Test Analysis with *jamovi*

Larry R Nelson[1]
Curtin University, Western Australia
Document date: 8 February 2024[2]
website: www.lertap5.com

Background

Lertap5 and Iteman are commercial apps used to assess the quality of tests and surveys. I am the author of Lertap5.

There are free versions of both – Iteman's requires Windows in order to operate and is limited to 100 data records. Lertap5's free version, known as the "Mini" version, runs with Windows and Macintosh computers with Microsoft Excel, and is limited to 250 data records.

A test analysis app might well be expected to cover both affective and cognitive domains, that is to say, capable of analysing both survey and test items. However, in this paper my focus will be on the latter, that is, on tests, exams, and quizzes. Furthermore, my discussion will be limited to multiple-choice items.

There have been free alternatives to Iteman and Lertap5 for some time now. Recently my attention has been drawn to two of them in particular, CRAN-based systems known as JASP and *jamovi*. They may not yet be quite as comprehensive, but they clearly have things to offer, are fairly easy to install and apply, and they're free.

I applied JASP in this study. Below I demonstrate the use of *jamovi*.

To do so, I'll use one of the sample datasets from Lertap known as "MathsQuiz", a 15-item multiple-choice test once widely used in a large American school district.

The Step-by-Step Process

Ready, set, go? I will demonstrate *jamovi*'s ability to undertake a classical test analysis using MathsQuiz results[3].

Three things will be needed: (1) a copy of student responses to the items as found in this downloadable MathsQuizData.xlsx file; (2) the correct answers to the 15 multiple-choice items in the quiz (they appear below) and (3), a copy of *jamovi* (get it from here).

(I opted for the "2.4.1 current" *jamovi* desktop version in Windows exe format as offered at the time. I have also used *jamovi* on a MacBook laptop with total success.)

The correct answers string for the 15 items is: 3,4,1,4,2,1,3,2,3,4,1,1,4,1,3

I began by opening *jamovi*. Figure 1 is what I saw.

I had undertaken some *jamovi* trials on another laptop before getting to this point and was already aware that I'd need to install a *jamovi* module called "snowIRT" if I wanted to undertake a bit of item analysis.

[1] Comments / questions may be sent to l.nelson@curtin.edu.au

[2] Click here to reload this page (there may have been an update).

[3] A companion document discusses the use of jamovi with the Rasch model.

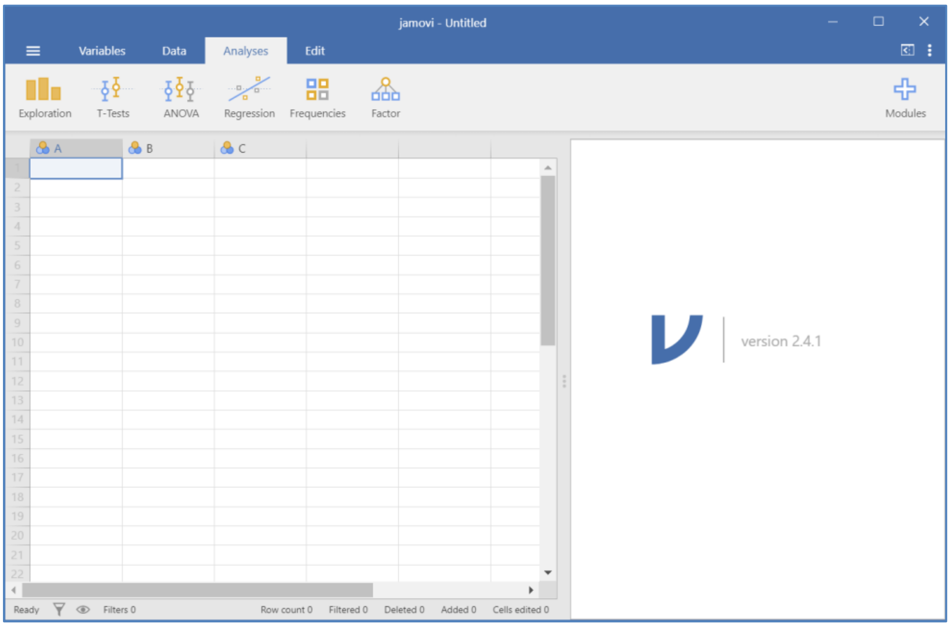


Figure 1

Towards the top of Figure 1, over on the right side, there is a "Modules" option with a large plus sign (+) above it (a bit hard to see). I take that, ask for the "*jamovi* library", and scroll down, down, until I find "snowIRT". I select it and then watch as it is installed. See Figure 2 where snowIRT is now seen as an option to the right of Factor.

