Latent Factor Structure and Measurement Invariance of the NIH Toolbox Cognition Battery in an Alzheimer's Disease Research Sample

Yue Ma[1], Cynthia M. Carlsson[1,2,3], Michelle L. Wahoske[1], Hanna M. Blazel[1], Richard J. Chappell[1,4,5], Sterling C. Johnson[1,2,3], Sanjay Asthana[1,3], and Carey E.Gleason[1,2,3]

Supplemental Material: Model Identification

Model identification is needed for CFA. Below we described several different approaches for model identification. These approaches would yield the same overall model fit and standardized solution, except that the indicator intercepts and factor means might be shifted. However, their relative difference (e.g., the difference between two indicator intercepts) within a factor would still be maintained.

Model identification for single group CFA
For each factor, make one specification by A and make one specification by B.
    (A1) Fix the factor loading for one indicator at one, or
    (A2) Fix the factor variance at one. And
    (B1) Fix the intercept for one indicator at zero, or
    (B2) Fix the factor mean at zero.

In general, at least three indicators are needed to identify a factor. However, when a factor is sufficiently correlated with other factor(s) in the model, two indicators may also identify the factor. For example, in this study, the fluid cognition factor had only two indicators. It was identified in the whole sample and each individual group, except for in the dementia/MCI group, because the correlation of the fluid cognition and crystalized cognition factors was much lower at 0.22 for this group. Thus an additional parameter (the residual variance of an indicator for the fluid cognition) was fixed at zero to identify the factor.

Model identification for two-group CFA
For each factor, make one specification by A and make one specification by B.
(A1) fix the factor loading for one indicator at one for both groups, or
(A2) fix the factor variance at one for both groups, or
(A3) fix the factor loading for one indicator at one for one group, and constrain the factor loading for this indicator to be equal across groups, or
(A4) fix the factor variance at one for one group, and constrain the factor loading for one indicator to be equal across groups. And
(B1) fix the intercept for one indicator at zero for both groups, or
(B2) fix the factor mean at zero for both groups, or
(B3) fix the intercept for one indicator at zero for one group, and constrain the intercept for this indicator to be equal across groups, or
(B4) fix the factor mean at zero for one group, and constrain the intercept for one indicator to be equal across groups.

Supplemental Material: Sample Mplus Codes

Flanker = Flanker Inhibitory Control and Attention
DCCS = Dimensional Change Card Sort
PicSeq = Picture Sequence Episodic Memory
ListSort = List Sorting Working Memory
PatnComp = Pattern Comparison Processing Speed
PicVocab = Picture Vocabulary
OralRead = Oral Reading Recognition

Part of model specification is not explicitly specified with codes for conciseness, because it is specified in Mplus by default. The Mplus default may change between different models, estimators, or versions. Thus, it is recommended to use "tech1" option (not available for EFA) in the Output statement to request Mplus to provide a list of all parameters to check that the fixed, constrained, and freely estimated parameters are specified as intended.

Title:
EFA for the whole sample (1 to 3 factors)

Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;

Analysis:
Estimator = MLR;
Type = EFA 1 3;

Output:
patterns modindices (all, 0);

Plot:
Type = plot3;

Title:
CFA for the whole sample (2-factor model)

Data:
File = toolbox.dat;

Variables:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;

Analysis:
Estimator = MLR;

2

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by PicVocab OralRead;

Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;

Title:
CFA for the group coded 1 (2-factor model with a cross-loading)

Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Useobservations = group==1;

Analysis:
Estimator = MLR;

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by ListSort PicVocab OralRead;

Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;

Title:
CFA for the group coded 1 (3-factor model)

Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Useobservations = group==1;

Analysis:

3

```
Estimator = MLR;

Model:
f1 by Flanker DCCS PatnComp;
f2 by PicVocab OralRead;
f3 by PicSeq ListSort;


Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;
```

Title:
Two-group CFA testing configural invariance (2-factor model)

```
Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Grouping = group (1 = group1  2 = group2);

Analysis:
Estimator = MLR;

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by PicVocab OralRead;


Model group2:
!Lamda X
f1 by Flanker@1 DCCS PicSeq ListSort PatnComp;
f2 by PicVocab@1 OralRead;

!Tau X
[Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead];

[f1@0];
[f2@0];

Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;
```

4

Title:
Two-group CFA testing metric invariance (2-factor model)

Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Grouping = group (1 = group1 2 = group2);

Analysis:
Estimator = MLR;

