

Scoring Mixed Cognitive Item Types with Lertap 5

[Larry R Nelson](#)

[Curtin University](#) & [Burapha University](#)

Last updated on: 13 June 2017

This paper demonstrates how to optimally score a test with Lertap 5 when the test has a mixture of supply¹ and multiple-choice test items.

The items are from an international study of numeracy among Year 7 students. A 15-item test was involved – the actual questions are seen at the end of this paper. The original study may be found at the following URL:

<http://www.edmeasurementsurveys.com/TAM/Tutorials/4CTT.htm>

Nine of the items were supply items scored by teachers as (0,1), that is, zero points for an incorrect answer, one point for a correct answer. These were items 1, 2, 5, 7, 8, 9, 10, 13, and 14. Here is an example of one of the supply items:

Science Tests[®]

In Mei Lin's school, her science teacher gives tests that are marked out of 100. Mei Lin has an average of 60 marks on her first four Science tests. On the fifth test she got 80 marks.

What is the average of Mei Lin's marks in Science after all five tests?

Average:

The other six items were multiple-choice items; these were items 3, 4, 6, 11, 12, and 15. The first five of these items had four options (1,2,3,4). The last item, 15, had five options but for some unexplained reason was coded by teachers as being right or wrong, (0,1).

The actual data file for the study may be downloaded by visiting [this webpage](#) and, under the section labelled "Data files", clicking on "Numeracy data file in csv format". On most up-to-date

¹ Supply items are also known as "constructed response" items.

computers, csv² files will automatically open in Microsoft's Excel spreadsheet program (assuming, of course, that the computer has Excel installed on it). The screen snapshot below displays the first 10 data records as seen in Excel 2016:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15
2	1	0	4	4	1	2	1	1	1	1	3	1	1	1	1
3	0	1	4	1	1	1	1	1	1	1	3	1	1	0	1
4	1	0	1	2	0	1	1	1	1	1	1	2	1	1	1
5	0	0	3	2	0	3	0	0	0	0	2	2	1	1	1
6	0	1	4	1	1	2	0	1	1	1	3	4	0	1	0
7	0	1	4	1	1	2	1	1	1	1	3	1	1	1	1
8	0	1	4	1	1	1	0	0	0	0	4	1	0	1	1
9	0	1	4	1	1	1	0	0	1	1	3	1	1	1	1
10	1	1	4	1	1	1	0	0	1	1	3	1	1	1	1

A study of the csv file will reveal that student responses to all of the supply items were coded as (0,1), that is, zero for an incorrect answer, one for a correct answer. It turns out that item 15 was also coded in this manner even though it was a multiple-choice question with five options.

Responses to the other multiple-choice items were coded as (1,2,3,4), reflecting the option selected by students on those items. Above, for example, we see that the 5th student chose option 3 on item I6 (row 5, column F).

We can say, therefore, that the set (0,1,2,3,4) was used to code the item responses. Many of the items used just the first two members of this set, (0,1), while others used the last four, (1,2,3,4).

I will now use [Lertap 5](#) to score the test and to examine item performance. Lertap 5 is an item and test analysis system which runs as an Excel "app". It derives directly from a mainframe package I originally created in 1973 using the FORTRAN programming language.

² csv means "comma-separated values". To see a csv file in its raw state, open it with a text editor such as Notepad (Windows) or TextEdit (Mac).

As I go about applying Lertap 5, I will point out how the response coding scheme used in the study involved a bit of intricate maneuvering in order to get the program's output to look its best. More likely than not other item analysis programs would require at least a bit the same sort of effort.

CCs lines in Lertap 5

Taken as a group, the items can be said to use [response codes](#) of (0,1,2,3,4). Many of the items use just (0,1) while the others use (1,2,3,4). No item actually uses (0,1,2,3,4) as possible response codes.

