A Review of CAT Review

Renata Sekula-Wacura

The American Society of Clinical Pathologists administers 20 fixed length (100 item) registry examinations for laboratory professionals. Until 1993, the testing was of the paper and pencil variety, and a candidate was free to review items and change answers up to the time limit of the test. A computer adaptive test (CAT) administration was adopted in 1994. During a CAT, each examinee is administered a unique 100 item test (selected from an item bank of 500+ items) that is tailored specifically to their ability. Each item in the item bank has been calibrated for difficulty using a Rasch model (Wright and Stone, 1979). The candidate is first presented with an item whose calibration value is near or at the pass cutoff point for that exam. If the item is answered correctly, the computer program next presents a more difficult item. If the item is answered incorrectly, an easier item is presented, and so on.

The ASCP CAT program incorporates a review session. During the computer adaptive portion of the test, a candidate is required to answer all 100 items in the order presented. During this portion, any item can be marked for later review. After completing all 100 items, the computer adaptive portion is over and the program shifts into a review session. During this session, the candidate is free to look at any question in any order and to change answers until the time limit of the test is reached.

What effect does the review session have on the final score (person ability measure) and pass/fail decision? To answer this, ability measures pre and post review were examined for a

Table 1. Summary of test outcome before and after review.

BEFORE REVIEW PASS	BEFORE REVIEW FAIL
19,517 67%	9,776 33%
AFTER REVIEW PASS	AFTER REVIEW FAIL
20,197 69%	9,096 31%

period of three years. Table 1 summarizes the data. Out of 29,293 candidates, 67% passed before review and 69% after review. Table 2 summarizes the effect of the review session on the pass/fail decision. From Table 2 it can be determined that 1300

Table 2. Sumary of test pass/fail outcome before and after review

PASS TO PASS	PASS TO FAIL
19,207 66%	310 1%
FAIL TO PASS	FAIL TO FAIL
990 3%	8,786 30%

candidates had their decision altered due to the review session (pass to fail, or fail to pass). The good news is that three times as many candidates who changed answers improved their scores by doing so as opposed to those that lower their scores.
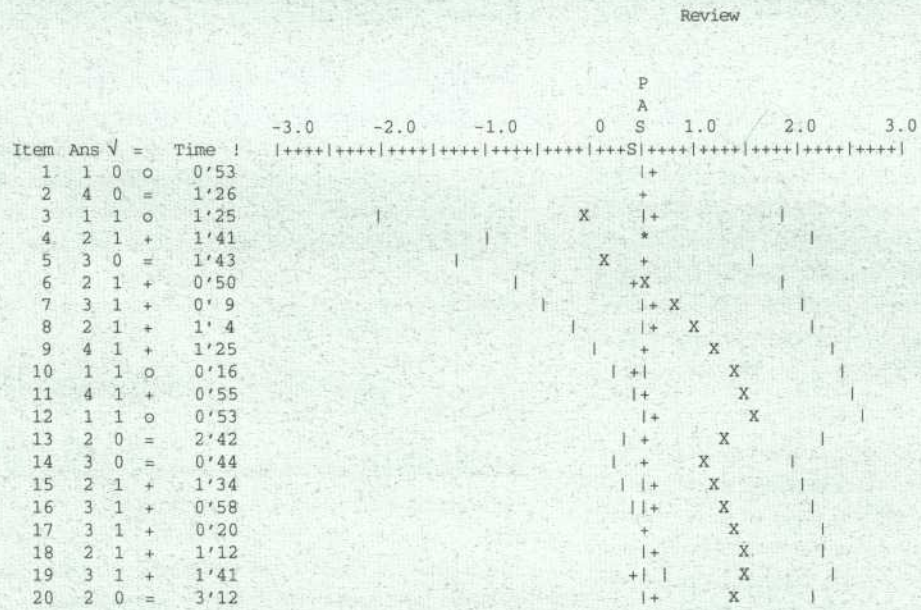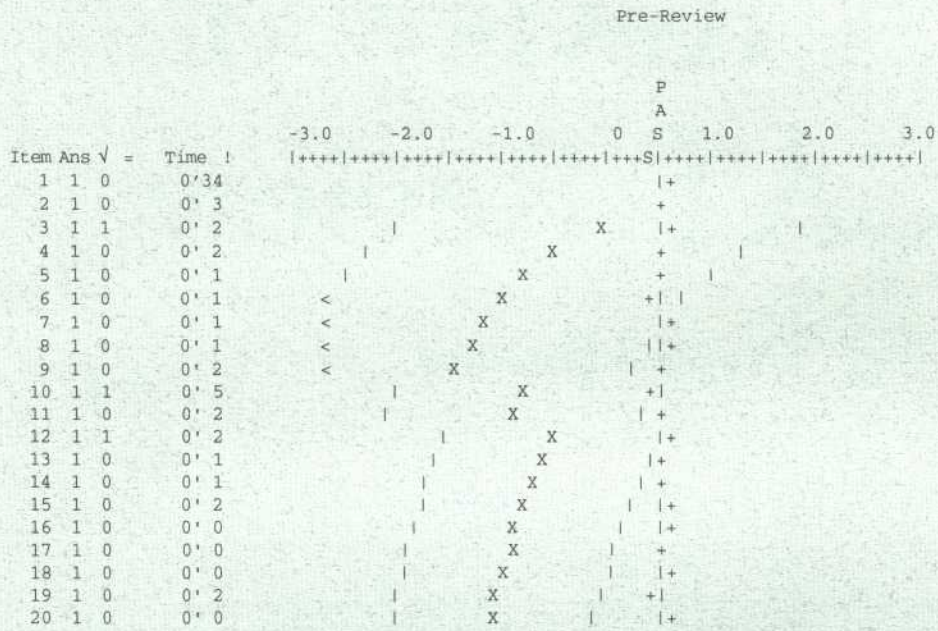
What are the candidates doing in the review session? Are they changing many answers or just a few? To answer this a table was created based on all candidates. A difference in the candidate pre and post review measure and the deviation of their difference (based on the candidate standard error) was calculated.

