Stop Interrupting Me! Examining the Relationship Between Interruptions, Test Performance and Reactions

Amie D. Lawrence
Select International, Inc., alawrence@selectintl.com

Ted B. Kinney
Select International, Inc., tkinney@selectintl.com

Matthew S. O'Connell
Select International, Inc., moconnell@selectintl.com

Kristin M. Delgado
Select International, kdelgado@selectintl.com

STOP INTERRUPTING ME! EXAMINING THE RELATIONSHIP BETWEEN INTERRUPTIONS, TEST PERFORMANCE AND REACTIONS

Amie D. Lawrence[1], Ted B. Kinney[1], Matthew S. O'Connell[1] and Kristin M. Delgado[1]

1. Select International, Inc.

ABSTRACT

Unproctored testing (UIT) is common, and mobile testing is increasing rapidly, which means applicants are completing assessments in a variety of test environments. Little is known about how differences in the test environment are related to candidate test performance and reactions. This study examines interruptions in the UIT test environment to better understand what interruptions candidates are experiencing and how they influence candidate outcomes. The results show that candidates are being interrupted in a UIT context. Interrupted candidates scored lower on test performance and reported less favorable applicant reactions. Mobile device moderated the relationship between interruptions and test performance such that, non-mobile test takers' scores were more influenced by interruptions than mobile test takers. Interruptions happen and they do matter. Implications for organizations and practitioners are discussed.

As soon as psychological assessments began being used for selection purposes, personnel psychologists realized the importance of controlling the test environment. At first, proctored paper-and-pencil testing was the preferred method of administration because of the ability to standardize the test environment for all candidates and secure the test content. At the time of the first proctored paper-and-pencil test administration, no one knew how advancing technology would impact the nature of the test environment. With the advent of computers, many paper-and-pencil tests were converted to computer-based assessments, which led to immediate efficiencies in scoring procedures. As computer ownership increased and the Internet became ubiquitous, unproctored Internet testing (UIT) became the norm for most organizations. UIT has been a topic of discussion for over a decade. The pros and cons of allowing individuals to complete assessments in this setting have been debated at length (e.g., Tippins et al., 2006). Researchers have concluded that UIT does not lead to differences in measurement (O'Connell, Delgado, & Kung, 2012; Templer & Lange, 2008) or validities (Beaty et al., 2011).

Just as personnel psychologists accepted UIT, mobile devices entered the test environment landscape. Mobile testing (using a mobile device to complete an assessment) has raised a new set of challenges and questions about the impact of an uncontrolled test environment. Until recently,

practitioners assumed that most unproctored candidates were using a similar device with similar characteristics (e.g., keyboard, mouse) in somewhat similar environments. As users of UIT processes are starting to track device usage, it is apparent that these assumptions no longer hold true (Kinney, Lawrence & Chang, 2014; McClure-Johnson & Boyce, 2015). Further, as device-blocking technology is not widely understood in our field or, for that matter, implemented accurately, it is important to consider the impact that varying user contexts based on devices might have (Petor, Kinney, Chang, Lawrence & Moretti, 2016). For example, factors such as screen size, resolution, connection speed, and so forth, may impact the candidate test taking experience in a variety of ways (Sanchez & Branaghan, 2011; Sanchez & Goolsbee, 2010). Now, it's not just UIT but varying user contexts within a UIT environment that we need to examine. When allowing UIT, candidates not only choose their own test environment but also, with many types of mobile devices available, it further opens the universe of testing environments previously unavailable to most candidates (e.g., bus, train, park). The less controlled the environment is, the more likely it is that unexpected noises and/or events can occur that can distract or interrupt test takers. So, what

Corresponding author:
Amie D. Lawrence
Address: Select International 5700 Corporate Drive, Ste 250
    Pittsburgh, PA 15237
Email: alawrence@selectintl.com
Phone: 412-847-5306

is happening to candidates while they are testing? Are candidates being interrupted? If so, by what and how does it relate to outcomes? The purpose of this study is to explore these questions and others in order to help practitioners understand the unproctored and mobile unproctored testing environments with specific emphasis on interruptions. Our goal is to examine aspects of the UIT test environment and investigate how they relate to test performance and applicant reactions.

Test Environment

Although researchers have focused their energy on the psychometric integrity and validity of unproctored and mobile unproctored assessments (Arthur, Doverspike, Munoz, Taylor & Carr, 2014; Beaty et. al, 2011; Illingworth, Morelli, Scott & Boyd, 2015; Kinney, Lawrence & Chang, 2014; O'Connell et. al, 2012; O'Connell, Chang, Lawrence & Kinney, 2016; Templer & Lange, 2008), very little is known about the UIT test environment. There is scant published literature on unproctored test environments and no published literature that investigates test environment with mobile test takers. In two field studies examining overall favorability of test environments (comparing proctored and unproctored), Weiner and Morrison (2009) found that test takers rated proctored environments as most favorable for workspace and rated unproctored test environments as more favorable for noise (less noise). Although asking about noise, this study did not further investigate how differences in the environment (e.g., amount of noise) were related to test performance or favorability ratings. Wasko, Lawrence, and O'Connell (2015) took a closer look at test environment factors in proctored and unproctored settings. They did not find test performance differences related to environmental factors, but they did find differences in applicant reactions. Specifically, there were significantly lower ratings from individuals who were distracted, noted an inability to focus, and experienced technical problems. Similarly, Gray and his colleagues (Gray, Morelli, & McLane, 2015) investigated the mobile testing environment and found that mobile test takers are most likely to complete assessments in an indoor, static (as opposed to moving) location. Test takers who were moving and distracted rated less enjoyment than those who were static and not distracted. Other researchers have also noted lower applicant reactions in unproctored environments (Wasko, Raymark, & Moore, 2008; Fallow & Stokes, 2004) and mobile, unproctored environments (Gutierrez & Meyer, 2013; King, Ryan & Kantrowitz, 2014; Landers, Reddock, Cavanaugh & Proaps, 2014).

