

**RESULTS of an INTERNATIONAL ROUND TRIAL using
AGRITEST SB2 INSTRUMENTS to MEASURE STAPLE LENGTH AND STRENGTH
in accordance with IWTO-30**

B. P. Baxter, SGS Wool Testing Services, Wellington, New Zealand

SUMMARY

At the request of the raw wool group meeting held in Boston, in May 1997, an international round trial was carried out to assess the between-laboratories precision of IWTO-30 when the Agritest SB2 instruments are used to measure staple length and strength. The trial involved 6 instruments located in 5 separate laboratories in 3 countries. The results of the round trial show appreciably better precision than is documented currently in the Standard.

INTRODUCTION

Following a number of presentations to the raw wool group concerning the use of Agritest instruments within IWTO-30, a request was made at the Boston meeting, held in May 1997, that, in order to complete the information, a round trial should be carried out to determine the between-laboratories precision of IWTO-30 when Agritest SB2 instruments are used.

Following discussions with a number of possible participants and the chairman of the raw wool group, it was agreed that the trial would be widened beyond the initial intent of involving one additional laboratory. It was also agreed that for convenience the trial would utilise wools that had already been sampled for a prior trial between the two New Zealand test houses. The identity of the wools in this trial remained known only to the chairman of the raw wool group.

During the initial discussions it became clear that the trial would become more relevant if an additional laboratory could be included that was known to operate to certification standards. Hence the trial was delayed until additional instruments were manufactured and installed. Since one of the participating test houses also belonged to the ILRT group, it was also necessary that the trial design be formally approved by all members of this group.

Three of the final six instruments would be used in laboratories that operated only in the fleece testing arena, and hence it was agreed that some initial trials would be run prior to the main trial, to confirm that standards of sample preparation were adequate in all laboratories. This entailed some additional delay.

TRIAL DESIGN

The participating laboratories were (in alphabetical order) as follows:

- Australian Fibre Testing (York, West Australia)
- INTA (Bariloche, Argentina)
- INTA (Rawson, Argentina)

- SGS Wool Testing Services (Wellington, New Zealand) – 2 instruments
- University of New England, Armidale, NSW, Australia

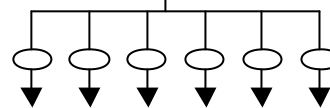
The trial samples comprised grab samples from each of 30 wools, of which 5 were blind-replicated, thereby giving 35 samples for measurement in the trial reported here.

The wools used were grab samples that had been retained from an earlier trial. The original identities were known only to the chairman of the raw wool group. 60 tufts were taken by SGS technicians from each grab sample for each instrument. The tufts were packaged each in separate compartments of specially-made tuft sample bags (which are used by SGS to transport tuft samples from branch laboratories to their main test laboratory). Each of the 6 sets of tufts from each sample was allocated at random to the 6 instruments:

Grab samples (30 in total):



Tuft samples (6 sets from each grab sample):



Instruments (at random):

A B C D E F

All tuft samples were sent out to the laboratories loosely rolled in their sample bags and packed in boxes to avoid any abnormal flattening of the staples. The laboratories were asked to store the samples until a few days prior to being required, when they were to prepare staples in accordance with IWTO-7.

Prior to the trial, each laboratory was supplied with the version of Appendix A to IWTO-30 that was approved at the Nice IWTO meeting. However, it should be noted that only 2 laboratories had access to the SLM instrument that is required to calibrate the SB2 for length. The other laboratories therefore relied on a factory calibration (based on the SLM reference).

All measurements were carried out on the trial samples during the last 2 weeks of March 1998. The laboratories used assumed core test results of 99.99% woolbase, 0.0% VM base, and category 1 as input data for the measurements. Each laboratory then reported their results directly to the chairman of the raw wool group, who converted the data using the real core test data, before transmitting the recalculated data to the author for final analysis.

ANALYSES

The analyses were conducted in general accordance with IWTO-0. Due to the difficulties of preparing large numbers of replicates from the available samples for all 6 instruments, it was agreed that only the 5 largest samples would be used for this purpose. This was considered adequate since the primary aim of the trial was to examine overall precision, with particular emphasis on between-laboratory precision. [Within-laboratory, between instrument precision was reported to this committee earlier¹]

Four analyses were therefore conducted:

- An examination of the standard deviation of differences from grand means for each of the 30 primary samples.

