Title	Student's reactions towards computerised adaptive testing
Author(s)	Chew Lee Chin
Source	*ERA Conference, Singapore, 24-26 November 1997*
Organised by	Educational Research Association of Singapore (ERAS)

Student's Reactions Towards Computerised Adaptive Testing

Chew Lee Chin
National Institute of Education
Nanyang Technological University

Introduction

Computerised adaptive testing (CAT) represents a new concept in testing. It has its theoretical foundations in item response theory (IRT); and with advances in computer technology, it has become a viable testing method where computers are used as devices for delivering tailored tests to individual students. That is, the computer sequentially selects items that are geared towards the individual examinee's performance, and this is done on the basis of the examinee's responses to previous items. The adaptive nature of the testing is made possible within an item response framework which provides for both item and examinee parameters to be measured on a common scale. The benefits alluded to CAT are its greater precision in measurement, and increased efficiency as less items are needed in the testing and therefore testing time is reduced.

Presently, there are few implementations of CAT in the field. The United States are taking the lead in most initial CAT programmes such as the large-scale testing of the Graduate Record Examination (Vogel, 1994), and the educational testing at some school or college levels (Legg & Buhr, 1992; Moe & Johnson, 1988, Oslen et al., 1989). Several researchers view CAT as a promising and viable alternative to paper-and-pencil testing (Bunderson et al., 1989, Kim, 1993; Moe & Johnson, 1988, Oslen et al., 1989). But CAT is a new experience to most students. In comparison to a paper-and-pencil testing which all students are accustomed to, CAT uses different interfaces, namely, a computer medium to administer tests instead of a paper medium, and a tailored test of variable length for each individual student instead of a fixed-item fixed-length test for a group of students. This means that any implementation of CAT for school testing will need to consider its suitability as well as usability. Of special interest will be students' acceptability of CAT.

Not surprisingly, accompanying CAT research has focused on the effects of administering tests via computers (Wise & Plake, 1989) especially that concerning students' reactions or attitudes towards computerised testing. Some investigators hypothesise that this psychological factor can affect students' performance on the test. For example, a study by Ward et al. (1989) reported an overall negative attitude of college students towards the use of computers for testing. Although they found no differences in test performance using a computer test or a paper-and-pencil test, they reported a significant difference in anxiety level with those tested by computer having a higher anxiety level. On the other hand, other research studies indicate overwhelmingly positive student attitudes towards computerised testing including CAT (Baghi et al., 1992; Legg & Buhr, 1992; Moe & Johnson, 1989). But the effects of this attitudinal component on test performance are less clear.

The research to this time has probed further to address a concern that computerised testing and CAT in particular may produce differential effects for different groups of students in terms of their reactions to the testing method (Baghi et al., 1992; Legg & Buhr, 1992) or test performance (Kim & Mclean, 1994; Parshall & Kromrey, 1993). These studies have examined a number of student variables ranging from demographic characteristics such as age, gender, ethnicity; to computer-related characteristics such as computer ownership, prior computing experience, frequency of computer use; to psychological characteristics such as computer anxiety, attitude towards the curriculum. Research findings from Legg and Buhr's study (1992) indicate that some differences in students' reactions to CAT exist for variables such as ethnic, gender, age, ability and computer-experience; but these differences do not appear to affect their test performance. All in all, much of this research is influenced by equity concerns. In her research paper entitled *Equity Issues*

in High Stakes Computerised Testing, Sutton (1993) contends two perspectives of equity concerns. One viewpoint questions whether the use of computerised testing will maintain or exaggerate inequities in education; the other questions how new technology can be used to reduce inequities.

This paper, which is based on a CAT implementation, first attempts to provide an evaluation of students' attitudes towards the new testing method. Next, comparison groups formed by eight variables are examined for any differential effects on students' reactions to the CAT administration.

Method

Participants

The participants were 113 secondary students (aged 16+) from a Singapore school who were enrolled for the GCE 'O' level 1995 biology examination. They had just completed the two-year biology course.

