretation Guide):

The TPVT is a measure of general vocabulary knowledge and is considered to be a strong measure of crystallized abilities (those abilities that are more dependent upon past learning experiences and are relatively consistent across the adult life span).

Key Points:

- Measure of receptive vocabulary
- Presented with 4 pictures, hears audio recording saying
- Participant is instructed to select the picture that most closely resembles the meaning of the word
- Test uses a variable length CAT with a max of 25 items, some will have fewer
- Words presented depend on performance

2.9 – Processing Speed

Pattern Comparison Processing Speed Test (ages 5+)

Pattern Comparison Processing Speed Test Ages 3-6 v2.0, Pattern Comparison Processing Speed Test Ages 7+ v2.0

Description (NIH Toolbox Website)

This test measures speed of processing by asking participants to discern, as quickly as possible, whether two side-by-side pictures are the same or not. The items are presented one pair at a time on the iPad screen, and the participant is given 85 seconds of response time (excluding any time needed for the given iPad to "load" the items) to respond to as many items as possible (up to a maximum of 130). The items

are designed to be simple so as to most purely measure processing speed. Overall, the test takes approximately three minutes to administer. This test is recommended for ages 7-85. HCP administered this task to participants in both studies who were ages 5+, as well as to the parents of participants in the Development study.

Interpretation (Scoring and Interpretation Guide):

The Pattern Comparison Test is a measure of speed of processing, which typically improves steadily (time to complete task decreases) throughout childhood and adolescence, then begins to decline in adulthood, becoming much slower in older adults. As such, it is considered a "fluid ability" measure. Slowed processing speed has been associated with normal aging, with decreases in processing speed being a significant contributor to age-related decline in other cognitive domains. Processing speed declines have also been found to impact several aspects of mental functioning in older age groups, including driving skills. Processing speed has also been shown to be highly vulnerable to brain damage, and multiple clinical populations demonstrate diminished processing speed. Assessments of processing speed have been found consistently to be among the most sensitive of neuropsychological measures (along with measures of episodic memory); typically, measures of processing speed are able to differentiate between clinical groups and healthy groups. Scores for this task can be found in the NDA data structure **pcps01**. Scores derived from all NIH Toolbox Cognition tasks can be found in the NDA data structure **cogcomp01**.

Key Points:

- Designed to measure processing speed
- Participants are required to discern whether 2 pictures are the same or not
- Younger children either choose a smiley face or frown face
- Yes or No buttons
- There are different practice items and instructions for young children and older
- Practice items are not timed
- During testing, examiner can give participants prompts to stay on task

2.10 – Working Memory

List Sorting Working Memory Test (ages 5+)

List Sorting Working Memory Test Ages 3-6 v2.0, List Sorting Working Memory Test Age 7+ v2.0

Description (NIH Toolbox Website):

The List Sorting test requires immediate recall and sequencing of different visually and orally presented stimuli (i.e., "working memory"). Pictures of different foods and animals are displayed with accompanying audio recording and written text (e.g., "elephant"), and the participant is asked to say the items back in size order from smallest to largest, first within a single dimension (either animals or foods, called 1-List) and then on two dimensions (foods, then animals, called 2-List). The test takes approximately seven minutes to administer and is recommended for ages 7-85, though a supplemental test designed for ages 3-6 is also available, if desired. This supplemental test also provides normative scores, though it may not

appropriate for all children in this age range (especially 3-year-olds). HCP administered this task to participants in both studies who were ages 5+, as well as to the parents of participants in the Development study.

Interpretation (Scoring and Interpretation Guide):
List Sorting is a measure of working memory, tapping both information storage and processing (manipulation). It is considered a "fluid ability" measure, with performance tending to peak in early adulthood and then declining across the life span. Scores for this task can be found in the NDA data structure **lswmt01**. Scores derived from all NIH Toolbox Cognition tasks can be found in the NDA data structure **cogcomp01**.

Key points:

- Assesses working memory or the capacity to process information across a series of tasks and modalities
- It requires the participant to sequence sets of visually and orally presented stimuli in size order from smallest to largest
- Pictures of different foods and animals are displayed with both a sound clip and written text
- Has 2 different conditions: 1 – list and 2 – list
  - 1 list – participant orders a series of objects, either a food or animals, in size order from smallest to largest
  - 2 list – participant is presented with both food and animals and first orders the food from smallest to largest and then orders the animals from smallest to largest

2.11 – Self-regulation/Impulsivity

Delay Discounting (ages 5+)

Description:
The Delay Discounting Task asks participants to choose whether they would want some money now or more money at a later date (e.g., "Would you rather receive $20 today or $100 in one week?"). HCP administered this task via a computer program on UPenn CNP to participants in both studies (i.e., ages 5+). Before the task, an RA read directions from the computer screen to the participant, administered practice questions, and asked if the participant has any questions to ensure participants' understanding of the task. After this practice, the task was completed by the participant without further clarification from the RA. Although they do not receive any additional compensation from this task, participants are asked to pretend that all money amounts are really being offered to them.

This task uses an adjusting-amount approach where delays are fixed and reward amounts are adjusted on a trial-by-trial basis based on participant's choices, to rapidly hone in on an indifference point. In the Development study an RA helps participants who are between the ages 5 and 7 complete the task by reading out the options to them and then selecting what the participant would prefer on the computer. Upon completion of the task, participants between the ages of 5 and 7 are also given a set of questions

to see their understanding of the task (e.g., "Which is more $20 or $50?" or "Which is longer, 1 week or 1 year?").

Scoring:

Discounting is illustrated by the fact that humans, and other animals (Green et al., 2007), will often choose a smaller immediate reward over an objectively larger, but delayed reward. Self-regulation and impulsivity are measured by how long participant can wait to receive a larger reward rather than take an immediate reward. Delayed discounting has been associated with impulsivity and has been shown as a risk factor for pathological gambling, substance abuse, and cigarette use (Hariri et al., 2006). This task measures discounting by calculating the area under the empirical discounting function (as seen in Myerson, Green & Warusawitharana, 2001). This method allows for a measure of discounting that is not tied to theoretical framework, and is simple/univariate (Myerson, Greene & Warusawitharana, 2001). Scores for this task can be found in the NDA data structure **deldisk01**.

Key Points:

- Measures participants' discounting by calculating area under an empirical discounting function
- Requires participant to choose between hypothetical amounts of money
- After directions and practice, hypothetical money amounts and wait time to receive are only displayed via text on the computer screen (unless the participant is between the ages 5 and 7 where their options are also read aloud to them)

2.12 – Executive Function/Switching

Trails A & B (ages 36+)

Description (Reitan, 1986):

The Trail Making Test is a neuropsychological test of visual attention and task switching. It consists of two parts in which the subject is instructed to connect a set of 25 dots as quickly as possible while still maintaining accuracy. The test can provide information about visual search speed, scanning, speed of processing, mental flexibility, as well as executive functioning. It is sensitive to detecting cognitive impairment associated with dementia, for example, Alzheimer's disease. There are two different versions of this task; Trail A and Trail B. Both trails were administered at each visit for subjects in the Aging study (i.e., ages 36+).

Method and Interpretation (Reitan, 1986):

The task requires a subject to connect a sequence of 25 consecutive targets on a sheet of paper or computer screen, in a similar manner to a child's connect-the-dots puzzle. There are two parts to the test: in the first, the targets are all numbers (1, 2, 3, etc.) and the test taker needs to connect them in sequential order; in the second part, the subject alternates between numbers and letters (1, A, 2, B, etc.). If the subject makes an error, the test administrator corrects them before the subject moves on to the next dot.

The goal of the test is for the subject is to finish both parts as quickly as possible, with the time taken to complete the test being used as the primary performance metric. The error rate is not recorded in the paper and pencil version of the test; however, it is assumed that if errors are made it will be reflected in the completion time. The second part of the test, in which the subject alternates between numbers and letters, is used to examine executive functioning. The first part is used primarily to examine cognitive processing speed. Scores for this task can be found in the NDA data structure **trail_ca01**.

	Perfectly Normal	Normal	Mild/Moderately Impaired	Moderately/Severely Impaired
Trails A	0-26	27-39	40-51	52+
Trails B	0-65	66-85	86-120	121+

Emotion Measures

3.1 – Negative Emotions

NIH Toolbox Emotion Surveys - Negative Affect

Negative affect is a phrase used to describe unpleasant feelings or emotions that exist on a continuum ranging from common and normal feelings of sadness, fear and anger to more extreme feelings along the same continuum. Negative affect is understood as comprising important underlying dispositions (e.g., neuroticism, negative emotional style) and more transient negative feeling states. The focus of this subdomain is on three principal negative emotions: sadness, fear and anger.

3.1.a – NIH Toolbox – Sadness

Sadness is distinguished by low levels of positive affect and comprised of symptoms that are primarily affective (poor mood) and cognitive (negative perceptions of self, the world and the future) indicators of depression.

Sadness Survey (ages 5+ and parents)

Description (Toolbox Website)

HCP administered this survey as a self-report measure for participants ages 18+ using a CAT format for participants in both studies as well as to parents of participants in the Development study. HCP also administered the survey as a self-report 8-item fixed-length form for participants ages 8-17; a parent-about-child report measure for participants ages 8-12 using a CAT format; and a parent-about-child report version for participants ages 3-7 that is a 7-item fixed-length form. Each survey version assessed negative mood, negative views of the self, and negative social cognition.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Sadness Survey, higher scores are indicative of more sadness. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for the parent-about-child report survey and the Development study self-report survey (ages 8-17) can be found in the NDA data structure **tlbx_sadness01**. Scores for the self-report administered to participants in both studies ages 18+ can be found in the NDA data structure **predd01**.

3.1.b NIH Toolbox – Fear

Fear is best characterized by symptoms of anxiety that reflect autonomic arousal and perceptions of threat.

Overview of Surveys:

For adult (ages 18+) self-report, fear is measured by the NIH Toolbox Fear-Affect Survey and the NIH Toolbox Fear-Somatic Arousal Survey; for child (ages 8-17) self-report, the NIH Toolbox Fear Survey is used. For parent report, NIH Toolbox Fear-Over Anxious Survey (ages 5-7) and NIH Toolbox Fear-Separation Anxiety Survey are used for ages 5-7 and the NIH Toolbox Fear Survey for ages 8-12.

Fear-Over Anxious Survey (ages 5-7)

Description (Toolbox Website)

This is a parent-about-child report measure for ages 3-7, assessing fear, worry and hyperarousal. It is a 6-item fixed-length form. HCP administered this survey to parents of participants ages 5 to 7.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Fear-Over Anxious Survey, higher scores are indicative of more parent-reported child fear, worry and hyperarousal. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **tlbx_fearanx01**.

Fear-Separation Anxiety Survey (ages 5-7)

Description (Toolbox Website)

This is a parent-about-child report measure for ages 3-7, assessing fear of being separated from home and from loved ones. It is a 7-item fixed-length form. HCP administered this survey to parents of participants ages 5 to 7.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Fear-Separation Anxiety Survey, higher scores are indicative of more parent-reported child separation anxiety. Additional work remains to examine the predictive and

concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **tlbx_fearanx01**.

Fear Survey (ages 8-17)

Description (Toolbox Website)

HCP administered this survey as a self-report measure for ages 8-17 as well as a parent-about-child report measure for ages 8-12. It assesses fear, anxious misery and hyperarousal. The self-report version is an 8-item fixed-length form; the parent-about-child report version for ages 8-12 uses a CAT format.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Fear Survey, higher scores are indicative of more child fear. Scores 1 SD or more below the mean (T ≤ 40) suggest low levels of fear and scores 1 SD or more above the mean (T ≥ 60) suggest high levels of fear. T-scores ≥ 60 may warrant heightened surveillance or concern. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **tlbx_fearanx01**.

Fear-Affect *S*urvey (ages 18+ and parents)

Description (Toolbox Website)

This self-report measure assesses fear and anxious misery. HCP administered this survey to participants in both studies ages 18+, using a CAT format. Parents of participants in the Development study also completed this survey as a self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Fear-Affect Survey, higher scores are indicative of more feelings of fear (fearfulness, panic). Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **preda01**.

Fear-Somatic Arousal Survey (ages 18+ and parents)

Description (Toolbox Website)

This self-report measure assesses somatic symptoms related to arousal. HCP administered this survey to participants in both studies ages 18+, using a 6-item fixed-length form. Parents of participants in the Development study also completed this survey as a self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Fear-Somatic Arousal Survey, higher scores are indicative of more somatic arousal. Additional work remains to examine the predictive and concurrent validity of these

measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **preda01**.

3.1.c NIH Toolbox – Anger

Anger is characterized by attitudes of hostility and cynicism and is often associated with experiences of frustration impeding goal-directed behavior.

Overview of Surveys:

For adult self-report, anger is comprised of three components: anger as an emotion, aggression as a behavioral component, and hostility as a set of cynical attitudes and mistrust of others and their motives. Anger is measured by the NIH Toolbox Anger-Affect Survey, NIH Toolbox Anger-Hostility Survey and the NIH Toolbox Anger-Physical Aggression Survey. For children, anger is measured by the NIH Toolbox Anger Survey.

