SSW REPORTS

IF YOU RETEST USING THE SSW - READ THIS

Vol. 20 No. 3 August 1998

PITFALLS OF RETEST

Jack Katz, Kim Tillery, Patricia Amorim

Abstract

Audiologists often retest using the SSW or other Central Auditory Processing (CAP) procedures, assuming that they can use the standard norms when the person is re-tested. There are reasons to believe that this is not the best approach. We propose a Predicted Total NOE Score when giving the SSW for a second time.

There are many reasons why an audiologist would retest an individual on the central test battery, just as he or she would give a retest on a peripheral test battery. However, the characteristics of a CAP test are such that they often give misleading information on retest. Therefore, we believe that audiologists should be aware of these pitfalls and we provide a potential solution.

When we administer a puretone threshold audiometric test to a patient we must first prepare the person with our instructions and allow for learning on the first few thresholds. We often *recheck* the first threshold later to be sure that significant learning did not take place after that. If the retest threshold is significantly better, then we retest other thresholds until we are satisfied that we have the correct information.

The situation with CAP testing is *more complex*, both literally and figuratively.

Because of the complex SSW task that is presented to the patient, it is likely that it will take more than a half dozen items to establish optimum performance. For normal listeners it generally requires a shorter time to figure out how to take the SSW test. Fortunately, it takes the person with CAPD longer. Even then their maximum score is not as good as the normal group, but the difference between their test and retest scores is greater.

Is this fair to test a person on CAP tasks before they reach their maximum score? Yes, even with the SSW, a repetitive auditory task, it takes the person with CAPD twice as long (or four times as long) to catch on. Under everyday, variable conditions, these people may never catch on to an optimal system for listening.

Jim Jerger pointed out many years ago that, as diagnosticians, we may not know exactly how or why our tests work, the main thing is for them to differentiate certain populations. This concept applies here. Do CAP tests work, yes. For example, the SSW identified 92% of people referred for CAP testing compared to only 20% of the normative population that failed (Katz et al., 1988).

The problem we face is *not* one of validity or even reliability. It is a learning factor (LF) with which we must deal. It is not unlike the learning factor on the puretone tests, it is simply a larger factor and more common, because the SSW is more complex, giving the person much more to learn.

You may recall that the SSW test was validated on site-of-lesion cases (Katz & Pack, 1975; Katz et al., 1981) and that the same response biases were found to predict the performance of those with CAPD (Katz & Smith, 1991; Katz, Kurpita et al., 1992; Katz, Smith, Kurpita, 1996). Also, the SSW has been shown to be reliable using a variety of methods (Katz & Arndt, 1979; Katz & Schultz, 1985; Katz & Kram, 1993; Katz, Tillery, Batheja, 1997).

In fact, it will be seen that all of the studies presented here demonstrated high correlations between test and retest. It was the *absolute scores* that differed. That is to say, those who were best, worst, and in between on the test were very likely to hold the same position on retest. It is just that the scores they got were different. You can assume that because of learning the retest score was generally *better* than the original score.

Two Major Factors Influencing Retest

The learning factor (LF) has been discussed above. Unfortunately, some people show a great deal of learning while others show very little. Occasionally, a person will have a slightly poorer score on retest.

Because of the unpredictability of the LF we *cannot* use a simple correction for retest (e.g., subtract 2 points for the retest). However a statistical prediction for the LF will be presented.

The second major factor, *maturation* (MF), must be considered as well. Normal control children continue to improve in their central test performance until about 11 years-of-age. However, children with problems may continue to improve (mature?) for a longer time before they reach a plateau.

We can estimate the maturation effect by looking at the normal improvement found with age or by calculating the improvement seen in learning disabled (LD) children from year to year. For normal control children studied in the National Sample (1985) there was approximately a 0.5 improvement for the mean Total NOE score per month between age 5 and 10 years. Over 10 years-of-age the improvement was .25 per month.

To determine how maturation tends to alter the NOE, a sample was collected of 225 patients seen for CAP testing at UB. They were 4 to 55 years of age (see Table 1).

