

Select Psychometric Properties and Predictive Validity of Scores on the SAT Writing Section

Annual Meeting of AERA
San Diego, CA
April 14, 2009

Thomas P. Proctor
YoungKoung Rachel Kim
The College Board

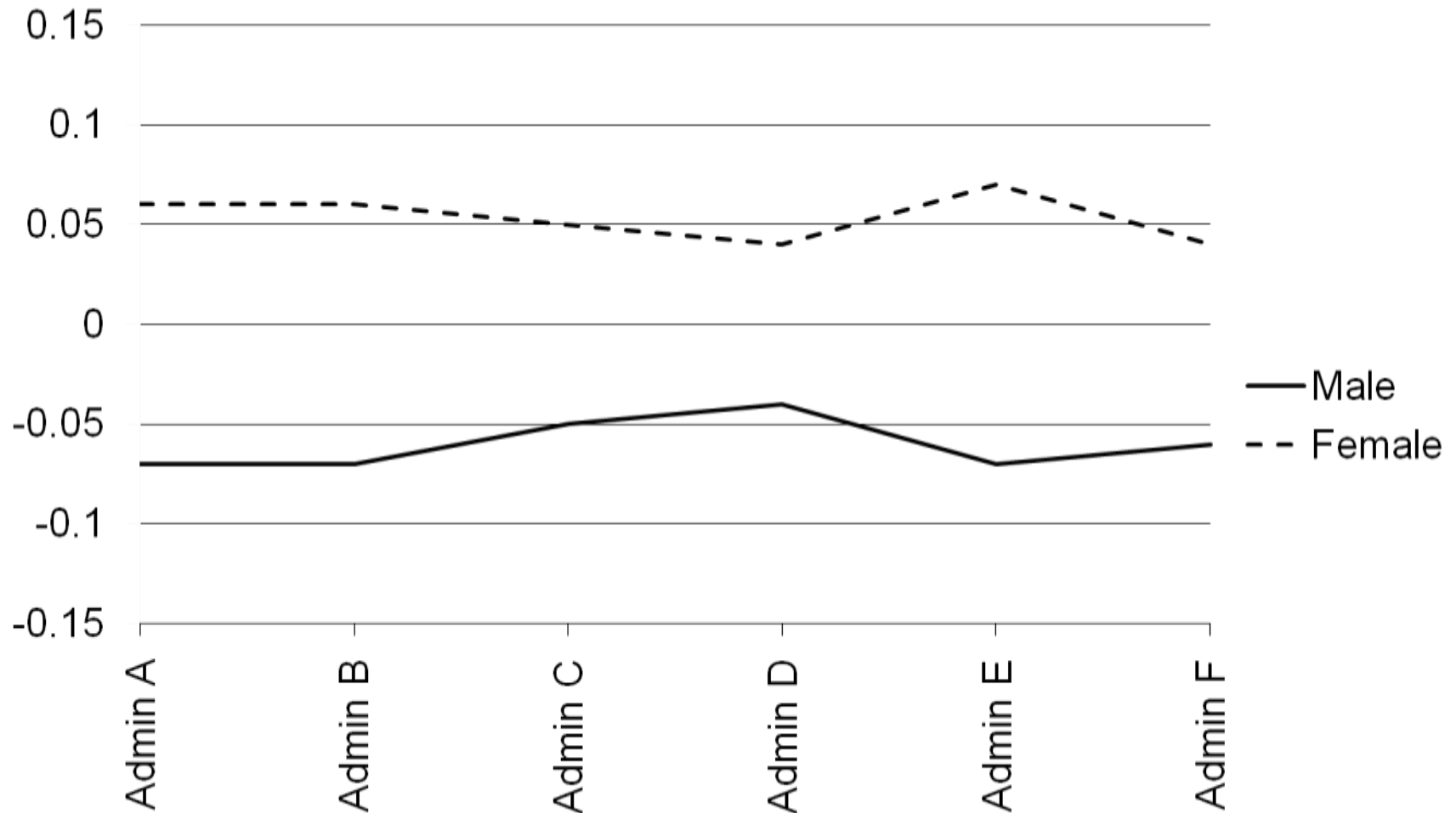
Background

- Purpose of adding the writing section:
 - Strengthen writing in our nation's schools and colleges
 - Reinforce the importance of writing skills throughout a student's education
- SAT Writing first administered March 2005
 - 30 regular administrations
 - 3 graduating cohorts
- Some institutions adopted immediate use of scores & others adopted a wait-and-see approach
 - Waiting for research on the utility of writing scores

