

Scoring Short-Answer Test Items with Lertap

Larry R Nelson

Curtin University & Burapha University

Original Document date: 10 May 2004

Updated: 11 July 2017

website: www.larrynelsonstuff.com/lertap/

July 2017 update: an interesting international study of numeracy among junior high-school students involved a short test with 20 cognitive items of mixed types, multiple-choice and constructed response. Click [here](#) to see how it was handled with Lertap 5.

In March 2004 I received the message shown below. The question raised in the message is one which crops up with some frequency, having to do with scoring test items which are not of the multiple-choice type. Have a read of the message. Then, if the general issue might also apply to your own work, take in the response.

(**Note:** while this document mainly addresses the matter of scoring open-ended items, it also provides more insights on the way Lertap uses response codes, and how the Res= declaration on *sub lines is interpreted. Part of this discussion is complex, for adults only, that is, not for youngsters, not for people first making their way with Lertap.)

Hi!

My name is Candy Veas, I'm a graduate student at the University of Minnesota in the US, and I'm currently working on my thesis project. I'm getting my MA in Teaching English as a foreign language, and I'm trying to use your fabulous program to run some statistical analysis of my thesis project's instrument, a 18 item English reading comprehension test that I administered to 53 students.

I'm writing to you in hope that you will be able to help me with a little problem that my classmates and professors haven't been able to help with.

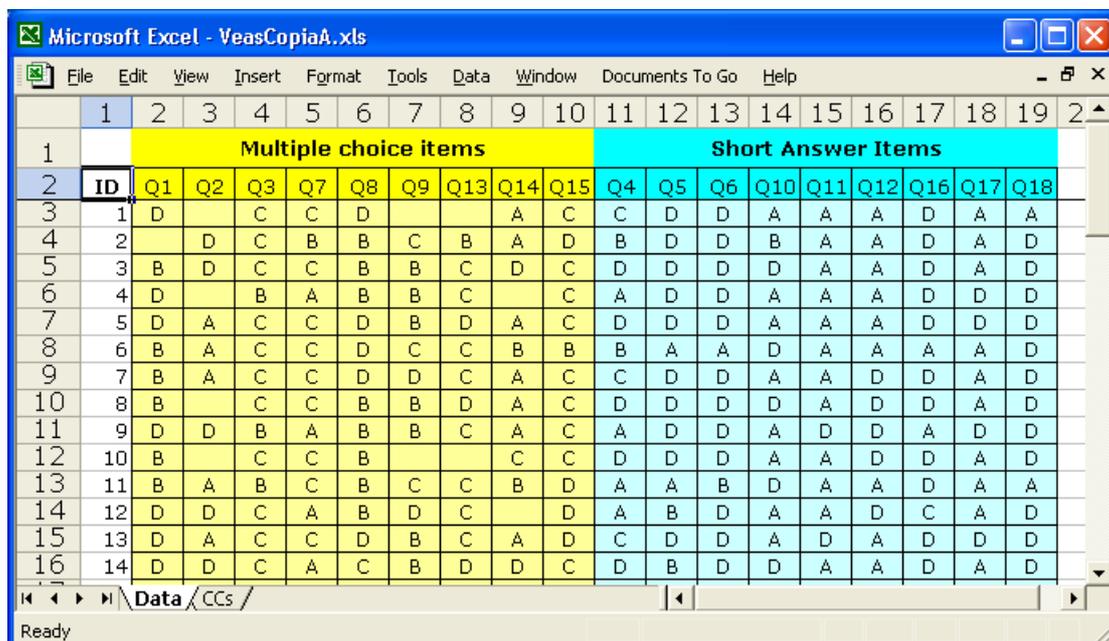
This is that my test has two different types of questions, one is multiple choice (which I haven't had problems with since Lertap is very clear on what I do with those), and the other type is short answer questions, which allow for partial credit, say .5, .75, 0, or 1. The two types of questions that I've found in the CCs are either multiple choice or "affective", neither of which apply to these short answer questions that I have. My question is, Is it possible to use your program to run analysis for my whole test including those items?

Thanks so much for your help!

Is she out of luck? Is Lertap incapable of scoring items which are not multiple-choice in nature?

No, and no. She's probably not out of luck at all; providing certain conditions are met, Lertap will score items which are not of the multiple-choice variety, including short-answer and essay questions.