Interruptions: Intrusions and Distractions

An underlying assumption of the aforementioned studies is that unproctored environments, especially mobile unproctored, provide more opportunities for test takers to be interrupted, thus making it more difficult to concentrate. In

fact, one study shows that mobile applicants are interrupted (Gutierrez, Meyer & Fursman, 2015). In their sample, 72% of their participants cited receiving pop-up notifications during the testing process, and 12% took or received a phone call. Interruptions occur during unproctored test administration; as such, it is important to investigate how these interruptions impact applicant reactions and test performance.

Jett and George (2003) set forth a model for categorizing and studying interruptions. They describe four types of interruptions (intrusions, breaks, distractions, and discrepancies) and discuss the positive and negative consequences of each in a typical work environment. Although not specifically discussing interruptions in the context of testing, their rationale fits nicely into this literature. The two interruption types that most likely come into play with testing are intrusions (interruptions that stop the flow of work and bring it to a halt) and distractions (background noise and other external stimuli that interfere with concentration). Examples of intrusions could be interruptions by other people who require the full attention of the candidate to participate in a conversation or help with a task. If a candidate is completing an assessment on a smartphone, an intrusion could also be a phone call or text message that appears during the assessment. In these cases, the candidate would need to stop the assessment and restart at another time or keep the assessment running while they attend to the other task. Given that many assessments are timed, an intrusion during one of these assessments could be a large detriment to the candidate's performance. Other negative consequences of intrusions discussed by Jett and George (2003) are stress and anxiety from lost time and the loss of mental involvement or concentration in the task at hand. It is easy to see how intrusions could be a part of some candidates' test environment and how they could influence candidates' testing experiences.

What seems even more commonplace in UIT is the second type of interruption– distractions. In an uncontrolled environment, external stimuli (e.g., background noise) are likely to be present during testing. Research has shown that distractions can negatively affect one's performance on certain tasks due to cognitive interference and the conflict over working memory resources (Wickens & Hollands, 2000). When a primary task, such as a pre-employment assessment, is new or unfamiliar, people are more susceptible to the negative effects of distractions (Jett & George, 2003). Kinney et al. (2014) observed lower test scores for candidates who were distracted and that the difference was even larger for candidates using mobile devices.

It is reasonable to assume that distractions could also result in more negative reactions to preemployment assessments. Additional research on interruptions (which includes intrusions and distractions) has shown that individuals tend to have more negative perceptions about their work

experience when they have been interrupted regardless of how much it affected their performance (Speier, Valacich, & Vessey, 1999). This finding aligns with current applicant reactions research (e.g., Gray, et. al, 2015; Wasko et. al, 2015). Candidates tend to rate home as a favorable environment, and most unproctored candidates are completing assessments at home (Cober, Wasko, Smedley & Chan, 2008; Mastrangelo, Safran & Haaland, 2008; Petor, Lawrence & Kinney, 2017). With so many candidates choosing to complete assessments in a place where interruptions are possible, if not likely, knowing more about how these interruptions relate to outcomes is important for applied psychologists implementing tests in organizations. The current research suggests that intrusions and distractions have negative consequences on test performance and reactions. As such, this study intends to examine this further by exploring a set of research questions about candidate testing and interruptions.

Purpose and Research Questions

The purpose of this paper is to examine interruptions more closely by investigating three hypotheses in a large, unproctored applicant sample. All hypotheses are examined with the full sample and then split into mobile and nonmobile groups for an increased understanding of how results might differ for candidates choosing to participate using mobile devices.

Hypothesis 1: Number of interruptions differ by testing location, such that we expect testing center locations to have the least number of interruptions, and public locations to have the greatest number of interruptions.

Hypothesis 2: Interruptions are negatively related to test performance.

Hypothesis 3: Interruptions are negatively related to applicant reactions.

METHOD

Sample and Procedure

5,594 job applicants from 17 organizations who applied for entry-level manufacturing positions were included in the sample. As part of the application process, candidates completed an Internet-based assessment. The sample consisted of 33.1% women, 45.4% White/Caucasian, 37.6% African American, 10.4% Hispanic, and 0.9% Asians. Prior to launching the assessment, candidates were asked which device they were using to complete it. Approximately 16% of applicants completed the assessment on a mobile device (smartphone or tablet). This question was used to create a dichotomous mobile/nonmobile variable for use in the mobile analyses.

