

## Chapter 3 Setting Up a New Data Set

• 14 November 2000 •

---

### Contents

Workbooks, worksheets, rows & columns .....	1
Piet Abik's class data set .....	2
A UCV data set.....	6
A class survey .....	9
A large-scale survey .....	10
Setting up a new workbook.....	12
Entering data .....	12
Creating the CCs lines.....	14
Making changes to CCs lines .....	15
Just Freqs? .....	15
Getting results .....	16

---

The last chapter was a breeze, wasn't it? You didn't have to enter any test results, they were already there. Nor did you have to create any job definition statements in the CCs worksheet as they too were supplied.

We're about to change this scene, and get you into setting up your own job. We'll start by walking you through some examples, real-life data sets from a variety of countries.

As we do this, our assumption will be that you have read through the last chapter. If you're being naughty, and starting out here instead of there, you may run into problems. 'Nuff said.

### **Workbooks, worksheets, rows & columns**

In the last chapter you read how Lertap 5 is based on Excel, a spreadsheet program, one of the four or five applications found in the Microsoft Office suite. Microsoft Word is another Office application, probably the most popular one.

Excel's files are referred to as workbooks. A workbook is a collection, a set, of worksheets. Each worksheet is a spreadsheet. Spreadsheets are based on "pages" of rows and columns. Rows run across the page (or screen), while columns run up and down. The intersection of a row and column defines what is called a "cell".

These rows and columns are very much like the rows and columns of the tables which may be used in Word. If you've used tables in Word, you know that you can do all sorts of things with the table's rows and columns, such as hide their gridlines, change column widths, change row heights, and add background colours. You can do the same, and more, with the rows and columns of an Excel worksheet.

One of the big differences between Word's tables and Excel's worksheets is that the rows and columns in an Excel sheet have labels, or "headers", such as 2 for row 2, 3 for row 3, and so forth. Excel's columns may be headed by upper case letters, or by digits, depending on the user's preference. If letters are used, Excel says it's using the "A1" referencing system, and the first column's header will be "A", the second's "B". Under this header system, the cell found in the upper left corner of a worksheet is called cell "A1", denoting the intersection of column "A" with row "1".

On the other hand, if column headers are digits, Excel says it's using the "R1C1" referencing system, and what was cell "A1" becomes known as cell "R1C1", for row 1 with column 1. The cell referencing system, A1 or R1C1, is an option which users set in Excel. Lertap uses the R1C1 system.

When you open a new document in *Word*, what is it usually called? Something like Document1, or Document2, a name which sticks until you save the new document with a different name. Under Windows and NT, Word documents almost always have an extension of "doc".

When you open a new workbook in Excel, it will be called Book1, or Book2, or BookX until you save it with a different name. Under Windows and NT, Excel workbooks usually have an extension of "xls" (the "XL" means Excel, while the "s" means "sheets").

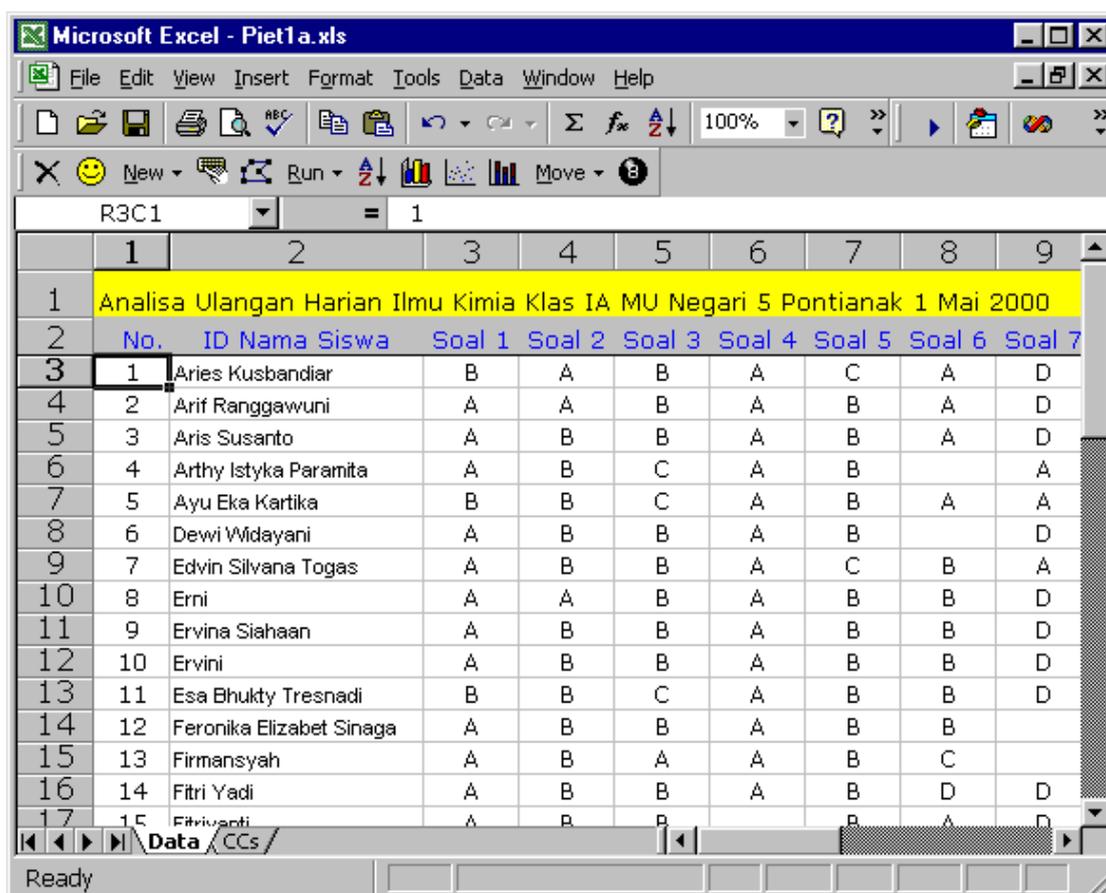
Excel is a very powerful system, and a flexible one. Users of Lertap 5 will find that increasing their familiarity with Excel will be of real benefit. Do you know how to copy a single worksheet from a given workbook to another workbook? It's easy, as is renaming worksheets. How about something as simple as changing the width of an Excel column, or the height of an Excel row? Easy. How about setting the background colour of a cell, or a row, or a column? Easy. How to find out how to do these things? Excel's on-line Help is one way, and a good one—but it can take some practice to master—be patient. How about an introductory book on the use of Excel? There are many excellent ones (for example, see Online Press (1997)).