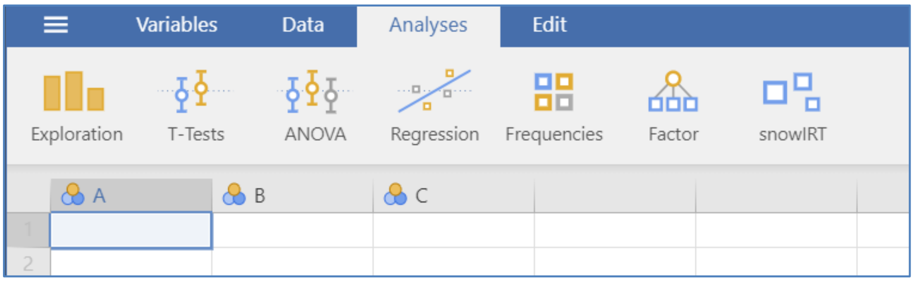


Figure 2

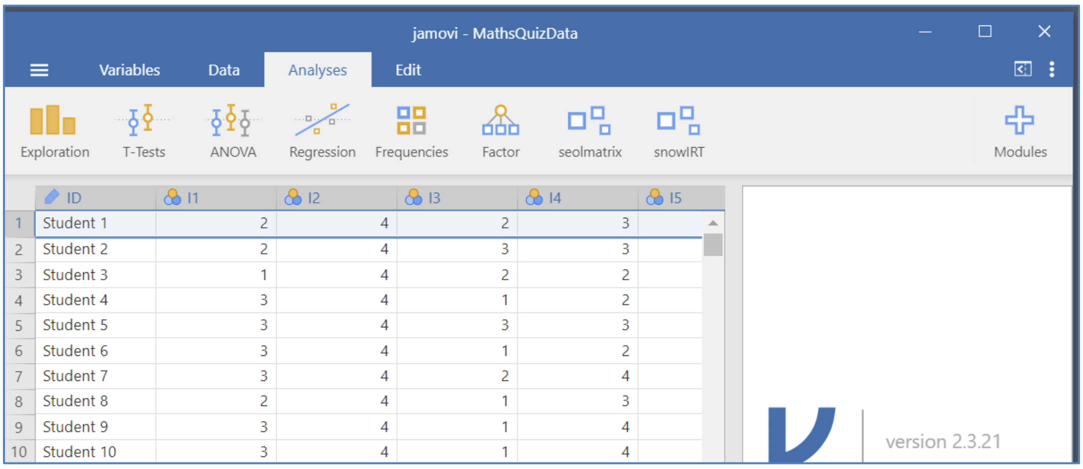
At this stage I was ready to open MathsQuizData.xlsx, the file I had downloaded.

To do so, I click on the three little white bars, above the "Exploration" icon, in the blue band, to the left of "Variables" (see Figure 2).

It gives me an option to "Open" a file. I take it and then "Browse" my computer for the downloaded file.

Figure 3 displays the outcome.

I see partial records for Student 1 to Student 10, and note that all of these students chose option 4 on "I2", the second of the quiz's 15 items.



	ID	I1	I2	I3	I4	I5
1	Student 1	2	4	2	3	
2	Student 2	2	4	3	3	
3	Student 3	1	4	2	2	
4	Student 4	3	4	1	2	
5	Student 5	3	4	3	3	
6	Student 6	3	4	1	2	
7	Student 7	3	4	2	4	
8	Student 8	2	4	1	3	
9	Student 9	3	4	1	4	
10	Student 10	3	4	1	4	

Figure 3

Figure 4 displays the options available once I had selected the snowIRT module.

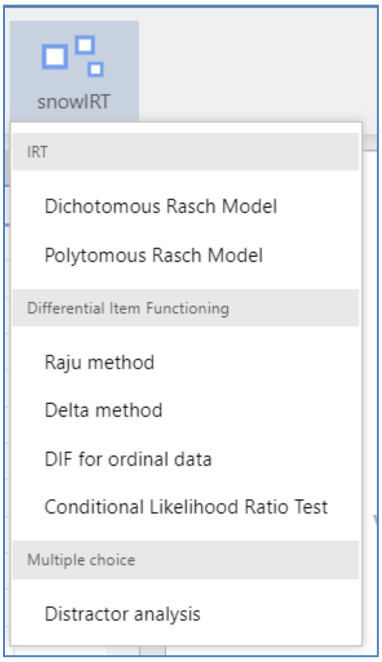
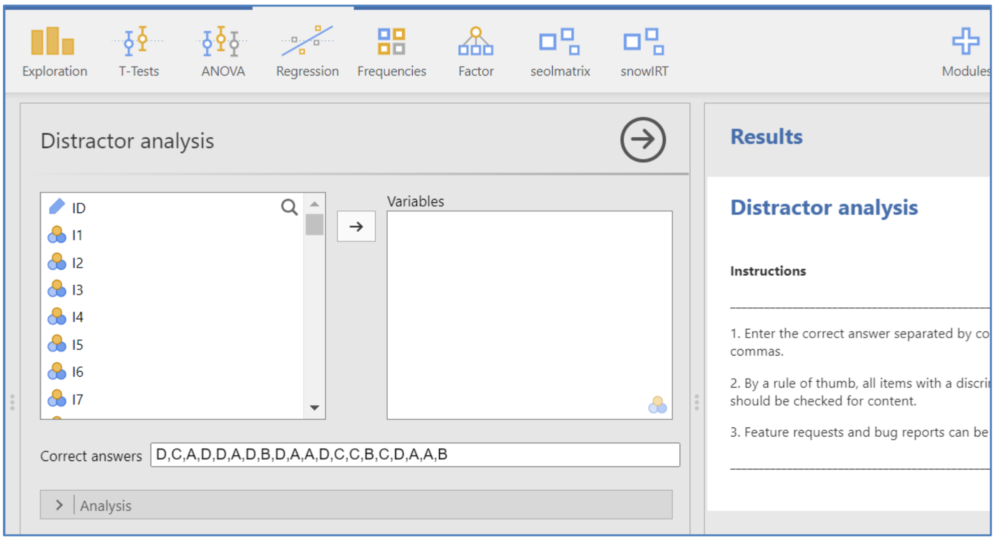


Figure 4

I selected the "Distractor analysis" option seen at the bottom of Figure 4.

