An Investigation of Differential Mode Effects When Comparing Paper and Online ACT Testing

Lu Wang, PhD, and Jeffrey Steedle, PhD

In recent ACT mode comparability studies, students testing on laptop or desktop computers earned slightly higher scores on average than students who tested on paper, especially on the ACT® reading and English tests (Li et al., 2017). Equating procedures adjust for such "mode effects" to make ACT scores comparable regardless of testing mode. However, it remains possible that the mode effects are different for different groups of students. For example, differences in performance between paper and online testing may be different for groups with different levels of comfort taking tests on computers. Thus, a general mode adjustment may be inappropriate. The purpose of this study was to explore the possibility of differential mode effects by gender, race/ethnicity, and ability using data from three recent mode comparability studies (Steedle, Pashley, & Cho, 2020). Results indicated that mode effects in English, reading, math, and science did not vary significantly between genders or race/ethnicity groups. Analyses detected significant interactions between mode effects and ability because mode effects tended to be smaller for lower ability examinees. Fortunately, equating processes appropriately adjust scores for differential mode effects by ability.

Prior Research

Several previous studies compared mode effects for specific examinee groups. MacCann (2006), for example, detected no significant interaction between gender and mode on a computing skills test. There was, however, a significant interaction between socioeconomic status (SES) and mode wherein low SES examinees scored higher on paper and high SES examinees scored higher online. Karkee, Kim, and Fatica (2010) calculated mode effects by gender and ethnicity on a social studies test, but they did not test for significant differences. The mode effects were similar for males and females (effect size d = 0.06 and -0.04 standard deviations, respectively, with positive effect sizes indicating higher online scores). The mode effect was relatively small for White examinees (-0.06) compared to Black (0.15), Hispanic (0.17), and Asian (0.23) examinees.

In another study, Kim and Kim (2013) examined reading comprehension tests administered to high school students. When comparing paper to a scanned version administered on computers, there were statistically significant mode effects favoring paper for females and males, though the magnitudes of the effects differed somewhat (d = 0.77 and 0.57, respectively). Jeong (2014) detected significant mode effects on a Korean language test for males and females, but only females exhibited significant mode effects on mathematics and science tests. The greatest difference occurred on the mathematics test, where the male mode effect was only -0.07 standard deviations (non-significant), but the female mode effect was -0.48 ($p < 0.05$). Jerrim and his colleagues (2018) analyzed 2015 PISA results in mathematics, reading, and science from three countries. In two cases, there was a significant interaction between gender and mode: females performed better on the paper reading test in Ireland and males did not, and males performed better on the paper science test in Sweden and females did not.

Prior studies provide evidence of differential mode effects by gender, SES, and race/ethnicity, but the evidence was not consistent across testing contexts (e.g., content areas, tested populations, cultures, test designs, and testing environments). The current study adds to the research literature by examining differential mode effects by gender, race/ethnicity, and ability using large samples of students who took the ACT English, math, reading, and science tests for the purpose of college admissions.

Method

Sample

The data for this analysis were collected during three mode comparability studies that coincided with Saturday national ACT administrations in October 2019, December 2019, and February 2020. The total sample sizes for the studies were 3,583, 6,352, and 6,645, respectively. Examinees in each study were randomly assigned to the paper and online testing conditions. Table 1 describes the demographics of the participating students in all three studies combined (the distributions were highly similar across studies). Females outnumbered males in the study sample, and White participants outnumbered the other race/ethnicity groups. Compared to a recent national sample of ACT examinees, the study sample had a similar percentages of female examinees (58% vs. 56%) and White examinees (58% vs. 54%). The small percentage differences between the paper and online groups were consistent with effective randomization to the paper and online conditions.

Table 1. Sample Demographics

Group	Paper	Online	Difference
Gender			
Male	41.1%	41.2%	-0.1%
Female	58.2%	58.1%	0.1%
Others	0.7%	0.7%	0.0%
Ethnicity			
Black/African American	14.0%	14.0%	0.0%
White	57.9%	57.8%	0.1%
Hispanic/Latino	15.8%	16.0%	-0.2%
Asian	4.1%	3.8%	0.3%
Others	8.2%	8.4%	-0.2%

Analysis

Analyses were designed to explore differential mode effects—that is, differences between groups in terms of the mode effect, which is the difference in performance between those who tested on paper and online. Separate analyses were conducted for gender, race/ethnicity, and ability. The race/ethnicity analysis included only the four largest groups: Black/African American, White, Hispanic/Latino, and Asian.

No independent measure of ability was available, so ability in each subject area was predicted from performance in the other subject areas. First, as Equation (1) shows, the target subject scale scores were regressed on the other three subjects' scale scores using only data from the paper samples (to avoid of the influence of mode effects). For example, if the target subject was English, then the English scale scores for examinees who tested via paper were regressed on their math, reading, and science scale scores.

$$Y_{target} = \beta_0 + \beta_1 X_{subj1} + \beta_2 X_{subj2} + \beta_3 X_{subj3} + \varepsilon \quad (1)$$

where Y_{target} is the paper scale score for the target subject and X_{subj1}, X_{subj2}, and X_{subj3} are the corresponding paper scale scores for the other three subjects. The proportion of variance of the dependent variable explained by the model (i.e., the R^2 statistics) for English, math, reading, and science were 0.72, 0.65, 0.70, and 0.72 respectively.

The estimated intercept and slope parameters (i.e., β_0, β_1, β_2, and β_3) in Equation (1) were then applied to all participants (i.e., students who tested on paper or online) to calculate predicted scale scores for the subject. To ensure the predicted scale scores were comparable regardless of testing mode, adjusted scale scores served as predictors for the online group (i.e., scale scores adjusted for mode effects via equipercentile equating).

Three-way analysis of variance (ANOVA) was conducted, using test administration (October, December, or February) as one factor, testing mode (paper or online) as the second factor, and student group or ability as the third factor. The initial ANOVA model was designed to detect potential interactions among test administration, testing mode, and student group, though the main outcome of interest was the mode by group interaction (i.e., differential mode effect). If certain interactions were non-significant, a simpler model (without the non-significant interactions) was fit to the data.