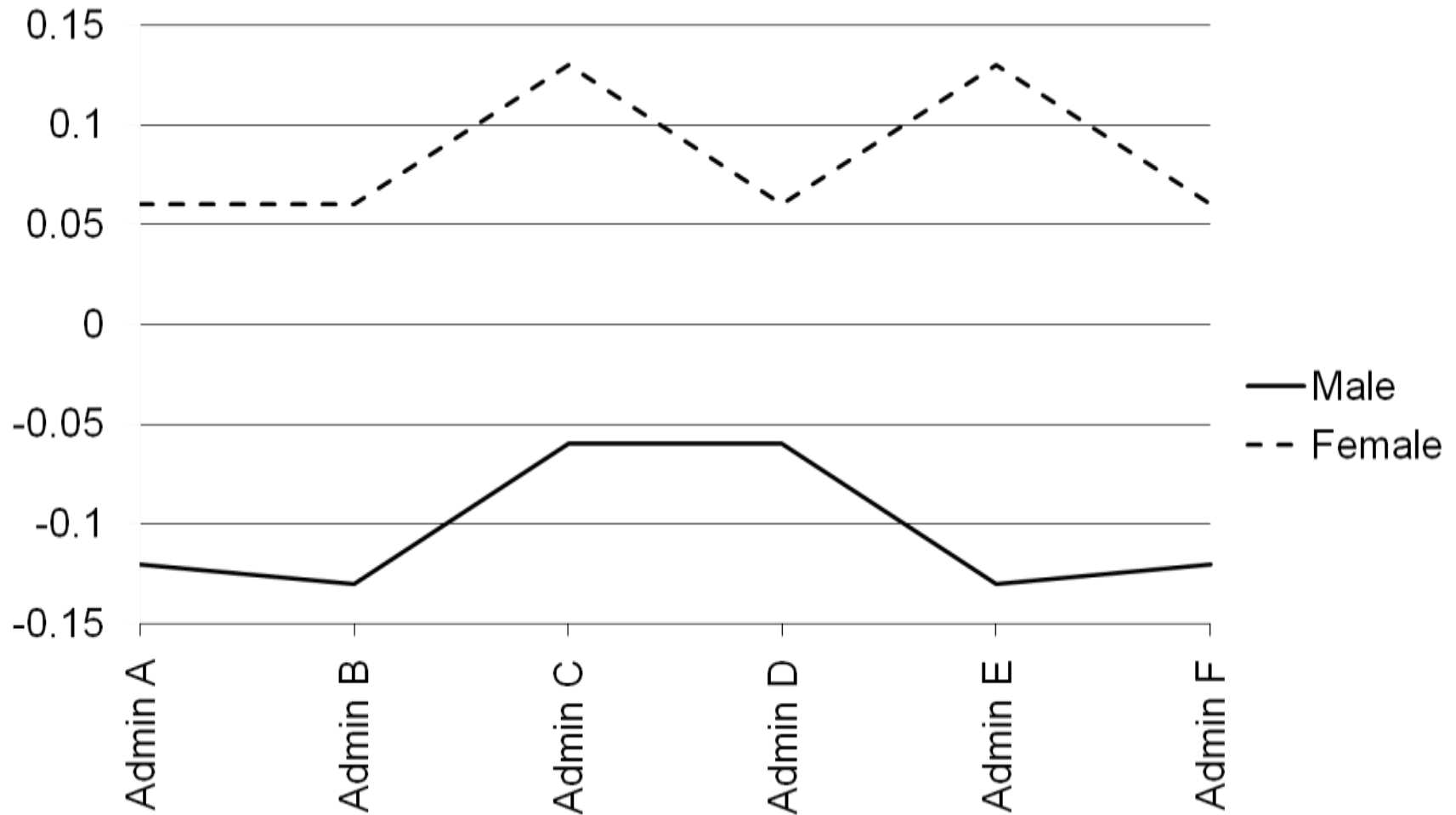
Description of SAT Writing

- 49 multiple-choice items & 1 25-minute essay
- MC Items are scored +1 for correct answer, -1/4 for an incorrect answer, and not scored for omitted or not-reached items
- Essay is scored 0 to 6 by two independent readers
 - Final Essay is scored 0, 2 – 12
 - A score of 0 is given if essay is blank, off topic, etc...
 - If 2 readers differ by more than 1 point, a third reader is used and that score is doubled for final score