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by PicVocab OralRead;

Model group2:
!Tau X
[Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead];

[f1@0];
[f2@0];

Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;

Title:
Two-group CFA testing scalar invariance (2-factor model)

Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Grouping = group (1 = group1 2 = group2);

Analysis:
Estimator = MLR;

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by PicVocab OralRead;

Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;

Title:
Two-group CFA testing residual invariance (2-factor model)

Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Grouping = group (1 = group1 2 = group2);

Analysis:
Estimator = MLR;

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by PicVocab OralRead;

Model group1:
!Theta-Delta
Flanker(13);
DCCS(14);
PicSeq(15);
ListSort(16);
PatnComp(17);
PicVocab(18);
OralRead(19);

Model group2:
!Theta-Delta
Flanker(13);
DCCS(14);
PicSeq(15);
ListSort(16);
PatnComp(17);
PicVocab(18);
OralRead(19);

Output:

6

patterns standardized tech1 modindices (0);

Plot:
Type = plot3;

Title:
Two-group CFA testing factor variance-covariance invariance (2-factor model)

Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Grouping = group (1 = group1 2 = group2);

Analysis:
Estimator = MLR;

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by PicVocab OralRead;

Model group1:
!Theta-Delta
Flanker(13);
DCCS(14);
PicSeq(15);
ListSort(16);
PatnComp(17);
PicVocab(18);
OralRead(19);

!Phi
f1 (21);
f2 (23);
f1 with f2 (22);

Model group2:
!Theta-Delta
Flanker(13);
DCCS(14);
PicSeq(15);
ListSort(16);
PatnComp(17);
PicVocab(18);
OralRead(19);

!Phi

7

```
f1 (21);
f2 (23);
f1 with f2 (22);

Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;
```

Title:
Two-group CFA testing factor mean invariance (2-factor model)

```
Data:
File = toolbox.dat;

Variable:
Names = id group Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Usevariables = Flanker DCCS PicSeq ListSort PatnComp PicVocab OralRead;
Missing = .;
Grouping = group (1 = group1  2 = group2);

Analysis:
Estimator = MLR;

Model:
f1 by Flanker DCCS PicSeq ListSort PatnComp;
f2 by PicVocab OralRead;

Model group1:
!Theta-Delta
Flanker(13);
DCCS(14);
PicSeq(15);
ListSort(16);
PatnComp(17);
PicVocab(18);
OralRead(19);

!Phi
f1 (21);
f2 (23);
f1 with f2 (22);

!Kappa
[f1@0];
[f2@0];

Model group2:
!Theta-Delta
```

8

```
Flanker(13);
DCCS(14);
PicSeq(15);
ListSort(16);
PatnComp(17);
PicVocab(18);
OralRead(19);

!Phi
f1 (21);
f2 (23);
f1 with f2 (22);

!Kappa
[f1@0];
[f2@0];

Output:
patterns standardized tech1 modindices (0);

Plot:
Type = plot3;
```

Supplemental Table S1

Means, Standard Deviations, and Ranges of the Test Scores for Each Demographic Group

| | Sex | | | | | |
| | Male | | | Female | | |
Test	% Missing	*M (SD)*	Range	% Missing	*M (SD)*	Range
Flanker	1.8%	7.3 (1.2)	3.5 - 9.6	0.8%	7.4 (1.2)	2.8 - 9.2
DCCS	2.9%	7.3 (1.6)	2.1 - 9.9	1.7%	7.6 (1.5)	1.5 - 10.0
Processing Speed	1.8%	36.8 (8.7)	6.0 - 57.0	1.3%	37.7 (9.5)	4.0 - 56.0
Working Memory	5.8%	15.7 (3.5)	4.0 - 23.0	2.1%	16.0 (3.6)	4.0 - 24.0
Episodic Memory	11.1%	-1.0 (0.8)	-2.2 - 1.3	7.9%	-0.7 (0.9)	-2.2 - 1.6
Vocabulary	1.8%	6.6 (2.0)	-0.2 - 11.9	1.3%	6.7 (2.3)	-0.5 - 10.6
Reading	1.2%	6.4 (2.8)	-7.0 - 11.5	2.1%	6.5 (2.7)	-3.0 - 11.5

