Wang, C., & Lu, H. (2018). Mediating Effects of Individuals' Ability Levels on the Relationship of Reflective-Impulsive Cognitive Style and Item Response Time in CAT. *Educational Technology & Society, 21* (4), 89–99.

Mediating Effects of Individuals' Ability Levels on the Relationship of Reflective-Impulsive Cognitive Style and Item Response Time in CAT

Chao Wang and Hong Lu*

Department of Educational Technology, Shandong Normal University, China // wxcstudent@126.com // sdnulh@163.com

*Corresponding author

(Submitted December 9, 2016; Revised April 13, 2017; Accepted April 17, 2017)

ABSTRACT

This study focused on the effect of examinees' ability levels on the relationship between Reflective-Impulsive (RI) cognitive style and item response time in computerized adaptive testing (CAT). The total of 56 students majoring in Educational Technology from Shandong Normal University participated in this study, and their RI cognitive styles were diagnosed using the Matching Familiar Figures Test-20 (MFFT-20). Examinees' ability values and average item response time were recorded by the computerized adaptive testing system. Then mediation analysis was implemented and the findings revealed that there was direct and indirect effect between RI cognitive style and item response time in CAT. What's more, RI cognitive style also directly affected the ability levels, and then the ability levels impacted on item response time. So, examinee's ability level was a partly mediator between RI cognitive style and item response time. Furthermore, RI cognitive style of the examinees might also be diagnosed according to ability values and average item response time. The relevant research and implications were further discussed.

Keywords

Computerized adaptive testing, Reflective-Impulsive cognitive style, Ability level, Item response time, Mediation

Introduction

Compared with the traditional linear tests, each examinee in computerized adaptive testing (CAT) takes a unique test that is dynamically adjusted to his/her performance level. After each response in CAT, the ability estimate is updated and the next item is selected such that it has optimal properties according to the new estimate (van der Linden & Glas, 2003). So the item administration would be neither too hard nor too easy for each examinee to keep them appropriately challenged and more likely to stay engaged (Green, 1983; Wainer, 2000). Additionally, by selecting items of particular relevance to an individual response, CAT applications can concurrently reduce the number of questions, increase measurement precision, decrease response burden, and cover a wide measurement range (Devine et al., 2016). It also takes less time to complete the gain of the accurate results by using more efficient and precise assessments. Because of the above mentioned advantages, CAT is becoming more common in high-stake assessment. For example, in the United States, several large tests have an operational CAT version, e.g., the Graduate Record Examination (GRE) and the Computerized Placement test. Several licensure boards have also implemented CAT versions of their tests, including the National Council of State Boards of Nursing and National Board of Medical Examiners. In addition, the US Department of Defense has implemented a CAT version of the Armed Services Vocational Aptitude Battery. CAT is also becoming increasingly popular outside the US. For example, in the Netherlands, the National Institute for Educational Measurement has recently released two CATs, one for assigning examinees to different levels of a mathematics course and the other for assessing achievement in a specific mathematics course (Meijer & Nering, 1999; Verschoor & Straetmans, 2000).

Review of literature

CAT has multiple theoretical advantages over standard static assessments and is becoming an evolutionary step to future testing methodologies, therefore, there are an increasing number of researchers dedicating to it. Most of the researches on CAT were focused on the technical aspects, such as the construction of item pool (Ariel, Veldkamp, & van der Linden, 2004; Barla et al., 2010; Reckase, 2010), the judgment of initial conditions (Kozierkiewicz-Hetmańska & Nguyen, 2010; Mansoor, 2007; Wauters, Desmet, & Van den Noortgate, 2010), the strategies of item selection which included content balancing (Chen & Ankenmann, 2004; Cheng, Patton, & Shao, 2014; Su, 2016; Yi & Chang, 2003) and item exposure control (Leroux, Lopez, Hembry, & Dodd, 2013; Leung, Chang, & Hau, 2005), the estimate of ability (Balas-Timar & Balas, 2009; Huang, Lin, & Cheng, 2009) and the terminal condition of the test (Kozierkiewicz-Hetmańska & Nguyen, 2010; Mansoor, 2007; Triantafillou, Georgiadou, & Economides, 2008). Besides the technical aspects, the psychological reactions of examinees to

CAT also had been investigated and described in detail. For example, Kim and McLean (1994) declared that test anxiety was generally found to be negatively related to test performance on the CAT. And the following researches paid attention to the question whether or not CAT would be more anxiety-producing for examinees than conventional testing and whether or not CAT could produce a disadvantage for examinees with higher anxiety. Fritts and Marszalek (2010) noted that traditional paper-and-pencil tests (P&P) examinees had higher anxiety than CAT examinees when controlling for trait test anxiety and computer anxiety. Some researchers also paid attention to the potential of CAT to activate examinees' motivations (Lu, Hu, Gao, & Kinshuk, 2016; Ortner, Weißkopf, & Koch, 2014; Rheinberg, Vollmeyer, & Burns, 2001; Tonidandel, Quiñones, & Adams, 2002). Compared with large number of researches which focused mainly on the technical aspect and psychological effects of CAT, a little attention was paid to item response time of different examinees in CAT.

