Date Updated: 11/08/2017

NIH Toolbox Analysis Guide: Cognition Battery

I. Background

a. ***Platforms****:* The NIH Toolbox is a comprehensive set of neurobehavioral tests that quickly assess cognitive, emotional, sensory, and motor functions. The Cognition Battery measures Attention, Episodic Memory, Working Memory, Language, Executive Function, and Processing Speed. It was originally developed for web-based administration, and it was normed using the web-based platform. However, it has been adapted for iPad app-based administration.

b. ***Score Updates****:* Two updates affect scores produced by the NIH Toolbox. First, updated normative scores were developed and published[1], but were implemented (i.e., made available to users automatically) only on the NIH Toolbox iPad app and not on the web. This affected all four NIH Toolbox domains. Second, a study was conducted in 2016 to evaluate the equivalence of Cognition Battery tests taken on the web vs. iPad. It resulted in score adjustments to selected tests on the iPad app. These two updates affect current and prospective NIH Toolbox users differentially, and are described in this Analysis Guide.

c. ***User Groups****:* Research and clinical users of the NIH Toolbox can be grouped, largely based on how they have collected data: First, there is a group of people who have not collected data, but are considering using the NIH Toolbox. The web-based platform is no longer available to them and they are concerned about starting to use the app without knowing how well it compares to the original test battery and norms (referred to below as "Future Users"). Among existing users, some have only used the web-based NIH Toolbox ("Web-Only Users"), while others are using only the app ("iPad-Only Users"). Finally, some research and clinical users have started collecting data on the web but have now transitioned to the iPad app ("Multi-Platform Users"). *We recommend that users identify the group to which they belong and read the relevant sections below. Help is available to users at any step in this process by contacting* help@nihtoolbox.org.

[1] Casaletto KB, Umlauf A, Beaumont J, Gershon RC, Slotkin J, Akshoomoff N, & Heaton RK. (2015). Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society,* 21, 378-391. PMID 26030001

II. Future Users

a. ***What has been done?*** The NIH Toolbox was developed for use in research and clinical settings. It offers brief, portable assessments of cognitive, emotional, sensory, and motor functions. The NIH Toolbox utilizes a lifespan developmental approach, maximizing efficiency through state-of-the-science methodology. It has been adapted for iPad-based administration. We have conducted a mode-of-administration equivalence study for the Cognition Battery, comparing web- vs. app-based participant performance. Score adjustments within the app based upon this study have ensured appropriate equivalency across both platforms. The published norms apply fairly and accurately regardless of mode of administration. This NIH Toolbox iPad app is available through the Apple app store. New users can be confident that the data they collect with it will match what would be obtained via the web-based version. The existing web-based norms are applicable to the NIH Toolbox iPad app.

b. ***What will be done automatically?*** The NIH Toolbox iPad app[2] provides normative scores, as described by Casaletto et al. (2015), and includes all rescoring and equivalency adjustments needed to utilize these norms.

c. ***What do you need to do?*** The iPad app is ready for you. We encourage you to download it from the app store and start using it!

[2] New users should be using app version 1.8 or later, which includes all equivalency adjustments (including rescoring) and norm score calculations. If you are not sure what version of the app that you are currently running, click the question mark on the bottom right corner of the home screen and scroll down to the App Version number in the Diagnostic Information section. Version 1.8 and later is optimal.

III. Web-Only Users

a. ***What has been done?*** As of January 2016, Assessment Center[SM] stopped allowing new users access to the web-administered NIH Toolbox. However, existing users could finish their in-process studies, some of which are longitudinal. (We recommend that web-based users who have projects that will continue beyond 2017 consider transitioning to the NIH Toolbox iPad app. Please see Section V. below to understand scoring and data considerations associated with this transition.)

b. ***What will be done automatically?*** No changes will be automatically implemented on the web-administered NIH Toolbox.