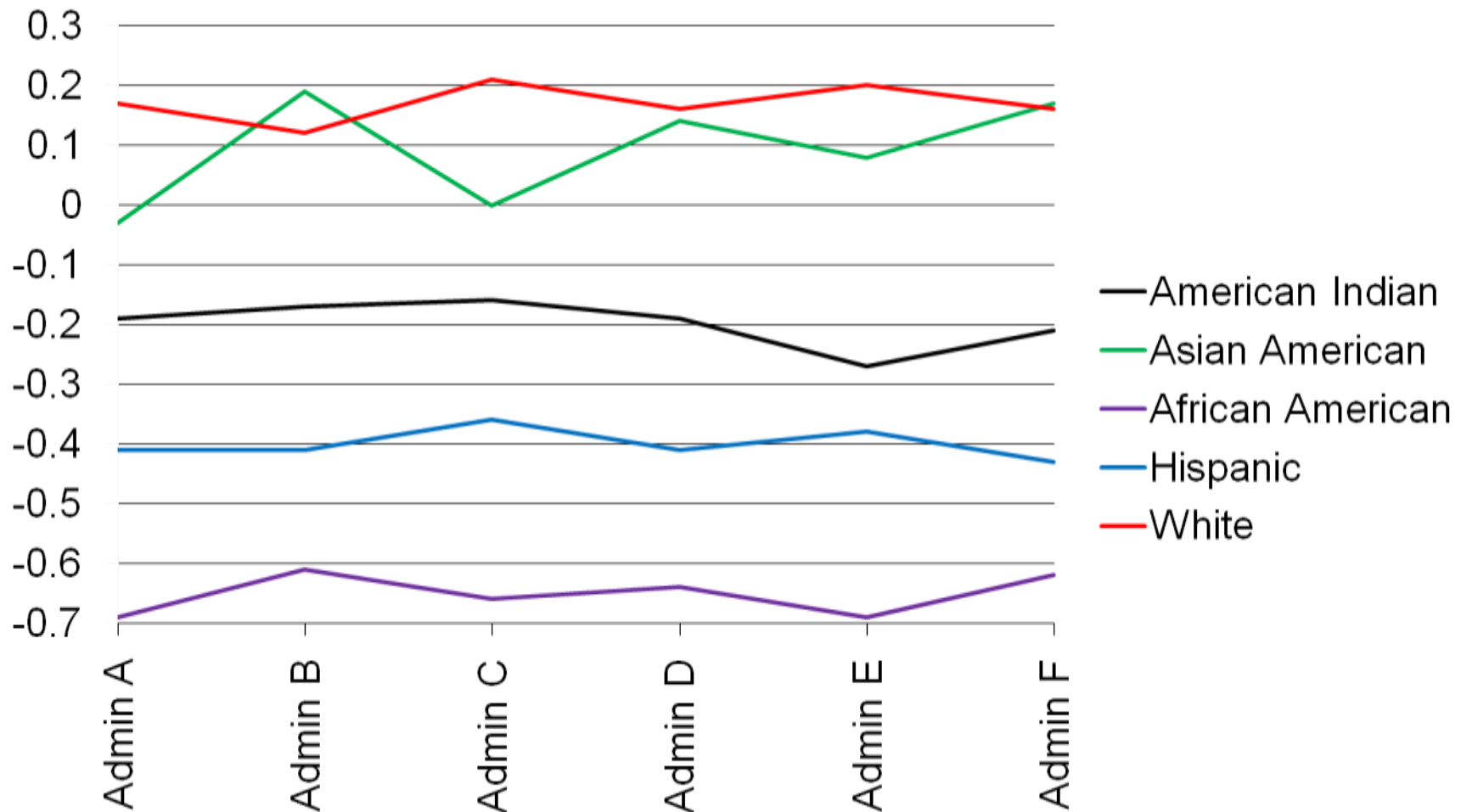
Standardized Differences for Males & Females Writing Composite Score



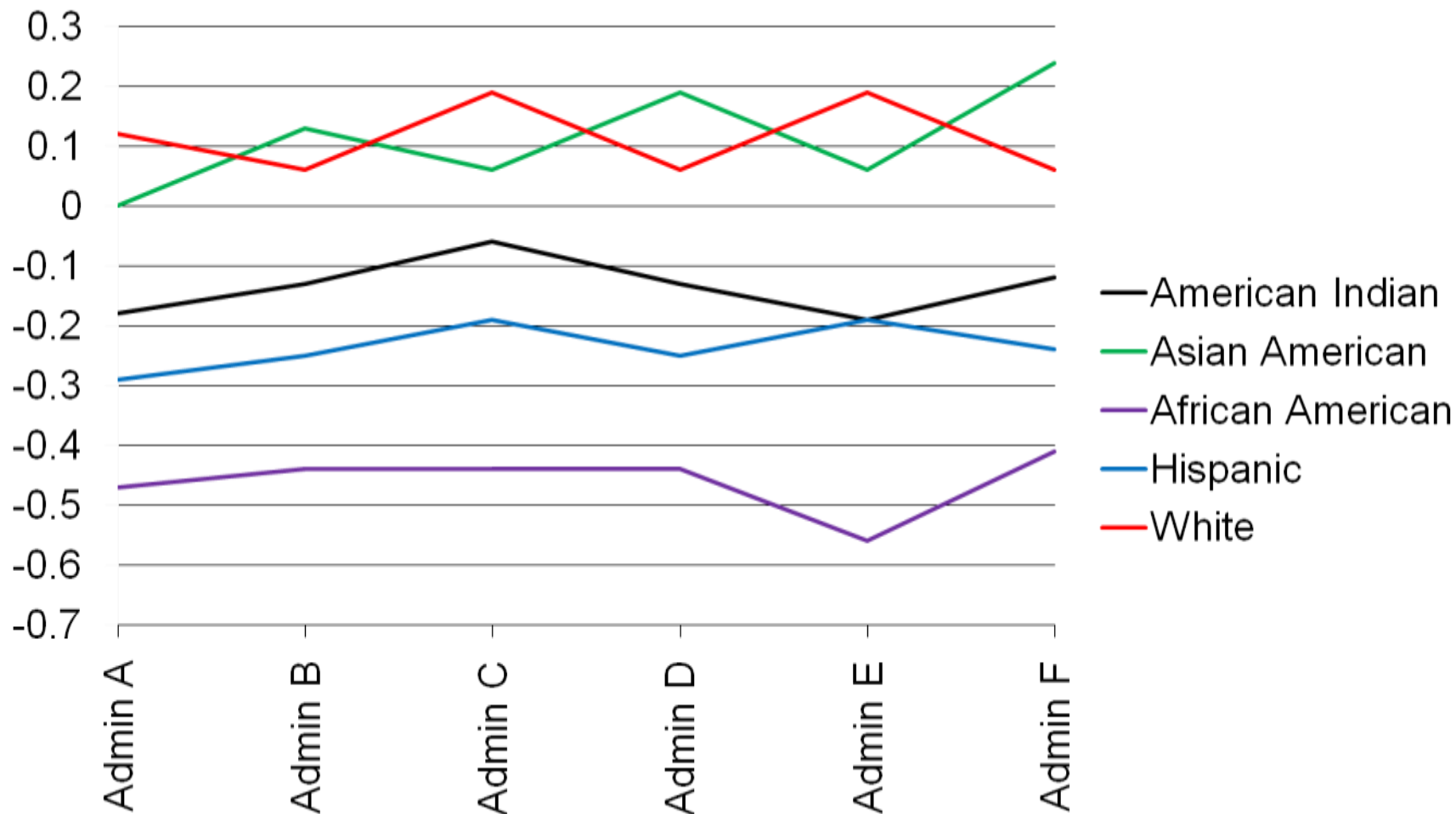
Standardized Differences for Males & Females Essay Raw Score