CAT provides not only examinees' ability values but also item response time. Typically, in the test, different examinees have different item response time in different items. It is therefore particularly interesting to investigate if there are differences of item response time between different conditions and what additional meaning can be gained from item response time. The evidences that incorrect answers took more time than correct were often reported. For example, Hornke (2000) tested 5912 young men with a computerized adaptive test and showed that item response time for wrong solutions was noticeably longer than for correct solutions. The reason of this difference was also be expounded by Hornke (2005). He indicated that the correct answer seemed to "catch the eye" of the examinee, while wrong answers were rather the result of a longer process. Perhaps item details were repeatedly considered, then discarded, and finally forgotten. The effort to find the answer dragged on, and in the end it might be terminated by a random guess. Chang, Plake and Ferdous (2005) investigated time demands on a time restricted, fixed-length CAT and found that across item blocks, high ability examinees spent more time on test questions than did low ability examinees. Beside this difference between higher and low ability examinees, another difference was that high ability examinees spent more time averaged across items they answered incorrectly than ones they answered correctly, but low ability examinees systematically spent nearly equivalent time on items that they answered correctly and incorrectly. Some studies demonstrated that item response time, on average, increased with the item difficulty (Bergstrom, Gershon, & Lunz, 1994; Bridgeman & Cline, 2000). The adaptive algorithm often provided more difficult items for examinee who had high level of competency. So, high ability examinees received more time-consuming items. However, Chang, Plake and Ferdous (2005) employed pretest items which were not tailored to the examinees' ability level to verify that viewpoint and found that regardless of whether the examinees answered the items correctly or incorrectly, the fact that high ability people spent more time on all items was not necessarily related to the difficulty of the item. And they suggested that high ability examinees might have a higher persistence with test questions.

Item response time is considered the most important measure used to investigate the hypotheses about mental processing in cognitive psychology (Eysenck & Keane, 2010). In CAT, examinees are administered test questions that are matched to their ability levels, so this item selection algorithm would create a challenging and optimally motivating assessment situation in which examinees feel neither over- nor under-challenged. Therefore, it's reasonable to assume that in CAT, the difference of item response time among the examinees should also be reflected by some elements of cognitive psychology, such as the Reflective-Impulsive (RI) cognitive style, which is one of the most frequently studied cognitive styles and is related with the speed of solving problems. Hence, the present study attempts to verify the direct and indirect effect of RI cognitive style on item response time.

Theoretical foundations

Everyone has different approaches and propensities of how to learn new things. Some people prefer to learn from theoretical information while others rather tend to learn from practical experience. The above-mentioned difference is embodied in the implication of cognitive style in the field of psychology. Cognitive style refers to individual differences when perceiving, attending, remembering, deciding, and solving problems (Quiroga, Martínez-Molina, Lozano, & Santacreu, 2011). RI cognitive style, first proposed by Kagan, Rosman, Day, Albert and Phillips (1964) using the Matching Familiar Figures Test (MFFT), is one of the most constantly researched cognitive styles. When solving problems, there are many situations where students have to make decisions under great uncertainty. RI cognitive style is defined as a property of the cognitive system that refers to individual differences of response speed and accuracy in the situations of individuals' information-processing, hypothesis-constructing, decision-making as well as problem-solving with response uncertainty (Kagan & Kogan, 1970).

Reflective individuals employ an analytic processing mode, while impulsive people apply a holistic processing mode (Ancillotti, 1984, 1985; Quiroga, Martínez-Molina, Lozano, & Santacreu, 2011). Reflection means a delay in response latency until being sure about the correct choice. Accordingly, reflective individuals show higher accuracy but higher response latencies. On the contrary, impulsivity implies a quicker choice of a response alternative. Impulsive individuals tend to accept the first hypothesis that emerges in their minds without checking its accuracy, therefore, they show lower response latencies but greater inaccuracy (Quiroga, Martínez-Molina, Lozano, & Santacreu, 2011).

The present study

A short summary that impulsive individuals tended to response much faster than reflective ones but make more mistakes in their decisions could be got from theoretical foundations section. Accordingly, Hypothesis 1 and Hypothesis 2 were as following.

H1. The difference of item response time between reflective and impulsive individuals was significant in CAT.

H2. The difference of ability levels between reflective and impulsive individuals was significant in CAT.

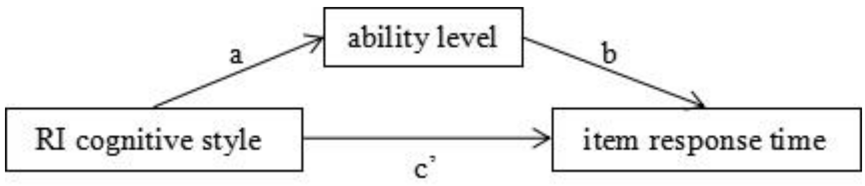
From the section of research review, a conclusion that high ability examinees spent more time on the items than did low ability examinees in CAT could be drawn. In other words, the examinees' abilities would influence their item response time in CAT. Therefore, this study constructed Hypothesis 3.

H3. The difference of item response time between different ability individuals was significant in CAT.

According to the above three assumptions, the fourth hypothesis was presented as following.

H4. The ability of the examinee had a mediating effect between RI cognitive style and item response time in CAT.