I asked Candy if she might be able to send me a copy of her data set. She did. Below is a snapshot of part of her data, after I modified it.



The screenshot shows a Microsoft Excel spreadsheet titled "VeasCopiaA.xls". The spreadsheet has a header row (row 2) with columns labeled "ID", "Q1", "Q2", "Q3", "Q7", "Q8", "Q9", "Q13", "Q14", "Q15", "Q4", "Q5", "Q6", "Q10", "Q11", "Q12", "Q16", "Q17", and "Q18". The first row (row 1) is highlighted in yellow and contains the text "Multiple choice items" for columns 2-10 and "Short Answer Items" for columns 11-19. The data rows (rows 3-16) contain student IDs in column 2 and their responses (A, B, C, D) in the subsequent columns. The responses for the first 10 columns (MC items) are in yellow cells, and the responses for the last 9 columns (SA items) are in cyan cells.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		Multiple choice items									Short Answer Items									
2	ID	Q1	Q2	Q3	Q7	Q8	Q9	Q13	Q14	Q15	Q4	Q5	Q6	Q10	Q11	Q12	Q16	Q17	Q18	
3	1	D		C	C	D			A	C	C	D	D	A	A	A	D	A	A	
4	2		D	C	B	B	C	B	A	D	B	D	D	B	A	A	D	A	D	
5	3	B	D	C	C	B	B	C	D	C	D	D	D	D	A	A	D	A	D	
6	4	D		B	A	B	B	C		C	A	D	D	A	A	A	D	D	D	
7	5	D	A	C	C	D	B	D	A	C	D	D	D	A	A	A	D	D	D	
8	6	B	A	C	C	D	C	C	B	B	B	A	A	D	A	A	A	A	D	
9	7	B	A	C	C	D	D	C	A	C	C	D	D	A	A	D	D	A	D	
10	8	B		C	C	B	B	D	A	C	D	D	D	D	A	D	D	A	D	
11	9	D	D	B	A	B	B	C	A	C	A	D	D	A	D	D	A	D	D	
12	10	B		C	C	B		C	C	C	D	D	D	A	A	D	D	A	D	
13	11	B	A	B	C	B	C	C	B	D	A	A	B	D	A	A	D	A	A	
14	12	D	D	C	A	B	D	C		D	A	B	D	A	A	D	C	A	D	
15	13	D	A	C	C	D	B	C	A	D	C	D	D	A	D	A	D	D	D	
16	14	D	D	C	A	C	B	D	D	C	D	B	D	D	A	A	D	A	D	

She's got nine MC (multiple-choice) and nine SA (short-answer) items. For some reason, she grouped the answers so that the MC item responses appear first, followed by codes which represent the scores earned by the students on each of the SA items.

The key to understanding how Lertap processes open-ended items, such as Candy's short-answer questions, is to realise that the letters showing in columns 11 through 19 of the Data worksheet do not represent item responses – rather, they're codes. Candy derived them.

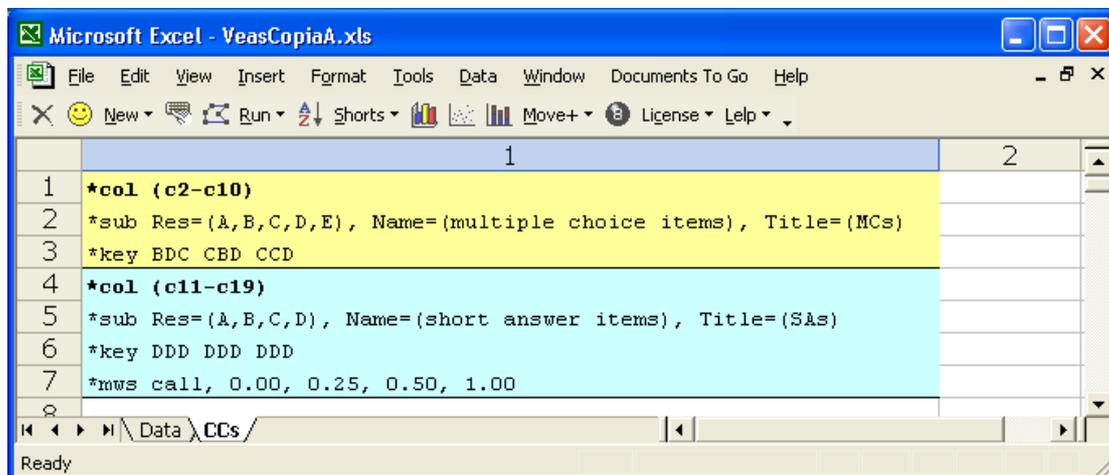
An "A" in these columns means that a student got zero points for their response to an item; "B" means they got 1/2 a point (0.50 points); "C" means they got 3/4 of a point (0.75 points); "D" means they got the item totally correct, and earned one point (1.00

points). How do I know this? Candy told me; these are scoring codes which she herself defined.

