eAppendix 1

Creation and psychometrics of prorated scores

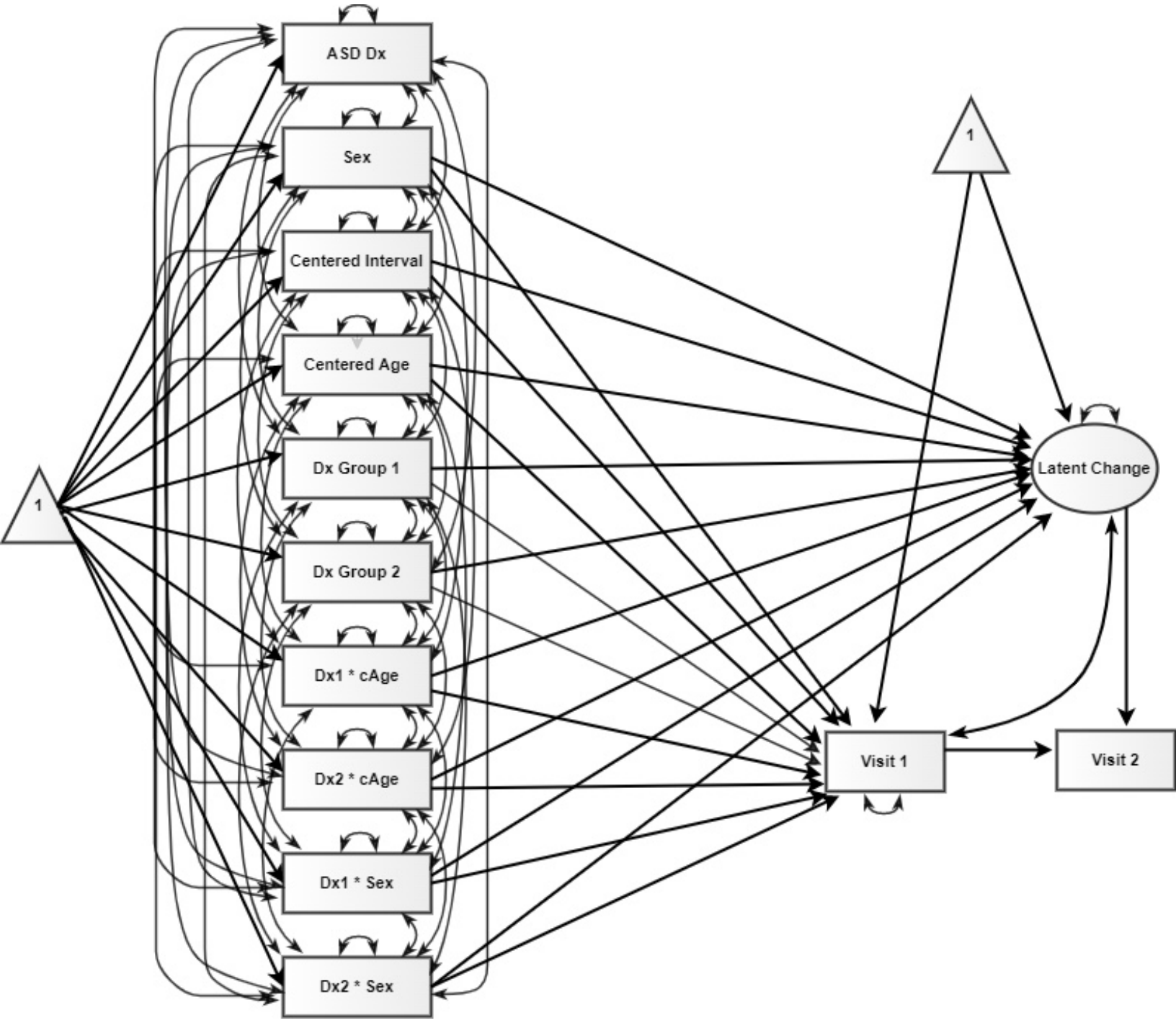
Two of the NIHTB-CB tests, Flanker (FICA) and Dimensional Change Card Sort (DCCS), required creation of alternative scores due to the multiple phases involved in these tests. The FICA test is designed to advance the participant to a second, more difficult test portion (arrows rather than fish) based on a certain level of performance in the first half. This second portion incorporates reaction time in scoring. Some participants received different test lengths at their two visits, depending on performance in the initial test portion. Because of these differing opportunities to receive points and the differing components of the score (i.e., reaction time incorporated only in some scores), evaluating longitudinal performance was challenging. Inspection of the raw FICA scores showed that some participants had what appeared to be unrealistic large increases or declines after the two-year interval, and these were in part due to receiving different items (with different opportunities for scoring) at their visits.