Results

Score Equivalency Analysis

Descriptive statistics for each ACT subject test and each administration were calculated for the total, gender, and race/ethnicity groups (see Appendix A). In each case, 1–36 scale scores were obtained using the raw-to-scale score conversion table for paper testing (with no mode adjustment). Therefore, the mean differences between paper and online reflect mode effects. In general, the average online scores were slightly higher than the corresponding paper scores, suggesting that online was “easier” than the corresponding paper versions of the tests.

Figures 1–3 plot the mean scale scores for different administrations, modes, and student groups, with ability divided into quantiles. Group is represented by color, and line type indicates mode (dashed = paper, solid = online). The vertical distance between the dashed and solid lines represents the average mode effect. As shown in Figure 1, English and reading scores showed fairly consistent mode effects for females and males across administrations. The mode effects for math and science were generally smaller in magnitude. For those subjects, the mode effects for females were slightly larger than the corresponding mode effects for males in October. In the December administration, however, the opposite result was observed.

For the race/ethnicity groups, different subjects exhibited different patterns in mode effects within each administration. For example, Black/African American students' English scores showed larger mode effects in October than other groups. This was also true for math in October, but Asian and White students' math scores showed slightly larger mode effects than the other two groups in December. For reading, the mode effects for different groups appeared to be consistent within and across administrations. Science showed patterns similar to math in October and December. However, in February, science scores for all groups except Black/African American showed mode effects favoring online.

For the ability quantiles, the mode effect was very small for AB1 (ability group 1, or 1st to 20th percentile), especially in December and February. English scores for AB4 showed larger mode effects than the other ability groups in December but not in other administrations. Math scores for AB5 showed slightly larger mode effects than the other ability groups in December, but mode effects in math were very small in general. For the two lowest ability groups (AB1 and AB2), reading scores displayed slightly smaller mode effects than the other ability groups in October and December. In February, however, scores for AB2 and AB4 showed smaller mode effects than the other ability groups. Science scores showed consistent patterns in mode effects across different ability groups across administrations. Specifically, science scores for the three lowest ability groups (AB1, AB2, and AB3) indicated almost no mode effect, but scores for higher ability groups revealed larger mode effects.

