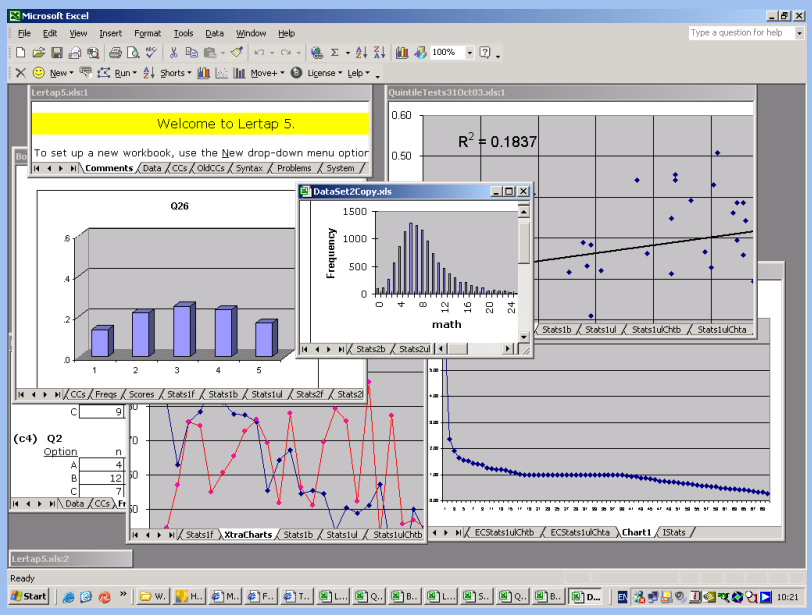


Lertap 5 with Excel

What it's all about

Interactive PDF version

Made in Western Australia

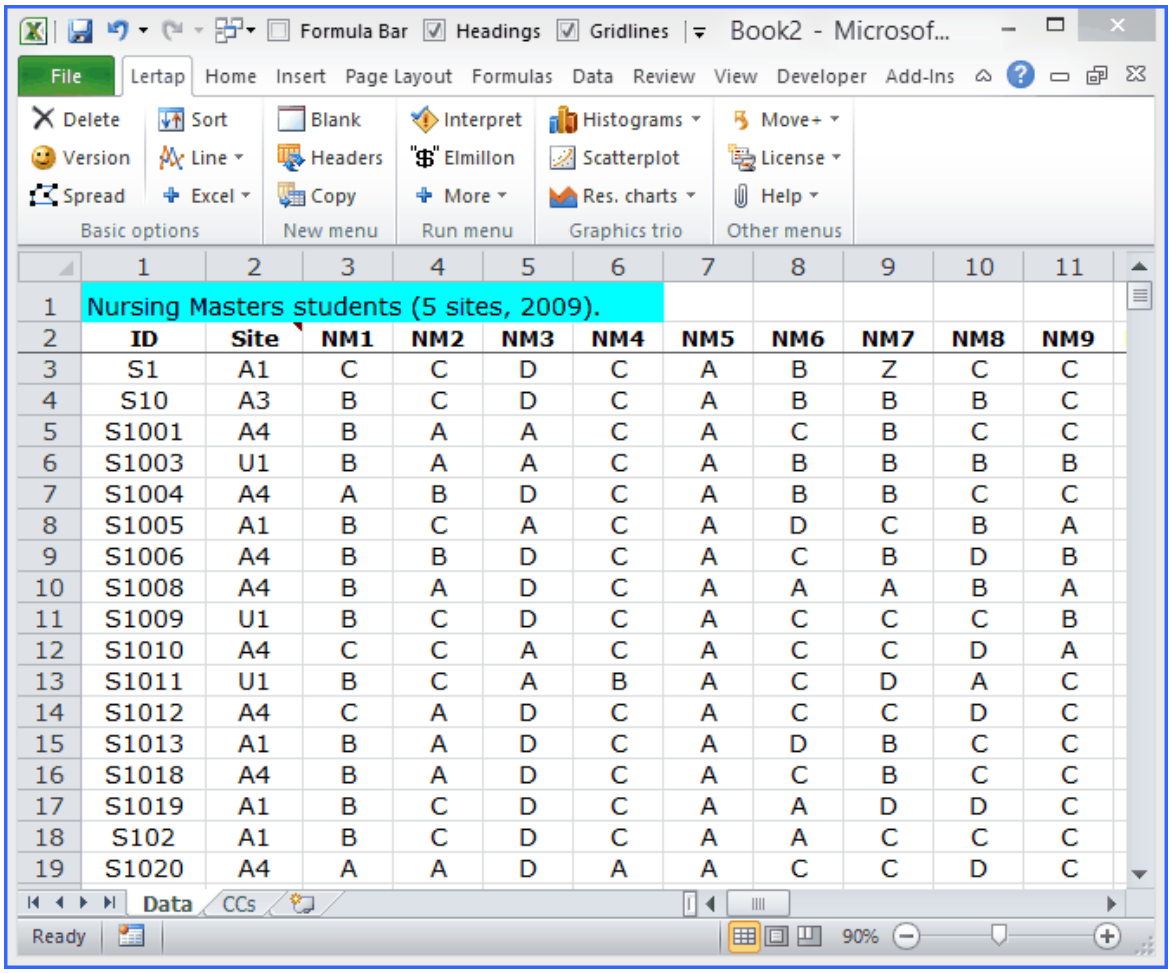
Table of Contents

1 Introduction

Lertap is a computer program used to process and analyze data collected from quizzes, exams, tests, and surveys. Lertap is an acronym for the Laboratory of Educational Research Test Analysis Package.

Lertap5, the latest "genre", works on both Windows and Macintosh computers using Microsoft's spreadsheet program, "**Excel**". How to get it? Refer to this link.

A screen shot of a typical Excel worksheet set up to work with Lertap5 is seen below. The test items (in this case) were numbered from NM1 to NM40 (more common labels would be Q1 to Q40, or perhaps I1 to I40):

