Technical Brief

2021-R2142

An Investigation of Speededness as a Possible Explanation for Mode Effects on the ACT

SHICHAO WANG, PHD, DONGMEI LI, PHD, AND JEFFREY STEEDLE, PHD

ACT.org

An Investigation of Speededness as a Possible Explanation for Mode Effects on the ACT

Shichao Wang, PhD, Dongmei Li, PhD, and Jeffrey Steedle, PhD

Testing time is an important aspect of test design that could impact examinee performance. Speeded tests set time limits so that few examinees can reach all items, and power tests allow most test-takers sufficient time to attempt all items. Educational achievement tests are sometimes described as "timed power tests" because the amount of time provided is intended to allow nearly all students to complete the test, yet this makes such tests speeded to some extent. Thus, speededness is often a matter of degree.

Speededness can be impacted by factors in the test administration process such as test delivery mode, technology, and devices (Camara & Harris, 2020). For example, it takes less time to select an option in online testing than it takes to bubble in an answer choice on a paper-and-pencil test, but examinees' reading rates might slow down when testing on computers if a significant amount of scrolling is required to read long passages on a small screen (Pommerich, 2004). However, the extent to which a test is speeded can affect how much impact these factors can have on student performance. For a test that is generously timed, the differences in response time caused by test delivery mode may have little impact on student performance. For aggressively timed or intentionally speeded tests, however, the differences in response time may have a greater impact. For example, Mead and Drasgow (1993) identified speededness as a moderator of mode differences; that is, larger mode differences were found in speeded tests than in timed power tests. Pomplun, Frey, and Becker (2002) also found greater mode differences for more speeded tests and hypothesized that differences in response mechanism were the primary causes of mode effects. Few studies, however, have directly investigated the relationship between the extent that a test is speeded and the observed mode differences of the test.

The purpose of the study was to investigate the potential interactions between the degree of speededness and observed mode differences in student performance for the ACT® test, a timed achievement test often used for academic planning and placement, college admissions, and scholarship eligibility. Based on prior research, it was hypothesized that more speeded tests would exhibit larger test delivery mode differences in test performance (i.e., paper testing versus online testing).

Method

Data

Data from three mode comparability studies that coincided with Saturday, National ACT administrations in October 2019, December 2019, and February 2020 were used for this study. Participants in each mode study were randomly assigned to paper or online testing conditions, and they took the four subject tests of the full ACT multiple-choice test (English, math, reading, and science). All participants received college-reportable scores. In each study, the same form was administered on paper and online, but a different form was used in each study. After data cleaning, sample sizes for the three studies were 3,583, 6,352, and 6,645. The score distributions and demographic breakdown of the samples were similar to a typical ACT National testing sample.

Speededness Detection

The change-point analysis procedure proposed by Shao, Li, and Cheng (2016) was used in this study to detect speededness. The procedure not only classifies an examinee as speeded or non-speeded, but it also identifies the point at which the examinee apparently started to exhibit speeded responding. Another advantage of this approach is that it does not require latency data (i.e., time spent on each item), which is impossible to gather for examinees testing on paper. For speeded examinees, the item response patterns were expected to exhibit a significant decrease in estimated ability starting at a certain item. Consider examinee i who sped through the last s_i items on a J-item test. Treating d_i as the examinee's decrease in estimated ability, the 3-parameter logistic item response model is updated to

$$P_{ij}(\theta) = \begin{cases} c_j + (1 - c_j)\dfrac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]}, & if\ j \le (J - k) \\ c_j + (1 - c_j)\dfrac{\exp[a_j(\theta_i - b_j - d_i)]}{1 + \exp[a_j(\theta_i - b_j - d_i)]}, & otherwise \end{cases}, \quad (1)$$

where a_j, b_j, and c_j are the item parameters of item j. The log-likelihood ratio test was adopted to pinpoint the change-point. That is,

$$H_0: s_i = 0$$

$$H_a: s_i > 0$$

Specifically, the procedures are described below:

1. Obtain the log-likelihood under H_0, $l_i^{H_0}$, and the log-likelihood under H_a, $l_i^{H_a}$, to compute

$$\Delta l_i = 2\left(l_i^{H_a} - l_i^{H_0}\right). \quad (2)$$

And $l_i^{H_0}$ is obtained by plugging in the maximum likelihood estimation (MLE) of θ_i given the entire response vector, $u_{i1}, u_{i2},\ldots, u_{iJ}$, for examinee i. Suppose $s_i = k > 0$, $l_i^{H_a}$ is obtained by the MLE of θ_i using responses of the first $J-k$ items and the MLE of $\theta_i - d_i$ obtained using responses of the last k items. That is,

$$l_i^{(k)} = \sum_{j=1}^{J}\left[u_{ij}\ln P_{ij}^{(k)} + (1-u_{ij})\ln Q_{ij}^{(k)}\right], \quad (3)$$

where $P_{ij}^{(k)}$ is P_{ij} in equation (1) with the respective MLEs, and $Q_{ij}^{(k)} = 1 - P_{ij}^{(k)}$. Thus,

$$\hat{s}_i = \arg\max_{k=1,2,\ldots,(J-1)}\left\{l_i^{(k)}\right\},$$

where $\hat{s}_i$ is the estimated number of speeded responses for examinee i; and

$$l_i^{H_a} = \max_{k=1,2,\ldots,(J-1)}\left\{l_i^{(k)}\right\}.$$

As Shao, Li, and Cheng (2016) pointed out, to avoid homogeneous responses and cases with $s_i = 1$ or $J-1$ (where there is no finite MLE), the MLE is set to be bounded by -4 and +4, which are typically used in the literature.

2. Derive the null distribution of Δl_i (no change point) by permutations of item order. Since it is assumed that an examinee started speeding from a certain item, and that the examinee has an ability that is the same for all items before the change point and a decreased ability that is the same after the change point, such a point would not exist anymore when the item responses are permuted. Thus, the distribution of Δl_i of the permutated data would be the null distribution of Δl_i,

$$\Delta l_i^{(b)} = 2\left(l_i^{H_{a,b}} - l_i^{H_0}\right), i = 1,2,\ldots,N,$$

where $l_i^{H_{a,b}}$ is the $l_i^{H_a}$ for each permutation b and $l_i^{H_0}$ stays the same with or without permutation.

3. Finally, examinees may be flagged for apparently speeded responding. False discovery rate (FDR) as suggested by Shao, Li, and Cheng (2016) is used to correct for multiple comparisons and to find the cutoff value T based on the null distribution of Δl_i. For any pre-specified cutoff T, the FDR is estimated by

$$\text{FDR} = \frac{B^{-1}\sum_{b=1}^{B}\sum_{i=1}^{N} I(\Delta l_i^{(b)} > T)}{\sum_{i=1}^{N} I(\Delta l_i^{(b)} > T)}, \quad (4)$$

where $I(\cdot)$ represents the indicator function, and B is the total number of permutations. With a given FDR level, we can find the smallest T with estimated FDR at or below this level. An FDR level of 0.2 (widely used by researchers) was used in this study. Examinees for which $\Delta l_i > T$ were flagged for speeded responding. If an examinee was flagged as speeded, the estimated change point was $(J - \hat{s}_i)$ as calculated in step 1.

In this study, R (R Core Team, 2020) was used to implement the change-point analyses. Permutation steps were very time consuming, so multithreading was used to reduce the computing time.

The change-point analysis results were summarized for each test mode in terms of the proportions of examines identified as speeded as well as the distributions and descriptive statistics of the change-point positions across all examinees. These statistics indicated the extent to which a test was speeded, and they were compared with observed mode differences at both the test level and the item level to examine the relationship between speededness and mode differences. Additional analyses were conducted to examine changes in mode effects after removing speeded examinees. Finally, survey results were analyzed for possible explanations of differential speededness.

Results

The number and proportion of examinees flagged by change-point analyses as speeded along with mode differences are summarized in Table 1. The mode difference results came from Table 4.1.1 in Steedle, Pashley, and Cho (2020). For the purposes of this research, the same raw-to-scale conversions are used for online and paper tests. Operationally, different online and paper raw-to-scale conversions were used operationally to adjust for mode differences and ensure that scale scores from online and paper tests are comparable.

Mode effects were generally consistent across three studies. Less than 1% and 5% of examinees were flagged as speeded for each mode in science and math, respectively, and the differences in percentages were small between modes. However, higher percentages of examinees were identified as speeded for English and reading for both modes, and the percentages were higher for examinees testing on paper compared to those testing online. On average across studies, 18.3% of the examinees who tested on paper and 5.3% of examinees who tested online were flagged as speeded in English. In reading, the percentages were 13% for paper and 6% for online. Greater mean differences between paper and online test scores were observed for the English and reading tests, which was consistent with the hypothesis that the greater the extent of speededness, the greater the observed mode effects would be.