Measures

Test performance. The assessment battery used in this research is a 70-minute, propriety assessment call the Select Assessment® for Manufacturing. It was developed to measure 11 competencies (e.g., applied learning, teamwork, safety orientation, quality focus, responsibility, work ethic) relevant to success in entry-level manufacturing jobs. The battery used three measurement methods: self-report personality measures, situational judgment scenarios, and interactive simulations. This assessment has been used in numerous peer-reviewed studies (e.g., Bott, O'Connell, Ramakrishnan, & Doverspike, 2007; O'Connell, Hartman, McDaniel, Grubb & Lawrence, 2007; O'Connell, Kung, & Tristan, 2011; Peterson, Griffith, Isaacson, O'Connell, Mangos, 2011) and has strong validation evidence to support it (Kung, Lawrence, & O'Connell, 2012). A final test score was calculated by averaging the 11 final competencies scores. All competency scores were on a 10-point scale (M = 5.52, SD = 1.15).

Test environment. At the end of the assessment, candidates were presented with a postassessment survey that included several questions about test environment and reactions towards the assessment/assessment process. Although the survey was optional, 97% of the candidates completed it. The postassessment questions relevant to the current study asked candidates about the test environment, including test location, presence of others, background noise, phone call/text message/notifications/pop-ups received during the assessment, Internet connection, and technical problems. These items are used in the research question investigating the kinds of interruptions.

Applicant reactions. The postassessment survey also asked candidates to rate their perceptions of the assessment and assessment process. Using a 6-point agreement scale, they rated whether they thought the assessment was fair, allowed them to show their skills and abilities, was related to the job of interest, and provided positive impressions of the company. A total of six applicant reaction questions were included. The alpha reliability of the scale was (α = .90). In this study, an overall applicant reactions score was created by averaging the responses (M = 4.87, SD = .78).

Interruptions. The postassessment survey, which also included the test environment and applicant reaction questions, contained two questions that measured interruptions - distractions and intrusions. Distractions were measured with one self-report question: Did you experience any distractions while completing the assessment? (Yes/No). Intrusions were measured with one self-report question: How many times were you interrupted when completing the assessment? (None, 1, 2, 3, 4 or more). The intrusion variable was dichotomized into experienced intrusion (1) and did not experience intrusion (0). A final dichotomous interruption variable was created by combining the two variables. Any candidate who experienced a distraction or an intrusion was coded as experiencing an interruption.

The main dependent variable used in subsequent anal-

yses is the dichotomized interruption variable, which relies on self-reported distractions and intrusions as opposed to actual experiences. As suggested by a reviewer, we examined the proportions of candidates who indicated an interruption within each of the test environment experience items to ensure that individuals who endorsed certain testing experiences also reported interruptions. For example, someone speaking directly to you during testing is an example of an intrusion as defined by Jett and George (2003). We confirmed that individuals who experienced events like those defined by Jett and George (2003) endorsed that they were interrupted. Tables 1 and 2 show the results of this analysis.

Table 1 shows the proportion of candidates who reported an interruption by testing experience. Using a one-sample goodness of fit test in which we compared the interruption rates among those candidates who reported specific testing experiences, we observed that a larger proportion of candidates reported an interruption when other people were around, $\chi^2 (1, N = 1753) = 71.08, p <.001$, after being spoken to, $\chi^2(1, N = 1272) = 306.11, p <.001$, experiencing a technical issue, $\chi^2(1, N = 838) = 125.27, p <.001$, or receiving a phone call, $\chi^2(1, N = 996) = 84.44, p <.001$. However, there were no significant differences between the proportions of candidates who did and did not report interruptions while hearing background noise or receiving a text message, pop-up, push notification, or email.

Similarly, we examined the differences in interruption rates among the overall number of testing experiences reported. Results indicated that interruption rates were significantly lower when reporting one testing experience $\chi^2(1, N = 1663) = 297.18, p <.001$. There were no significant differences between interruption rates for candidates who reported two testing experiences, $\chi^2 (1, N = 1168) = 0.99, p = .32$, meaning that individuals who reported two testing experiences were equally likely to report an interruption as not. However, after three or more testing experiences were

reported, candidates were significantly more likely to report an interruption. Taken together, these data suggest that the dichotomous interruption variable used in this study accurately represents individuals who were interrupted during testing (likely multiple times).

RESULTS

Interruptions and Test Location

To begin our analyses, we calculated the percentage of candidates experiencing distractions, intrusions and interruptions (distractions or intrusions). Table 3 shows the prevalence of interruptions. The results show that 23.9% of candidates reported being distracted, 36.3% of candidates experienced one or more intrusions, with 18.7% of those citing one, 10.2% citing two, 4.4% citing three and 3% reporting four or more intrusions. When combined into a final interruption variable (distraction or intrusion), 41% of candidates experienced an interruption.

To investigate kinds of interruptions, we examined the test environment items to learn more about candidate experiences during testing. Specifically, candidates were asked whether they experienced the following: hearing background noise, having people around, experiencing a technology issue, being spoken to directly, or receiving a phone call, pop-up reminder, text message, email, or push notification. Candidates responded yes or no to all interruption types. Table 4 summarizes the prevalence of each experience, with hearing background noise as the most common at 50.3% followed by having people around (33.4%) and being spoken to (24.3%). Less than 10% of candidates reported receiving pop-ups, push notifications, and e-mails. See Table 5 for means, standard deviations, and intercorrelations among study variables.

To examine the relationship between test location and interruptions addressed in our first hypothesis, a chi-square test was used to assess whether interruption rates differed among the five testing locations: office, home, public,

TABLE 1.

