

# Don't rush me!

## How Time Limits Affect Candidates' Testing Experience and Scores



Mary Ann Hanson, Christine Carvalho, and Matt Stewart  
Pearson; Bloomington, MN and London, England

### Introduction

Most cognitive assessments limit the time applicants have to answer test questions, so it is important to understand how a time limit affects what is being measured, test-taker reactions, and the validity of the test.

The distinction between how quickly one can solve problems (**speed**) and how difficult the problems one can solve (level or **power**) has been made in ability testing for years. In order to measure abilities in the power realm with fidelity, it is important that test takers have time to consider and attempt all items; in other words, it is important that speed is not a factor (e.g., Lu & Sireci, 2007, p. 29).

**Experts recommend that test publishers assess and report how time limits affect scores.**

### Approach

The third edition of the Watson-Glaser Critical Thinking Appraisal (Watson-Glaser III), a well known measure of critical thinking, includes a the 30-minute time limit (Watson & Glaser, 2018). Because this test focuses primarily on power, we conducted research to address the following questions:

- Is the 30-minute time limit sufficiently generous?
- How does imposing a time limit affect **perceptions of the testing experience**?
- Does the time limit affect some respondents more than others** (e.g., those with high test-taking anxiety)?

A total of 137 respondents completed the Watson-Glaser III on Mechanical Turk twice, once with a 30-minute time limit and once with unlimited time (half completed the timed first and half completed the untimed first).

### Does the 30-minute time limit affect Watson-Glaser III scores?

Across both conditions, with and without the time limit, all participants attempted all items with only one exception (see Figure 2). Also, the vast majority of test takers reported they had adequate time to complete the test to the best of their ability. The percent reporting they had sufficient time did not differ significantly across the timed versus untimed condition.