Table 1. Number and Proportion of Examinees Flagged as Speeded and Observed Mode Differences

Speededness Effect Difference							Mode Difference	
Subject	Study	Paper		Online		Prop Difference (Online - Paper)	Scale Score Mean Difference (Online - Paper)	Effect Size
		N	Prop	N	Prop			
English	Oct 2019	289	0.16	48	0.03	-0.13	-0.80	0.13
	Dec 2019	664	0.21	251	0.08	-0.13	-0.70	0.12
	Feb 2020	602	0.18	161	0.05	-0.14	-0.63	0.10
Math	Oct 2019	2	0.00	19	0.01	0.01	-0.29	0.06
	Dec 2019	9	0.00	4	0.00	0.00	-0.25	0.05
	Feb 2020	15	0.01	23	0.01	0.00	0.07	-0.01
Reading	Oct 2019	314	0.17	116	0.07	-0.11	-1.50	0.22
	Dec 2019	421	0.13	262	0.08	-0.05	-1.06	0.16
	Feb 2020	307	0.09	98	0.03	-0.06	-1.19	0.18
Science	Oct 2019	86	0.05	50	0.03	-0.02	-0.62	0.12
	Dec 2019	164	0.05	94	0.03	-0.02	-0.19	0.04
	Feb 2020	34	0.01	8	0.00	-0.01	-0.39	0.07

Further analyses were conducted to investigate changes in mode differences after removing speeded examinees. It is important to note that after removing speeded examinees from the groups testing on paper and online, the two groups were no longer randomly equivalent, so the revised mode effects may have reflected both group differences and mode differences. Therefore, caution is warranted when interpreting these results. Because of the small percentages of examinees flagged as speeded and the small differences across modes in the math and science tests, the following analyses focused on the English and reading tests.

Table 2. Raw Score Descriptive Statistics and Comparisons for English and Reading

Subject	Study	Paper			Online			Mean Difference (Online-Paper)
		N	Mean	SD	N	Mean	SD	
English	**Total**							
	Oct 2019	1807	38.71	14.45	1776	40.65	14.26	1.94
	Dec 2019	3147	42.44	14.23	3205	44.16	14.28	1.72
	Feb 2020	3297	42.82	14.14	3348	44.28	14.45	1.45
	Speeded							
	Oct 2019	289	38.8	12.72	48	37.06	11.03	-1.74
	Dec 2019	664	42.58	11.83	251	41.04	12.29	-1.54
	Feb 2020	602	42.06	11.58	161	40.7	11.27	-1.36
	Non-Speeded							
	Oct 2019	1518	38.7	14.76	1728	40.75	14.33	2.06
	Dec 2019	2483	42.41	14.8	2954	44.43	14.41	2.02
	Feb 2020	2695	42.99	14.65	3187	44.46	14.57	1.46
Reading	**Total**							
	Oct 2019	1807	21.55	7.69	1776	23.29	7.82	1.73
	Dec 2019	3147	24.09	7.49	3205	25.32	7.44	1.22
	Feb 2020	3297	22.88	7.71	3348	24.31	7.67	1.43
	Speeded							
	Oct 2019	314	23.22	5.81	116	23.4	6.2	0.18
	Dec 2019	421	24.9	5.9	262	24.68	6.32	-0.21
	Feb 2020	307	23.68	5.66	98	23.69	5.44	0.01
	Non-Speeded							
	Oct 2019	1493	21.2	7.98	1660	23.28	7.92	2.08
	Dec 2019	2726	23.97	7.7	2943	25.37	7.53	1.4
	Feb 2020	2990	22.79	7.89	3250	24.33	7.73	1.54

Table 2 provides descriptive statistics for the ACT English and reading raw scores of paper and online examinees in the total group, the speeded group, and the non-speeded group. For both tests, the online mean scores were greater than the paper means for the total group, indicating that online examines performed better than paper examinees. Since the total paper and online groups were randomly equivalent, the observed differences can be attributed to mode effects. For the non-speeded group, online examinees also had higher means than the paper examinees, with the differences across modes being slightly larger than the total group. The trend was reversed for the speeded group, wherein online examinees had lower means than paper examines. The paper and online groups within the speeded and non-speeded groups, however, were not randomly equivalent, so the observed differences might have reflected a combination of group differences and mode differences.