Now have a look at Figure 5.

I had nothing to do with the "Correct answers" seen in Figure 5. They appeared automatically. I will have the chance to replace them in a minute.



Distractor analysis

ID
I1
I2
I3
I4
I5
I6
I7

Variables

Correct answers D,C,A,D,D,A,D,B,D,A,A,D,C,C,B,C,D,A,A,B

> Analysis

Results

Distractor analysis

Instructions

1. Enter the correct answer separated by co
commas.

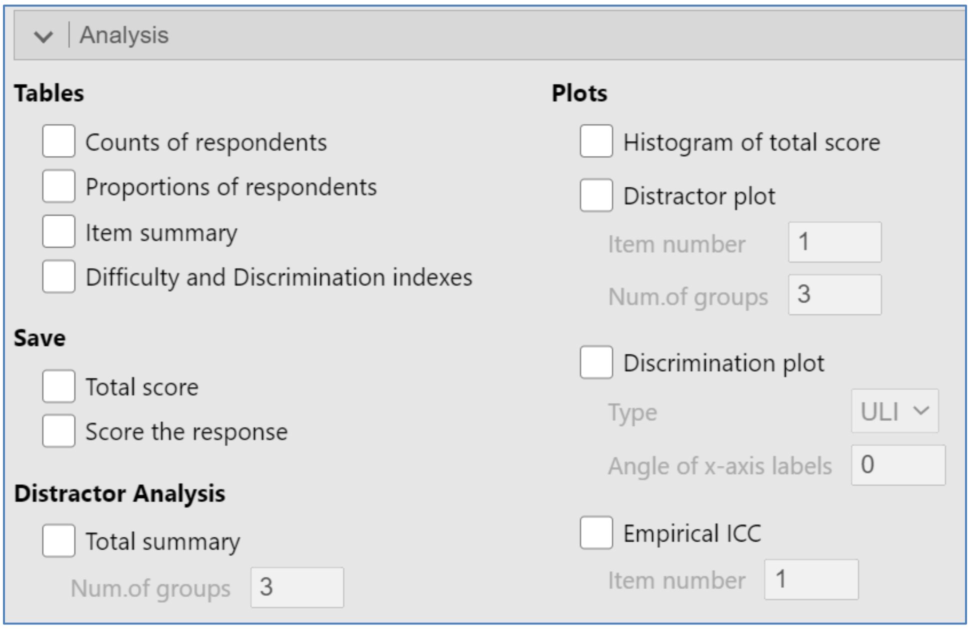
2. By a rule of thumb, all items with a discri
should be checked for content.

3. Feature requests and bug reports can be

Figure 5

I clicked on the >Analysis option seen at the bottom left of Figure 5 (it's not easy to see in the figure, having a grey background and presented in a very small font).

Figure 6 then resulted and my demonstration can *almost* begin. I say almost because I first have to enter the correct answers to the 15 multiple-choice items in MathsQuiz. I do that, replacing the correct answer string seen at the base of Figure 5 with the correct answers I've given above (3,4,1,4,2,1,3,2,3,4,1,1,4,1,3).



Analysis

Tables

□ Counts of respondents
□ Proportions of respondents
□ Item summary
□ Difficulty and Discrimination indexes

Save

□ Total score
□ Score the response

Distractor Analysis

□ Total summary
Num.of groups 3

Plots

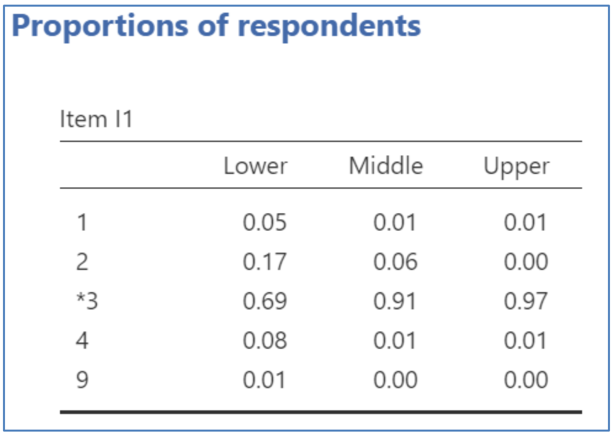
□ Histogram of total score
□ Distractor plot
Item number 1
Num.of groups 3
□ Discrimination plot
Type ULI
Angle of x-axis labels 0
□ Empirical ICC
Item number 1

Figure 6

Proportions of respondents

This is one of the Analysis options seen in Figure 6 (under "Tables") . I click on it and then, looking above at Figure 5, I select I1 I2 I3 down to I15. Following that, using the arrow in the little box seen in Figure 5 between the two display panels, I move all of the items, I1, I2, I3, … I15 over to the right, so that they're in the "Variables" panel.