c. ***What do you need to do?*** Casaletto et al. (2015) published revised formulas for calculating normative scores, which were developed based on the web-administered NIH Toolbox; however, they *were not implemented online*. The formulas are available as an appendix to the publication or are available online at http://nihtoolbox.desk.com as a **Python NIH Toolbox Cognition Battery Norming Program**[3]. To obtain updated normative scores for web-administered data, web users must program these formulas or use the Python NIH Toolbox Cognition Battery Norming Program. For more information or additional help, please email help@nihtoolbox.org.

d. ***Special instructions when locally programming normative scores.*** The formulas for the updated norms provided by Casaletto et al. (2015) use specific fields and values that may not be clear from the web-based exports and the published formulas. First, all types of normative scores for two of the Cognition Battery tests (Picture Vocabulary and Oral Reading) use theta values in the formulas, but the score exports report a computed score instead. Second, a different coding scheme is used for demographics in the Fully Corrected T-Scores formulas than the codes exported in the web-based NIH Toolbox registration file. NOTE: The Python NIH Toolbox Cognition Battery Norming Program uses the Assessment Center exports and automatically applies the adjustments (below) prior to calculating the Casaletto et al. (2015) norming formulas. If you are **not** using the Python NIH Toolbox Cognition Battery Norming Program and are instead locally programming the formulas, the transformations described below are required.

   i. *Converting Computed Scores and Theta.* All published formulas for Oral Reading Recognition Test and Picture Vocabulary Test utilize theta values but the scores export report computed scores. The final theta value from a data file would be acceptable. Otherwise, one can convert between these values using the linear transformations in Table 1. NOTE: these conversions only apply to the English version of the NIH Toolbox.

[3] See the instructions accompanying the Python NIH Toolbox Cognition Battery Norming Program for specific instructions. Editing source data files can cause the program to malfunction. If editing is necessary, best practice is to use a text editor.

Table 1: English Computed Score and Theta Score Conversions

Test	Theta to Computed	Computed to Theta
Oral Reading	Computed = (theta + 16.1) * 100	Theta = (computed / 100) – 16.1
Picture Vocabulary	Computed = (theta + 12.0) * 100	Theta = (computed / 100) – 12.0
Picture Sequence Memory	Computed = (theta + 5.4) * 100	Theta = (computed / 100) – 5.4

ii. *Gender.* In the Fully Corrected T-Score formulas, gender is coded {0 = Female; 1 = Male}. However, in the Assessment Center registration file, gender is coded {1 = Male; 2 = Female}. This needs to be recoded prior to normative score calculation if one is not using the Python NIH Toolbox Cognition Battery Norming Program.

iii. *Educational Level*. Highest level of education completed is also used for calculating Fully Corrected T-Scores. For adults (ages 18+ years), it is the individual's own education; for children (ages 3-17 years), maternal education is used. Table 2 provides a mapping between the variable coding in the Assessment Center registration file and the "number of years completed" variable used when calculating Fully Corrected T-Scores. Web-based users need to map the education code to the years of education if they are not using the Python NIH Toolbox Cognition Battery Norming Program.

Table 2: Levels of Education Transformations for Fully Corrected T-Score Formulas

Education Code in Registration File	Description	Years of Education Completed for Norm Calculation	Education Code in Registration File	Description	Years of Education Completed for Norm Calculation
1	none	0	14	11th grade	11
2	preschool	0	16	High School Graduate	12
3	kindergarten	0	17	GED	**
4	1st grade	1	18	Some college credit but less than 1 year	12
5	2nd grade	2	20	Associates degree (e.g., AA, AS)	14
6	3rd grade	3	21	Bachelor's degree (e.g., BA, AB, BS)	16
7	4th grade	4	22	Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)	18
8	5th grade	5	23	Professional degree (e.g., MD, DDS, DVM, LLB, JD)	20
9	6th grade	6	24	Doctorate degree (e.g., PhD, EdD)	20
10	7th grade	7	25	One or more years of college at a 2-year program, no degree	13
11	8th grade	8	26	One year of college at a 4-year program, no degree	13
12	9th grade	9	27	Two years of college at a 4-year program, no degree	14
13	10th grade	10	28	Three years or more of college at a 4-year program, no degree	15

**Individuals who obtained a GED (education code 17 in Assessment Center) but did not go on to complete additional years of schooling beyond high school should be mapped onto their last grade COMPLETED (e.g., if a person dropped out *during* 10th grade and later obtained a GED but did not go on to complete one or more years of college, he or she should be mapped to 9 years of education as 9th grade was the final year completed).