Testing Experiences and Reported Interruption

Testing experience	Interruption reported	
	Yes (%)	No (%)
Heard background noise	49.4	50.6
Had people around*	60.1	39.9
Spoken directly to*	74.5	25.5
Experienced technical issue*	69.3	30.7
Received phone call*	64.6	35.4
Received text message	46.9	53.1
Received pop-up	58.9	46.1
Received push notification	49.8	50.2
Received email	49.8	50.2

Note. *indicates significant difference between proportions

TABLE 2.

Number of Testing Experiences and Reported Interruption

Count of testing experiences	n	Yes (%)	No (%)
0*	955	9.0	91.0
1*	1663	28.9	71.1
2	1168	48.5	51.5
3*	730	66.8	33.2
4*	347	74.4	25.6
5*	161	80.7	19.3
6*	63	85.7	14.3
7*	26	73.1	26.9
8*	4	100.0	0.0
9*	2	100.0	0.0

Note. *indicates significant difference between proportions

TABLE 3.

Percent of Interruptions Reported by Candidates

Type of interruption	n	Yes (%)	No (%)
Interruption (distraction or intrusion)	5,119	40.8	59.2
Distraction	5,136	23.0	76.1
Intrusion	5,531	36.3	63.7

TABLE 4.

Type of Interruptions Experienced by Candidates

Testing experience	n	Yes (%)	No (%)
Heard background noise	5,485	50.3	47.8
Had people around	5,515	33.4	65.2
Spoken directly to	4,531	24.3	56.7
Experienced technical issue	5,594	22.6	77.4
Received phone call	4,531	19.0	76.5
Received text message	4,531	14.2	66.8
Received pop-up	4,531	7.4	73.6
Received push notification	4,531	4.3	76.7
Received email	4,531	4.3	76.7

testing center, and other. Chi-squares are the statistical procedure of choice when both variables are categorical. In addition, with the large sample size, the available sample size per cell was more than five. Therefore, the assumptions for utilizing a chi-square were met.

Chi-square results show a statistically significant difference in interruption rates among the five testing locations χ^2 (4, N = 4647) = 30.35, p <.001. The strength of this relationship, as indexed by Cramer's V, was .08, indicating a weak effect. As seen in Table 6, interruption rates were lowest among office, testing center, and Other locations. As hypothesized, public locations had the highest interruption rates. Post hoc analysis using 2x2 chi square tests with Bonferroni correction revealed that public locations had significantly higher interruption rates than all other locations. The interruption rates between the rest of the location types were not significantly different from one another. These results suggest that there is a relationship between the probability of interruption and testing location. It appears that probability of interruption in public locations is higher than the other locations (office, home, testing center, other). However, there appears to be no differences in interruption rates among the remaining four locations. Hypothesis 1 is partially supported.

Next, we examined the relationship of mobile usage with interruption rate and testing location. A chi-square test revealed there was not a statistical difference between interruption rates for mobile and nonmobile users, χ^2 (1, N = 5076) = 2.23, p =.14, indicating that candidates testing on mobile devices were just as likely to be interrupted as candidates on PCs.

Interestingly, chi-square results showed a statistically significant difference in mobile device usage among the five testing locations χ^2 (4, N = 5104) = 116.75, p < .001. The strength of this relationship, as indexed by Cramer's V, was .12, indicating a moderate effect. Mobile device usage rates were significantly lower than PC usage for all locations, except home, where mobile device usage (96%) was significantly higher than PC usage (81%). These results may be surprising to many practitioners who believe that candidates using mobile devices are more likely to be in public places and are more likely to experience interruptions. Our results contradict this belief and show that, although most unproctored test takers chose to complete the assessment at home, all but 4% of mobile test takers tested at home. This observation further supports past research (Petor et al., 2017) that candidates, especially those using mobile devic-

TABLE 5.

Correlations Among and Descriptive Statistics for Key Study Variables

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Interruption	0.41	0.49													
2. Distraction	0.24	0.43	.67**												
3. Intrusion	0.22	0.41	.86**	.66**											
4. Heard background noise	0.51	0.50	.17**	.20**	.17**										
5. Had people around	0.34	0.47	.28**	.29**	.34**	.25**									
6. Spoken directly to	0.30	0.46	.35**	.34**	.42**	.06**	.41**								
7. Experienced technical issue	0.23	0.42	.25**	.15**	.24**	.01	.02	.02							
8. Received phone call	0.24	0.42	.19**	.14**	.21**	-.14**	-.03*	.03**	.02						
9. Received text message	0.18	0.38	-.01	.00	.02	-.06**	-.02	.02	-.01	.18**					
10. Received pop-up	0.09	0.29	.04**	.02	.07**	-.03**	.01	.00	.03*	.02	.06**				
11. Received push notification	0.05	0.22	.01	-.01	.01	-.01	.00	.01	.03*	.03*	.19**	.17**			
12. Received email	0.05	0.22	.01	-.01	.01	-.01	.00	.01	.03*	.03*	.19**	.17**			
13. Test score	5.48	0.93	-.05**	-.04**	-.06**	.05**	-.06**	.01	-.01	-.03**	-.01	.00	.01	.01	
14. Applicant reaction score	4.87	0.78	-.12**	-.11**	-.15**	-.07**	-.06**	-.04**	-.04**	-.03*	-.03	-.03*	.01	.01	.30**

TABLE 6.