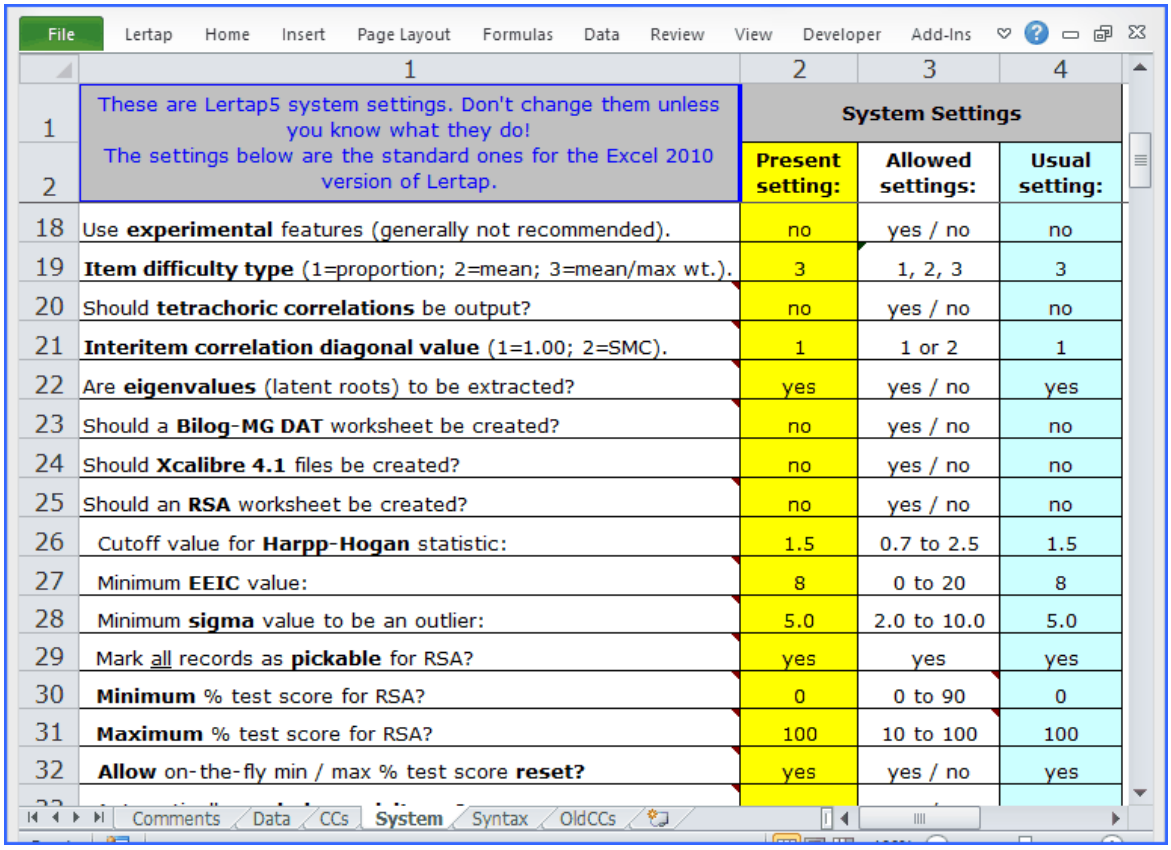


	1	2	3	4	5	6	7	8	9	10	11
1	Nursing Masters students (5 sites, 2009).										
2	ID	Site	NM1	NM2	NM3	NM4	NM5	NM6	NM7	NM8	NM9
3	S1	A1	C	C	D	C	A	B	Z	C	C
4	S10	A3	B	C	D	C	A	B	B	B	C
5	S1001	A4	B	A	A	C	A	C	B	C	C
6	S1003	U1	B	A	A	C	A	B	B	B	B
7	S1004	A4	A	B	D	C	A	B	B	C	C
8	S1005	A1	B	C	A	C	A	D	C	B	A
9	S1006	A4	B	B	D	C	A	C	B	D	B
10	S1008	A4	B	A	D	C	A	A	A	B	A
11	S1009	U1	B	C	D	C	A	C	C	C	B
12	S1010	A4	C	C	A	C	A	C	C	D	A
13	S1011	U1	B	C	A	B	A	C	D	A	C
14	S1012	A4	C	A	D	C	A	C	C	D	C
15	S1013	A1	B	A	D	C	A	D	B	C	C
16	S1018	A4	B	A	D	C	A	C	B	C	C
17	S1019	A1	B	C	D	C	A	A	D	D	C
18	S102	A1	B	C	D	C	A	A	C	C	C
19	S1020	A4	A	A	D	A	A	C	C	D	C

A quick overview of how Lertap5 operates, and samples of the reports it produces, may be seen by looking at this set of slides. (Also available as a PDF file.)

Lertap5's forte is item analysis and test scoring. The majority of the statistics it produces are solidly based on **CTT**, classical test theory; **IRT** analyses are also supported. It will handle polytomous cognitive and affective items (such as rating scales), with extensive support for partial-credit item scoring. A virtual cornucopia of charts and graphs is a salient feature of Lertap5.

Users may tailor Lertap5's operations and output by setting their preferred options in the "**System worksheet**". A display of part of this worksheet is provided below:



	1	2	3	4
1	These are Lertap5 system settings. Don't change them unless you know what they do! The settings below are the standard ones for the Excel 2010 version of Lertap.	System Settings		
2		Present setting:	Allowed settings:	Usual setting:
18	Use **experimental** features (generally not recommended).	no	yes / no	no
19	**Item difficulty type** (1=proportion; 2=mean; 3=mean/max wt.).	3	1, 2, 3	3
20	Should **tetrachoric correlations** be output?	no	yes / no	no
21	**Interitem correlation diagonal value** (1=1.00; 2=SMC).	1	1 or 2	1
22	Are **eigenvalues** (latent roots) to be extracted?	yes	yes / no	yes
23	Should a **Bilog-MG DAT** worksheet be created?	no	yes / no	no
24	Should **Xcalibre 4.1** files be created?	no	yes / no	no
25	Should an **RSA** worksheet be created?	no	yes / no	no
26	Cutoff value for **Harpp-Hogan** statistic:	1.5	0.7 to 2.5	1.5
27	Minimum **EEIC** value:	8	0 to 20	8
28	Minimum **sigma** value to be an outlier:	5.0	2.0 to 10.0	5.0
29	Mark all records as **pickable** for RSA?	yes	yes	yes
30	**Minimum** % test score for RSA?	0	0 to 90	0
31	**Maximum** % test score for RSA?	100	10 to 100	100
32	**Allow** on-the-fly min / max % test score **reset?**	yes	yes / no	yes

The next page, "Features", has more information on many of Lertap5's capabilities.

A paper discussing and comparing some of the features of several free or inexpensive item analysis programs, Lertap5, SAS University, and jMetrik among them, is here. The paper also makes mention of Iteman, Xcalibre, and Bilog-MG.

Links to a variety of Lertap resources are listed below.

1 A PDF copy of this website's topics. A CHM copy (compiled help file for Windows). An iBook copy, ready for reading on an iPad or an iPhone. A link to

the website itself.

2 A small set of PowerPoint slides with an introduction to Lertap. These are also available as a PDF file.

3 Some "Tips & Tricks" for users. Demonstrates the use of quintile plots; showcases selected Excel and Lertap features.

4 The online help system for Lertap. A primary source for finding out how to obtain Lertap, how to get it running, and understanding the features added after the manual was printed years ago.

5 Sample datasets for downloading. These also show off selected special features, and showcase some of Lertap's most popular charts.

6 The Lertap 5 e-store.

Please direct questions or comments to: lertap5@gmail.com

Last update: **5 November 2023**

1.1 Features

Feature-packed is Lertap 5. Have a browse of the topics listed below to get an idea of some of its features and capabilities.

- It's easy to use
- Heaps of documentation
- Simple but really powerful control language
- Graphs? (*we got 'em, even with a gestalt capability*)
- Context-sensitive help
- Sample datasets? *You bet!*
- Certification and licensing is here
- Looking for group differences (DIF+)
- Cheating detection
- Comprehensive item flags

- Advanced stuff

1.1.1 Not hard to use

Lertap 5 is an Excel "app". Data may be entered directly into an Excel spreadsheet via the computer's keyboard. Data may also be imported by reading files produced by, for example, bubble sheet scanners.

Excel has very good facilities for data "snooping", providing handy tools, such as "Filter" and "Pivot tables", for looking at data quality. Lertap has some too -- see this little discussion with examples.

Lertap 5's output, its numerous tables and charts, are all nested in well-formatted Excel worksheets. Print them easily. Save them to PDFs if you want. Add more charts and graphs readily. Reformat and re-colour them when wanted.

Use of Excel for data input, output, and manipulation results in an "all in one" application. This is worth noting as a great number of other item analysis programs require that data be prepared by using another application -- some even make it necessary to use an external application (such as a word processor) to read their output.

An "all-in-one" app is considerably more convenient and very often substantially easier to use.

Data analysts frequently want to use more than one software system in order to look at their data in other ways. Excel worksheets have become a "lingua franca", a common "language", facilitating data sharing among a variety of "apps", such as **SPSS** and **SAS**. The factor analysis routines in these programs will happily look at a Lertap 5-created, Excel-based, matrix of inter-item correlations. Should you be using a program which has not been trained to read Excel worksheets, you might well find that Lertap's **Iteman**, **Xcalibre**, and **Bilog-MG** interfaces will suit your needs well. More information about these interfaces may be found at this topic.

Not an Excel expert? Worry not. Have a gander and a chuckle with Lertap 5's user guide and up and running you'll be. (You may be likely to chuckle as it's not a conventional manual.)

