Split-Half Study
Jack Katz

A question from an audiologist on the *Simple & Effective* Web-group stimulated quite a bit of interest with a variety of opinions from our group. He tested a 6-year-old, but just gave the first 20 items of the SSW because of issues with the child that he observed. Having done so; he was unsure how to handle the data.

To briefly answer his question I suggested multiplying the scores by 2 to extrapolate the approximate results so that they could be compared to the 6 year norms. But, I could only remember one study on split-half SSW (Katz & Arndt, 1982) that was carried out and that was with adults who had either brain lesions or peripheral hearing losses. That just now triggered a memory of a study with children that I found in SSW Reports (Katz & Cummings, 1985).

The adult study showed for the 120 subjects in 6 anatomical groups (from conductive to non-auditory reception in the brain) that the total scores for the two halves were very similar (difference = 0.6). For reversals, the groups that had significant reversals, each had about 1 more reversal on the second half. For the children's study (that I forgot about) there were 60 subjects with 15 for each age group 7-10 years. For these children the number of errors were equal for the 2 halves of the SSW and the reversals showed one more on the second half. So these 2 studies in the 80's (using the same version of the SSW test) showed that you could multiply by 2 for the total score and reversals and get a very good approximation.   Thus, doubling the errors will give you a good estimate of the total score, and number of reversals, in most of the cases. Therefore, I have considered these (half-test) extrapolated scores as "approximate" or "the best estimate". Of course, for those who are under 6 years of age the test is only based on the first 20 items so no additional testing is needed to compare the results to the age norms.

Over the years, to insure that the child (or adult) was able to perform at an acceptable level (e.g., not significantly fatigued, nor distracted), I would give just the first 20 items. The same is true when I saw a person struggling with the test; I would discontinue it, because I don't like to torture anyone. In those occasional cases I would just give 20 items and double the scores with impressive success. But, when the full test was

completed at another time; I have noticed that Ear and Order Effects and Type-As often were not accurate.

So that's why, I decided to run this little study to see how the children I tested recently have performed. I compared the results for children who were given all 40 items the first time with those tested with 20 and later on the other 20. Those of you who would like to see the details of this little study can read the next few pages. Those who just want the summary can go to the summary at the end of this article.

Procedures

The children who took the first 20 items and later took the last 20 (Gr20) were hard to find. I came up with a total of 8 Gr20 kids. For the typical Gr40 cases I took an incidental sample of children whose files were at hand at home and at work. Then a few more were needed so they were drawn at random from my "cold cases". Table 1 below provides the characteristics of the 2 groups.

Group	# Subjects	% Females	Initial Age	Test-Retest Interval
Gr40	30	37%	10.2	0
Gr20	8	38%	7.2	2.5 mo

Table 1. Mean demographics of those who took 40 SSW items initially and those who finished the last 20 items on another date.

Table 1 shows that both groups were about 2/3 male that is commonly seen for the general CAPD population. Although all subjects were 6 to18 years the Gr20 children were 3 years younger. This makes sense, especially if they were having a lot of difficulty on the first 20 items, or at least more so than the Gr40 children. While many of the Gr40 children would have had breaks during the test, it was felt that they had sufficient attention to complete the test on the same day. The Gr20 children demonstrated enough fatigue, inattention or frustration to require completion of the test on the following visit to avoid unnecessary errors or a very negative experience for the child. The interval between the initial test and completion was on average 72 days.

Results and Discussion

Table 2 shows the results for this split-half study for the 2 groups. We will first discuss errors on the 2 halves of the test when given on the same day for Gr40. As we found in the past the results were essentially the same with differences of less than 1 error for the 2 halves. This was the same for both the adults with lesions and the children with CAPD in the 2 previous studies. Therefore, it is a bit curious that for 8 subjects in Gr20 there were 4.7 more errors on the second half. This could be sample size or suggest younger children, or those with more severe scores, might tend to be more sensitive to the differences on the 2 halves. However, the take away message is doubling the first-half score should give a good approximation of the total score.