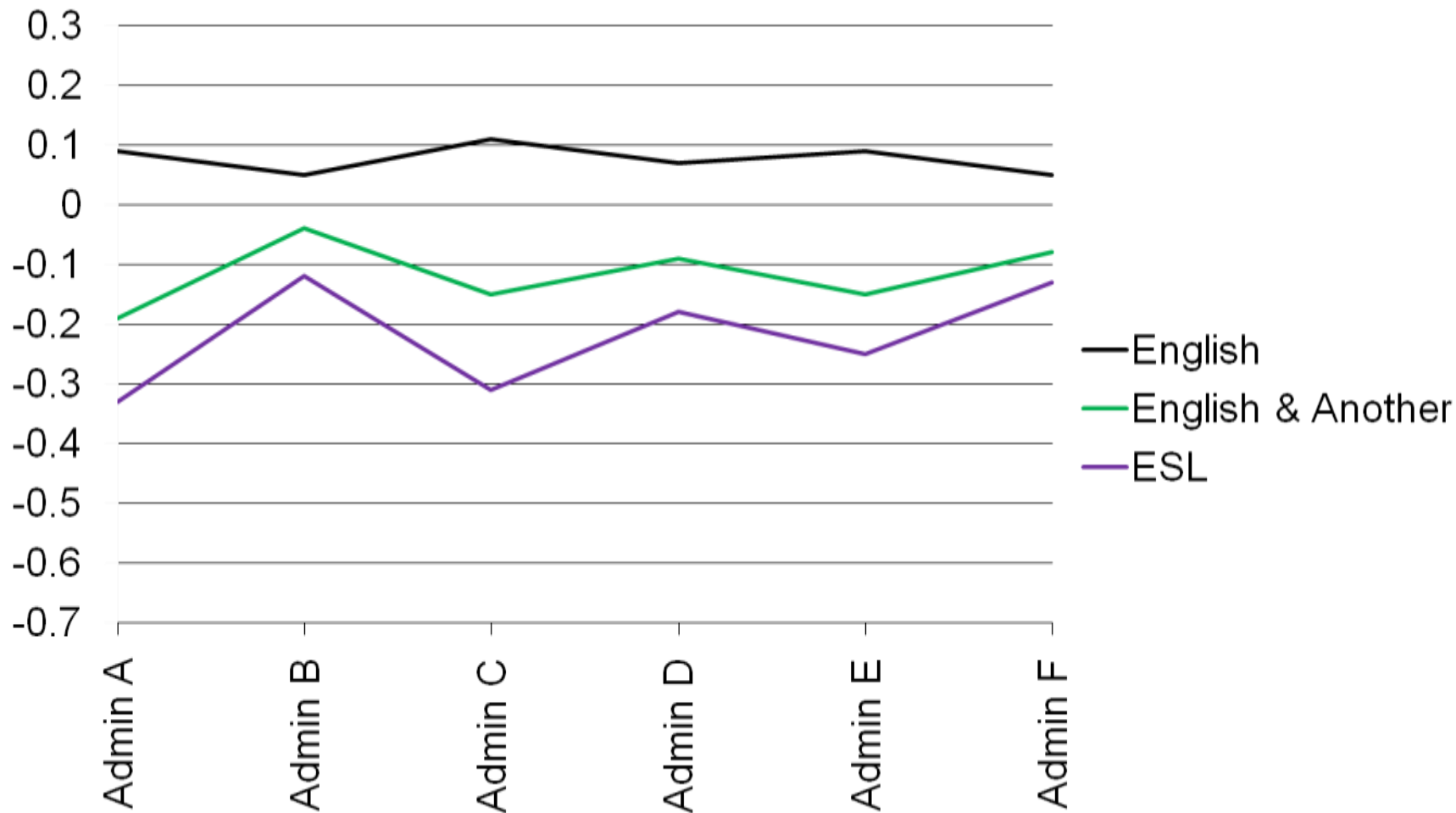
Std. Differences for Ethnic/Racial Background Writing Composite Score