¹ Baxter, B.P. (1996), Repeatability data on length and strength (IWTO-30) determined by the Agritest Staple Length and Staple Breaker model 2 instruments, IWTO raw wool group report, Capetown

- Two-way analysis of variance with samples and laboratories as the factors
- Two-way analyses of variance on the samples for which there were blind replicates.
- An additional verification of the statistics was undertaken by examining all possible pairs of comparisons between instruments for each parameter (i.e. 15 pairs for each of 30 samples).

RESULTS

The raw data is tabulated in Appendix A.

Initial analyses were undertaken to identify outliers and to assess homogeneity of variances. No outliers were found for mean staple length (SL) or coefficient of variation of staple length (CVL).

However, the staple strength and position of break results from sample 29 were clearly anomalously variable, as may be seen in Appendix A. Careful examination of the retained portion of the sample showed it to contain a mix of very sound mature-aged fleece wool, some hogget wool, and some very tender necks, with a tendency to break very near the tip. It was therefore clear that the results from this sample would need to be excluded from the analyses. Unfortunately this sample had also been used to prepare one of the blind replicates, and the results from this other sample (31) were also widely variable in strength and position of break. Examination of the variances from the replicates showed unambiguously that the sample pair was an outlier. The results have therefore been excluded from all staple strength and position of break analyses. Nevertheless, it is worth noting that the inclusion of this sample would make only a very minor difference to the final overall precision estimates.

The overall mean results for the samples are summarised in Table 1. (figures in parentheses for staple strength and position of break include sample 29):

Table 1: Overall mean results for each instrument

Instr.	SL	CvL	SS	Tip	Mid	Base
1	87.0	13.6	20.2 (20.5)	29 (30)	58 (57)	13 (13)
2	86.2	13.5	21.3 (21.7)	26 (28)	61 (60)	13 (12)
3	85.9	13.7	21.2 (21.9)	26 (26)	61 (61)	14 (13)
4	85.9	13.3	22.8 (23.5)	26 (27)	62 (61)	13 (12)
5	86.3	13.6	21.5 (22.0)	24 (24)	63 (63)	13 (13)
6	87.1	13.5	19.9 (20.3)	30 (31)	57 (56)	13 (13)
Range	1.2	0.4	2.9 (3.2)	6 (7)	6 (7)	1 (1)

For comparison, the overall ranges between 4 laboratories in the trials that were used to determine the precision statistics currently in IWTO² were 0.8 mm SL, 0.4% CvL, and 3.1 N/ktex respectively. These results are therefore comparable.

95% confidence limits were calculated for each sample as follows:

$$95\% \text{ CL} = 1.96 * (\text{SD of differences individual lab results from the grand mean})$$

These are tabulated in Appendix B.

² Stubbs, C.G., Marler, J.W., and Bow, M.R. (1991), The precision of average staple length and average staple strength measurements, IWTO tech. report 9, Nice

In order to establish whether there are any significant level-dependencies in the precision values, these can be plotted, as is suggested in IWTO-0. The following sections show the precision plots for each of the relevant parameters. On each plot, the least squares linear regression line is shown, together with the 95% confidence boundaries for that regression. As an easy guideline, if a horizontal line can be drawn within the two confidence boundaries, it can be assumed that there is no obvious level-dependency with respect to the parameter plotted on the X axis. Also shown for convenience is the 95% confidence limit for 2 staple preparation operators currently shown in IWTO-30, where available.