Results appear immediately – *jamovi* opens a results summary box to the right of the screen. Figure 7 displays results for the first of the MathsQuiz items, Item I1. The asterisk denotes the correct answer to the item.



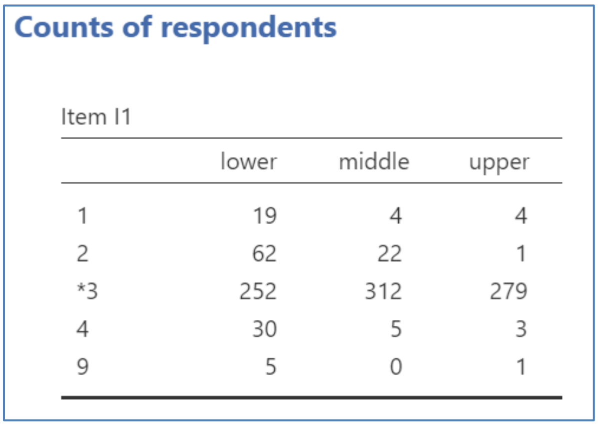
Proportions of respondents

Item I1

	Lower	Middle	Upper
1	0.05	0.01	0.01
2	0.17	0.06	0.00
*3	0.69	0.91	0.97
4	0.08	0.01	0.01
9	0.01	0.00	0.00

Figure 7

As seen in Figure 7, *jamovi* has presented response proportions for three groups: Lower, Middle, and Upper. It wasn't too clear to me how these three groups had been formed so I went back to the Analysis options and selected "Counts of respondents". Figure 8 resulted.



Counts of respondents

Item I1

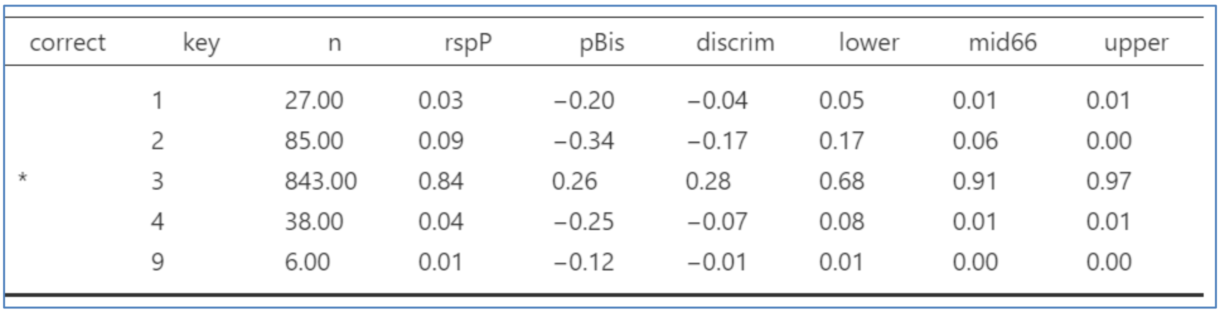
	lower	middle	upper
1	19	4	4
2	62	22	1
*3	252	312	279
4	30	5	3
9	5	0	1

Figure 8

Of the total of 999 students, *jamovi* has put 368 (37%) in the lower group, 343 (34%) in the middle, and 288 (29%) in the upper. The disparity in group sizes results from the restricted range of possible test scores in MathsQuiz, stemming from the small number of test items.

What I might draw attention to at this point is that Item 1 was pretty easy for the 999 students: 97% of the students in the top group got it correct, and even the lower group did pretty well with 69% selecting 3, the correct answer (Figure 7).

Now back to Figure 6 where I then selected "Item summary". This produced a new set of tables. Figure 9 displays the results for the first item, I1.