Comparison groups were obtained based on eight independent variables to investigate students' attitudes to the CAT administration. This procedure has been used to classify groups in past research (Baghi et al. 1992; Legg & Buhr, 1992). Three of these variables were specified in the research design of the study. These include: CAT treatment (*with and without performance diagnosis*), gender (*male and female*), and ability level (*low, average, and high*). Another five independent variables were based on five aspects of student individual differences. These variables were three student characteristics related to computers namely, home computer *ownership (with and without computers)*, computing experience (*low and high experience levels*), frequency of computer use (*low, average, and high frequencies*), and two student affective variables, one related to attitude towards *computers (least positive, averagely positive, and most positive)*, and the other related to attitude towards science (*least positive, averagely positive, and most positive*).

Instruments

Computerised Adaptive Test

The computerised adaptive test was constructed from a pre-calibrated item bank under the item response model. A modal bayesian item selection procedure was specified and also to obtain estimates of student ability. The testing was terminated when a bayesian posterior variance of 0.2 was reached or when a maximum of 30 items had been administered.

Two versions of the CAT were used: one with performance diagnosis and the other without diagnosis. The performance diagnosis strategy gave immediate feedback to students about their success or failure on each item taken during the adaptive testing. Typically, the computer screen presented the correct or the best answer for an item together with an explanation for the rightness or wrongness of each answer option.

Paper-and-Pencil Test

This fixed- item, fixed-length paper-and-pencil test was constructed from the same IRT-calibrated bank. It consisted of 30 items selected by the computer using a test specification based on item characteristics.

Student Attitudes toward Computer-Adaptive Test Administration (SACATA)

An adapted version of the instrument, entitled *Student Attitudes toward Computer-Adaptive Test Administration (SACATA)* which was developed by Baghi, Ferrara and Gabrys (1992), was used in this study. It consisted of 13 items which tapped on one of the following aspects: (a) method of testing, (b) use of equipment, (c) test anxiety, and (d) difficulty of test questions.

The first twelve items were made on a 4-point Likert type scale ranging from 'strongly agree' to 'strongly disagree'. For positively worded items, a score range of 4 to 1 was assigned. These values were reversed for negatively worded items. The last item was made on a 3-point Likert scale.

Student Information Questionnaire

A student questionnaire was used to collect some information concerning their personal and computer literacy background. This survey questionnaire also included two attitudinal instruments to capture their affective characteristics. The *Computer Attitude Scale (CAS)*, developed by Gressard and Loyd (1987), was used to assess students' attitudes towards computers; and the *Attitude to Science Instrument (ASI)*, developed by Foong & Lam (1988), was used to assess student's attitudes towards science.

Procedure

Using stratified random sampling based on gender and ability, students were assigned to two CAT treatment groups, namely, with performance diagnosis (Group 1) and without performance diagnosis (Group 2). For each group, students were administered both the computer-adaptive and the paper-printed biology test. The order of test administration was also counterbalanced. The computerised testing was conducted using the school's computer facilities. The Examination Subsystem of the MicroCAT software programme (Assessment Systems Corporation, 1994) was used to administer the adaptive tests. Students responded to the survey questionnaire, *SACATA*, on their computer terminals immediately after taking the computer-adaptive test.

Descriptive statistics were employed to yield profiles of students' attitudes to CAT administration of the study sample. Test for statistically significance differences in students' attitude to CAT administration were conducted for students across the various groups formed according to gender, ability, CAT treatment, computer literacy indicators, and affective characteristics. Analyses were carried out using one-way analysis of variance (ANOVA). The a priori level of significance was set at .05.

Results

Table 1 shows the percentage of students selecting each response on the *SACATA* questionnaire items, and the mean item ratings for the whole sample. Ten items out of the first twelve items have a mean rating of 3.0 and greater, and the last item has a mean rating of 2.0. These results indicate an overall positive student reactions to CAT. That is, the study sample responded favourably to all four aspects of CAT administration.