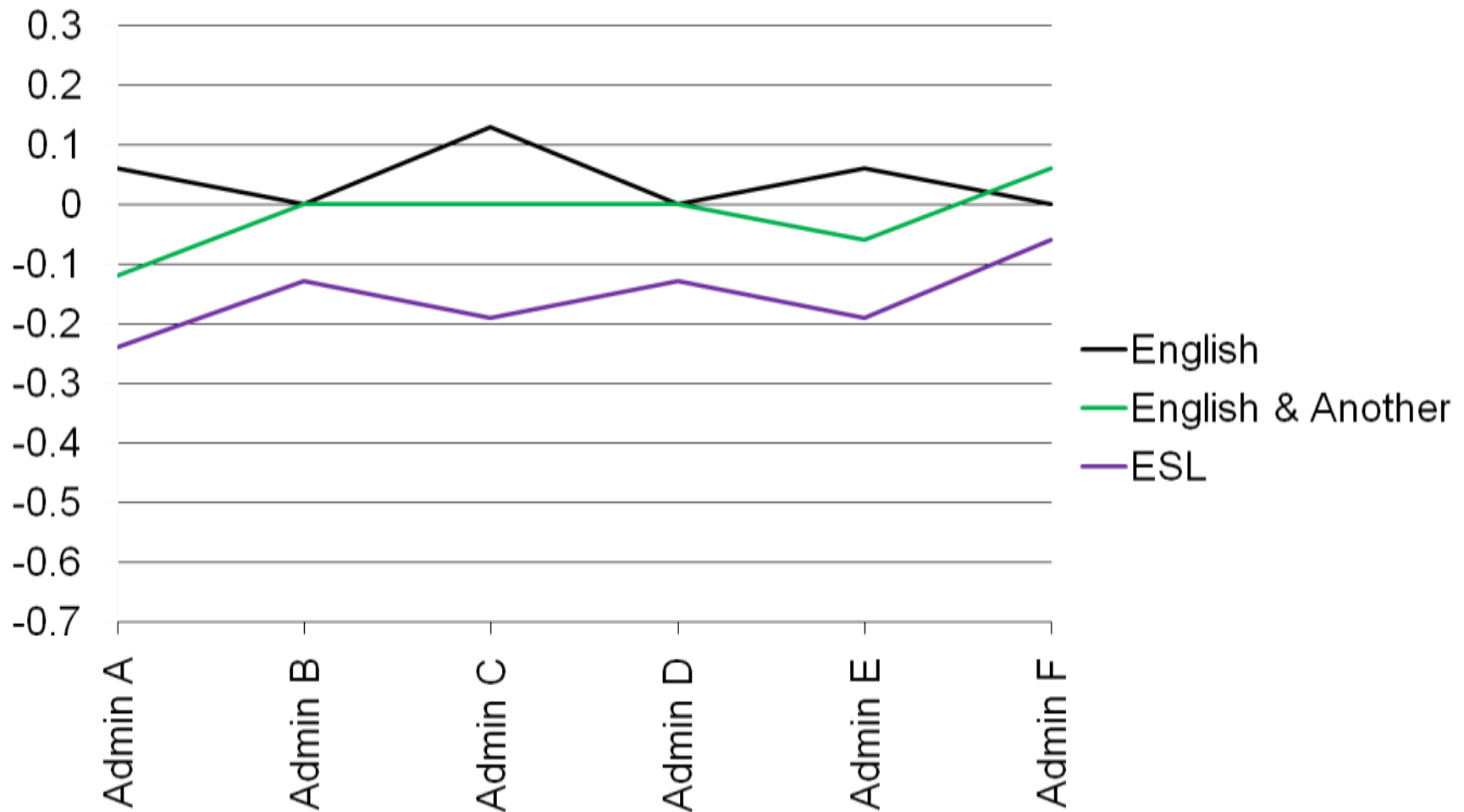
Std. Differences for Ethnic/Racial Background Essay Raw Score