Percent of Interruptions Reported by Candidates

	Test location				
	Office	Home	Public	Testing center	Other
No interuption	183 (69.6%)	2,423 (59.7%)	142 (48.6%)	58 (70.7%)	15 (68.2%)
Interuption	80 (30.4%)	1,632 (40.3%)	150 (51.4%)	24 (29.3%)	7 (31.8%)

Note. χ^2 = 30.35*, *df* = 4. Numbers in parentheses indicate column percentages.*$p < .001$.

es, are completing assessments at home.

Interruptions and Test Performance

To test Hypothesis 2, we compared the mean test scores of candidates who reported an interruption to those who did not report an interruption. We found that candidates who reported an interruption ($M = 5.44$, $SD = 1.10$) scored significantly lower than those who did not ($M = 5.57$, $SD = 1.16$), ($t(5117) = 5.16$, $p < .001$). To better understand the practical value, Cohen's d (Cohen, 1988) was calculated where d = .12. When interpreted in the context of Bosco, Aguinas, Singh, Field, and Pierce's (2015) recent recommendations for effect size benchmarks, this effect size is moderate. These results suggest that interruptions are negatively related to test scores. Hypothesis 2 is supported.

To understand how interruptions relate to test performance in the context of mobile device usage, we conducted a stepwise multiple regression analysis with test performance regressed on interruptions and mobile status as main effects (See Table 7). Both main effects were significant. Candidates who experienced an interruption scored lower than those who did not. For the mobile main effect, candidates who took the assessment on a mobile device scored

lower than those who did not. In Step 2 of the multiple regression, the interaction term was significant. This model accounted for 1.3% of the variability in test performance. After plotting the interaction, we observed that the negative relationship between interruptions and test performance is significantly stronger for nonmobile test takers compared to mobile test takers (See Figure 1).

Interruptions and Applicant Reactions

The same analysis approach used to test Hypothesis 2 was used for Hypothesis 3. Results showed significantly lower (i.e., less favorable) applicant reactions for those who were interrupted ($M = 4.95$, $SD = .78$) as compared to those who were not ($M = 4.75$, $SD = .78$, ($t(5101) = 8.79$, $p < .001$). The effect size was moderate for this outcome ($d = .24$; Bosco, et al., 2015). Hypothesis 3 is supported; interruptions have a negative relationship with applicant reactions.

To compare the mobile and nonmobile groups, we conducted a stepwise multiple regression with applicant reactions as the criterion (See Table 8). The main effect for interruptions was significant, such that the interrupted group reported less favorable applicant reactions compared to the

TABLE 7.

Summary of Hierarchical Moderated Regression Analyses for Interruptions, Mobile, and Test Performance (n = 5,075)
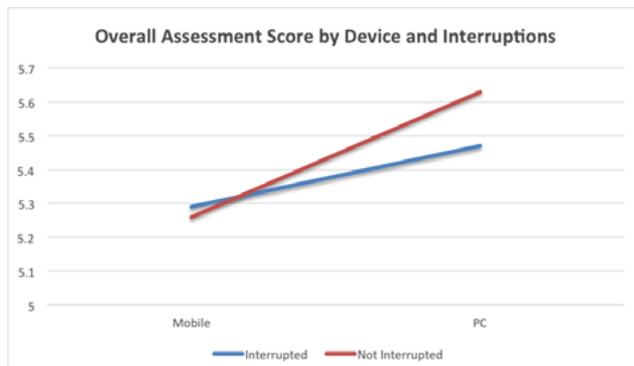
	Model 1			Model 2		
Variable	*B*	*SE B*	ß	*B*	*SE B*	ß
Interruptions	-.13	.03	-.05***	.37	.06	.12***
Mobile/nonmobile	.29	.04	.10***	.03	.08	.01
Interruptions x Mobile			-.18	-.18	.09	-.08*
R^2		.01			.01	
F for change in R^2		31.96***			4.52*	

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

TABLE 8.

Summary of Hierarchical Moderated Regression Analyses for Interruptions, Mobile, and Test Performance (n = 5,075)

	Model 1			Model 2		
Variable	*B*	*SE B*	ß	*B*	*SE B*	ß
Interruptions	-.13	.03	-.05***	.37	.06	.12***
Mobile/nonmobile	.29	.04	.10***	.03	.08	.01
Interruptions x Mobile			-.18	-.18	.09	-.08*
R^2		.01			.01	
F for change in R^2		31.96***			4.52*	

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

FIGURE 1.
Overall Assessment Score by Device and Interruptions

noninterrupted group. However, the mobile main effect and the interaction term were not significant. Overall, the model accounted for 2% of the variance in applicant reactions. Thus, these results suggest that applicant reactions do not differ based on mobile device usage, but they are negatively related to interruptions, such that interruptions are associated with less favorable applicant reactions.

Post Hoc Analyses: Distractions and Intrusions

We conducted several post hoc analyses in line with the Jett and George (2003) model that distinguishes distraction and intrusions as two types of interruptions. The previous analyses used a dichotomized interruption variable that combined distractions and intrusions. Post hoc analyses examined differences in test performance and applicant reactions with distraction and intrusion as the grouping variables. For test performance, candidates who reported an intrusion scored significantly lower ($M = 5.41$, $SD = 1.10$) than candidates who did not report an intrusion ($M = 5.56$, $SD = 1.17$), $t(4493) = 3.73$, $p < .001$. Similarly, candidates who reported that they were distracted scored significantly lower ($M = 5.42$, $SD = 1.10$) than candidates who were not distracted ($M = 5.55$, $SD = 1.14$), $t(5134) = 3.27$, $p < .001$. The effect sizes for these differences were $d = .14$ and $d = .11$, respectively, indicating a moderate effect for both, although differences are slightly larger in magnitude for intrusions.

