How long is a piece of string?

Ralph Riddiough
Kyle Academy, Ayr, Scotland.

John H. McColl
University of Glasgow, Scotland.

◆INTRODUCTION ◆

THE glee on the faces of young children as they investigate how many of their friends can squeeze inside a cubic metre play frame, or the amazement of older pupils as they are reminded how large a cubic metre is, provide ample proof that learning about quantities can be great fun. The importance of the practical skill of estimating quantities is recognised in mathematics and science curricula with statements such as “make estimates of weight, time, volume and length”. In this article we describe a statistical activity based on the estimation of length, which provides interesting data for analysis and can develop an appreciation of experimental design. The approach taken can easily be adapted to suit the level of pupil. We have successfully used the idea as a junior secondary class activity, and as the basis of project work for senior pupils. The question of interest is “How well can we estimate length without the assistance of any measuring instruments?” In a further development of the idea, we pose the question “Does our ability to estimate length improve if, after each attempt, we are told how ‘well’ we are doing?”

◆COLLECTING THE DATA: ◆ A BASIC DESIGN

Take a large ball of string. Hide all measuring instruments. Invite a pupil to cut off a 20 cm length of string. Ask the same pupil to do this a further 9 times. Each time the pupil cuts off a piece of string, remove it from sight, but note the order in which the pieces of string are cut. After all the cutting is complete, measure the pieces of string to the nearest mm.

A demonstration of this basic design can stimulate a lively discussion of the issues involved. May we assume that each piece of string provides evidence of an independent estimate of the target length by the subject? Are there subtle influences at work? Should the ball of string be held by some sort of mechanical dispenser, or should the end of the string be unwound from the ball and presented to the subject by an assistant? Our approach has been to employ an assistant, but questions remain as to how much string should be unwound and in what manner it should be presented to the subject. Questions such as “Why 20 cm?” and “Why 10 cuts?” need to be asked. Furthermore, to what extent may we consider our chosen subjects to be representative of people in general?

◆ORGANISING, DISPLAYING◆ AND INTERPRETING THE DATA

On one occasion we asked members of a group of 14 and 15 year-old school pupils to choose a table tennis ball from a large bag. The balls in the bag were well mixed, and fourteen pupils drew balls with a special mark. (No doubt, other teachers will have their own favourite ways of randomly selecting a sub-group from a class.) These 14 pupils worked in pairs, one holding the string while the other estimated 20 cm ten times as described above. Measuring did not commence until both pupils in each pair had completed the string cut-

Table 1. Estimated lengths (cm) - no feedback.

Pupil	Cut 1	2	3	4	5	6	7	8	9	10
1	29.0	30.3	30.9	29.0	26.7	28.3	30.0	28.9	28.6	23.4
2	30.0	26.2	27.2	27.5	29.4	30.5	25.0	24.6	26.5	27.5
3	20.0	19.4	23.0	20.0	22.0	20.2	23.8	34.0	22.5	20.5
4	17.9	23.7	24.5	19.2	18.8	37.5	23.0	28.2	23.3	24.0
5	20.2	20.6	19.0	20.9	24.0	20.5	20.6	20.1	18.4	19.3
6	18.5	20.5	19.4	23.0	21.6	21.5	21.8	23.0	21.1	21.8
7	21.5	18.8	17.8	15.6	21.1	20.6	21.1	18.0	19.3	19.0
8	23.5	20.6	18.5	23.1	20.2	22.0	17.9	22.3	19.7	20.2
9	28.2	28.4	26.8	27.6	18.6	20.4	23.6	23.4	19.9	28.0
10	25.5	22.0	28.0	28.8	28.5	27.2	28.4	27.5	28.6	25.2
11	29.0	21.8	20.7	22.4	26.5	24.8	25.3	28.1	26.1	21.3
12	22.6	22.5	24.3	21.5	18.9	24.8	24.3	23.8	21.5	23.3
13	18.0	16.9	18.4	18.9	19.0	19.1	17.1	17.3	21.9	20.0
14	26.0	21.0	24.6	26.1	25.5	28.0	29.0	27.8	30.7	26.6

ting. Table 1 records the data produced by these 14 pupils.