Table 1
Percentage of Students Selecting Each Response on the SACATA Questionnaire Items, and the Mean Item Ratings for the Whole Sample

Item	N	Strongly Agree (%)	Agree (%)	Disagree (%)	Strongly Disagree (%)	Mean (sd)
Method of Testing						
1. Directions were clear.	113	75.2	23.9	0.9	0.0	3.7 (0.5)
2. (Not) bothered by inability to review.	113	10.6	33.6	34.5	21.2	2.3 (0.9)
3. Graphics were clear.	113	60.2	35.4	3.5	0.9	3.6 (0.6)
4. Tutorials/feedback were useful.	58	43.1	53.4	3.4	0.0	3.4 (0.6)
Use of Equipment						
5. Using keyboard was (not) difficult.	113	29.2	43.4	21.2	6.2	3.0 (0.9)
6. Using space bar was (not) difficult.	113	48.7	43.4	5.3	2.7	3.4 (0.7)
7. (No) problem/scrolling.	113	38.9	38.9	21.2	0.9	3.2 (0.8)
8. Reading as easy as from booklet.	113	24.8	48.7	23.9	2.7	3.0 (0.8)
9. (No) problem/brightness of screen.	113	38.1	57.5	3.5	0.9	3.3 (0.6)
10. (No) problem/crowded screen.	113	33.6	57.5	8.0	0.9	3.2 (0.6)
Anxiety						
11. Compared with PP, (not) more anxious.	113	31.0	53.1	12.4	3.5	3.1 (0.8)
12. Prefer computer over booklet.	113	24.8	44.2	27.4	3.5	2.9 (0.8)

Item	N	Too Easy (%)	About Right (%)	Too Difficult (%)	Mean (sd)
Question Difficulty					
13. Perception of question difficulty.	113	1.8	93.8	4.4	2.0 (0.3)

For the 'method of testing' category, 99% of the students indicated that the directions for taking the test were clear, and 95% indicated that the graphics for the items were clear. However, 55% of them were bothered by the inability to review their answers after they had pressed the enter key. This item also has the lowest mean rating. For students who took the CAT with performance diagnosis, majority of them (96%) found the immediate feedback and tutorial useful.

Results reveal that most of the students did not have problems using the computer equipment. Only 8% of students expressed problems using the spacebar, 5% found letters on the screen too bright, and 9% found the screen too crowded. However, it is noted that 27% of the students indicated having general difficulty using the keyboard, 22% had problems with scrolling, and 27% found that their reading comprehension were affected by reading on a monitor screen as compared to a test booklet.

Results show that 84% of them reported that they were not more anxious with a computer-administered test than with a paper-administered test, and 69% preferred the computer-adaptive test over the paper-and-pencil version.

Lastly, the results for question difficulty show that majority of students perceived the test questions to be of just the right difficulty (94%). Only about 2% of them perceived the questions too easy and 4% too difficult.

Table 2
Analysis of Variance for the Gender, Ability Level and CAT Treatment Variables and the SACATA Mean Item Ratings

	Gender			Ability Level				CAT Treatment		
	Male	Female	F-ratio	Low	Average	High	F-ratio	with pd	w/o pd	F-ratio
Method of Testing										
1. Directions were clear.	3.8	3.7	0.58	3.8	3.6	3.8	3.19	3.7	3.7	0.00
2. (Not) bothered by inability to review.	2.1	2.6	8.76**	2.4	2.4	2.2	0.34	2.4	2.3	0.82
3. Graphics were clear.	3.6	3.5	0.53	3.5	3.6	3.5	0.26	3.6	3.5	0.95
4. Tutorials/feedback were useful.	3.4	3.4	0.02	3.4	3.4	3.3	0.25	3.4	-	-
Use of Equipment										
5. Using keyboard was (not) difficult.	2.9	3.0	0.02	3.2	2.9	2.9	1.61	3.1	2.8	2.04
6. Using space bar was (not) difficult.	3.4	3.4	0.17	3.4	3.3	3.4	0.31	3.3	3.5	1.81
7. (No) problem/scrolling.	3.3	3.1	2.38	3.2	3.0	3.3	1.83	3.1	3.2	0.29
8. Reading as easy as from booklet.	2.8	3.1	2.64	3.0	3.0	2.9	0.23	3.0	2.9	0.14
9. (No) problem/brightness of screen.	3.4	3.2	2.94	3.3	3.3	3.3	0.08	3.3	3.3	0.00
10. (No) problem/crowded screen.	3.3	3.2	1.50	3.2	3.2	3.3	0.58	3.2	3.3	0.31
Anxiety										
11. Compared with PP, (not) more anxious.	3.2	3.0	2.12	3.2	2.9	3.3	2.50	3.2	3.0	1.79
12. Prefer computer over booklet.	2.9	3.0	0.40	3.1	2.7	3.0	2.19	3.0	2.8	1.16
Question Difficulty										
13. Perception of question difficulty.	2.0	2.0	0.18	1.9	2.0	2.0	1.65	2.0	2.0	0.17