For applicant reactions, candidates who reported an intrusion conveyed significantly less favorable reactions ($M = 4.66$, $SD = 0.77$) than candidates who did not report an intrusion ($M = 4.95$, $SD = 0.78$), $t(4481) = 10.18$, $p < .001$, $d = .32$, indicating a strong effect. Similarly, candidates who reported that they were distracted conveyed significantly less favorable reactions ($M = 4.91$, $SD = 0.78$) than candidates who were not distracted ($M = 4.92$, $SD = 0.79$), $t(5118) = 8.26$, $p < .001$, $d = .23$, indicating a moderate effect. Again, the differences in applicant reactions was larger in magnitude for intrusions. Taken as a whole, these results suggest that the presence of intrusions or distractions are associated with less favorable applicant reactions and lower test scores. From a practical standpoint, the effect sizes

indicate a moderately negative relationship with applicant reactions and, to a lesser degree, with test performance. These relationships appear to be larger in magnitude in the event the candidates experiences an intrusion versus a distraction.

DISCUSSION

As more organizations use preemployment assessments as part of their hiring processes, ensuring that all candidates have an equal chance of performing their best is important. A lot of attention has been placed on the psychometric properties of the assessments, but little focus has been directed at the unproctored candidate test environment and factors outside of the test that might influence outcomes. As such, this study addressed this topic by investigating testing experiences that could lead to interruptions and whether those interruptions relate to test performance and/or how candidates react to the assessment.

After examining the prevalence of interruptions, results show that about 40% of candidates are being interrupted either by a distraction or an intrusion. Intrusions are more common than distractions at 34%. This result means that over a third of unproctored candidates in this sample stopped the assessment at some point and restarted or relaunched. Our analysis of testing location showed that candidates in public environments were most likely to be interrupted (51%), and those in public settings were interrupted significantly more than all other test locations. Interestingly, regarding location, a large majority of candidates completed assessments at home (84%), and significantly more candidates using mobile devices (96%) tested at home as compared to nonmobile (81%). Despite the frequency of interruptions at home (40%), candidates seem to prefer that location.

We were also interested in learning more about the kind of disruptive events candidates were experiencing while testing. The most commonly reported test experiences were hearing background noise, having people around, being spoken to, experiencing technical problems, and receiving phone calls. To a lesser extent, candidates were receiving notifications, pop-up reminders, e-mail, and text messages. Additionally, we examined the total number of testing experiences reported by candidates. 81% of candidates experienced one or more reasonably disruptive event during their testing. Our data also suggested that candidates, in general, did not endorse the question about being interrupted until they had experienced three or more of the testing experiences, which was about 26% of the sample. So, whereas 40% of candidates endorse being interrupted, 81% report experiencing an event that could be considered reasonably disruptive, but it may take up to three of those to "feel" interrupted. Future research should further investigate disruptive events and their ability to cause a candidate to perceive an interruption.

A closer look at the interruption rates observed by mobile and nonmobile candidates showed no significant differences. This study suggests that although candidates using mobile devices have more contexts available to them, they are highly likely to be at home, and they are not more likely to be interrupted. This is good news from a practitioner's perspective in that it seems that the two device groups have similar test environments and experiences from an interruption standpoint.

Knowing the prevalence of interruptions is interesting and important it its own right, but of greater interest to most practitioners is the influence that interruptions have on candidates. This study examined test performance and applicant reactions as two relevant outcomes. Candidates who perform poorly are less likely to be hired and applicant reactions are related to important job relevant attitudes, intentions, and behaviors (see McCarthy et al., (2017) for a comprehensive review). Our results suggest that interruptions are moderately negatively related to test performance. Additionally, our results show an interaction effect with mobile devices. Interruptions are more strongly related to test performance among nonmobile candidates than among mobile candidates. Mobile candidates performed lower on the assessment regardless of the presence of interruptions; the mobile/interrupted and mobile/not interrupted group mean scores are highly similar, but this is not true for candidates using nonmobile devices. Nonmobile/interrupted test takers show a meaningful drop in their test scores as compared to nonmobile/not interrupted test takers.

Because the assessment used in this study contains multiple measurement methods, we hypothesized that the lower assessment scores shown by mobile candidates could be a result of the cognitive components and simulation methods included in the assessment, which studies suggest may not be equivalent across devices (O'Connell et. al, 2016). Therefore, we re-analyzed test performance using only the personality scores to see if we would observe the same results. Results of the personality-only measure of test performance resulted in the same pattern of results, (R^2 = .006, $F(3,5072) = 10.406$, $p < .001$), with a significant interaction between mobile usage and interruptions, ($\beta = -.270$, $p < .05$), such that the reduction in test scores for candidates who experienced an interruption were larger in magnitude when using a PC compared to a mobile device (See Figure 2). Observing the same pattern of results after removing the cognitive component of the assessment further supports the finding that there is a larger drop in test scores when an interruption is encountered by candidates on nonmobile devices than those on mobile devices. It is possible that candidates on mobile devices have a greater expectation of interruptions or a greater tolerance for interruptions. Future research should examine this finding further.