1.1.2 Lots of documentation

You'd expect a software system to have a manual, wouldn't you? That's how our Lertap 5 documentation started. You ought to make it a point to read at least a bit of it -- it's more than a manual, more than a how-to, and, perhaps best of all, the author frequently had his funny bone active when typing. It's got real examples, and

in numerous spots serves as a comprehensive technical reference. You can get to it via this link.

After the system was initially released in the year 2001 (officially), a series of technical documents, referred to as "erudite epistles" began to emerge. These explain more about using Lertap 5 in practical contexts.

Lertap 5 has continued to grow; many enhancements have been added since its initial release. Instead of updating the manual with every enhancement, a dedicated website called "**Lelp**" (for Lertap help) was brought online to cover changes and provide related reference material. It's updated fairly often, and it's all yours in a variety of formats. Don't miss it, see it by clicking here.

In 2011 a special website was developed with numerous sample datasets. Almost all of these are from real-life applications of Lertap 5, and all but one may be downloaded as an Excel workbook, ready for practice, experimentation, and class research projects. They've become very popular.

Another dedicated website came online in 2013 to demonstrate a variety of Lertap 5 and Excel "tips and tricks" in the context of an authentic national assessment of science achievement.

1.1.3 Easy control language

Lertap 5 uses both a "command language" and menu-based clicks to accomplish the tasks users have in mind.

Basic commands and instructions are entered in an Excel worksheet called "**CCs**" (for "control cards").

These two command lines would set up scoring instructions for a ten-item multiple-choice test:

```
*col (c2-c11)
*key ACCDBBCCAD
```

These two lines would define scoring instructions for a ten-item rating scale:

```
*col (c12-c21)
*sub affective
```

These three command lines would set up instructions for scoring a 40-item multiple-choice criterion-referenced test with a cut-off score of 80%:

```
*col (c2-c41)
*sub mastery=80
*key ACCDBBCCAD BCDBBAACCC CABDADBCDA BCDBBAACCC
```

Pretty simple, eh? In all of these examples, the *col line tells Excel where item responses have been entered in the Data worksheet. The *key line gives the keyed-correct answer for cognitive items. The optional *sub line is used for a variety of purposes. In the second example above it's used to tell Excel that items are to be scored in the "affective" manner (as seen in, for example, rating scales). In the third example it has been employed in order to activate the mastery-test scoring option.

The control language is not complicated. Several examples may be seen here. Note that the use of a control language such as Lertap 5's provides real parsimony in most cases -- had Iteman 4 been used in the third example, forty rows would have to be entered in a table or spreadsheet simply to describe the items (where they are in the input file, and how to score them).

Once scoring instructions have been entered in the CCs worksheet, the options in the Lertap 5 Excel tab come into play. A complete example is presented in these on-line PowerPoint slides, also available as a pdf file.

1.1.4 Flexible item scoring

Any test item may use as many as 26 responses. For cognitive questions, default responses are {A,B,C,D} while for affective questions such as those found on rating scales, default responses are {1,2,3,4,5}.

An item response may have any "weight", or scoring points, and these may be negative.

Cognitive test items may have more than one keyed-correct answer.

Rating scale items may have their scoring polarity reversed via the simple use of plus and minus signs. Examples may be browsed at this topic.

"EMQs", extended-matching questions, are supported (more information is here).

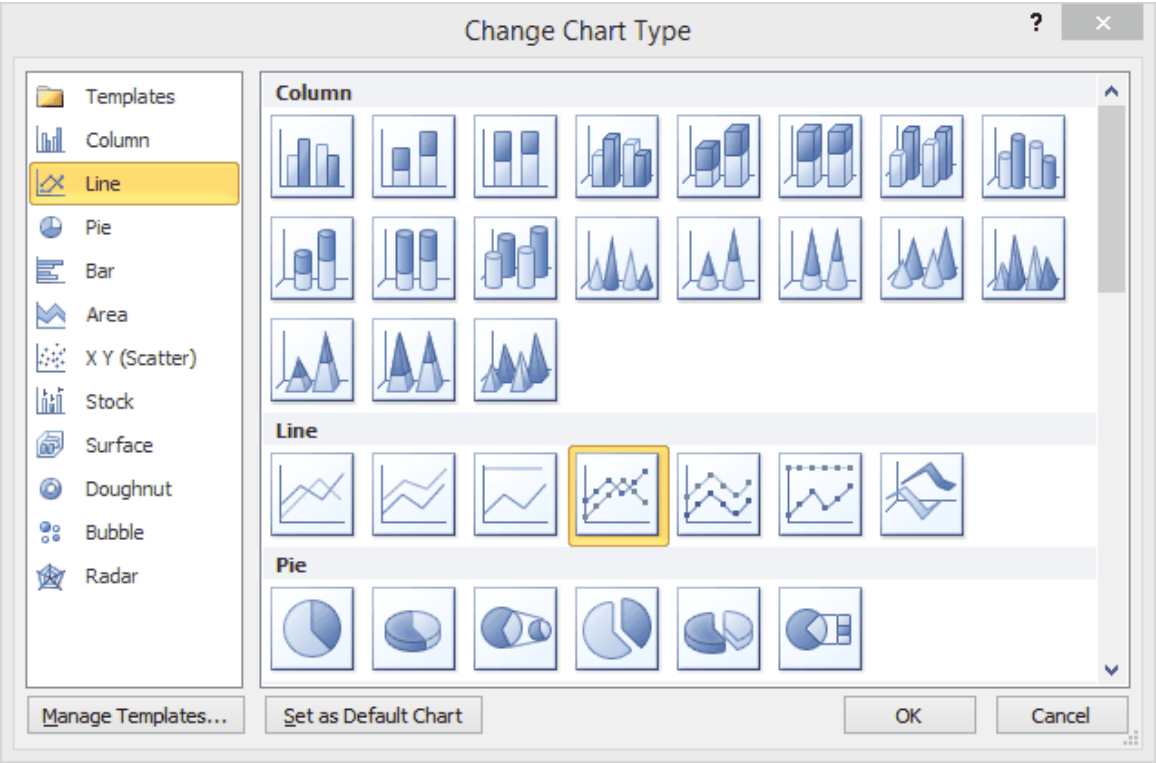
See this topic for more about responses, responses code, and response scoring.

We take dinkum pride in suggesting that Lertap 5 has the most flexible and easiest to use item scoring of any test analysis program.

1.1.5 Graphs, charts, and plots

A major feature of Lertap 5 is its Excel base. Excel is well known for its ability to make what it calls "**charts**", frequently referred to in Lertap 5 as graphs or plots.

A browse of this topic will serve to exemplify some of the most commonly-used Lertap 5 charts. **Quintile plots** of response trace lines, a crowd pleaser, are demonstrated here, and, more recently in this January 2015 presentation of enhancements.