Figure 1. Change-Point Position Distribution for English

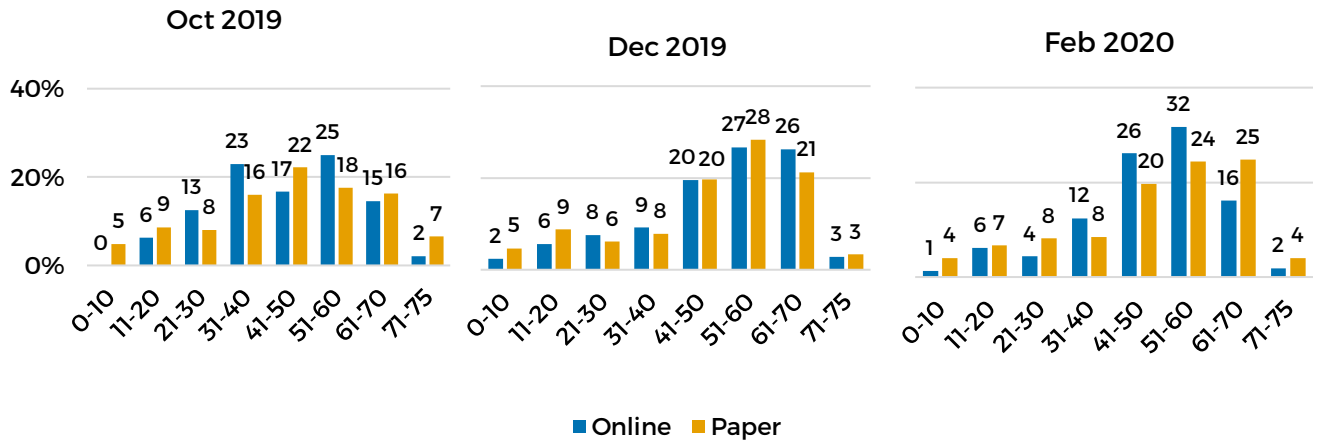
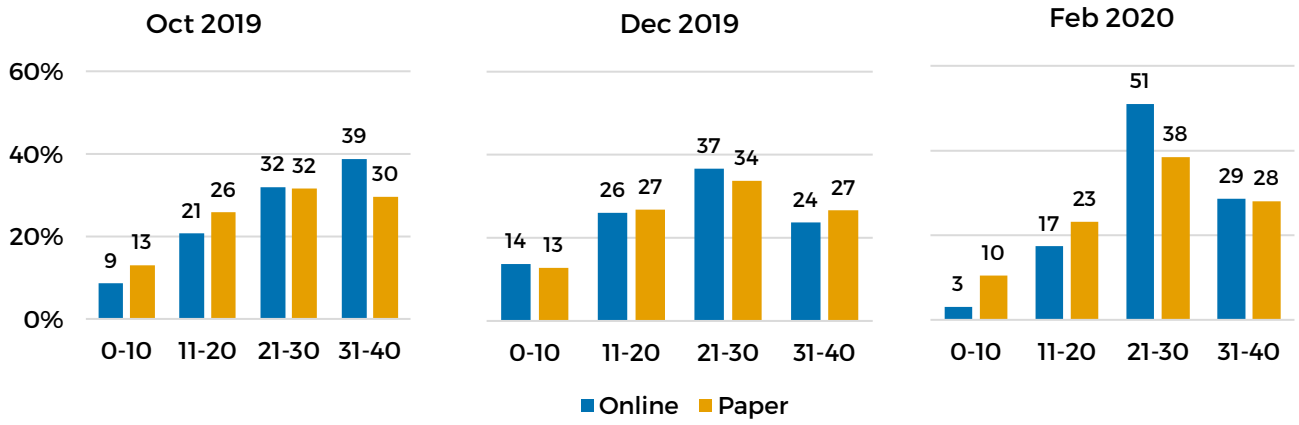


Figure 2. Change-Point Position Distribution for Reading



Distributions and descriptive statistics of the change point positions (i.e., where students began to speed) are shown in Figures 1 and 2 and Table 3. The mean change-point positions were greater for online test-takers compared to paper test-takers, which indicates that the point at which examinees started to speed occurred slightly later for those who took the online test. On average, the differences were within 1 to 3 item positions for the English test and 0 to 1 item position for the reading test. In addition, distributions of change point positions were bell-shaped, and trends were very similar for paper and online testing.

Table 3. Change-Point Position for English and Reading

Subject	Study	Paper		Online		Average Difference
		Mean	SD	Mean	SD	
English	Oct 2019	44.79	18.34	45.02	15.77	2.57
	Dec 2019	47.15	17.63	49.41	16.31	1.32
	Feb 2020	48.23	17.72	48.47	14.25	3.47
Reading	Oct 2019	22.52	9.48	25.28	9.17	0.31
	Dec 2019	22.49	9.34	22.37	9.30	0.05
	Feb 2020	23.31	8.47	25.90	7.13	1.35