The following lines would score the test in Lertap 5, giving one point for each correct answer³:

```
*col (c2-c16)
*sub res=(0,1,2,3,4), title=(Scoring1)
*key 11411 11111 31111
```

The lines above assume that item responses have been coded in columns 2 (c2) through 16 (c16) of Lertap's Data worksheet⁴. The res=(0,1,2,3,4) assignment tells Excel which item responses are valid; if a column is found to have another response, such as, for example, a null response (empty cell) or, for example, a 5, then Lertap will process that response as missing data. (For readers unfamiliar with Lertap 5, [this document](#) is suggested viewing.)

Of the 15 items, only two, the 3rd item and the 11th item, have a correct response which is not 1 (one). This is indicated on the *key line. As seen on that line, the correct answer to the 3rd item is 4, while 3 is the correct answer on the 11th item.

The lines above are referred to as "CCs lines" as they are always found in an Excel worksheet called CCs – the letters stand for "control cards". This is an [historical anomaly](#).

The purpose of CCs lines in Lertap is to score tests. The three lines above will certainly do that, but they have a slight problem: as they stand, they imply that all of the test's items use the set of response codes given in the res= assignment. This isn't the case, and, because it isn't,

³ The 15 entries on the *key line are the "keyed-correct" answers to the items. I have grouped them in three sets of five keys just because that is my preference. It's not necessary to group the keys; many people would just use one string, 15 characters long.

⁴ I added an ID field in the first column, moving all the item responses one column to the right.

Lertap's output will not be what I would call "optimal". It will be accurate in terms of the statistics it produces, but the format of some important sections of the output will be somewhat messy.

Let me take two of the items to show the limitations that will result by running Lertap with the CCs lines shown above.

The first item is a supply item coded (0,1). Lertap's "[Freqs](#)" report for this item indicates that the 876 students who took the test were about equally split on this item. Slightly less than half of them got the item right:

I1 (c2)

Option	n	/876
0	454	51.8%
1	422	48.2%

The third item is a multiple-choice question:

I3 (c4)

Option	n	/876
1	99	11.3%
2	49	5.6%
3	15	1.7%
4	713	81.4%

The "Freqs" report is a simple one. Its job is to report on the actual responses found for each test item.

Now have a look at Lertap's "[Stats1f](#)" report for the two items:

I1 (c2)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0.00	454	0.52	-0.57	-0.71	6.01	-0.55
<u>1</u>	<u>1.00</u>	<u>422</u>	<u>0.48</u>	<u>0.48</u>	<u>0.60</u>	<u>10.66</u>	<u>0.59</u>
2	0.00	0	0.00	0.00	0.00	0.00	0.00 <-no
3	0.00	0	0.00	0.00	0.00	0.00	0.00 <-no
4	0.00	0	0.00	0.00	0.00	0.00	0.00 <-no

I3 (c4)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0.00	0	0.00	0.00	0.00	0.00	0.00 <-no
1	0.00	99	0.11	-0.32	-0.54	4.53	-0.91
2	0.00	49	0.06	-0.21	-0.42	4.76	-0.85
3	0.00	15	0.02	-0.12	-0.36	4.60	-0.89
<u>4</u>	<u>1.00</u>	<u>713</u>	<u>0.81</u>	<u>0.34</u>	<u>0.50</u>	<u>9.08</u>	<u>0.20</u>

The Stats1f report has assumed that all items use (0,1,2,3,4) as response codes. But we know this wasn't the case, and it shows above: three of these codes (2, 3, 4) had no takers on I1, while the first response code (0) had no takers on I3.

The various statistics reported for each item are in fact correct. But the summaries seen in Stats1 look rather bad in my opinion. Things would be cleaner and look better if we got in and told Lertap that the items use different response codes (as readily seen in the Freqs report).

This is done by making changes to the CCs lines. The following example uses a `*alt` line to specify the last alternative in the `res=` assignment that was used by each item.

```
*col (c2-c16)
*sub res=(0,1,2,3,4), title=(Scoring2)
*key 11411 11111 31111
*alt 11441 41111 44111
```

*alt lines are interpreted with regard to the `res=` assignment. The last alternative in the `res=` assignment used by the first two items was 1 (one) according to the *alt line seen above, while the last alternative used by the second and third items was 4 according to the *alt line above.