Compare this with the meaning of the letters found in columns 2 through 10 of the Data worksheet. If you were able to see all of the rows of the data set, you'd notice that the multiple-choice items used five options, coded as (A,B,C,D,E). An "A" in columns 2 through 10 simply means that a student selected option A on a multiple-choice item; an "E" means option E was chosen. How many points would a student get if s/he selected option A on the first question, Q1? We don't know. We'd have to ask Candy what the right answer was.

Candy could have used any letters or digits as the scoring codes for the short-answer items. In fact, her first inclination was to use codes of (0,1,2,3). But (and this is really a big but!), things turn out to be much easier, Lertap-wise, if the scoring codes bear close resemblance to the response codes used for the options to the multiple-choice questions. I twisted her arm (by email), and got her to change the short-answer scoring codes from (0,1,2,3) to (A,B,C,D).

Now, how to get Lertap to score this test? Good question. Have a look at the CCs worksheet which I helped Candy create:



If the screen snapshot above is hard to read on your computer, I've copied the CCs rows from Excel and pasted them below:

```
*col (c2-c10)
*sub Res=(A,B,C,D,E), Name=(multiple choice items), Title=(MCs)
*key BDC CBD CCD
*col (c11-c19)
*sub Res=(A,B,C,D), Name=(short answer items), Title=(SAs)
*key DDD DDD DDD
*mws call, 0.00, 0.25, 0.50, 1.00
```

The seven CCs lines above show that Candy wanted to process the MC and SA items separately.

I would think you'd be able to understand the first three lines without problem: we've got nine columns of the Data worksheet, 2 through 10, being scored in a conventional manner. The right answer to the first item is B. The right answer to the second item is D. The right answer to the last of the nine multiple-choice items is D.

Why are spaces used to set off every three letters in the *key line? Dunno – we'd have to ask Candy. Spaces on Lertap's CCs lines are usually used just to make the line easier to read.

Now, what about the last four lines above? They're tricky – they get Lertap to convert Candy's scoring codes to points.

The line which does this is the last one, the *mws line. The "call" part of this line means "columns all" – it says to Lertap that the points (or numbers) shown on the line are to apply to all of the columns used by this group of items, that is, columns 11 through 19 of the Data worksheet.

There are four numbers on the *mws line, after the "call" word. And, definitely **not** by coincidence, there are four letters in the Res= declaration on the *sub line.

There's a one-to-one correspondence between the numbers and the letters. The first number, 0.00, links to the first letter, A. The second number, 0.25, links to the second letter, B.

In essence, the *mws line is saying this to Lertap (and Excel):

$$(A,B,C,D) = (0.00, 0.25, 0.50, 1.00)$$

Neat, eh? This is the trick to use with open-ended test items, such as short-answer and essay questions: develop scoring codes, single letters or digits, and enter these in Lertap's Data worksheet. Then use *mws lines to map the scoring codes to actual points.

Now, as I look out at the audience, I see several hands in the air. Some will be from those needing to go to the toilet, but no doubt at least one of the hands will be attached to a person who wants to ask this question: *"I think I understand, but – what the devil is that *key line doing for the group of short-answer questions?"*.

This is a really good question. *The *key line is used to indicate the correct answer to a multiple-choice question, isn't it? (Everyone nods, even those who've gone to the toilet.) But these are short-answer items! They don't have a keyed-correct answer! True. You could try and run this job without a *key line, but I wouldn't – when Lertap sees a *key line in the CCs worksheet, it first of all says to itself: Ah-ha! The user is saying these items are cognitive ones, after which it looks for the keyed-correct answer for each of the items.*

In this case, the *mws line gets Lertap to abandon the idea of a correct answer for the items, but, because it's seen a *key line, Lertap will not forget that the items are cognitive ones, and it'll format its various reports, such as the Stats1f report, in the cognitive style (as opposed to the affective style).