Figure 1. Mean Scale Scores by Gender

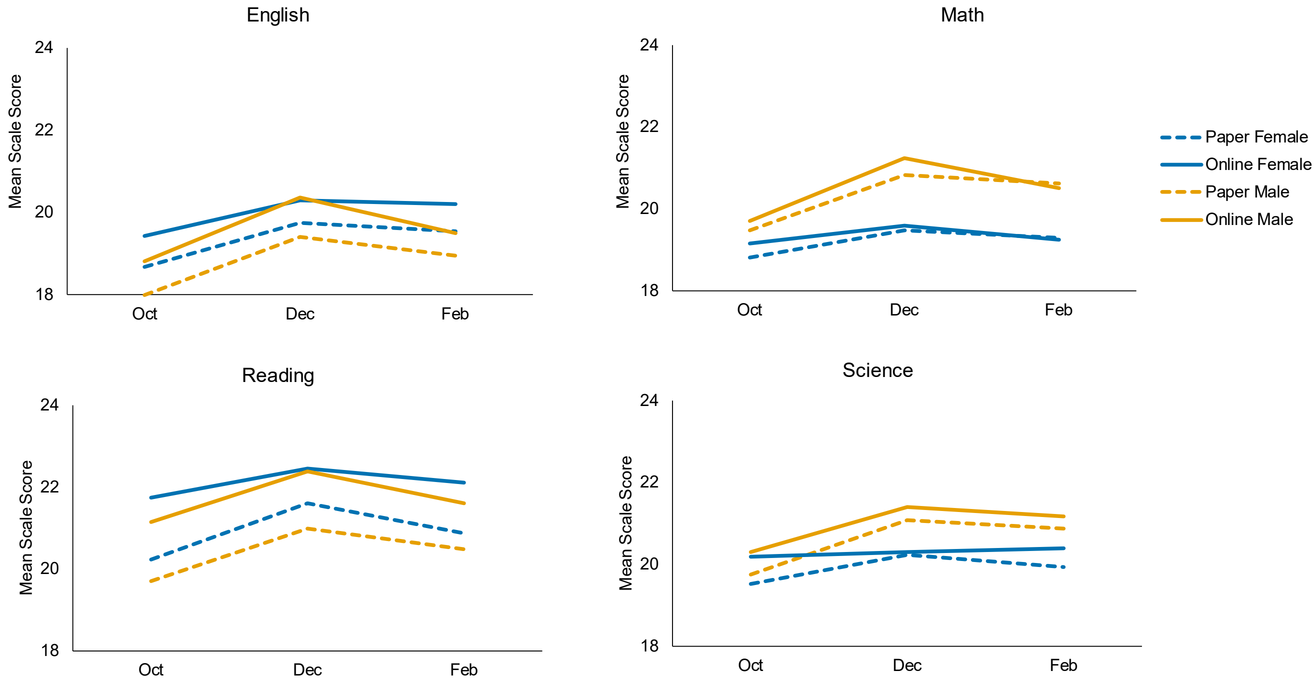


Figure 2. Mean Scale Scores by Race/Ethnicity

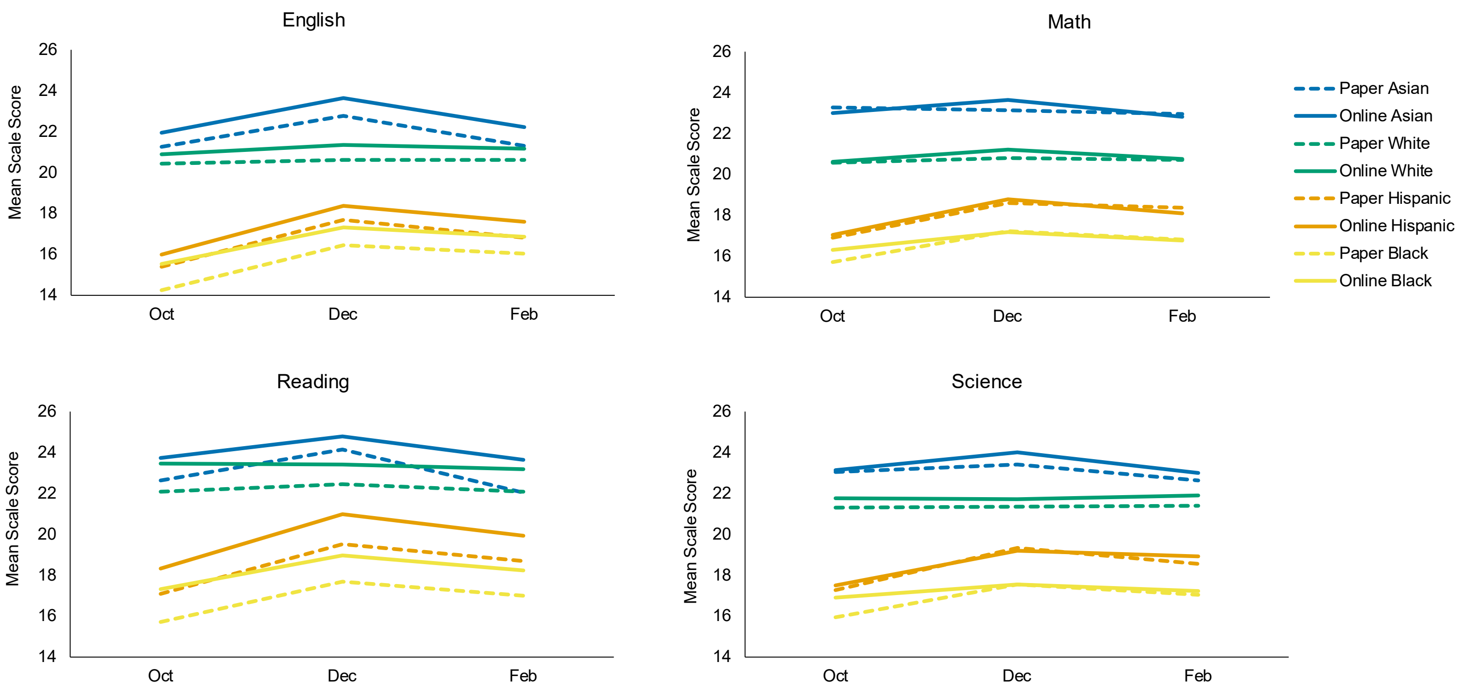
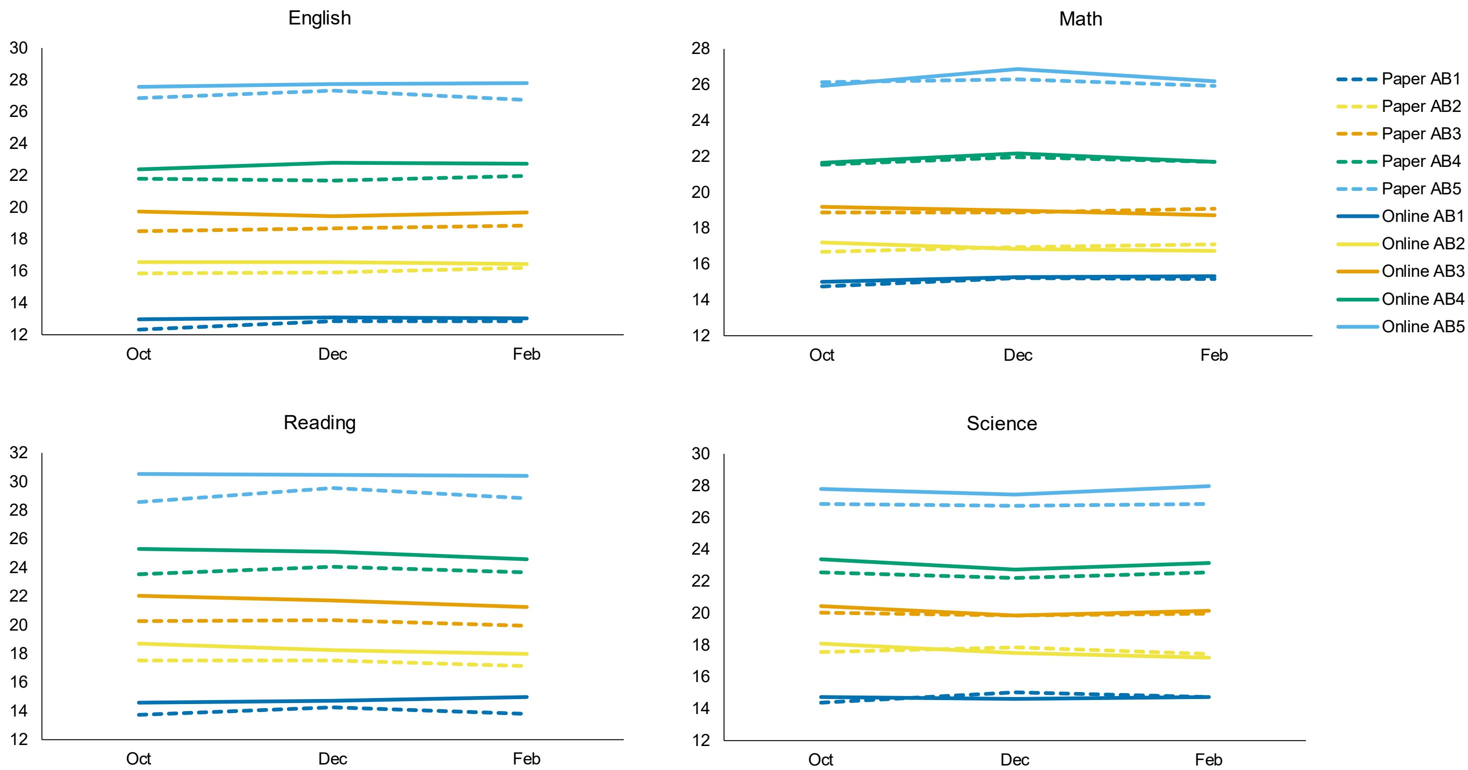


Figure 3. Mean Scale Scores by Ability Quantiles



ANOVA Analysis

The descriptive trends shown in Figures 1–3 may not indicate systematic, statistically significant differences. For that reason, ANOVA was applied to identify statistically significant main effects and interactions. Appendix B presents all three-way ANOVA tables for the four subjects. For the gender and ethnicity groups, none of the three-way interactions were statistically significant at the 0.05 level nor were any two-way mode by administration interactions. Therefore, simpler models with two two-way interaction terms (mode by group interaction and administration by group interaction) were fit. Results from those simpler models are described here (see Appendix B, Tables B.1 and B.2). For the ability ANOVA, predicted scale scores were entered in the model as a continuous independent variable for each subject. Except for the reading analysis, all three-way interactions were statistically significant at the 0.05 level, so no simpler models were fit (see Table B.3).