Figures 3 and 4 show plots of *p*-value changes after removing speeded examinees for the English and reading tests, respectively. The vertical axis indicates the *p*-value (proportion correct) for the total group minus the *p*-value after removing the speeded examinees. Figure 3 shows a consistent downward trend that crosses the horizontal axis across studies and modes. That is, removing speeded examinees caused *p*-values to decline near the beginning of the English test, but removing speeded examinees caused *p*-values to increase toward the end of the English test. This pattern is consistent with the notion that speeded examinees gave good effort at the beginning of the test (and therefore contributed to higher *p*-values "before"), but they did not give good effort toward the end of the test (and therefore caused lower *p*-values "after"). In general, the changes in *p*-values were small (less than 0.02 in magnitude), but this would be expected since speeded examinees made up relatively small fractions of the total paper and online groups. The downward trend was stronger for examinees testing on paper, and this could simply reflect the fact that a greater number of examinees testing on paper were removed due to speeded responding. The trends in *p*-value changes were not as clear for the reading test. For examinees testing online, there was a weak downward trend (like the English test), but the changes were very small in magnitude (almost always < 0.01). For examinees testing on paper, there was no discernable downward trend; rather, *p*-values were generally higher (by a very small degree) before removing the speeded examinees. That is, the speeded examinees apparently contributed positively to the reading *p*-values for paper testing.