So there. How are you doing? What would Candy now do to get results? She'd go to the Run menu, and click on Interpret CCs lines. She'd look at Lertap's Freqs report, and, if everything's in order, she'd trot back to the Run menu, and click on Elmillon item analysis.

How many scores will Lertap make for each student? Three. There will be a score for the MC items, a score for the SA items, and a total score. Look:

The screenshot shows a Microsoft Excel window titled "Microsoft Excel - VeasCopiaA.xls". The spreadsheet displays a table with the following data:

	1	2	3	4
1	Lertap5 Scores worksheet, last update			
2	ID	MCs	SAs	Total
3	1	2.00	3.50	5.50
4	2	4.00	4.50	8.50
5	3	6.00	6.00	12.00
6	4	2.00	5.00	7.00
7	5	2.00	6.00	8.00
8	6	4.00	2.25	6.25
9	7	5.00	5.50	10.50
10	8	4.00	7.00	11.00
11	9	3.00	6.00	9.00
12	10	5.00	6.00	11.00
13	11	5.00	2.25	7.25

The whole test, and nothing but.

Hmmm ... Candy's out there; she's speaking to us through the ether (via Ethernet, of course), saying that she'd really like all the items to be scored as a single test.

Can do? Certainly.

Are you in a good mood? If not, take a break – we're going to get into a mildly complex CCs worksheet, and you may want to be wearing your best suit of patience.

Look:

```
*col (c2-c19)
*sub Res=(A,B,C,D,E), Name=(MA reading test), Title=(Reading)
*key BDC CBD CCD DDD DDD DDD
*mws c11, 0.00, 0.25, 0.50, 1.00, *
*mws c12, 0.00, 0.25, 0.50, 1.00, *
*mws c13, 0.00, 0.25, 0.50, 1.00, *
*mws c14, 0.00, 0.25, 0.50, 1.00, *
*mws c15, 0.00, 0.25, 0.50, 1.00, *
*mws c16, 0.00, 0.25, 0.50, 1.00, *
*mws c17, 0.00, 0.25, 0.50, 1.00, *
*mws c18, 0.00, 0.25, 0.50, 1.00, *
*mws c19, 0.00, 0.25, 0.50, 1.00, *
```

The *col line is now covering all of the items in the test, the nine MC items plus the nine SA items.

The *key line now has 18 entries.

There is now one *mws line for each of the SA items. Before there was just one *mws line, and it made use of the "call" statement. But now "call" is not right. Remember that the "all" part of "call" means all columns as found in the *col line, but this is no longer correct – we don't want the mapping seen in the *mws lines to apply to all items, only the latter nine, the ones found in columns 11 through 19. Since "call" can't be used, I am forced to use an *mws line for each of the nine affected columns.

What's that asterisk doing at the end of each *mws line?

Recall that the *mws lines are mapping the Res= characters to points. Now the Res= declaration has five letters – before it had only four for the SA items, but that was when the short-answer questions had been grouped on their own. Now all items have been lumped together, and the *mws lines would ordinarily be expected to have five numbers, one for each of the five letters seen in the Res= declaration. But this wouldn't be right, would it? The *mws columns correspond to items with four scoring codes, (A,B,C,D).

The asterisk at the end of each *mws lines tells Lertap to forget about the fifth Res= letter as there is not a scoring code corresponding to the letter E for the short-answer items.

Providing certain conditions are met?

"I say, Larry, this is really neat stuff. Lertap is really a marvel, just like Mother said. But I know I heard you imply that Lertap will do this type of scoring only when certain conditions are met. What are those conditions?"

Simple. There can be a maximum of 30 characters in the Res= declaration, so the mapping of scoring codes to points will only work when the number of scoring codes is 30 or less.

Note that the mapping does not have to be the same for all items. Have a gander at these CCs lines:

```
*col (c2-c19)
*sub Res=(A,B,C,D,E), Name=(MA reading test), Title=(Reading)
*key BDC CBD CCD DDD DDD DDD
*mws c11, 0.00, 1.00, *, *, *
*mws c12, 0.00, 0.50, 0.75, 2.00, *
*mws c13, 0.00, 0.25, 0.50, 1.00, *
```

The *mws c11 line tells Lertap that just two scoring codes, A and B, were used by the item whose responses have been coded in column 11 of the Data worksheet. Why? Because there are asterisks corresponding to the third, fourth, and fifth Res= letters: C, D, and E.