Table B.1 shows the ANOVA results for the gender groups for the four subjects. Overall, mode, administration, and gender (and their interactions) were poor predictors of ACT scores. Indeed, the R^2 values ranged from only 0.01 to 0.02 across the four subjects. Despite this, all main effects (mode, administration, and gender) were statistically significant except for the math mode effect. Specifically, online scores were significantly higher in English, reading, and science; scores for females were significantly higher in English, lower in math, higher in reading, and lower in science; and average scores varied by administration—an expected result, since the study samples varied in ability. However, none of the mode by gender interaction terms were statistically significant, suggesting that the observed mode effects between paper and online did not differ between female and male examinees. The interaction between administration and gender was also statistically significant for English, math, and science, but this finding simply reflects sample differences across studies.

Compared to gender, race/ethnicity was a better predictor of ACT scores (R^2 ranged from 0.12 to 0.13 for the ANOVA models). As in the gender ANOVA, the only non-significant main effect was the mode effect for math (Table B.2). All other main effects for mode, administration, and race/ethnicity were statistically significant for all the subjects. The mode by race/ethnicity interaction terms were non-significant for all four subjects, indicating that testing mode had a similar effect on scores across race/ethnicity groups. The administration by race/ethnicity interaction was also statistically significant, but this simply reflected sample differences across administrations.

The ANOVA results for ability (Table B.3) indicated that the main effects for mode, administration, and ability were statistically significant for all four subjects. Even the math mode effect was statistically significant, and this may be related to greater estimation precision provided by the statistical control for ability. The R^2 values for the ANOVA models ranged from 0.66 to 0.73 because ability was a strong predictor of ACT scores. Unlike the gender and race/ethnicity analyses, there was a significant three-way interaction for English, math, and science, so no simpler models were fit to the data. The ω^2 statistic shown in Equation (2) was used to measure effect size for the interaction terms.

$$\omega^2 = \frac{SS_{effect} - df_{effect} \times MS_{residual}}{SS_{total} + MS_{residual}} \qquad (2)$$

All of the statistically significant three-way interactions had negligible effect sizes (< 0.0002), indicating that none of the three-way interactions had practical significance. Regarding differential mode effects, the mode by ability interaction was statistically significant for English, reading, and science but not for math (p = 0.06). However, none of those two-way interactions were practically significant (the effect sizes for English, reading, science were 0.0001, 0.0001, and 0.0015, respectively).

The potentially complex interactions between mode, administration, and ability are best illustrated graphically. The English mode effect increased slightly as ability level increased, and this trend was most noticeable for the February administration (Figure 4). The math mode effect was small in October and February, and it decreased slightly as ability increased (Figure 5). In contrast, the mode effect was larger in December, and it tended to increase with ability. Mode effects were greatest on the reading test, especially in October and February (Figure 6). In each administration, the mode effect increased with ability. For Science, the mode effect was larger for examinees with high ability, in particular for the December administration, which also had the largest science mode effect overall (Figure 7). The lines plotted for the February administration cross, which indicates that paper scores were higher for low-ability students and online scores were higher for high-ability students.