Groups	Errors 1st Half	Errors 2nd Half	Errors Total Test	Reversals 1st Half	Reversals 2nd Half	Reversals Total Test
Gr40						
mean	**12.7**	**13.1**	25.8	**1.1**	**2.0**	3.2
sd	**6.8**	**8.7**	7.7	**1.7**	**2.7**	2.3
r	**0.86 (<0.01)**			**0.82 (<0.01)**		
Gr20						
mean	**18.6**	**22.3**	40.9	**3.1**	**3.5**	6.6
sd	**8.3**	**8.2**	8.20	**3.6**	**3.1**	3.3

Table 2.   The means and standard deviations for errors and reversals. For Gr40 correlations are also shown. For our purposes the main interest is how the children performed on first vs. second halves of the test.

Unexpectedly high correlations were found for Gr40 for both errors and reversals. But just for fun (too small a sample for r) I ran the correlations for Gr20. The correlation for errors was 0.87 which was significant at the <0.01 level! For reversals the correlation was 0.75 and it was significant at the <0.05 level. The strong correlations for both groups are important because the means do not tell us if the children who performed well on one half were the same ones that performed well on the second half. These data certainly support that if we double the score it is likely to predict the performance for the person.

For number of reversals in Table 2, the Gr40 difference for the 2 halves was quite small. For the 8 subjects in Gr20 the difference was also somewhat poorer for the second half. When you compare the predicted number of reversals from the first half for the 8 children the prediction was correct for each of them and surprisingly when looking at the 30 subjects in Gr40 all but one would have been correctly rated as significant or not!

	Errors Significant		Reversals Significant	
	Hit	Miss	Hit	Miss
Gr40	30	0	29	1
Gr20	7	1	8	0

Table 3. Hit-Miss comparison for the predicted significance vs. the actual significance.

Table 3 shows that for the 38 subjects in this study that just one was misclassified as significant-or-not for number of errors and one was misclassified for their number of reversals. So in actual practice for these children the prediction was 97% accurate. Table 4 shows that the median errors were rather small.

	Gr40	Gr20
Errors	2	5
Reversals	1	2.5

Table 4. Median values for the difference between the predicted and actual error scores/reversals for 40 items and prediction based on the first half test.

Figure 1 provides a visual understanding of the first and second halves for errors and reversals. You can see that Gr40 had just about the same scores for the first and second half-tests for both errors and reversals, so doubling the scores should give us a very good prediction for significance and a good prediction of the actual number of errors and reversals. This interpretation is supported by Table 3. For reversals Gr20 predictions should be pretty good and they were (although one was misclassified for reversals). But the increase of almost 4 points for errors for Gr20 suggests there might be more misclassifications but there were none in this sample.
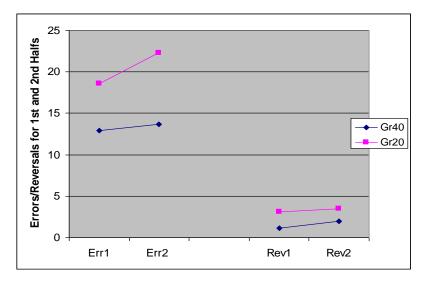

Figure 1. For both errors and reversals Gr40 provides a very good estimate for total score from the first half because both halves are so similar. For the 8 Gr20 subjects there were more errors on the second half of the SSW.

The take away message is, based on our findings it looks like the first half-test errors and reversals provide good-or-better estimates of the actual total tests. So when it appears necessary to test just 20 items one should be fairly confident that it will not misclassify the child. But of course the exact number may differ somewhat.

Ear and Order Effects
Now let's briefly consider Ear and Order Effects. First we should recognize that a split-half test compares one half-test with the other half (i.e., 20 vs. 20 instead of the full 40 items). Now a split-half Ear or Order Effect reduces the comparison to 10 items vs. 10 items. So reliability is surely reduced. So far I have not figured out how to predict the total Ear and Order Effect scores from the first 20 items. Unlike the total score and reversals Ear and Order for the half-tests are not significantly correlated, so it is difficult to make a reasonable prediction. Based on this you may not want to consider Ear and Order Effects when you only have half of the test. I have not worked on the Type-A.