IV. iPad-Only Users

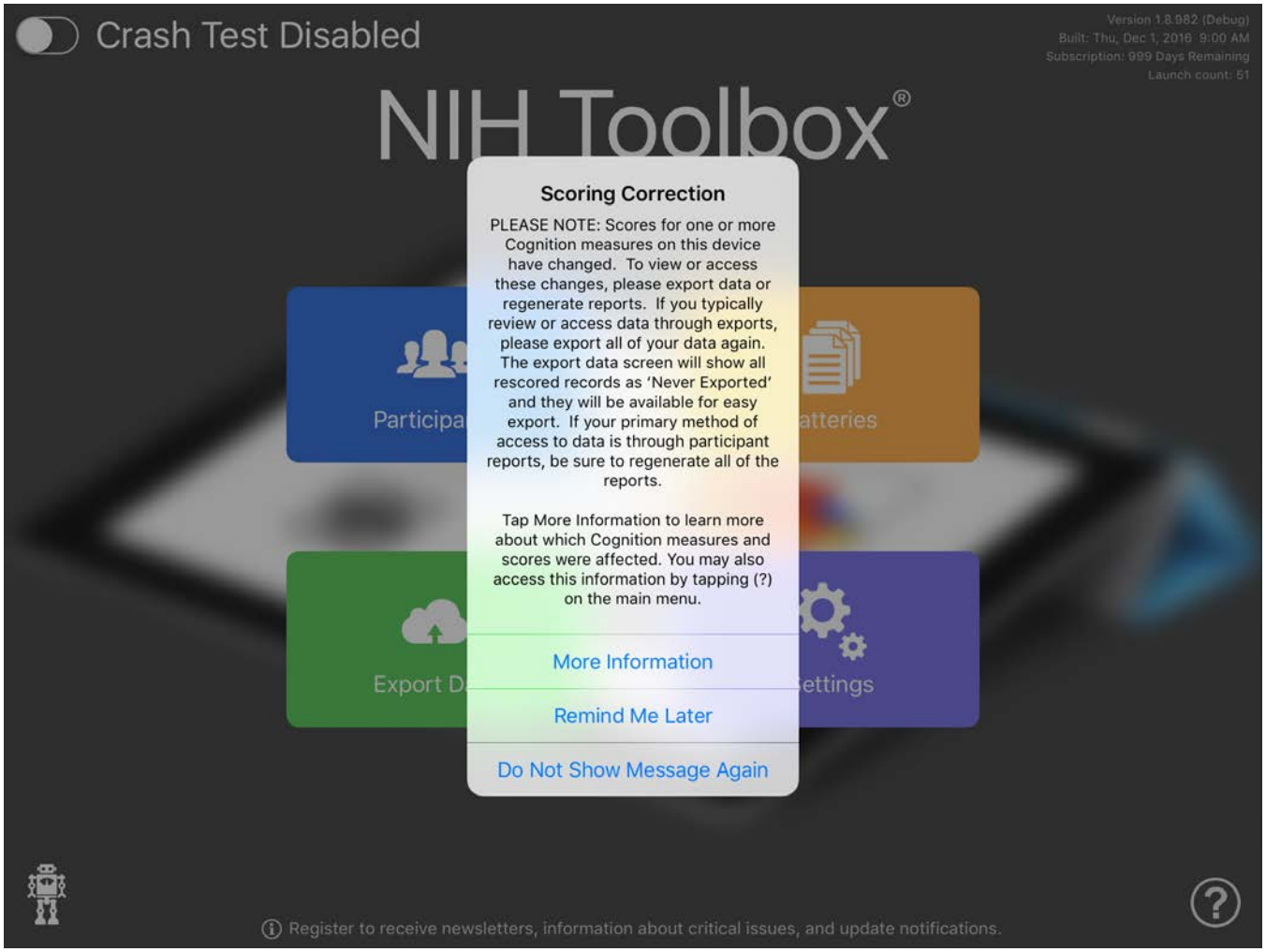
a. ***What has been done?*** The NIH Toolbox was adapted from web-based administration to app-based administration. The normative formulas developed and described in Casaletto et al. (2015) are fully implemented in the iPad version of the NIH Toolbox. These normative formulas were based on a national norming study using the web-administered NIH Toolbox.