This isn't quite right, but let's see how the Stats1f report will look now:

I1 (c2)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0.00	454	0.52	-0.57	-0.71	6.01	-0.55
<u>1</u>	<u>1.00</u>	<u>422</u>	<u>0.48</u>	<u>0.48</u>	<u>0.60</u>	<u>10.66</u>	<u>0.59</u>

I3 (c4)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0.00	0	0.00	0.00	0.00	0.00	0.00 <-no
1	0.00	99	0.11	-0.32	-0.54	4.53	-0.91
2	0.00	49	0.06	-0.21	-0.42	4.76	-0.85
3	0.00	15	0.02	-0.12	-0.36	4.60	-0.89
<u>4</u>	<u>1.00</u>	<u>713</u>	<u>0.81</u>	<u>0.34</u>	<u>0.50</u>	<u>9.08</u>	<u>0.20</u>

Things are a bit better; the results for I1 are cleaner than before as only the two options used by that item are listed. There aren't any <no lines showing for I1.

I3's display, on the other hand, has not been corrected. This item did not use 0 (zero) as a possible option. Can I get rid of that <no line for I3?

Yes, I can. To clean up items which do not use one or more of the characters seen at the start of the `res=` assignment, `*mws` lines are employed. The best CCs lines for scoring the test are seen below:

```
*col (c2-c16)
*sub res=(0,1,2,3,4), title=(Scoring4)
*key 11411 11111 31111
*alt 11441 41111 44111
*mws c4, *, 0, 0, 0, 1
*mws c5, *, 1, 0, 0, 0
*mws c7, *, 1, 0, 0, 0
*mws c12, *, 0, 0, 1, 0
*mws c13, *, 1, 0, 0, 0
```

The first `*mws` line above refers to the third item, the item whose responses are found in `c4` (column 4). The five `*`, `0`, `0`, `0`, `1` characters refer directly to the five characters found in the `res=(0,1,2,3,4)` assignment. They say that the first character found in the `res=` assignment was not used by the item, the item score to be given to students selecting the second entry in the `res=` assignment is to be zero, the item score to be given to students selecting the third entry in the `res=` assignment is to be zero, the item score to be given to students selecting the fourth entry in the `res=` assignment is to be zero, while the item score to be given to students selecting the fifth entry in the `res=` assignment is to be one.

The correct answer to the third item, coded in column 4, is 4, the fifth entry in the `res=` assignment. One point will be given to students who select the correct option.

The correct answers to items 4, 6, 11, and 12, coded in columns 5, 7, 12, and 13, are 1, 1, 3, and 1 respectively. One point will be given for each correct answer.

`*mws` lines are used as a scoring map. They have values which directly score each of the entries found in the `res=` assignment. These values do not have to be integers. See more examples starting at “Example C10” on [this webpage](#).

After making these changes and running Lertap again⁵, the Stats1f output is as follows:

⁵ By taking the Interpret and Elmillon options as mentioned [here](#).

I1 (c2)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0.00	454	0.52	-0.57	-0.71	6.01	-0.55
<u>1</u>	<u>1.00</u>	<u>422</u>	<u>0.48</u>	<u>0.48</u>	<u>0.60</u>	<u>10.66</u>	<u>0.59</u>

I2 (c3)

option	wt.	n	p	pb(r)	b(r)	avg.	z
0	0.00	483	0.55	-0.65	-0.81	5.87	-0.58
<u>1</u>	<u>1.00</u>	<u>393</u>	<u>0.45</u>	<u>0.57</u>	<u>0.71</u>	<u>11.18</u>	<u>0.72</u>

I3 (c4)

option	wt.	n	p	pb(r)	b(r)	avg.	z
1	0.00	99	0.11	-0.32	-0.54	4.53	-0.91
2	0.00	49	0.06	-0.21	-0.42	4.76	-0.85
3	0.00	15	0.02	-0.12	-0.36	4.60	-0.89
<u>4</u>	<u>1.00</u>	<u>713</u>	<u>0.81</u>	<u>0.34</u>	<u>0.50</u>	<u>9.08</u>	<u>0.20</u>

The Stats1f results are cleaner than they were. The first item has just two option lines now instead of five, and the third item now has four option lines instead of five.