Staple length:

Figure 1: 95% confidence limits on a sample by sample basis for Mean Staple Length

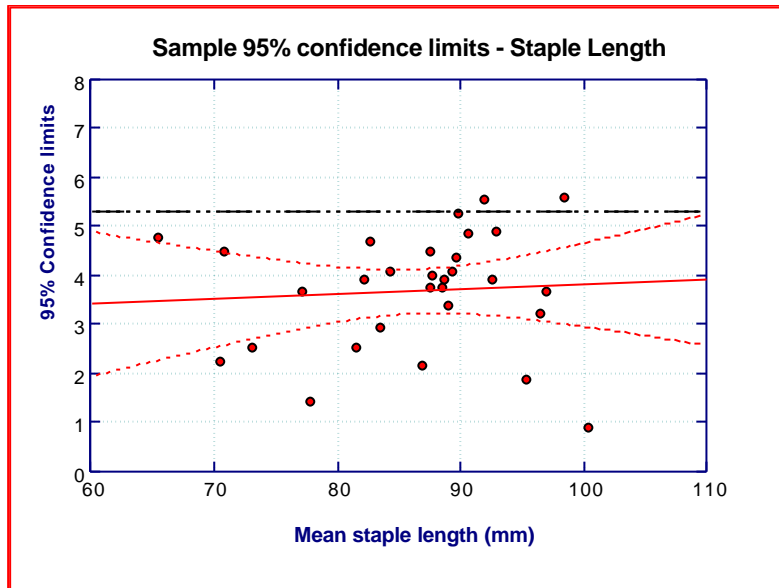
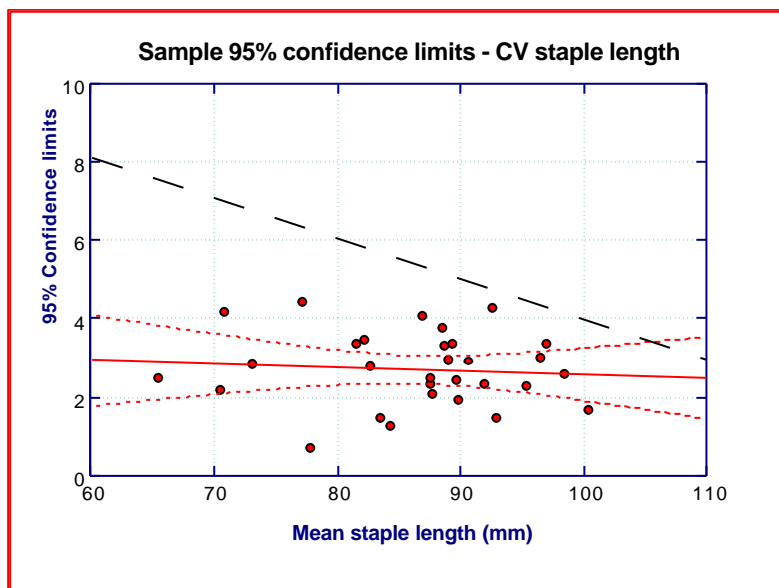


Fig. 1 shows that the precision is not level-dependent, and averages ± 3.7 mm. This compares very favourably with the figure of ± 5.3 mm currently shown in table 2 of IWTO-30.

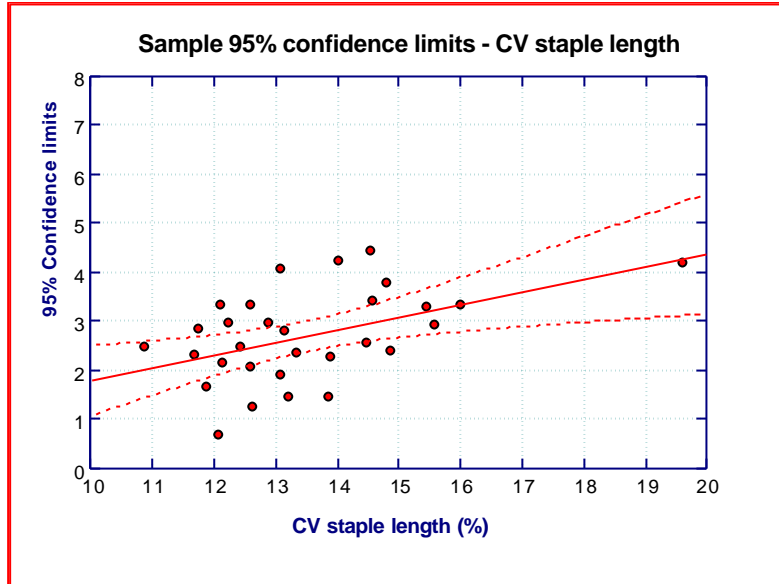
Coefficient of variation of staple length (CvL):