| | Race / Ethnicity | | | | | |
| | URG | | | Non-URG | | |
Test	% Missing	*M (SD)*	Range	% Missing	*M (SD)*	Range
Flanker	1.1%	7.2 (1.1)	4.4 - 9.2	1.3%	7.4 (1.2)	2.8 - 9.6
DCCS	2.2%	7.2 (1.6)	2.1 - 9.9	2.2%	7.5 (1.5)	1.5 - 10.0
Processing Speed	2.2%	35.3 (7.5)	15.0 - 54.0	1.3%	38.0 (9.4)	4.0 - 57.0
Working Memory	1.1%	14.9 (3.5)	6.0 - 24.0	4.5%	16.2 (3.5)	4.0 - 24.0
Episodic Memory	4.4%	-1.1 (0.8)	-2.2 - 1.6	10.5%	-0.8 (0.9)	-2.2 - 1.6
Vocabulary	1.1%	4.8 (2.2)	-0.5 - 9.6	1.6%	7.1 (1.9)	-0.0 - 11.9
Reading	3.3%	4.4 (3.5)	-7.0 - 9.9	1.0%	7.0 (2.2)	-1.3 - 11.5

10

Supplemental Table S1 (Continued)

	Age					
	< 65 years			≥ 65 years		
Test	% Missing	M (SD)	Range	% Missing	M (SD)	Range
Flanker	0.00%	8.0 (0.7)	4.6 - 9.6	0.7%	7.5 (0.8)	5.0 - 9.4
DCCS	0.00%	8.2 (0.8)	6.3 - 10.0	0.7%	7.5 (1.0)	2.4 - 9.9
Processing Speed	0.61%	42.5 (6.3)	27.0 - 57.0	0.7%	36.8 (6.8)	17.0 - 53.0
Working Memory	0.00%	17.8 (2.4)	10.0 - 24.0	0.7%	16.0 (2.7)	10.0 - 23.0
Episodic Memory	1.21%	-0.4 (0.8)	-2.2 - 1.6	1.3%	-0.9 (0.8)	-2.2 - 1.6
Vocabulary	1.21%	7.0 (2.0)	-0.5 - 11.9	0.7%	7.1 (2.0)	0.9 - 10.6
Reading	0.61%	7.1 (2.2)	-0.2 - 11.5	1.3%	6.8 (2.4)	-1.3 - 11.5

	Education					
	Low			High		
Test	% Missing	M (SD)	Range	% Missing	M (SD)	Range
Flanker	2.1%	7.1 (1.3)	3.5 - 9.2	0.7%	7.5 (1.1)	2.8 - 9.6
DCCS	4.9%	7.0 (1.7)	2.1 - 9.8	0.7%	7.6 (1.4)	1.5 - 10.0
Processing Speed	3.5%	35.8 (9.3)	4.0 - 54.0	0.4%	38.2 (9.1)	4.0 - 57.0
Working Memory	4.2%	15.0 (3.9)	4.0 - 24.0	3.4%	16.4 (3.3)	4.0 - 24.0
Episodic Memory	10.4%	-1.1 (0.8)	-2.2 - 1.1	8.6%	-0.7 (0.9)	-2.2 - 1.6
Vocabulary	1.4%	5.4 (2.2)	-0.5 - 10.2	1.5%	7.3 (1.9)	0.4 - 11.9
Reading	2.1%	5.0 (3.2)	-7.0 - 10.7	1.5%	7.3 (2.0)	-0.2 - 11.5

Note. Reasons for missingness included the following: (1) The participant was unable to complete the test because of limited cognitive abilities or other limitations such as poor vision or hearing; (2) The test was automatically skipped if the participant failed on the sample items before the test; (3) The participant refused the test; (4) There was lack of time to administer the test.

Supplemental Table S2

Pearson Correlations of the Test Scores for Each Demographic Group

Lower diagonal: male (*n*s: 151 to 167)						
Upper diagonal: female (*n*s: 219 to 236)						

Test	1	2	3	4	5	6	7
1. Flanker	-	0.66	0.69	0.58	0.39	0.40	0.42
2. DCCS	0.68	-	0.68	0.58	0.34	0.34	0.36
3. Processing Speed	0.62	0.59	-	0.58	0.35	0.37	0.41
4. Working Memory	0.67	0.57	0.48	-	0.45	0.43	0.45
5. Episodic Memory	0.45	0.40	0.30	0.47	-	0.38	0.33
6. Vocabulary	0.37	0.41	0.17	0.44	0.28	-	0.78
7. Reading	0.37	0.36	0.18	0.46	0.25	0.74	-