A similar clean up will be found when the [Stats1b](#) and [Stats1ul](#) reports are examined.

It might be worth repeating that the very first set of CCs lines seen above, at the start of this document, will work fine and result in correct statistics for all of Lertap's reports. All I've done is indicate what I would do to make the output easier to read and more reflective of how many responses each item actually employed.

There is even another way to accomplish these things. The CCs lines tell Lertap and Excel how items are to be scored when the [Elmillon](#) option is taken. They do so by creating hidden Excel worksheets called "Sub" sheets – when Elmillon runs it gets all item scoring instructions from these sheets. When the "[user run mode](#)" option is used, the Sub sheets are not hidden. This makes it possible to re-code the scoring instructions directly in the Sub sheets without having to add *mws and *alt line on the CCs sheet. For more information, write to support@lertap.com.

Iteman4 and jMetrik

Iteman is a program developed by Assessment Systems Corporation. It's a solid item and test analysis program with a pedigree nearly as long as Lertap's.

As good as Iteman4 is, it could not analyze the 15 items featured in this paper. This is because Iteman assumes that the responses for cognitive test items will consist of a set starting with a "1" such as (1,2,3,4), or starting with an "A", such as (A,B,C,D). As we've seen, many of the items used in this study used response codes of (0,1). The zero would throw Iteman off. We'd have

to recode the item responses if we wanted to analyze them with IteMan; for example, we could convert (0,1) to (1,2).

jMetrik is a free item analysis program authored by Patrick Meyer at the University of North Carolina. Its ability to score items is similar to Lertap's in that it has the capability of working with any response code set, even those which include zeros.

The statistics for the first three items as produced by jMetrik are shown below. Note that the correct number of response codes is found for each of these items⁶.

```
=====
```

Item	Option (Score)	Difficulty	Std. Dev.	Discrimin.
i1	Overall	0.4817	0.5000	0.4762
	0.0(0.0)	0.5183	0.5000	-0.6424
	1.0(1.0)	0.4817	0.5000	0.4762
i2	Overall	0.4486	0.4976	0.5654
	0.0(0.0)	0.5514	0.4976	-0.7083
	1.0(1.0)	0.4486	0.4976	0.5654
i3	Overall	0.8139	0.3894	0.3437
	1.0(0.0)	0.1130	0.3168	-0.3913
	2.0(0.0)	0.0559	0.2299	-0.2605
	3.0(0.0)	0.0171	0.1298	-0.1487
	4.0(1.0)	0.8139	0.3894	0.3437

I have more comments on IteMan and jMetrik in [this paper](#).

The actual test items are shown starting on the following page. They're from the following website:

<http://www.edmeasurementsurveys.com/TAM/Tutorials/4CTT.htm>

⁶ jMetrik doesn't need the "Overall" row when right/wrong items are processed but there is no way to turn it off (note how the Overall line and that for the correct answer are the same).

1. Area [ID:N904]

smeebu

Area

The diagram on the right shows a floor plan of a bedroom in a house. The measurements shown are in millimetres. Estimate the area of the bedroom in square metres.

m²

Acceptable numeric values Between 22 and 25

2. Tables [ID:N905]

smeebu

Tables

The photo on the right shows six tables placed together to make a hexagon. Each table is shaped as a trapezium as shown below (not drawn to scale). What is the angle x in the diagram?