Figure 2: 95% confidence limits on a sample by sample basis for CvL as a function of SL



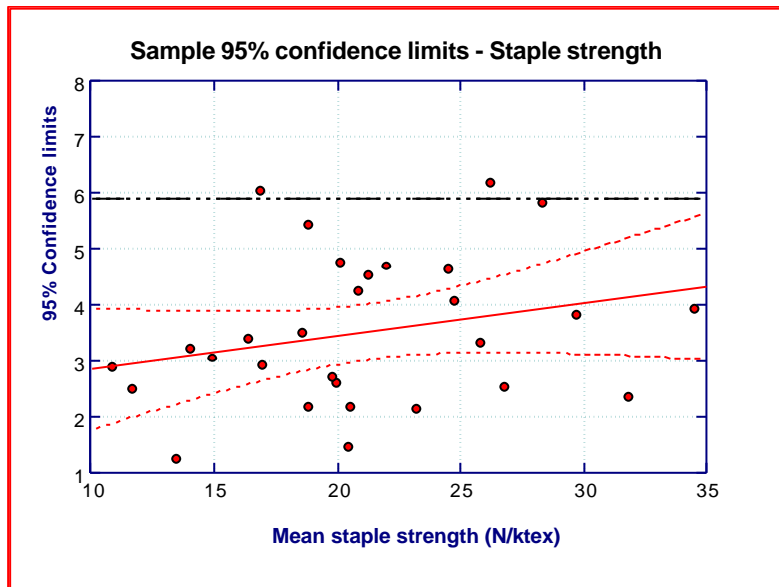
This shows no dependence on mean staple length, giving an average 95% confidence limit for CvL of $\pm 2.7\%$. This again compares very favourably with the figures in IWTO-30 of $\pm 5\%$ at a mean staple length of 90mm to $\pm 7\%$ at a mean staple length of 70 mm. There was a tendency for the precision to increase with increasing CvL, as is shown in Figure 3.

Figure 3: : 95% confidence limits on a sample by sample basis for CvL as a function of CvL



Staple strength:

Figure 4: 95% confidence limits on a sample by sample basis for Mean Staple Strength

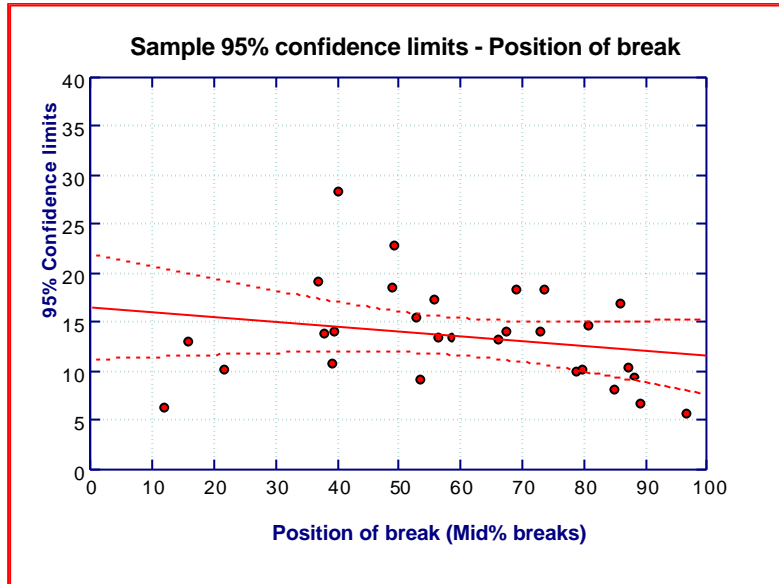


The trend shown is not significantly different to a zero gradient. The mean 95 % confidence limits for staple strength, ± 3.5 N/ktex, lie well within the value of ± 5.9 N/ktex quoted in IWTO-30, and the most extreme values encountered are virtually identical with this average value shown in the Standard. Taking into account the scope of the trial, the non-instrument factors which contribute to the variation between laboratories, and the geographic and operational diversity of the laboratories involved, these results offer strong support for the high precision of the SB2 technology (as alluded to in previous presentations).