Age	n	Mean	SD
5	26	61.5	21.4
6	27	43.7	18.9
7	26	27.7	21.1
8	26	21.6	10.4
9	30	21.3	15.9
10	31	18.8	11.8
11	24	17.5	14.9
12-18	17	10.1	7.3
21-55	17	6.6	4.9

Table 1. Sample of 225 cases seen for CAP evaluation at UB. This shows the mean Total NOE score (mean) and standard deviation (SD) as well as the sample size (n) for each age group.

The CAPD data may underestimate the MF for the older subjects because those who improved with age (due to maturation or therapeutic measures) were less likely to be seen when they were older. So this is likely to represent the more serious or tenacious problem group.

When looking at ages 5 to 12 years, the improvement is very similar to that found in the normal 5 to 10 year group, an average of about 0.5 errors per month. Between 12-18 years the difference is approximately 1 error per year. The two sets of data for control and LD cases are similar. Thus, the latter data were chosen for the MF because they are 1) simpler, and 2) represent the population in question.

Studying Test-Retest

You may wonder why it has taken so many years to find a solution to the retest question for the SSW and other central tests. For one thing it requires that a group of individuals with a CAPD be retested 1) more than a few days after the initial test (so the specific answers do not influence the retest) and 2) less than a few months after the first test to minimize any maturational or therapeutic effects.

These requirements make the research difficult to obtain in a clinical population. No study has been reported for the SSW or any other CAP test (to our knowledge) so far. [Ooops last week, during preparation of this article, JSHR published an article on retest of the SCAN.]

We have tried to garner similar data from other SSW studies with only fair success. The best study that we have seen which may offer appropriate data is the study of Tillery (1997).

Tillery's Test-Retest Study

In 1997 Tillery carried out a study with 34 children who were diagnosed as having Attention Deficit Hyperactivity Disorder (ADHD) and Central Auditory Processing Disorder (CAPD). These children were given a battery of CAP tests including the SSW. They were tested three times. The first was for a baseline with no medication, a second test was when they were taking either Ritalin (a drug to control hyperactivity and improve attention) or a placebo (an inactive) pill, and then for the third test they were on the reversed condition (placebo or Ritalin). Only the first two tests are relevant here.

Between the test and retest (the inter-test-interval or ITI), was a period of 54 days or approximately 7.7 weeks apart. The range was 26 to 137 days. This is a reasonable ITI, although it is slightly longer than ideal. It has a more satisfactory ITI than any of the other studies.

Seventeen of Tillery's subjects had a no-drug baseline followed by a placebo condition. Thus, the retest was not influenced by an active drug and therefore a rather benign treatment. The Total NOE score for the initial test was 26.9 and on retest was 22.4, an improvement of 4.5 points (or 17% improvement). It is likely that this change was due primarily to learning, although some maturation was likely involved (e.g., the person who was retested 137 days later had 20 weeks of maturation). Table 2 shows a summary of Tillery's test and retest results.

Test	Mean	SD	r
Baseline	26.9	13.4	---
Retest	22.4	14.1	---
Base vs Retest	4.6	6.6	.92

Table 2 shows the means (M), standard deviations (SD) and correlation (r) between the Total NOE test and retest scores.

The correlation between test and retest (r = 0.92) was a strong one. As expected a t-test showed the two means to be significantly different (t =2.89, p = 0.01).

Although the SSW results were highly reliable, the retest scores were not equal to the baseline scores. The NOE differences in this relatively short ITI, ranged from 15 points better to 5 points poorer. The SD between test and retest for Tillery's data was 5.6 errors. Using a >1 SD improvement (or 6 points) as a positive change, 6 subjects met that criterion. There is no reason to believe that her subjects actually improved in their CAP abilities between test and retest as they were not receiving therapy and the maturation period was minimal. The primary or exclusive reason was the LF. None of them appeared to be significantly worse using the 6 point criterion.