Gender variable :   Male (n=54), Female (n=59)
Ability Level variable: Low (n=32), Average (n=47), High (n=34)
CAT Treatment variable: with performance diagnosis (n= 58), without performance diagnosis (n=55).
**   p<.01

Comparison by Gender

Inspection of the mean ratings by gender (see **Table 2**) reveal that both groups were overall positive towards CAT administration. Results show a significant gender difference on item 2 (F=8.76, df=1/111, p<.01). Compared to girls, boys were more bothered by the inability to review items during the adaptive testing.

Comparison by Ability Level

Results by ability level show no significant differences on any of the items (see **Table 2**). However, inspection of the mean ratings reveal some interesting patterns for items 1, 11, and 12. The lowest mean ratings are obtained for the average ability students. These results seem to indicate that average ability students had more problems with the directions for testing; they were more anxious about the computerised testing, instead they preferred the paper version over the computer test.

Comparison by CAT Treatment

Results by CAT treatment also show no significant differences on any of the items (see **Table 2**). However, inspection of the mean ratings reveal lower mean ratings on three items, namely, items 5, 11 and 12, against the group who were administered CAT without performance diagnosis. These results seem to indicate that the students had more difficulties with the keyboard, and they were more anxious when taking the CAT, and also preferred the test booklet to the computer. For item 4, a mean rating of 3.4 was obtained. It can be inferred that students who took the CAT with performance diagnosis found the feedback very useful.

Table 3
Analysis of Variance for the Three Student Computer Literacy Variables and the SACATA Mean Item Ratings

	Home Comp			Computing Exp			Frequency of Comp Use			
	With	Without	F-ratio	Low	High	F-ratio	Low	Average	High	F-ratio
Method of Testing										
1. Directions were clear.	3.8	3.7	1.58	3.7	3.8	0.93	3.6	3.9	3.8	2.72
2. (Not) bothered by inability to review.	2.3	2.4	0.01	2.4	2.3	0.82	2.5	2.4	1.9	3.67*
3. Graphics were clear.	3.7	3.4	7.85**	3.5	3.6	1.15	3.5	3.5	3.8	2.47
4. Tutorials/feedback were useful.	3.4	3.4	0.05	3.4	3.4	0.02	3.4	3.2	3.6	2.20
Use of Equipment										
5. Using keyboard was (not) difficult.	3.0	2.9	0.47	2.9	3.0	0.25	3.0	2.9	3.0	0.46
6. Using space bar was (not) difficult.	3.5	3.3	2.16	3.2	3.6	7.79**	3.2	3.6	3.5	4.28*
7. (No) problem/scrolling.	3.1	3.2	0.28	3.2	3.2	0.01	3.1	3.2	3.2	0.02
8. Reading as easy as from booklet.	2.9	3.0	0.05	2.9	3.0	0.65	2.9	3.0	2.9	0.11
9. (No) problem/brightness of screen.	3.4	3.2	2.84	3.2	3.5	8.84**	3.2	3.4	3.6	5.38**
10. (No) problem/crowded screen.	3.3	3.1	1.74	3.1	3.4	3.98	3.1	3.3	3.4	2.35
Anxiety										
11. Compared with PP, (not) more anxious.	3.2	3.0	1.03	3.0	3.3	4.15**	3.0	3.2	3.2	0.46
12. Prefer computer over booklet.	2.9	2.9	0.04	2.8	3.0	2.00	2.9	3.0	2.9	0.15
Question Difficulty										
13. Perception of question difficulty.	2.0	2.0	0.26	2.0	1.9	1.52	2.0	2.0	2.0	0.06

