These tables illustrate a couple interesting technical issues related to ROC and logistic regression. The tables are reproduced in their entirety in this version, with markup not included in the published version that calls out the teaching points.

Key issues:

- Collinearity of predictors (due to shared source variance)
- Interaction effects
- Clinical interpretability

Executive summary:

- The PCSI self-report ROC is significant and the AUC substantial (as were the *t*-tests and *d* values).
- The PCSI and BRIEF are highly correlated (not surprising – shared method variance, similar constructs).
- Putting them into the same block of the logistic regression changes the interpretation of the *b* for each. It's no longer a test of whether the scale predicts the outcome (as the reviewer interpreted it); it's a test of whether the scale makes an incremental contribution above and beyond what both self-report scales are doing. "Team Self-report" is highly significant (see the 2 *df* chi-squared for the block, $p < .0005$), but neither scale is a super-star (or ball hog!) that shines above the other. (The Venkatraman test is a more powerful approach for asking whether one does better than the other; the incremental efficacy is a related issue, but not exactly the same thing.)
- The interaction effect tells us that the effect is stronger in High School for the PCSI, but doesn't mean that it is not significant in the younger age group. There are several ways of looking for whether it still works in the younger. Alison is right that simple slopes (and a test of significance for the slope in the younger group) is one way of doing it. Two alternatives are probably easier and faster in practice (though technically less elegant). I put them in the annotations. One is running the regression again with only the PSCI; the other is subsetting the file and running the ROC for PCSI only in the "High School = no" subsample. Both are quick, and they would be the last bullet point in allaying the reviewer's concern.

Table 1

Descriptive Statistics for Clinical and Demographic Variables and Bivariate Tests of Association with Perceived School Problems Status at Visit One (N=142)

Variable	**School Problems: High (*n* = 63)**	**School Problems: Low (*n* = 79)**	**Test statistic**	***p***	**Effect size**
Age in years					
Mean	15.14	14.79	*t* (140 *df*) = 1.14	.26	*d* = .19
SD	1.68	1.89			
Female	*n* = 28 (44%)	*n* = 30 (38%)	χ^2 (1 *df*) = 0.61	.44	phi = .07
Race (Caucasian %)	*n* = 50 (83%)	*n* = 58 (77%)	χ^2 (1 *df*) = 0.75	.39	phi = .07
Level of schooling (High School)	*n* = 41 (65%)	*n* = 49 (62%)	χ^2 (1 *df*) = 0.14	.71	phi = .03
Pre-Injury History					
ADHD/LD	*n* = 17 (27%)	*n* = 15 (19%)	χ^2 (1 *df*) = 1.28	.26	phi = .09
Anxiety/Mood disorder	*n* = 16 (25%)	*n* = 16 (20%)	χ^2 (1 *df*) = 0.53	.47	phi = .06
Headaches/migraines	*n* = 31 (49%)	*n* = 28 (35%)	χ^2 (1 *df*) = 2.73	.10	phi = .14
At least one of the above	*n* = 46 (73%)	*n* = 42 (53%)	χ^2 (1 *df*) = 5.86	.02	phi = .20
At least two of the above	*n* = 16 (25%)	*n* = 14 (18%)	χ^2 (1 *df*) = 1.24	.27	phi = .09
All three of the above	*n* = 2 (3%)	*n* = 2 (3%)	χ^2 (1 *df*) = 0.05	.82	phi = .02
Injury: sport-related concussion	*n* = 54 (87%)	*n* = 61 (77%)	χ^2 (1 *df*) = 2.26	.13	phi = .13
Injury characteristics					
Loss of consciousness	*n* = 9 (15%)	*n* = 7 (9%)	χ^2 (1 *df*) = 1.12	.29	phi = .09
No recall of impact	*n* = 26 (41%)	*n* = 22 (28%)	χ^2 (1 *df*) = 2.82	.09	phi = .14
Retrograde amnesia	*n* = 12 (19%)	*n* = 11 (14%)	χ^2 (1 *df*) = 0.69	.41	phi = .07
Anterograde amnesia	*n* = 18 (27%)	*n* = 14 (18%)	χ^2 (1 *df*) = 2.12	.15	phi = .12
Seizures	*n* = 1 (2%)	*n* = 1 (1%)	χ^2 (1 *df*) = 0.02	.89	phi = .01
Days since injury *Mean (SD)*	18.29 (6.04)	15.81 (6.04)	*t* (140 *df*) =2.43	.02	*d* = 0.41
PCSI Self-Report mean score					
Mean	1.26	.43	*t* (140 *df*) = 6.78	<.001	*d* = 1.18
(SD)	(.83)	(.55)			
PCSI Parent Report mean score					
Mean	[illegible]	[illegible]	[illegible]	<.001	*d* = .98
(SD)	[illegible]	[illegible]			
BRIEF Self-Report raw total					
Mean	22.48	7.41	*t* (140 *df*) = 6.50	<.001	*d* = 1.12
(SD)	(15.70)	(10.79)			
BRIEF Parent-Report raw total					
Mean	14.37	6.03	*t* (140 *df*) = 3.77	<.001	*d* = .63
(SD)	(14.64)	(11.70)			
Exertional Effects Index					
Mean	3.75	2.10	*t* (140 *df*) = 3.41	<.001	*d* = .48
(SD)	(3.13)	(3.75)			
Cognitive measures (ImPACT/MACS):					