Each pupil's performance can be summarised in a number of ways. A measure of bias is given by each individual's sample mean minus the target length. Most subjects have been found to over-estimate the required length, and we might review the design to see if there is any methodological reason why this should be so. The variability in an individual's performance is readily appreciated to be important as a measure of his or her consistency. Depending on the level of the class, this might be indicated by a simple measure such as the range, or a more sophisticated statistic like the standard deviation, of each subject's attempts. Since the attempts are recorded in order, it is easy to motivate pupils to draw a profile plot of an individual's attempts over time. Figure 1 illustrates the profiles of two pupils in this group.

As well as the variability within each pupil's performance, we are interested in the variability between pupils, and the efforts of these 14 pupils might be compared as shown in Figure 2.

In Figure 2, each line covers the range of estimates for an individual, and on this line the mean is clearly marked. Readers may recognise this type of plot as essentially an adaptation of the Hi-Lo-Close line graph for stock market prices, available with certain widely used computer packages. An alternative display, suitable for older pupils, is the multiple boxplot shown in Figure 3.

The order in which pupils' results are displayed on Figures 2 and 3 is completely arbitrary. Clearly, imposing an ordering (e.g. according to the pupil's mean or median estimate) would introduce a spurious appearance of trend. It is also worth discussing with pupils why it would not be appropriate to pool all the results from all the pupils and display them on,

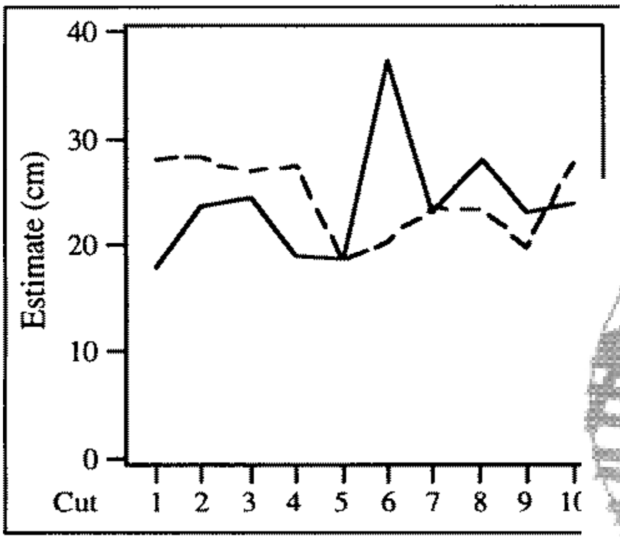
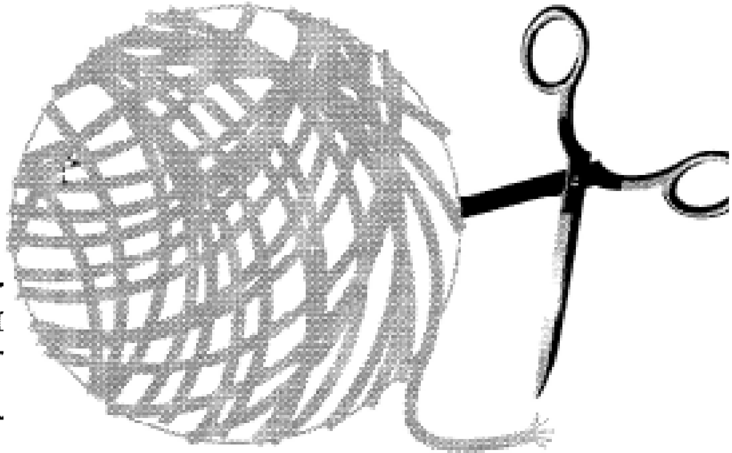


Fig 1. Individual profiles for two pupils.