We therefore created a prorated score as a metric that is based on a set number and type of items and that incorporates accuracy with reaction time for all participants. The prorated score (FICA Pro) was created from the 20-item fish portion of the test. The FICA Pro score equals the number of correct items per second from incongruent fish trials. This score had similar reliability and convergent validity to the FICA computed score in our original sample of DS, FXS, and OID (FICA Pro reliability, df=140, ICC=.72, p<.001, 95% CI: .63 to .79; convergent validity, df=123, r=-.42, p<.001, 95% CI: -.56 to -.26).[16]

Similarly, the DCCS test also has different test phases based on performance on simpler phases. The DCCS uncorrected standard score (USS) showed a pattern of large increases or decreases over the two-year period for some participants, primarily but not solely for those who

received different test items at visits. In DCCS, participants first learn an original matching rule (color), then switch to a new matching rule (shape), and finally learn to match by either shape or color based on each trial's cue (mixed phase). We created a prorated score (DCCS Pro) from the 30 mixed test items, calculated as the number of correct items per second. The mixed test phase was used because the earlier and more feasible phases are practice and training phases, with only 5 trials per phase. However, due to the difficulty of learning this task for many of our participants, numbers of participants with a DCCS Pro score were more limited than for the DCCS USS score.

Similar to FICA Pro, DCCS Pro reliability and convergent validity were comparable to the DCCS computed score in our original sample (DCCS Pro reliability, df=74, ICC=0.76, p<.001, 95% CI: .62 to .84; convergent validity, df=80, r=.58, p<.001, 95% CI: .41 to .70). The prorated score showed a potential advantage for longitudinal modeling in that there were fewer unusual large increases/decreases based on different test phases attempted at each visit. However due to the lower number of participants who reached the mixed phase (and have a prorated score) at both visits, we present both the DCCS USS and DCCS Pro score.


ASD Dx
Sex
Centered Interval
Centered Age
Dx Group 1
Dx Group 2
Dx1 * cAge
Dx2 * cAge
Dx1 * Sex
Dx2 * Sex
1
1
Latent Change
Visit 1
Visit 2

eFigure 1.

Structural equation modeling path diagram representing the latent change score models. One model was run for each measure (Stanford Binet 5 Full Scale Change Sensitive Score and all NIH Toolbox tests: Flanker Inhibitory Control and Attention, Dimensional Change Card Sort USS and prorated score, List Sorting Working Memory, Pattern Comparison Processing Speed, Picture Sequence Memory, Picture Vocabulary, and Oral Reading Recognition). The triangles indicate an intercept for the relevant covariates and outcomes. The left column of rectangles shows observed covariates, including an interaction between age and diagnostic group and an interaction between sex and diagnostic group, all of which are correlated with each other. The rectangles on the right are observed scores on the measure at Visit 1 and 2. The oval is the latent change score, representing the estimate of mean change on the measure over two years. Two-headed arrows are variances or covariances; single-headed arrows are regressions. Table e-2 provides additional model specification details for each measure.