Enough of that, let's get cooking, eh?

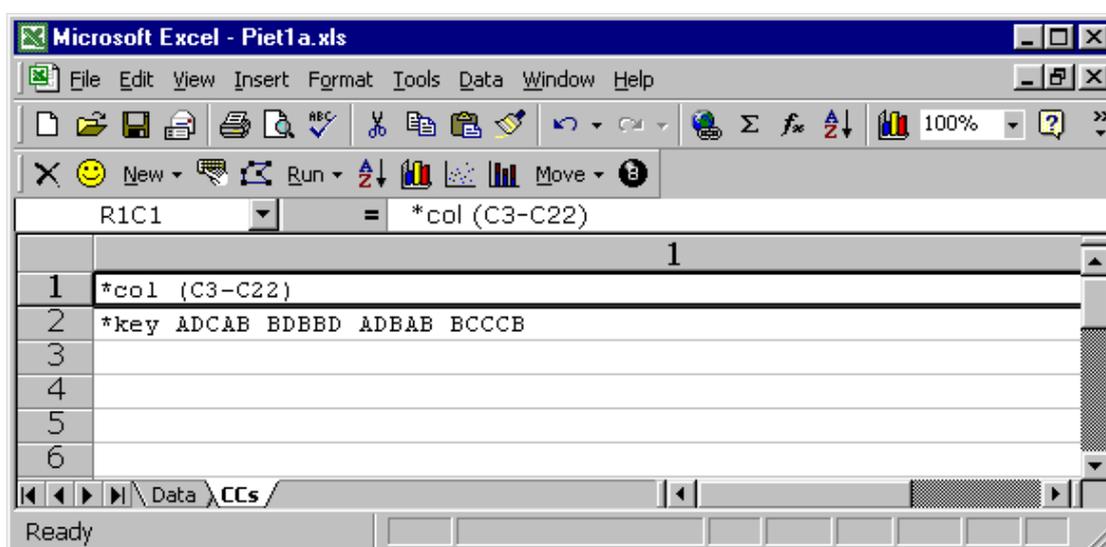
### **Piet Abik's class data set**

Piet Abik is a senior teacher of science at a large high school in Pontianak, an Indonesian city on Borneo's west coast. He gave one of his classes a 20-item multiple-choice test on chemistry, and used Lertap to look at results.

Below we show the partial contents of Piet's Data and CCs worksheets:



Analisa Ulangan Harian Ilmu Kimia Klas IA MU Negari 5 Pontianak 1 Mai 2000									
No.	ID Nama Siswa	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Soal 6	Soal 7	
1	Aries Kusbandiar	B	A	B	A	C	A	D	
2	Arif Ranggawuni	A	A	B	A	B	A	D	
3	Aris Susanto	A	B	B	A	B	A	D	
4	Arthy Istyka Paramita	A	B	C	A	B		A	
5	Ayu Eka Kartika	B	B	C	A	B	A	A	
6	Dewi Widayani	A	B	B	A	B		D	
7	Edvin Silvana Togas	A	B	B	A	C	B	A	
8	Erni	A	A	B	A	B	B	D	
9	Ervina Siahaan	A	B	B	A	B	B	D	
10	Ervini	A	B	B	A	B	B	D	
11	Esa Bhukty Tresnadi	B	B	C	A	B	B	D	
12	Feronika Elizabet Sinaga	A	B	B	A	B	B		
13	Firmansyah	A	B	A	A	B	C		
14	Fitri Yadi	A	B	B	A	B	D	D	
15	Fitriyanti	A	B	B	A	B	A	D	



Analisa Ulangan Harian Ilmu Kimia Klas IA MU Negari 5 Pontianak 1 Mai 2000									
No.	ID Nama Siswa	Soal 1	Soal 2	Soal 3	Soal 4	Soal 5	Soal 6	Soal 7	
1	Aries Kusbandiar	B	A	B	A	C	A	D	
2	Arif Ranggawuni	A	A	B	A	B	A	D	
3	Aris Susanto	A	B	B	A	B	A	D	
4	Arthy Istyka Paramita	A	B	C	A	B		A	
5	Ayu Eka Kartika	B	B	C	A	B	A	A	
6	Dewi Widayani	A	B	B	A	B		D	
7	Edvin Silvana Togas	A	B	B	A	C	B	A	
8	Erni	A	A	B	A	B	B	D	
9	Ervina Siahaan	A	B	B	A	B	B	D	
10	Ervini	A	B	B	A	B	B	D	
11	Esa Bhukty Tresnadi	B	B	C	A	B	B	D	
12	Feronika Elizabet Sinaga	A	B	B	A	B	B		
13	Firmansyah	A	B	A	A	B	C		
14	Fitri Yadi	A	B	B	A	B	D	D	
15	Fitriyanti	A	B	B	A	B	A	D	

Every Lertap workbook has worksheets named Data and CCs. Test results go into the Data sheet, while job definition statements go into CCs.