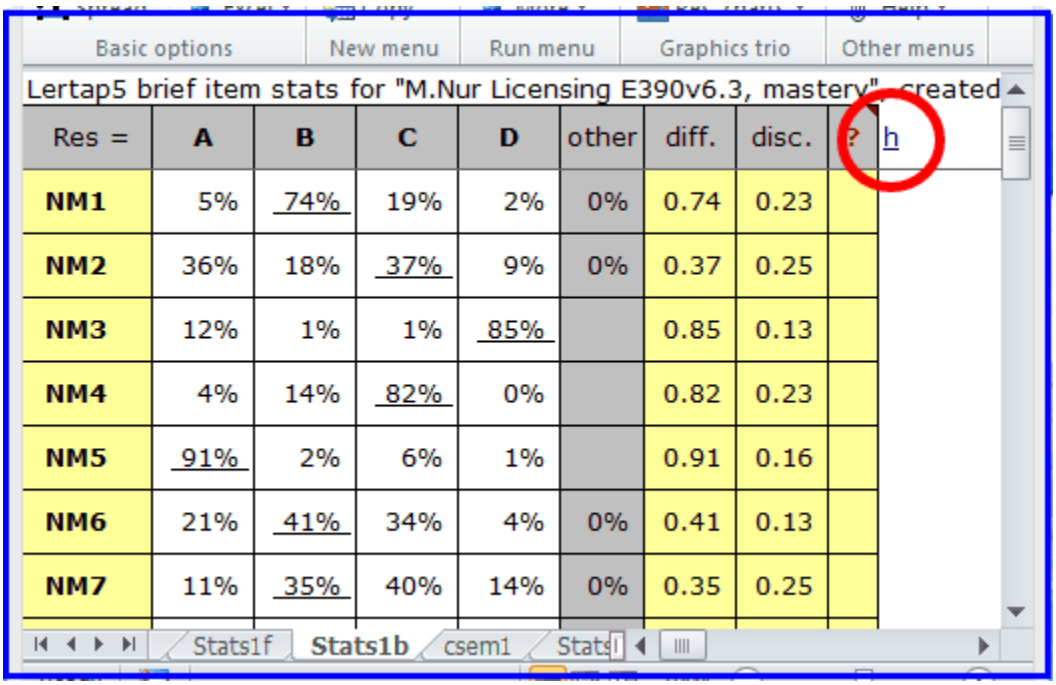


Any of the charts, graphs, and plots created by Lertap 5 may be edited in a manner which many will find familiar: select the chart and then choose from what is often an extensive list of options, such as resizing, colouring, or changing even the "type" of the chart (without losing its link to the data it's based on). Lertap 5 at times provides its own options for changing charts -- here is an example.

1.1.6 Context-sensitive help

The snapshot below is from one of Lertap 5's standard reports for cognitive test items, "**Stats1b**".

How to interpret Stats1b's output? Look for the little h as seen in the screen snapshot below-- doing so will activate the web browser on your computer, getting it to display supporting documentation as you may see by clicking here. (Please note that the little h captured in the red circle below is not "hot" -- it's not an active hyperlink, just an example.)



Res =	A	B	C	D	other	diff.	disc.	?	h
NM1	5%	74%	19%	2%	0%	0.74	0.23		
NM2	36%	18%	37%	9%	0%	0.37	0.25		
NM3	12%	1%	1%	85%		0.85	0.13		
NM4	4%	14%	82%	0%		0.82	0.23		
NM5	91%	2%	6%	1%		0.91	0.16		
NM6	21%	41%	34%	4%	0%	0.41	0.13		
NM7	11%	35%	40%	14%	0%	0.35	0.25		

The great majority of Lertap 5 reports give "little-h" access to help in this manner.

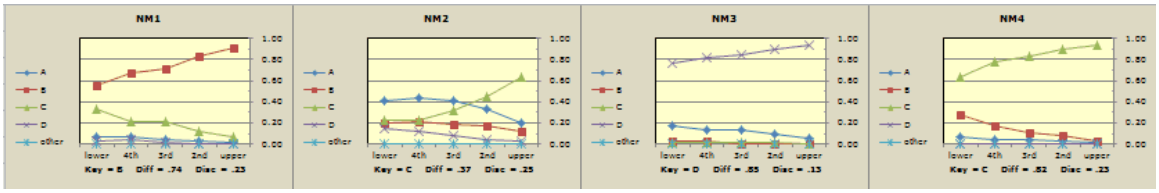
1.1.7 Gestalt output

There are undoubtedly numerous item analysis programs available these days: search the internet -- there will be thousands of "hits". Of them, **Lertap** and **Iteman** are unique in that they have pedigrees dating back to the 1970s. They're probably among the most popular and feature-rich item analysis programs in use today.

Lertap 5 and Iteman 4 are similar in that they output similar statistics and, to some extent, similar graphs. But they differ greatly in both the format and the quantity of their output. Lertap 5 is "gestalt oriented"; it presents output in a relatively concise,

colorful manner organized in what may be called "**themes**". In contrast, Iteman 4's output is organized by items, and is without doubt comparatively verbose.

As an example, Lertap 5 will collect item-response plots and present them as a group (in this example, the 'theme' is "**packed plots**"):



Iteman 4, in contrast, presents similar plots on a one-by-one basis, one page for each test item. Examples of Iteman4 output may be seen here.

More comments on Lertap 5 and Iteman 4 may be found in this document. Mention of other item analysis programs is made in this topic, and also in this one, a special for students and instructors.

1.1.8 Datasets galore

A number of **real-life datasets** are available for browsing, giving insight as to how Lertap 5 has been applied in a variety of settings. It's possible to download all but one or two of these datasets, providing the chance to experiment with authentic test and survey results -- for instructors, useful class exercises may be built around these datasets.

You don't have to have Lertap 5 running on your computer in order to be able to appreciate and caress the datasets -- all you need is a spreadsheet program, with Excel the most logical (but perhaps not the only) one to use.

Our sample datasets website also has a free **Spearman-Brown** calculator, and discussions about interfacing Lertap results with **SPSS** and **Xcalibre**.

All these wonders may be found by admiring this website.

1.1.9 Certification and licensing

Lertap 5's **mastery mode** is used extensively by sites engaged in certification and licensing.

Click here to read about mastery mode and its numerous associated features / options. Also, have a look at this topic to see how to link Lertap 5's output to a form used by **NCAA**, the National Commission for Certifying Agencies.

A practical example of a nursing school's use of mastery mode may be found at this webpage.

1.1.10 Group differences

There are options in Lertap 5 which make it possible to compare results from groups of respondents (such as students). Results for groups may be "broken out" on the basis of a test score, or on the basis of item responses. Lertap 5 presents tables and corresponding graphs.

A "**DIF**" analysis, differential item functioning, may also be undertaken using Lertap 5.

Here are related links:

How will Lertap 5 breakout scores by groups? Have a look at two examples which demonstrate how to get results for groups of test takers: a comparison of undergraduate student responses to a study-habits inventory; and looking at age differences on a test of science achievement.

What about getting "**breakouts**" for item responses? Well, you might take in an example from a national assessment of science achievement, showing how item responses differed by grade level. You could also admire another example: comparing item responses by private- and public-school students.

An example showing Lertap 5's support for **DIF**, differential item functioning, may be seen in this topic.

1.1.11 Cheating detection

Lertap 5 has a feature which may be used to see if student test results may show evidence of cheating. It's called "**RSA**", response similarity analysis. Read about RSA by clicking here, and then perhaps take in a practical, real-life example by having a look at this topic where RSA suggested that two students had overly-similar results, leading to a confession of cheating.

1.1.12 Comprehensive item flags

Statistical programs sometimes use special "**flags**" to draw attention to an unusual or unexpected outcome. When there are pages and pages of results, these flags facilitate the task of finding out if anything went awry, making it possible to quickly find areas in need of attention.

Lertap 5 has numerous flags particularly useful when it comes to assessing how **item distractors** have worked. Examples are presented in this paper.

It also has options which control how selected statistical indices are flagged -- for example, should a chi-square value be flagged when it reaches the .01 level, the .05, or .10? Options such as this are set in Lertap 5's "System" worksheet.

1.1.13 Advanced features

Estimates of 2PL **IRT** (item response theory) parameters may be obtained by activating a special system option. Getting Lertap 5 to prepare data and control files for **Xcalibre 4**, **Iteman 4**, and **Bilog-MG**, is discussed in this topic.

A **Rasch** analysis for cognitive test items, such as those found in multiple-choice quizzes and tests, may be processed using this special macro.

A comprehensive interface for the **IRT** routines in **SAS** is discussed in this topic, while a handy **IRT** routine made to use with Excel, "**EIRT**", works very well with Lertap and is easy to use.