Anger Survey (ages 5 - 17)

Description (Toolbox Website)

HCP administered this survey as a self-report measure for participants ages 8-17 as well as a parent-about-child report measure for participants ages 3-12. It assesses angry mood and aggression (verbal and physical). The self-report version is a 6-item fixed-length form; the parent-about-child report version for ages 8-12 uses a CAT format; and the parent-about-child report version for ages 3-7 is a 9-item fixed-length form.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Anger Survey, higher scores are indicative of more child anger. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **prang01**.

Anger-Affect Survey (ages 18+ and parents)

Description (Toolbox Website)

This self-report measure assesses anger as an emotion. HCP administered this survey to participants in both studies ages 18+, using a CAT format. Parents of participants in the Development study also completed this survey as a self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Anger-Affect Survey, higher scores are indicative of more feelings of anger (irritability, frustration). Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **prang01**.

Anger-Hostility Survey (ages 18+ and parents)

Description (Toolbox Website)

This self-report measure assesses attitudes of hostility and cynicism. HCP administered this survey to participants in both studies ages 18+ using a 5-item fixed-length form. Parents of participants in the Development study also completed this survey as a self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Anger-Hostility Survey, higher scores are indicative of more hostility. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **prang01**.

Anger-Physical Aggression Survey (ages 18+ and parents)

Description (Toolbox Website)

This self-report measure assesses aggression as a behavioral component. HCP administered this survey to participants in both studies ages 18+ using a 5-item fixed-length form. No versions for other ages are available. Parents of participants in the Development study also completed this survey as a self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Anger-Physical Aggression Survey, higher scores are indicative of more reported physical aggression. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **prang01**.

Negative Affect Surveys Key Points:

- Comprised of 3 components: Anger, Fear, and Sadness
  - Anger has 3 aspects
    1. Anger as an emotion
    2. Aggression as a behavioral component
    3. Hostility as a set of cynical attitudes and mistrust
  - Fear is characterized by symptoms of anxiety
  - Sadness is characterized by low levels of positive affect
    - Comprised of symptoms that are primarily affective and cognitive, which are indicators of depression

3.2 – Positive Emotions

NIH Toolbox Emotion Surveys **- Psychological Well-being**

Psychological well-being includes both hedonic and eudaimonic aspects of well-being. Hedonic aspects are more subjective and experiential and emphasize pleasure and positive affect (happiness, serenity and cognitive engagement). Eudaimonic well-being is more evaluative in nature and emphasizes fulfillment and purpose (e.g., meaning, life satisfaction). Psychological well-being composite scores (general life satisfaction, meaning and purpose, and positive affect) for ages 8+ and parents can found in the NDA data structure **tlbx_wellbeing01**.

3.2.a NIH Toolbox - Positive Affect

Positive Affect refers to feelings that reflect a level of pleasurable engagement with the environment, such as happiness, joy, excitement, enthusiasm and contentment. It is measured by the NIH Toolbox Positive Affect Survey.

Positive Affect Survey (ages 5+ and parents)

Description (Toolbox Website)

This measure assesses both activated (i.e., happiness, joy) as well as inactivated (i.e., serenity, peace) aspects of positive affect. HCP administered this survey as a self-report to participants in both studies. Separate CAT versions were used for respondents ages 13-17 and 18+; a 9-item fixed-length self-report form was used for ages 8-12. CAT versions of this survey were also administered as a parent-about-child report when participants were ages 5-7 and 8-12. Parents of participants in the Development study also completed this survey as a self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Positive Affect Survey, higher scores are indicative of more positive affect. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can be found in the NDA data structure **tlbx_wellbeing01**.

3.2.b NIH Toolbox – Life Satisfaction

Life Satisfaction is one's cognitive evaluation of life experiences; this measure is concerned with whether people like their lives or not. Life satisfaction includes both general (e.g., my life is going well) and domain-specific (e.g., I am satisfied with my family life) aspects.

General Life Satisfaction Survey (ages 5+ and parents)

Description (Toolbox Website)

This self-report measure assesses global feelings and attitudes about one's life. HCP administered this report as a CAT is used for adult participants in both studies (ages 18+), a separate CAT version is used for participants ages 13-17, and a 5-item fixed-length form is used for ages 8-12, as well as for the parent-about-child report version when participants are ages 5-12. Parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox General Life Satisfaction Survey, higher scores are indicative of more general life satisfaction. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Item-level data for this survey can found in the NDA data structure **tlbx_wellbeing01**.

3.2.c NIH Toolbox - Meaning and Purpose

Meaning and Purpose is characterized by the extent to which people feel their life matters or makes sense.

Meaning and Purpose Survey (ages 18+ and parents)

Description (Toolbox Website)

This is a self-report measure HCP administered to participants in both studies ages 18+ as a CAT. Parents of participants in the Development study also completed this self-report survey.

Interpretation (Scoring and Interpretation Guide):

Higher scores indicate more self-reported meaning and purpose. Additional work remains to examine the predictive and concurrent validity of these measures and to identify clinically meaningful thresholds. Scores for this survey can found in the NDA data structure **tlbx_wellbeing01**.

Positive Emotion Surveys Key Points:

- Comprised of 4 components: Positive Affect, Life Satisfaction, and Meaning and Purpose
  - Positive affect – feelings that reflect a level of pleasurable engagement with the environment
  - Emotion Control – one's ability to regulate frequency and intensity of positive and negative emotions
  - Life Satisfaction – one's cognitive evaluation of life experiences
  - Meaning and Purpose – the extent to which people feel their life matters or makes sense

3.3 – Emotion Recognition

Penn Emotion Recognition Test (ER40; ages 8+)

Description (Gur et. al., 2010):

The ER40 is a measure of emotion recognition. HCP administered ER40 via a computer program on UPenn CNP to participants ages 8+ in both studies. During the task participants are shown a series of 40 faces, one at a time, and asked to determine what emotion the face is showing for each trial. There are 5 answer choices: happy, sad, anger, fear and no emotion.

Participants respond to each trial by clicking with the mouse on the word describing the emotion each faces expresses. There are 4 female faces for each emotion (4 x 5 = 20) and 4 male faces for each emotion

(4 x 5 = 20). ER40 is derived from the University of Pennsylvania Emotion Recognition Task 96 faces version, and balanced for equality and intensity of emotion, age, gender and ethnicity.

Practice Trial(s): The participant practices with one emotional face displaying intense anger.

- Presentation: 1 question.
- Description: one question where the participant must choose "Anger".
- Time: forced-choice - the participant must click with the mouse on one of the emotion-description buttons; otherwise, the question will remain on the computer screen.
- Feedback: correct or incorrect feedback pages are presented after the practice question. Both correct and incorrect feedback pages let the participant know that anger was the correct answer choice. If the participant chose an incorrect emotion, he/she will return to the practice question and answer it again. The incorrect feedback page is the same and will continue to show until the participant answers the practice question correctly. Once the correct answer is given, the participant will move to the test trials

Test Trials: 40 questions for which the participant determines the emotion each face is showing.

- Presentation: 40 questions, randomized.
- Description: the participant must choose, by clicking with the mouse, one of the emotions to the right of the page labeled happy, sad, anger, fear and no emotion for each of the 40 test trials.
- Time: forced-choice - the participant must click with the mouse on one of the emotion descriptions buttons; otherwise, the question will remain on the computer screen.
- Feedback: none

Interpretation:

Rules & Variables: The scores are based on the number of correct responses for female versus male faces; the number of correct happy, sad, anger, fear and no emotion faces; the number of false positives for happy, sad, anger, fear and no emotion faces; and the number of mild and number of intense emotion expressions correctly identified. Median response times are given for all of these categories. Scores for this task can be found in the NDA data structure **er4001**.

3.4 – Positive and Negative Affect

Positive and Negative Affect Schedule (PANAS; ages 8+)

Description (Watson et al., 1988):

The Positive and Negative Affect Schedule – short form (PANAS-S) used is a self-administered, self-report 20-item Likert scale assessment that measures degree of positive or negative affect. PANAS-S is administered to participants online (cross-sectional participants) or via an iPad (longitudinal participants at visit 2) on the REDCap platform in both studies during their follow-up surveys (approximately 12-15 months after their original participation in the study). Participants are asked to rate 10 adjectives that measure positive feelings such as joy or pleasure (e.g. "Interested"), and 10 adjectives that measure

negative feelings, such as anxiety or sadness, on a scale of how closely the adjective describes them on average at present. The response options are on scale of 1 to 5 (where 1= "Very Slightly or Not at all" and 5= "Extremely"). In the Development study ages 8+. This assessment is also given to parents as a parent-about-child report for participants ages 8-17.

Interpretation:

Responses to questions are then scored for a Positive Affect Score as well as a Negative Affect Score. Higher scores correlate with higher ratings of Positive/Negative affect. Scores for this instrument are not available in the HCP Release 2.0, but will be available in the NDA data structure **panas01.**

3.5 – Impulsivity

UPPS-P Impulsive Behavior Scale (ages 5-21)

Description (from PhenX):

The UPPS-P Impulsive Behavior Scale for Children (Zapolski et al. 2010) includes 40 self-administered items that are scored as five subscales (urgency, premeditation, perseverance, sensation seeking, and positive urgency). These dimensions reflect distinct personality traits that lead to impulsive-type behavior. Participants ages 8-17 completed the report in a self-report survey. Parents of participants ages 5-17 also completed the UPPS-P as a parent-about-child report. During this survey, the respondent is asked to read each item and circle the number on a scale that best describes him/her (or his/her child). The 4-item Likert scale includes values for 1 (not at all like me), 2 (not like me), 3 (somewhat like me), and 4 (very much like me). The survey was taken from PhenX, and was added to the REDCap server so that it could be administered via iPad. Responses to questions from this instrument can be found in the NDA data structure **upps01**.

3.6 – Behavioral Inhibition and Activation

Behavioral Inhibition/Behavioral Approach System (BIS/BAS; ages 5-21)

Description:

Several theorists have argued that two general motivational systems underlie behavior. A behavioral approach system (BAS) is believed to regulate appetitive motives, in which the goal is to move toward something desired. A behavioral avoidance (or inhibition) system (BIS) is said to regulate aversive motives, in which the goal is to move away from something unpleasant (Carver & White, 1994). The survey was taken from PhenX, and was added to the REDCap server so that it could be administered via iPad. The child version of the behavioral inhibition system (BIS)/behavioral activation system (BAS) scale is a 24-item questionnaire that measures individual differences in motivation. This scale is based on the original BIS/BAS system, which was later revised by Gray and McNaughton (2000). This survey is self-reported for participants in the Development study ages 8 to 21; and is also administered as a parent-about-child report for participants ages 5 to 17. Respondents indicate how much they agree or disagree with the statements on a four-point scale (very true to not true).

Scoring (from PhenX):

Scores can be derived for three behavioral activation system subscales (Drive, Fun Seeking, and Reward Responsiveness) and one behavioral inhibition system subscale. Various psychological disorders-such as attention-deficit and hyperactivity disorder, bipolar disorder, and substance abuse-may result from an imbalance in one motivational system or the other (Bijttebier et al., 2009). Responses to questions from this instrument can be found in the NDA data structure **bisbas01**.

3.7 – Temperament and Risk-taking Behavior

Children's Behavior Questionnaire (CBQ; ages 5-8)

Description (Rothbart et al., 2001):

The CBQ is a caregiver/parent-about-child report assessing the temperament of the child (Capaldi & Rothbart 1992). This report can be used to describe the temperament of children between the ages of 3 and 7. Individual differences are assessed on 15 primary temperament characteristics which can be found in Rothbart et al. (2001). Factor analyses of CBQ scales reliably (including cross-culturally) recover a three-factor solution indicating three broad dimensions of temperament: Extraversion/Surgency, Negative Affectivity, and Effortful Control. The CBQ in reference to this study is given to parents of participants ages 5 to 8. The survey is administered via an iPad on the REDCap platform. Parents are given descriptions (e.g. "Gets quite frustrated when prevented from doing something s/he wants to do.") and asked how well the statement describes their child.

Scoring:

Ratings are given on a 7-point scale where 1= "extremely untrue of your child" 4= "neither true nor false of your child" and 7= "extremely true of your child." There was also an N/A option for situations that did not apply to their child. Responses are then scored based on the parents' response to the questions, and separated into the separate dimensions of temperament analyzed by CBQ (e.g., Activity Level). Scores for this instrument can be found in the NDA data structure **cbq01**.

Strengths and Difficulties Questionnaire (SDQ; ages 5-12)

Description:

The Strengths and Difficulties Questionnaire (SDQ) has been shown to be a reliable (Goodman, 1997; Goodman, Meltzer, & Bailey, 1998, Goodman, 2001) and brief behavioral screening questionnaire about 3- to 16-year-olds. It exists in several versions to meet the needs of researchers, clinicians and educationalists. Includes between 1-3 of components. We only used the first component: 25 items on psychological attributes that make up 5 scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behavior. SDQ was given as a parent-about-child report for participants ages 5 to 12. Parents were given descriptions (e.g., "Often complains of headaches, stomach-aches or sickness") and asked how well that describes their child. This report was administered to respondents via an iPad survey from the REDCap platform.

Scoring:

Ratings are given via the options "Not true", "Somewhat true", and "Certainly True". These ratings are then scored across the domains outlined above (e.g., emotional symptoms, conduct symptoms, etc.). Responses to questions from this instrument can be found in the NDA data structure **sdq01**.