Figure 4. Interaction Between Mode and Ability by Administration for English

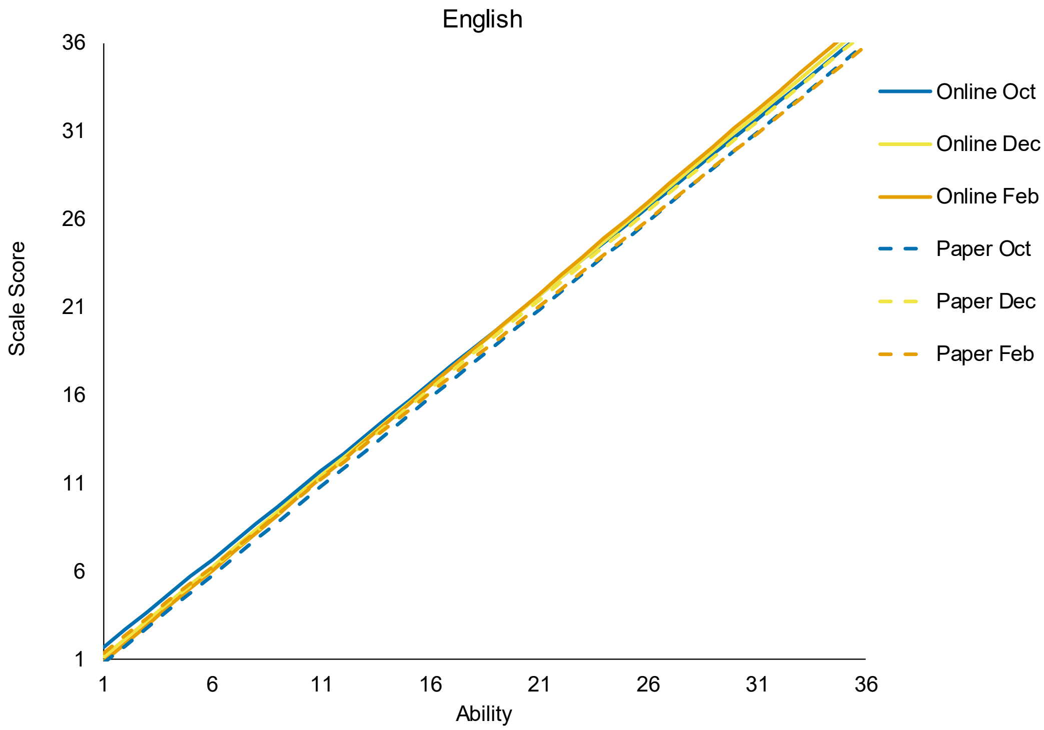


Figure 5. Interaction Between Mode and Ability by Administration for Math

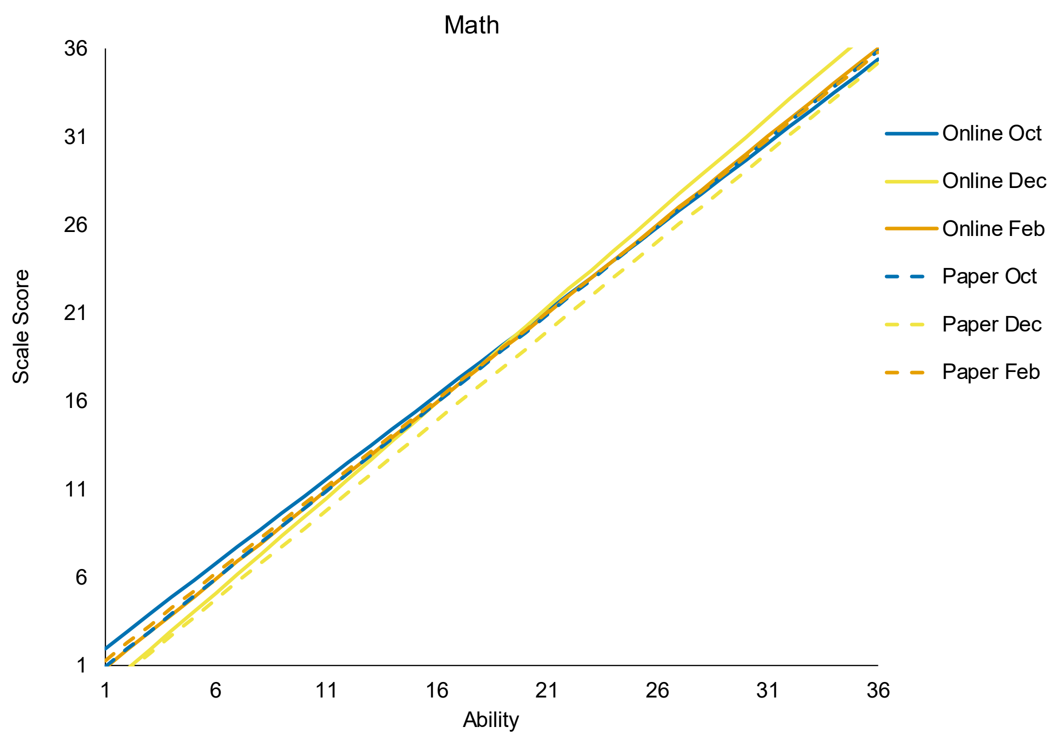


Figure 6. Interaction Between Mode and Ability by Administration for Reading

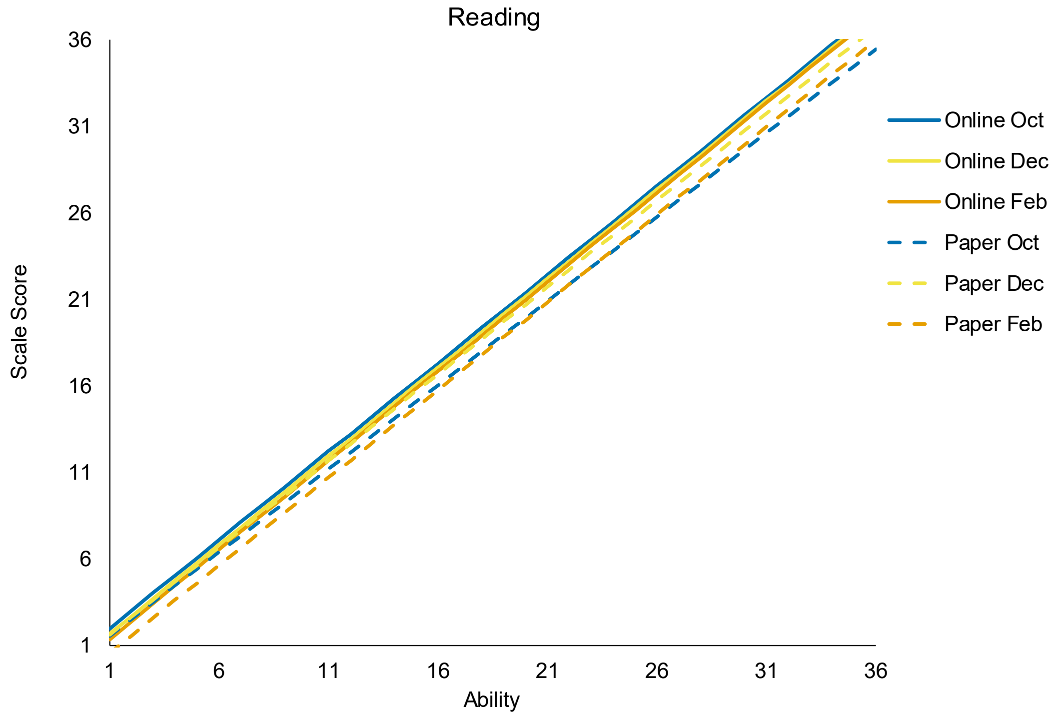
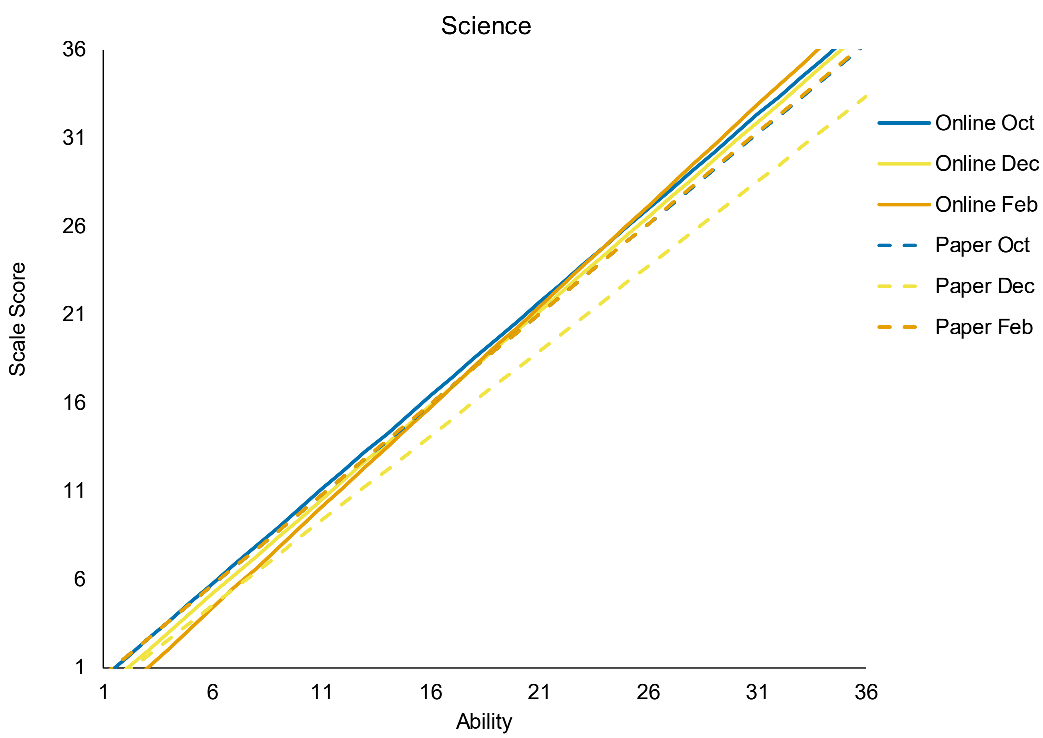