eTable 1. Model specification and fit indices

Test	Model	Group covariates in final model		*df*	χ^2 (*p*)	*BIC*
		Visit 1 score regression	Change score regression			
SB5 CSS	Initial			2	2.19 (.33)	9849
	Final	All group effects, sex effect, FXS X sex, OID X age	All group effects, OID X age	8	4.57 (.80)	9818
FICA Pro	Initial			2	0.92 (.63)	6870
	Final	All group effects, OID X age	All group effects, OID X age	11	7.03 (.80)	6810
DCCS USS	Initial			2	0.05 (.98)	9339
	Final	All group effects, sex effect, FXS X sex	All group effects, DS X age	9	9.83 (.37)	9270
DCCS Pro	Initial			2	2.55 (.28)	6751
	Final	All group effects	All group effects	12	6.73 (.88)	6671
LSWM	Initial			2	0.61 (.74)	9179
	Final	All group effects; sex effect, FXS X sex	All group effects; DS X age	9	7.66 (.57)	9148
PCPS	Initial			2	0.41 (.81)	9361
	Final	All group effects; sex effect, FXS X sex	All group effects	10	9.09 (.52)	9325
PSM	Initial			2	2.44 (.30)	9511
	Final	All group effects, sex effect, FXS X sex	All group effects	10	4.96 (.89)	9469
PV	Initial			2	1.58 (.45)	9904
	Final	All group effects, sex effect, FXS X sex, DS X age	All group effects, sex effect, FXS X sex, DS X age	6	2.77 (.84)	9884
ORR	Initial			2	0.37 (.83)	9843
	Final	All group effects, sex effect, FXS X sex, OID X age	All group effects, OID X age	8	3.91 (.87)	9813

Initial models included all covariates and group-specific effects except autism diagnosis, which covaried with other observed variables but regression weight was fixed to zero. From the initial models, two modification steps were taken to evaluate fit. The $\chi 2$ test of model fit was used; a significant p-value indicates a significantly poorer fit than the baseline model. First, the effect of sex was evaluated; if fixing the regression weight to zero did not harm model fit, the effect was fixed to avoid over-constraining the model and increase power. If the effect of sex was included, the FXS*sex interaction was evaluated. Second, the same was done for group*age

interactions. The final models for FICA and DCCS Pro eliminate the effect of sex. We also provide BIC indices as an additional model fit metric. A lower BIC value represents better fit; when evaluating competing models, a BIC difference of at least 2 indicates a better fit in the model with smaller BIC.

BIC, Bayesian information criterion; SB5, Stanford Binet Intelligence Scales, Fifth Edition; FICA, Flanker Inhibitory Control and Attention; DCCS, Dimensional Change Card Sort; LSWM, List Sorting Working Memory; PCPS, Pattern Comparison Processing Speed; PSM, Picture Sequence Memory; PV, Picture Vocabulary; ORR, Oral Reading; CSS, Change Sensitive Score; USS, uncorrected standard score.

eTable 2. Age frequency of participants included in analysis (i.e., with valid score) from Visit 1

	n																				
	Age at Visit 1 (years)																				
Test	6 - 8.99			9 – 11.99			12 – 14.99			15 – 17.99			18 – 20.99			21 – 23.99			24 – 25.99		
	O	F	D	O	F	D	O	F	D	O	F	D	O	F	D	O	F	D	O	F	D
SB5	14	7	6	15	13	21	22	14	17	12	21	16	8	9	13	9	8	10	6	6	9
FICA Pro	12	4	4	9	8	16	18	9	14	5	8	12	5	7	12	6	5	9	2	4	7
DCCS USS	9	3	1	11	7	10	20	8	12	10	9	11	5	6	7	8	5	9	5	4	5
DCCS Pro	4	2	0	5	5	7	15	3	8	9	4	5	3	4	5	6	5	8	4	4	4
LSWM	11	3	4	11	8	13	19	8	7	9	8	9	5	8	11	8	7	8	5	4	6
PCPS	10	6	4	11	10	15	20	11	9	10	11	9	6	8	7	8	8	9	5	4	7
PSM	10	6	5	11	9	16	19	5	11	10	14	9	6	5	11	8	7	10	6	4	7
PV	13	7	6	15	13	20	22	12	16	12	20	15	8	9	13	9	8	10	6	6	9
ORR	12	6	6	15	11	20	22	13	15	12	20	14	6	9	13	9	8	10	6	6	9