   A new study evaluating the equivalency between the two modes of administration (iPad and web) was completed in 2016. It noted that there were some differences when using the iPad app on four tests: Dimensional Change Card Sort Test, Flanker Inhibitory Control and Attention Test, Pattern Comparison Processing Speed Test, and Picture Sequence Memory Test. The Cognition Fluid and Cognition Total Composite Scores, which include these tests, were also affected as a result. The differences were primarily due to timing and speed of response on a keyboard (web) vs. a touchscreen (iPad). An adjustment was required to accurately utilize the Casaletto et al. (2015) norms. App version 1.8 and later of the NIH Toolbox iPad app reflects these adjustments.

b. ***What will be done automatically?*** The NIH Toolbox iPad app does two things: 1. It takes your existing iPad data from prior assessments and rescores it (also providing normative corrections and composite scores) automatically; 2. It will automatically apply the scoring adjustments for all future test administrations.
   i. *For existing data on the iPad.* All EXISTING data will be rescored. This will be a one-time occurrence and not require additional rescoring for future administrations (described below). Data that have not been previously exported will be rescored and available for export. Data that have been exported but are still stored on the iPad (including archived data) will also be rescored and available to export again. Finally, an audit log will be created by the app, documenting every change (i.e., test scores before and after rescoring for each participant affected by the adjustments). Note that if you have exported data and deleted it from the iPad app (or if you do not wish to re-export data), the **SAS NIH Toolbox Cognition Battery Rescoring Program** is available online at http://nihtoolbox.desk.com; see Section IV.c. below for more details.
   ii. *For future administrations.* The NIH Toolbox iPad app will automatically apply all equivalency adjustments and normative scores to future administrations. Instrument versions within the app reflect this change. **The test items and administration procedures have not changed.** The only change is in the scoring, to account for the mode of administration (iPad vs. Web).

c. ***What do you need to do?*** Most people using the iPad version of the NIH Toolbox will see a pop-up after upgrading to version 1.8 of the app. This will inform them about rescoring, which they will need to acknowledge (see Figure 1). Users can get more information in the app ("More Information"); they can choose to be reminded about this every single time the app opens (until they change their response "Remind Me Later"); or they can acknowledge that they are aware of rescoring and do not need to be reminded about it again ("Do Not Show Message Again"). Rescoring will then occur automatically (see Section IV.b.i.). ***Users will then need to re-export the results from these existing assessments***.

Figure 1: In-App Notification for Rescoring



In rare cases where a user has deleted data from the app, or in cases when a user has already exported data and does not wish to re-export it, the SAS NIH Toolbox Cognition Battery Rescoring Program has been developed and is available online at http://nihtoolbox.desk.com. This script requires ALL THREE exports (scores, data, and registration files) to rescore Dimensional Change Card Sort Test, Flanker Inhibitory Control and Attention Test, Pattern Comparison Processing Speed Test, and Picture Sequence Memory Test. Computed scores and normative scores equivalent to the web-administered NIH Toolbox are calculated in this process. Instructions and an example on how to use the SAS NIH Toolbox Cognition Battery Rescoring Program are also available online to assist users. For more information or additional help, please email help@nihtoolbox.org.