Figure 2 shows the results for Ear Effect on the first 20 items. The figure shows that for both groups there was a clear negative score. This tells us that initially both groups had LEF scores that were about 5 points poorer than the REF. However, on the second half (for both Gr40 and Gr20) the score tended to go in the opposite direction, from about a -5 Ear Effect to a +2 Ear Effect. For the full test it would be an average of a -3 that would not be significant for most groups. This suggests that on average REF is a little poorer than LEF, but the important thing is that doubling the first half will <u>not</u> provide a decent approximation. For Order Effect it can be seen that doubling the score would be chancy. It will be interesting to study this situation further to see if there is a reasonable method to make a prediction from the first 20 items. At first glance the answer would be 'no'.
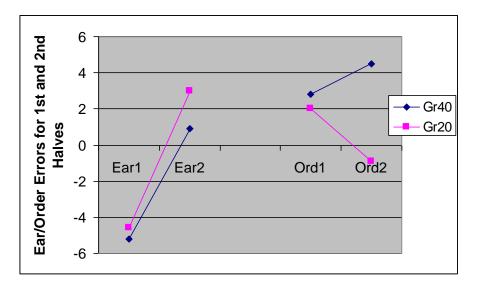

Figure 2. This shows the first half-test performance (1) with the second (2) for Ear and Order Effects.

It is likely, based on the mean scores that doubling the Ear or Order Effect would not provide a reasonable estimate of the full test.

8 Cardinal Numbers

It was interesting to study the 8 Cardinal Numbers for the 2 groups. Figure 3A shows the errors for the first 20 items.

Figure 3A shows the Condition error scores, which can be compared to the combined data in Figure 1. It was interesting to see the major peaks for column-F for LEF items. It seems logical that the left competing condition for left-ear-first items (that are used in the Type-A pattern) would be especially sensitive. It would make sense that the younger children would have more immature Integration (e.g., corpus callosum) function than the older children. The age difference between the 2 groups likely accounts for some of the difference. In addition, there were other young children who were able to

take all 40 items. So the Gr20 group may have more severe auditory problems or simply attention/anxiety etc. factors.
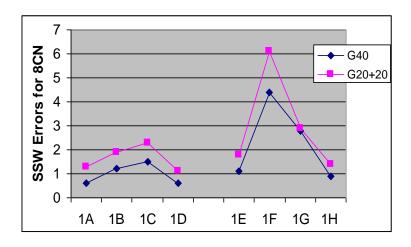

Figure 3A. The 8 CN for both groups on the first 20 items. Please ignore the '+20' in the caption above.

Figure 3B shows the 8 CN for the last 20 items for the Gr40 and Gr20 groups. The second half of the SSW had poorer scores for the REF items than the first half. Also the LC condition for REF items tend to reduce the Type-A as you no doubt have noticed. While the errors for both halves are quite close, the REF patterns are quite different. This contributes importantly to why we see a problem in predicting the Ear and Order Effects from one half to the other.
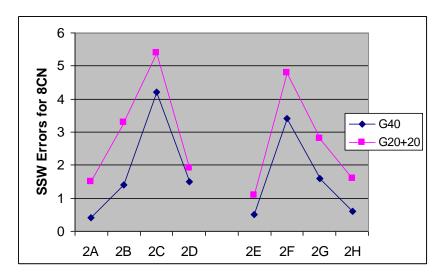

Figure 3B. This figure shows the data for the 2 groups on the second half of the SSW 8 CN. (Please ignore the '+20' in the caption above.)