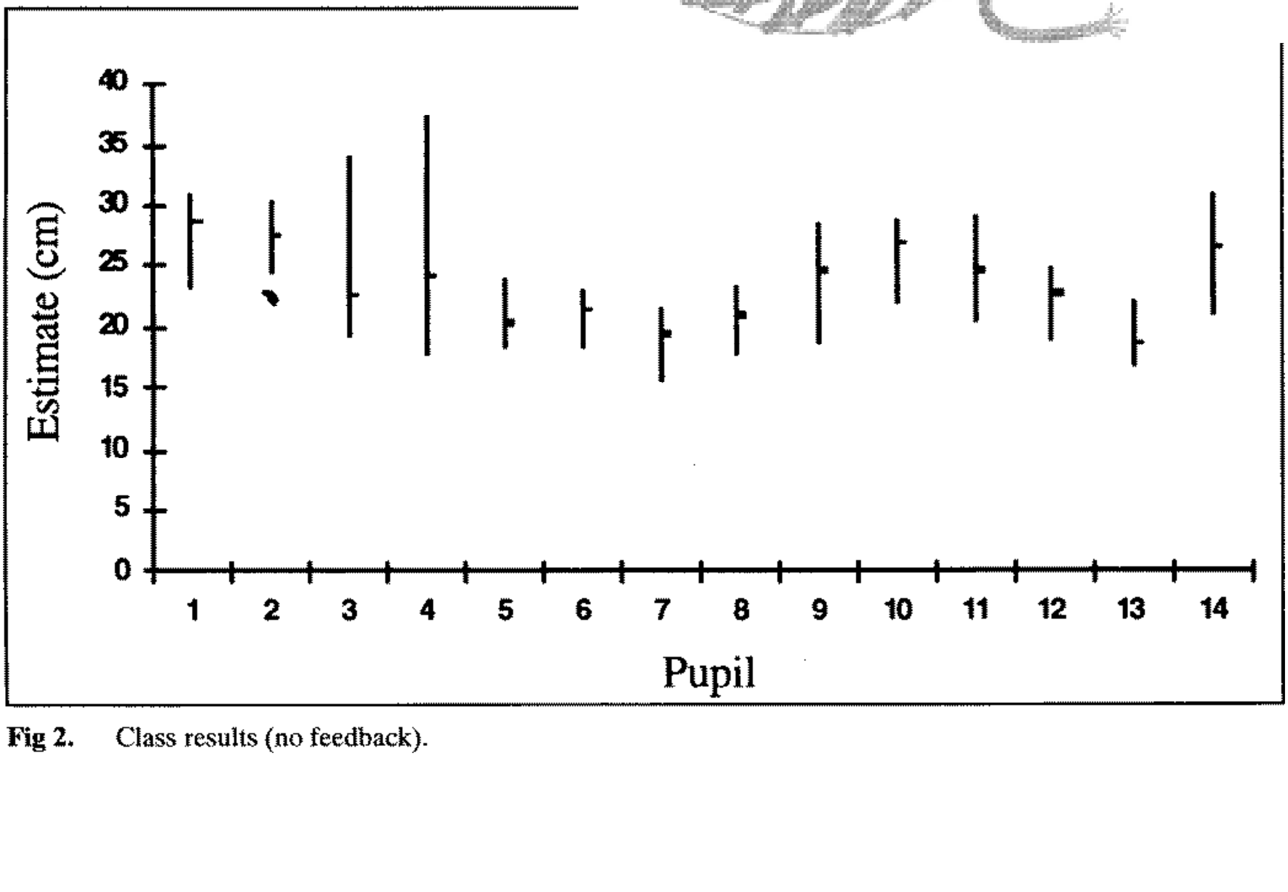


Fig 2. Class results (no feedback).

for example, a single boxplot. They might also be led to consider why it would be appropriate to plot, for example, the average results for the individuals on one graph. Here, we are beginning to approach the issue of "sources of variation", but the discussion need not become overly technical.