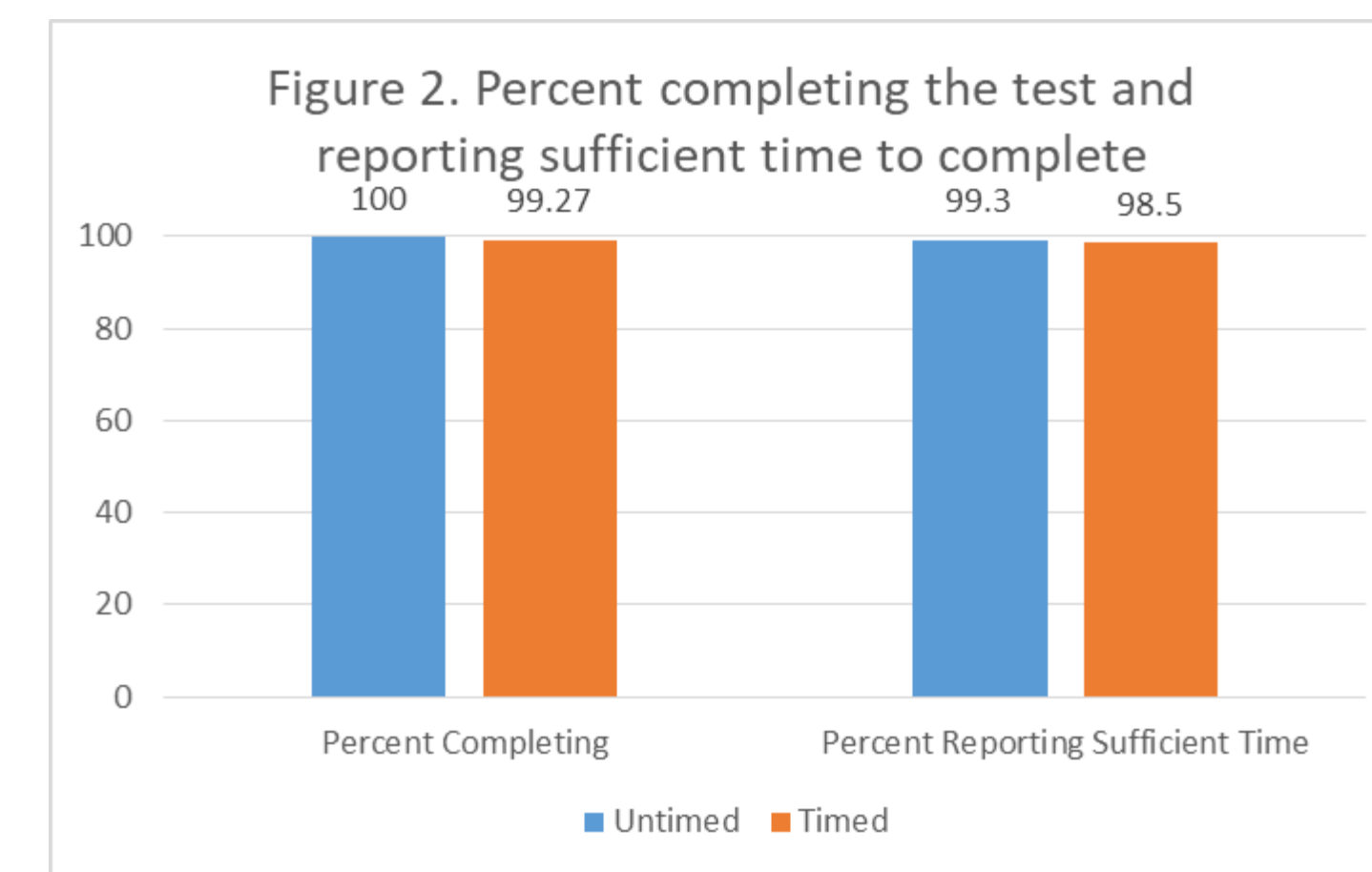


Table 1 shows the mean Watson-Glaser III score obtained in the timed administration across both samples and the mean for untimed administration. The mean score did not differ across the timed versus untimed condition any more than would be expected by chance.

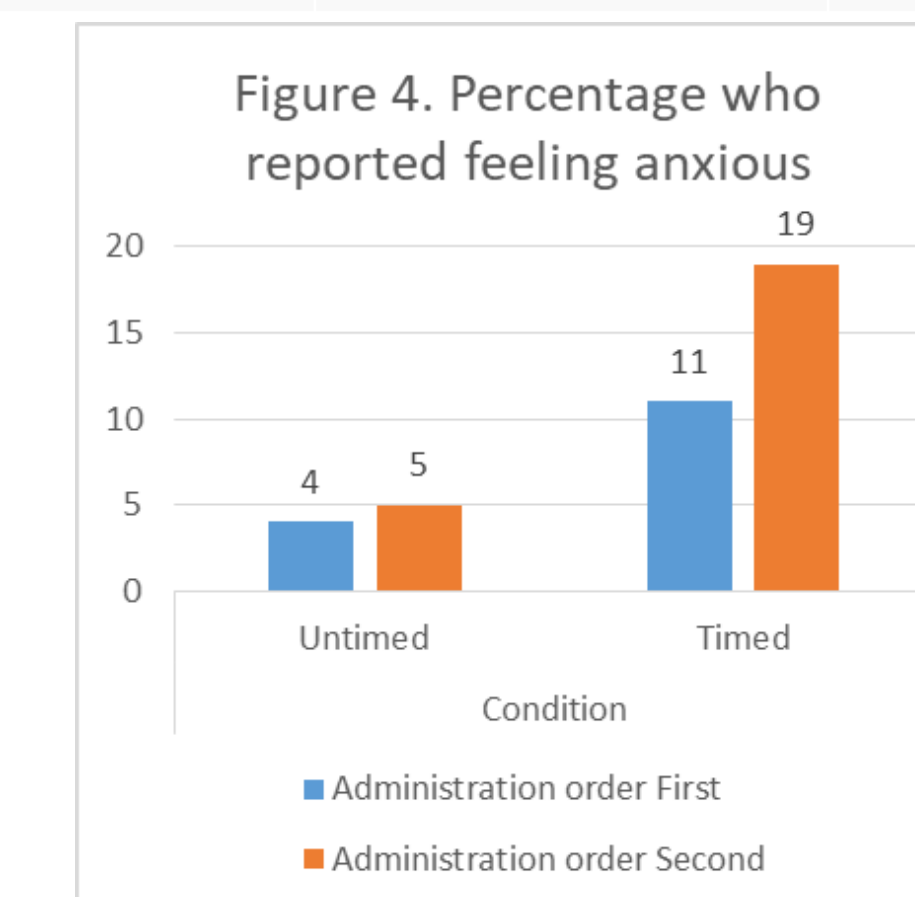
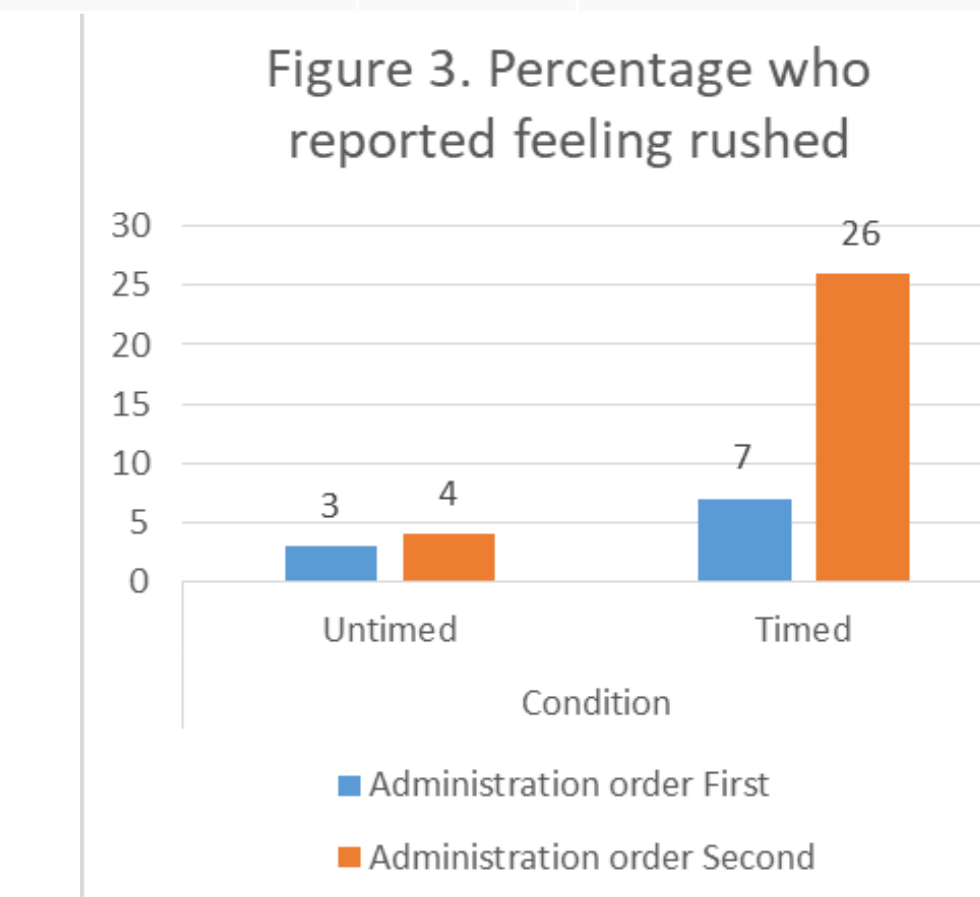
Table 1. Watson-Glaser III scores timed versus untimed