Lower diagonal: URG (*n*s: to 84 to 89)						
Upper diagonal: Non-URG (*n*s: 279 to 308)						

Test	1	2	3	4	5	6	7
1. Flanker	-	0.67	0.70	0.59	0.39	0.36	0.36
2. DCCS	0.65	-	0.65	0.56	0.38	0.31	0.31
3. Processing Speed	0.45	0.56	-	0.53	0.31	0.29	0.36
4. Working Memory	0.62	0.55	0.46	-	0.45	0.41	0.42
5. Episodic Memory	0.36	0.28	0.22	0.37	-	0.29	0.25
6. Vocabulary	0.53	0.52	0.18	0.42	0.37	-	0.71
7. Reading	0.55	0.49	0.21	0.52	0.33	0.76	-

12

Supplemental Table S2 (Continued)

	Lower diagonal: < 65 years (*n*s: 161 to 165)						
	Upper diagonal: ≥ 65 years (*n*s: 148 to 151)						
Test	1	2	3	4	5	6	7
1. Flanker	-	0.49	0.34	0.39	0.19	0.28	0.35
2. DCCS	0.55	-	0.40	0.45	0.11	0.34	0.35
3. Processing Speed	0.29	0.43	-	0.36	0.23	0.29	0.18
4. Working Memory	0.15	0.19	0.19	-	0.16	0.47	0.42
5. Episodic Memory	0.17	0.18	0.04	0.27	-	0.25	0.19
6. Vocabulary	0.32	0.19	0.13	0.25	0.23	-	0.78
7. Reading	0.23	0.17	0.24	0.36	0.16	0.70	-
	Lower diagonal: low education (*n*s: 127 to 141)						
	Upper diagonal: high education (*n*s: 242 to 265)						
Test	1	2	3	4	5	6	7
1. Flanker	-	0.73	0.66	0.60	0.34	0.28	0.27
2. DCCS	0.57	-	0.64	0.58	0.24	0.28	0.21
3. Processing Speed	0.64	0.64	-	0.53	0.22	0.22	0.26
4. Working Memory	0.60	0.54	0.54	-	0.38	0.34	0.29
5. Episodic Memory	0.45	0.48	0.47	0.53	-	0.24	0.14
6. Vocabulary	0.46	0.38	0.36	0.48	0.38	-	0.65
7. Reading	0.47	0.43	0.34	0.54	0.38	0.79	-

Note. Insignificant correlations ($p > .05$) are underscored.

Supplemental Figure S1.