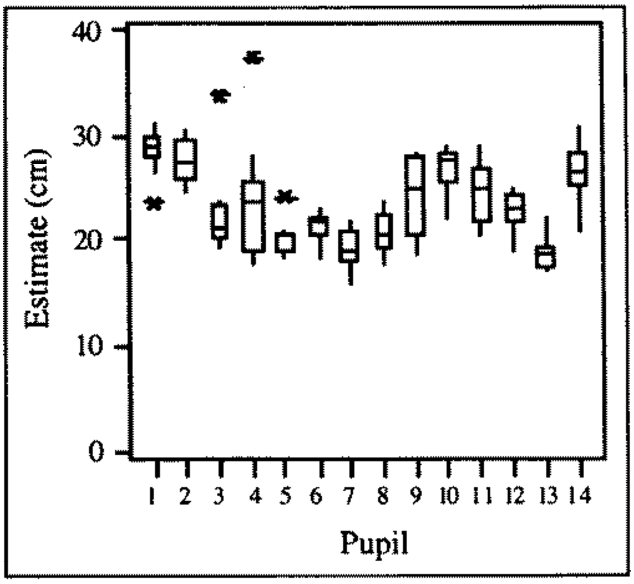


Fig 3. Class results (no feedback).

◆INCORPORATING FEEDBACK ◆

We now turn our attention to the second question we posed: "Does our ability to estimate length improve if, after each attempt, we are told how 'well' we are doing?" Forty-five pupils discovered that the table tennis ball they had drawn from the bag did not have the special mark. They were formed into groups of three, one of whom was designated a Recorder who did not make any estimates. The other two pupils followed the above procedure but this time, after each cut, the Recorder measured the piece of string out of view of the subject, who was told the result before making another cut. (Since Recorders do not make any estimates themselves, teachers will probably want to allocate at least 50% more pupils to the group with feedback than to the group without feedback.)

Table 2 lists the data produced by 30 pupils following this revised scheme. A glance at the data reveals that feedback improves performance in two ways : there is less bias and less variability in these data. While pupils might start by considering displays of the data that are similar to Figures 2 and 3, the~should be led to realise that these are not so appropriate for this group. From the simplest point of view, such plots miss out the element of improvement over time that is now an important part of the story. Older pupils should realise that an individual's attempts are no longer independent of one another. Individual profiles are very useful here.

Table 2. Estimated lengths (cm) - with feedback.