Figure 7. Interaction Between Mode and Ability by Administration for Science



Summary and Conclusions

Consistent with previous ACT mode comparability studies (Li et al., 2017), the studies conducted in October 2019, December 2019, and February 2020 indicated that online scores were systematically higher than paper scores, especially on the English and reading tests (Steedle, Pashley, & Cho, 2020). Descriptive analyses identified small differences in mode effects for different examinee groups. For example, females were more affected by mode in October, and males were more affected in December. The magnitudes of mode effects on race/ethnicity groups also differed slightly by subject and administration. Mode effects appeared to increase for higher ability examinees, but the magnitudes of those increases were small.

Subsequent analyses determined whether the observed differences in mode effects were statistically significant. The three-way ANOVA results detected no statistically significant differential mode effects for gender and race/ethnicity groups. There were significant interactions between examinee group (gender or race/ethnicity) and administration, but this result was due to sample differences between administrations. Three-way interactions between mode, administration, and ability were statistically significant for English, math, and science. In most cases, the interaction plots indicated that higher ability examinees were more affected by testing mode compared to lower ability examinees. This finding is possibly explained by higher ability students taking greater advantage of whatever benefits are offered by online testing compared to paper testing (e.g., an on-screen timer to help with pacing through the test).

In sum, the current study detected no statistically significant evidence of differential mode effects between gender or race/ethnicity groups. However, there was evidence that mode effects tended to be greater for higher ability examinees. Online testing offers advantages such as faster scoring and greater convenience, but there is always a risk of introducing score comparability issues for examinees testing in different modes. For this reason, tests administered on paper and online may be equated to make scores comparable regardless of testing mode. Fortunately, equating across modes for the ACT need not account for gender or race/ethnicity, and current equating processes adjust for differential mode effects for examinees of differing ability. The end result is scores with the same meaning for examinees testing in different modes and from different demographic groups.

References

Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour & Information Technology, 33*(4), 410–422. https://doi.org/10.1080/0144929X.2012.710647

Jerrim, J., Micklewright, J., Heine, J.-H., Salzer, C., & McKeown, C. (2018). PISA 2015: How big is the 'mode effect' and what has been done about it? *Oxford Review of Education, 44*(4), 476–493. https://doi.org/10.1080/03054985.2018.1430025

Karkee, T., Kim, D.-I., & Fatica, K. (2010, April-May). *Comparability study of online and paper and pencil tests using modified internally and externally matched criteria*. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO. https://www.measurementinc.com/sites/default/files/2017-05/Online%20and%20Paper%20and%20Pencil%20Comparability%20Study%20with%20Alternate%20Design.pdf

Kim, H. J., & Kim, J. (2013). Reading from an LCD monitor versus paper: Teenagers' reading performance. *International Journal of Research Studies in Educational Technology, 2*(1). https://doi.org/10.5861/ijrset.2012.170

Li, D., Yi, Q., & Harris, D. (2017). *Evidence for paper and online ACT® comparability: Spring 2014 and 2015 mode comparability studies*. Iowa City, IA: ACT. https://www.act.org/content/dam/act/unsecured/documents/Working-Paper-2016-02-Evidence-for-Paper-and-Online-ACT-Comparability.pdf

MacCann, R. (2006). The equivalence of online and traditional testing for different subpopulations and item types. *British Journal of Educational Technology, 37*(1), 79–91. https://doi.org/10.1111/j.1467-8535.2005.00524.x

Steedle, J., Pashley, P., & Cho, Y. (2020). *Three studies of comparability between paper-based and computer-based testing for the ACT.* Iowa City, IA: ACT.

Appendix A

Table A.1. ACT English Descriptive Statistics by Demographic Group

Admin	Subgroup	Online			Paper			Mean Diff.
		N	Mean	SD	N	Mean	SD	
Oct	Total	1776	19.17	6.02	1807	18.37	6.08	0.80
	Male	737	18.81	6.28	780	17.98	6.06	0.83
	Female	1030	19.41	5.81	1011	18.66	6.10	0.75
	Black/African American	330	15.50	4.96	362	14.24	4.03	1.27
	White	990	20.89	5.64	982	20.45	5.74	0.45
	Hispanic/Latino	236	15.97	5.03	248	15.38	5.59	0.59
	Asian	85	21.95	6.57	86	21.23	6.20	0.72
Dec	Total	3205	20.34	6.08	3147	19.63	6.02	0.71
	Male	1337	20.36	6.05	1286	19.39	6.02	0.97
	Female	1841	20.28	6.07	1839	19.75	5.98	0.54
	Black/African American	418	17.32	5.74	406	16.45	5.23	0.87
	White	1850	21.36	5.91	1839	20.61	5.85	0.75
	Hispanic/Latino	528	18.35	5.43	491	17.66	5.27	0.69
	Asian	142	23.64	6.35	153	22.76	6.53	0.88
Feb	Total	3348	19.94	6.12	3297	19.31	5.90	0.63
	Male	1358	19.50	6.13	1327	18.94	5.84	0.56
	Female	1965	20.20	6.10	1953	19.54	5.94	0.67
	Black/African American	415	16.86	5.37	384	16.01	4.82	0.85
	White	1974	21.15	5.91	1958	20.61	5.66	0.54
	Hispanic/Latino	572	17.59	5.64	564	16.79	5.62	0.79
	Asian	91	22.21	6.68	100	21.30	6.30	0.91

Note. Mean Diff. was calculated as online mean minus paper mean.