Regarding applicant reactions, interruptions matter. Our results show that candidates who were interrupted reported lower applicant reactions across the board. For distractions or intrusions, in all cases applicant reactions were significantly lower and the effect sizes were practically meaningful. So, whereas candidates' test scores may or may not have been lower, interrupted candidates reported more negative reactions. The relationship did not differ for candidates using mobile devices. This finding means regardless of the medium used, candidates who are interrupted are more likely to have lower applicant reaction scores. This study does not test the directional/causal nature of the relationship between interruptions and applicant reactions; thus, future researchers should examine this more closely. However, past research with applicant reactions and literature around interruptions suggest that experiencing interruptions may lead to lower confidence and more anxiety about test performance, which could influence reaction ratings (Jett & George, 2003).

Last, our study examined distractions and intrusions separately to determine if a particular type of interruption was more or less related to test performance or applicant reactions. The results of our analyses suggest they are not. The results for each separate interruption were the same as the results found with the final interruption variable.

Implications

Our study suggests that candidates are likely to be interrupted (distractions or intrusions) in a UIT context; our data suggest as many as 81% will experience a reasonably disruptive event, and 40% of our sample felt interrupted. Interrupted candidates are more likely to have lower test performance and less favorable applicant reactions. Interruptions happen and they do matter.

Organizations can use these results to provide guidance and recommendations to their candidates before they test. Based on our study, to maximize their chances of performing well and having a good assessment experience, candidates should complete the assessment in a nonpublic location, on a nonmobile device, and be free of interruptions. Although this may not always be feasible for all candidates, organizations may want to add this to their instructions and
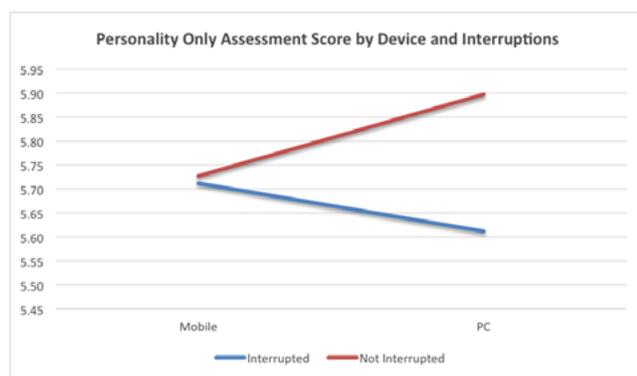
FIGURE 2. Personality Only Assessment Score by Device and Interruptions

suggestions before applicants launch the assessment.

Limitations and Future Research

This study was based on a large job applicant sample, which on the one hand made the results highly generalizable but on the other hand inevitably introduced issues typically associated with between-subject designs. Individual differences in skills and abilities, test motivations, and personal attributes may likely contribute to performance and reaction differences. It would be valuable to see if our findings can be replicated with within-subject designs in the future. Additionally, this study examines some common environmental stimuli, but we were not able to draw any conclusions about their interruption strength. Gaining a better understanding of types of interruptions often experienced by candidates and the level at which they are disruptive would be helpful in further research. Also, future studies should investigate additional environment factors that can engender distractions, especially factors that interplay with specific devices. Qualitative content analysis on self-reported distractions could be a good approach to identify distracters that have gone unnoticed.

This study represents a first look into the challenges that emerge for candidates in UIT environments and how they differ across devices. It is encouraging that candidates, by and large, do control their testing environment (e.g., it appears that "mobile candidates" are not actually mobile, despite the capability of their chosen device). Most mobile device research to date has investigated measurement equivalence across different modes of measurement. Measurement equivalence research is important in understanding the viability of continued usage of unrestricted UIT processes; however, researchers should also investigate the particular user contexts that may impact the user experience. Although this study represents a first foray into answering questions about how events in the UIT environment impact candidates across devices, our hope is that these initial answers spark other researchers to continue down this path. Lessons learned about the UIT environment across devices help practitioners to gain a greater understand the consequences associated with unproctored testing as candidates continue to "go mobile" when participating in selection processes.

REFERENCES

Arthur, W., Doverspike, D., Muñoz, G. J., Taylor, J. E., & Carr, A. E. (2014). The use of mobile devices in high-stakes remotely delivered assessments and testing. International *Journal of Selection & Assessment, 22*(2), 113-123. doi:10.1111/ijsa.12062

Beaty, J. C., Nye, C. D., Borneman, M. J., Kantrowitz, T. M., Drasgow, F., & Grauer, E. (2011). Proctored versus unproctored internet tests: Are unproctored noncognitive tests as predictive of job performance? International *Journal of Assessment and Selection, 19*(1), 1-10.

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100*(2), 431.

Bott, J. P., O'Connell, M. S., Ramakrishnan, M., & Doverspike, D. D. (2007). Practical limitations in making decisions regarding the distribution of applicant personality test scores based on incumbent data. *Journal of Business & Psychology, 22*, 123-134.

Cober, R., Wasko, L., Smedley, M., & Chan, S. (2008, April). *Impact of test-taking environment on test performance and validity.* Symposium presented at the 23rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Fallow, S. S., & Stokes, G. S. (2004, April). *Reactions to online selection systems: Examining differences by location.* Paper presented at the 19th Annual Conference of the Society for Industrial Organizational Psychology, Chicago, IL.