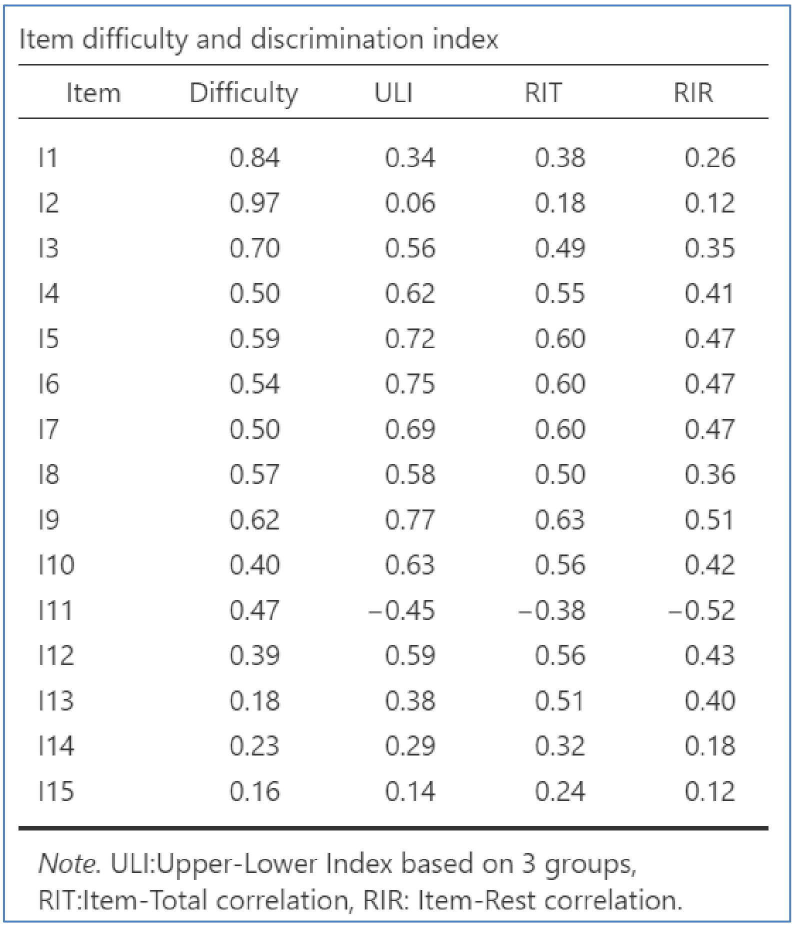


correct	key	n	rspP	pBis	discrim	lower	mid66	upper
	1	27.00	0.03	−0.20	−0.04	0.05	0.01	0.01
	2	85.00	0.09	−0.34	−0.17	0.17	0.06	0.00
*	3	843.00	0.84	0.26	0.28	0.68	0.91	0.97
	4	38.00	0.04	−0.25	−0.07	0.08	0.01	0.01
	9	6.00	0.01	−0.12	−0.01	0.01	0.00	0.00

Figure 9

Note the "lower, mid66, and upper" labels in Figure 9. They agree with Figure 7, and now the "discrim" figure has been derived by subtracting "lower" from "upper". "mid66" would be the proportion of students in the middle 66% (*roughly*) of the test score range who selected each item option.

I now return to the Analysis options shown above in Figure 6 and opt for "Difficulty and Discrimination indexes". I am rewarded with a table I very much like, see Figure 10.



Item difficulty and discrimination index

Item	Difficulty	ULI	RIT	RIR
I1	0.84	0.34	0.38	0.26
I2	0.97	0.06	0.18	0.12
I3	0.70	0.56	0.49	0.35
I4	0.50	0.62	0.55	0.41
I5	0.59	0.72	0.60	0.47
I6	0.54	0.75	0.60	0.47
I7	0.50	0.69	0.60	0.47
I8	0.57	0.58	0.50	0.36
I9	0.62	0.77	0.63	0.51
I10	0.40	0.63	0.56	0.42
I11	0.47	−0.45	−0.38	−0.52
I12	0.39	0.59	0.56	0.43
I13	0.18	0.38	0.51	0.40
I14	0.23	0.29	0.32	0.18
I15	0.16	0.14	0.24	0.12

Note. ULI:Upper-Lower Index based on 3 groups, RIT:Item-Total correlation, RIR: Item-Rest correlation.

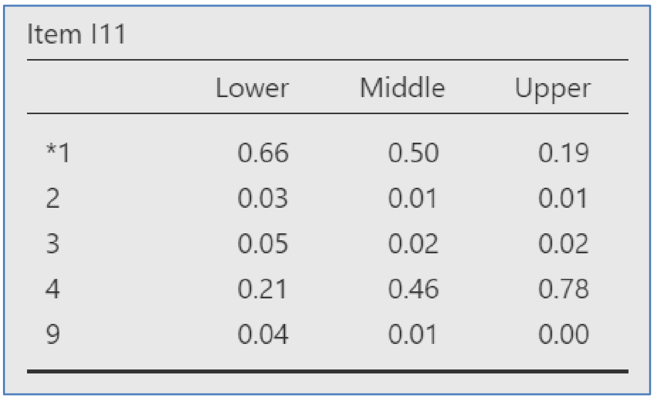
Figure 10

The item performance figures in Figure 10 are the classics of "CTT", classical test theory. I've written about them in the Lertap manual, where the "RIR" in Figure 10 is referred to by its classic label of "item discrimination", being the point-biserial correlation between the item score (0 for wrong, 1 for right) and the test score with the item score partialed out (seen in Lertap's Stats1b and Stats1f reports). The "ULI" in Figure 10 is the upper-lower index found in Lertap's StatsUL reports.

Now, while we're here, let me draw attention to Figure 10's statistics for the eleventh item, I11. The three discrimination indices are all negative!

This is highly unexpected and unwanted, indicating that the students selecting this option, said to be the correct answer, were ones having low test scores.

Keep that in mind as I now go out and get *jamovi*'s "Proportion of Respondents" results for I11 (Figure 11).



Item I11

	Lower	Middle	Upper
*1	0.66	0.50	0.19
2	0.03	0.01	0.01
3	0.05	0.02	0.02
4	0.21	0.46	0.78
9	0.04	0.01	0.00

Figure 11

What does Figure 11 say? The keyed-correct answer was 1 according to the figure. In the strongest group of students, the "Upper" group with the highest test scores, only 19% chose what has been defined as the correct answer. Almost 80% of the students in the Upper group regarded 4 to be the correct answer to item I11.

And they were right! The string of correct answers, seen just above Figure 6 has an error. The correct answer was indeed 4, not 1.