Scree plot of eigenvalues in the exploratory factor analysis for the whole sample.
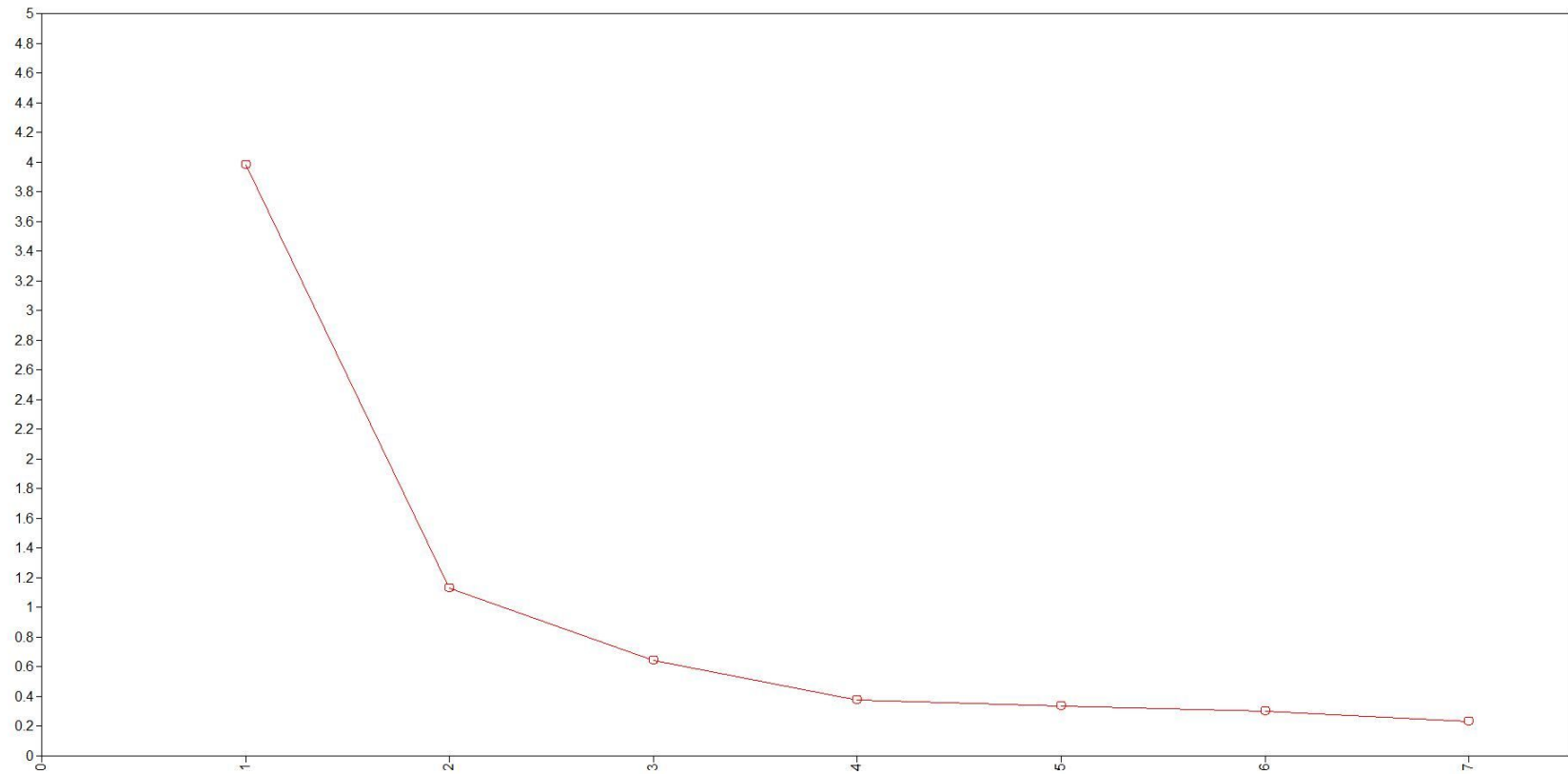
Logistic Regression Analyses for Predicting Missingness on One Test from the Observed Scores on Other Tests

Working Memory		Episodic Memory	
(*n* = 391, *c* = 0.96)		(*n* = 391, *c* = 0.87)	
Predictor	*OR* (95% *CI*)	Predictor	*OR* (95% *CI*)
Flanker	1.01 (0.38, 2.71)	Flanker	**0.61 (0.38, 0.99)**
DCCS	0.64 (0.33, 1.25)	DCCS	0.80 (0.56, 1.14)
Processing Speed	**0.84 (0.71, 0.98)**	Processing Speed	**0.91 (0.85, 0.98)**
Vocabulary	1.22 (0.62, 2.41)	Vocabulary	1.04 (0.75, 1.46)
Reading	1.06 (0.61, 1.83)	Reading	1.07 (0.82, 1.39)
Vocabulary		Reading	
(*n* = 367, *c* = 0.97)		(*n* = 368, *c* = 0.90)	
Predictor	*OR* (95% *CI*)	Predictor	*OR* (95% *CI*)
Flanker	0.35 (0.02, 5.53)	Flanker	0.27 (0.02, 3.85)
DCCS	0.84 (0.05, 12.96)	DCCS	1.53 (0.15, 15.24)
Processing Speed	1.40 (0.96, 2.04)	Processing Speed	0.93 (0.64, 1.34)
Working Memory	0.87 (0.42, 1.80)	Working Memory	1.12 (0.51, 2.48)
Episodic Memory	1.98 (0.43, 9.13)	Episodic Memory	1.45 (0.11, 18.87)
Reading	2.75 (0.55, 13.85)	Vocabulary	1.52 (0.52, 4.48)