The model of the mediating effect in this present study was shown in Figure 1.



a: The regression coefficient was obtained by regressing the ability values on RI cognitive style.
b: The regression coefficient between item response time and ability levels was obtained by regressing the average item response time on RI cognitive style and the ability values.
c': The regression coefficient between item response time and RI cognitive style was obtained by regressing the average item response time on RI cognitive style and the ability values.

Figure 1. Ability level as mediator of RI cognitive style and item response time

Methodology

Participants and procedure

The procedure of the present study was divided into three phases. The first phase was to develop a computerized adaptive testing system which could record the ability values as well as the average item response time of each examinee. The ability value was a good indicator of examinee's ability level and the average item response time was used as the metric of item response time. In order to create the item bank of the adaptive testing, five professors developed 180 multiple choice items in respect of the probability theory and mathematical statistics. The 180 items were divided into four parallel tests. Each test was composed of 40 independent items and 20 anchor items. During the course of confirming the item parameters based on item response theory (IRT), 3524 students from four Chinese universities in Shandong province were recruited to take the pretest. Of course, the students had taken the probability theory and mathematical statistics as compulsory course. In each university, the participants were randomly separated into four groups and each group was assigned one test. On the basis of the students' responses, the parameters of each item were calculated by employing the Bayesian Expected A

Posterior method with BILOGMG 3.0. The consequences turned out that 22 items failed to fit the two parameter logistic model, which was one of the most widely used IRT models, and were deleted in terms of the value of chi-square value and the degree of freedom. The parameters of the remaining 158 items were connected on the same scale by the mean and sigma method. Eventually, 158 items that varied according to their difficulty and discrimination constituted the item bank of the adaptive testing system.

The second phase was measuring RI cognitive style of 73 sophomore students who majored in Educational Technology in Shandong Normal University, and had taken the course of probability theory and mathematical statistics. The average age of the students was 19.09, and *SD* was 0.90. RI cognitive style was measured in the students using the Matching Familiar Figures Test-20 (MFFT-20; Cairns & Cammock, 1978) presented on the computer screen in this study. Before the test, a notebook computer installed with MFFT-20 was got ready. During the process of the test, each student participated in the test separately under the supervision of the experimenter. Specifically, after one examinee finished the test and went out, next one could enter the laboratory and start the test. The reason why the examinees took the test separately was to ensure the testing process of each examinee without other students' influence and lead to more accurate data. The purposes of this phase was to choose the reflective and impulsive individuals among the 73 students, and the result was that there were 29 reflective students, 27 impulsive students, 8 fast-accurate students and 9 slow-inaccurate students. Ultimately, the results of this phase showed that the examinees who would participate in the third phase were 29 reflective and 27 impulsive students.

The third phase was the implementation of CAT. 56 students (19 male, 37 female) whose cognitive style was either reflection or impulsivity took part in the adaptive testing. CAT was conducted on a notebook computer in a quiet laboratory at Shandong Normal University. The students were tested separately just as the test process of the second phase for more accurate data. Experimental instruction was presented via the screen of the notebook computer. And the instruction explained to the examinees that they were about to take a fixed-length adaptive testing which designed to evaluate their knowledge of probability theory and mathematical statistics, and the test comprised 25 multiple choice items. After the instruction was presented, the examinees had 75 minutes to take the adaptive test, which was sufficient for them to finish the testing. At the end of the testing, the examinees would receive their ability values regarding their actual level of performance on the adaptive testing. In addition, the average item response time would also be recorded by the computerized adaptive testing system and it could only be viewed by the experimenters.

Measurement development

Reflective-Impulsive cognitive style

The examinees were asked to complete the MFFT-20 developed by Cairns and Cammock (1978), and in the present study the test was in the form of stand-alone version accomplished by the researchers. The feasibility of computer controlled MFFT-20 administration was demonstrated by van Merriënboer and Jelsma (1988). Twenty stimulus items were comprised in the MFFT-20, which was suitable for use with children and youths in the age range 7–21 yrs. Each stimulus item consisted of a standard figure and eight alternatives, and it required a visual match of the alternatives to the standard. Only one of the alternatives was identical to the standard. That was, the examinees were required to choose the alternative that was identical to the standard figure. If they chose wrongly, the borders of computer screen would flash red three times and the message would appear on the information window to require the examinee to try again until the right one was clicked. The total number of errors on the test would be collected and the average time to first response would be calculated. A double median-split method was employed to classify individuals according to the average time to first response and the total number of errors (Kagan, 1966a; Kagan, 1966b; Kagan, Pearson, & Welch, 1966). Reflective individuals had an average time to first response score above the median of the sample and a total errors score below the median of sample. In contrast, impulsive individuals had an average time to first response score below the median of the sample and a total errors score above the median of the sample. There were still another two types of individuals, and they were fast-accurate and slow-inaccurate. The former had an average time to first response score below the median of sample and a total errors score below the median of the sample. The latter one had an average time to first response score above the median of the sample and a total errors score above the median of the sample. The average time to first response as well as the total number of errors of each examinee was recorded in the data base. After the whole test was completed, the researchers figured out the median of the average time to first response score and the median of the total errors score of the sample. Then, RI cognitive style was distinguished with the above mentioned double median-split method.