V. Special Concerns for Multi-Platform Users

a. ***What has been done?*** Multi-platform users should be aware of what changes have occurred that may affect both the web (see Section III.a.) and the iPad (see Section IV.a.).

b. ***What will be done automatically?*** Rescoring app-based data will happen automatically, as described in Section IV.b.i. However, no other changes will be implemented automatically that are exclusively relevant to multi-platform users.

c. ***What do you need to do?*** Researchers and clinicians who have collected data on both platforms have multiple steps to follow, reflective of each platform. For assistance with any of these steps, please email the NIH Toolbox team at help@nihtoolbox.org.
   i. *Data from the web.* If you are not already applying the Casaletto et al. (2015) formulas to your web-administered data, please see Section III.c. These are not implemented on the web and must be calculated locally, using the formulas in the published appendix, or through the Python NIH Toolbox Cognition Battery Norming Program (which we recommend using), if accurate comparisons are to be made across platforms. If you are not using the Python NIH Toolbox Cognition Battery Norming Program, please be sure to first apply the transformations described in Section III.d.
   ii. *Data from the app.* Multi-platform users need to be aware that the data on their iPad is being rescored, and potentially use the SAS NIH Toolbox Cognition Battery Rescoring Program if data has been deleted from the app, as described in Sections IV.b.i. and IV.c.
   iii. *Aggregated data across platforms.* In most cases, clinicians and researchers want to be able to combine data from both the web- and app-administered NIH Toolbox. Once you have completed the above-mentioned steps, **all Age-Corrected Standard Scores and Fully Corrected T-Scores and almost all other scores** (see "d" and Table 2 below) **will be comparable across the web- and iPad-administered NIH Toolbox platforms, and can be combined for analytical purposes**. The general recommendation is that the normative scores should be used in analyses. For more information on what each type of score means and their utility, please refer to the *Scoring and Interpretation Manual* and Casaletto et al. (2015).

d. ***What scores are not comparable across platforms?*** Some values reported in the scores export are not comparable across platforms, even after completion of the steps in Section V.c. These include selected scores on Pattern Comparison Processing Speed, Picture Sequence Memory Test, the Cognition Fluid Composite, and the Cognition Total Composite.
   i. *Pattern Comparison Processing Speed* – Timing differences result in fewer trials on the iPad and a lower score. Therefore, raw scores are not comparable. The NIH Toolbox iPad app now produces a Pattern Comparison Processing Speed *computed* score that can be compared to the web-administered *raw* score. Note that all normative scores on Pattern Comparison Processing Speed are comparable.
   ii. *Picture Sequence Memory Test* – The web- and iPad-administered Picture Sequence Memory Test report different scores, though the test is the same. The Picture Sequence Memory Test Age-Corrected and Fully Corrected normative scores are comparable across platforms. The computed score is also comparable, when

reported.[4] However, the Picture Sequence Memory Test standard error (SE), which is not recommended for analyses, and the Uncorrected Standard Score are not comparable. Formulas are provided on the final page of this document to facilitate comparisons of this and other Uncorrected Standard Scores.[5]

iii. *Cognition Fluid Composite Uncorrected Standard Score* – Because the Picture Sequence Memory Test Uncorrected Standard Score is used to calculate the Cognition Fluid Composite Uncorrected Standard Score, it is also not comparable across platforms. The Age-Corrected and Fully Corrected Cognition Fluid Composite scores are comparable.

iv. *Cognition Total Composite Uncorrected Standard Score* – The Cognition Total Composite Uncorrected Standard Score utilizes the Cognition Fluid Composite Uncorrected Standard Score. Thus, it is also not comparable across platforms. The Cognition Total Composite Age-Corrected and Fully Corrected normative scores are comparable.

e. ***Score comparability across platforms (following the Score Export format).*** Table 3 displays the score comparability described in Section V.d., utilizing the Score Export file format to indicate which scores are comparable. The instrument and composite names are listed in the default order. Column names reflect the order in which they appear on the export (though not all columns are shown). Grayed-out cells indicate scores that do not exist. "Not Comp" indicates non-comparable scores. "Comp" indicates that scores are comparable, and superscripts indicate special cases.

[4] The Picture Sequence Memory Test computed score is always reported on the web-administered NIH Toolbox. On the iPad-administered version, it is reported for individuals aged 3-7 years and for individuals age 8+ when date of birth is in the registration file. (Some users have elected not to include date of birth in the registration file, to avoid what might be considered personally identifiable information in the file.)