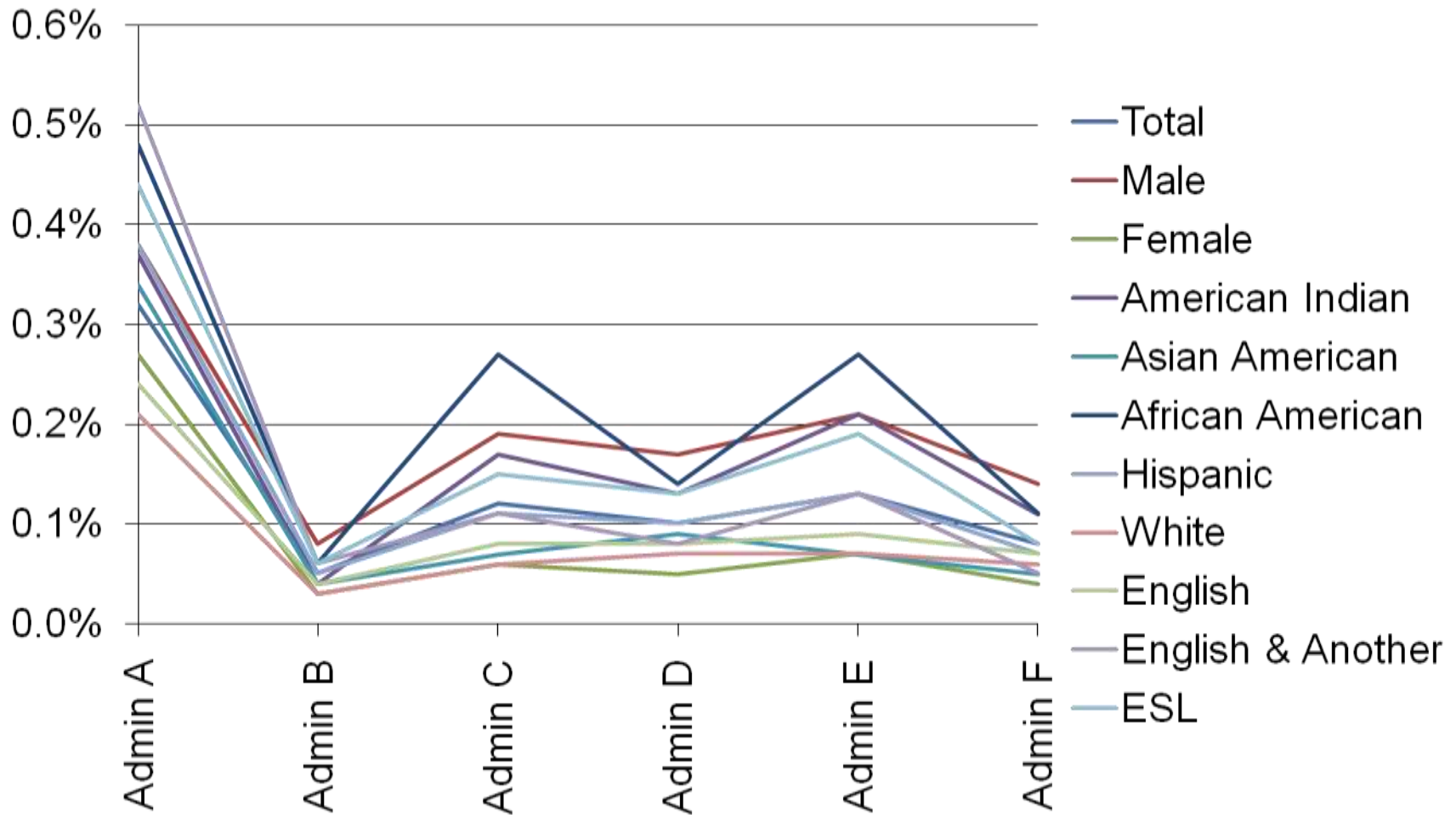
Std. Differences for Language Background Writing Composite Score