Figure 3. *P*-Value Change after Removing Speeded Examinees for English

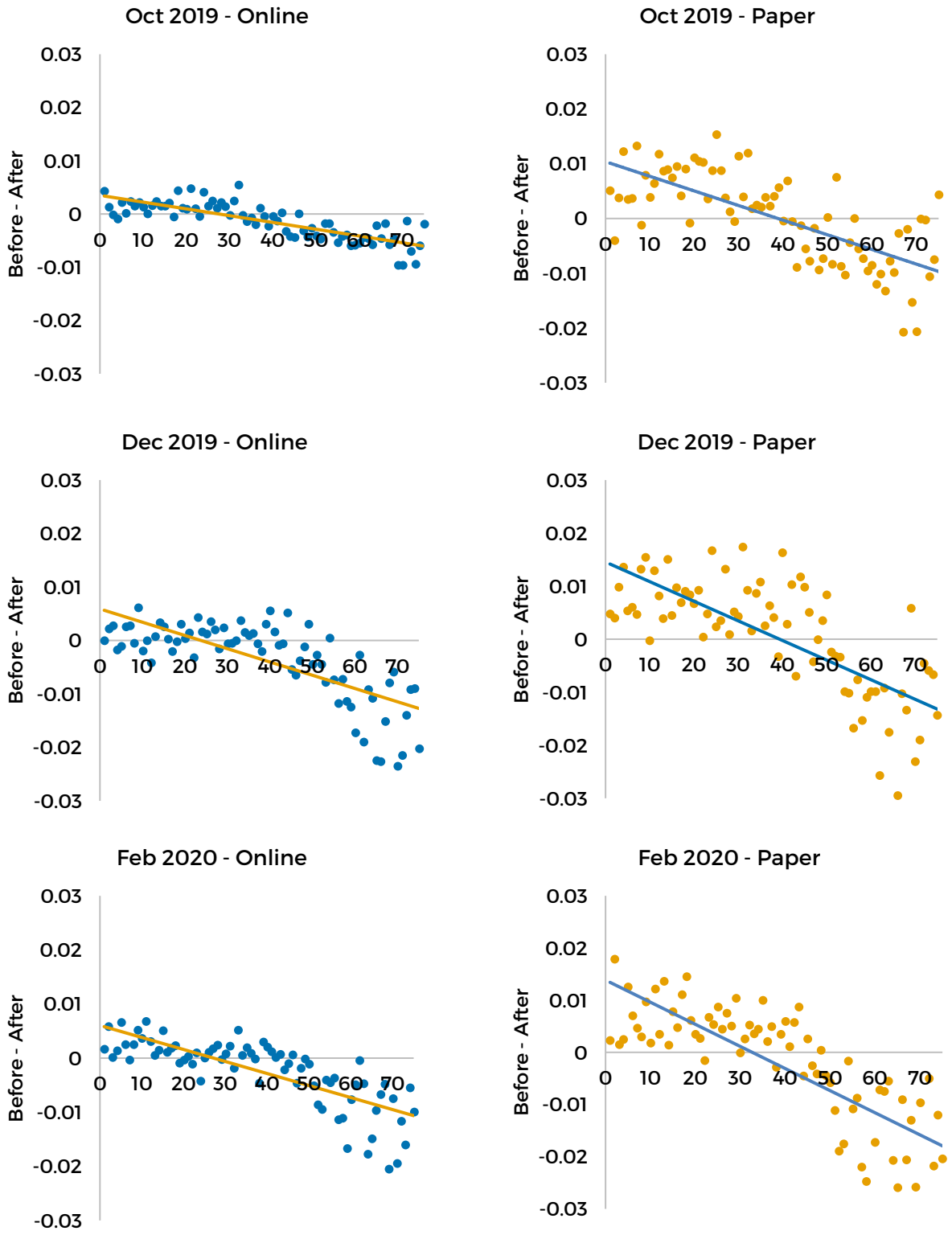


Figure 4. *P*-Value Change after Removing Speeded Examinees for Reading

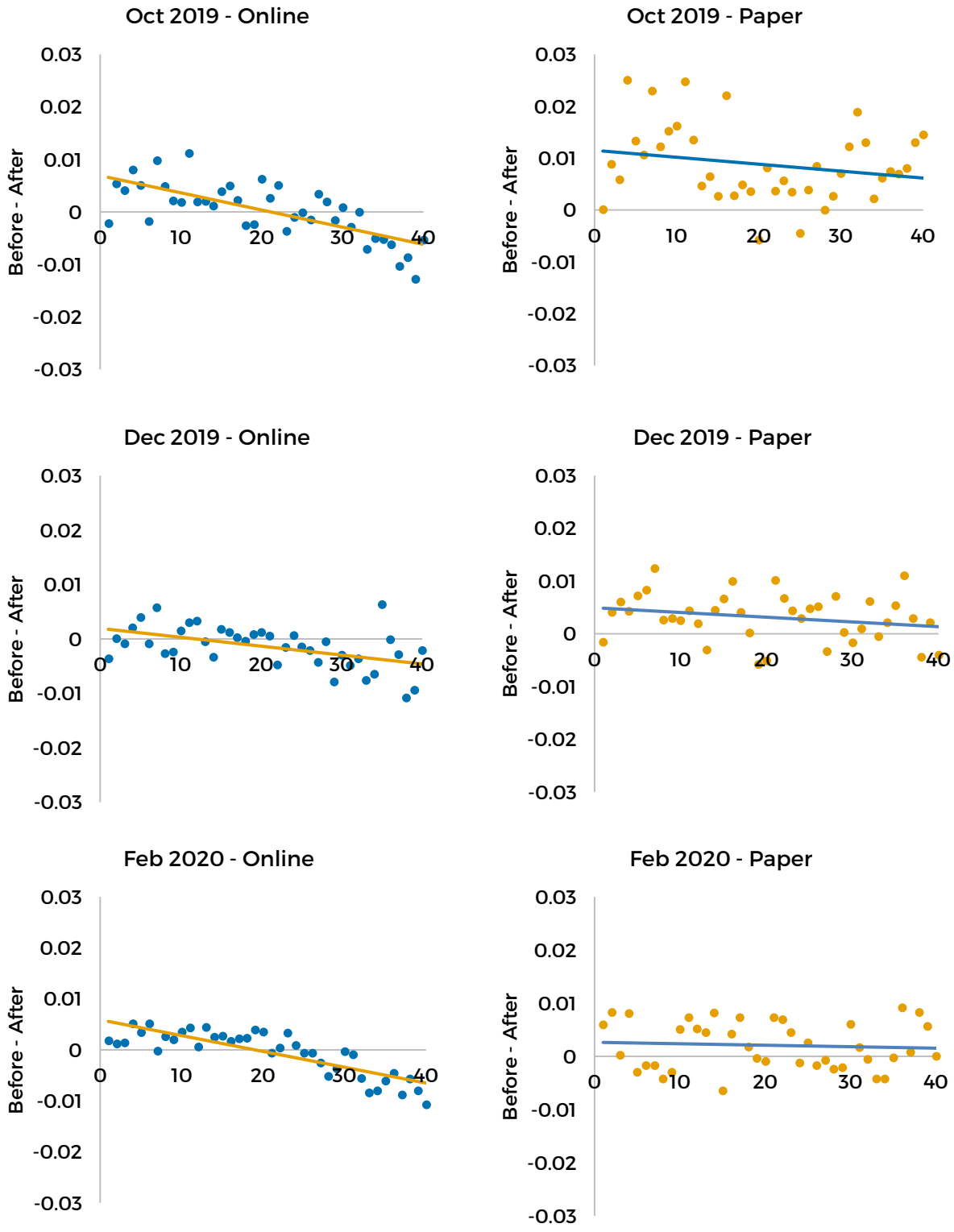
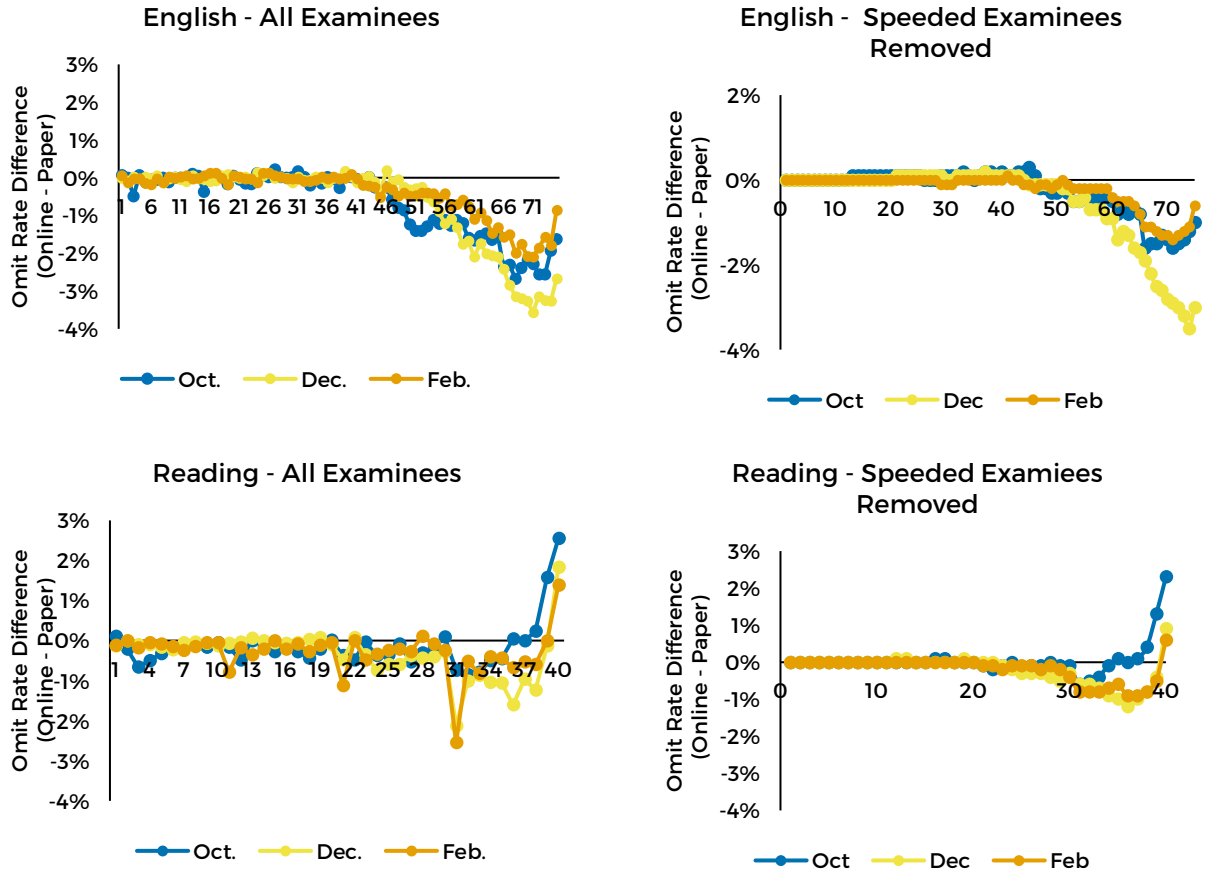