Factor analysis is supported via a capable **principal-components** option. We have an introductory discussion, and a practical example. Interfacing Lertap 5's inter-item correlation matrix to **SPSS** is demonstrated here.

Interested in calculating **McDonald's omega** statistics (another way of estimating reliability)? Lertap has support, have a look.

Support for online testing systems and LMS, learning management systems, has been developed, including Angel Learning (now part of Blackboard) and Pearson VUE. Special support modules are available for working with such systems. An example based on Angel Learning is mentioned here, while another example, focused on **PearsonVUE**, may be enjoyed with a click here.

We have support for users of Pearson VUE's **Exam Developer** system too. A Lertap option will see that item statistics and test results are carried over to the special Excel workbook used in Exam Developer.

Test scores may be **rescaled** or transformed. Read how to do this.

Categorical variables may be **recoded** readily, using a Lertap 5 option, or by using standard Excel features. Categorical variables include, for example, Gender, Native Language, Grade Level, School District, and so on. Read more about recoding such variables.

An **external score** may be used as the criterion measure for calculating item statistics. This is accomplished by activating the "external criterion option".

It is possible to actually program Lertap 5 so as to incorporate site-specific capabilities. Programs are written using "**VBA**". VBA Excel macros are placed in a Lertap 5 companion workbook, **Lertap5MacroSetA.xlam**. Each macro is then linked to Lertap5's Excel tab. For help and examples, see these links the Macs menu; and a very practical example based on an online testing package.

Preparing "**ASCII**" data files for importing into Excel / Lertap 5 is exemplified in a couple of places. Excel has a "data importer" which is powerful and not difficult to use. A more involved importing procedure is exemplified in this technical paper.

A "did-not-see" option may be used to adjust test scores when students are not presented with all test items. Read about the use of this option when **missing data** are also present.

A correction for chance scoring option, "**CFC**", may be applied to adjust test scores for the effects of guessing. A correction of this sort is also commonly known as "**formula scoring**". Mention of this option is made in Example C6, and it's also covered in Chapter 10 of the manual.

1.2 How to Lertap

How to Lertap? How's it done? Is it difficult to learn?

Nah, it's not difficult. It's simple, really. You start it up, tell it to make a new workbook with two worksheets, and away you go.

Into one of these worksheets, the one called "**Data**", you put test or survey results. Into the other worksheet, "**CCs**", you type up the so-called "control cards" which tell Excel how to process your stuff.

Could we have an example, please?

And why not. Here's a common setup:

Calculus quiz of 15 April 2014

ID	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15
Student 1	2	4	2	3	1	1	4	3	1	4	3	5	2	4	2
Student 2	2	4	3	3	2	3	4	1	3	2	1	3	1	3	1
Student 3	1	4	2	2	1	2	4	4	1	2	1	2	3	1	2
Student 4	3	4	1	2	1	1	3	2	3	2	5	3	5	2	1
Student 5	3	4	3	3	2	1	4	1	1	2	1	3	1	1	3
Student 6	3	4	1	2	3	2	4	1	3	2	1	1	1	1	4
Student 7	3	4	2	4	1	2	4	1	1	2	1	2	1	2	1
Student 8	2	4	1	3	1	2	3	2	1	4	4	4	2	3	1
Student 9	3	4	1	4	2	2	3	2	3	4	1	1	2	1	1
Student 10	3	4	1	4	1	3	4	5	1	2	4	2	3	2	3
Student 11	4	4	2	4	2	4	4	1	1	4	1	4	1	1	1
Student 12	3	4	2	4	1	2	1	1	2	4	1	3	1	4	1
Student 13	3	4	1	3	2	2	2	2	3	4	4	3	3	2	1
Student 14	3	4	3	1	1	5	2	1	4	3	1	3	3	5	4
Student 15	2	4	1	3	2	1	1	2	3	4	1	3	3	4	9

```
*col (c2-c16)
*sub res=(1,2,3,4), title=(Calculus7)
*key 34142 13234 11413
```

In this example, students answered 15 multiple-choice items having to do with calculus. Their responses were entered in columns 2 through 16 on the Data worksheet -- the *col (c2-c16) line tells Excel this.

Each question (or item) presented four response options for the student to pick from. The res=(1,2,3,4) tells Excel that the response possibilities were labeled 1 to 4.

The *key line gives the correct answer for each of the 15 questions.

Simple enough? *Sure thing.*

How to get results, such as test scores and item statistics?

Got it covered -- you've got your internet connection up and running? Good, that means you can have a look at some slides -- if you're not particularly keen on watching PowerPoint slides via the internet, well, here's a link to a pdf version of the same slides. Some videos are also available, just click here.

Might there be some sample data already set up so we can see how it runs?

But of course. Lertap 5 comes with sample Data and CCs sheets built in. Why not take the "Cooks Tour"? Another idea, ideal for when you might be caught out in a bus, train, or plane and you've got a tablet with you (other than a sleeping tablet): a pdf copy of Chapter 2 of the manual (if the sleeping tablet didn't put you to sleep, Chapter 2 might; but, if you stick with it, you'll be able to give lectures on Lertap at the end of your journey). ***But wait!*** There's even more: a whole website with more sample stuff; some of it is pretty neat.

Yes, it looks just the thing we need, but what about some sample data we can have some practice with, you know, some item responses we can type into Excel, just to get a real feel for this marvelous Lertap thing?

Neat idea. Here's a link to sample data from a survey, and here's another with sample data from a chemistry quiz. (Enjoy, take your time, I'll just nip out for some coffee.)

Alright, already! We did those exercises, and Lovely Lertap gave us lots of results to look at. Is there some sort of guide to help interpret the output?

Not a problem. There are some whiz-bang webpages with guides and snapshots of some of Lertap 5's results screens -- start here.

And there's a manual, too. The survey references would be, above all, Chapter 8 of the manual, and a more-recent real-life example of a **rating-scale**. As to looking at results from quizzes and exams, a must-read, *you-gotta-read-it!*, is the plaudit-earning, acclaimed Chapter 7 of the manual.

Okay, just great stuff, but one more thing if you don't mind -- we often use scanners to process student test results. How do we get our scanner's output into Excel for use with Lertap 5?

Another good question, thanks, you people are clearly on the ball. Have you not trained your scanner to create results files in Excel format? That's commonly done; have the scanner save results in "**csv**" format and then make use of the Lertap csv importer.

Wonderful, wonderful, thank you. All schools in our district will have to have Lertap 5!

Hmmmm, I guess we just might be able to handle that. Let me refer you to another topic, how-to-get-it.

1.3 Recent happenings

Developments of possible interest:

- **Free alternatives to Lertap5**

  Readers may know that the free version of Lertap5, the so-called "**Mini**" version, is limited to processing no more than 250 data records. Yes, one can have a Lertap5 Data sheet with many more records, but the Mini version will stop reading them after the 250-th record.

  It is possible to get around this limitation with the purchase of a license and corresponding "unlock code" at a current cost of $78.

  Another expense concerns the need to have Microsoft Excel in order to run Lertap5. Some sites may be able to obtain a site license for all staff/students to freely use, subject to licensing conditions. And, at times, old versions of Excel may be offered by vendors at reduced prices.

  But there are definitely alternatives, *free ones*, fairly capable up to a point. This paper presents one (***Jamovi***), and refers to another (**JASP**).

- **Added two "practical exhibits"**

  These "exhibits" showcase examples of Lertap5 in action, one for a cognitive test and another for an affective scale on depression. They are primarily introductory in nature but do delve a bit into more advanced topics. Both provide the chance to download copies of actual data.

  The cognitive test is called "Test13" and is available here. The depression inventory (or scale) is a very popular, true-to-life scale and may be accessed via this link.