The two self-report measures both produce large effect sizes comparing the High and Low school problems groups

Processing Speed SS					
Mean	90.87	93.52	*t* (140 *df*) = .92	.36	*d* = .16
(SD)	(15.79)	(17.99)			
Memory SS					
Mean	91.75	94.96	*t* (140 *df*) = 1.18	.24	*d* = .19
(SD)	(16.50)	(15.73)			

Table 2

Correlations among Variables

Variable	Female	High School	Pre-injury History[a]	PCSI-Self	PCSI-Parent	BRIEF-Self	BRIEF-Parent	EEI	Processing Speed SS[c]	Memory SS[c]
School Problems: High[b]	-.07 [d]	.03 [d]	.20** [d]	.51*** [e]	.44*** [e]	.50*** [e]	.30*** [e]	.28*** [e]	-.08 [e]	-.10 [e]
Gender: Female		.13 [d]	.14 [d]	.25***[e]	.18*[e]	.21** [e]	.21** [e]	.09 [e]	.03 [e]	.06 [e]
School Level: High School			-.01[d]	-.09[e]	.07[e]	.11 [e]	.13 [e]	.09 [e]	.17* [e]	.07 [e]
Pre-injury History[a]				.04[e]	.17*[e]	-.04 [e]	-.01 [e]	-.06[e]	.15 [e]	.04 [e]
PCSI Self					.58***	.68***	.42***	.39***	-.28***	-.25**
PCSI Parent						.52***	.65***	.30***	-.14	-.15
BRIEF Self							.55***	.26**	-.23**	-.19*
BRIEF Parent								.22**	-.14	-.15
Exertional Effects Index									-.16	.11
Processing Speed SS[c]										.52***

The two self-report measures are highly correlated with each other – shared source variance, as well as measuring similar constructs (incidentally, the parent-youth agreement is exceptionally high!)

Note. [a]Pre-Injury History includes diagnoses of ADHD, Learning Disability, Anxiety, Depression, or personal history of headaches/migraines.
[b] coded such that low academic problems = 0 and high academic problems = 1
[c]Standard Score; all others are raw scores, adjusted for retrospective ratings of pre-injury functioning.
[d]Phi Coefficient.
[e]Point-biserial correlation; all others are Pearson *r* correlations.
*p < .05, **p < .01, ***p < .001, two-tailed.

The high correlations among the PCSI and BRIEF scales are consistent with the idea of them being converging measures of the same construct. We didn't report a structural equation model in the paper, but conceptually it helps show what is going on in the later regression models.