The average item response time in CAT

The average item response time used as the metric of item response time meant the average time from stimulus onset to answer execution by an examinee completing a test item. During the process of adaptive testing, each examinee's response time on each item was recorded in the database table. The system was designed to calculate the average item response time of each examinee and record it in the database as long as the examinee finished the adaptive testing.

The ability values obtained in CAT

The ability value recorded by the system was a good indicator of examinee's ability level. In CAT, on the basis of IRT, the ability values depended on which items were responded correctly rather than on the number of items responded correctly. That was to say, the ability values of the examinees would be greater if they answered more difficult questions correctly compared with easier questions. In this study, the ability value was finally presented on the range of -3 to 3.

Statistical analysis

SPSS version 20.0 (Chicago, IL, USA) was used for analyzing the data. Significance test was done using independent-sample T test to verify the first three hypotheses. The statistical significance was set at $p < .05$. Then, linear regression analysis was performed to test the mediating effect.

Results

The difference of the average item response time between reflective and impulsive examinees

The mean of the average item response time of the examinees was 44.09 seconds, and the *SD* was 15.77. According to the theoretical foundations section, it was known that the reflective and impulsive individuals needed different time to answer the questions. In this study, the mean of the average item response time of the reflective examinees was 50.39 seconds, and the *SD* was 15.04. The mean of the average item response time of the impulsive examinees was 37.32, and the *SD* was 13.81. There was statistically significant difference in the average item response time between reflective and impulsive examinees (t = 3.38, df = 54, p = .010). That was, the average item response time of the reflective examinees were significantly longer than the impulsive examinees'.

The difference of the ability values between reflective and impulsive examinees

The ability values received by the examinees at the end of the adaptive testing represented their ability levels in the probability theory and mathematical statistics. The mean ability value of all examinees was 0.61, and the standard deviation was 0.44. Furthermore, the mean ability value of the sample of the reflection (M = .73, SD = .39) was found to be higher than the mean ability value of the sample of the impulsivity (M = .47, SD = .46). Statistically significant difference in the ability values between reflective and impulsive examinees was detected at this independent sample t test (t = 2.34, df = 54, p = .023). So, the ability values of the reflective examinees were significantly higher than the impulsive examinees'.

The difference of the average item response time between examinees with different ability levels

The third independent sample *t* test was performed to test the difference of the average item response time between examinees with different ability levels. At the end of the experiment, the examinees were divided into two groups according to their ability values. The examinee whose ability value was higher than the mean ability value of the sample was classified as high ability examinee, otherwise he or she would be classified as low ability examinee. In this study, the average time spent by each high ability examinee on each item was 51.84 seconds, and the standard deviation was 14.79. The average time spent by each low ability examinee on each item was 36.34 seconds, and the standard deviation was 12.79. There was statistically significant difference in the average item response time between high ability and low ability examinees (t = 4.19, df = 54, p = .000). In other words, the average item response time of the high ability examinees was significantly longer than the low ability examinees'.

The procedure and result of mediation analysis

According to the above three results that proved H1, H2 and H3 to be true, it was meaningful to test the H4. So, whether ability level of examinee acted as a mediating variable in the relation of RI cognitive style and item response time was tested. Following the approach suggested by Wen, Chang, Hau and Liu (2004), after centering of the massive measured data, three regression analyses were conducted. The findings from the three regression analyses were summarized in Table 1, Table 2 and Table 3. From Table 1, it would be found that the first linear regression model about the average item response time on RI cognitive style (Reflection = 1, Impulsivity = 0) accounted for 17.5% (R^2 = .175) of the variance. And the second linear regression model about the ability values on RI cognitive style account for 9.2% of the variance. In the third multiple linear regression model, RI cognitive style and the ability values explained 39.8% (R^2 = .398) of the average item response time. In Table 2, the value of *F* was the mean square regression divided by the mean square residual. The probability of the *F*-values in the three regression models showed that the probability of the given correlation occurring by chance was less than 0.05. It meant that all the three linear regression equations were significant. In Table 3, the values of *B* were the constant and the coefficients of the linear regression equation. Beta was the *B*-value for standardized scores of the independent variables. From Table 3, it would be found that all the regression coefficients were statistically significant.

Table 1. Results of regression analysis (Model summary)

Model	*R*	R^2	Adjusted R^2	*SE* of the estimate
1[A]	.418	.175	.159	14.4597
2[B]	.303	.092	.075	.4240
3[C]	.631	.398	.375	12.4696

Note. [A] = Predictors: (Constant), RI cognitive style. Dependent Variable: the average item response time. [B] = Predictors: (Constant), RI cognitive style. Dependent Variable: ability vales. [C] = Predictors: (Constant), RI cognitive style, ability values. Dependent Variable: the average item response time.

Table 2. Results of regression analysis (*ANOVA*)

Model		Sum of squares	*df*	Mean square	*F*	*Sig.*
1[A]	Regression	2390.426	1	2390.426	11.433	.001
	Residual	11290.505	54	209.083		
	Total	13680.931	55			
2[B]	Regression	.983	1	.983	5.467	.023
	Residual	9.708	54	.180		
	Total	10.690	55			
3[C]	Regression	5439.912	2	2719.956	17.493	.000
	Residual	8241.019	53	155.491		
	Total	13680.931	55			

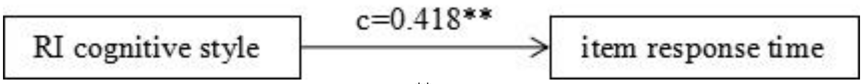
Note. [A] = Predictors: (Constant), RI cognitive style. Dependent Variable: the average item response time. [B] = Predictors: (Constant), RI cognitive style. Dependent Variable: ability values. [C] = Predictors: (Constant), RI cognitive style, ability values. Dependent Variable: the average item response time.