Home Computer Ownership variable : With (n=70), Female (n=43)
Computing Experience variable: Low (n=61), High (n=52)
Frequency of Computer Use variable: Low (n=49), Average (n=40), High (n=24).
* p<.05
** p<.01

Comparison by Home Computer Ownership

Analyses of mean item ratings by home computer ownership variable show that both groups are overall positive towards CAT administration (see **Table 3**). A significant difference between the two comparison groups is obtained for item 3 (F=7.85, df=1/111, p<.01). Compared to their counterparts, students who did not have computers found the graphics presented during the testing to be less clear. They also indicated less favourable ratings for item 6 (using the space bar), items 9 and 10 (monitor screen), and item 11 (anxiety taking computerised test).

Comparison by Computing Experience

Analyses of mean item ratings by computing experience variable show higher item means, except items 2 and 13, in favour of the group with higher computing experience (see **Table 3**). Significant differences are found on three items: item 6 (F=7.79, df=1/111, p<.01), item 9 (F=8.84, df=1/111, p= <.01) and item 11 (F=4.15, df=1/111, p<.01). That is, compared to their counterparts, students with higher computing experience were overall more positive towards the computerised testing. In general, they had less problems with the method of testing, and the use of computer equipment. They were also less anxious towards the computerised testing; instead they preferred it over the paper version. However, the two groups did not differ in their reactions towards the inability to review questions during the testing. They also perceived the question difficulty to be just right for them.

Comparison by Frequency of Computer Use

Analyses by frequency of computer use variable show significant differences on three items (see **Table 3**). These include: item 2 (F=3.37, df=2/110, p<.05), item 6 (F=4.28, df=2/110, p= <.05) and item 9 (F=5.38, df=2/110, p<.01). The Scheffe test shows that for item 2, the "high" group has a lower mean rating than the "low" group (1.9 versus 2.5); for item 6, the "low" group has a lower mean rating than "average" group (3.2 versus 3.6); and for item 9, the "low" group has a lower mean rating than the "high" group (3.2 versus 3.6). In other words, the high

frequent users were most dissatisfied with the inability to review items during the testing, but low frequent users had most problems using the computer equipment, specifically with the space bar and the computer screen.

Table 4
Analysis of Variance for Two Student Affective Characteristic Variables and the SACATA Mean Item Ratings

	Computer Attitude				Attitude to Science			
	Least positive	Average positive	Most positive	F-ratio	Least positive	Average positive	Most positive	F-ratio
Method of Testing								
1. Directions were clear.	3.7	3.8	3.8	0.63	3.8	3.7	3.7	0.02
2. (Not) bothered by inability to review.	2.3	2.5	2.2	0.76	2.3	2.4	2.4	0.17
3. Graphics were clear.	3.6	3.5	3.6	0.78	3.4	3.4	3.8	5.94**
4. Tutorials/feedback were useful.	3.5	3.4	3.3	0.52	3.3	3.4	3.5	1.03
Use of Equipment								
5. Using keyboard was (not) difficult.	2.8	3.0	3.2	2.03	3.0	2.9	3.0	0.32
6. Using space bar was (not) difficult.	3.3	3.3	3.5	1.23	3.4	3.3	3.5	1.06
7. (No) problem/scrolling.	3.1	3.2	3.2	0.16	3.0	3.2	3.3	1.05
8. Reading as easy as from booklet.	2.8	3.0	3.1	1.94	2.9	3.0	3.0	0.11
9. (No) problem/brightness of screen.	3.3	3.2	3.4	0.93	3.3	3.2	3.5	2.00
10. (No) problem/crowded screen.	3.2	3.1	3.4	2.63	3.2	3.1	3.5	3.57*
Anxiety								
11. Compared with PP, (not) more anxious.	2.8	3.1	3.5	5.00**	3.0	3.1	3.3	2.89
12. Prefer computer over booklet.	2.8	2.9	3.0	1.05	2.9	2.8	3.0	1.01
Question Difficulty								
13. Perception of question difficulty.	2.0	2.0	2.0	0.42	2.0	1.9	2.0	3.09