D, Down syndrome group; DCCS, Dimensional Change Card Sort; F, Fragile X syndrome group; FICA, Flanker Inhibitory Control and Attention; LSWM, List Sorting Working Memory; O, Other intellectual disability group; ORR, Oral Reading; PCPS, Pattern Comparison Processing Speed; Pro, prorated score; PSM, Picture Sequence Memory; PV, Picture Vocabulary; SB5, Stanford Binet Intelligence Scales, Fifth Edition; USS, uncorrected standard score.

eAppendix 2

Regression effects of sex and sex*group on latent intercepts and change scores

Stanford-Binet 5 (SB5)

Regarding Visit 1 score, there was a significant main effect of sex on Visit 1 intercept (b=-4.0, SE=1.5, p=.007), such that males had a lower score than females. There was also a significant sex*group interaction (p<.001); in FXS, males had a lower SB5 starting level than females (b=-16.6, SE=2.7, p<.001); there was no effect of sex in the other groups (b=0.5, SE=1.6, p=.74).

Dimensional Change Card Sort (DCCS)

There was no main effect of sex on Visit 1 intercept (b=-2.84, SE=2.74, p=.30). However, there was a significant sex*group interaction, such that effect of sex on Visit 1 DCCS USS score differed for FXS compared to the other groups (p=.020). In FXS, males had a lower DCCS USS starting level than females (b=-13.4, SE=5.3, p=.011); there was no effect of sex in the other groups (b=0.95, SE=3.15, p=.76).

List Sorting Working Memory (LSWM)

There was no main effect of sex on Visit 1 score (b=-1.98, SE=2.01, p=.32). but there was a significant sex*group interaction between FXS and the other groups (p<.001). In FXS, males had a lower LSWM starting level than females (b=-15.4, SE=3.6, p<.001); there was no effect of sex in the other groups (b=3.15, SE=2.26, p=.74).

Pattern Comparison Processing Speed (PCPS)

There was a significant main effect of sex on Visit 1 PCPS intercept, with males having a lower starting level than females (b=-6.35, SE=2.65, p=.02). As seen with other measures, there was a significant sex*group effect on Visit 1 PCPS score; in FXS, males had a lower score at

Visit 1 than females (b=-15.5, SE=4.90, p=.001); this comparison differed from the other groups (p=.027), in which there was no effect of sex on Visit 1 intercept (b=-2.69, SE=3.10, p=.39).

Picture Sequence Memory (PSM)

There was a significant main effect of sex on Visit 1 PSM level (b=-3.96, SE=1.62, p=.014), with males having a lower starting score than females; there was also a sex*group effect on Visit 1 score. Males with FXS had a lower score at Visit 1 than FXS females (b=-13.26, SE=3.57, p<.001); this comparison differed from the other groups (p<.001), in which there was no effect of sex on Visit 1 intercept (b=-0.48, SE=1.83, p=.79).

Picture Vocabulary (PV)

There was no main effect of sex on PV starting level (b=-2.01, SE=1.55, p=.19); however, there was a significant sex*group interaction (b=-13.89, SE=3.39, p<.001). Visit 1 PV intercept was lower for FXS males than FXS females, and this differed from OID and DS. The OID and DS groups had no effect of sex on PV starting level (b=1.68, SE=1.75, p=.34). There was no main effect of sex on PV change score (b=-1.75, SE=2.13, p=.41). The sex*group interaction effect on PV change score was also not significant (b=-0.17, SE=2.54, p=.95).

Oral Reading Recognition (ORR)

Sex predicted Visit 1 intercept, with higher ORR score in females (b=-4.82, SE=1.73, p=.005). There was again a sex*group effect on Visit 1 score. Males with FXS had a lower starting level than FXS females (b=-17.03, SE=3.23, p<.001); this differed from other groups (p<.001), in which there was no effect of sex on Visit 1 intercept (b=-0.40, SE=1.94, p=.84).

eTable 3. β-coefficients and *SE*s for latent change scores regressed on group covariates at centered ages