Look at Piet's Data sheet for a moment. The first row has its background colour set to yellow, and contains a title. The background colour of the second row is grey, and this row is used to contain headers for the columns of data.

Piet headed his columns of item responses with the word "Soal" ("Item", in English), followed by a space and then the item's sequential number. This is fine. However, had we set up this data set at Lertap headquarters, we would not have put the space after Soal. Some of the Lertap output is easier to read without that space. For example, the bands of item difficulties and discrimination coefficients found in the Stats1f sheet look better if the items have short names, and no spaces.

Another reason for eliminating spaces from item headers is that this makes it easier to pass the data to other systems, such as SPSS, a popular statistics package. SPSS will call each of the item columns a "variable", and it doesn't like spaces in variable names. (Nor does it like variable names which begin with a digit, or ones which are longer than eight characters.)

Had we been in Pontianak, sitting next to Piet as he set up his Data sheet, we would have urged him to refer to the items as "S1", "S2", "S3" and so on, not "Soal 1", "Soal 2", and "Soal 3". However, this is entirely up to users. Lertap imposes no restrictions on item names. In fact, as we'll see in a later example, you don't even have to have headers above the columns of item responses—this is not recommended, but you can leave the header row empty and Lertap will invent its own item numbers.

This is a good spot to mention that Excel has a superb shortcut for putting item headers into the Data sheet. Above Piet has his first item header in cell R3C3 (row 3, column 3), where he's typed "Soal 1". Once this is done, Excel makes it possible for "Soal 2", "Soal 3", ..., "Soal 20" to be automatically entered in the columns to the right. How? Select cell R3C3 by clicking on it. A box forms around the cell to indicate that it's selected. Grab the lower right corner of this box and drag it across the columns to the right. This is called "extending the series". If it doesn't work for you, use Excel's on-line Help for more instructions; suitable keywords are "extend series".

What else can we note about Piet's Data worksheet? He's used the first two columns as a record "No.", and as an ID field with student names ("ID Nama Siswa"). The names used in the second column will be picked up by Lertap and used to report scores. Had the second column been headed "Name Siswa", this would not have happened—the scores would have simply been numbered. When Lertap finds a column header beginning with the letters "ID" it uses the information in this column to label the results in its Scores worksheet. Note that this column must be either the first one, or the second one.

Piet's "No." column is just a sequential number. He could have used Excel's "extend series" shortcut to fill this column. There's no need for this column from Lertap's point of view, but Piet obviously found it useful for something.

Now let's get to the CCs sheet shown above—this is where the data analysis was defined. The first line, \*col (C3-C22), tells Lertap that item responses start in column 3 and end in column 22, a span of 20 columns. The \*key line has 20 letters. This is the string of right answers to the items. The correct answer to the first item, for example, was "A". The correct answer to the sixth item was "B". The correct answer to the 17<sup>th</sup> item was "C". The spaces which Piet typed in this line are not necessary, but they help read the line, we use them too.

With his results in the Data sheet, and job definition statements in the CCs sheet, what's necessary to get results from Lertap? This is what we demonstrated in the last chapter—Lertap's Run option is used twice, first to "Interpret CCs lines", and

then to get the “Elmillion item analysis”. This will make all those delectable, luscious worksheets of scores and statistics which you read about in Chapter 2.

Okay, shortly we’ll flash up another example. But before we do, we respond to those readers who want to know why Piet’s CCs sheet did not include a \*sub card.

What’s a \*sub card? You saw examples of them in Chapter 2. Piet’s CCs lines could have been these:

```
*col (C3-C22)
*sub title=(ChemTest), name=(Class test on 20 May)
*key ADCAB BDBBD ADBAB BCCCB
```

Here (above) a \*sub line, often called a “card” by veteran Lertap users, has been used to provide a title for the test scores, and a name for the item analysis. The title will be seen in Lertap’s Scores worksheet, while the name will be found at the top of all the item statistics sheets, such as Stats1f, Stats1b, and Stats1ul.

If a subtest has no title, Lertap will label its scores as “Test1”, and will give it the same name, “Test1”, when it creates all its statistical summaries.

Piet’s CCs lines could have been these:

```
*col (C3-C22)
*sub res=(A,B,C,D), title=(ChemTest), name=(Class test on 20 May)
*key ADCAB BDBBD ADBAB BCCCB
```

Now Piet is telling Lertap that the items in his multiple-choice test all used A B C and D as possible responses. However, this is Lertap’s default assumption for cognitive subtests<sup>1</sup>. The res=() declaration is required on a \*sub line only when the test’s items use a different set of responses. In the last chapter, for example, there was a \*sub card with Res=(A,B,C,D,E,F). This told Lertap that the items used as many as six different responses, from A through F.