- **Missing data now highlighted in two ways**

  The first way: cells with missing data will have a background highlight in the Data sheet. This makes it very easy to get an idea of the extent and location of missing data.

  The second way results when users ask Lertap5 to create a "csv" file for subsequent input to R and RMD scripts, and/or to JASP and SPSS. When Lertap5's Omega1 macro creates this file, "NA" values will be written to indicate missing data. (NA is interpreted as "value not available" in a number of scripts and apps. It's a universal missing data code.)

  Read more about it here.

- **Updated the TAM Rasch analysis Rmd module**

  The TAM Test Analysis Module is a CRAN Project package described here. Rmd code modules are scripts written to work with the RStudio system.

  An example of the output created by the updated Rasch-Analysis-TAM.Rmd module may be admired here, while a general presentation regarding the use of the module is here.

- **Documented the performance of Lertap5's Rasch item analysis**

  Of all the data analysis capabilities in Lertap5, the Rasch item analysis routine is unique in that it involves an iterative procedure, making substantial demands on a computer's central processing unit. Examples of sample running times may be seen here.

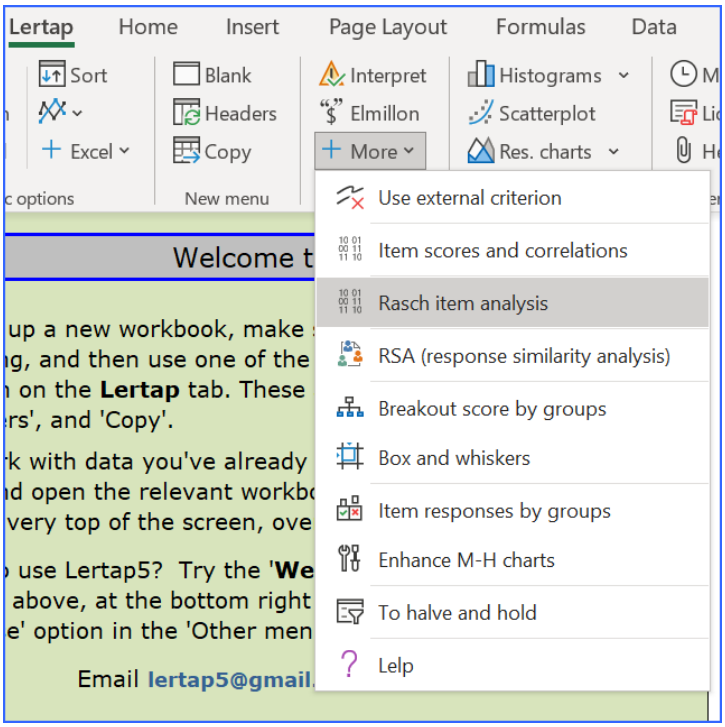
- **Updated the "Alpha Omega Spot the Difference paper"**

  See it here. The paper is no longer sitting on the fence. Limitations previously found in the the JASP app have been addressed and corrected, making it the presently-recommended method for calculating coefficient omega.

  The revised paper also addresses use of the "closed-form" method for calculating omega, a method now built into **SPSS** and Lertap5 itself, finding limitations that will rule out its use in some cases.

- **Improved Lertap5's Rasch support**

  Lertap5's Rasch analysis capability has been enhanced. Now considerably easier to use, and a bit faster, it has been added as an entry in the "Run" menu (formerly it was nested within the special macros list and was not all that easy to find).



- **Created Lertap5 Version 5.11**

  This new version runs on both Windows and MacOS computers. (The former versions were 5.10.9.4 for Windows, and 5.10.99.4 for MacOS. These versions are no longer supported.)

  Having now just a single version for both platforms has been made possible by Microsoft working to bring the Mac version of Excel up to par with the Windows version (or very nearly so).

  Lertap 5.11 may be downloaded from this page.

- **Updated the set of introductory slides**

  Formerly they just focused on cognitive instruments (tests, exams, quizzes). Now they also include affective ones (surveys, scales, questionnaires).

  They're found here (nested in a PDF file).

- **Added plots of item correlation values for affective items**

  Cognitive items have had a difficulty-by-discrimination plotter for years. Now there's a plotter for affective item correlations. Rush to see examples here.

- **Lertap5 now has its own omega reliability estimator**

  The closed-form method of estimating McDonald's omega reliability figure has been added to the Windows version. Refer to Appendix D in this working paper for a discussion and example. The "Item scores and correlation" option is used to produce it.

- **Added a paper related to the calculation of alpha and omega reliability estimates**

  The paper uses SPSS, JASP, and the psych R package to derive alpha and omega estimates for 16 datasets. As has been found in other studies, alpha and omega most often turn out to be very similar in value. The paper goes on to include omega hierarchical and exemplifies its interpretation. Despite the similarity between alpha and omega values (in most cases), the paper suggests that omega is preferred.

  JASP is recommended (with reservations). Read the paper by following this link.

- **Updated documents having to do with using R and RStudio.**

  R and RStudio are free data analysis tools capable of advanced statistical analyses currently beyond Lertap5's ken.

  There are special macros in Lertap5 charged with preparing R and RStudio scripts for IRT and instrument reliability work -- they will, for example, ease the process of getting ICCs (item characteristic curves) and computing coefficient omega.

  A link to documentation on one of these special macros is here, while another link exemplifies how to use files prepared by the macro with R and RStudio. An example of the output produced by one of the RStudio scripts is available if you click here (results in a Word document downloading to your computer).

- **Assisted with the application of Lertap5 to compare exam results from sites using on-site proctoring to results collected under remote proctoring conditions during the COVID19 pandemic.**

  The research question in this study concerned exam integrity: did it appear that remote proctoring may have made it easier for examinees to cheat?

  Exam results were analyzed using Lertap5's RSA, "response similarity analysis", with Wesolowsky's "similarity checking" method used, where possible, for corroboration.

Data were collected pre-COVID19, and during the March to May period of 2020 when the pandemic was in full stride. The pre-COVID results were collected at certified test centers with on-site proctoring. Many of these centers closed down during the COVID19 isolation period, giving rise to an increased reliance on the use of "remote proctoring".

The results from several thousand students, pre-COVID19 with on-site proctoring, did not show evidence of cheating. However, a near-equal number of examinees sitting the same exams under remote proctoring conditions did exhibit signs of possible cheating in a relatively small (but not insignificant) number of cases.

A link to the complete "**updates summary**" page is here.

2 Resources

A distinguishing characteristic of Lertap 5 is its extensive variety of supporting documents, samples, and examples. Just page ahead, resourcefully, to find out.

2.1 The manual

Links to the Lertap 5 manual and related references

The original manual for Lertap 5 was printed by the School of Education, Curtin University, in the year 2001.

A second edition of the manual has been anticipated for some time, but it's not finished yet. In the meantime, the original manual is still very highly recommended. It remains one of the best Lertap 5 resources, especially when it comes to understanding the results produced by the program (Chapters 7 and 8 talk about how to make sense of Lertap 5's various reports). It is also recommended as an entertaining read -- try one of the first chapters to see why.

Of course Lertap 5 has grown since the manual was printed. Many enhancements have been added. These are presented and discussed in "**Lelp**", the Lertap help system. Many of them are also showcased in the "**Sample datasets**" website.

Use the links below to download chapters from the manual and/or to have a look at related information as found in Lelp and the sample datasets.