Discussion

This study turned out to be quite interesting. It reinforced some of the observations over the years about the 2 halves of the SSW test and provided some new information. First it should be reiterated that if possible and profitable give all 40 items initially. However, when a person is getting sluggish, tired, inattentive, overloaded (e.g., giving strange errors) consider giving the person a break, to take a walk, get a drink, do jumping jacks or whatever. However, if it appears that the person cannot give a good amount attention and effort, then do not continue the test as it will be tainted and may not reflect their CAPD. In recent years I suspect that I have given half tests more frequently. The fatigue and frustration gives us a window on why these children so often are easily tired. Listening to speech (especially) is not an easy task for them, because they must first figure out what the teacher actually said and remember it. Then they find they have to try to catch up on what they have missed. So they are constantly working to figure things out, filling in the blanks and catching up with the rest of the class. When we see some responses that bear almost no relationship to the words presented it adds overload errors to the actual CAP errors. Of course, it is what often happens to them in life but our main job is to find out what their CAPD is like and not their fatigue.

You may have developed a strategy to keep the kids in the game. I give big or small breaks, have the child take a walk or get a drink (no alcohol! – but a power drink would be okay in this case). When a response is slow I stop the CD to give the person time to get ready for the next item. If it appears laborious I often give some extra time between items to get the best CAPD information. And, of course, I do jumping jacks with the kids so I get a little boost too.

If you do stop after 20 items of the SSW you should be able to estimate the full test errors by doubling the scores in order to compare them to the age norms (except for those below 6 years who only get 20 items). The same is true for reversals. There are 4 good reasons why we can have good confidence in our estimate for the full test errors/reversals from the first half test: 1) The scores for both half tests are quite similar so doubling the score is likely to be a good estimate based on the mean scores. 2) The scores for the individuals are highly correlated for the 2 half tests. 3) When the correction of doubling the errors and reversals we found in 97% of the cases the correction correctly predicted the significance of the scores. 4) Finally, we know for sure the score on the first half so we are only predicting for the other half.

Unfortunately, Ear and Order Effects do not have good correlations and the scores are often quite different on the second half of the test. For this reason it is a good idea to omit these 2 scores until the last 20 items can be obtained. I hope to have some further information on the Type-A in the future.

Summary and Conclusions

1. The readers of the *SSW Reports* know that the SSW is a powerful test and is also the most widely used test by audiologists who do central testing (Emanuel et al. 2011, pp 51-53).
2. The complexity of the test enables us to analyze a broad span of auditory processing skills.
3. Because of the complexity of the SSW we have been able to derive some important measures that are included in the 4 categories of the Buffalo Model.
4. We have known for a long time that the test's complexity leads to "internal dynamics" in which one aspect influences another. For example, a Type-A pattern will contribute to Ear and Order Effects. For this reason when there is a Type-A then void the significance of Ear and Order Effects.
5. A small percentage of children can only handle the first 20 items. For these individuals it is well to complete the other 20 items as soon after to make the best determination.
6. But, when that is not feasible and we have to make statements based on the first 20 items, we can make some very good predictions, certainly about the total score and number of reversals by simply doubling the scores.
7. I have found very good (not perfect) success in predicting the 4 Conditions, delays etc. by doubling their scores. It is not so critical that we predict the exact numbers but to make the correct diagnosis based on the significant factors.
8. However, Ear and Order Effects are too hard to predict from the first half. Also the Type-A numbers are likely to be reduced in the second half of the test.
9. Whenever possible do what you can to maintain a person's attention, reduce pressure and fatigue to enable a complete SSW test in the same visit.

References

Emanuel, D., Ficca, K. and Korczak, P. (2011). Survey of the diagnosis and management of Auditory Processing Disorder. Am. J. Audiol., 20, 48-60.

Katz, J. and Arndt, W.B. (1982). A split-half evaluation of the SSW test. In D. Arnst and J. Katz, The SSW Test: Development and Use. San Diego, CA., College Hill Press.

Katz, J. and Cummings, D. (1985). Split-half reliability of the SSW EC list with LD Children. SSW Reports, 7, 11-12.