Figure 5 plots omit rate differences between online and paper for English and reading. The two plots on the left are omit rate differences for the whole group, and

the two plots on the right are omit rate differences for the non-speeded examines only. In the overall sample, it is obvious that online examinees were more likely to respond to most items and they were less likely to leave items blank. The difference was typically less than 1%, but the differences were substantially greater at the end of the English and reading tests. Omit rate differences across modes dropped slightly after removing speeded examinees. This finding is consistent with expectations, assuming that some examinees flagged for speededness left items blank near the end of the test because they ran out of time.

Figure 5. Omit Rate Differences Between Online and Paper for English and Reading



After the test, examinees were asked to fill out a survey about their testing experiences. The response rates were 13%, 8%, and 18% for the October 2019, December 2019, and February 2020 studies, respectively. Note that survey results must be interpreted with caution because the survey respondents may not have been representative of the full study samples. Results from the two survey questions related to the research question of this study are presented in Tables 4 and 5. Despite the statistical findings of this study, respondents apparently perceived the English test to be the least speeded. Specifically, around 73% of respondents agreed that they had enough time to finish the English test and around 20% disagreed. On the other

subject tests, there was greater balance between agreement and disagreement: 40% agree vs. 48% disagree for math, 46% agree vs. 44% disagree for reading, and 44% agree vs. 41% disagree for science. The responses to Q17 indicate that approximately 94% of survey respondents found the on-screen timer helpful for pacing during the test. In addition, a few students mentioned that testing online was faster and easier because of clicking responses compared to bubbling answers on paper. Another student commented that testing online could make math more difficult because it took more time to copy the answer down or mark things on graphs and visuals. These results must be interpreted with caution because the survey respondents were not necessarily representative of the full study samples.