- **Chapter 1**
  "*Lertap Comes Alive as Version 5*"
  (Similar information found here: **Sample datasets**)

- **Chapter 2** (updated October 2015)
  "A Cook's Tour of Lertap 5"
  (Another view of the Tour is here: **Sample datasets**)

- **Chapter 3**
  "Setting Up a New Data Set"
  (More details related to this topic are here: **Lelp**)

- **Chapter 4**
  "An Overview of Lertap 5 Control Cards"
  (Another presentation of this topic: **Lelp**)

- **Chapter 5**
  "Control Cards for Cognitive Subtests"
  (More examples and explanations: **Lelp**)

- **Chapter 6**
  "Control Cards for Affective Subtests"
  (Another approach with more examples: **Lelp**)

- **Chapter 7**
  "Interpreting Lertap Results for Cognitive Tests"
  (Chapter 7 is a top reference, but more information is in **Lelp**)

- **Chapter 8** (updated 3 September 2019)
  "Interpreting Lertap Results for Affective Tests"
  (Chapter 8 is a top read, another reference is found in the **Sample datasets**)

- **Chapter 9** (updated 4 September 2019)
  "Lertap, Excel, SPSS, and others"
  (A related discussion is here: **Sample datasets**)

- **Chapter 10**
  "Computational Methods Used in Lertap 5"
  (For more information on this topic, see **Lelp**)

- **Chapter 11**
  "A History of Lertap"
  (A very similar presentation is here, in this very document)

- **Appendix A**
  "The Original (1973) Lertap Quiz"
  (Not covered anywhere else.)

- **All**
  The complete original manual (203 pages)

2.2 Other guides

The manual is far from the only user guide for Lertap 5. The references and links below point to additional resources for both tyros and experienced users.

- "**Lelp**"

  Lelp stands for "**Lertap help**". It's an extensive online reference for Lertap 5, serving to update the manual in numerous spots, and to cover new features and capabilities added after the manual was printed.

- **Sample datasets**

  A collection of datasets is available for online viewing. Most are in Excel workbooks, ready to use with Lertap. The majority of the datasets are from authentic studies, and many may be downloaded.

- **Tips & tricks for Lertap 5 users**

  This document traces the analysis of data from the administration of a national assessment of science achievement in a developing country. As it does, a variety of special Excel and Lertap features are highlighted.

2.3 Erudite epistles

This section expands on a few particularly popular themes related to the application of Lertap 5.

- In visual item analysis you'll see an example of the use of response trace lines. **Haladyna & Rodriguez** (2013) (and others) have suggested that these trace lines are a very useful way to judge how well cognitive items have performed, an alternative to looking at tables of item statistics.

- Mastery tests often use one or more cut-scores to determine whether or not candidates have met minimum performance criteria. Lertap has had support in this area for decades. Read more.

- An examination of how item responses and test scores vary over groups of respondents is the subject of two epistles, one that includes DIF, differential item functioning, and another which gets into such things as boxplots.

- Cheating? Have test respondents perhaps been able to "share" answers? Read more.

- Test reliability is a classic topic in measurement. This epistle delves more into this time-tested topic.

- Lertap began its career as a tool for applying "CTT", classical test theory, to analyze responses to tests and surveys. But it also supports "IRT", item response theory.

- We all know that tests and surveys commonly use multiple-choice question formats. This topic discusses ways to score "supply" questions, ones where respondents write (or type) their answers.

2.3.1 Visual item analysis

Visual item analysis with quintile plots

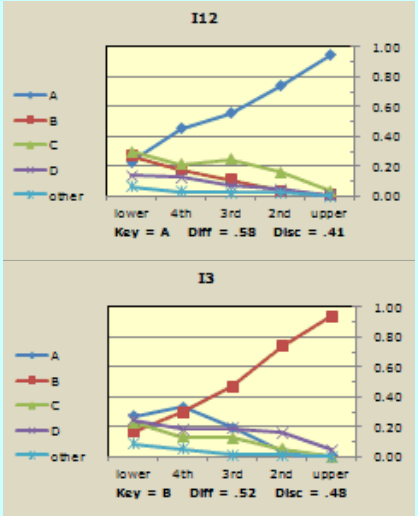
The matter of assessing the quality of cognitive test items has traditionally been based on tables of numeric data, embroidered with core global measures, such as estimates of test reliability.

Lertap's **quintile plots** provide an alternative method: pictures.

Have a look at some examples, and see if you too might not be found wearing smiles for quintiles. Click here to have an initial look. See "**packed plots**" in action here. See fancy new capabilities introduced in January 2015.

Then, pour yerself another cuppa something, sit back, and take in a short practical example from some quality achievement tests developed and used in **Central Java** ('Jateng').

Make a leap for **Lelp** (Lertap help) to see even more about quintiles and associated Lertap 5 options.



2.3.2 Certification and licensing

The *Standards for Educational and Psychological Testing* (1999), published by the American Educational Research Association, recommend the use of 'special' statistics when the measurement process involves the use of cut scores. Mastery, **licensing**, and **certification** tests are examples of applications which typically use cut scores, often on a pass-fail basis.

Lertap 5 produces all of the 'special' statistics recommended in the *Standards,* and then some. We've got a whiz-bang, top-flight paper which you'll not want to miss if you hope to make the cut.

NCCA, the National Commission for Certifying Agencies, has a special report form used to summarize results from mastery (or pass/fail) exams. A paper which indicates how Lertap's output links in with the information requested by NCCA is here.

Lelp (Lertap help) has these topics covered too.

The screen snapshot below displays part of Lertap 5's report for a certification test.

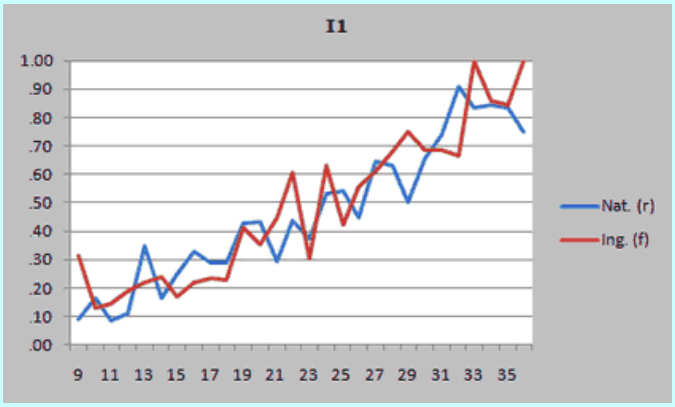
Variance components

	df	SS	MS
Persons	1768	1751.56	0.99
Items	59	3340.98	56.63
Residual	104312	18124.56	0.17

Hoyt's reliability coefficient:	0.825	
CSEM at the cut score:	3.288	
Livingston's coefficient:	0.830	
Index of dependability:	0.799	
Estimated error variance:	0.003	
For 68% conf. intrvl. use:	0.059	
Prop. consistent placings:	0.810	(Estimated number of incorrect classifications: 336)
Prop. beyond chance:	0.617	

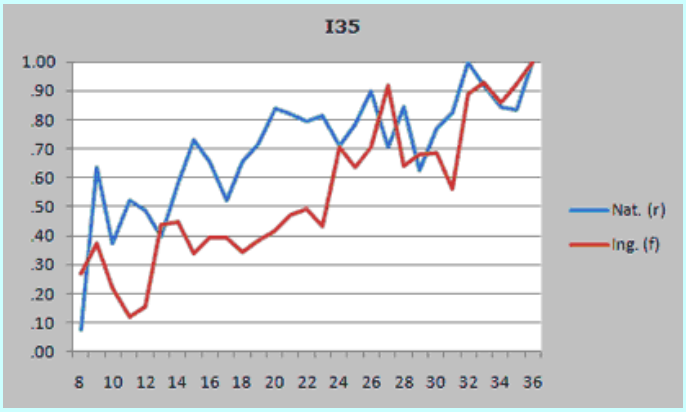
2.3.3 Group DIFferences

"**DIF**", differential item functioning, refers to procedures used to determine whether or not test items may have unexpectedly favored one test-taking group over another.