Test Reliability

To this point I have been using the Distractor Analysis option as found in the version of the snowIRT module available at the time I was working on this paper (version 4.9.4). It did not have an option for calculating the reliability of the MathsQuiz test.

But all is not lost. Following the kind advice of the snowIRT author, I go back to the Analysis options seen in Figure 6 and select both options under 'Save': 'Total score' and 'Score the Response'. Something seemed to happen after doing this, there was some flickering on the screen, however I was unable to detect any new results.

But then, as advised, I understood that I could at this stage use the 'Reliability Analysis' option as found in the Factor module. (This paper exemplifies the use of that option when processing student responses to survey items.)

So it is that I now go to the Factor module and click on 'Reliability Analysis'. I found that, when it started up, it came ready to work my item scores, as seen in Figure 12.

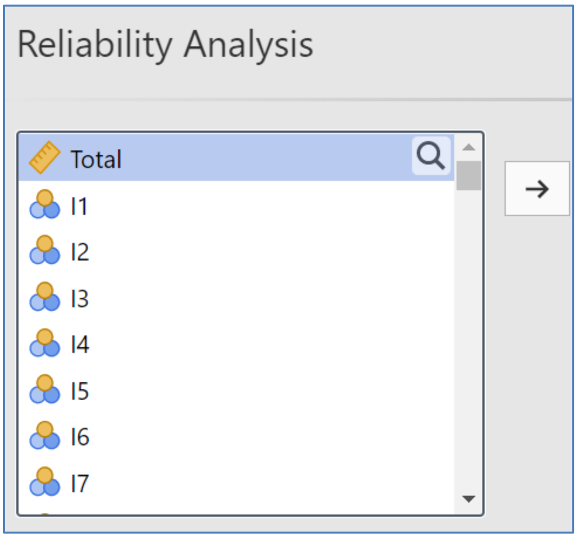
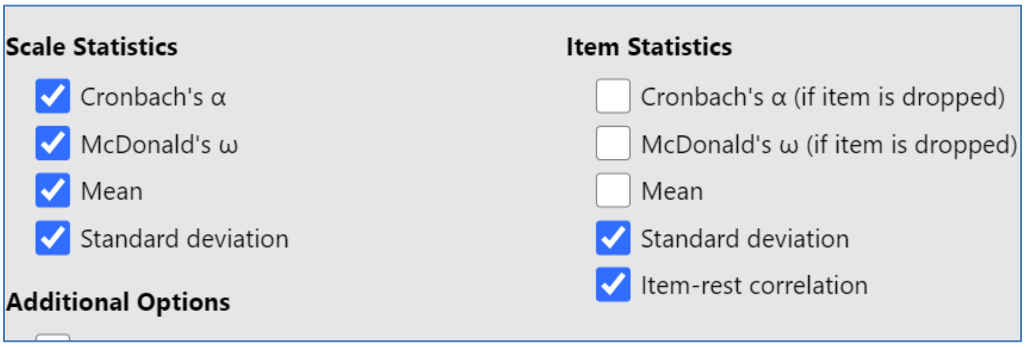


Figure 12

I select the options shown below in Figure 13; these options appeared beneath the panel shown above in Figure 12.



Scale Statistics

- ☑ Cronbach's α
- ☑ McDonald's ω
- ☑ Mean
- ☑ Standard deviation

Additional Options

Item Statistics

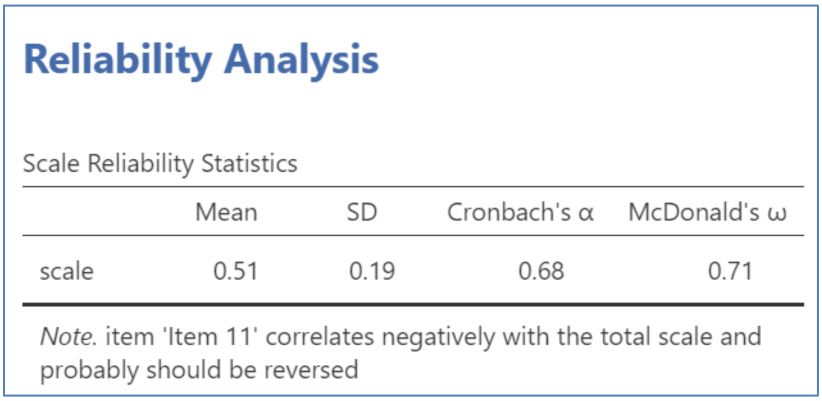
- □ Cronbach's α (if item is dropped)
- □ McDonald's ω (if item is dropped)
- □ Mean
- ☑ Standard deviation
- ☑ Item-rest correlation

Figure 13

Next, exactly as I did in the Distractor Analysis, I move Item1 through Item15 from the left-hand side over to the 'Items' box on the right.

The output seen below in Figures 14 and 15 resulted.

The little *Note* seen at the base of Figure 14 is instructive. There is a problem with Item 11.



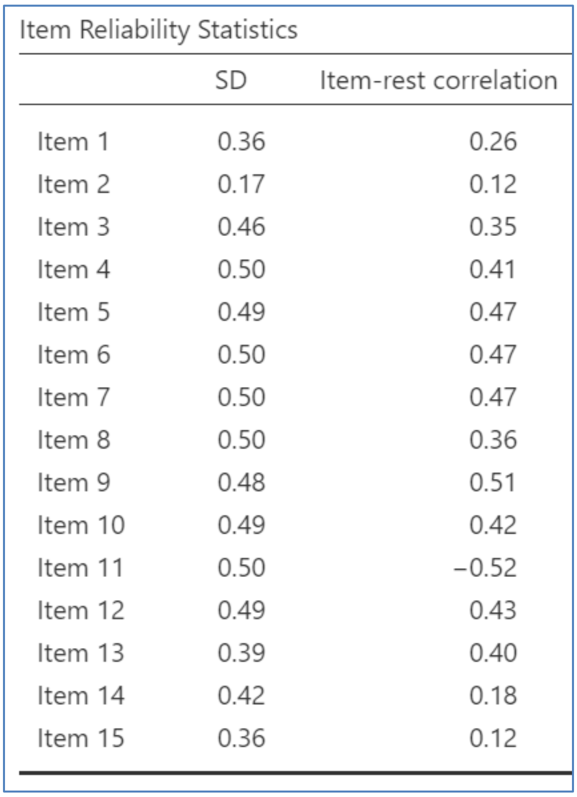
Reliability Analysis

Scale Reliability Statistics

	Mean	SD	Cronbach's α	McDonald's ω
scale	0.51	0.19	0.68	0.71

Note. item 'Item 11' correlates negatively with the total scale and probably should be reversed

Figure 14



Item Reliability Statistics

	SD	Item-rest correlation
Item 1	0.36	0.26
Item 2	0.17	0.12
Item 3	0.46	0.35
Item 4	0.50	0.41
Item 5	0.49	0.47
Item 6	0.50	0.47
Item 7	0.50	0.47
Item 8	0.50	0.36
Item 9	0.48	0.51
Item 10	0.49	0.42
Item 11	0.50	−0.52
Item 12	0.49	0.43
Item 13	0.39	0.40
Item 14	0.42	0.18
Item 15	0.36	0.12

Figure 15

Now, continuing to focus on Item 11 (I11), I'd like to find out what the test reliability figures would be if I reversed the scoring for that item. Doing so will correct the error made in the item keys: an item score of zero (for "wrong") will become a score of 1 (for "right"), and a previous score of one will become a zero.

It's very easy to reverse score items with the Reliability Analysis option in the Factor module. Below the "Additional Options" seen at the bottom of Figure 12 there's a panel to reverse scale selected items, see Figure 16.

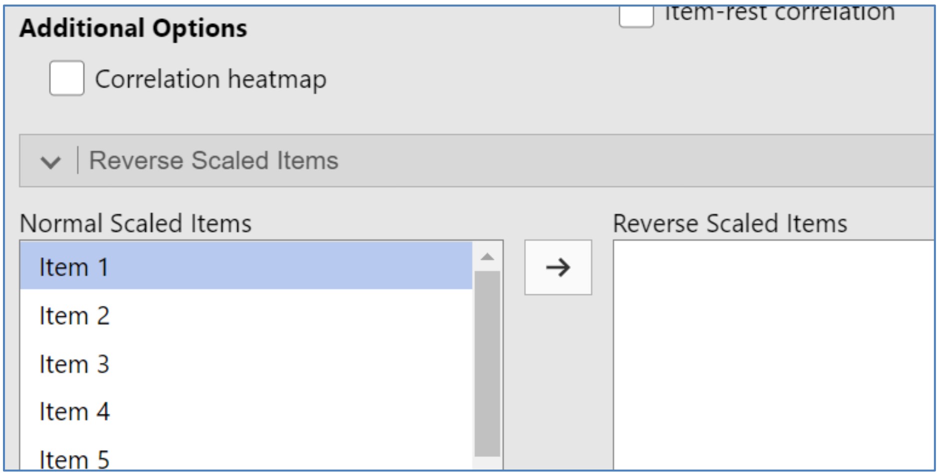
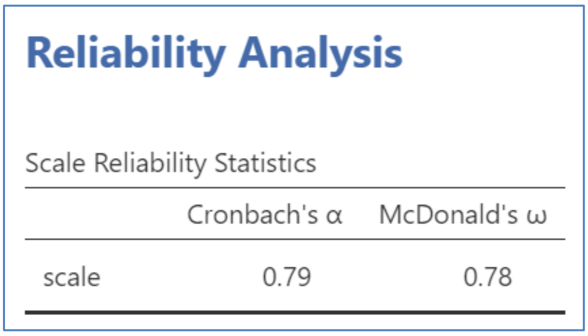


Figure 16

Using the panel on the left in Figure 16, I move down to Item 11 and then, using the arrow, move it over to the "Reverse Scaled Items" side. Figure 17 displays the result: both alpha and omega reliability estimates have increased. (In theory alpha will never be greater than omega but that doesn't always hold, as mentioned in this paper.)



Reliability Analysis

Scale Reliability Statistics

	Cronbach's α	McDonald's ω
scale	0.79	0.78

Figure 17

Error!

Now, just a minute here. I've made an error in this work and I would bet some readers may have spotted it.

I started off with a string of correct item answers, a string having an error.