Early Adolescent Temperament Questionnaire – Revised (EATQ-R; ages 9-17)

Description:

As can be found on the PhenX website, the Early Adolescent Temperament Questionnaire (EATQ-R) is a 65-item self- or parent-about-child report questionnaire that measures the 11 dimensions of temperament (activation control, activity level, affiliation, attention, fear, frustration, high-intensity pleasure, inhibitory control, perceptual sensitivity, pleasure sensitivity, and shyness) and the 2 dimensions of behavior (aggression and depressive mood). Participants are asked to read each description of behavior and indicate how many times in the past week the behavior was exhibited (never, very rarely, less than half the time, half the time, more than half the time, almost always, always). EATQ-R has been shown to be a reliable measure of temperament (Capaldi & Rothbart, 1992). This study used both parent-reported EATQ-R (ages 9-17) and the self-reported EATQ-R (ages 9-15) versions for participants ages 9 to 17. This report was administered to respondents via an iPad survey from the REDCap platform. Responses to questions from this instrument can be found in the NDA data structure **eatq01**. Note: for the LS 2.0 Release, the eatq01 data structure has rows for parent about child data for subjects aged 5-8, but no parent data was collected for these ages.

3.8 – Personality

NEO-FFI Short Form (ages 16+)

Description:

The HCP uses the 60-question version of this questionnaire. It is a quick, reliable, and accurate measure of the 5 broad traits of personality: neuroticism, extraversion, openness, agreeableness, conscientiousness. These 5 broad traits have been generally accepted as describing the personality (Goldberg, 1993; McCrae & Costa, 2004; Costa & McCrae, 2012). NEO was obtained as a self-report from Development participants ages 16+ and all Aging participants ages 36+. This report was administered to participants via an iPad survey from the REDCap platform. Responses to questions from this instrument can be found in the NDA data structure **nffi01**.

3.9 – Mania Symptoms

General Behavior Inventory (GBI; ages 5-17)

Description:

This survey was given to the parent of participants in the Development study ages 5-17 as a parent-about-child report. GBI is a 10-item mania scale assessing symptoms such as elated mood, high energy, irritability, and rapid changes in mood and energy as indicators of potential juvenile bipolar disorder

(Youngstrom et al., 2008). GBI contains questions about experiences (e.g., "Have there been periods of several days or more when your child's friends or other family members told you that your child seemed unusually happy or high - clearly different from his/her usual self or from a typical good mood?") and asks the parent to rate how often these experiences occur on a 4-point scale from "Never or hardly ever" to "Very often." This report was administered to participants via an iPad survey from the REDCap platform. Responses to questions from this instrument can be found in the NDA data structure gbi01.

7 Up 7 Down Inventory (ages 12-21)

Description:

The 7 Up 7 Down Inventory is a self-reported measure of manic and depressive tendencies carved from the full-length General Behavior Inventory (Youngstrom et al., 2013). This instrument contains 14 questions about experiences (e.g., "Have there been times lasting several days or more when you felt you must have lots of excitement, and you actually did a lot of new or different things?") and asks the participant to rate how often these experiences occur on a 4-point scale from "Never or hardly ever" to "Very often." HCP administered the inventory as a self-reported survey to participants in the development study ages 12 to 21. This report was administered to participants via an iPad survey from the REDCap platform. Responses to questions from this instrument can be found in the NDA data structure **gbi01**.

Motor Measures

4.1 – Endurance

NIH Toolbox - 2-minute Walk Test (ages 5+ & Parents)

Description (Toolbox Website):

This test is adapted from the American Thoracic Society's 6-Minute Walk Test Protocol. This test measures sub-maximal cardiovascular endurance by recording the distance that the participant is able to walk on a 50-foot (out and back) course in two minutes. The participant's raw score is the distance in feet and inches walked in two minutes. The test takes approximately four minutes to administer (including instructions and practice). This test is recommended for ages 3-85. Able participants in both the Aging and Development Studies (i.e., ages 5+) were given this task, as were all able parents from the Development Study.

Interpretation (Scoring and Interpretation Guide):

Cardiorespiratory and muscle endurance are important components of physical fitness and contribute to both performance and health status. On the 2-Minute Walk Endurance Test, greater distance walked is suggestive of better endurance. People with better endurance are able to complete daily tasks and are more fit to pursue leisure activities and accomplish higher-intensity workloads. The clinical significance of endurance as measured by timed walk tests to morbidity and mortality outcomes has been extensively reported in healthy and clinical populations across the age span. Results can be found in the NDA data structure **tlbx_motor01**.

Key Points:

- Assesses stamina and provides an assessment of a person's overall cardiovascular endurance
- A lap is every time the participant rounds the cone
- Measurements are taken from the back of the participant's feet to the preceding 10ft marker

4.2 – Locomotion

NIH Toolbox- 4-meter Walk Test (ages 7+ & Parents)

Description (Toolbox Website):

This test is adapted from the 4-meter walk test in the Short Physical Performance Battery. Participants are asked to walk a short distance (four meters) at their usual pace. Participants complete one practice and then two, timed trials. Raw scores are recorded as the time in seconds required to walk 4 meters on each of the two trials, with the better trial used for scoring. The test takes approximately 19 three minutes to administer (including instructions and practice). This test is recommended for ages 7-85. Able participants (ages 7+) in both the Aging and Development Studies were given this task, as were all able parents from the Development Study.

Interpretation (Scoring and Interpretation Guide):
On the 4-Meter Walk Gait Speed Test, higher computed scores are indicative of better gait speed (i.e., fewer seconds to walk four meters). One can evaluate the descriptive statistics in the NIH Toolbox Technical Manual to get a sense of how individual or group performance compares to results obtained from the national norming sample, though care should be exercised in specific interpretation. Gait speed as a measure of bipedal locomotion is both a good way to summarize the overall burden of disease as well as a generic indicator of health status, prognosis and the co-morbid burden of disease in older persons. The speed at which older individuals walk is relevant to their functioning in the community. Moreover, gait speed is an important predictor of outcomes such as: length of stay and discharge disposition of patients admitted for acute rehabilitation after stroke, mortality, incident ischemic stroke and incident dementia. Results can be found in the NDA data structure **tlbx_motor01**.

Key Points:

- Assesses the usual walking speed of participants
- 1 practice trial and 2 timed trials
- Timing starts when the participant steps over the starting line
- If the participant stumbles or tries to run, void the trial and ask the participant to do another trial.
- Timing Ends when one of the participant's feet is completely across the finish line

4.3 – Dexterity

NIH Toolbox - 9-hole Pegboard Test (ages 5-21 & Parents)

Description (Toolbox Website):

This simple test of manual dexterity records the time required for the participant to accurately place and remove nine plastic pegs into a plastic pegboard. The protocol includes one practice and one timed trial with each hand. Raw scores are recorded as time in seconds it takes the participant to complete the task with each hand (separate score for each). The test takes approximately four minutes to administer and is recommended for ages 3-85. Participants in the Development Study (ages 5+) and their parents were given this task.

Interpretation (Scoring and Interpretation Guide):

Dexterity is a central component of hand function and relates to both the speed and accuracy of hand movements during the manipulation of objects. For the 9-Hole Pegboard Dexterity Test, the raw score is commonly used for interpretation, with faster completion times (less time to complete) representing better manual dexterity. This also allows for raw score comparisons between dominant and non-dominant hand performance. Results be found in the NDA data structure **tlbx_motor01**.

Key Points:

- Assesses fine motor dexterity
- Timed with an on-screen stop watch
- One set of instructions for all ages
- Practice and Test for both hands, both are timed
- Hands start flat on the peg board and the table at the countdown

4.4 – Strength

NIH Toolbox - Grip Strength Dynamometry Test (ages 5+ & Parents)

Description (Toolbox Website):

This protocol is adapted from the grip strength testing protocol of the American Society of Hand Therapy. Participants are seated in a chair with their feet touching the ground. With the elbow bent to 90 degrees and the arm near the trunk, wrist at neutral, participants squeeze the Jamar Plus Digital dynamometer as hard as they can for a count of three. The dynamometer provides a digital reading of force in pounds. A practice trial at less than full force and one test trial are completed with each hand. The test takes approximately three minutes to administer and is recommended for ages 3-85. Able participants in both the Aging and Development Studies (ages 5+) were given this task, as were all able parents from the Development Study.

Interpretation (Scoring and Interpretation Guide):

Muscle strength is an essential element for humans to move against gravity and provide sufficient force to perform movements within the full range of motion. For the Grip Strength Test, the raw score has

commonly been used for interpretation, with greater force (in pounds) representing greater strength. This also allows for raw score comparisons between dominant and non-dominant hand performance. However, one can also evaluate performance with the dominant and/or non-dominant hand by looking at the normative standard scores provided. More generally, grip strength has been used to characterize total body strength and predict mortality, postsurgical complications and future disability. Muscle strength of the limbs and trunk declines with age and is associated with an increased risk of falls, hip fractures, loss of bone mineral density, long-term survival in severe congestive heart failure, functional dependence in people aged 75 years or older, and loss of functional status in hospitalized patients. Results can be found in the NDA data structure **tlbx_motor01**.

Key Points:

- Measure of how strong a person's hand grip is, specifically this test provides a good estimate of upper body muscle strength
- Dynamometer is set on the second notch for all participants
- First practice and test are done with the dominant hand. The second practice and test are done with the non-dominant hand

Sensory Measures

5.1 – Audition

Audition (hearing) is the processing of sound in the environment. It is necessary for navigating in the environment and communicating with others. Acoustic information is processed through three groups of peripheral structures (outer, middle and inner ears) and then through the central auditory nervous system to create auditory experience.

NIH Toolbox - Words-in-noise Test (ages 5+)

Description (Toolbox Website)

This test measures a person's ability to recognize single words presented amid varying levels of background noise. It measures how much difficulty a person might have hearing in a noisy environment. A recorded voice instructs the participant to listen to and then repeat words. The task becomes increasingly difficult as the background noise gets louder, thus reducing the signal-to-noise ratio. The test is recommended for participants ages 6-85 and takes approximately Six minutes to administer. HCP administered the task to participants in both studies (ages 5+), as well as to all able parents from the Development Study.

Interpretation (Scoring and Interpretation Guide):

Assessment of the ability to understand speech in a noisy background yields an ecologically valid measure of hearing because a substantial portion of communication in the real world occurs in less-than-ideal environments. Moreover, speech perception in noise is often difficult to predict from pure-tone thresholds or from speech perception in quiet settings. The range of possible

scores for each ear is -2.0 to 26.0 dB S/N, with lower scores indicative of better performance and, conversely, higher scores potentially suggestive of hearing difficulties. For score interpretation with ages 13 and above, a cutoff of 10 dB S/N is recommended for the NIH Toolbox version of this measure. Participants with a score higher than this cutoff should follow up with a hearing professional, specifically an otolaryngologist, who would then refer to an audiologist as needed. Results can be found in the NDA data structure **tlbx_sensation01**.

Key Points:

- Quantifies a participant's ability to hear words in a noisy environment
- Ages 6+
- Seven levels of word/noise ratio and 5 words at each level
- Speaker's voice gradually gets softer until it is the same level of the background noise
- Each ear tested separately
- If all 5 words are incorrect, it will switch to the other ear
- Hearing aids should be removed
- 2 versions: 3-9, and 10+

5.2 – Vision

Vision is a complex sensation that provides us with a personal conscious representation of our surrounding environment. Loss of vision or blindness may limit a person's ability to complete normal, daily activities and decrease overall quality of life. Visual impairment can impose various limitations on a person's functional ability, including reading, mobility (which includes driving), visual information processing (also called "seeing"), and visually guided motor behavior (also called "manipulation"). One core NIH Toolbox vision measure is available with national normative scores provided.

NIH Toolbox - Visual Acuity Test (ages 5+)

Description (Toolbox Website)

This test directly measures participants' visual acuity or distance vision. The participant is seated 3 meters away from the iPad screen at eye level, and letters (called "optotypes") are displayed one at a time on the screen for the participant to identify, using both eyes at the same time, with the participant wearing his/her normal corrective lenses for distance vision (glasses or contact lenses), if worn. As the participant successfully identifies optotypes of a given size, smaller ones appear on the screen, until the software ascertains the smallest-size optotype the participant can successfully see. Conversely, the program displays larger optotypes if the participant cannot see the size that is first displayed, until a size that he/she can accurately see is found. For participants ages 3-7, only the letters H, O, T and V are used, and children may point to a laminated card showing the letters if they cannot verbalize or recall the letter names. For participants ages 8 and above, the entire set of optotypes is used, following a common protocol used in professional vision testing. This test takes approximately three minutes to administer and is recommended for

ages 3-85. HCP administered the task to participants in both studies (ages 5+), as well as to all able parents from the Development Study.

Interpretation (Scoring and Interpretation Guide):

The Visual Acuity Test provides a reliable measure of participants' overall functional distance vision since it measures both eyes simultaneously. In everyday life, a Snellen equivalent of 20/40 or better (including corrective lenses) is typically a requirement of obtaining a motor vehicle license. Federal government definitions of "legally blind" refer to corrected vision in the individual's better eye of 20/200 and worse. Results can be found in the NDA data structure **tlbx_sensation01**.