Position of break (Mid%):

It is difficult to assess the position of break figures since there is no clear benchmark. The results are also influenced by the choice of samples, and in this case most of the samples tended towards tip or middle breaks rather than base breaks. This factor causes a skew on the distribution of PoB data. Since the Mid% breaks figure is the one used in the calculation of TEAM predictions, this has been chosen as also being most representative for this data set:

Figure 5: 95% confidence limits on a sample by sample basis for Position of Break (Mid%)



Once again this shows no level-dependency. The average sample by sample 95% confidence level for Mid% is $\pm 14\%$. (The corresponding figures for Tip% and Base% were 10% and 6%, respectively, but these should be used as guidance only in view of the skew of data in the samples used.) There is no published data with which to make a comparison. One notes that Standard requires the rounding of staple length and staple strength measurements to the nearest whole unit, i.e. 1mm or 1 N/ktex, and these figures are associated with confidence limits that are 5 or 6 times greater. In the case of PoB the Standard proposes rounding to the nearest 5%, suggestive of a confidence level of more than 25%. The observed value of 14% falls well below this figure.

ANALYSES OF VARIANCE – 2 WAY

Since there is no obvious or major level-dependency in the major parameters, 2 way analyses of variance may be undertaken on the differences from the grand means to obtain more precise estimates of the precision. In each case, homogeneity of variances analyses have been undertaken on both laboratories and samples effects, using Bartlett’s Chi-square, Bartlett-Box F, and Hartley’s F tests. Comments are made only when these tests indicate a significant non-null response.

Staple length:

Table 2: Two way ANOVA of average Staple Length

Due To	Sum of Squares	DoF	Mean Square	F-Stat	Signif
Main Effects	39.986	34	1.176	0.316	0.9999
Instrument	39.906	5	7.981	2.144	0.0634
Sample	0.079	29	0.003	0.001	1.0000
Explained	39.986	34	1.176	0.316	0.9999
Error	539.674	145	3.722		
Total	579.659	179	3.238		

This analysis confirms that differences between instruments on staple length are not significant, and the best estimate of the 95% confidence limits is $1.96 * \sqrt{\text{total mean square}} = \pm 3.5 \text{ mm}$.

Coefficient of variation of staple length:

Table 3: Two way ANOVA of CVL measurements

Due To	Sum of Squares	DoF	Mean Square	F-Stat	Signif
Main Effects	3.267	34	0.096	0.044	1.0000
Instrument	3.192	5	0.638	0.295	0.9149
Sample	0.076	29	0.003	0.001	1.0000
Explained	3.267	34	0.096	0.044	1.0000
Error	313.520	145	2.162		
Total	316.787	179	1.770		

This analysis confirms no significant difference between the instruments, and the best estimate of the 95% confidence limit for CvL is $\pm 2.6\%$.

Staple strength:

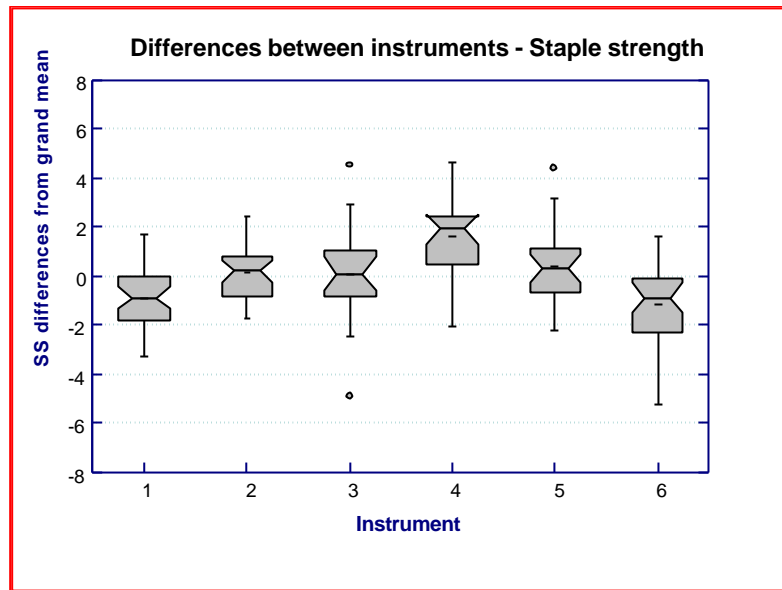
With all 30 samples included, the homogeneity of variance tests indicate that the variances are not homogenous for either samples or laboratories. (This also applies to the position of break figures) This can be taken as confirmation of the decision to exclude sample 29 from analyses for strength and PoB.