Table 3. Results of regression analysis (Coefficients)

Model		Unstandardized Coefficients		Standardized Coefficients	*t*	*Sig.* (2-tailed)
		B	*SE*	Beta		
1[A]	(Constant)	.031	1.932		.016	.987
	RI cognitive style	13.075	3.867	.418	3.381	.001
2[B]	(Constant)	.001	.057		.011	.991
	RI cognitive style	.265	.113	.303	2.338	.023
3[C]	(Constant)	.020	1.666		.012	.991
	RI cognitive style	8.376	3.499	.268	2.394	.020
	Ability values	17.724	4.002	.495	4.429	.000

Note. [A] = Predictors: (Constant), RI cognitive style. Dependent Variable: the average item response time. [B] = Predictors: (Constant), RI cognitive style. Dependent Variable: ability values. [C] = Predictors: (Constant), RI cognitive style, ability values. Dependent Variable: the average item response time.

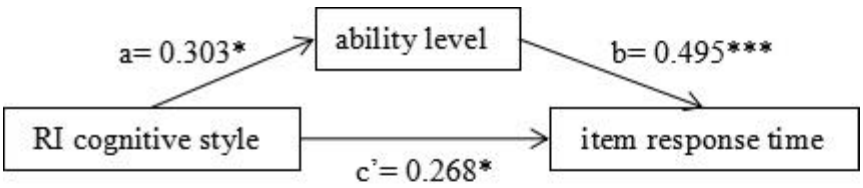
To be specific, when the average item response time was regressed on RI cognitive style without any mediator (path c in Figure 2), the coefficient was found to be significant (c = .418, p = .001). This result confirmed that there was a significant total effect of RI cognitive style on item response time. In the second regression, the ability values were regressed on RI cognitive style (path a in Figure 3), and the regression coefficient also arrived significance level (a = .303, p = .023), which revealed that RI cognitive style was significantly positively related to examinee's ability level. Finally, multiple regression analysis about the average item response time on RI cognitive style and the ability values (path b and c' in Figure 3) was done and both path coefficients were found to be significant (b = .495, p = .000; c' = .268, p = .020). From the third regression, it could be found that the relationship of RI cognitive style and item response time was still significant when the effect of ability level was also taken into account, and it could also be found that ability level was significantly positively related to item response time when controlling for RI cognitive style. According to Wen, Chang, Hau and Liu (2004), the results of this mediation analysis implied that ability level of the examinee was a partly mediator in the relationship between RI cognitive style and item response time. The resulting model was shown in Figure 2 and Figure 3.



Note. $^{**}p < .01$.

c = The regression was obtained by regressing the average item response time on RI cognitive style.

Figure 2. The relation of RI cognitive style and item response time without mediator



Note. $^{*}p < .05$; $^{***}p < .001$.

a: The regression coefficient was obtained by regressing the ability values on RI cognitive style.

b: The regression coefficient between item response time and ability levels was obtained by regressing the average item response time on RI cognitive style and the ability values.

c': The regression coefficient between item response time and RI cognitive style was obtained by regressing the average item response time on RI cognitive style and the ability values.

Figure 3. Mediation model showing the effect of RI cognitive style on item response time through ability level

Discussion

From the results, it would be found that the RI cognitive style of examinees affected their item response time and ability levels in CAT. Reflective examinees spent more time averagely on each item but showed higher accuracy than impulsive examinees. That was, the first two hypotheses, which stated that the difference of item response time and the difference of ability levels between reflective and impulsive individuals would all be significant in CAT, were supported. The results could be explained by the implication of the RI cognitive style. What needed further explanation was that the experiment in this study was carried out with computerized adaptive testing mode. That was, the present study provided examinees with the test environment where the item administration was dynamically adjusted to the estimate of the examinee's ability level. However, the previous studies (Haghighi, Ghanavati, & Rahimi, 2015; Spinella & Miley, 2003) were used to apply conventional linear tests with fixed item sequence, in which the examinees were forced to work on items that were either too hard or too easy for them, to detect the differences of mental characteristics between reflective and impulsive individuals. The accuracy of the results of testing the differences between reflective and impulsive individuals might be influenced by the fact that the difficulty of item was not adjusted to the examinee's ability level in traditional linear test. In contrast to the studies with fixed item sequence testing mentioned above, the present study with computerized adaptive testing would be more scientific and persuasive.

The third hypothesis described the influence of examinees' ability levels on item response time in CAT. This study did verify that the ability levels of the examinees influenced their response time. And this agreed with the previous researches which argued that the high ability individuals spent significantly longer time on solving problems than did low ability individuals (Chang, Plake, & Ferdous 2005; Hornke, 2000, 2005). In the interpretation of the results, Chang, Plake and Ferdous (2005) suggested that more able individuals were more likely to persevere on test questions. However, the interpretations were just speculations which needed to be validated in future research.