Key Points:

- Provides a quick and accurate assessment of distance vision
- Children and older adults have different sets of optotypes or letters
- Participants sit 3 meters away from an iPad at eye level
- Participants that normally wear glasses for distance vision should use them
- Children Ages 3-7
    - can point to a card to show what letter they saw on the screen (H O T V)
    - There is a training where the child is encouraged to name the letters. Letter naming is not required
    - There is also a practice where a set of 4 letters is presented at 20/400. If the child does not succeed after 3 times, testing is discontinued
- Ages 8+ the card is not used
- If the participant requires a very strong correction but does not have their corrective lenses, the test is likely invalid
- If the participant does wear lenses but are not as strong and forgot them, the examiner makes a judgement call
- The letter sets are different for younger and older participants
- The test progressively presents smaller letters if the participant is correct, and larger letters if they are wrong, until an acuity score is established and the test ends

5.3 – Olfaction

The primary purpose of the **olfactory** system in humans is to detect and perceive volatile airborne chemicals and thus provide information about our environment and food quality that is critical to our health, a nutritious diet and psychological well-being.

NIH Toolbox - Odor Identification Test (ages 5-21 & parents)

Description (Toolbox Website)

This task assesses a person's ability to identify various odors. Participants use scratch-and-sniff cards and after scratching them one at a time, are asked to identify which of four pictures on the iPad screen matches the odor they have just smelled. Participants ages 10-85 are administered

nine odor cards, while those ages 3-9 are administered five odor cards. Child participants (ages 3-9 years) are first asked to identify the eight pictures used as answer choices to ensure they can complete the task. Having identified the pictures, they are asked if they have tasted or smelled the objects or foods depicted. This test takes approximately four to five minutes to administer and is recommended for ages 3-85. HCP only administered measurements for olfaction to participants in the Development Study (ages 5+) as well as to the parents of participants in the Development study.

Interpretation (*Scoring and Interpretation Guide*)*:*
Olfactory testing can be a useful adjunct to comprehensive assessments of health and well-being. For example, impaired olfactory function is now recognized as one of the hallmark early signs of several neurodegenerative disorders, including Alzheimer's and Parkinson's disease. Nevertheless, individuals can vary widely in their ability to detect, recognize and identify odors, yet still be within the range of normal function. When evaluating normative standard scores for the Odor Identification Test, higher scores indicate better olfactory ability/functioning. Results can be found in the NDA data structure **tlbx_sensation01**.

Key points:

- Assesses the ability to identify various odors
- 2 versions: 3-9, and 10+
- Ages 3-9
    - Examiner scratches card for child and asks them to smell and identify 5 odors
    - Before identifying the odors, child is asked to identify 8 pictures that are used as answer choices
    - When the child identifies the pictures, the child is asked if he/she has smelled the objects or foods shown
- Ages 10+
    - Participants scratch cards and identify 9 odors

5.4 – Pain

Pain is an important component of health and function. Pain has been defined as an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage. Pain is a major symptom in many medical conditions and can significantly interfere with a person's quality of life and general functioning. The Toolbox measures of pain focus on a participant's reported pain experience, as well as the intensity of the pain experienced. Both measures used are derived from the published NIH PROMIS scales. Toolbox norms and scale scores are not available for the pain measures; however, descriptive statistics obtained from the sample of participants who were administered the pain scales during the Toolbox norming study are available in the NIH Toolbox Technical Manual.

NIH Toolbox - Pain Intensity Survey (ages 18+)

Description (Toolbox Website)

This measure consists of a single item measuring immediate (i.e., acute) pain in adults. It asks a participant to rate level of pain experienced "over the last seven days." It takes less than one minute to administer and is recommended for ages 18-85. HCP administered this survey to participants in both the Development and Aging studies (ages 18+) as well as to the parents of participants in the Development study.

Interpretation (Scoring and Interpretation Guide):

One could reasonably expect a large proportion of the normal population to obtain scores of zero on this measure. Regardless, it is an easily quantifiable piece of information on one's subjective pain experience. Additional work remains to examine the predictive validity of this measure and to identify any clinically meaningful thresholds. Results can be found in the NDA data structure **tlbx_sensation01**.

NIH Toolbox - Pain Interference Survey (ages 18+)

Description (Toolbox Website)

This brief self-report scale measures the degree to which pain interferes with other activities in life in adults. Pain interference items were developed as part of the NIH PROMIS. This measure is administered as a CAT and takes approximately three minutes. It is recommended for 18-85. HCP administered this survey to participants in both the Development and Aging studies (ages 18+) as well as to the parents of participants in the Development study.

Interpretation (Scoring and Interpretation Guide):

The pain interference item bank measures the self-reported consequences of pain on relevant aspects of one's life. This includes the extent to which pain hinders engagement with social, cognitive, emotional, physical and recreational activities. Higher theta and T-Scores represent greater participant report of pain interference in daily life. Thus, T-Scores ≥ 60 may be of concern. Additional work remains to examine the predictive validity of this measure and to identify any clinically meaningful thresholds. Results can be found in the NDA data structure **tlbx_sensation01**.

5.5 – Color Vision

Farnsworth Color Vision Test (ages 5-21)

Description:

HCP assessed color vision of Development study participants, ages 5 to 21, using the Farnsworth Test, a valid and reliable measure that provides more quantitative information than the commonly used Ishihara Test (Cole 2007). In this task, participants order 15 colored blobs as a function of what they think are the closest matching colors. Based on the results, participants are classified as having Normal color vision, Protan (reduced sensitivity to red light), Deutan (reduced sensitivity to green light) or Tritian (reduced

sensitivity to blue light) color vision problems. Results can be found in the NDA data structure **vision_tests01**.

5.6 – Contrast Sensitivity

Mars Contrast Sensitivity Test (ages 5-21)

HCP assessed contrast sensitivity, of participants in the Development Study, ages 5 to 21, using the Mars Contrast Sensitivity Test (Arditi 2005), a brief, valid and reliable measure that improves upon the traditional Pelli-Robson measure (Dougherty et al. 2005; Haymes et al. 2006; Thayaparan et al. 2007). Subjects aged 5- to 7-years old were asked to identify all of the letters used on a form before taking the test. If the subject was unable identify any of the letters, the test was skipped. Results can be found in the NDA data structure **vision_tests01**.

Physical Health and Medical History

6.1 – Pubertal Development

Pubertal Development Scale (PDS; ages 5-21)

Description (Petersen et al., 1988):

PDS has been shown to be a reliable self-reported measure of pubertal development in young adolescents. The survey begins with the statement, "Now I'm going to ask you some questions about physical development." During this survey participants are asked a rather global question about pubertal change. These characteristics included growth spurt in height, pubic hair, and skin change for both boys and girls; facial hair growth and voice change in boys only; and breast development and menarche in girls only.

HCP administered PDS as a self-response survey on REDCap to participants in the Development Study ages 9+, and as a parent-about-child report for Development participants ages 5-13. Since this survey is administered in the lab, younger participants are able to ask questions in order to clarify phrasing of questions (e.g., survey respondents are able to ask an RA to describe a growth spurt).

Scoring (Petersen et al., 1988):

These sets of items form scales consisting of five items for each gender, coded on a four-level ordinal response scale. An overall pubertal development score was computed by summing across the five items to obtain a total score. Results can be found in the NDA data structure **pds01**.

Morris Udry Puberty Survey (ages 5-21)

Description (Morris & Udry, 1980):

The Morris Udry Puberty Survey is a pubertal development survey that includes descriptions of the 5 stages of development as well as drawings made from Tanner's photos of the 5 stages of development

to clarify to participants what some of the question answers are. This has been shown to be a reliable measure of self-reported pubertal stages. This survey was administered on the iPad via REDCap as a self-report to participants in the Development Study ages 9+, and administered to the parent as a parent-about-child report for Development participants ages 5-13. Note: for the LS 2.0 Release, the pds01 data structure has rows for parent about child data for subjects aged 14-17, but no parent data was collected for these ages.

Scoring:
There are only two questions for this survey which are then summed together for a final score (e.g., a response of pubertal hair growth stage 1 followed by a response of stage one for breast development would result in a total score of 2). Results can be found in the NDA data structure **pds01**.

6.2 – History of Endocrine Disorders

(Females ages 8+) Female participants ages 8+ in both studies are asked in a self-report on the iPad via REDCap if they have ever been diagnosed with an endocrine disorder (e.g, hypothyroidism) and, if so, what age they were at the onset of the disorder. For female Development participants ages 5-13, the parents of participants were asked this question about their child in a parent-about-child report. Responses from participants in the Aging study can be found in the NDA data structure **mchq01**; responses for participants in the Development study can be found in the NDA data structure **pds01**.

6.3 – Menstrual Cycle

(Females ages 8-21) Information about menstrual cycle is obtained for all female participants in both Development and Aging studies via a survey on REDCap.

In the Development Study, questions about the child's menstrual cycle are covered in Pubertal Development Scale (PDS; Petersen et al., 1988). PDS is a self-report survey for participants ages 8+ and a parent-about-child report for participants ages 5-13. Responses to the PDS can be found in the NDA data structure **pds01**.

STRAW-10 Survey (females ages 36+)

Female participants in the Aging Study were asked about their menstrual cycle via the self-report survey Stages of Reproductive Aging Workshop (STRAW-10; Harlow et al., 2012). Answers to the STRAW-10 survey are used to calculate an overall code for menopause stage for aging female participants (Harlow et al., 2012). Codes are explained fully at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3580996/and reported in the straw_code data element as 1.11 = Postmenopausal: plus 1a; 1.12 = Postmenopausal: plus 1b ; 1.13 = Postmenopausal: plus 1c; 1.14 = Postmenopausal: plus 2; 1.21 = Postmenopausal; 1.22 = No periods for 12 months other reason; 2 = Late Transition; 2.2 = Skipped periods not necessarily due to natural menopause; 3 = Early Transition ; 4.1 = Late Reproductive (i.e., late premenopausal); 4.2 = Reproductive(i.e., premenopausal).

Responses to the Straw-10 survey and straw_code can be found in the NDA data structure **mchq01**.

6.4 – Developmental History

Parents of participants ages 5-17 in the Development Study were administered a parent-about-child survey on REDCap of participant developmental history including substance use during pregnancy, medical conditions during pregnancy, frequency of prenatal care, birth complications, and relative motor and speech development (Kessler et al., 2009a; Kessler et al., 2009b; Merikangas et al., 2009; Barch et al., 2018). This is the same instrument as used in the ABCD study (although with different exact variable names), ABCD NDA data structure **dhx01**. Responses can be found in the NDA data structure **mab01**.

6.5 – Parental History of Psychiatric or Neurological Disorders

Participants were asked if their mother or father have these conditions: schizophrenia or psychosis, depression, bipolar, anxiety that needed treatment, drug or alcohol problems, Alzheimer's disease or dementia, Parkinson's disease, or Tourette's syndrome is obtained via a survey on REDCap. This was a shortened two-question version of the Family History Assessment Module Screener (FHAM-S; Rice et al., 1995) from the National Consortium on Alcohol and Neurodevelopment in Adolescence(NCANDA) study Brown et al., 2015. FHAM-S has been shown to allow family history information to be used both as a proxy for a non-interviewed relative, as well as a second source of information to be used in the analysis of genetic family data (Rice et al., 1995). HCP obtained this information for all participants in the Development Study as a self-reported survey for participants ages 18-21 and as a parent-about-child report for participants ages 5-17. Responses can be found in the NDA data structure **fenvs01**.

6.6 – Medical History

Medical History Survey (ages 5-17)

Parents of participants ages 5-17 in the Development Study were administered a parent-about-child survey on REDCap of participant medical history including health professional visits, chronic conditions, illnesses, emergency room visits, and hospitalizations (Barch et al., 2018). Derived from the MAGIC Health Services Utilization Questionnaire (Todd, et al., 2003). This is the same instrument as used in the ABCD study (although with different exact variable names), ABCD NDA data structure **abcd_mx01**. This survey was administered on an iPad via REDCap. Responses can be found in the NDA data structure **mab01**.

Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA; ages 35+)

See below in 7.3 description and references. SSAGA includes questions about medical history and chronic conditions. **SSAGA medical history data has not been mapped to NDA data structures and is therefore not available in the Lifespan 2.0 Release.**

6.7 – Head Trauma

Boston Assessment of Traumatic Brain Injury-Lifetime Questionnaire (BAT-LQ, ages 36+)

Description:

Assess head trauma throughout the lifetime (Corrigan & Bogner, 2007; Bogner et al., 2017; Fortier et al., 2014). Participants in the aging study (ages 36+) were asked if they have ever experienced a blow to the head or been hospitalized or seen in a hospital emergency room for various reasons. If yes, they were then asked if they "lost consciousness or were knocked out," and "were dazed or had a memory gap." The BATL-Q uses a hybrid classification system for the diagnosis of TBI according to VA and DoD consensus criteria for TBI severity. Mild TBIs are further specified as mild Grade I, II, or III injuries. This survey was administered on an iPad via REDCap. Responses can be found in the NDA data structure **batbil01**.