[5] The equivalency study found age-based differences by mode of administration on the Picture Sequence Memory Test. A new Uncorrected Standard Score was developed from the normative study data (web-administered NIH Toolbox) based on expected performance on the iPad version. Thus, there are two Picture Sequence Memory Uncorrected Standard Score formulas: the original Casaletto et al. (2015) Picture Sequence Memory Test Uncorrected Standard Score from the theta score; and the new score reported by the app from the computed score. Both formulas require age at time of test administration. The Addendum to this document includes formulas to calculate the Uncorrected Standard Score composites for the Cognition Fluid Composite and Cognition Total Composite that utilizing the different versions of the Picture Sequence Memory Test Uncorrected Standard Score.

Table 3: Score Comparability (abbreviated* `comp`*) across the Web- and App-Administered NIH Toolbox

Instrument	Raw Score	Theta	Standard Error (SE)	Computed Score	Uncorrected Standard Score	Age-Corrected Standard Score	Fully Corrected T-score
Toolbox Picture Vocabulary Test (TPVT)		Comp	Comp		Comp	Comp[A]	Comp[D]
Flanker Inhibitory Control and Attention Test (Flanker)	Comp			Comp	Comp	Comp[A]	Comp[D]
List Sorting Working Memory Test	Comp				Comp	Comp[A]	Comp[D]
Dimensional Change Card Sort Test (DCCS)	Comp			Comp	Comp	Comp[A]	Comp[D]
Pattern Comparison Processing Speed Test (PCPS)	Not Comp			Not Comp	Comp	Comp[A]	Comp[D]
Picture Sequence Memory Test (PSM)	Comp	Not Comp	Not Comp	Comp[†]	Not Comp	Comp[A]	Comp[D]
Oral Reading Recognition Test		Comp	Comp		Comp	Comp[A]	Comp[D]
Cognition Fluid Composite					Not Comp	Comp[A]	Comp[D]
Cognition Crystalized Composite					Comp	Comp[A]	Comp[D]
Cognition Total Composite					Not Comp	Comp[A]	Comp[D]
Cognition Early Childhood Composite					Comp[†]	Comp[A]	Comp[D]

- `Not Comp` indicates that the score is not comparable across the web- and app-administered versions.
- `Comp` indicates that the score is always available and is comparable across platforms
- `Comp`[†] means the score is comparable, but is only output for ages 3-7 years or when investigators have included date of birth in the registration file for ages 8+.
- `Comp`[A] means the score is comparable, but is only output when date of birth is in the registration file (for all ages).
- `Comp`[D] means the score is comparable, but is only output when all demographics are in the registration file.

Addendum: Multi-Platform Uncorrected Standard Scores for Picture Sequence Memory and Associated Composites

*NOTE: Researchers and clinicians using Age-Corrected Standard Scores or Fully Corrected T-Scores do not need to reference this addendum. It is only relevant for **Multi-Platform Users** who want to be able to compare **Uncorrected Standard Scores** across the web- and iPad-based administrations.*

I. Background

When evaluating the equivalency between the web- and iPad-based NIH Toolbox administrations, there was a natural desire to utilize the enhanced normative formulas published by Casaletto et al. (2015). These norms were developed based upon web-administered data. After equating the tests, the web-based norms were equally applicable to both modes of administration.

However, there was an age-by-mode effect for the Picture Sequence Memory Test (PSM). It was therefore necessary to include age when making adjustments to the test. Because Date of Birth is not a required field within the NIH Toolbox iPad app, and because this information as needed to account for the PSM age-by-mode effect, the decision was made to "reverse" the equating procedure (for PSM only). Rather than utilizing the web-based norms with an age adjustment to iPad-based administrations, the original norming data (utilized by Casaletto et al., 2015) was converted to equivalent expected iPad scores and re-normed using the same procedures described by Casaletto et al. This created two types of norms: the web-based norms described by Casaletto et al., and the iPad-based norms from equivalent expected scores. This approach for PSM only affected the Uncorrected Standard Scores (USS).[6]

In the above sections, we recommended that web-users generate normative scores through the Python NIH Toolbox Cognition Battery Norming Program, which exports the Casaletto et al. Uncorrected Standard Scores. However the NIH Toolbox iPad App rescores and exports the new Uncorrected Standard Scores, as does the SAS NIH Toolbox Cognition Battery Rescoring Program. However, these scores *are not* comparable for multi-platform users because they do not take account of the age-by-mode effect. If users are analyzing these scores across platforms (for example, in a longitudinal study examining individual growth), additional conversions are necessary.