Figure 1. Candidate measure map

```
                                            Pre-Review


                                              P
                                              A
                        -3.0    -2.0    -1.0    0   S  1.0     2.0     3.0
Item Ans √ =   Time !   |++++|++++|++++|++++|++++|++++|+++S|++++|++++|++++|++++|++++|
  1  1  0      0'34                                     |+
  2  1  0      0' 3                                     +
  3  1  1      0' 2              |                 X    |+            |
  4  1  0      0' 2         |                   X       +         |
  5  1  0      0' 1        |                  X         +      |
  6  1  0      0' 1    <                  X             +| |
  7  1  0      0' 1    <                X               |+
  8  1  0      0' 1    <                X               ||+
  9  1  0      0' 2    <             X                  |  +
 10  1  1      0' 5        |             X              +|
 11  1  0      0' 2        |            X               |  +
 12  1  1      0' 2           |             X           |+
 13  1  0      0' 1         |             X             |+
 14  1  0      0' 1          |           X              | +
 15  1  0      0' 2         |            X              | |+
 16  1  0      0' 0          |          X               |  |+
 17  1  0      0' 0        |           X                | +
 18  1  0      0' 0        |           X                |  |+
 19  1  0      0' 2       |         X              |    +|
 20  1  0      0' 0        |          X             |   |+



                                            Review


                                              P
                                              A
                        -3.0    -2.0    -1.0    0   S  1.0     2.0     3.0
Item Ans √  =   Time !   |++++|++++|++++|++++|++++|++++|+++S|++++|++++|++++|++++|++++|
  1  1  0  o    0'53                                    |+
  2  4  0  =    1'26                                    +
  3  1  1  o    1'25            |                  X    |+             |
  4  2  1  +    1'41               |                    *            |
  5  3  0  =    1'43                  |           X     +           |
  6  2  1  +    0'50                    |              +X           |
  7  3  1  +    0' 9                  |               |+ X          |
  8  2  1  +    1' 4                      |           |+   X        |
  9  4  1  +    1'25                      |      +         X        |
 10  1  1  o    0'16                      |+|         X        |
 11  4  1  +    0'55                      |+           X       |
 12  1  1  o    0'53                      |+             X     |
 13  2  0  =    2'42                      | +         X        |
 14  3  0  =    0'44                      | +         X        |
 15  2  1  +    1'34                      | |+        X        |
 16  3  1  +    0'58                      ||+          X       |
 17  3  1  +    0'20                        +          X       |
 18  2  1  +    1'12                      |+            X.      |
 19  3  1  +    1'41                     +| |          X       |
 20  2  0  =    3'12                      |+           X       |
```