Note. For predicting missingness on a memory test, the observed score on the other memory test was not included as a predictor, because missingness on the two memory tests was highly related, and including the other memory test as a predictor would lead to much missingness on the outcome memory test to be excluded from the analysis. For Vocabulary and Reading, the missingness in the outcome only included the recoded missingness from the extremely high scores because of unreliable measurement associated with lack of items with high difficulty levels. *n* = sample size, *c* = concordance statistic (i.e., *c*-statistic). C-statistic indicates the model's discrimination ability in distinguishing between participants who were missing vs. who were not missing on the outcome test. It is calculated by taking all possible pairs of participants consisting one missing and the other not missing on the outcome test, and is equal to the proportion of these pairs in which the participant who was missing on the outcome test had a higher predicted probability of missing than the participant who was not missing. C-statistic equals to the area under the curve (AUC) in the receiver operating characteristics (ROC) curve. It ranges from 0 to 1, and being closer to 1 represents greater ability in predicting missingness. *OR* (95% *CI*) = odds ratio with 95% confidence interval. Higher observed scores (i.e., better performance) on a predictor test would predict higher probability in missing on the outcome test with *OR* > 1, and lower probability in missing with *OR* < 1. Statistically significant *OR* (*p* < .05) are in boldface.
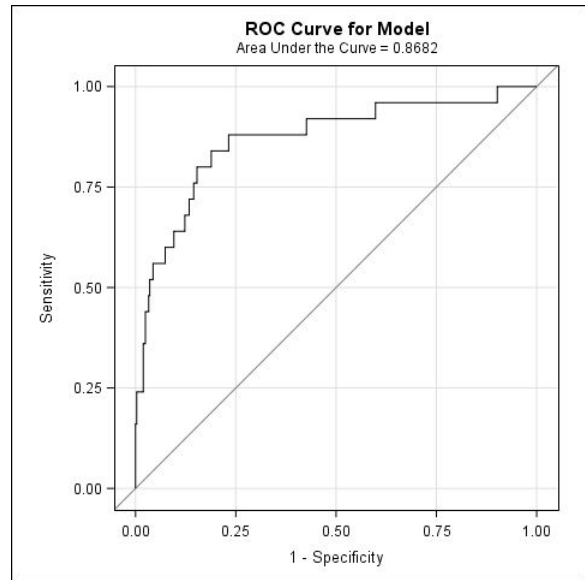
Supplemental Figure S2.

Receiver operating characteristic (ROC) curves and area under the curve (AUC, i.e., *c*-statistic) for logistic regression analyses predicting missingness on Working Memory (S2A) and Episodic Memory (S2B), and predicting recoded missingness on Vocabulary (S2C) and Reading (S2D).
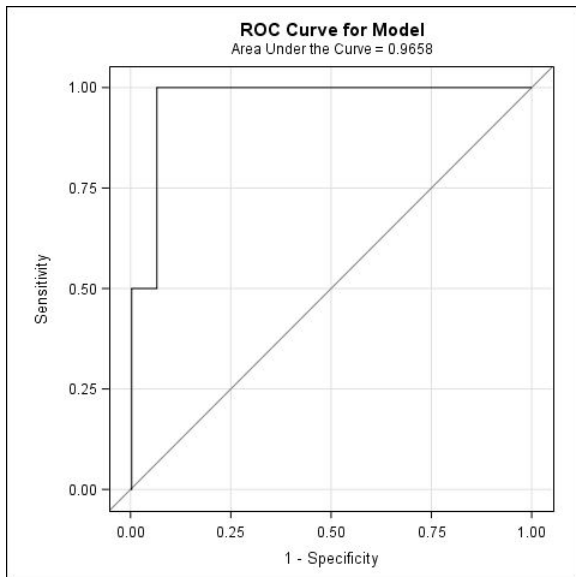
Supplemental Figure S2A: Working Memory
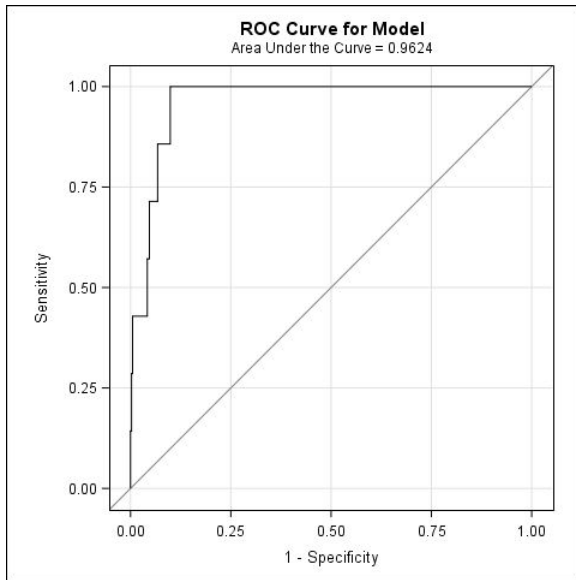
Supplemental Figure S2B: Episodic Memory

Supplemental Figure S2C: Vocabulary

Supplemental Figure S2D: Reading