Clinicians and researchers should choose to use scores based on *either* the original Casaletto et al. formulas or the iPad formulas, but not both. Either the iPad-based data needs to be changed to be comparable with the output from the Python NIH Toolbox Cognition Battery Norming Program (using Addendum Section II), or the web-based data needs to be changed to be comparable to scores exported by the app or the SAS NIH Toolbox Cognition Battery Rescoring Program (using Addendum Section III).

[6] The Age-Corrected and Fully Corrected normative scores *already* required age. The Casaletto et al. formulas are utilized for these corrected scores regardless of data collection platform and are thus comparable.

II. **Casaletto et al. (2015) Formulas:** These conversions are applied to iPad-administered data, to make the scores comparable to the output from the Python NIH Toolbox Cognition Battery Norming Program.

a. ***Picture Sequence Memory Computed Score.*** Note: This score is reported on the NIH Toolbox iPad app when age is available in the registration file. It must be calculated outside of the app if date of birth was not entered in the registration file.

i. *Age 3-11:* $PSM\ Computed\ Score = \left((theta + 0.217807) + 5.4\right) * 100$

ii. *Age 12-59:* $PSM\ Computed\ score = \left((theta + 0.447964) + 5.4\right) * 100$

iii. *Age 60+:* $PSM\ Computed\ Score = (theta + 5.4) * 100$

b. ***Picture Sequence Memory Uncorrected Standard Score.***

$$PSM\ USS = \left(\frac{PSM\ Computed\ Score - 505.59}{99.83}\right) * 15 + 100$$

c. ***Cognition Fluid Composite Uncorrected Standard Score***

$$Fluid\ USS = \left(\frac{\left(DCCS\ USS + Flanker\ USS + LSWM\ USS + PSM\ USS + PCPS\ USS/_5\right) - 100.39}{11.83}\right) * 15 + 100$$

d. ***Cognition Total Composite Uncorrected Standard Score***

$$Total\ USS = \left(\frac{\left(Fluid\ USS + Crystallized\ USS/_2\right) - 100.84}{12.04}\right) * 15 + 100$$

III. **Formulas Implemented on the iPad:** These conversions are applied to the web-administered data to make scores comparable to the output from the NIH Toolbox iPad app version 1.8 or the SAS NIH Toolbox Cognition Battery Rescoring Program.

a. ***Picture Sequence Memory Adjusted Theta.***

i. *Age 3-11:* $PSM\ adjusted\ theta = \left(\left(Computed\ Score/_{100}\right) - 5.4\right) - 0.217807$

ii. *Age 12-59:* $PSM\ adjusted\ theta = \left(\left(Computed\ Score/_{100}\right) - 5.4\right) - 0.447964$

iii. *Age 60+:* $PSM\ adjusted\ theta = \left(\left(Computed\ Score/_{100}\right) - 5.4\right)$

b. ***iPad-based Picture Sequence Memory Uncorrected Standard Score.***

$$PSM_{ipad}\ USS = \left(\frac{PSM\ adjusted\ theta + 0.69051}{0.94131}\right) * 15 + 100$$

c. ***Cognition Fluid Composite Uncorrected Standard Score***

$$Fluid_{ipad}\ USS = \left(\frac{\left(DCCS\ USS + Flanker\ USS + LSWM\ USS + PSM_{ipad}\ USS + PCPS\ USS/_5\right) - 100.89}{10.77}\right) * 15 + 100$$

d. ***Cognition Total Composite Uncorrected Standard Score***

$$Total\ USS = \left(\frac{\left(Fluid_{ipad}\ USS + Crystallized\ USS/_2\right) - 100.53}{12.60}\right) * 15 + 100$$