Psychopathology Clinical Assessments

7.1 – Achenbach assessments

Achenbach Child Behavior Checklist (CBCL; ages 5-17)

Description (Achenbach, 2009):

The CBCL is the original ASEBA instrument on which other forms have been modeled. It is a 4-page form to be completed by a parent or parent surrogate for children ages 4 to 18. Parents provide information for 20 social competence items, covering their child's sports participation, hobbies and activities, social organizations, jobs and chores, friendships, relationships with other people, ability to play and work alone, and school functioning. In addition, page 2 provides open-ended items for describing illnesses and disabilities, what concerns the parent most about the child, and the best things about the child. Parents' reports of their child's favorite activities, the degree and quality of the child's involvement in activities, friendships, and the best things about the child provide practitioners with details of the child's strengths for discussion in clinical interviews with the parents and child. Parents also rate their child on 118 specific problem items, such as acts too young for age, cries a lot, cruel to animals, gets in many fights, sets fires, and unhappy, sad, or depressed. Open-ended items are also provided for the respondent to add if the child has physical problems without known medical cause, and/or other problems that are not specifically described on the CBCL. Each item is rated on a 0-1-2-point scale (not true, somewhat or sometimes true, very true), based on the child's functioning during the preceding 6 months. The CBCL is completed by most parents in about 15 to 20 minutes. HCP administered the CBCL to parents as a parent-about-child report for participants ages 5 to 17 on an iPad via REDCap.

Interpretation (Achenbach, 2009):

The CBCL is scored on separate profiles for boys and girls for ages 5 to 11 and 12 to 17. The profile provides scores for total competence, 3 competence scales (Activities, Social, and School), plus total problems, internalizing, externalizing, and 8 syndrome scales (withdrawn, somatic complaints, anxious/depressed, social problems, thought problems, attention problems, delinquent behavior, and aggressive behavior). The syndrome scales were derived from principal components analyses of forms

completed by parents for 4455 children referred for mental health services. For more information on the CBCL see the ASEBA website. Responses for participants that are 5 years old can be found in the NDA data structure **cbcl1_501**; responses for participants that are ages 6 to 17 can be found in the NDA data structure **cbcl01**.

Achenbach Youth Self-Report (YSR; ages 11 - 17)

Description (Achenbach, 2009):

The YSR is a self-rating form for youths ages 11 to 17. It requires a mental age of 10 years and fifth-grade-level reading skills. However, if reading skills are below fifth-grade level, the YSR can be read aloud to the respondent. The YSR has most of the same social competence and problem items as the CBCL, but the items are stated in the first person. Sixteen CBCL problem items considered inappropriate to ask youths were deleted and replaced with 16 socially desirable items that enable respondents to say something favorable about themselves. The favorable items are omitted from the Total Problem score. The remaining 102 YSR specific problem items all have counterparts on the CBCL. Each item is rated on a 0-1-2-point scale (not true, somewhat or sometimes true, very true), based on the child's functioning during the preceding 6 months. The YSR takes about 15 to 20 minutes to complete and is administered in a survey on an iPad via REDCap.

Interpretation (Achenbach, 2009):

The YSR is scored on separate profiles for boys and girls ages 11 to 18. The profile provides scores for Total Competence and 2 competence scales (Activities and Social), plus Total Problems, Internalizing, Externalizing, and 8 syndrome scales comparable to those scored on the CBCL. The YSR syndrome scales were derived from principal components analyses of forms completed by 1272 youths referred for mental health services. For more information on the YSR see the ASEBA website. Responses can be found in the NDA data structure **ysr01**.

Achenbach Adult Self Report (ASR; ages 18 – 59 and parents)

Description (Achenbach, 2009):

The ASR used for HCP is a self-reported survey completed by the participant and has 110 specific problem items and 13 socially desirable items. The ASR can be completed in about 15 to 20 minutes and is administered in a survey on an iPad via REDCap.

Scoring (ASEBA Website):

The scoring the ASR include raw and normed scales (T-scores and percentiles) for empirically based syndromes (Internalizing: Anxious/Depressed, Withdrawn, Somatic Complaints, Thought Problems, Attention Problems, Externalizing: Aggressive Behavior, Rule-Breaking Behavior, Intrusive), Internalizing, Externalizing, and Total Problems. In addition, the ASR profiles feature DSM-oriented scales consisting of items that experts from many cultures identified as being very consistent with DSM-5 categories (Depressive Problems, Anxiety Problems, Somatic Problems, AD/H Problems, Antisocial Personality). The profiles also include a critical items scale consisting of items of particular concern to clinicians. Scale

scores (T-scores and percentiles) in relation to norms for each gender at ages 18-35 and 36-59, are based on national probability samples. Item-level responses and scores can be found in the NDA data structure **asr01**.

Older Adult Self-Report (OASR; ages 60+)

Description (ASEBA Website):

The OASR obtains older adults' self-reports of diverse aspects of adaptive functioning and problems. This survey is different than ASR in that is asks questions more relevant to participants ages 60+ such as: psychiatric and psychological evaluations; medical care, including routine care and evaluation of functioning following events such as strokes, falls, and illnesses; following significant life changes, such as loss of a loved one, moves to retirement communities, assisted living, and nursing homes; and evaluations before and after planned changes and interventions (Achenbach et al., 2004). The OASR also looks at the following scales: adaptive functioning (e.g., friends, personal strengths, etc.), syndrome scales (e.g., anxious/depressed, functional impairment, etc.), and DSM-oriented scales (e.g., depressive problems, antisocial personality problems, etc.). HCP administers OASR via a survey on the iPad to participants ages 60+. Each item is rated on a 0-1-2-point scale (not true, somewhat or sometimes true, very true), based on the adult's functioning during the preceding 6 months. The OASR takes about 20-30 minutes to complete depending on the age of the participant and is administered in a survey on an iPad via REDCap. Responses can be found in the NDA data structure **asr01**.

7.2 – Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS; ages 5-21)

Description (Barch et al., 2018b)

KSADS is a measure of current and past psychopathology. HCP administered the KSADS-COMP v1.1 NIMH03.31.2018 (version info) instrument on www.ksads.net on an iPad to participants ages 12-21 as a self-report and to the parents of participants ages 5-17 as a parent-about-child report. This is the same version of KSADS-COMP used in the ABCD Study *(Barch et al., 2018b)*. The survey has been shown to be a reliable and valid measure of psychopathology in children and adolescents (Orvaschel et al., 1982, Chambers et al., 1985, Kaufman et al., 1997, Kaufman et al., 2000). HCP used the updated 2013 version of KSADS that also assess DSM-5 diagnoses, and is administered as a computerized assessment (Kaufman et al. 2013). Dr. Kaufman and Dr. Kobak created three computerized versions (KSADS-COMP): a traditional clinician administered version, a self-administered parent version (report on youth), and a self-administered youth version (report on self) (Kobak and Kaufman 2015). Specifically, HCP used the two self-administered versions of this instrument as they have high concordance between current episode diagnoses using the computer self-administered KSADS-5 and clinician-administered paper-and-pencil version of the KSADS-5, with percent agreement in diagnostic categories ranging from 88%-96%, and kappas in the good to excellent range (Kobak et al. 2013). **KSADS data has not been mapped to NDA data structures and is therefore not available in the Lifespan 2.0 Release.**

7.3 – Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA; ages 35+)

Description (COGA study website):

SSAGA is a polydiagnostic psychiatric interview that covers demographics, mental health, substance use, emotional, and other medical conditions over the lifetime. All participants in the Aging Study (ages 36+) were administered SSAGA over the phone prior to (usually) or after their study visit. HCP-A used SSAGA-IV (cogastudy.org) with some modifications and omissions, very similar to what was used in the HCP-Young Adult study. The entire survey takes approximately one hour to complete and covers questions across the following areas: demographic questions, medical history, medications, mood disorders, and substance abuse. SSAGA has been shown to be a useful tool to look at not only familial substance abuse (e.g., alcoholism) but a variety of family studies (Bucholz et al., 1994). Administering the interview over the phone appears to not affect most of the results of the SSAGA (Kramer et al., 2009). SSAGA has also been shown to be a reliable and valid measure of a variety of different psychiatric disorders and has been used in multiple studies (Hesselbrock et al., 1999; Lynksey et al., 2005). It has been found that both substance abuse and psychiatric disorders are heritable, comorbid, and tend to emerge in early adulthood for at risk populations (Ehlers et al., 2013; Gilder et al., 2004; Munn-Chernnoff et al.,2013). **Only SSAGA demographic data (gender, marital status,** education, income, household size) **has been mapped to NDA data structures for the Lifespan 2.0 Release.** Responses can be found in the NDA data structure **ssaga_cover_demo01. Other SSAGA data has not been mapped to NDA data structures and is therefore not available in the Lifespan 2.0 Release.**

Substance Use

8.0 – Tobacco, Alcohol, and Substance Use

NIDA Substance Abuse and Alcohol Core (ages 12-21 and parents)

Description:

HCP administered this survey on an iPad via REDCap in the Development study to participants ages 12+, and parents of all participants as a self-report. NIDA Substance Abuse and Alcohol Core: Tier 1 assessments of Tobacco and Alcohol Use were adapted from PhenX (Conway et al., 2014; Somerville et al., 2018) which used resources such as the 2008 National Survey on Drug Use and Health to make a comprehensive measure of substance use. This survey has been shown to be reliable and valid measure of alcohol consumption, tobacco use, family history of depression and psychiatric disorder in the general population (Grant et al., 2003). Questions about alcohol, tobacco, and marijuana use and associated behaviors were asked as part of a survey on the iPad. Responses can be found in the NDA data structure **phenx_su01**.

Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA; ages 35+)

See above in 7.3 description and references. SSAGA includes questions about history and current use of tobacco, alcohol, marijuana, and other drugs used recreationally (outside prescribed treatment). **SSAGA**

substance use data has not been mapped to NDA data structures and is therefore not available in the Lifespan 2.0 Release.

Experiential and Behavioral Measures

9.1 – Sleep

Sleep Disturbances Scale for Children (SDSC; ages 5-12)

Description:

Parents of participants ages 5-12 in the Development study completed the SDSC as a parent-about-child report of their child's sleep on an iPad in a survey via REDCap. The SDSC is a 27-item inventory rated on a 5-point Likert-type scale created by Bruni et al. (1996); and has been shown to be reliable and valid in its measure of sleep disturbances in children (Ferreira et al., 2009). The instrument's purpose is to categorize sleep disorders in children. As well as giving an overall score the instrument uses five subdomains: disorders of initiating and maintaining sleep, sleep breathing disorders, disorders of arousal, sleep-wake transition disorders, disorders of excessive somnolence, and sleep hyperhidrosis. Responses can be found in the NDA data structure **sleepdis01**.

Munich Chrono Type Questionnaire (MCTQ; ages 8-17)

Description:

MCTQ is a self-reported survey analyzing sleep that HCP administered to participants ages 8 to 17 on an iPad via REDCap. If participants had questions about the phrasing of questions, they were able to clarify with an RA what something means (e.g., what time is 6PM on a 24-hour clock). MCTQ asks about participants' distinct preferences for various activities over the course of a day (time at which someone chooses to go to bed and get up) (Roenneberg et al., 2003). A key feature of the MCTQ is that it inquires information about sleep and activity times separately for work and work-free days. HCP uses what is currently the most recent version of MCTQ (version 2015-01 Till Roenneberg); however, MCTQ is constantly being updated and growing. The most updated version of MCTQ as well as a complete list of variables assessed/computed can be found on their website. HCP results from MCTQ can be found in the NDA structure **mctq01** with the question asked along with the participant's response. HCP does not report MCTQ calculated chronotype as described below.

Scoring (Roenneberg et al., 2015):

The MCTQ estimates chronotype based on the midpoint between sleep onset and offset on work-free days (midsleep on free days: MSF) corrected for "oversleep" due to the sleep debt that individuals accumulate over the workweek. This MCTQ-proxy for chronotype is based on the assumption that sleep timing on work-free days is highly influenced by the circadian clock. Therefore, chronotype should only be calculated when people do not use an alarm clock to wake up on work-free day. MCTQ-derived chronotype is a continuous, not a categorical, variable. It follows a near normal distribution in the population, like height or weight, with very few extreme early or late chronotypes. Since populations

differ in mean and width of their chronotype distribution—depending on many parameters such as position within a time zone, light exposure, and age—how "early" or "late" someone can be considered changes with the reference population. Therefore, MCTQ does not provide any general cutoff scores for "early" or "late" chronotypes. If one would like to compare different chronotypes within a sample, it's recommended to split groups according to the median or, better, working with quartiles or deciles. For example, one could compare the "earliest" 25% of a given sample with the "latest" 25%.

Pittsburgh Sleep Quality Index (PSQI; ages 18+)

Description (Buysse et al., 1989):

PSQI is a self-reported sleep quality scale that the HCP administered to participants in the Development and Aging studies who are ages 18+ on an iPad survey via REDCap. This survey assesses sleep quality and disturbances over a 1-month time interval. The PSQI consists of 19 self-rated questions and five questions rated by the bedpartner or roommate. HCP only used the 19 self-rated questions. The 19 self-rated questions assess a wide variety of factors relating to sleep quality, including estimates of sleep duration and latency and of the frequency and severity of specific sleep-related problems. The entire index requires 5-10 min for the subject to complete, and 5 min to score.

Scoring (Buysse et al., 1989)

The 19 questions asked on this questionnaire are grouped into seven component scores (each weighted equally on a 0-3 scale): subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, use of sleeping medication, and daytime dysfunction. The seven component scores are then summed to yield a global PSQI score, which has a range of 0-21; higher scores indicate worse sleep quality. The seven components of the PSQI are standardized versions of areas routinely assessed in clinical interviews of patients with sleep/ wake complaints. More information on exact scoring and which questions are grouped into each component can be found in the appendix of Buysse et al. (1989). Responses can be found in the NDA data structure **psqi01**.