Table 4: Two way ANOVA of average Staple strength measurements

Due To	Sum of Squares	DoF	Mean Square	F-Stat	Signif
Main Effects	146.474	33	4.439	1.631	0.0272
Instrument	146.454	5	29.291	10.760	0.0000
Sample	0.020	28	0.001	0.000	1.0000
Explained	146.474	33	4.439	1.631	0.0272
Error	381.106	140	2.722		
Total	527.580	173	3.050		

Instrument is confirmed as a significant effect, although the variances are homogenous. Examination of the means and t intervals shows that instrument 4 may be regarded as an outlier. This can be illustrated by a box and whisker plot as in Fig. 6, where the notches represent t intervals.

Figure 6: Differences between instruments on staple strength measurements



The best estimate of the 95% confidence limit from this trial is ± 3.4 N/ktex, but after adjustment it is probable that the limits could be reduced to approximately ± 3.2 N/ktex. Nevertheless, even without adjustment, the overall confidence limit is very much less than that quoted in the Standard.

Position of break – Mid%:

Table 5: Two way ANOVA of position of break measurements

Due To	Sum of Squares	DoF	Mean Square	F-Stat	Signif
Main Effects	839.900	33	25.452	0.511	0.9868
Instrument	839.776	5	167.955	3.370	0.0066
Sample	0.125	28	0.004	0.000	1.0000
Explained	839.900	33	25.452	0.511	0.9868
Error	6978.057	140	49.843		
Total	7817.958	173	45.191		

Again, instrument is a statistically significant effect, but the variances are homogenous. No one instrument can be regarded as an outlier, and, as can be see from the table of mean values on page 3, the total variation is, in practical terms, very small. The best estimate of the 95% confidence limits from this trial is ± 13 %.

ANALYSES OF VARIANCE - REPLICATES

In view of the problems found with sample 29 and its corresponding replicate, the ANOVA's could only be carried out on 4 sets of replicates. This, however, was sufficient to provide estimates of within and between instrument components of variance, as shown in Table 6.

Table 6: Components of variance estimated from ANOVA on replicate measurements

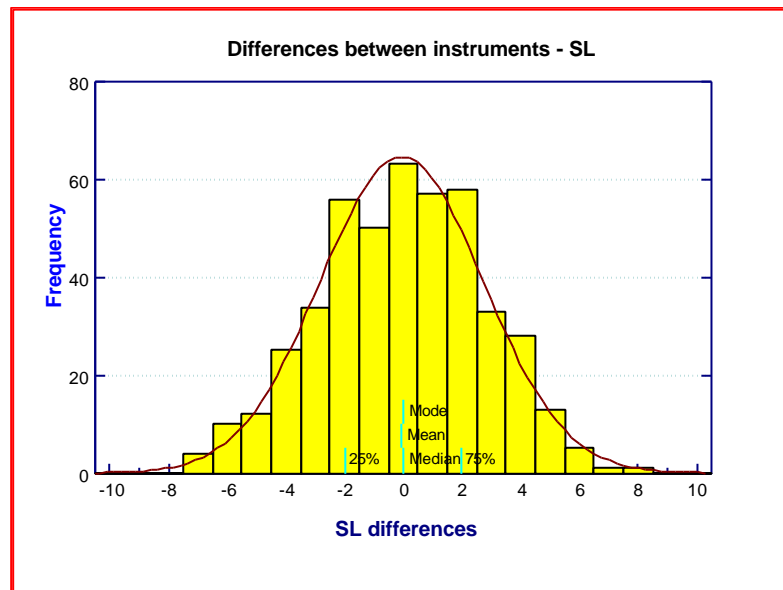
	Variance between	Variance within	95% CL
Staple length	0.94 mm ²	3.95 mm ²	4.3 mm
CV length	0.2 % ²	2.2 % ²	3.0 %.
Staple strength	0.7 (N/ktex) ²	1.5 (N/ktex) ²	2.9 N/ktex.
Mid% breaks	2 % ²	45 % ²	14%.