X is candidate's ability
+ is item difficulty level
√ column indicates if the question is answered right(1) or wrong(0)
= column shows how question was changed: + from wrong to right, = from wrong to wrong, - from right to wrong
Ans column indicates answer candidate choose
| indicates standard error limits

Table 3. Candidates who chanced more than 25 questions with deviation from their standard error greater than 2.

Deviation from standard error (in logits)	Difference between pre and post review measure	Number of questions changed	Final pass/fail outcome
2.09	0.45	28	CHANGED
2.09	0.45	28	NOT CHANGED
2.09	0.47	29	CHANGED
2.10	0.47	26	NOT CHANGED
2.10	0.46	29	NOT CHANGED
2.13	0.52	36	NOT CHANGED
2.13	0.45	96	CHANGED
2.14	0.55	43	NOT CHANGED
2.16	0.52	44	NOT CHANGED
2.16	0.53	27	NOT CHANGED
2.30	0.50	31	CHANGED
2.32	0.53	32	NOT CHANGED
2.49	0.55	55	NOT CHANGED
2.52	0.58	30	CHANGED
2.54	0.54	27	CHANGED
2.55	0.61	26	NOT CHANGED
2.57	0.54	37	CHANGED
2.70	0.59	38	CHANGED
2.84	0.60	38	NOT CHANGED
3.12	0.76	42	CHANGED
3.16	0.67	28	NOT CHANGED
3.28	0.69	30	NOT CHANGED
3.36	0.72	47	NOT CHANGED
3.47	0.73	45	CHANGED
3.51	0.86	40	CHANGED
4.08	1.10	26	CHANGED
4.20	0.89	49	NOT CHANGED
4.32	0.91	72	NOT CHANGED
4.92	1.04	67	NOT CHANGED
5.58	1.19	88	NOT CHANGED
6.08	1.31	58	CHANGED
6.95	1.49	81	CHANGED
7.13	1.73	54	CHANGED
7.84	1.68	87	CHANGED
12.38	3.33	89	CHANGED

Of the 29,293 candidates, 99% changed 25 or fewer answers. Of the 1% who changed more than 25 answers, 88% had a deviation of their difference in their measure equal to or less than 2 standard errors. The candidates with a difference of greater than 2 standard errors are summarized in Table 3.

Of interest are candidates with deviations greater than 4 logits and changing more than 50% of their answers. The CAT program can generate a Candidate Measure Map that provides information about both the computer adaptive (pre-review) and review session portions of the test. Several maps were printed and a "cheater" strategy was detected. Figure 1 shows the first twenty items of both the pre-review and review sessions for one of the candidates. In the pre-review session, the candidate selected the answer "1" for every item. The time column indicates that sufficient time did not elapse for the item and distracters to be read before proceeding to the next item! The review session is where the candidate actually "took" the test. Items were read and appropriate answers were selected to the best of their ability.

The presumed purpose of the "cheater" strategy is an attempt to get an easier test. The CAT algorithm can detect this. After 40% of the answers are incorrect, the program will automatically select items with a measure close to the pass point. However, a very able candidate will likely get a test with an average difficulty below their ability.

This review of the CAT review has raised some interesting topics for further research. For example, can incorporating a minimum time requirement before allowing presentation of the next question eliminate the "cheater" strategy? Is there a correlation between the pass/fail decision and the number of answers changed? Is the candidate's true ability underestimated when the "cheater" strategy is employed? And finally, does the cut-point level of the exam influence the percentage of candidates going from pass to fail or fail to pass?
Stay tuned!

Renata Sekula-Wacura, MS, is Manager of Database and Network Operations at the ASCP Board of Registry. She enjoys relaxing on the beach and climbing mountains in her free time.

TESTING - TESTING - TESTING - TESTING