Piet’s CCs lines could have been these:

```
*col (C3-C22)
*sub title=(ChemTest), name=(Class test on 20 May), per
*key ADCAB BDBBD ADBAB BCCCB
```

Now he’s using the “per” control word on the \*sub card, see? This asks Lertap to create a percentage score for each person, and causes the Scores worksheet to look like this:

---

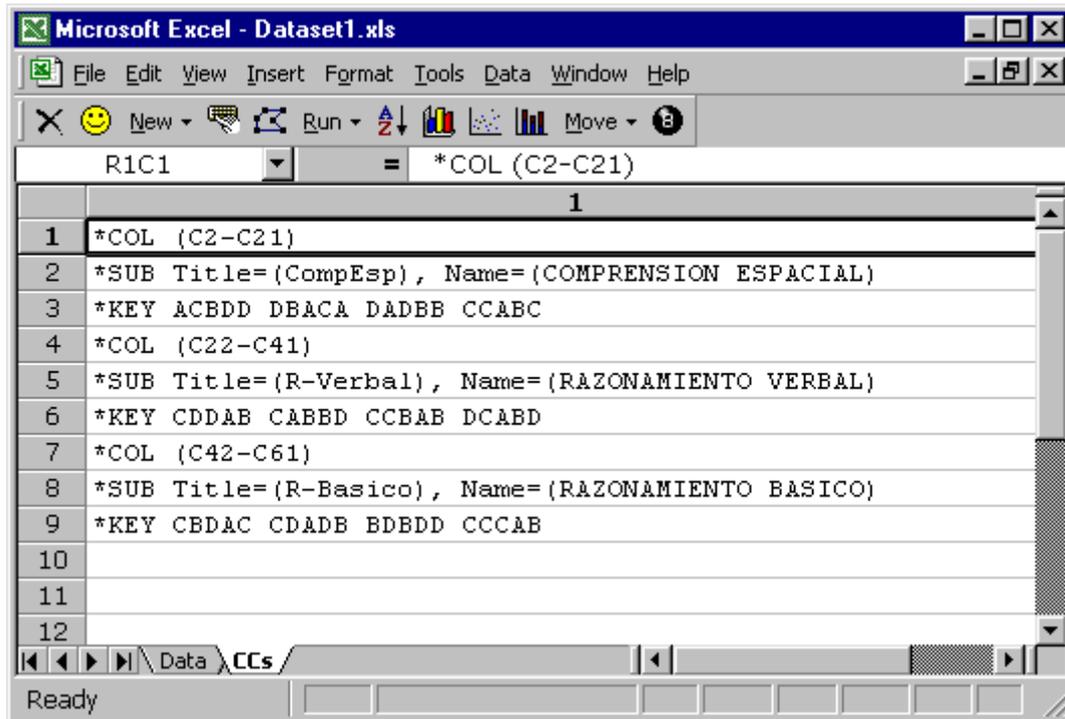
<sup>1</sup> For affective subtests, the default is res=(1,2,3,4,5).

	1	2	3
1	Lertap5 scores worksheet, last updated on: 25/09/00.		
2	<b>ID Nama Siswa</b>	<b>ChemTest</b>	<b>ChemTest%</b>
3	Aries Kusbandiar	5.00	25.0%
4	Arif Ranggawuni	5.00	25.0%
5	Aris Susanto	9.00	45.0%
6	Arthy Istyka Paramita	8.00	40.0%
7	Ayu Eka Kartika	5.00	25.0%
8	Dewi Widayani	7.00	35.0%
9	Edvin Silvana Togas	7.00	35.0%

### A UCV data set

Lertap was born in Venezuela, and has come to be frequently used at a number of Latin American sites. One of the most active of these is la Universidad Central de Venezuela. Have a look at one of their data sets:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	Data from Carlos Gonzales, received 18 March 2000 (n = 1,495).																											
2	ID																											
3	9911	D	B	C	D	C	B	A	B	C	B	D	C	A	B	A	A	B	D	C	B	A	A	B	D	C	B	A
4	9912	A	B	C	D	D					D			B	D		C	A	D	D	A	C	C	A	C	C	A	
5	9913	A	C	B	D	D	D	B	A	C	A	D	C	A	B	B	B	C	D	B	C	C	D	C	A	C	B	D
6	9914	B	B	C	D	B		A	B	D	B	D	C		D		A	D	A	B	C	C	D	C	A	B	B	C
7	9915	B		A	D	D		A		A		D	A		B	D						A	D	D	A	C	C	B
8	9916	B	C	B	D	D	D	B	A	C	B	D	A		B	B	C	B	A	B	C	C	D	C	A	A	C	D
9	9917	B	C	B	A	D	B	C	B	B	A	D	C	B	B	D	A	A	A	C	D	A	D	D	A	D	C	D
10	9918	A	B	C	D	D	D	C	A	B	A	D	C	D	B	B	D	C	D	A		B	D	C	D	C		D
11	9919	D	C	A	D	C	D	B		B	C	D	A		B	A												
12	9920	A	C	B	D	D	D	B	A	C	A	D	C	D	B	D	C	C	A	B	D	C	D	C	A	C	C	B
13	9921	A	A	C	D	D	A	A	A	A	D	C	A	B	D	A												



In this example, the Data sheet is not as extensively prepared as Piet Abik's was. For example, there are no headers at the top of the columns which contain the item responses. This is not recommended, but it's certainly okay. In its reports, Lertap will label the items as exemplified below:

The screenshot shows a summary table for 'COMPRENSION ESPACIAL' with the following data:

Res =	A	B	C	D	other	diff.	disc.	?
Item 1	<u>50%</u>	38%	2%	10%	1%	0.50	0.35	
Item 2	3%	23%	<u>52%</u>	6%	16%	0.52	0.22	
Item 3	6%	<u>58%</u>	28%	6%	2%	0.58	0.34	
Item 4	14%	6%	8%	<u>68%</u>	4%	0.68	0.29	
Item 5	13%	17%	9%	<u>56%</u>	4%	0.56	0.36	
Item 6	3%	18%	6%	<u>66%</u>	7%	0.66	0.30	
Item 7	15%	<u>57%</u>	8%	11%	9%	0.57	0.34	

Notice how the CCs sheet for this job has quite a number of lines? This is an example of a data set which involved more than one test. There were three subtests here, and the \*sub cards give their titles: CompEsp, R-Verbal, and R-Basico.

Each subtest has a \*col card, a \*sub card, and a \*key card. The item responses for the first subtest, CompEsp, start in column 2 of the Data worksheet, and end in column 21.

Lertap's Scores worksheet for this data set looked like this:

	1	2	3	4	5
1	Lertap5 scores worksheet, last updated on: 25/09/00.				
2	ID	CompEsp	R-Verbal	R-Basico	Total
3	9911	4.00	4.00	4.00	12.00
4	9912	7.00	7.00	10.00	24.00
5	9913	16.00	9.00	8.00	33.00
6	9914	5.00	9.00	9.00	23.00
7	9915	5.00	6.00	7.00	18.00
8	9916	16.00	12.00	12.00	40.00
9	9917	7.00	11.00	8.00	26.00
10	9918	11.00	7.00	9.00	27.00
11	9919	7.00	2.00	7.00	16.00
12	9920	17.00	13.00	11.00	41.00
13	9921	7.00	8.00	11.00	26.00
14	9922	12.00	12.00	8.00	32.00
15	9923	9.00	6.00	5.00	20.00

Whenever a job has more than one subtest, Lertap computes and displays a total score, as shown above. This score will be the sum of the individual subtest scores. It is possible to exclude subtests from the total, to combine them with different weights, or to turn off the total score computation altogether. These things are done by using Wt= specifications on \*sub cards—see the example shown in Chapter 2.

## A class survey

We've presented two examples so far, and both have been based on cognitive tests. Here's a simple example which from an instructor who had used a 10-item survey, an "affective" instrument, to get end-of-semester feedback from his class:

The screenshot shows an Excel spreadsheet with the following data:

	1	2	3	4	5	6	7	8	9	10
1	Ed 503 class survey, 8 September.									
2	No.	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9
3	1	3	3	3	4	3	3	3	3	2
4	2	3	2	3	3	2	4	4	4	4
5	3	3	3	2	2	2	2	4	4	3
6	4	1	2	3	4	4	2	1	2	2
7	5	2	2	2	2	3	3	1	3	2
8	6	2	3	2	3	3	3	2	3	4
9	7	2	3	2	3	3	3	1	2	4
10	8	2	4	3	3	3	2	3	2	2
11	9	1	3	3	3	3	2	2	3	2
12	10	2	4	1	1	1	1	3	2	2
13	11	1	3	2	2	2	2	3	2	3
14	12	3	2	2	2	3	2	3	3	2
15	13	3	3	5	1	1	1	1	3	2
16	14	2	2	1	1	3	3	2	2	2

The screenshot shows an Excel spreadsheet with the following data:

	1
1	*Col (C2-C13)
2	*Sub Affective
3	
4	
5	
6	
7	
8	
9	
10	

The two CCs lines for the class survey tell Lertap that item responses are found in columns 2 through 13, and that these responses are from an affective "test". The two CCs lines could also have looked like these:

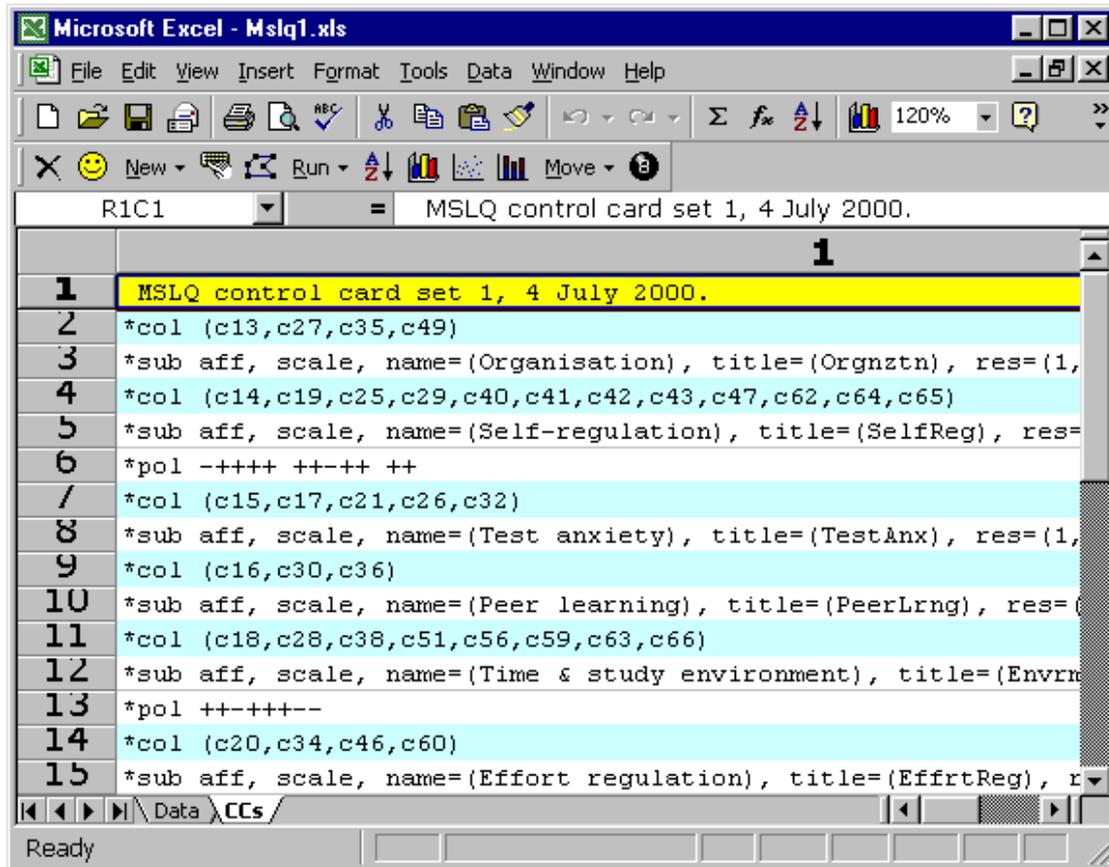
```
*col (c2-c13)
*sub aff, title=(ed503), name=(Ed503 survey, Nov. 1999)
```