To test the fourth hypothesis, the current study examined the mediating effects of ability levels on the relationship between RI cognitive style and item response time in CAT. The result of the mediation analysis indicated that RI cognitive style directly affected ability levels of the examinees, and that, ability levels then impacted on item response time. Therefore, ability level served as a mediator in the relationship between RI cognitive style and item response time in CAT, which was supportive for the fourth hypothesis. The result could be interpreted by the theoretical foundations part of this paper. It was clear that the reflective individuals used an analytic processing mode to solve problems, whereas the impulsive individuals used a holistic processing mode (Ancillotti, 1984; Ancillotti, 1985). Different methods of processing information not only made the individuals different in the response time on solving problems, but also made the individuals different in the accuracy of solving problems. Additionally, early studies had verified that individuals' ability levels influenced the response time on solving problems (Gvozdenko & Chambers, 2006; Schnipke & Scrams, 1999). So, ability levels might be the mediator between RI cognitive style and response time when individuals solved problems. What's more, the results also demonstrated that the effect of RI cognitive style upon item response time included both direct effect and indirect effect in CAT. And the ratio of indirect effect to direct effect was 56.0%, which could revealed that the direct effect of RI cognitive style on item response time was greater than the indirect effect.

Conclusion

This study provided novel preliminary evidence that RI cognitive style not only had direct relationship with item response time, but also affected item response time through influencing ability level of the examinee in CAT. More precisely, ability level played a partial mediation effect between RI cognitive style and item response time.

Previous studies had discussed how RI cognitive style affected response time when all the individuals faced one and the same task. However, this study investigated the effect of RI cognitive style on item response time while all the individuals resolved the test questions of which the difficulties were match to the individuals' ability levels. Obviously, the conclusion of this study could enrich and advance the theory of RI cognitive style.

The specialists in the field of educational measurement and evaluation had been expecting the issue whether more information of the examinees could be obtained from the test results, especially their psychological traits. From the results of the present study, besides the examinees' ability levels, the RI cognitive style of the examinees might also be diagnosed according to the ability values and average item response time recorded by the computerized adaptive testing system. On the basis of the foregoing, during the process of CAT it was reasonable to remind the examinees who belonged to impulsivity to make their decisions after careful considerations and remind the reflective examinees to pay attention to the item response time. It can not only help impulsive individuals to overcome the limitations of holistic information processing method to show their real ability levels, but also reduce the average item response time of the reflective examinees to some extent, which could improve the test economy. So, an interactive and cognitive style-friendly test environment could be established in CAT. The above discussion demonstrated that the potential of the vast information generated during a CAT could be designed to increase the precision of evaluation and to extend the utility of a test.

According to item response time of the examinees, some mathematical models or judgment criteria had been built to divided examinees' response behaviors into two sections, "rapid-guessing behavior" and "answer-giving behavior" (Chang, Plake, Kramer, & Lien, 2011; van der Linden, 2006). The results of this paper showed that the effects of RI cognitive style and ability levels on item response time were both significant. So, if the differences of RI cognitive style and ability levels could be taken into account to build different mathematical models or judgment criteria according item response time, it would be accurate and effective to distinguish "rapid-guessing behavior" from "answer-giving behavior."

Limitation and future study

One limitation of the present study lied in sample bias. The number of participants in this study was small and their ages were mostly between 19 and 21 years. In addition, female in this sample significantly outnumbered male. These imbalances of sample might lead to universal conclusions of the study subject to certain restrictions. In order to make the conclusions of similar study be more persuasive, the size of sample should be as large as possible, and the distribution of age and gender should be more balanced in the future.

Item bank of CAT in this study was all about the probability theory and mathematical statistics. The single type of testing item wouldn't be helpful for the universality of results. Thus in future study, another item types of CAT in experiment could be taken into consideration such as processes associated with verbal and numerical reasoning as well as perceptual discrimination tests and so on.

The success probability was set to 0.5 for the maximum-information item selection rule used in this study. However, earlier researches indicated that examinees might feel challenged by a constant success probability of $p = 0.5$ and therefore could not come to a sufficiently high answer certainly within a reasonable timeframe (Häusler, 2006; Häusler & Sommer, 2008). So, the chosen items in this study might be hard for examinees, which could influence the veracity of item response time. Häusler (2006) also found that within reasonable limits there was very little loss of information when the success probability deviated from $p = 0.5$. As a result, in the future study it was justifiable to set the success probability to 0.6 or 0.7, or use mixtures of highly informative ($p = 0.5$) and easier item ($p = 0.6$ or $p = 0.7$) in the item selection process.

Acknowledgments

This work has been funded by the Chinese National Education Science Foundation (BBA170069).

Reference

Ancillotti, J. P. (1984). Rythme conceptuel et apprentissage cognitif [Conceptual tempo and cognitive learning]. *Psychologie Française, 29*, 38-38.

Ancillotti, J. P. (1985). Dimension reflexive-impulsive de la personnalité et fonctionnement cognitif [Reflective–impulsivity style and cognitive functioning]. *Cahiers de Psychologie Cognitive, 5*, 667–686.