Test	FXS compared to OID[a]			DS compared to OID[a]			DS Group compared to FXS[a]		
	10 years	16 years	22 years	10 years	16 years	22 years	10 years	16 years	22 years
SB5 FS CSS	-2.4 (1.5)	-0.4 (1.1)	1.5 (1.6)	-0.8 (1.5)	1.2 (1.1)	3.1 (1.6)*	1.6 (1.1)[b]	1.6 (1.1)	1.6 (1.1)
FICA	0.1 (0.1)	0.0 (0.1)	0.1 (0.1)	0.0 (0.1)	0.1 (0.1)	0.2 (0.1)*	0.1 (0.1)[b]	0.1 (0.1)	0.1 (0.1)
DCCS USS	-1.9 (4.6)[b]	-1.9 (4.6)	-1.9 (4.6)	-10.6 (5.6)	-4.5 (4.3)	1.7 (5.8)	-8.8 (6.1)	2.6 (4.7)	3.6 (5.9)
DCCS Pro	0.0 (0.1)[b]	0.0 (0.1)	0.0 (0.1)	0.0 (0.1)[b]	0.0 (0.1)	0.0 (0.1)	-0.1 (0.1)[b]	-0.1 (0.1)	-0.1 (0.1)
LSWM	-3.0 (2.9)[b]	-3.0 (2.9)	-3.0 (2.9)	-3.9 (3.7)	0.2 (2.8)	4.3 (3.7)	-0.9 (3.9)	3.2 (2.9)	7.3 (3.7)
PCPS	2.7 (2.9)[b]	2.7 (2.9)	2.7 (2.9)	4.3 (2.9)	4.3 (2.9)	4.3 (2.9)	1.6 (3.0)[b]	1.6 (3.0)	1.6 (3.0)
PSM	-0.1 (2.6)[b]	-0.1 (2.6)	-0.1 (2.6)	1.8 (2.4)[b]	1.8 (2.4)	1.8 (2.4)	1.1 (4.3)[b]	1.1 (4.3)	1.1 (4.3)
PV	-1.5 (2.2)[b]	-1.5 (2.2)	-1.5 (2.2)	-4.4 (1.9)*	-1.2 (1.4)	2.0 (2.0)	-2.9 (2.5)	0.3 (2.1)	3.4 (2.6)
ORR	-4.4 (1.8)*	-3.3 (1.3)*	-2.2 (1.9)	-1.9 (1.8)	-0.9 (1.3)	0.2 (1.9)	2.4 (1.3)[b]	2.4 (1.3)	2.4 (1.3)

[a]Group effects on change score at centered age; a significant negative result represents a smaller change score in first group compared to second group.
[b]Model eliminated Age*Group interaction effect for this group comparison; group effect for this comparison is for full age range
*p<.05; **p<.01; ***p<.001

eTable 4. β-coefficients and *SE*s for change scores regressed on sex, interval, and age*group interaction

Test	Sex[a] β (SE)	Interval[b] β (SE)	Age*Group interaction[c] β (SE)			Sex*Group interaction[d] β (SE)		
			FXS:OID	DS:OID	DS:FXS	FXS:OID	DS:OID	DS:FXS
SB5 FS CSS	-	-0.1 (0.1)	0.3 (0.2)	0.3 (0.2)	-	-	-	-
FICA	-	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	-	-	-	-
DCCS USS	-	0.0 (0.3)	-	1.0 (0.6)	-1.0 (0.6)	-	-	-
DCCS Pro	-	0.0 (0.0)	-	-	-	-	-	-
LSWM	-	0.4 (0.2)	-	0.7 (0.4)	-0.7 (0.5)	-	-	-
PCPS	-	-0.1 (0.2)	-	-	-	-	-	-
PSM	-	0.0 (0.2)	-	-	-	-	-	-
PV	-1.8 (2.1)	-0.1 (0.1)	-	0.5 (0.2)*	0.5 (0.2)*	-0.2 (2.5)	-	0.2 (2.5)
ORR	-	0.1 (0.1)	0.2 (0.2)	0.2 (0.2)	-	-	-	-

[a]Reference group is female; a significant positive coefficient would represent a larger change score in males than females
[b]Interval between visits, centered at 24 months
[c]A significant age*group interaction represents a difference in change score between given groups depending on age
[d]A significant sex*group interaction represents a difference in change score between given groups depending on sex
*p<.05; **p<.01; ***p<.001