A \*sub card must be used with affective subtests so that the "aff" or "affective" control word can be passed to the program. Without this control word, Lertap would assume that the items were cognitive ones, and would then expect a \*key card with a string of correct answers.

### A large-scale survey

This example is based on the application of the University of Michigan's Motivated Strategies for Learning Questionnaire, MSLQ (Pintrich, *et al*, 1991):

	1	9	10	11	12	13	14	15	16	17	18
1	LEAP M3	(1.7)	(1.8)	(1.9)	(1.95)						
2	ID_code	StudyHrs	PT_job	WorkHrs	FamRes	Q1	Q2	Q3	Q4	Q5	C
3	1S	8.0	N	0	Y	2	4	2	3	4	
4	2S	4.0	N		N	5	6	7	4	5	
5	3S	10.0	Y	7	N	2	5	6	1	4	
6	4S	5.0	N		N	7	2	2	4	6	
7	5S	6.0	Y	17.5	N	5	5	2	2	6	
8	6S	5.0	Y	20	N	3	5	5	1	4	
9	7S	6.5	Y	9	N	3	2	4	2	5	
10	8S	12.5	Y	12	Y	6	5	2	4	5	
11	9S	20	Y	10	N	4	5	2	4	3	
12	10S	15	N		N	5	2	4	5	4	
13	11S	22	Y	18	Y	6	3	3	5	5	
14	12S	10	N		N	6	4	4	2	3	
15	13S	15	Y	8	N	3	1	2	4	2	
16	14S	15	Y	11	N	4	3	2	5	3	



The MSLQ is a good example of an instrument with numerous embedded scales. For example, the CCs lines above indicate that the scale whose title was Orgnzttn was comprised of four items, with corresponding responses found in columns 13, 27, 35, and 49 of the Data worksheet. Altogether, there were ten scales in this job—these are “subtests” to Lertap. Each subtest has a \*col card and a \*sub card; the subtests which correspond to scales with reversed items have \*pol cards in addition. (The matter of reversed items was introduced in the last chapter.)

The Data worksheet itself has many columns which contain more than item response information. Column 9, for example, is headed “StudyHrs”, while “PT\_job” indicates whether or not the respondents had a part-time job. This data set is an example of one which was destined for further processing with the SPSS statistical package. In this case, Lertap was used to form the 10 MSLQ scale scores, which were then copied to the Data worksheet using a Lertap option activated via the Move option on the toolbar. This enhanced Data worksheet was then made ready for SPSS by using the toolbar’s 8-ball icon.

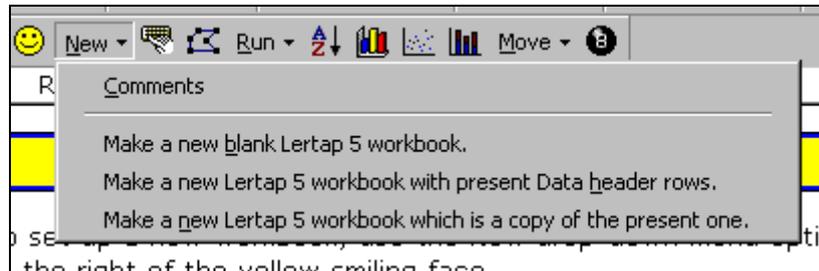
The CCs sheet above is the first we’ve seen with a comments line in it. It’s line 1. There’s no asterisk at the start of this line, which means that the line will be ignored by Lertap. Such lines may be used to add comments to the CCs worksheet.

Enough, enough? Let’s get you started on your own job. Do you have a sense of what’s needed? An Excel workbook with two sheets: Data and CCs. Item responses go into the Data worksheet, with a column for each item.

## Setting up a new workbook

If you're an experienced user of Excel, all that's required to set up a new Lertap workbook is to know that Lertap wants a book with two worksheets, one called Data, the other called CCs. The Data sheet must use its first two rows for title and header information, as seen in the examples above.