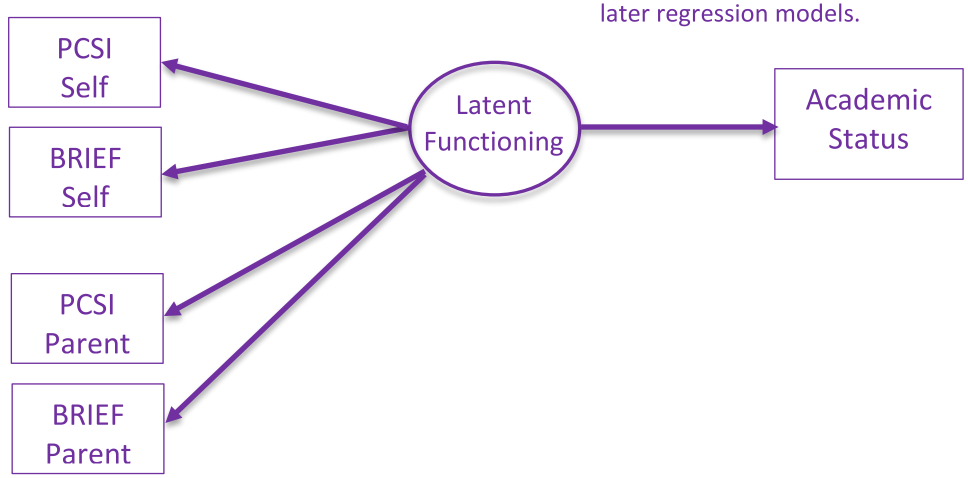

Table 3

Area under the Curve from Receive Operating Characteristic Analyses Identifying Students reporting School Problems at Visit One with Index Tests and Moderators

				95% Confidence Interval	
Index Test	**Area under curve**	**Standard error**	***p* value**	**Lower**	**Upper**
BRIEF Self-Report	.84	.03	<.001	.78	.91
PCSI Self-Report	.80	.04	<.001	.73	.87
PCSI Parent Report	.79		1	.72	.87
BRIEF Parent-Report	.74		1	.66	.83
Exertional Effects Index	.70		1	.61	.78
Cognitive performance:					
Processing Speed	.56			.47	.66
Memory	.57	.05	.17	.47	.66

The two self-report measures both produce large effect sizes comparing the High and Low school problems groups

Table 4

Logistic Regression Model Identifying Students reporting High Levels of Post-Injury School Problems

	B	**Standard error**	***p* value**	**Odds Ratio**
Block 0:				
Intercept	-0.99	0.54	0.06	0.37
Block 1: $\chi 2$ =7.00 , p= .07				
Gender (Female)				
School Level (High school)				
Pre-injury Characteristics[a]				
Block 2: $\chi 2$ = 54.43, p< .001				
BRIEF Self-Report	0.03	0.03	0.27	1.03
PCSI Self-Report	0.07	0.55	0.90	1.08
Bloc				
P				
B				
Block 4: $\chi 2$ =11.74 , p= .003				
High School x PCSI Self-Report	1.87	0.84	0.03	6.47
Pre-injury History x BRIEF Self-Report	0.13	0.06	0.02	1.14
Block 5: $\chi 2$ =1.13 , p= .				
Exertional Effects Inc				
Cognitive measures[b]:				
Processing Speed				
Memory				

Note. [a]presence of pre-inju
[b]ImPACT and MACS Pro
greater difficulty.

Putting the two self-report measures in as the same block produces a large improvement in prediction. We could emphasize this by adding an R^2 change for the block, thought the chi-squared and *p*-value are convincing, too.

Entering in the same block, the *b* for each is no longer a test of the main effect, but rather the incremental effect (or unique effect) of each scale above & beyond the other. Neither being significant independently tells us that it is a functional tie (and no value in giving both or interpreting both in combination for this dependent variable)

The significant interaction effect for High School tells us that the discriminative power is significantly better for the older age group. Just looking at this, we can't tell whether it shrinks to non-significance in the lower age group. One way of checking would be to estimate simple slopes for the regression. Another would be to split the file and run the ROC separately in the young age group. The second usually will be the easier approach, and often easier for audiences to understand.

Another quick approach would be to run the logistic regression again without the BRIEF self-report included. If High School is dummy coded, then the b for PCSI is the slope for the younger age group, and the interaction is how much the slope changes (improves, since the sign is positive) for High School. With the BRIEF excluded, the PSCI should be significant (it will be getting credit for all the covariance that was collinear with the BRIEF in the model that is tabled here); but if it isn't, that will be the same result that we would identify running the ROC separately in the subset.

An aside about collinearity: The significant correlation between predictors (gray in Table 2) means that they have some "collinearity." There are two ways this can affect the regression results. One is shared prediction, which is what we will see in the later tables. We can think of this as "conceptual collinearity."

The second way is when the collinearity is so high that the regression no longer can produce accurate results. Functionally, the predictors (or sets of predictors) are "identical twins" and the regression can't attribute the prediction to one variable with any confidence. The standard errors become huge, the confidence intervals blow up, and the betas "bounce" around a lot on cross-validation. This is when the tolerance is <.10 (or the R^2 for the overlap with a predictor is >.90). We are a not yet into that scenario here, so we can trust the regression results, and they show that we have "conceptual collinearity" in the sense that two scales from the same informant provide functionally redundant information.

We want to be careful that our presentation doesn't trigger a "kick us here" reaction from reviewers if we mention collinearity.