Computer Attitude variable :   Least positive (n=37), Average positive (n=38), Most positive (n=38)
Attitude to Science variable: Least positive (n=37), Average positive (n=38), Most positive (n=38)
* $p < .05$
** $p < .01$

Comparison by Computer Attitude

Analyses of mean ratings by computer attitude show that the "most positive" group had an overall most favourable attitudes towards the CAT administration (see **Table 4**). On the other hand, the "least positive" group had more negative attitudes toward several aspects of the testing. Results show a significant difference on item 11 ($F = 5.00$, $df = 2/110$, $p < .01$). The Scheffe test reveals that the "most positive" group has higher mean rating than the "least positive" group (3.5 versus 2.8). In other words, compared to their counterparts, students who were most positive towards computers were least anxious towards a computerised testing. This pattern of mean ratings is also noted for item 12.

Comparison by Attitude towards Science

Analyses by science attitude variable show significant differences on two items (see **Table 4**). These include: item 3 ($F = 5.94$, $df = 2/110$, $p < .01$) and item 10 ($F = 3.57$, $df = 2/110$, $p < .05$). The Scheffe tests reveal that for item 3, the "most positive" group has higher mean rating than both the "average and "least positive" groups (3.8 versus 3.4 and 3.4 respectively); and for item 10, the "most positive" group has significantly higher mean rating than the "average" group only (3.5 versus 3.1). In other words, compared to their counterparts, students who were most positive towards science were most satisfied with the graphics presented during the testing. Unlike their counterparts in the "average" group, they were more satisfied with the computer screen, that is, they did not find the letters crowded.

Inspection of the mean ratings on other items reveal some other interesting points about the students in the "most positive" group. Students in this group found the tutorial feedback most useful to them; they were also least anxious about taking the computerised tests, in fact they most preferred the computerised testing over the paper version. Another striking point is noted about students' perception of question difficulty. Whilst both the "least positive" and "most positive" groups perceived the questions presented to be just right, the "average" group perceived the questions to be slightly difficult.

Discussion and Conclusions

Based on the findings of this study, a number of conclusions could be made about students' reactions towards computerised adaptive testing, and its implications for Singapore schools will be discussed.

First, students' reactions towards CAT administration were overall positive. This result is congruent with the findings reported by previous studies in the United States (Baghi et al. 1992; Legg & Buhr, 1992; Moe & Johnson, 1989). Thus in considering a widespread use of computers for testing purposes in Singapore schools, this finding related to students' acceptance of CAT is supporting evidence of students' readiness for the innovative change.

However, an implementation of CAT in the field must take note of some differential effects of gender, home computer ownership, computing experience, frequency of computer use, computer attitude and attitude to science groups in their reactions to some aspects of CAT.

Students reacted less favourably towards their inability to review questions after a response had been entered. This result is similar to the findings in Baghi et al. (1992), and Legg and Buhr (1992) studies. In CAT, the ability to review items is withheld from students as the response to a previous item is used to estimate the student's current ability and to select the most informative item to be presented next. Interestingly, both the boys and the high frequent users of computers were most bothered by their inability to review questions during the testing. Two questions are posed here. Were the boys also the high frequent computer users in the group? Were the boys greater risk-takers when selecting responses in a multiple-choice question format? If so, then this could help to explain the more negative reactions observed with boys towards their inability to review questions during the testing. Further research will be needed to examine this supposition.

In general, students' familiarity with the computer seems to affect their reactions to the graphics presented and to the use of computer equipment. For instance, the group with home computers found graphics presented during the testing to be clear; and both the groups with high computing experience and of high frequent users had less problems with the computer keyboard or computer screen. It also affected students' anxiety when taking a computerised test. For instance, the group with higher computing experience reported less anxiety when taking a computer test compared to a printed test. These findings have indirect implications, which are reflected in the equity concerns highlighted by Sutton (1993), on the differential effects of computer literacy on computerised testing. Of final consequence would be students' test performance obtained by the use of this testing method. This concern must be investigated further.