Pupil	Cut									
	1	2	3	4	5	6	7	8	9	10
1	20.3	21.5	17.6	19.1	19.5	21.2	26.0	21.9	19.0	17.6
2	16.5	18.5	21.3	24.0	20.8	22.5	18.9	21.4	21.0	21.8
3	31.0	25.0	19.0	19.7	20.4	19.8	20.2	19.5	24.0	21.5
4	19.0	19.8	20.0	19.5	24.0	20.1	21.0	22.5	19.9	19.8
5	23.3	24.3	22.1	23.6	17.6	18.6	20.1	20.8	15.4	20.2
6	20.0	20.7	18.4	21.7	21.1	17.2	20.9	21.6	22.4	19.1
7	29.1	31.0	23.5	17.7	20.4	27.7	15.0	20.2	20.1	20.2
8	15.9	20.7	18.9	20.8	22.5	19.0	21.0	22.0	21.0	18.5
9	24.0	20.5	17.5	18.5	18.5	19.0	23.5	20.7	22.5	21.2
10	18.5	18.9	20.5	20.5	19.0	17.0	19.3	18.2	23.5	21.5
11	27.0	21.0	19.8	23.4	16.3	22.1	26.6	22.3	18.0	20.1
12	21.2	20.5	17.9	20.7	21.3	21.2	20.7	20.4	26.1	20.0
13	23.0	18.5	22.1	19.0	20.0	20.0	22.2	18.0	19.1	19.0
14	17.6	17.4	20.0	22.7	22.4	25.0	19.0	22.2	22.0	20.0
15	15.3	18.2	20.4	19.9	18.6	20.2	19.0	19.8	20.1	20.3
16	18.5	20.6	19.4	19.4	21.2	18.8	20.3	20.6	19.3	20.9
17	35.9	31.0	30.0	22.4	22.9	24.1	18.2	20.7	18.0	19.8
18	21.6	23.6	15.8	24.4	20.4	18.6	21.4	24.2	19.8	20.8
19	29.5	18.4	16.8	20.2	15.0	22.5	18.1	18.3	23.2	21.2
20	18.5	19.0	20.2	19.8	20.7	19.1	18.9	18.2	17.5	18.5
21	24.1	20.0	18.4	20.4	19.5	21.2	19.8	20.7	20.4	20.8
22	19.5	19.9	18.6	19.9	21.9	20.3	17.5	19.5	21.6	18.3
23	26.5	23.9	22.2	20.5	19.6	19.7	20.4	18.4	21.8	20.1
24	20.6	20.1	18.3	21.0	20.0	17.6	22.1	20.1	22.5	19.8
25	22.6	22.1	19.0	21.3	19.7	23.7	21.4	16.9	16.7	23.0
26	19.7	16.4	17.4	16.5	17.2	16.5	20.1	18.2	20.3	19.6
27	33.0	19.0	22.0	21.0	21.0	22.0	20.0	22.0	21.0	17.0
28	21.0	19.0	25.0	23.0	17.0	20.0	19.0	19.0	20.0	18.0
29	20.6	16.5	14.9	20.3	18.5	22.0	23.3	19.1	25.5	20.9
30	21.6	22.3	18.5	20.3	20.1	20.1	18.5	18.8	19.5	24.2

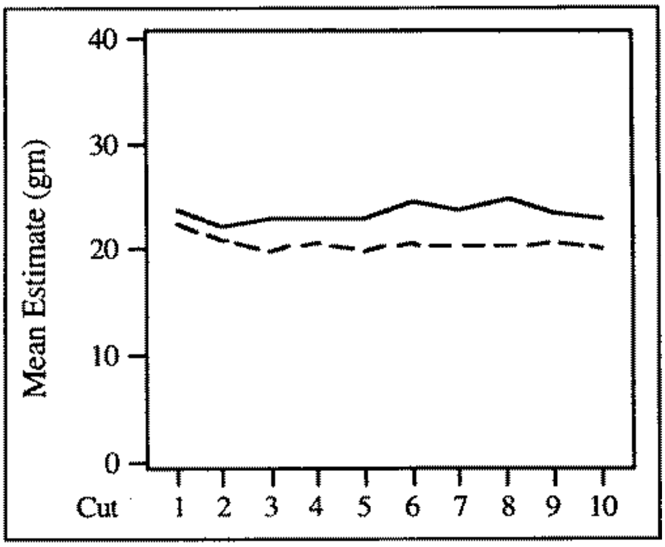


Fig 4. Mean profile plots to compare groups.

Mean profile plots are an effective way to display the data for this group of pupils and to compare the performance of the two groups. How did the performance of the 30 pupils who received feedback compare with that of the 14 pupils who were not told how 'well' they were doing? A mean estimate for each of the ten cuts is calculated for each group and plotted as shown in Figure 4.

The profile shown as a solid line is for the group that did not receive feedback, and on average they are clearly overestimating the target length at each attempt. The profile plotted as a broken line shows that, at least on average, the group with feedback very quickly "homed in" on the target length. Figure 5 compares the standard deviations for the two groups at each attempt, and shows that as well as becoming more accurate, those who were given feedback also became more consistent with one another.