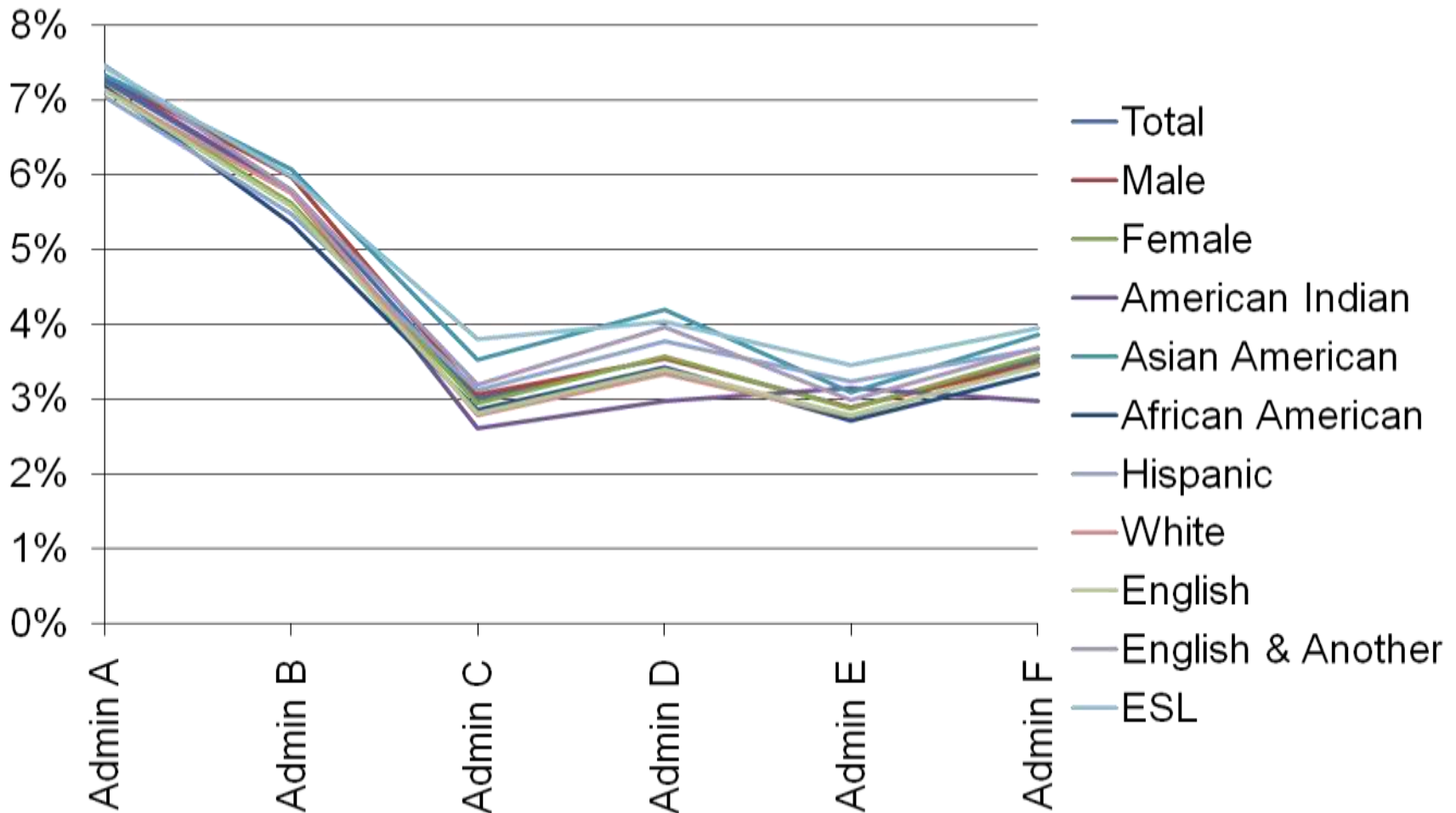
Std. Differences for Language Background Essay Raw Score



Percent of Essays Scored Zero



Percent of Essays Referred to 3rd Reader



Estimates of Reliability

- Challenges:
 - Mixed format exam
 - Multiple rater pairs without knowledge of pairs
 - Examinees answer different numbers of MC items