The significant point is that in each case the between-laboratories component of variance was a small part of the overall variance, and the overall confidence limits were all less than are quoted in the Standard.

INDIVIDUAL DIFFERENCES

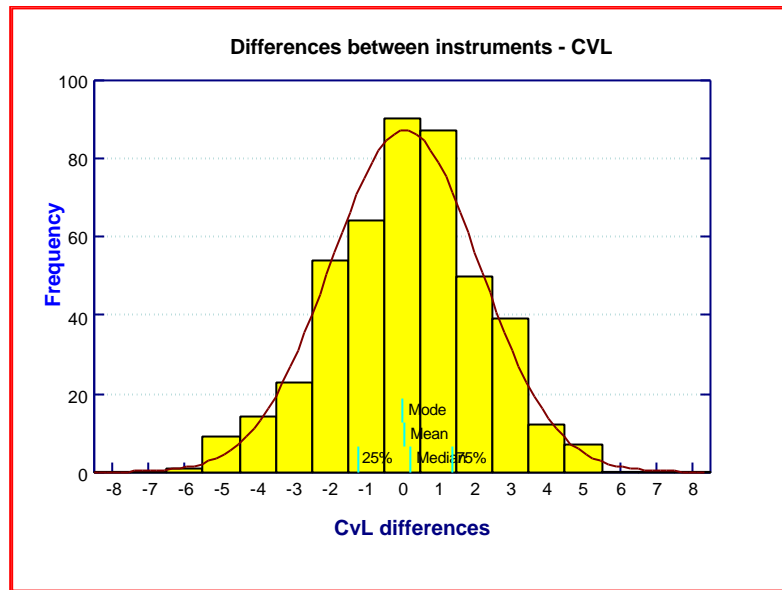
As verification of the statistics, for each sample the difference was calculated for all possible pairs of instruments for all parameters. This is equivalent to calculating the distribution of probable differences between laboratories. Two standard deviations either side of the mean should be equivalent to the “Maximum Retest Range” as defined in the Staple Test Regulations, and less than 1 in 20 results should fall outside the range ± (Maximum Retest Range). The distributions of differences are shown in figures 7 through 10.

Figure 7: Distribution of all possible differences on mean Staple Length



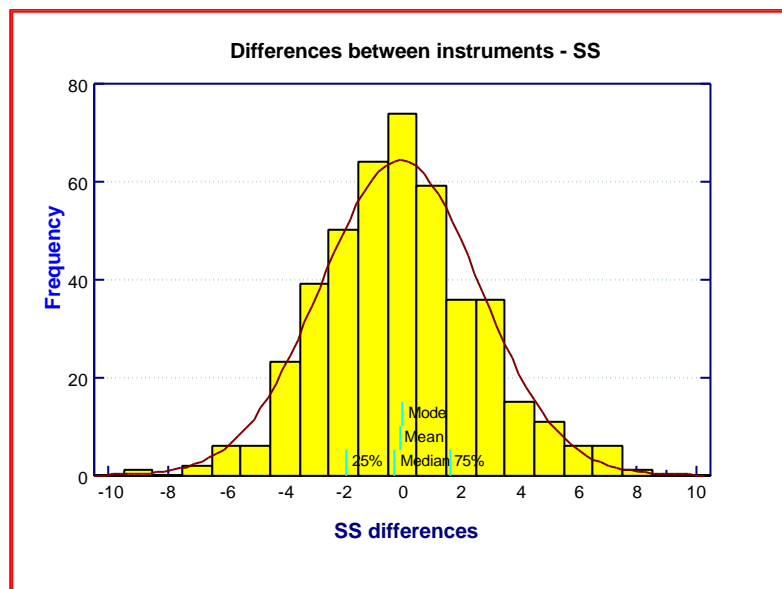
According to the staple test regulations, one would expect 1 in 20 results to fall outside the range of differences –9 to +9 mm. None of the 450 comparisons exceeded this range (the actual minimum and maximum differences were – 7 and + 8 mm respectively)