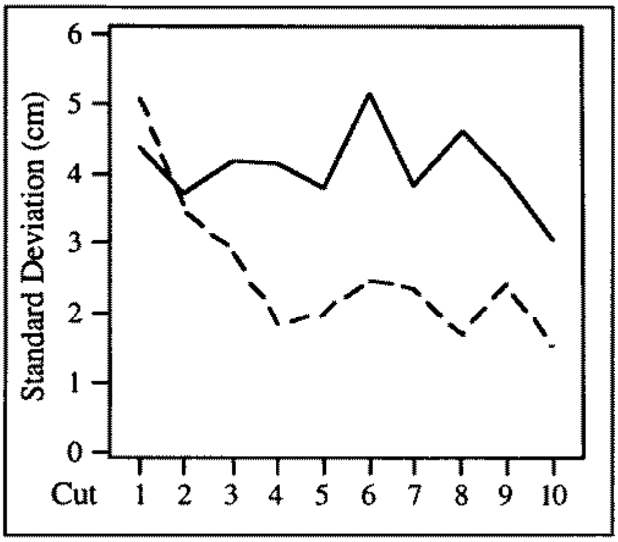


Fig 5. Profile plot to compare standard deviations.

◆DESIGN CONSIDERATIONS ◆

When comparing the two groups, the teacher can arouse in pupils the objection "How do we know that the individuals in the two groups are comparable?" Could it be that the data in Table 2 came from pupils who had a greater talent for estimating length? Such an objection can motivate consideration of the role of randomisation, and of more sophisticated experimental designs.

For example all individuals could be tested before, during and after the intervention of providing feedback. This would reveal whether or not there is a residual effect once feedback is withdrawn. An alternative might be provided by a multiple base-line design where subjects are assigned at random to a number of treatment groups. Here, each group commences estimation without feedback, but the intervention of providing feedback is introduced at different stages during the experiment, as illustrated in Figure 6. Implementing this design in the classroom will provide data that give rise to interesting profile plots but are difficult to analyse formally.

Group 1 N N N | F F F F F F F F F
Group 2 N N N N N N | F F F F F F
Group3 N N N N N N N N N | F F F

Fig 6. Multiple base-line design (N = no feedback, F = feedback)

Although cutting string is an ethically neutral activity, pupils might be led to reflect on more sensitive situations where removing or delaying an intervention thought to be beneficial could pose a moral dilemma.

We have also used this activity as the basis of project work with senior pupils. In this case, the investigators gathered data from subjects individually. This avoided some subjects benefiting from watching other pupils work before trying the experiment themselves. For example, the data in Table 2 are arranged in pairs (so that Pupils 1 and 2 worked together). The odd-numbered pupil always worked before his or her even-numbered partner. There is some suggestion that the second partner to cut was nearer the target, especially in early attempts. (If this is so, then teachers might wish to have three times as many pupils in the group with feedback than in the group without. There would then be scope for comparing just the results for the first partners to cut in the former group with all the results from the latter group.)

With older pupils, especially using more complicated experimental designs, the formal analysis of data must be handled carefully. A naive form of repeated-measures analysis is recommended, where each subject's response is the result from one Cut (e.g. Cut 10) or the average of a few Cuts (e.g. Cuts 8-10). There is also scope for paired comparisons (e.g. Cut 1 with Cut 10).

As an alternative to string and scissors, we have also successfully used paper and pencil. Sheets of A3 paper are cut into 4 strips. A black line is drawn down the centre of each strip and one end of the line is marked with a '0'. Subjects are asked to imagine that the line is a ruler with its scale markings missing. They are asked to place a pencil mark on the line where they judge the 20cm scale marking should be. We feel that the experiment with string and scissors has a greater impact on pupils, but colleagues have told us that they are unwilling to provide pupils with potentially offensive weapons!

◆CONCLUSION ◆

The experimental activity described here has several advantages to commend it. Teachers can easily adapt it to suit a wide range of levels. The data produced can motivate the practise of a wide variety of statistical techniques chosen to suit the pupils’ stage. Important statistical concepts may also be introduced am reinforced, such as the notion of variance as a measure of consistency, independence and dependence, sources of variation. Further investigations of factors such a age and gender are easily accommodated. The discussion of experimental design can be treated lightly or in depth. The materials required are inexpensive. Most important of all, learning can be fun