Ariel, A., Veldkamp, B. P., & van der Linden, W. J. (2004). Constructing rotating item pools for constrained adaptive testing. *Journal of Educational Measurement, 41*(4), 345–359.

Balas-Timar, D. V., & Balas, V. E. (2009, October). *Ability estimation in CAT with fuzzy logic*. Paper presented at the 4th International Symposium on Computational Intelligence and Intelligent Informatics, Luxor, Egypt. doi:10.1109/ISCIII.2009.5342278

Barla, M., Bieliková, M., Ezzeddinne, A. B., Kramár, T., Šimko, M., & Vozár, O. (2010). On the impact of adaptive test question selection for learning efficiency. *Computers & Education, 55*(2), 846–857.

Bergstrom, B., Gershon, R. C., & Lunz, M. E. (1994, April). *Computerized adaptive testing exploring examinee response time using hierarchical linear modeling.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Bridgeman, B., & Cline, F. (2000). Variations in mean response times for questions on the computer-adaptive GRE® General test: Implications for fair assessment. *ETS Research Report Series, 2000*(1), i–29. doi:10.1002/j.2333-8504.2000.tb01830.x

Cairns, E. D., & Cammock, T. (1978). Development of a more reliable version of the Matching Familiar Figures Test. *Development Psychology, 14*(5), 555–560. doi:10.1037/0012-1649.14.5.555

Chang, S. R., Plake, B. S., & Ferdous, A. A. (2005, April). *Response times for correct and incorrect item responses on computerized adaptive tests.* Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Montreal, Quebec, Canada.

Chang, S. R., Plake, B. S., Kramer, G. A., & Lien, S. M. (2011). Development and application of detection indices for measuring guessing behaviors and test taking effort in computerized adaptive testing. *Educational and Psychological Measurement, 71*(3), 437–459. doi:10.1177/0013164410385110

Chen, S. Y., & Ankenmann, R. D. (2004). Effects of practical constraints of item selection rules at the early stages of computerized adaptive testing. *Journal of Educational Measurement, 41*(2), 149–174. doi:10.1111/j.1745-3984.2004.tb01112.x

Cheng, Y., Patton, J. M., & Shao, C. (2014). a-Stratified Computerized Adaptive Testing in the presence of calibration error. *Educational and Psychological Measurement, 75*(2), 260–283. doi:10.1177/0013164414530719

Devine, J., Fliege, H., Kocalevent, R., Mierke, A., Klapp, B. F., & Rose, M. (2016). Evaluation of Computerized Adaptive Tests (CATs) for longitudinal monitoring of depression, anxiety, and stress reactions. *Journal of Affective Disorders, 190*, 846–853. doi:10.1016/j.jad.2014.10.063

Eysenck, M. W., & Keane, M. T. (2010). *Cognitive psychology: A Student's handbook* (5th ed.). Hove, UK: Psychology Press.

Fritts, B. E., & Marszalek, J. M. (2010). Computerized adaptive testing, anxiety levels, and gender differences. *Social Psychology of Education, 13*(3), 441–458. doi:10.1007/s11218-010-9113-3

Green, B. F. (1983). The Promise of tailored tests. In H. Wainer, & S. Messik (Eds.), *Principals of modern psychological measurement* (pp. 69–80). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gvozdenko, E., & Chambers, D. P. (2006, June). Insights into learners' understanding through monitoring test takers' temporal behaviors. In *Proceedings of the International Federation of Information Processing conference: Imagining the future for ICT and Education conference* (pp. 142–151). Ålesund, Norway: Hogskolen i Alesund/Alesund University College.

Haghighi, M., Ghanavati, M., & Rahimi, A. (2015). The Role of gender differences in the cognitive style of impulsivity/reflectivity and EFL success. *Procedia-Social and Behavioral Sciences, 192*, 467–474. doi:10.1016/j.sbspro.2015.06.072

Häusler, J. (2006). Adaptive success control in computerized-adaptive testing. *Psychology Science, 48*(4), 436–450.

Häusler, J., & Sommer, M. (2008). The Effect of success probability on test economy and self-confidence in computerized adaptive tests. *Psychology Science*, *50*(1), 75–87.

Hornke, L. F. (2000). Item response times in computerized adaptive testing. *Psicológica, 21*(1-2), 175–189.

Hornke, L. F. (2005). Response time in computer-aided testing: A "Verbal Memory" test for routes and maps. *Psychology Science, 47*(2), 280–293.

Huang, Y. M., Lin, Y. T., & Cheng, S. C. (2009). An Adaptive testing system for supporting versatile educational assessment. *Computers & Education, 52*(1), 53–67. doi:10.1016/j.compedu.2008.06.007

Kagan, J. (1966a). Developmental studies in reflection and analysis. In A. H. Kidd, & J. L. Rivoire (Eds.), *Perceptual Development in Children* (pp. 487–522). New York, NY: International University Press.

Kagan, J. (1966b). Reflection-impulsivity: The Generality and dynamics of conceptual tempo. *Journal of Abnormal Psychology, 17*(1), 17–24. doi:10.1037/h0022886

Kagan, J., & Kogan, N. (1970). Individual variation in cognitive processes. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (pp. 1273–1365). New York, NY: Wiley.