About the easiest way to establish a new workbook, be you Excel experienced or not, is to use the New option on Lertap's toolbar. Here's a picture of its options:



The first of the three "Make a new ..." options sets up a new Excel workbook with two worksheets. They will be named Data and CCs, and both will be completely blank, or empty.

The second of these three options also creates a new workbook with two worksheets, Data and CCs, but now the Data sheet will not be empty—it'll have its first row with a title in it, and its second row set up with column headers. The title and column headers will be copied from the workbook which was active when the option was taken.

At this point we toss the ball over to you. You're here because you've got data to enter. You first need to open a new workbook, and see that it has a worksheet named Data, and another named CCs. Please do this, using the New option which seems to best suit your needs (we would think it would be the first option, but it could be the second).

## Entering data

The first row in the Data worksheet is where you may type a title for your job. Make sure you're in R1C1 of the Data sheet, that is, row 1, column 1, and type whatever you want. Lertap makes no use of the information in this line, so you can type absolutely anything you wish. (You could even leave the line blank, if you want.)

The second row in the Data worksheet is meant for column titles. If you have ID information, such as student names or ID numbers, put this in either the first column, or the second, and make sure the corresponding column title begins with the letters "ID". Of the four examples given above, the first two, and the last, have such a column. In two of these examples, it's the first column, while in one, Piet Abik's, it's the second.

The columns with item responses may begin anywhere in the Data sheet. Our suggestion (for English speakers), is that the item titles follow a pattern such as "Q1, Q2, Q3, ...", or "I1, I2, I3, ...", or "Itm1, Itm2, Itm3, ...", or "Item1, Item2, Item3, ...". The briefer the title, the better some of Lertap's output will look. If it's likely data will be exported to another program, such as SPSS, it will be very useful to make sure item titles conform to the naming conventions of the other

system—SPSS, for example, spits out item titles which are longer than eight characters, which begin with a digit, and which contain a space.

### Entering item responses

Let's imagine that you have a 10-item test which uses response codes of A B C and D. Let's suppose that a person named Arkin has answered the 10 items this way:

BCDBAACCCDB

To type in Arkin's responses, you'd ordinarily go to the column set aside for the first item, type an A, move to the next column to the right, type a C, move to the next column to the right, type a D, and so on.

How you move from one column to the next column to the right depends on you. Excel lets you use either the Tab key or the right-arrow key for this. If you set Excel up correctly, you can also use the Enter key (in Excel 2000, the "Move selection after Enter" option on the Tools / Options Edit tab defines what happens after the Enter key is pressed).

Lertap has a couple of shortcuts available which make it possible to type in a person's responses as one long string, have the string dissected, and its contents spread over columns to the right. These options are represented by two icons on

Lertap's toolbar: .

It would be worthwhile experimenting with these two options—they can save you quite a bit of typing time. To do so, go to a cell in a row where you'd usually start typing item responses. However, instead of starting to type, click on the first of these two icons, and then type the string of item responses into the data entry box which will appear—do not use spaces in the string. See what happens.

Then do the same in the next row, but use the second of the icons, the "Spreader".

If your item responses are letters, not digits, then both of these icons result in the same action: the string of responses is taken apart, and distributed, one by one, over the columns to the right. If you're using the Spreader, it will spread the responses in the row you start from, and then walk down the rows, one by one, continuing to spread responses until it finds an empty row. (Press the <Esc> key to get the Spreader to stop its walk. Use the toolbar's smiling yellow face to restore the Status Bar so that it says "Ready".)

If, however, your item responses are digits, only the first of the two icons can be guaranteed to work as wanted—the Spreader (the second icon) may fail.

The reason the Spreader fails is that it lets you type directly into the cell on the worksheet. If you do this, and if what you type begins with a digit, then Excel thinks you're entering a number into one of its cells, and things can come undone—your string of item responses will often get converted to a number in "scientific" notation, such as the following: 3E+10 (which is what Excel displayed after we typed a string of 10 responses, 32451245, into a cell on the worksheet). In order to defeat this and have the Spreader work correctly with a string of digits, type a single quotation mark before the first digit ('32451245). This tells Excel to treat your string as textual information, not a number. Another way to defeat this, a more convenient way, is to highlight the column where you start

typing item responses, and get Excel to format all cells in the column as text (use Format / Cells). Once you've done this you no longer need to use the single quote before the string of digits.

### **How many rows and columns?**

The Data worksheet may have as many rows and columns as Excel permits. In Excel 2000, there may be 65,536 rows, and 256 columns. To date the largest Data worksheets we have heard of have been processed at la Universidad Central de Venezuela, where data sets with up to fourteen thousand records have been used with Lertap 5 on Pentium-based computers, running with Excel 97, under Windows 98. At Lertap headquarters we have processed data sets with just under ten thousand records on a Pentium III, using Excel 2000 and NT 4.0.

To be noted is that Lertap stops reading information from Data and CCs worksheets when it encounters a row whose first column is empty. In large data sets, with thousand of rows in the Data sheet, it is sometimes useful to test Lertap by inserting a blank row after, say, row 200. This will let you speedily try the "Interpret CCs lines" and "Elmillion item analysis" options to see if you've got things set up correctly before analysing all data. (For information on Lertap time trials, please refer to the Computational Methods chapter. Data sets with thousands of records can take several minutes to process on some computers.)

