The effect of item response time patterns on ability estimates in high stakes computer adaptive testing

Gerald Giraud, Ph.D.
Russel Smith, Ph.D.

June, 2005

Contact information:

Gerald Giraud, Ph.D.
708 N. Main Street
Gordon, Nebraska 69343

email: ggiraud@olc.edu

Abstract

This study examines the effect of item response time across 30 items on ability estimates in a high stakes computer adaptive graduate admissions examination. Examinees were categorized according to 4 item response time patterns, and the categories are compared in terms of ability estimates. Significant differences between response time patterns were observed. Highest ability estimates were associated with a response time pattern that was consistent across items, while lowest ability estimates were associated with long response times on items early in the test and short response times late in the test. These results suggest that teaching examinees to manage time effectively can maximize ability estimates. An alternative interpretation is that more able examinees require less time to respond to items.

The effect of item response time patterns on ability estimates in high stakes computer adaptive testing

Objectives of inquiry

Response time has been referred to as "psychology's ubiquitous dependent variable" (Luce, 1986, p. 1). Cognitive psychologists have researched response times "because how long it takes someone to process something is thought to indicate something about *how* the person processed it" (Schnipke & Scrams, 1998, p.4). Item response time in a testing situation refers to the amount of time it takes an examinee to select his or her response once the item has been presented. In addition to gaining a better understanding of *how* examinees process information, there are other practical reasons for studying item response times in testing situations. Recent research has focused on the effect of response time strategies on ability estimates (e.g. Narayanna, Durso and Roussos, 2000). This paper examines the relationship between item response time patterns and ability estimates in a high stakes graduate school admissions test.

Whether response time patterns affect ability estimates for examinees is a concern when tests are timed, requiring examinees to choose how best to utilize allotted time. This study seeks to address the following questions:

Do different time use strategies result in different ability estimates?

Are some time use strategies more or less effective, in terms of maximizing examinee ability estimate, than others?

Does time use strategy affect ability estimate, taking into account item difficulty, item word count, and the language fluency of the examinee?

Does time use affect the rate of correctly answering items?

Methods

Source of data

Item response data from the verbal section of a large scale, high stakes graduate entrance examination were utilized in this study. Test items were intended to assess reading comprehension, vocabulary and other verbal skills. The test was computer adaptive, and ability estimates resulting from the test were determined using item response theory. An algorithm that supplied items based on the ability estimate of the examinee determined by items previously answered determined item selection during test administration. The algorithm was designed to minimize ability estimation error. The test was terminated after 75 minutes.

The data set included 30 item responses for 5,447 examinees. Relevant variables in the analysis reported here were examinee ability estimate, and response time for each item. Examinees were divided into 4 groups. Assignment to group was determined by item response time pattern.

Categorization of examinees
The algorithm for categorization of examinees in terms of time use was as follows:

The 30 items were divided into 5 parts of 6 items each. The average response time per item was determined for each 6-item part. The difference in time use per part was computed by subtracting average time per part from the immediately subsequent part, resulting in 4 time difference measurements. This allowed for the identification of time use patterns across test parts. The examinees were grouped into 4 general patterns of time use:

Pattern 1: Time use decreased precipitously for each test part. This pattern is characterized by the longest response time in the first test sections, and a sequential decline in response times to the end of the test.

Pattern 2: Response time decreased but less precipitously than pattern 1.

Pattern 3. Response time was flat for the first 4 test parts, but dropped significantly for the last part.

Pattern 4. Response time was initially less than the other patterns, and remained similar over all 5 test parts, with a slight but insignificant increase in time use as the test progressed.

Figure 1 displays the 4 time use patterns used for comparison in this study.

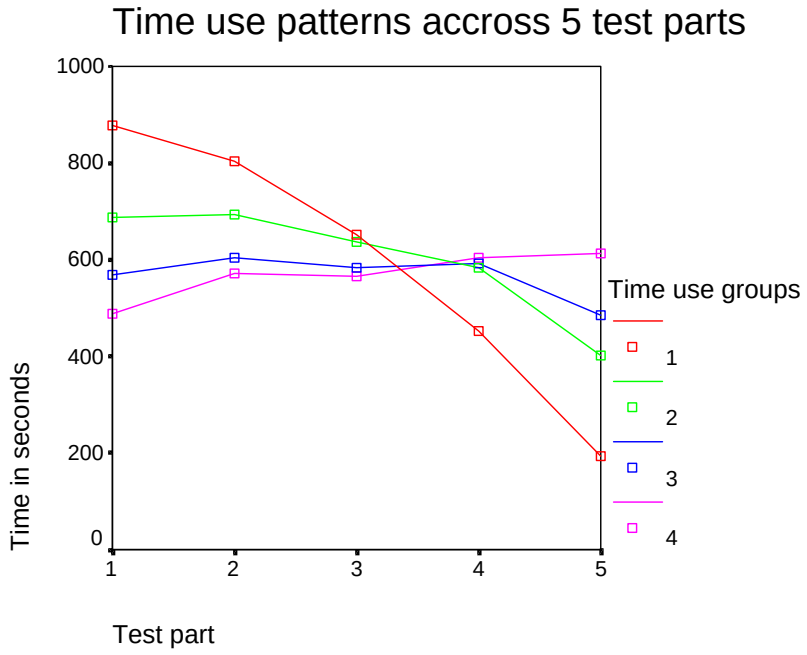


Figure 1.