Lastly, the groups were unanimous in their perception of question difficulty. Especially with the low and high ability groups, they did not find the test questions too easy or too difficulty. Rather, they perceived them to be about right. This finding attests to the adaptive nature of the testing. In addition, students who took the CAT with performance diagnosis found the feedback useful. Two conclusions can be drawn from these results. First, the tailored testing can be exploited as a motivating factor for students during the testing. That is, students are encouraged and/or challenged to perform when they are administered test questions that are geared to their ability level. Second, feedback is the essence of testing, and this is also part and parcel of the whole instructional process. Thus, when students receive immediate feedback on the correctness/incorrectness of their responses to test questions, it brings about a cognitive concord/conflict. That is, it serves to either confirm or reveal gaps in students' knowledge and understanding elicited in a test question. Using a strategy of test performance diagnosis, students can be encouraged to actively participate in self-evaluation and independent learning.

Given the favourable reactions of the present results and the results of others, it may well advocate a more widespread use of computers for testing purposes. For Singapore schools, the prospects for CAT replacing paper-and-pencil testing in some instances are good. The Education Ministry's recent launch of a $2 billion master plan for information technology (IT) will mean increased availability of computers and a broader base of access to computers. As schools begin to explicitly use computer technology for instructional purposes, they will also need to rethink about

their assessment modes. CAT can be considered as an alternative assessment method incorporating the use of computers.

References

Baghi, H., Ferrara, S. F., & Gabrys, R. (1992). Student attitudes toward computer-adaptive test administrations. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, LA, April 12-16.

Bunderson, C. V., Inouye, D. K., & Oslen, J. B. (1989). The four generations of computerised educational measurement. In R. L. Linn (Ed.), Educational Measurement (3rd ed.), (p367-407). New York: American Council on Education: Macmillan.

Foong, Y. Y., & Lam T. L. (1988). Development and validation of the attitudes to science instrument. Institute of Education Research Paper, SSV1, Singapore.

Gressard, C. P., & Loyd, B, H. (1986). Validation studies of a new computer attitude scale. AEDS Journal, 19(4), 295-301.

Kim, J. G. & McLean, J. E. (1994). The relationships between individual difference variables and test performance in computerised adaptive testing. Paper presented at the annual meeting of the Mid-South Educational Research Association, Nashville, TN, November 9-11.

Kim, J. G. (1993). Individual differences in computerised adaptive testing. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans, LA, November 10-12.

Legg, S. M., & Buhr, D. C. (1992). Computerised adaptive testing with different groups. Educational Measurement: Issues and Practice, 11(2), 23-27.

Moe, K. C. & Johnson, M. F. (1988). Participants' reactions to computerised testing. Journal of Educational Computing Research, 4(1), 79-86.

Oslen, J. B., Maynes, D. D., Slawson, D. Ho, K. (1989). Comparisons of paper-administered, computer-administered and computerised adaptive achievement tests. Journal of Educational Computing Research, 5(3), 311-326.

Parshall, C. G., & Kromrey, J. D. (1993). Computer testing versus paper-and-pencil testing: An analysis of examinee characteristics associated with mode effect. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA., April 12-16.

Sutton, R. E. (1993). Equity issues in high stakes computerised testing. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA., April 11-16.

Vogel, L. A. (1994). Explaining performance on p&p versus computer mode of administration for the verbal section of the graduate record exam. Journal of Educational Computing Research, 11(4). 369-383.

Ward, T. J., Hooper, S. R., & Hannafin, K. M. (1989). The effects of computerised tests on the performance and attitudes of college students. Journal of Educational Computing Research, 5(3), 327-333.

Wise, S. L., & Plake, B. S. (1989). Research on the effects of administering tests via computers. Educational Measurement: Issues and Practice, 8(3). 5-10.