### **Missing data**

What should be done when someone does not answer a question? We suggest that non-answers be processed as a space, or what Excel refers to as a "blank". For example, let's say that Arkin did not answer item 5. Then his string of item responses would be (using the sample 10-item response string from above):

BCDB ACCDB

When typing a string such as this into Excel, we'd use a space as the fifth response, that is, we'd use the keyboard's Space Bar.

### **Importing data**

Data may be imported to the Data worksheet using Excel's standard options. In Excel 2000 these are found under Data / Get External Data options. If the data are already in an Excel worksheet in some other workbook, go to that workbook, select the sheet with the data (by clicking on its tab, for example), and then copy this worksheet to your new Lertap workbook. Once it's there rename it as Data.

### **Creating the CCs lines**

By far the most laborious part of a Lertap job is getting the Data worksheet set up with records of item responses. Once this is done, job definition "statements" must be entered in the lines of the CCs worksheet.

All Lertap subtests have to have a \*col () line in the CCs worksheet.

Cognitive subtests also require a \*key card, a line in the CCs worksheet which contains a string with the correct answers. Affective subtests require a \*sub card with the control word "Aff" or "Affective" on it.

It is, then, always the case that a minimum of two CCs lines are required for each subtest to be processed—cognitive tests require \*col and \*key, affective tests require \*col and \*sub. Many times additional "cards", or lines, are used. At

Lertap headquarters, for example, we always use a \*sub card for cognitive tests in order to give the test a name and a title.

You've already seen several examples of CCs lines in this chapter, and in the last. There are more in the chapters which follow. Don't forget that "on-line" examples of CCs lines may be found in the Syntax worksheet in the Lertap5.xls file, as mentioned in Chapter 2.

CCs lines must all begin in column 1 of their respective rows. Lines with an asterisk at the front, as the very first character in the line, are parsed by Lertap (examined by the CCs interpreter). Lines which do not have an asterisk as their first character are considered to be comments, and are totally ignored by the program.

It is sometimes useful to type up CCs lines for a subtest, but have Lertap ignore them temporarily. This happens in complex jobs, such as the MSLQ example above, when users want to limit attention, for the moment, to just some of a job's subtests. At Lertap headquarters we take subtests "off line" by putting a space in front of the asterisks of the corresponding CCs lines. Another way to do this is to put a single quote in front of the asterisk. For example, in the following job the second subtest, SelfReg, has been taken "off line":

```
*col (c13,c27,c35,c49)
*sub aff, scale, name=(Organisation), title=(Orgnztn), res=(1,2,3,4,5,6,7), wt=0
`*col (c14,c19,c25,c29,c40,c41,c42,c43,c47,c62,c64,c65)
`*sub aff, scale, name=(Self-regulation), title=(SelfReg), res=(1,2,3,4,5,6,7), wt=0
`*pol -++++ ++-++ ++
*col (c15,c17,c21,c26,c32)
*sub aff, scale, name=(Test anxiety), title=(TestAnx), res=(1,2,3,4,5,6,7), wt=0
*col (c16,c30,c36)
*sub aff, scale, name=(Peer learning), title=(PeerLrng), res=(1,2,3,4,5,6,7), wt=0
```

### **Making changes to CCs lines**

It is common for users to change their CCs lines. They might do so, for example, only to correct an error as simple as the one seen below, where there's a spelling mistake in the \*sub card's name= assignment:

```
*col (C3-C22)
*sub title=(ChemTest), name=(Clas test of 20 May)
*key ADCAB BDBBD ADBAB BCCCB
```

Once changes have been made to CCs lines, both the "Interpret CCs lines" and "Elmillion item analysis" options have to be applied again in order for the changes to be effected.

### **Just Freqs?**

We mentioned above how inserting a blank row in the Data worksheet is sometimes useful. The same holds true for the CCs worksheet. For example, the following four CCs lines

```
*col (C3-C22)

*sub title=(ChemTest), name=(Class test of 20 May)
*key ADCAB BDBBD ADBAB BCCCB
```

would result is a "Freqs only" analysis.

When the "Interpret CCs lines" is taken with CCs cards such as these four, Lertap reads only the first line, creates its Freqs worksheet, and stops without going on to make the Sub worksheet mentioned in the last chapter. It will do this because all processing halts once a blank line is encountered in the CCs worksheet.

## **Getting results**

What's the title of this chapter? "Setting Up and Processing a New Data Set".

What is a data set? An Excel workbook tailor-made for use with Lertap. Such workbooks have item response data in a worksheet called Data, and job definition statements in another worksheet called CCs.

Do you need Lertap to set up a data set? No, not really. Anyone with sufficient nous can get Excel to create a new workbook, and then see that it has two worksheets, one named Data, the other CCs.

Lertap can definitely assist in setting up a new Excel workbook which meets the requirements of a Lertap data set. For example, Lertap's toolbar has its New option which simplifies the creation of an Excel workbook. And Lertap's toolbar has two data entry assistants, one of them called "the Spreader", which can ease the process of typing in all those item responses. But we repeat: Lertap is not really required to set up a data set. Helpful at times, but not critical.

It's when you want to get results that Lertap becomes indispensable. You have to open the Lertap5.xls file in order to have access to the all-conquering Run option, the gateway to "Interpret CCs lines" and "Elmillion item analysis". It's only when these tools are available that you can get those wonderful, information-laden Lertap worksheets with scores and item statistics.

At this point we'll power down, hoping that this chapter, and Chapter 2, have given you all that's required to get set up and running. When your analyses become more complex, or you want to turn on more of Lertap's options, the information in the following two chapters on Lertap control "cards" will be useful.