Analysis

To determine whether different time use strategies result in different ability estimates, and whether some time use strategies are more or less effective, in terms of maximizing examinee ability estimate, response time patterns were compared for differences in examinee ability estimates using an ANOVA procedure. Post-hoc follow up was done using a Tukey test. Effect size estimate (eta squared) is reported as well as significance test results.

To determine the effect of time use strategy on ability estimate, taking into account item difficulty, word count, and the language fluency of the examinee, a multiple regression analysis was done. Time use strategy was represented in the regression analysis by the summed time use difference across five 6-item test sections, as reported above. Item difficulty was represented by the average item difficulty computed for 30 items for each examinee and word count was represented by the average word count for 30 items for each examinee. Language fluency was represented on a 15-point scale with higher numbers indicating greater fluency. Variables were entered in the following order into the regression equation: item difficulty, word count, language fluency, and time use.

To determine whether time use pattern affected the rate of correct answers, a one-way ANOVA procedure was conducted, using the number correct on the final 6 test items as the dependent variable, and time use pattern as the independent variable. In addition, the difficulty of the final 6 six items was compared across time use patterns.

Results

Mean ability estimates (reported as Theta) for response pattern groups are shown in Table 1.

Table 1. Mean ability estimate (Theta) for 4 time use pattern groups.

Time use pattern	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	-.360	.028	-.414	-.306
2	.115	.028	6.098E-02	.169
3	.463	.028	.409	.518
4	.576	.028	.522	.630

Result for ANOVA test of ability estimate differences among time use patterns is shown in Table 2. A significant difference was found among time use patterns in terms of ability estimate. The effect of time use pattern on ability estimate was moderate.

A Tukey follow up procedure revealed that response pattern 1 yielded a lower ability estimate than all other patterns, and that ability estimates significantly increased through patterns 2 through 4. Pattern 4 yielded the highest ability estimate.

Table 2. ANOVA result for comparison of time use pattern by ability estimate

	Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Item response pattern	722.637	3	240.879	231.297	.000	.113
Error	5668.501	5443	1.041			

Regression analysis revealed that time use strategy had a small but statistically significant effect on ability estimate, taking into account item difficulty, word count and language fluency. R-square and R-square change statistics are reported in Table 3.

Table 3. Multiple regression results: R-square and R-square change

Model*	R square	R Square Change	F Change	df1	df2	Sig. F Change
Item difficulty	.705	.705	11362.435	1	4748	.000
Word count	.712	.007	117.719	1	4747	.000
Language fluency	.716	.003	53.787	1	4746	.000
Time use	.759	.043	856.930	1	4745	.000

*Variables listed are entered in order, each R-square includes all previous variables

One-way ANOVA revealed a significant difference among time use patterns in terms of items correct in the final 6-item test part. A post-hoc examination of time use pattern group means indicated that time use pattern 1 resulted in significantly lower number correct than all other patterns. Patterns 3 and 4 resulted in the highest number correct. These results were observed while average item difficulty was lowest for pattern 1, and highest for pattern 4. See Table 4.

Table 4. Average item difficulty and number correct by time use pattern group for final 6 test items.

	Time use pattern	N	Mean	SD
Difficulty	1	1171	-.3473	.4289
	2	1303	-.1591	.4491
	3	1338	-8.1353E-03	.4446
	4	1345	7.823E-02	.4415
# Correct	1	1171	1.7011	1.2511
	2	1303	2.8849	1.3101
	3	1338	3.4738	1.1380
	4	1345	3.5584	1.1239

Discussion

Results suggest that ability estimates are affected by response time patterns (or that time use pattern is affected by ability). Examinees who take longer to respond initially and hurry to respond as the test time winds down have significantly lower ability estimates than examinees who distribute item response time evenly across the test time. Departure from an evenly distributed use of time seems to diminish ability estimates, as evidenced by differences in ability estimates among the 4 response time patterns.

After taking into account several variables that logically could mediate the effect of time use on ability estimate (item difficulty, word count and language fluency of the examinee) time use pattern accounts for about 5% of the variance in ability estimate. While this represents a small effect, it is a substantial increase in explained variance when compared to word count and language fluency. Of course, item difficulty explains most of the variance in examinee ability estimate.

A revealing result is that as the examinees approached the end of the test, those in time use pattern 1 answered about 25% of the items correctly, a rate close to chance. This result was observed even though item difficulty for the final items on the test were lower than for the other time use pattern groups. This suggests that these examinees, cognizant of time, were guessing. Examinees in time use pattern 3 and 4, on the other hand, answered end of test items at a much higher than chance rate, even though these items were more difficult.