Figure 8: Distribution of all possible differences on CVL



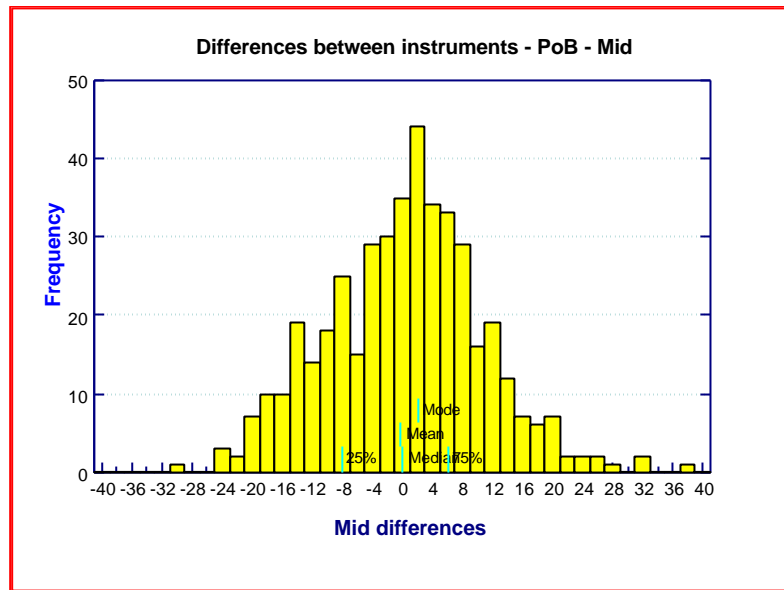
There are no criteria in the regulations for the maximum probable difference between two coefficient of variation results. We can, however, using the quoted mean of 6 % for the 95%CL value, calculate a value of 9 % (i.e. one in 20 results are expected to fall outside the range -9 to $+9$ %). None of the 450 differences approach this value – the minimum and maximum of the distribution are -6 and $+5$ % respectively.

Figure 9: Distribution of all possible differences on mean Staple Strength



The corresponding staple test regulation figure for the maximum probable difference on staple strength is 9 N/ktex. One in 20 results are expected to fall outside the range -9 to $+9$. In the above plot, no results fall outside this range, with the minimum and maximum being -9 and $+8$ respectively. (Even with the results included from the anomolous sample 29, the number of results exceeding the ± 9 range would only be approximately 1%)

Figure 10: Distribution of all possible differences on Position of Break (Mid%)



There is no corresponding reference information for PoB. The plot is shown for information only.

CONCLUSIONS

This international round trial involved 6 instruments in 5 separate laboratories in 3 countries. Only 2 of these laboratories (3 instruments in all) operate to certification standard. This makes this staple length and strength trial one of the most diverse, in terms of participation, published to date, and therefore provided a significant challenge for the technology in terms of its operational efficiency, ease of use and accuracy.

The best estimates of overall precision (95% CL) from this trial can be summarised as follows:

	Staple Length mm	CV Length %	Staple Strength N/ktex	PoB Mid%
This trial	3.5	2.6	3.4	13
IWTO-30, table 2	5.3	5 to 7	5.9	N/a

Irrespective of how the data from this trial is analysed, the precision statistics for the use of the Agritest SB2 under IWTO-30 were all significantly better than the figures currently published in the Standard. In view of these results, the SB2 clearly and unambiguously meets all the requirements for inclusion in IWTO-30 as a suitable instrument.

ACKNOWLEDGEMENTS

The author is very grateful to all participants, Agritest Ltd, and the chairman of the raw wool group, for their patience and considerable help with this trial.