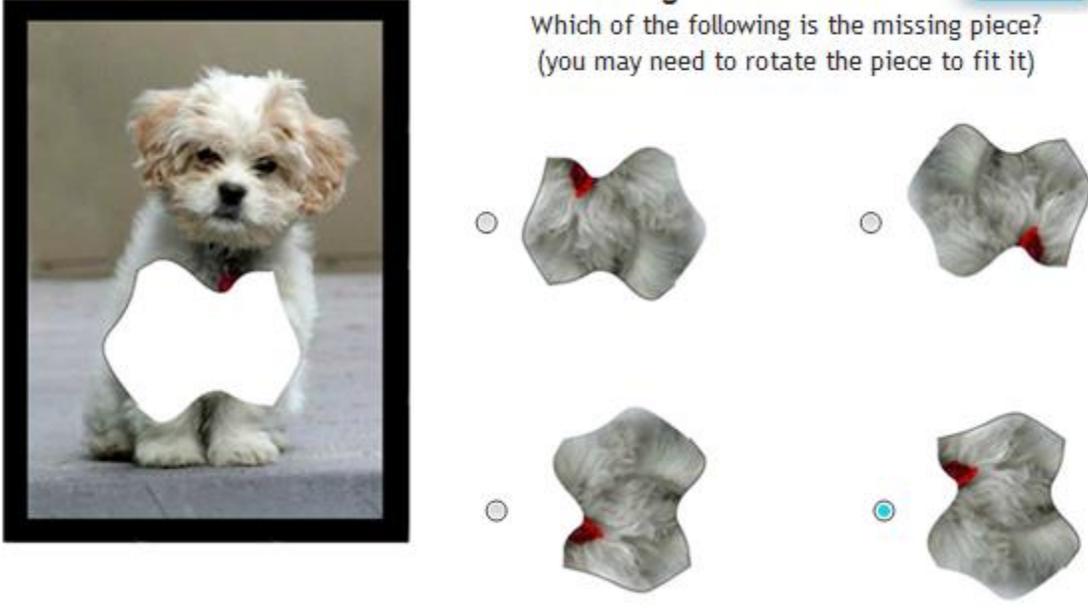
degrees

Acceptable numeric values Equal to 60

3. Jigsaw Puzzle [ID:N907]

Jigsaw Puzzle smeebu

Which of the following is the missing piece?
(you may need to rotate the piece to fit it)

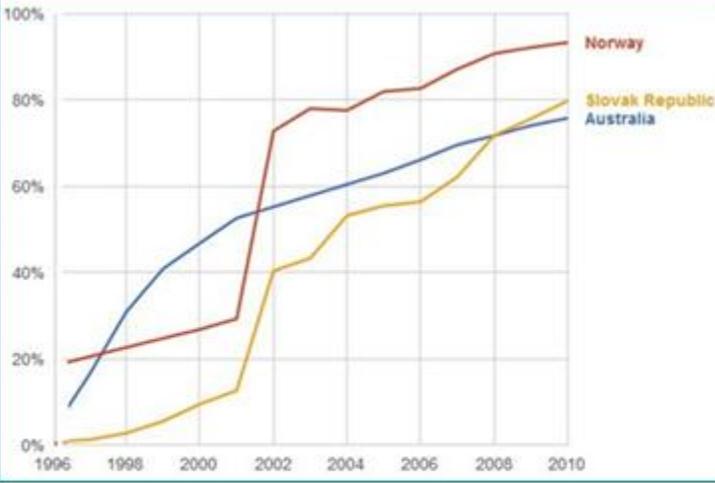


4. Internet Usage [ID:N911]

Internet Usage

The graph on the right shows internet users as percentage of population for three countries, between 1996 and 2010 (Public data from Google).

Between 2002 and 2010, the rate of increase in Slovak Republic is about



Year	Norway (%)	Slovak Republic (%)	Australia (%)
1996	20	0	10
1998	25	5	35
2000	28	10	45
2002	30	15	55
2003	75	40	58
2004	78	45	60
2006	82	55	65
2008	90	70	70
2010	95	80	75

twice the rate in Australia
 3/4 the rate in Australia
 the same rate as in Australia
 half the rate in Australia

5. Grand Prix [ID:N914]

smeebu



Grand Prix

In a Formula 1 race, drivers complete a number of laps with a total distance adding to about 300 km.