Estimates of Reliability

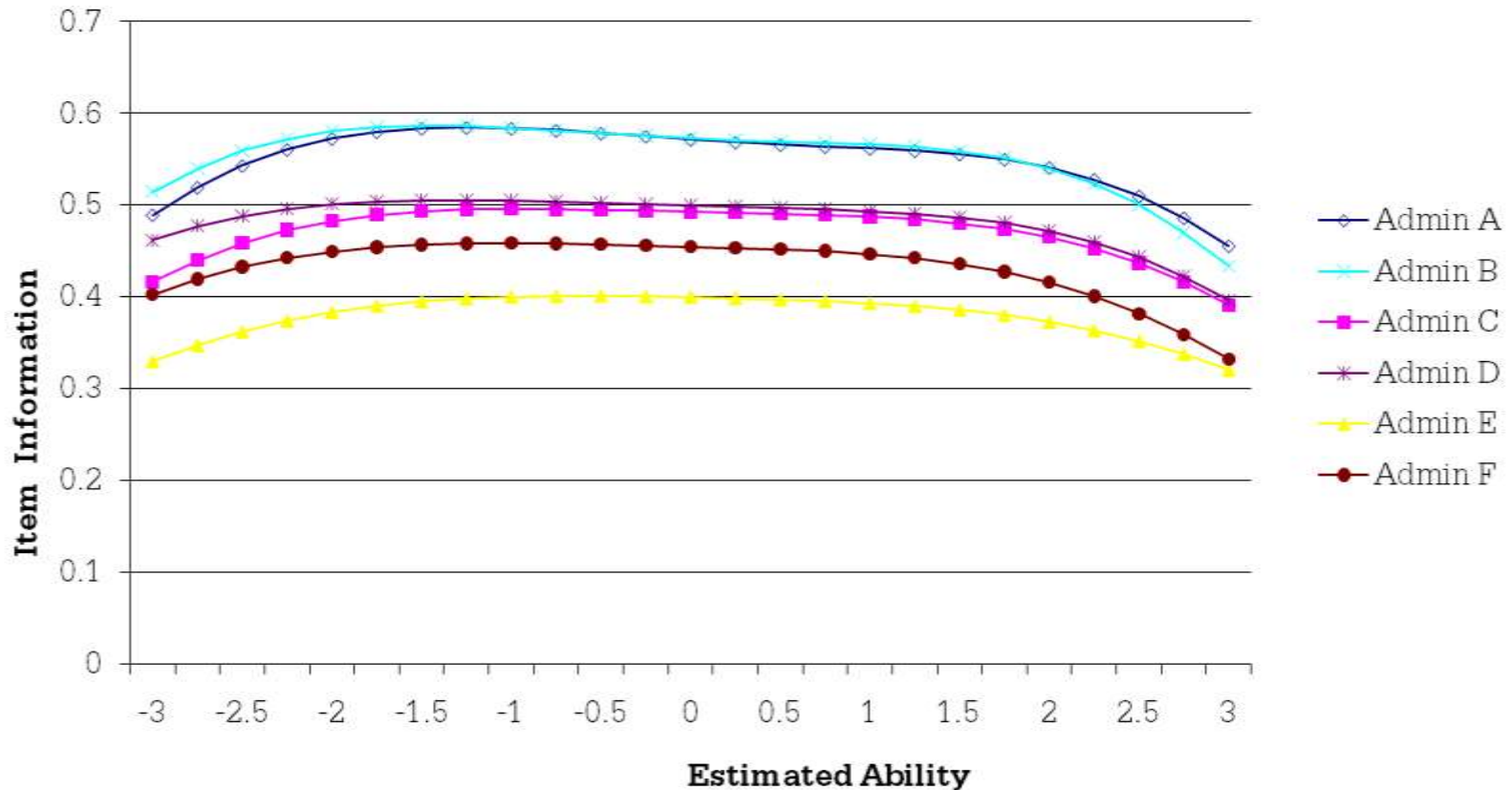
- Methods
 - Generalizability Theory
 - Univariate designs: unbalanced $p \times i$ for MC items, $r:p$ for essay
 - Multivariate design for composite by combining UV designs and obs. covariance
 - Mixed IRT model
 - 2PL for MC and GRM for Essay

$$\bar{\rho} = \frac{\text{var } \theta - \overline{SE}^2 \theta}{\text{var } \theta}$$

Estimates of Reliability

	$E\rho^2$			IRT Composite
	<u>Multiple-Choice</u>	<u>Essay</u>	<u>Composite</u>	
Admin A	0.96	0.77	0.89	0.91
Admin B	0.95	0.81	0.90	0.91
Admin C	0.98	0.81	0.92	0.89
Admin D	0.96	0.81	0.92	0.91
Admin E	0.98	0.76	0.92	0.91
Admin F	0.97	0.80	0.91	0.90

Essay Item Information Functions



Review of Validity Studies of SAT Writing

- Kobrin, Patterson, Shaw, Mattern & Barbuti (2008) & Mattern, Patterson, Shaw, Kobrin, & Barbuti (2008)
 - 110 institutions, over 150,000 students

	HSGPA	SAT - CR	SAT - M	SAT - W
Total	0.54	0.48	0.47	0.51
Male	0.52	0.44	0.45	0.47
Female	0.54	0.52	0.53	0.54
American Indian	0.49	0.41	0.41	0.42
Asian American	0.47	0.41	0.43	0.44
African American	0.44	0.40	0.40	0.43
Hispanic	0.46	0.43	0.41	0.46
White	0.56	0.48	0.46	0.51
English	0.55	0.49	0.47	0.52
English & Another Language	0.42	0.41	0.43	0.45
ESL	0.35	0.28	0.34	0.32

Review of Validity Studies of SAT Writing

- Cornwell, Mustard, & Van Parys (2008): U. of Georgia
 - Found SAT writing scores were a significant predictor of:
 - 1st year college gpa
 - Course grades in English but not mathematics
 - Number of credit hours enrolled in
 - Number of credit hours earned
 - Number of credit hours withdrawn from
 - Loss of HOPE scholarship
- Consequential Validity (Noeth & Kobrin) to be presented shortly

Conclusions

- Scores on the SAT Writing Test:
 - Display more than adequate and stable estimates of reliability for use in high-stakes decisions
 - Differences in scores within groups are stable from administration to administration
 - Percents of essay scores receiving zeros and being sent to a 3rd reader have stabilized and are fairly small across groups
 - Are slightly more predictive of 1st year college gpa than other SAT variables across groups
 - Are useful for predicting course grades in appropriate content areas
 - Useful for predicting other academic outcomes
- Institutions should conduct their own validity studies

Thank-you

- Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board presentations do not necessarily represent official College Board position or policy.
- Access this presentation online at <http://professionals.collegeboard.com/data-reports-research/cb/presentations>
- Please forward any questions, comments, and suggestions to:
Thomas Proctor at: tproctor@collegeboard.org