That string was: 3,4,1,4,2,1,3,2,3,4,1,1,4,1,3

I have highlighted the error in yellow. The keyed-correct answer to the 11th item was not 1, it was 4.

The correct string is: 3,4,1,4,2,1,3,2,3,4,4,1,4,1,3

Reversing the scoring for Item11 is not the best idea. Think of how things are at this stage: I asked one of the Distractor Analysis options seen in Figure 6 to "Score the response". That will create a string of 15 item scores for each student. Item scores will be (0,1) for (wrong, right) for every item. Now, look again at the response frequencies:

Item 11 Response	Number of Students
1	473
2	18
3	30
4	459
9	19

With incorrect scoring, that is, with Item 11's correct answer said to be 1, there will be 473 students getting an Item 11 score of one and 526 getting a score of zero. If I ask the Reliability routine to *reverse* the scoring for this item, there will then be 526 getting a score of one, and 473 a score of zero.

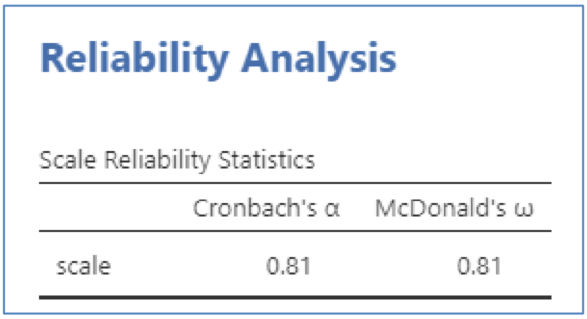
However, that's not the correct picture.

The number who got Item 11 right was 459, not 526. The extra 67 students (526 – 459) come from 18 + 30 + 19, from the other incorrect answers.

To correct this situation, I will go all the way back to the start, way back to the Distractor Analysis options, put in the new string of correct answers, and undertake all of the analyses again (easy to do).

That will then see me coming into the Reliability routine in the Factor module with the right number of correct answers for Item 11.

Once I have done this, I find the following reliability estimates:



Reliability Analysis

Scale Reliability Statistics

	Cronbach's α	McDonald's ω
scale	0.81	0.81

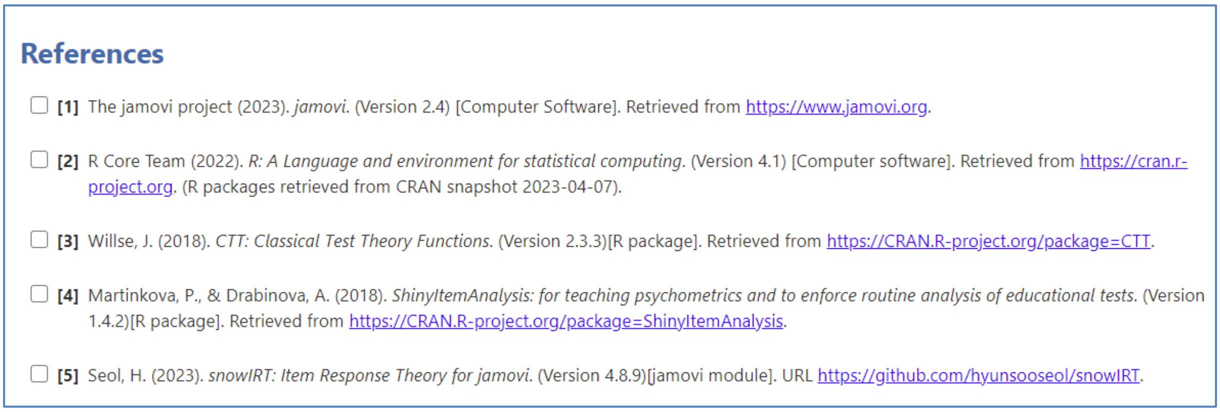
Figure 18

Wrapping up

I would use *jamovi* with item analysis classes if it were not possible to have students use Lertap5, an app which requires Microsoft Excel. And, having said this, of note is that the free version of Lertap5 is restricted to a maximum of 250 data records. There is no limit on the number of data records in *jamovi* that I know of.

Something I have not mentioned above is that the Distractor Analysis routine in the snowIRT module has support for application of the Rasch IRT model. Lertap5 may also be used for Rasch analysis as discussed here.

References for the snowIRT module are seen in the screen shot below.



References

☐ **[1]** The jamovi project (2023). *jamovi*. (Version 2.4) [Computer Software]. Retrieved from https://www.jamovi.org.

☐ **[2]** R Core Team (2022). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from CRAN snapshot 2023-04-07).

☐ **[3]** Willse, J. (2018). *CTT: Classical Test Theory Functions*. (Version 2.3.3)[R package]. Retrieved from https://CRAN.R-project.org/package=CTT.

☐ **[4]** Martinkova, P., & Drabinova, A. (2018). *ShinyItemAnalysis: for teaching psychometrics and to enforce routine analysis of educational tests*. (Version 1.4.2)[R package]. Retrieved from https://CRAN.R-project.org/package=ShinyItemAnalysis.

☐ **[5]** Seol, H. (2023). *snowIRT: Item Response Theory for jamovi*. (Version 4.8.9)[jamovi module]. URL https://github.com/hyunsooseol/snowIRT.