Based on these preliminary findings, it appears that time use has some effect on test scores in a computer adaptive, IRT driven test environment. Whether examinees might benefit from instruction on time use during high stakes testing is open to question. It is likely that if all examinees maximized time use, ability estimates for all examinees would remain relatively constant. However, the results do suggest that time limitations affect examinee behavior, especially near the end of the test.

Further analysis

These results are preliminary, and require further analysis. For example, it might be that true ability determines time use patterns: Examinees with higher ability might require less time to answer more difficult items, and examinees with lower ability take longer to answer, leaving less time to devote to each item at the end of the test.

The patterns found in the test of verbal ability may or may not apply in tests of other domains, such as quantitative ability. An analysis of quantitative items is in the works.

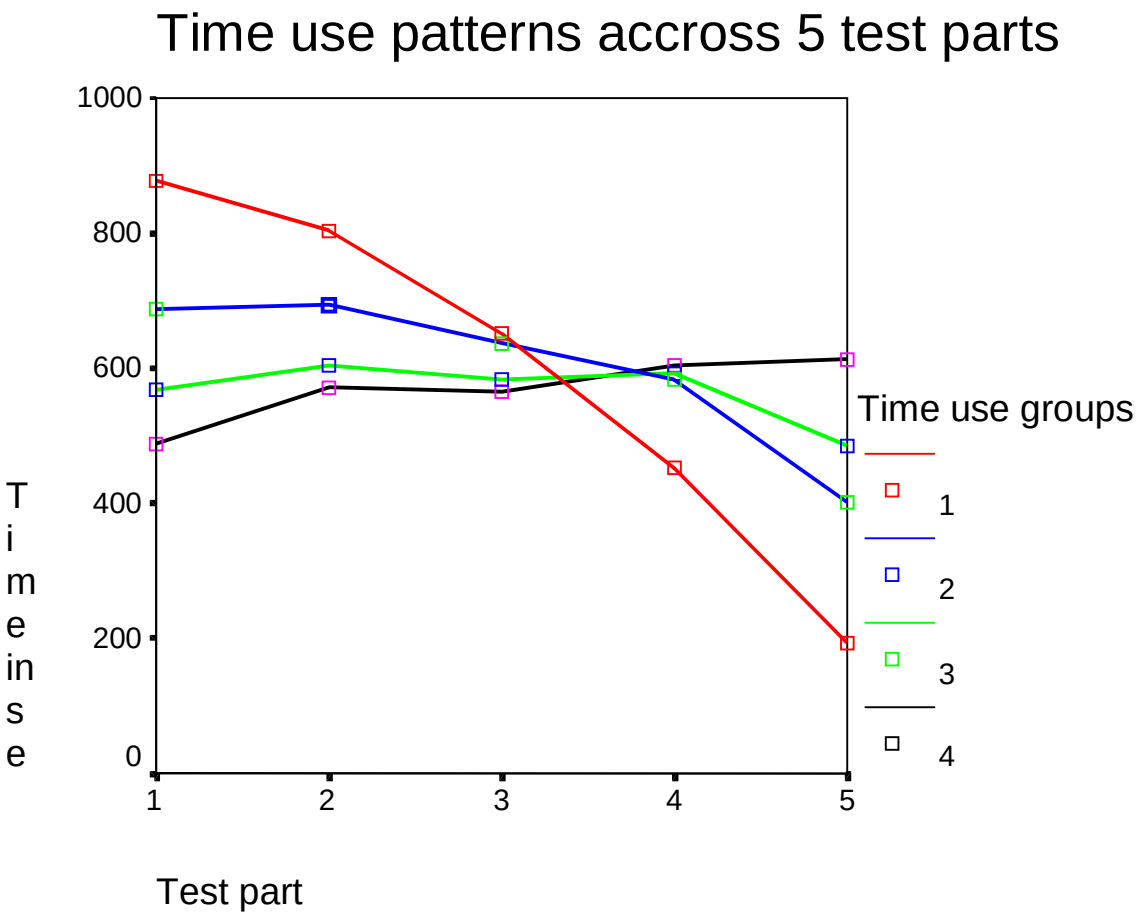
References

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Narayanna, P., Durso, R., and Roussos, L. (2000). *Impact of time-usage strategies on examinee performance: An investigation using examinee item response time data.* Paper presented at the annual meeting of NCME: New Orleans, LA.

Schnipke, D. L., & Scrams (1998, September). *Exploring Issues of Examinee Behavior: Insights Gained from Response-Time Analyses.* Paper presented at the ETS Colloquium, Computer-Based Testing: Building for the Future Assessments, Philadelphia, PA.

Figure 1.


Time use patterns accross 5 test parts
1000
800
600
400
200
0
Time in se
1
2
3
4
5
Test part
Time use groups
1
2
3
4