Table A.2. ACT Math Descriptive Statistics by Demographic Group

		Online			Paper			
Admin	Subgroup	N	Mean	SD	N	Mean	SD	Mean Diff.
Oct	Total	1776	19.37	4.82	1807	19.08	4.99	0.29
	Male	737	19.70	5.15	780	19.47	5.44	0.23
	Female	1030	19.14	4.56	1011	18.80	4.60	0.34
	Black/African American	330	16.30	3.19	362	15.72	2.68	0.58
	White	990	20.63	4.59	982	20.57	4.87	0.06
	Hispanic/Latino	236	17.02	3.86	248	16.92	4.38	0.10
	Asian	85	22.99	6.20	86	23.26	5.64	-0.27
Dec	Total	3205	20.30	5.20	3147	20.05	4.93	0.25
	Male	1337	21.23	5.53	1286	20.82	5.26	0.41
	Female	1841	19.58	4.79	1839	19.47	4.58	0.12
	Black/African American	418	17.17	3.96	406	17.21	3.97	-0.04
	White	1850	21.22	5.19	1839	20.80	4.83	0.43
	Hispanic/Latino	528	18.77	4.41	491	18.58	4.16	0.19
	Asian	142	23.64	5.78	153	23.13	5.86	0.51
Feb	Total	3348	19.76	4.94	3297	19.83	4.88	-0.07
	Male	1358	20.50	5.38	1327	20.62	5.30	-0.11
	Female	1965	19.24	4.54	1953	19.29	4.50	-0.06
	Black/African American	415	16.74	3.05	384	16.82	3.14	-0.07
	White	1974	20.74	5.08	1958	20.72	4.87	0.01
	Hispanic/Latino	572	18.09	3.88	564	18.37	4.38	-0.28
	Asian	91	22.81	5.92	100	22.97	5.34	-0.16

Note. Mean Diff. was calculated as online mean minus paper mean.

Table A.3. ACT Reading Descriptive Statistics by Demographic Group

Admin	Subgroup	Online			Paper			Mean Diff.
		N	Mean	SD	N	Mean	SD	
Oct	Total	1776	21.50	6.81	1807	20.00	6.53	1.50
	Male	737	21.14	7.04	780	19.69	6.62	1.45
	Female	1030	21.74	6.64	1011	20.22	6.45	1.52
	Black/African American	330	17.29	5.54	362	15.72	4.67	1.57
	White	990	23.47	6.44	982	22.08	6.14	1.39
	Hispanic/Latino	236	18.29	5.95	248	17.06	6.04	1.24
	Asian	85	23.73	7.01	86	22.63	7.00	1.10
Dec	Total	3205	22.45	6.52	3147	21.39	6.42	1.06
	Male	1337	22.39	6.65	1286	20.99	6.61	1.39
	Female	1841	22.45	6.42	1839	21.61	6.24	0.84
	Black/African American	418	18.95	5.77	406	17.67	5.90	1.27
	White	1850	23.39	6.39	1839	22.44	6.18	0.95
	Hispanic/Latino	528	20.95	6.20	491	19.53	5.84	1.43
	Asian	142	24.79	6.42	153	24.14	6.71	0.64
Feb	Total	3348	21.92	6.49	3297	20.73	6.39	1.19
	Male	1358	21.60	6.65	1327	20.48	6.46	1.12
	Female	1965	22.11	6.37	1953	20.87	6.34	1.24
	Black/African American	415	18.23	5.33	384	16.96	5.12	1.27
	White	1974	23.18	6.37	1958	22.06	6.23	1.13
	Hispanic/Latino	572	19.92	5.96	564	18.69	6.19	1.23
	Asian	91	23.63	6.62	100	22.04	6.82	1.59

Note. Mean Diff. was calculated as online mean minus paper mean.

Table A.4. ACT Science Descriptive Statistics by Demographic Group

Admin	Subgroup	Online			Paper			Mean Diff.
		N	Mean	SD	N	Mean	SD	
Oct	Total	1776	20.23	5.35	1807	19.61	5.24	0.62
	Male	737	20.30	5.86	780	19.75	5.72	0.55
	Female	1030	20.18	4.96	1011	19.52	4.85	0.66
	Black/African American	330	16.88	4.36	362	15.93	3.93	0.95
	White	990	21.75	4.96	982	21.32	4.76	0.44
	Hispanic/Latino	236	17.47	4.71	248	17.25	5.00	0.22
	Asian	85	23.15	5.60	86	23.03	5.42	0.12
Dec	Total	3205	20.78	5.45	3147	20.59	5.00	0.19
	Male	1337	21.40	5.80	1286	21.06	5.37	0.33
	Female	1841	20.29	5.10	1839	20.22	4.66	0.07
	Black/African American	418	17.55	4.69	406	17.52	4.58	0.04
	White	1850	21.72	5.38	1839	21.34	4.84	0.37
	Hispanic/Latino	528	19.18	4.81	491	19.30	4.43	-0.13
	Asian	142	23.99	5.31	153	23.39	5.21	0.59
Feb	Total	3348	20.71	5.51	3297	20.31	5.06	0.39
	Male	1358	21.16	5.88	1327	20.87	5.46	0.28
	Female	1965	20.38	5.22	1953	19.94	4.76	0.45
	Black/African American	415	17.21	4.27	384	17.05	4.09	0.16
	White	1974	21.87	5.38	1958	21.39	4.80	0.48
	Hispanic/Latino	572	18.92	4.80	564	18.55	4.89	0.37
	Asian	91	22.98	5.63	100	22.62	5.38	0.36

Note. Mean Diff. was calculated as online mean minus paper mean.