9.2 – Adverse Life Events

Life Events-Child (ages 5+)

Description (PhenX Toolkit):

HCP administered a self-report survey measuring adverse life events for participants in the Development study ages 8+. The adverse life events survey was pulled from the PhenX Toolkit and is 25-item questionnaire that reports events experienced during the previous year over which the participant had little or no control. The respondent reviews the list of items and indicates which events have occurred. These questions are phrased as yes or no questions. If the event had occurred for the participant then the respondent is asked how the event made them feel and how much the event affected them. Participants answered these questions in a survey on the iPad via REDCap. Responses can be found in the NDA data structure **phenx_su01**.

Geriatric Adverse Life Events Scale (GALES; ages 36+)

Description (Devanand et al., 2002):

HCP administered the GALES to all participants in the Aging Study (ages 36+). This survey involves the participant to complete an initial self-report checklist of 26 adverse life events with a clearly specified time-frame of inquiry. If the participant indicates that they have experienced an event the survey prompts them with the follow-up questions "how stressful was this event" on the 3-point scale: Not At All, Somewhat, or Very Stressful; then "how much did the event impact you" with responses on the 5-point scale: Felt Much Better, Better, The Same, Worse, or Much Worse. Participants answered these questions in a survey on the iPad via REDCap. Responses can be found in the NDA data structure **gales01**.

9.3 – Stress and Self-efficacy

9.3.a NIH Toolbox – Perceived Stress

Perceived Stress is defined by individual perceptions about the nature of events and their relationship to the values and coping resources of an individual.

Perceived Stress Survey (ages 8+ and parents)

Description (Toolbox Website):

This measure assesses how unpredictable, uncontrollable and overloaded respondents find their lives. HCP administered the survey as a self-report measure to participants in both studies ages 13-17 and 18+; as well as a parent-about-child report measure for Development participants ages 8-12. The self-report versions are administered as 10-item fixed-length forms. The parent-report version is administered as a CAT. Parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Perceived Stress Survey, higher scores are indicative of more perceived stress. Scores can be found in the NDA data structure **pss01**.

9.3.b NIH Toolbox - Self-efficacy

Self-Efficacy can be described as a person's belief in his/her capacity to manage functioning and have control over meaningful events.

Self-Efficacy Survey (ages 8+ and parents)

Description (Toolbox Website)

This is a self-report measure administered to ages 8-12, 13-17, and 18+, as well as a parent-about-child report measure for ages 8-12. All versions are administered as a CAT. It assesses respondents' sense of global self-efficacy. Parents of participants in the Development study also completed this self-report survey.

Interpretation (Scoring and Interpretation Guide):
For the NIH Toolbox Self-Efficacy Survey, higher scores are indicative of more general self-efficacy. Scores for this survey can be found in the NDA data structure **self_effic01**.

9.4 – Friendships and Social Support

Social relationships have several dimensions, including their structure, extent and quality. NIH Toolbox focuses on three aspects of social relationships: perceived social support, positive social development, companionship and distress. Social satisfaction summary composite scores (friendship, emotional support, instrumental support, and reverse coded loneliness and perceived rejection) for ages 8+ and parents can found in the NDA data structure **promisgl01**.

9.4.a NIH Toolbox – Perceived Social Support

Perceived social support - is the extent to which an individual views his/her social relationships as available to provide aid in times of need or when problems arise. It includes instrumental and emotional types of perceived social support. Emotional Support refers to the perception that people in one's social network are available to listen to one's problems with empathy, caring and understanding. It is measured by the NIH Toolbox Emotional Support Survey. Instrumental Support refers to the perception that people in one's social network are available to provide material or functional aid in completing daily tasks, if needed; it is measured by the NIH Toolbox Instrumental Support Survey.

Emotional Support Survey (ages 8+ and parents)
Description (Toolbox Website):
This self-report measure assesses emotional support through two fixed-length forms. HCP administered an 8-item form for ages 18+ and a 7-item form for ages 8-17. No parent-about-child-report versions are available. However, parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):
For the NIH Toolbox Emotional Support Survey, higher scores are indicative of more emotional support. Scores can be found in the NDA data structure **tlbx_emsup01**.

Instrumental Support Survey (ages 18+ and parents)
Description (Toolbox Website):
This self-report measure assesses instrumental support. HCP administered this to participants in both studies ages 18-85, using an 8-item fixed-length form. No versions for other ages are available. However, parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Instrumental Support Survey, higher scores are indicative of more reported support. Scores can be found in the NDA data structure **tlbx_emsup01**.

9.4.b NIH Toolbox – Positive Social Development

Positive Social Development is characterized by parents' evaluation of their children's empathic behaviors. It is an indicator of a child's current emotional health and a predictor of positive and supportive social relationships in adolescence and adulthood. It is measured by the NIH Toolbox Empathic Behaviors Survey.

Empathic Behavior Survey (ages 5-12)

Description (Toolbox Website):

This parent-report measure for children ages 3-12 assesses parent perceptions of children's prosocial behaviors using a CAT. HCP obtained this as a parent-about-child report for participants ages 5-12.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Empathic Behaviors Survey, higher scores are indicative of more parent-reported child prosocial behaviors. Scores can be found in the NDA data structure **tlbx_empbeh01**.

9.4.c NIH Toolbox - Companionship

Companionship is characterized by self-reported perceptions of the availability of friends or companions with whom to interact or affiliate (i.e., friendship) and perceptions that one is alone, lonely or socially isolated from others (i.e., loneliness). Companionship is measured by the NIH Toolbox Friendship Survey, NIH Toolbox Loneliness Survey, NIH Toolbox Social Withdrawal Survey and NIH Toolbox Positive Peer Interaction Survey.

Friendship Survey (ages 8+ and parents)

Description (Toolbox Website):

This self-report measure assesses perceptions of friendship. HCP administered an 8-item fixed-length form for participants in both studies ages 18+ and a 5-item fixed-length form for Development participants ages 8-17. Parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Friendship Survey, higher scores are indicative of a greater perceived availability of companions with whom to interact or affiliate. Scores can be found in the NDA data structure **tlbx_friend01**.

Loneliness Survey (ages 8+ and parents)

Description (Toolbox Website):

This self-report measure assesses perceptions of loneliness. HCP administered this survey using a 5-item fixed-length form for participants in both studies ages 18+ and a 7-item fixed-length form for Development participants ages 8-17. Parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):
For the NIH Toolbox Loneliness Survey, higher scores are indicative of more loneliness. Scores can be found in the NDA data structure **prsi01**.

Social Withdrawal Survey (ages 5-12)
Description (Toolbox Website):
As an analogue to the NIH Toolbox Loneliness Survey, the NIH Toolbox Social Withdrawal Survey is a parent-about-child report measure for children ages 3-12. It is a 4-item fixed-length survey that assesses how often a child avoids or withdraws from social activities with peers. HCP obtained this parent-about-child report for participants ages 5 to 12.

Interpretation (Scoring and Interpretation Guide):
For the NIH Social Withdrawal Survey, higher scores are indicative of higher levels of social withdrawal. Scores can be found in the NDA data structure **tlbx_socwit01**.

Positive Peer Interaction Survey (ages 5-12)
Description (Toolbox Website):
As a conceptual analogue to the NIH Toolbox Friendship Survey, the NIH Toolbox Positive Peer Interactions Survey is a parent-about-child report measure for children ages 3-12. It is a 4-item fixed-length survey that assesses how often a child plays with friends and gets along with peers. HCP obtained this parent-about-child report for participants ages 5 to 12.

Interpretation (Scoring and Interpretation Guide):
For the NIH Toolbox Positive Peer Interaction Survey, higher scores are indicative of more positive peer interactions. Scores can be found in the NDA data structure **tlbx_socwit01**.

9.4.d NIH Toolbox - Social Distress

Social distress is the extent to which an individual perceives his/her daily social interactions as negative or distressing. This can include aspects of perceived hostility (e.g., how often people argue with me, yell at me, or criticize me) and perceived insensitivity (e.g., how often people don't listen when I ask for help, or do not pay attention to me). Self-reported perceived hostility is measured by the NIH Toolbox Perceived Hostility Survey; perceived insensitivity is measured by the self-report NIH Toolbox Perceived Rejection Survey and the parent-about-child report NIH Toolbox Peer Rejection Survey.

Peer Rejection Survey (ages 5-12)

Description (Toolbox Website):

As an analogue to the NIH Toolbox Perceived Rejection Survey, the NIH Toolbox Peer Rejection Survey is a parent-about-child report measure for children ages 3-12. It is a 9-item fixed-length form that assesses how often a child is left out, avoided or teased by peers. HCP obtained this parent-about-child report for participants ages 5 to 12.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Peer Rejection Survey, higher scores are indicative of greater peer rejection. Scores can be found in the NDA data structure **tlbx_rej01**.

Perceived Rejection Survey (ages 8+ and parents)

Description (Toolbox Website):

This self-report measure assesses perceptions of rejection. HCP administered this survey using an 8-item fixed-length form for participants in both studies ages 18+ and a 5-item fixed-length form for Development participants ages 8-17. Parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Perceived Rejection Survey, higher scores are indicative of greater perceived rejection. Scores can be found in the NDA data structure **tlbx_rej01**.

Perceived Hostility Survey (ages 8+ and parents)

Description (Toolbox Website):

This self-report measure assesses perceptions of hostility. HCP administered this survey using an 8-item fixed-length form for participants in both studies ages 18+ and a 5-item fixed-length form for Development participants ages 8-17. Parents of participants in the Development study also completed this survey as the adult self-report.

Interpretation (Scoring and Interpretation Guide):

For the NIH Toolbox Perceived Hostility Survey, higher scores are indicative of greater perceived hostility. Scores can be found in the NDA data structure **tlbx_perhost01**.

9.5 – Social Responsiveness

Social Responsiveness Scale - 2 (SRS-2; ages 5-21)

Description (Constantino & Gruber, 2012):

SRS-2 is 65-item rating scale that measures deficits in social behavior associated with Autism Spectrum Disorder (ASD), as outlined by the DSM-IV. This scale can be completed by raters who have at least 1-months experience with the rated individual when the rated individual is under the age of 18 and can be completed as a self-report when the person being rated is an Adult. This survey takes about 15 to 20

minutes to complete and requires an average reading ability. HCP administered this form as a parent-about-child report on the iPad via a REDCap survey to the parents of participants in the Development Study ages 5 to 17 and as a self-report survey to Development participants ages 18+. SRS-2 makes a unique contribution toward a comprehensive assessment of ASD, because it focuses specifically on aspects of social reciprocity and social communication.

Interpretation (WPS website):

HCP reports Total Raw score, Total Raw Score School Agehis instrument survey will provide: a total score, scores for 5 treatment subscales and scores for 2 DSM-5 Compatible Subscales. Norms for these scores are provided separately for each form; based on sample of 1,906 individuals (4,709 ratings); nationally representative in regard to gender, ethnicity, education, and geographic region; separated by rater type as well as age and gender of individual rated. Item-level responses and Scores for this instrument can be found in the NDA data structure **srs02**.

9.6 – Family Conflict

Family Environment Scale - Family Conflict Subscale (ages 5-21)

Description (PhenX Toolkit):

The 9-item family conflict subscale of the Family Environment Scale (FES), Fourth Edition (Moos & Moos, 2009), examines each family member's perceptions of family relationships as it is currently ("real" form of FES). This subscale taps the degree of commitment and support family members provide for one another, the extent to which family members are encouraged to express their feelings directly, and the amount of openly expressed anger and conflict among family members (Moos 1990). HCP administers this survey on the iPad via REDCap as a parent-about-child report to the parents of participants in the Development study ages 5 to 17 as well as to the participants themselves, who are ages 8 to 21, as a self-report. This is the same instrument as used in the ABCD study (although with different exact variable names), NDA data structure **abcd_fes01**. Responses to this survey can be found in the NDA data structure **fenvs01**.

9.7 – Screen Time

Screen Time Survey (ages 5-21)

Description:

HCP administered a screen time survey to participants in the Development Study. Parents of participants ages 5 to 17 were given a parent-about-child report asking questions on the child's total time spent on typical weekdays and weekend days on a computer, cell phone, tablet, other electronic device, watching TV/movies or videos, excluding school-related activities (similar to Sharif et al., 2010). Participants ages 8 to 21 were given a self-reported survey asking how many hours per weekday or weekend day they watch TV or movies, watch internet videos, play video games, text or message, visit social media, or video chat. Participants were also asked how often they play mature-rated video games and watch R-rated movies (similar to Hull et. al., 2014) in the self-reported version (Barch et al., 2017). Both of these

surveys (parent- and child-reports) are administered on an iPad via REDCap. This is the same instrument as used in the ABCD study (although with different exact variable names), NDA data structure **abcd_fes01**.

Interpretation:
Studies have found correlations between screen exposure and sensation seeking behaviors resulting in an indirect effect of poor school performance (Sharif et al., 2010) as well as deviant behavior (Hull et. al., 2014). Responses to the screen time survey can be found in the NDA data structure **screentime01**.