This finding supports the data of Katz (1968) who carried out a similar study with LD children (any with ADHD were not specifically identified because ADHD was unknown at that time). The mean ITI in that study was 86 days, a month longer than Tillery's group. The period between test and retest in the Katz study was 21 to 217 days for an average of 12 weeks. Despite this longer inter-test-interval the correlation was quite impressive (r = 0.88). The mean difference of 5 (no-drug vs. placebo retest) was slightly greater than Tillery's subjects. Perhaps the difference between the two studies is due to the slightly longer ITI for Katz '68.

Prediction of Learning on First Test

Because of the good correlations found between test and retest and the relatively small standard deviations there seemed to be a good rationale for calculating a prediction factor for SSW retest score (Guilford, 1956, p366-368). The formula is:

$$(1)\quad X = r\,(SD_{pre}/SD_{post})\;(Y - M_y) + M_x$$

where ***X*** is the Total NOE Score to be predicted for the learning factor; ***r*** is the correlation between Tillery's test and retest scores; SD_{pre} is the pretest (baseline) SD; SD_{post} is the posttest SD; **Y** is the pretest score for the patient in question, M_y is the mean Total NOE for the pretest for Tillery's subjects and M_x is the mean Total NOE for their retest.

Substituting into the equation we have:

$$X = 0.92\,(13.4/14.1)\,(Y - 27) + 22$$

This can be simplified to:

$$(2)\quad X = (Y-27)\,(0.95) + 22$$

Let us assume a person had a score of 42. We would get:

$$X = (15)\,(0.95) + 22 = 14 + 22 = 36.$$

Thus a person with a first test of 42 would be predicted to have a retest of 36 solely on the basis of learning.

Accounting for Inter-Test-Interval (ITI)

We now have a formula for predicting the learning factor, so the next job is to account for maturation and subtract the number of errors based on the ITI. For example, if we retest in 6 months, one year, or three years we must reduce the prediction accordingly. Note that we must use the *initial test* as a baseline and subtract, rather than work from the next year's age norms, even if the person is a year older. The next year's norms assume a naïve subject (the person was naïve only the first time).

We will use the LD norms from Table 1 to account for maturation. If the person was 6-0 years old for the first test and is now being retested at age 6-8 years we simply subtract 4 points from the statistical prediction of learning to get the Predicted Total NOE Score (PTNS). For any child 5 to 11 years we subtract .5 errors per month of ITI.

If the child at 6 years had a Total NOE score of 42 and the prediction for the learning factor is 36 (as we have calculated above) then we must subtract 4 to get a Predicted Total NOE Score which is 32. If there was *no* significant therapeutic change we would predict 32 ±5 based on both learning and maturation.

Applying Predicted Total NOE Score

To see if this prediction really works we will apply it to the placebo data of Katz (1968).

As in the case of Tillery's placebo data we assumed that no therapy and no maturation took place. Let's see what we would have concluded if these were unknown clients

who had gotten foot message therapy for their CAPD.

Table 3 shows the results for 6 illustrative subjects. They were chosen to show the variety of findings in the Katz study.

Test	Retest	Sig-1?	PTNS	Sig-2?
47	47	n	40	n
50	38	Y	43	n
35	19	Y	29	Y
33	21	Y	26	n
8	0	Y	4	n
7	4	n	1	n

Table 3. Initial Total NOE score, retest score, Sig-1 (whether there is improvement of $\geq$ 6 points), Predicted Total NOE Score (based on both LF and MF), and Sig-2 (whether there is improvement of $\geq$ 6 points).

The first and last cases (above) are simply to show that some people have no significant improvement (by 6 errors or more) on retest. Of course, the predicted score was unnecessary for these cases, but was carried out for fun. Subjects 2, 4 and 5 all would have appeared to have benefited from the foot message therapy as they improved by 6 points or more on retest. However, the PTNS demonstrated (correctly) that foot message (although wonderful) does not correct CAPD (or improve SSW results). In one case (#3) there was a significant improvement in both the typical comparison (difference =16) as well as when comparing baseline to the PTNS (difference =10).