The *mws c11 line also says that A maps to 0.00 points, while B maps to 1.00 points (for this item).

Now look at the *mws c12 line. B maps to 0.50 points for this item; C to 0.75 points; and D to 2.00 points.

Two hands are up at the back of the room, both connected to arms which lead down to the same over-excited individual. Yes, sonny? *"What restrictions are there on the numbers found in a *mws line?"*

Just about none. The following line is often encountered in scoring a certain style of affective items, one having seven scoring codes:

```
*mws call, -3.00, -2.00, -1.00, 0.00, 1.00, 2.00, 3.00
```

There are essentially no limits to what the *mws numbers may be.

Candy's original setup.

This small section is for adults only. It gets into dirty CCs lines.

Look at the lines shown below; they indicate how Candy originally thought of processing her data:

```
*col (c2-c10)
*sub Res=(A,B,C,D,E), Name=(multiple choice items), Title=(MCs)
*key BDC CBD CCD
*col (c11-c19)
*sub Res=(0,1,2,3), Name=(short answer items), Title=(SAs)
*key 333 333 333
*mws call, 0.00, 0.25, 0.50, 1.00
```

I have mentioned how Candy first coded the short-answer questions: she used 0 as a code for 0.00 points, 1 as a code for 0.25 points, 2 as a code for 0.50 points, and 3 as a code for 1.00 points.

The (0,1,2,3) codes in the Res= declaration map to the scoring points seen in the *mws call line, and there's nothing at all wrong with this – results will be *absolutely identical* to those found by changing the short-answer scoring codes from (0,1,2,3) to the (A,B,C,D) codes which I have used.

The problem with the (0,1,2,3) codes arises when I want to score all items, the multiple choice **and** the short answer, as one set. I'll require CCs lines like these:

```
*col (c2-c19)
*sub Res=(A,B,C,D,E,0,1,2,3), Title=(Reading)
*key BDC CBD CCD 333 333 333
*alt 555 555 555 999 999 999
*mws c11, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c12, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c13, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c14, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c15, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c16, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c17, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c18, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
*mws c19, *, *, *, *, *, 0.00, 0.25, 0.50, 1.00
```

In this set of CCs lines, the Res= declaration has been extended to show all of the codes used by the 18 items, and a *alt line has been included to tell Lertap that the first 9 items use just the first 5 Res= codes (that is: A,B,C,D,E).

The *alt line gives the impression that the last of the items use all of the Res= codes (of which there are 9); this isn't right, and the asterisks at the beginning of the *mws lines are used to correct the picture. They tell Lertap that the items whose codes have been

recorded in columns 11 through 19 of the Data worksheet do not make use of the first five characters found in the Res= declaration.

Some might look at the CCs lines above and say "*no worries, mate – I got no problem writing lines like these*". And good on them. But they're unnecessarily complex. I think lines such as these are dirty, messy, unsightly, and too complicated. I've gone into them only to help explain why it's a good idea to try and use the same codes for both multiple-choice and open-ended questions. (Way at the back of my mind, I've also thought that maybe this discussion would be enlightening for some readers. Elsewhere I have been brash enough to suggest that Lertap will "score just about anything", and I think, within the limits cited above, that this might be a fairly fair statement. The *mws line has extensive power. Readers who want to get down to an even dirtier level will be pleased to know that it's possible to do so: open **Lelp**, the Lertap help file, and search for topics having to do with Lertap's Sub worksheets, and the "advanced toolbar".)

More reading.

You say you'd like to know even more about scoring items in Lertap? Great. I recommend two sources: "**Lelp**", the Lertap help file, and the manual. Lelp comes in three forms: a pdf file, a chm file, and a website.

Users with Windows computers automatically get a copy of the chm version of Lelp – it's a normal Windows help file, accessed via the Lelp menu found at the extreme right of the Lertap toolbar.

The Lertap website is: www.larrynelsonstuff.com/lertap/. The various versions of Lelp are available via hyperlinks in the documentation section. The same section also has links to the manual.

My thanks to Candy Veas for prompting this little paper, and for providing a very practical example to work with. If you have questions, email them to: larry@lertap.com. Thanks.

Larry Nelson
School of Education
Curtin University
Perth, Western Australia