9.8 – Sports and Activities

Sports and Activities Involvement Questionnaire (ages 5-17)

Description:
HCP obtained a survey about participants in the Development Study sports and activities involvement. This survey was administered on the iPad, via a REDCap survey, as a parent-about-child report for participants ages 5 to 17. Whether the participant has ever participated in a range of sports or activities continuously for 4 months or more was assessed. If they had participated in a sport, school affiliation, organization, instruction method, and time spent were assessed in follow up questions. Frequency of listening to music and reading for pleasure were also asked. This measure was modified from the Vermont Health Behavior Questionnaire (Huppertz et al., 2016; Barch et al., 2017).

Scoring:
Many studies have tied child and adolescent sports and activities involvement to psychological phenomenon (Vella et al., 2016) such as self-image and behavior problems (Dolenc, 2015; Kirkaldy et al., 2002), family relationship quality (Bohnert et al., 2007), self-esteem (McClure et al., 2010), health risk behaviors (Taliaferro et al., 2010), and other parts of mental health in young adulthood (Godfrey et al., 2015; Jewett et al., 2014; Kremer et al., 2014). Responses to the exercise and activity survey can be found in the NDA data structure **saiq01**.

9.9 – Physical Activity

Short International Physical Activity Questionnaire (IPAQ; ages 36+)

Description (Booth, 2000):
The International Physical Activity Questionnaires (IPAQ) comprises a set of 4 questionnaires. Long (5 activity domains asked independently) and short (4 generic items) versions for use by either telephone or self-administered methods are available. The HCP used the short version in the Aging study for all subjects (i.e., ages 36+); IPAQ was not used in Development study. The purpose of the questionnaires is to provide common instruments that can be used to obtain internationally comparable data on health–related physical activity. Participant responses were collected by an RA and recorded on REDCap.

Interpretation:

Responses to the exercise and activity survey can be found in the NDA data structure **ipaq01**. Three levels (categories) of physical activity are proposed:

Category 1: Low

- This is the lowest level of physical activity. Those individuals who not meet criteria for categories 2 or 3 are considered low/inactive.

Category 2: Moderate

Any one of the following 3 criteria:

- 3 or more days of vigorous activity of at least 20 minutes per day

OR

- 5 or more days of moderate-intensity activity or walking of at least 30 minutes per day

OR

- 5 or more days of any combination of walking, moderate-intensity or vigorous intensity activities achieving a minimum of at least 600 MET-min/week (MET = Metabolic equivalent of task, an objective measure of energy expended per minute).

Category 3: High

Any one of the following 2 criteria:

- Vigorous-intensity activity on at least 3 days and accumulating at least 1500 MET-minutes/week

OR

- 7 or more days of any combination of walking, moderate-intensity or vigorous intensity activities achieving a minimum of at least 3000 MET-minutes/week

9.10 – Activities of Daily Living

Lawton-Brody Instrumental Activities of Daily Living (IADL) scale (ages 60+)

Description (Lawton & Brody, 1969):

HCP administered a self-report survey to older participants in the Aging study (i.e., ages 60+) where participants were asked about their level of independence when completing certain tasks associated with daily living (e.g., "Which of the following applies to your current ability to use a telephone?"). Participant responses were collected by an RA and recorded on REDCap. The survey was also self-administered via REDCap online at annual follow-up(s).

Options are given as a scale from being able to complete a task completely alone to not being able to complete the task at all.

Ex: Which of the following applies to your current ability to use a telephone?

1. I can operate the telephone on my own initiative, looking up and dialing numbers
2. I can dial only a few well-known numbers
3. I can answer the telephone, but I do not dial
4. I do not use the telephone at all

Interpretation (Lawton & Brody, 1969):

A higher score means that the participant is more autonomous in this area. What counts as the person being autonomous will vary from question to question but if a response scores as autonomous then they are given 1 point and if a response scores as non-autonomous it is given 0 points. Responses and Scores of this survey can be found in the NDA data structure **lbadl01**.

References

1. Achenbach TM, Newhouse PA, & Rescorla LA. (2004) Manual for the ASEBA Older Adult Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
2. Achenbach TM. (2009) The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.
3. Arditi A. (2005) Improving the design of the letter contrast sensitivity test. *Investigative Ophthalmology & Visual Science*, 46(6), 2225-2229.
4. Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, Hudziak JJ, Jernigan TL, Tapert SF, Yurgelun-Todd D, Alia-Klein N, Potter AS, Paulus MP, Prouty D, Zucker RA, & Sher KJ. (2017) Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Development cognitive neuroscience*, 32, 55–66. https://doi.org/10.1016/j.dcn.2017.10.010
5. Bijttebier P, Beck I, Claes L, & Vandereycken W. (2009) Gray's reinforcement sensitivity theory as a framework for research on personality-psychopathology associations. *Clinical Psychology Review*, 29, 421-430.
6. Bogner JA, Whiteneck GG, MacDonald J, et al. (2017) Test-Retest Reliability of Traumatic Brain Injury Outcome Measures: A Traumatic Brain Injury Model Systems Study. *J Head Trauma Rehabil.* 32(5): E1-E16. https://doi.org/10.1097/HTR.0000000000000291
7. Bohnert AM, Martin NC, & Garber J. (2007) Predicting Adolescents' Organized Activity Involvement: The Role of Maternal Depression History, Family Relationship Quality, and Adolescent Cognitions. *Journal of Research on Adolescence*, 17(1), 221–244. https://doi.org/10.1111/j.1532-7795.2007.00520.x
8. Booth M. (2000) Assessment of Physical Activity: An International Perspective, *Research Quarterly for Exercise and Sport*, 71:sup2, 114-120, https://doi.org/10.1080/02701367.2000.11082794
9. Brown SA, Brumback T, Tomlinson K, et al. (2015) The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): A Multisite Study of Adolescent Development and Substance Use. *J Stud Alcohol Drugs*. 76(6):895-908. https://doi.org/10.15288/jsad.2015.76.895
10. Bruni O, Ottavianio S, Guidetti V, Romoli M, Innocenzi M, Cortesi F, & Giannotti F. (1996) The Sleep Disturbance Scale for Children (SDSC): Construction and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. *J. Sleep Rrs*, 5, 251-261.
11. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger Jr. JI, Reich T, Schmidt I, Schuckit MA. (1994) A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol* 55: 149-158.
12. Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. (1989) The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 28: 193-213.
13. Capaldi DM & Rothbart MK. (1992) Development and validation of an early adolescent temperament measure. *The Journal of Early Adolescence, 12*(2), 153–173. https://doi.org/10.1177/0272431692012002002
14. Carver CS, & White TL. (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology*, 67, 319-333.

15. Chambers WJ, Puig-Antich J. (1985) The assessment of affective disorders in children and adolescents by semistructured interview. Test-retest reliability of the schedule for affective disorders and schizophrenia for school-age children, present episode version. *Arch. Gen. Psychiatry.* 42(7):696–702.
16. Cole BL. (2007) Assessment of inherited colour vision defects in clinical practice. *Clinical and Experimental Optometry*, 90(3), 157-175.
17. Constantino JN, & Gruber CP. (2012) Social Responsiveness Scale–Second Edition (SRS-2). Torrance, CA: Western Psychological Services.
18. Conway KP, Vullo GC, Kennedy AP, Finger MS, Agrawal A, Bjork JM, Farrer LA, Hancock DB, Hussong A, Wakim P, Huggins W, Hendershot T, Nettles DS, Pratt J, Maiese D, Junkins HA, Ramos EM, Strader LC, Hamilton CM, & Sher KJ (2014). Data compatibility in the addiction sciences: an examination of measure commonality. *Drug and Alcohol Dependence*, 141, 153-158.
19. Corrigan JD, Bogner J. (2007) Initial reliability and validity of the Ohio State University TBI Identification Method. *J Head Trauma Rehabil.* 22(6):318-329. https://doi.org/10.1097/01.HTR.0000300227.67748.77
20. Costa P & McCrae R. (2012) The Five-Factor Model, Five-Factor Theory, and Interpersonal Psychology. *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions*. 91-104. https://doi.org/10.1002/9781118001868.ch6
21. Crum RM, Anthony JC, Bassett SS, Folstein MF. (1993) Population-Based Norms for the Mini-Mental State Examination by Age and Educational Level. *JAMA.* 269(18):2386–2391. https://doi.org/10.1001/jama.1993.03500180078038
22. Depue, R. A., Krauss, S., Spoont, M. R., & Arbisi, P. (1989). General Behavior Inventory identification of unipolar and bipolar affective conditions in a nonclinical university population. *Journal of Abnormal Psychology*, 98(2), 117.
23. Devan and DP, Kim MK, Paykina N, Sackeim HA. (2002) Adverse life events in elderly patients with major depression or dysthymic disorder and in healthy-control subjects. *Am J Geriatr Psychiatry*. 10(3):265-274.
24. Diamond A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750
25. Dolenc, P. (2015). PHYSICAL SELF-CONCEPT IN SLOVENIAN ADOLESCENTS: DIFFERENCES BY GENDER AND SPORTS PARTICIPATION. *Facta Universitatis - Series Physical Education and Sport*. 13. 57-66.
26. Dougherty BE, Flom RE, Bullimore MA. (2005) An evaluation of the Mars Letter Contrast Sensitivity Test. *Optom Vis Sci*. 82(11):970-975. https://doi.org/10.1097/01.opx.0000187844.27025.ea
27. Ehlers CL, Gizer IR, Gilder DA, Yehuda R. (2013) Lifetime history of traumatic events in an American Indian community sample: heritability and relation to substance dependence, affective disorder, conduct disorder and PTSD. *J. Psychiatr. Res.* 47, 155–161
28. Ferreira VR, Carvalho LB, Ruotolo F, de Morais JF, Prado LB, Prado GF. (2009) Sleep disturbance scale for children: translation, cultural adaptation, and validation. *Sleep Med.* 10(4):457-463. https://doi.org/10.1016/j.sleep.2008.03.018
29. Folstein MF, Folstein SE, McHugh PR. (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 12(3):189-198. https://doi.org/10.1016/0022-3956(75)90026-6
30. Fortier CB, Amick MM, Grande L, et al. (2014) The Boston Assessment of Traumatic Brain Injury-Lifetime (BAT-L) semistructured interview: evidence of research utility and validity. *J Head Trauma Rehabil.* 29(1):89-98. https://doi.org/10.1097/HTR.0b013e3182865859

31. Geary DC, Berch DB, & Mann Koepke K. Chapter 1 - Introduction: Cognitive Foundations for Improving Mathematical Learning, Editor(s): Geary DC, Berch DB, & Mann Koepke K, In *Mathematical Cognition and Learning, Cognitive Foundations for Improving Mathematical Learning*, Academic Press, Volume 5, 2019, Pages 1-36, ISSN 22142568, ISBN 9780128159521, https://doi.org/10.1016/B978-0-12-815952-1.00001-3

32. Gilder DA, Wall TL, Ehlers CL (2004) Comorbidity of select anxiety and affective disorders with alcohol dependence in southwest California Indians. *Alcohol Clin. Exp. Res.* 28, 1805–1813

33. Godfrey C, Devine-Wright H & Taylor J. (2015) The positive impact of structured surfing courses on the wellbeing of vulnerable young people. *Community Practitioner.* 88.

34. Goldberg LR. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26

35. Goodman R. (1997) The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38: 581-586. d https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

36. Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of Child and Adolescent Psychiatry*, 40(11), 1337-1345.

37. Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *Eur Child Adolesc Psychiatry*. 1998;7(3):125-130. https://doi.org/10.1007/s007870050057

38. Grant KE, Compas BE, Thurm AE, McMahon SD, Gipson PY. (2004) Stressors and child and adolescent psychopathology: measurement issues and prospective effects. *J Clin Child Adolesc Psychol.* 33(2):412-425. https://doi.org/10.1207/s15374424jccp3302_23

39. Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R. (2003) The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug Alcohol Depend.* 71(1):7-16. https://doi.org/10.1016/s0376-8716(03)00070-x

40. Gray JA, & McNaughton N. (2000) The neuropsychology of anxiety: An inquiry into the functions of the septo-hippocampal system. Oxford: Oxford University Press.