For Katz' 20 subjects 7 (35%) appeared to have had significant improvement on retest before taking learning and maturation into account. When we used the LF and subtracted the MF only one case appeared significant (#3 above). Thus, the retest prediction procedures produced a much improved view of foot message therapy as applied to CAPD.

Discussion

If you wish to give retests but have not had criteria to assess therapeutic or other types of improvement, there is some relief here. Tillery's data can be used to calculate the learning factor on the SSW test in an almost pure fashion. This accounts for most of the reduction in our prediction in the Katz (1968) data, because the ITI was relatively short. However, in one year we could anticipate an improvement of 6 points for most ages. So this factor becomes a greater influence as the ITI increases.

We tried out the PTNS on Tillery's original group and found 6 subjects (38%) initially improving by 6 or more, but when PTNS was used only one (6%) appeared to be significant. This provides some more support for this approach.

On Katz's Ritalin subjects 9 (38%) out of 24 improved by $\geq$ 6 errors. When PTNS was used, 8 of the 9 were no longer significant, only 5% appeared to improve due to the Ritalin. This supports the findings of the study that there was no significant difference between those who took Ritalin vs. placebo.

Retest Prediction Table

To simplify the effort in arriving at a the LF, Table 4 can be used. Table 4 shows the initial score and the predicted learning factor when taking the test again. This score must be reduced by the ITI to arrive at a proper prediction.

For example a 7-year-old child is seen and initially gets a score of 30 (Total NOE). The child is seen for retest 14 months later to determine if he still needs therapy as the parents thought he might have outgrown the CAPD. On retest he gets a score of 16 which is just at the limit of normal for an 8-year old. Is he really okay now? Please check Table 4 and subtract for the ITI before you agree.

PreTest NOE	Learning Factor	PreTest NOE	Learning Factor
0 TO 4	0	32	27
5	1	33	28
6	2	34	29
7	3	35	30
8	4	36	31
9	5	37	32
10	6	38	32
11	7	39	33
12	8	40	34
13	9	41	35
14	10	42	36
15	11	43	37
16	12	44	38
17	12	45	39
18	13	46	40
19	14	47	41
20	15	48	42
21	16	49	43
22	17	50	44
23	18	51	45
24	19	52	46
25	20	53	47
26	21	54	48
27	22	55	49
28	23	56	50
29	24	57	51
30	25	58	51
31	26	59	52

Table 4. To predict retest based on learning alone, find pretest NOE score to locate the prediction in the right-hand column. Then subtract for the maturation factor.

The solution to the practice case from page 17 follows. With a score of 30 for the child we get a PTNS of 25 based on the LF. Because the ITI was 14 months and the child was between 5 and 11 years, we subtract 7 points for the MF. This gives us a prediction of 18. To be significantly improved he would need a score of 12 or better (as 6 is 1 SD). So, it looks like the retest improvement was a function of the LF and MF and not that CAPD suddenly disappeared.

Summary of Retest Prediction

If you give the SSW test for the second time to a person:

- it is chancy to use the usual age norms
- roughly 35% (or more depending on the ITI) of the cases are likely to be better by 6 points or more
- they will *appear* to have improved when there was no real improvement
- improvement was likely due to learning and maturation in these cases

Children who take the test for a second time will likely have learned (LF) from the first test and therefore perform better on the retest. If a significant amount of time has elapsed between test and retest, then this too (MF) may have an important effect because of maturation.

To figure out the appropriate score on retest we can use the Predicted Total NOE Score (PTNS):

- calculate the learning factor (LF)
  a) use formula **2** on pg 16, *or*
  b) use table 4 on this page
- then subtract the maturation factor (MF) from the LF
  a) .5/month - ages 5 to 11
  b) 1/year - ages 12 to 18
- a score of 6 or more below this prediction would show significant improvement
- even with this procedure about 5% of the control Ss appeared to improve when there was no reason to believe it was anything more than LF and MF

Beware, even though the person is a year, or two, older we do not use new age norms. Instead we subtract from the <u>original</u> Total NOE Score to get the PTNS.