	Watson-Glaser Scores <sup>1</sup>		
	Mean (theta)	Standard Deviation	F-test
Timed	-.35	0.81	.78 (ns)
Untimed	-.41	0.80	

Table 2 shows that the correlations across the timed and untimed conditions are high, higher than the reliability of the test forms in this sample, so when they are corrected for unreliability the "true" correlation would be greater than 1.0. Another way to look at this would be to use Spearman-Brown to show that the observed correlation of .73 is not significantly different in magnitude than the split-half reliability that would be expected from simply administering a test that is twice as long (.77 plus or minus .06).

Table 2. Watson-Glaser III internal consistency reliabilities and intercorrelations

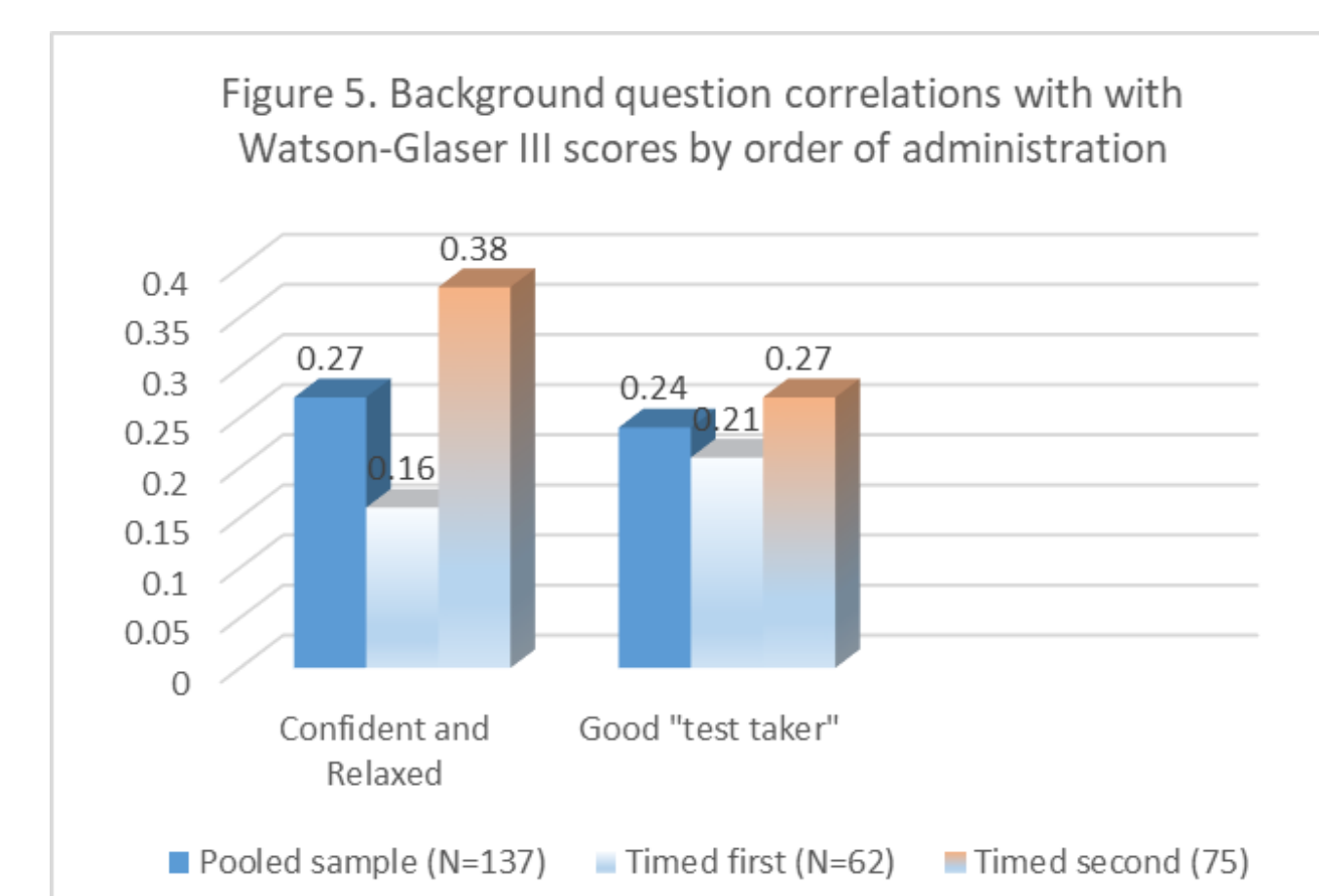
	N	Form 1 Reliability	Form 2 Reliability	Correlation Between Timed/Untimed	Pooled SD theta
Sample A	75	.61 (untimed)	.58 (timed)	.66	0.76
Sample B	62	.67 (timed)	.62 (untimed)	.79	0.85
Pooled	137	.64	.60	.73	0.80



### Does knowing there is a time limit affect test-takers' perceptions?

The results related to test takers' perceptions are a bit more complicated. Figures 3 and 4 show that test takers generally reported feeling more anxious and rushed in the timed condition. However, this difference only reached significance for those who completed the untimed version before the timed version. For the sample who completed the timed version first the difference was small and not statistically significant.

### Does the time limit affect some test-takers differently than others?



For more information visit  
[TalentLens.com](http://TalentLens.com)  
or Booth 527.

Finally, within each of the samples and administration conditions, we examined the correlations of self-reported test-taking confidence and test-taking ability with Watson-Glaser III scores. These variables correlated significantly with Watson-Glaser III scores in the timed condition (.27 and .24 respectively) but not in the untimed condition. Figure 5 shows these timed condition results for the pooled sample, then **separately for those who completed the timed version first versus second**. This figure shows that the effect was more pronounced for those who completed the timed version first (.38 and .27 respectively). That is, timing has a stronger effect on test taker perceptions and also on how self-perceptions affect scores when examinees complete the timed test **after** completing an untimed administration.

### Discussion and Conclusions

**Based on all of these results, the 30-minute time limit appears to allow plenty of time.**

From both the score-oriented (mean differences) and fairness-oriented (correlational) perspective the 30-minute time limit for the Watson-Glaser III does not introduce an unintended speeded component to Watson-Glaser scores. Results indicate that the ability measured with the time limit is the same as the ability measured without the time limit (Watson & Glaser, 2018).

**Even so, knowing that there is a time limit affects respondent perceptions.**

Those who completed an untimed test before the timed test felt more rushed and anxious, and test taking confidence and skill were more likely to affect scores. These were surprising results. One possible explanation is that completing the test with unlimited time first makes the time limit instructions more salient. This suggests that efforts to make time limits less salient may reduce any unintended or unwanted effects on test scores.

How applicants react to testing is complicated and understudied (see for example Truxillo et al., 2009). Applicant reactions can have an effect on motivation and on testing outcomes (Chan & Schmitt, 2004). The present research adds to this discussion by illuminating how one somewhat complicated aspect of test instructions impacts test takers' scores and their perceptions of the testing experience.

**While the type of research reported here is stipulated in the AERA Standards for testing, it is virtually absent in the literature.**