Kagan, J., Pearson, L., & Welch, L. (1966). Conceptual impulsivity and inductive reasoning. *Child Development, 37*(3), 583–594. doi:10.2307/1126680

Kagan, J., Rosman, B. L., Day, D., Albert, J., & Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs: General and Applied, 78*(1), 1–37. doi:10.1037/h0093830

Kim, J. G., & McLean, J. E. (1994, November). *The Relationships between individual difference variables and test performance in computerized adaptive testing.* Paper presented at the Annual meeting of the Mid-South Educational Research Association, Nashville, TN.

Kozierkiewicz-Hetmańska, A., & Nguyen, N.T. (2010). A Computer adaptive testing method for intelligent tutoring systems. In R. Setchi, I. Jordanov, R. J. Howlett, & L.C. Jain (Eds.), *Knowledge-Based and Intelligent Information and Engineering Systems. Lecture Notes in Computer Science 6276* (pp. 218–289). Heidelberg, German: Springer, Verlag.

Leroux, A. J., Lopez, M., Hembry, I., & Dodd, B. G. (2013). A Comparison of exposure control procedures in CATs using the 3PL model. *Educational and Psychological Measurement, 73*(5), 857–874. doi:10.1177/0013164413486802

Leung, C. K., Chang, H. H., & Hau, K. T. (2005). Computerized adaptive testing: A Mixture item selection approach for constrained situations. *British Journal of Mathematical & Statistical Psychology, 58*(2), 239–257. doi:10.1348/000711005X62945

Lu, H., Hu, Y. P., Gao, J. J., & Kinshuk. (2016). The Effects of computer self-efficacy, training satisfaction and test anxiety on attitude and performance in computerized adaptive testing. *Computers & Education, 100*, 45–55. doi:10.1016/j.compedu.2016.04.012

Mansoor, A. A. (2007). Implementation of an improved adaptive testing theory. *Educational Technology & Society, 10*(4), 80–94.

Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement, 23*(3), 187–194. doi:10.1177/01466219922031310

Ortner, T. M., Weißkopf, E., & Koch, T. (2014). I will probably fail: Higher ability students' motivational experiences during adaptive achievement testing. *European Journal of Psychological Assessment, 30*(1), 48–56.

Quiroga, M. Á., Martínez-Molina, A., Lozano, J. H., & Santacreu, J. (2011). Reflection-impulsivity assessed through performance differences in a computerized spatial task. *Journal of Individual Differences, 32*(2), 85–93. doi:10.1027/1614-0001/a000038

Reckase, M. D. (2010). Designing item pools to optimize the functioning of a computerized adaptive test. *Psychological Test and Assessment Modeling, 52*(2), 127–141.

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen [QCM: A questionnaire to assess motivation in learning situations]. *Diagnostica, 47*(2), 57–66. doi:10.1026//0012-1924.47.2.57

Schnipke, D. L., & Scrams, D. J. (1999). Exploring issues of test taker behavior: Insights gained from response time analyses. *Law School Admission Council Computerized Testing Report. LSAC Research Report Series.* Princeton, NJ: Law School Admission Council. Retrieved from http://files.eric.ed.gov/fulltext/ED467811.pdf

Spinella, M., & Miley, W. M. (2003). Impulsivity and academic achievement in college students. *College Student Journal. 37*(4), 545–549.

Su, Y. H. (2016). A Comparison of constrained item selection methods in multidimensional computerized adaptive testing. *Applied Psychological Measurement, 40*(5), 346–360. doi:10.1177/0146621616639305

Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The Impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology, 87*(2), 320–332. doi:10.1037/0021-9010.87.2.320

Triantafillou, E., Georgiadou, E., & Economides, A. (2008). The Design and evaluation of a computerized adaptive test on mobile devices. *Computers & Education, 50*(4), 1319–1330. doi:10.1016/j.compedu.2006.12.005

van der Linden, W. J. (2006). A Lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31*(2), 181–204. doi:10.3102/10769986031002181

van der Linden, W. J., & Glas, C. A. W. (2003). *Computerized adaptive testing: Theory and practice.* Boston, MA: Kluwer Academic Publishers.

van Merriënboer, J. J. G., & Jelsma, O. (1988). The Matching familiar figures test: Computer or experimenter controlled administration? *Educational and Psychological Measurement, 48*(1), 161–164.

Verschoor, A. J., & Straetmans, G. J. J. M. (2000). MathCAT: A Flexible testing system in mathematics education for adults. In W. J. van der Linden, & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 101–116). Boston, MA: Kluwer Academic Publishers.

Wainer, H. (2000). *Computerized adaptive testing: A Primer* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wauters, K., Desmet, P., & Van den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. *Journal of Computer Assisted Learning, 26*(6), 549–562. doi:10.1111/j.1365-2729.2010.00368.x

Wen, Z. L., Chang, L., Hau, K. T., & Liu, H. Y. (2004). Testing and application of the mediating effects. *Acta Psychologica Sinica, 36*(5), 614–620. Retrieved from http://www.cnki.com.cn/Article/CJFDTotal-XLXB200405016.htm

Yi, Q., & Chang, H. H. (2003). a-Stratified CAT design with content blocking. *British Journal of Mathematical and Statistical Psychology, 56*(2), 359–378. doi:10.1348/000711003770480084