Appendix B

Table B.1. Three-Way ANOVA for Gender Group

Subject	Factor	df	Sum of Squares	Mean Squares	*F*	*p*
English	Mode	1	2018.477	2018.477	55.629***	0.00000
	Admin	2	3239.927	1619.964	44.646***	0.00000
	Gender	1	838.900	838.900	23.120***	0.00000
	Mode × Gender	1	21.162	21.162	0.583	0.44506
	Admin × Gender	2	251.011	125.505	3.459*	0.03149
	Residual	16456	597099.353	36.285	--	--
Math	Mode	1	72.141	72.141	2.973	0.08467
	Admin	2	1959.954	979.977	40.390***	0.00000
	Gender	1	5943.678	5943.678	244.973***	0.00000
	Mode × Gender	1	6.184	6.184	0.255	0.61367
	Admin × Gender	2	452.033	226.016	9.315***	0.00009
	Residual	16456	399265.499	24.263	--	--
Reading	Mode	1	6063.635	6063.635	143.726***	0.00000
	Admin	2	3102.049	1551.025	36.764***	0.00000
	Gender	1	767.570	767.570	18.194***	0.00002
	Mode × Gender	1	22.502	22.502	0.533	0.46521
	Admin × Gender	2	29.750	14.875	0.353	0.70288
	Residual	16456	694260.913	42.189	--	--
Science	Mode	1	537.327	537.327	19.472***	0.00001
	Admin	2	1315.742	657.871	23.841***	0.00000
	Gender	1	2234.552	2234.552	80.979***	0.00000
	Mode × Gender	1	0.378	0.378	0.014	0.90683
	Admin × Gender	2	392.860	196.430	7.119***	0.00081
	Residual	16456	454090.809	27.594	--	--

* $p < .05$, ** $p < .01$, *** $p < .001$

Table B.2. Three-Way ANOVA for Race/Ethnicity Group

Subject	Factor	df	Sum of Squares	Mean Squares	*F*	*p*
English	Mode	1	1795.046	1795.046	55.812***	0.00000
	Admin	2	2160.526	1080.263	33.588***	0.00000
	Race/Ethnicity/Ethnicity	3	63444.135	21148.045	657.544***	0.00000
	Mode × Race/Ethnicity	3	81.439	27.146	0.844	0.46953
	Admin × Race/Ethnicity	6	1791.285	298.547	9.283***	0.00000
	Residual	15188	488479.144	32.162	--	--
Math	Mode	1	60.751	60.751	2.805	0.09401
	Admin	2	1190.219	595.110	27.474***	0.00000
	Race/Ethnicity	3	46649.508	15549.836	717.879***	0.00000
	Mode × Race/Ethnicity	3	19.479	6.493	0.300	0.82560
	Admin × Race/Ethnicity	6	606.814	101.136	4.669***	0.00009
	Residual	15188	328984.170	21.661	--	--
Reading	Mode	1	5282.384	5282.384	139.984***	0.00000
	Admin	2	1810.903	905.451	23.995***	0.00000
	Race/Ethnicity	3	66911.799	22303.933	591.058***	0.00000
	Mode × Race/Ethnicity	3	53.568	17.856	0.473	0.70096
	Admin × Race/Ethnicity	6	2147.700	357.950	9.486***	0.00000
	Residual	15188	573128.491	37.736	--	--
Science	Mode	1	500.914	500.914	20.694***	0.00001
	Admin	2	590.311	295.156	12.194***	0.00001
	Race/Ethnicity	3	52348.217	17449.406	720.883***	0.00000
	Mode × Race/Ethnicity	3	35.600	11.867	0.490	0.68905
	Admin × Race/Ethnicity	6	1180.905	196.818	8.131***	0.00000
	Residual	15188	367634.825	24.206	--	--

* $p < .05$, ** $p < .01$, *** $p < .001$

Table B.3. Three-Way ANOVA for Ability

Subject	Factor	df	Sum of Squares	Mean Squares	*F*	*p*
English	Mode	1	2038.020	2038.020	206.655***	0.00000
	Admin	2	3434.850	1717.425	174.147***	0.00000
	Ability	1	440122.159	440122.159	44628.434***	0.00000
	Mode × Admin	2	18.134	9.067	0.919	0.39878
	Mode × Ability	1	82.479	82.479	8.363**	0.00383
	Admin × Ability	2	18.697	9.349	0.948	0.38756
	Mode × Admin × Ability	2	82.400	41.200	4.178*	0.01535
	Residual	16568	163392.331	9.862	--	--
Math	Mode	1	78.151	78.151	9.168**	0.00247
	Admin	2	2057.506	1028.753	120.682***	0.00000
	Ability	1	267457.908	267457.908	31375.184***	0.00000
	Mode × Admin	2	110.120	55.060	6.459**	0.00157
	Mode × Ability	1	29.121	29.121	3.416	0.06458
	Admin × Ability	2	223.652	111.826	13.118***	0.00000
	Mode × Admin × Ability	2	106.887	53.444	6.269**	0.00190
	Residual	16568	141233.995	8.525	--	--
Reading	Mode	1	6082.719	6082.719	481.019***	0.00000
	Admin	2	3268.664	1634.332	129.242***	0.00000
	Ability	1	491346.232	491346.232	38855.502***	0.00000
	Mode × Admin	2	113.527	56.763	4.489*	0.01125
	Mode × Ability	1	96.008	96.008	7.592**	0.00587
	Admin × Ability	2	50.115	25.058	1.982	0.13789
	Mode × Admin × Ability	2	20.868	10.434	0.825	0.43820
	Residual	16568	209510.208	12.645	--	--
Science	Mode	1	565.796	565.796	75.045***	0.00000
	Admin	2	1381.028	690.514	91.588***	0.00000
	Ability	1	334422.878	334422.878	44356.796***	0.00000
	Mode × Admin	2	107.709	53.854	7.143***	0.00079
	Mode × Ability	1	679.417	679.417	90.116***	0.00000
	Admin × Ability	2	270.191	135.095	17.919***	0.00000
	Mode × Admin × Ability	2	70.301	35.151	4.662**	0.00946
	Residual	16568	124912.498	7.539	--	--

* $p < .05$, ** $p < .01$, *** $p < .001$

Lu Wang, PhD

Lu Wang is a research scientist I in Assessment Transformation. Her research interests include speededness detection, statistical modeling and practical issues in testing.

Jeffrey Steedle, PhD

Jeffrey Steedle is a lead psychometrician in Assessment Transformation directing the team responsible for statistical analyses for the ACT test and guiding research studies related to maintaining measurement quality while making changes to the assessment program. Jeff holds advanced degrees in education, statistics, and educational psychology, and his research interests include assessment validation and motivation on achievement tests.