Gray, C. J., Morelli, N. A., & McLane. W. L. (2015). Does use context affect selection assessments delivered via mobile devices? In N.A. Morelli (Chair), *Mobile devices in talent assessment: The next chapter.* Symposium at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Gutierrez, S. L. & Meyer, J. M. (2013, April). Assessments on the go: Applicant reactions to mobile testing. In N. A. Morelli (Chair), *Mobile devices in talent assessment: Where are we now?* Symposium presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Gutierrez, S. L. & Meyer, J. M., & Fursman, P. (2015, April). What exactly drives positive reactions to mobile device administration? In N. A. Morelli (Chair) *Mobile devices in talent assessment: The next chapter.* Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Illingworth, A., Morelli, N., Scott, J., & Boyd, S. (2015). Internet-based, unproctored assessments on mobile and non-mobile devices: Usage, measurement equivalence, and outcomes. *Journal of Business & Psychology, 30*(2), 325-343. doi:10.1007/s10869-014-9363-8.

Jett, Q. R. & George, J. M. (2003). Work interrupted: A closer look at the role of interruptions in organizational life. *Academy of Management Journal, 28*(3), 494-507.

King, D. D., Ryan, A. M., & Kantrowitz, T. (2014, May). Mobile and PC delivered assessments: comparison of scores and reactions. In T. Kantrowitz & C. M. Reddock (Chairs), *Shaping the future of mobile assessment: Research and practice update.* Symposium presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Kinney, T.B., Lawrence, A. D., & Chang, L. (2014, May).

Understanding the mobile candidate experience: reactions across device and industry. In T. Kantrowitz & C. M. Reddock (Chairs), *Shaping the future of mobile assessment: Research and practice update.* Symposium presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Kung, M-C., Lawrence, A. D., & O'Connell, M. S. (2012). *Select assessment for manufacturing technical manual.* Technical manual. Pittsburgh, PA: Select International, Inc.

Landers, R. N., Reddock, C. M., Cavanaugh, K. J., & Proaps, A. B. (2014, May). Talent assessment using mobile devices. In T. Kantrowitz & C. M. Reddock (Chairs), *Shaping the future of mobile assessment: Research and practice update.* Symposium presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Mastrangelo, L., Safran, A., & Haaland, D. (2008, April). *I can apply from home? Applicant reactions at Frito-Lay.* Symposium presented at the 23rd Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing "so what?", "what's new?" and "where to next?" *Journal of Management.* Advance online publication. doi: 10.1177/0149206316681846

McClure-Johnson, & Boyce (2015, April). Selection testing: An updated look at trends in mobile device usage. In N. Morelli (Chair), *Mobile devices in talent assessment: The next chapter.* Symposium presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

O'Connell, M., Kung, M. C., & Tristan, E. (2011). Beyond impression management: Evaluating three measures of response distortion and their relationship to job performance. *International Journal of Selection and Assessment, 19*, 340-351.

O'Connell, M.S., Chang, L.C., Lawrence, A.D., & Kinney, T.B. (2016). PC-Mobile equivalence of four interactive simulations: A within-subject design. In J. Ferrell (Chair), *Going mobile: Empirical evidence from higher fidelity mobile simulations.* Symposium presented at the 31st Annual Conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.

O'Connell, M. S., Delgado, K., & Kung, M. C. (2012, April). *Does proctoring impact measurement methods differently? An evaluation in a high stakes testing environment.* Paper presented at the 26th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A, Grubb, W. L. III, & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment, 15,* 19-29.

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance, 24,* 270-290.

Petor, J. R., Kinney, T. B., Chang, L., Lawrence, A. D., & Moretti D. (2016). Auto-Detection Versus Self-Report: Best Practices in Mobile Device Research. *The Industrial-Organizational Psychologists, 54* (1).

Petor, J. R., Lawrence, A.D., & Kinney, T. B. (2017, April). *Painting the picture: What is the mobile test environment?* Paper presented at the 32nd Annual Conference of the Society for Industrial Organizational Psychology, Orlando, FL.

Sanchez, C. A., & Branaghan, R. J. (2011). Turning to learn: Screen orientation and reasoning with small devices. *Computers in Human Behavior, 27*, 793–797.

Sanchez, C. A., & Goolsbee, J. Z. (2010). Character size and reading to remember from small displays. *Computers & Education, 55*, 1056–1062.

Speier, C., Valacich, J. S., & Vessey, I. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences, 30*, 337–360. doi:10.1111/ j.1540-5915.1999.tb01613.x

Templer, K. J., & Lange, S. R. (2008). Internet testing: Equivalence between proctored lab and unproctored field conditions. *Computers in Human Behavior, 24*(3), 1216-1228. doi:10.1016/j.chb.2007.04.006.

Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., & Shepherd, W. (2006). Unproctored Internet testing in employment settings. *Personnel Psychology, 59*, 189–225. http://doi.org/10.1111/j.1744-6570.2006.00909.x

Wasko, L. E., Raymark, P., & Moore, D. (2008). *Antecedents and consequences of applicant perceptions within an internet-based testing context.* Unpublished Dissertation Manuscript. Clemson University, Clemson, SC.

Wasko. L. E., Lawrence, A. D., & O'Connell, M. S. (2015, April). *What matters in the test environment?* Paper presented at the 30th Annual Conference of the Society for Industrial Organizational Psychology, Philadelphia, PA.

Weiner, J. A. & Morrison, J. D. (2009). Unproctored online testing: environmental conditions and validity. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*(1), 27-30.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.