It takes about $1\frac{1}{2}$ hours to complete the race.

Estimate the average speed.

Average speed = km / h

Acceptable numeric values Between 180 and 220

6. Liechtenstein [ID:N915]

smeebu

Liechtenstein

Liechtenstein is one of the smallest countries in the world. The total area of Liechtenstein is about 160 km^2 . The distance from north to south is about 24 km. See map. A plane flies across Liechtenstein from west to east at a cruising speed of 600 km/h. Estimate the time taken for the plane to fly across Liechtenstein (from the Switzerland border to Austrian border) .



- 1 minute
- 6 minutes
- 16 minutes
- 30 minutes

7. Wagon Wheel [ID:N917]

smeebu



Wagon Wheel

A wagon wheel has a diameter of 1m.

For each complete turn of the wheel, how many metres does the wagon move?

metres

Acceptable numeric values Between 3 and 3.2

8. Height Estimation [ID:N919]

smeebu



Height Estimation

The ratio of the shadow of a small pole to its real height (see red circle in picture) is 3:5.

To estimate the height of the Hilton flag pole, we measure the length of its shadow.

If the length of the shadow is 4.5 metres, what is the length of the flag pole?

metres

Acceptable numeric values Between 7 and 8

9. Science Tests© [ID:N970]

smeebu

Science Tests©

In Mei Lin's school, her science teacher gives tests that are marked out of 100. Mei Lin has an average of 60 marks on her first four Science tests. On the fifth test she got 80 marks.

What is the average of Mei Lin's marks in Science after all five tests?

Average:

© Source from PISA maths test

Acceptable numeric values

Equal to 64

10. Time Difference [ID:N922]

smeebu

Time Difference

The "world clocks" in the foyer of a hotel are shown below.



Washington (pm)



Sydney (am)



Tokyo (am)



Berlin (am)

When it's 10 o'clock at night in Berlin, what time is it in Sydney? o'clock in the morning

Acceptable numeric values

Equal to 7

11. Probability [ID:N924]

smeebu

Probability

There are 100 Scrabble® letter tiles in the Scrabble game. See picture below.

The letter tiles are placed in a bag. One letter is drawn from the bag at random.

What is the probability that an "A" is drawn?

- 1 / 9
- 1 / 100
- 9 / 100
- 9 / 91

Source: Wikimedia Commons

12. Equation [ID:N927]

smeebu

Equation

Which equation below is illustrated by the above diagram?

- $(a+b)^2 = a^2 + 2ab + b^2$
- $(a+b)^2 = a^2 + b^2$
- $a \times b + a \times b = a^2 + 2ab + b^2$
- $a^2 + b^2 = 2ab$

Source: Wikimedia Commons

13. Choices© [ID:N967]

smeebu

Choices©

In a pizza restaurant, you can get a basic pizza with two toppings: cheese and tomato. You can also make up your own pizza with extra toppings. You can choose from four different extra toppings: olives, ham, mushrooms and salami.

Ross wants to order a pizza with two different extra toppings.

How many different combinations can Ross choose from?

Answer: combinations

© Source from PISA maths test

Acceptable numeric values

Equal to 6

14. Olive Oil [ID:N930]

smeebu

Olive Oil



5 kg of olives are needed to make 1 litre of olive oil.

A picker can pick 6 kg of olives per hour.

How long does it take for the picker to pick enough olives to make 1 litre of olive oil?
(calculate time in minutes)



minutes

Acceptable numeric values

Equal to 50

15. Shapes [ID:N935]

Shapes

smeebu

Ten shapes make up a rectangle as shown on the right.

Which of the following pairs of pieces can be swapped without moving other pieces, and maintaining the shape of the rectangle? The pieces can be rotated, but not flipped over.

(There may be more than one pair)



- 1 and 9
- 5 and 6
- 1 and 10
- 2 and 10
- 2 and 5