41. Green L, Myerson J, Shah AK, Estle SJ, & Holt DD. (2007) Do adjusting-amount and adjusting-delay procedures produce equivalent estimates of subjective value in pigeons? *Journal of the Experimental Analysis of Behavior*, 87(3), 337–347. https://doi.org/10.1901/jeab.2007.37-06

42. Gur RC, Richard J, Hughett P, Calkins ME, Macy L, Bilker WB, Brensinger C, Gur RE. (2010) A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *J Neurosci Methods*. 187(2):254-262. https://doi.org/10.1016/j.jneumeth.2009.11.017

43. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, Hammond JA, Huggins W, Jackman D, Pan H, Nettles DS, Beaty TH, Farrer LA, Kraft P, Marazita ML, Ordovas JM, Pato CN, Spitz MR, Wagener D, Williams M, … Haines J. (2011) The PhenX Toolkit: get the most from your measures. *American Journal of Epidemiology*, *174*(3), 253–260. https://doi.org/10.1093/aje/kwr193

44. Hariri AR, Brown SM, Williamson DE, Flory JD, de Wit H, Manuck SB. (2006) Preference for Immediate over Delayed Rewards Is Associated with Magnitude of Ventral Striatal Activity. *Journal of Neuroscience*, 26 (51) 13213-13217; https://doi.org/10.1523/JNEUROSCI.3446-06.2006

45. Harlow SD, Gass M, Hall JE, Lobo R, Maki P, Rebar RW, Sherman S, Sluss PM, de Villiers, TJ, & STRAW+10 Collaborative Group. (2012) Executive summary of the Stages of Reproductive Aging Workshop +10:

addressing the unfinished agenda of staging reproductive aging. *Climacteric: The Journal Of The International Menopause Society*, 15(2), 105–114. https://doi.org/10.3109/13697137.2011.650656

46. Haymes SA, Roberts KF, Cruess AF, et al. (2006) The letter contrast sensitivity test: clinical evaluation of a new design. *Invest Ophthalmol Vis Sci*. 47(6):2739-2745. https://doi.org/10.1167/iovs.05-1419
47. Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. (1999) A validity study of the SSAGA— a comparison with the SCAN. *Addiction* 94: 1361-1370
48. Huppertz C, Bartels M, de Zeeuw EL et al. (2016) Individual Differences in Exercise Behavior: Stability and Change in Genetic and Environmental Determinants From Age 7 to 18. *Behav Genet* 46, 665–679. https://doi.org/10.1007/s10519-016-9799-x
49. Hull JG, Brunelle TJ, Prescott AT, & Sargent JD. (2014). A longitudinal study of risk-glorifying video games and behavioral deviance. *Journal of Personality and Social Psychology*, 107(2), 300–325. https://doi.org/10.1037/a0036058
50. Jewett R, Sabiston CM, Brunet J, O'Loughlin EK, Scarapicchia T & O'Loughlin, J. (2014) School Sport Participation During Adolescence and Mental Health in Early Adulthood. *Journal of Adolescent Health*. 55. https://doi.org/10.1016/j.jadohealth.2014.04.018.
51. Kaufman J, Birmaher B. (1997) Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J. Am. Acad. Child Adolesc. Psychiatry*. 36(7):980–988
52. Kaufman J, Birmaher B. K-Sads-Pl. (2000) *J. Am. Acad. Child Adolesc. Psychiatry*. 39(10):1208
53. Kaufman J, Birmaher B. Yale University; New Haven, CT: 2013. KSADS-PL.
54. Kaushanskaya M, Blumenfeld HK, Marian V. (2019) The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition 1–6*. https://doi.org/10.1017/S1366728919000038
55. Kessler RC, Avenevoli S, Costello EJ, et al. (2009) Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *Int J Methods Psychiatr Res.* 18(2):69-83. https://doi.org/10.1002/mpr.279
56. Kessler RC, Avenevoli S, Costello EJ, et al. (2009) National comorbidity survey replication adolescent supplement (NCS-A): II. Overview and design. *J Am Acad Child Adolesc Psychiatry*. 48(4):380-385. https://doi.org/10.1097/CHI.0b013e3181999705
57. Kirkcaldy BD, Shephard RJ, Siefen RG. (2002) The relationship between physical activity and self-image and problem behaviour among adolescents. *Soc Psychiatry Psychiatr Epidemiol.* 37(11):544-550. https://doi.org/10.1007/s00127-002-0554-7
58. Kobak KA, Kaufman J. Center for Telepsychology; Madison, WI: (2015) Ksads-comp.
59. Kobak KA, Kratochvil CJ. Anxiety Disorders and Depression. (2013) Computerized screening of comorbidity in adolescents with substance or psychiatric disorders. (La Jolla, CA).
60. Kramer, JR, Chan G, Kuperman S, Bucholz KK, Edenberg HJ, Schuckit MA, Polgreen LA, Kapp ES, Hesselbrock VM, Nurnberger JI, Bierut LJ. (2009) A comparison of diagnoses obtained from in-person and telephone interviews, using the semi-structured assessment for the genetics of alcoholism (SSAGA). *J. Stud. Alcohol Drugs* 70, 623–627.

61. Kremer P, Elshaug C, Leslie E, Toumbourou JW, Patton GC, Williams J. (2014) Physical activity, leisure-time screen use and depression among children and young adolescents. *J Sci Med Sport*. 17(2):183-187. https://doi.org/10.1016/j.jsams.2013.03.012

62. Lawton, M.P., Brody, E.M. (1969) Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontol* 9, 179–186.

63. Lynskey MT, Nelson EC, Neuman RJ, Bucholz KK, Madden PA, Knopik VS, Slutske W, Whitfield JB, Martin NG, Heath AC. (2005) Limitations of DSM-IV operationalizations of alcohol abuse and dependence in a sample of Australian twins. *Twin Res. Hum. Genet.* 8, 574–584

64. Marian V, Blumenfeld HK, Kaushanskaya M. (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J Speech Lang Hear Res.* 50(4):940-967. https://doi.org/10.1044/1092-4388(2007/067)

65. McClure AC, Tanski SE, Kingsbury J, Gerrard M, Sargent JD. (2010) Characteristics associated with low self-esteem among US adolescents. *Acad Pediatr.* 10(4):238-44.e2. https://doi.org/10.1016/j.acap.2010.03.007

66. McCrae, R. R., & Costa Jr, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36(3), 587-596

67. Merikangas KR, Nakamura EF, & Kessler RC. (2009) Epidemiology of mental disorders in children and adolescents. *Dialogues in clinical neuroscience*. 11(1), 7–20.

68. Moos R. (1990) Conceptual and empirical approaches to developing family-based assessment procedures: Resolving the case of the Family Environment Scale. *Family Process*, 29, 199-208.

69. Moos R, & Moos B. (2009). Family Environment Scale Manual and Sampler Set: Development, Applications and Research (4th ed.). Palo Alto, CA: Mind Garden, Inc.

70. Morris NM and Udry JR. (1980). Validation of a self-administered instrument to assess stage of adolescent development. *J Youth Adolescence* 9, 271–280. https://doi.org/10.1007/BF02088471

71. Munn-Chernoff, MA, Duncan AE, Grant JD, Wade TD, Agrawal A, Bucholz KK, Madden PA, Martin NG, Heath AC. (2013) A twin study of alcohol dependence, binge eating, and compensatory behaviors. *J. Stud. Alcohol Drugs.* 74, 664–673

72. Myerson J, Green L, & Warusawitharana M. (2001) Area under the curve as a measure of discounting. *Journal of the experimental analysis of behavior*. 76(2), 235–243. https://doi.org/10.1901/jeab.2001.76-235

73. Nasreddine ZS, Phillips NA, Bédirian V, et al. (2005) The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment [published correction appears in J Am Geriatr Soc. 2019 Sep;67(9):1991]. *J Am Geriatr Soc*. 53(4):695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

74. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia.* 1971 Mar;9(1) 97-113. https://doi.org/10.1016/0028-3932(71)90067-4. PMID: 5146491.

75. Orvaschel H., Puig-Antich J. (1982) Retrospective assessment of prepubertal major depression with the Kiddie-SADS-e. *J. Am. Acad. Child Psychiatry*. 21(4):392–397.

76. Pagliaccio D, Luking KR, Anokhin AP, et al. Revising the BIS/BAS Scale to study development: Measurement invariance and normative effects of age and sex from childhood through adulthood. *Psychol Assess*. 2016;28(4):429-442. https://doi.org/10.1037/pas0000186

77. Petersen AC, Crockett L, Richards M, Boxer A. A self-report measure of pubertal status: Reliability, validity, and initial norms. *J Youth Adolesc.* 1988;17(2):117-133. https://doi.org/10.1007/BF01537962

78. Reitan, RM. (1986) Trail Making Test: Manual for Administration and Scoring. United States: Reitan Neuropsychology Laboratory.

79. Reitan RM. (1958) Validity of the Trail Making test as an indicator of organic brain damage. *Percept Motor Skills*. 8: 271-276.

80. Rey, A. (1941) L'examen psychologique dans les cas d'encephopathie traumatique (The psychological examination of cases of traumatic encephalopathy), *Archives de Psychologie*, 28, 286-340. Corwin, J. and Bylsma, F.W., Translated (1993) *The Clinical Neuropsychologist*, 7, 4-9.

81. Rice JP, Reich T, Bucholz KK, et al. (1995) Comparison of direct interview and family history diagnoses of alcohol dependence. *Alcohol Clin Exp Res.* 19(4):1018-1023. https://doi.org/10.1111/j.1530-0277.1995.tb00983.x

82. Roenneberg T, Wirz-Justice A, Merrow M. Life between clocks: daily temporal patterns of human chronotypes. *J Biol Rhythms*. 2003;18(1):80-90. https://doi.org/10.1177/0748730402239679

83. Roenneberg T, Keller LK, Fischer D, Matera JL, Vetter C, Winnebeck EC. (2015) Human activity and rest in situ. *Methods in Enzymology*. 552:257-283. https://doi.org/10.1016/bs.mie.2014.11.028

84. Rothbart MK, Ahadi SA, Hershey KL, & Fisher P. (2001) Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, 72(5), 1394-1408.

85. Schoenberg MR, Dawson KA, Duff K, Patton D, Scott JG, Adams RL. (2006) Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Archives of clinical neuropsychology.* 21. 693-703. https://doi.org/10.1016/j.acn.2006.06.010

86. Sharif I, Wills TA, Sargent JD. (2010) Effect of visual media use on school performance: a prospective study. *The Journal of adolescent health: official publication of the Society for Adolescent Medicine*, 46(1), 52–61. https://doi.org/10.1016/j.jadohealth.2009.05.012

87. Somerville LH, Bookheimer SY, Buckner RL, Burgess GC, Curtiss SW, Dapretto, M, Elam JS, Gaffrey MS, Harms MP, Hodge C, Kandala S, Kastman EK, Nichols TE, Schlaggar BL, Smith SM, Thomas KM, Yacoub E, Van Essen DC, & Barch DM. (2018). The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5-21 year olds. NeuroImage, 183, 456–468. https://doi.org/10.1016/j.neuroimage.2018.08.050

88. Taliaferro LA, Rienzo BA, Donovan K. (2010) Relationships Between Youth Sport Participation and Selected Health Risk Behaviors From 1999 to 2007. *The Journal of school health*. 80. 399-410. https://doi.org/10.1111/j.1746-1561.2010.00520.x

89. Thayaparan K, Crossland MD, Rubin GS. (2007) Clinical assessment of two new contrast sensitivity charts. *The British journal of ophthalmology*, *91*(6), 749–752. https://doi.org/10.1136/bjo.2006.109280

90. Tiet QQ, Bird HR, Davies M, Hoven C, Cohen P, Jensen PS, Goodman S. (1998) Adverse life events and resilience. *Journal of the American Academy of Child and Adolescent Psychiatry*, *37*(11), 1191-1200.

91. Todd RD, Joyner CA, Heath AC, Neuman RJ, & Reich W. (2003) Reliability and stability of a semistructured DSM-IV interview designed for family studies. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(12), 1460-1468.

92. Vella SA, Braithwaite RE, Gardner LA, Spray CM. (2016) A systematic review and meta-analysis of implicit theory research in sport, physical activity, and physical education. *International Review of Sport and Exercise Psychology*. 1-24. https://doi.org/10.1080/1750984X.2016.1160418

93. Watson D, Clark LA, Tellegen A. (1988) *Development and validation of brief measures of positive and negative affect: the PANAS scales*. Journal of Personality and Social Psychology, 54(6), 1063.

94. Wechsler D. (2003) *Wechsler adult intelligence scale - fourth edition*. Bloomington, MN: Pearson.

95. Wechsler, D. (2003). *Wechsler intelligence scales for children–fourth edition*. Bloomington, MN: Pearson.

96. Wechsler D, Kaplan E, Fein D, Kramer J, Morris R, Delis D, et al. (2004) *WISC-IV Integrated: Administration and scoring manual*. Harcourt Assessment, San Antonio, TX.

97. Wechsler D. (2014) *Wechsler intelligence scale for children®–fifth edition*. Bloomington, MN: Pearson.

98. Wheeler M. Episodic and Autobiographical Memory: Psychological and Neural Aspects, Editor(s): Smelser NJ, Baltes PB, *International Encyclopedia of the Social & Behavioral Sciences*, Pergamon, 2001, Pages 4714-4717, ISBN 9780080430768, https://doi.org/10.1016/B0-08-043076-7/03514-2.

99. Youngstrom EA, Frazier TW, Demeter C, Calabrese JR, Findling RL. (2008). Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. *The Journal of clinical psychiatry*, *69*(5), 831–839. https://doi.org/10.4088/jcp.v69n0517

100. Youngstrom EA, Murray G, Johnson SL, Findling RL. (2013) The 7 up 7 down inventory: a 14-item measure of manic and depressive tendencies carved from the General Behavior Inventory. *Psychological assessment*, *25*(4), 1377–1383. https://doi.org/10.1037/a0033975

101. Zapolski, T. C., Stairs, A. M., Settles, R. F., Combs, J. L., & Smith, G. T. (2010). The measurement of dispositions to rash action in children. Assessment, 17(1), 116-125.

102. Zhao Q, Lv Y, Zhou Y, Hong Z, Guo Q. (2012) Short-term delayed recall of auditory verbal learning test is equivalent to long-term delayed recall for identifying amnestic mild cognitive impairment. *PloS one*, 7(12), e51157. https://doi.org/10.1371/journal.pone.0051157