Table 4. Survey Question and Responses of Q8

Q8: Please provide your level of agreement with the following statements: I had enough time to finish the English/mathematics/ reading/science test						
Subject	**Study**	**Strongly Agree**	**Agree**	**Neither Agree nor Disagree**	**Disagree**	**Strongly Disagree**
English	Oct 2019	31%	38%	8%	16%	6%
	Dec 2019	35%	40%	6%	13%	6%
	Feb 2020	35%	39%	7%	13%	6%
Math	Oct 2019	9%	25%	12%	34%	20%
	Dec 2019	17%	34%	13%	24%	13%
	Feb 2020	10%	26%	12%	33%	19%
Reading	Oct 2019	11%	34%	11%	30%	16%
	Dec 2019	15%	29%	10%	30%	17%
	Feb 2020	16%	34%	12%	24%	14%
Science	Oct 2019	15%	38%	15%	23%	10%
	Dec 2019	7%	23%	16%	32%	22%
	Feb 2020	14%	35%	16%	25%	10%

Table 5. Survey Question and Responses of Q17

Q17: How helpful or unhelpful was the timer for managing your pace throughout the test? (Question displayed only if respondent indicated that the ACT was taken online)					
Study	**Very Helpful**	**Helpful**	**Unhelpful**	**Very Unhelpful**	**Did not Use the Timer**
Oct 2019	65%	26%	6%	3%	1%
Dec 2019	70%	24%	4%	2%	1%
Feb 2020	72%	24%	3%	1%	0%

Summary and Conclusions

This study examined speededness and mode effects for the four sections of the ACT test. Change-point analyses were used to detect examinees that may have speeded through items. Using data from three mode studies, the proportions of examinees identified as speeded from paper and online testing were calculated, change-point position and descriptive statistics of raw scores were analyzed, and changes in *p*-values and omit rates were compared.

Results showed that few examinees were flagged for speededness in either mode on the math and science tests. Yet, a greater proportion of examinees testing on paper were flagged as speeded on the English and reading tests, which was consistent with the greater mode effects observed on the English and reading tests in Steedle et al. (2020). With speeded examinees removed, the mode differences in the non-speeded examinees were very close to those in the overall sample. This seems to indicate that, even though differential speededness across modes exists for some of the ACT tests, differences in speededness across modes may not be the only factor contributing to mode differences. Note that an FDR of 0.2 was used in this study. A change in the FDR value would have resulted in different numbers of examinees being flagged for speeded responding.

One limitation of the study is that speededness was only detected using the change-point analysis procedure, which focused on examinees whose ability decreases after a specific test item. Thus, for example, an examinee who rushes through the entire test at the same rate would be less likely to be detected. Other behaviors of speededness such as spending less time on the last few items or guessing at the end of the test were not considered. Future studies can use different approaches to detect speededness so that results from different procedures can be compared and validated against each other. In addition, latency data could be used to validate the speededness classifications for online examinees.

The survey results revealed that examinees perceived the English test to be less speeded than the other tests, though results from this study suggested that the English test may have been the most speeded test. Further investigations are needed to explain this inconsistency. In addition, the majority of online examinees cited the on-screen countdown timer as beneficial, and this could partly explain a difference in speededness between paper and online testing.

References

Camara W., & Harris, D. J. (2020). Impact of technology, digital devices, and test timing on score comparability. In M. J. Margolis & R. A. Feinberg (Eds.), *Integrating timing considerations to improve testing practices* (pp. 104–121). Abingdon, VA: Routledge.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*(3), 449–458.

Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *The Journal of Technology, Learning, and Assessment, 2*(6), 1–44.

Pomplun, M., Frey, S., & Becker, D. F. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement, 62*(2), 337–354.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing [software], Vienna, Austria. Retrieved from *https://www.R-project.org/*.

Shao, C., Li, J., & Cheng, Y. (2016). Detection of test speededness using change-point analysis. *Psychometrika*, *81*(4), 1118–1141.

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.

About the Authors

Shichao Wang is a psychometrician at ACT specializing in test equating, scaling, mode comparability and programming.

Dongmei Li is a lead psychometrician at ACT specializing in test equating, scaling, and growth modeling.

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.

Acknowledgements

The authors would like to thank Lu Wang for her contribution on earlier drafts of this report.