The plot above indicates the proportion of students with test scores ranging from 9 to 36 who got "I1" right. The plot shows that for all those with a test score of 9, 31% of the students in the "Ing." group got the item right, compared to 9% in the "Nat." group. However, over the range of scores plotted, there is not a consistent pattern: one group will be on top at a particular score level, but then the other group will be on top at another score level. This plot (and its associated statistics) is an example where no DIF was detected.

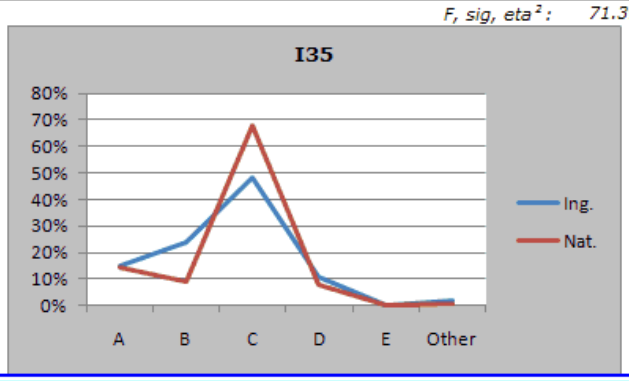
The plot below differs: it shows some evidence of DIF as the Nat. group rather consistently outperformed the Ing. group, especially at score levels from 14 to 23.



Lertap 5 uses the **Mantel-Haenszel** method for assessing DIF. A technical paper on this is available, as are related topics in "**Lelp**", Lertap help.

There are many times when test developers and users are not necessarily concerned with DIF, per se, but want to have plots which trace the performance of item responses in order to discern how groups may have differed in their "endorsements" to each of an item's options. For example, the table below, and its corresponding graph, indicate that many of the 844 students in the "Ing." group were distracted by option B: 24% endorsed B, compared to just 9% in the "Nat." group. This almost certainly accounts for the considerable group disparity seen at option C, which was the correct answer to I35.

I35	A	B	C	D	E	Other	n	mean	s.d.
Ing.	15%	24%	48%	11%	0%	2%	844	0.48	0.50
Nat.	14%	9%	68%	8%	0%	1%	876	0.68	0.47
						F, sig, eta²:	*71.39*	*0.00*	*0.04*



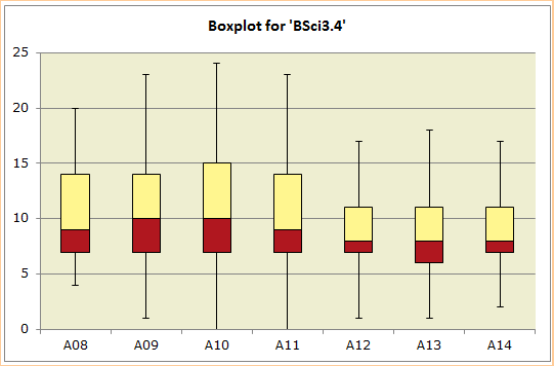
Tables and plots such as the one above are discussed in another Lelp topic.

2.3.4 Breakout scores

Assessing test score differences among groups

The table and plot for I35 seen in the box above exemplifies Lertap 5's capability for displaying group differences at the item level. A similar capability is available for expressing the extent to which two or more groups differ with regard to a score, such as a total test score.

For example, the boxplot to the right indicates how students in seven age levels differed on a rigorously-developed test of science achievement. It shows that the greatest spread of test scores was found in group A10 (ten-year olds), while older students had the lowest median test scores. It appears that the test may have been on the difficult side as the median test score for all groups never exceeded 10 (the maximum possible test score was 25).



Lertap 5's "**Breakout scores by groups**" options are in charge of making boxplots, other graphs, and a corresponding detailed analysis of variance.	

2.3.5 Cheating detection

Response similarity analysis, RSA

Lertap 5 has a **cheat-checker** referred to as "**RSA**". It's used to determine if the answers a pair of students gave on a multiple-choice test were exceptionally and unexpectedly similar.

When an RSA analysis is run, Lertap gets Excel to create a variety of tables with descriptive and inferential statistics.

Pairings	
Suspect:	3
Not suspect:	73,533
Total:	73,536
Inclusions	
Number of items:	59
Number of students:	384
Run control	
EEIC minimum:	8
H-H index minimum:	1.5
H-H sigma minimum:	5
Items excluded:	1
Minimum score setting:	32
Maximum score setting:	52

Lertap5 RSA cases list with EEIC min = 8, produced on: 11/04/2011.

ID	Data row	Responses	Score	EEIC	D	Index	Log	Sigma
210515	DataRow437	.2..4.444...1.x.4.2.33..4....4....23..44..3.3.4.3.	34	25	1	25.00	-53.96	17.64
210516	DataRow438	.2..4-444...1.x.4.2.33..4....4....23..44..3.3.4.3.	33					
		4.1..2..44						
		4.1..2..44						
212290	DataRow138	.....x.2.2.....2....2...2.1.2.....	50	9	6	1.50	-25.90	6.27
212297	DataRow143	..2.....44...2x.2.2.....2.4.42...2.1.2.....	44					
		.....4.2.						
		.....4.2.						
210123	DataRow158	.....434.....x.....44.....13..4...	47	10	3	3.33	-24.14	5.56
210126	DataRow160	.....444.....x.....1.....4.....13..4...	47					
		..2..342..						
		..2..342..						

Total number of cases displayed above: 3.

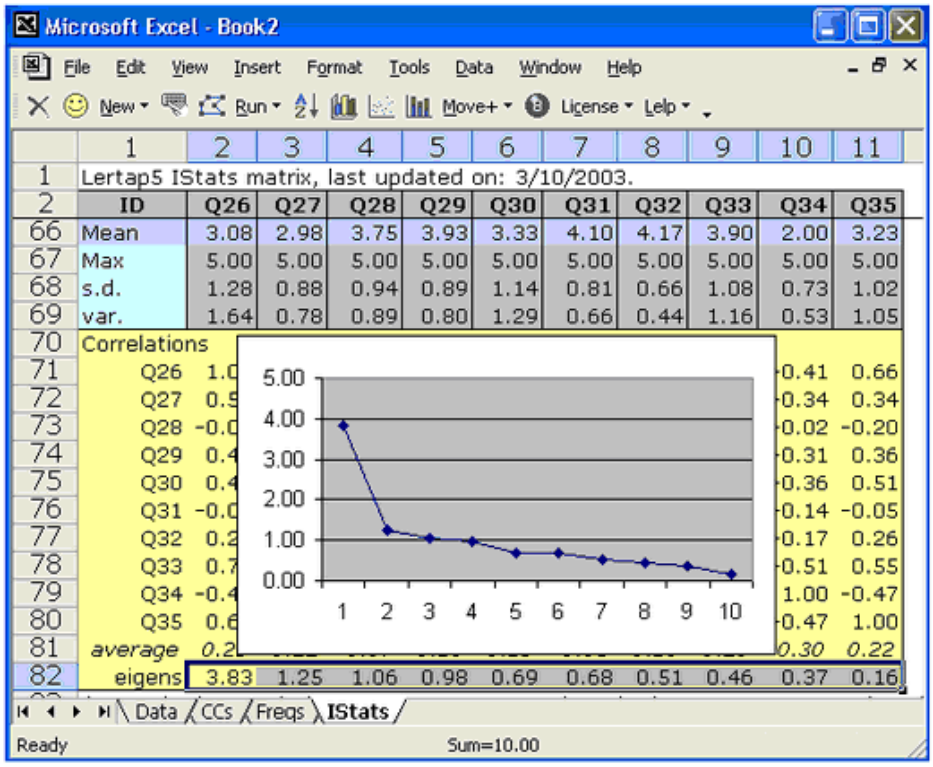
In this example, a real-life case, 384 students sat a 59-item test. RSA indicated that, out of a total of 73,538 pairs of students, 3 student pairs had response patterns which could be regarded as "suspicious". Seating charts from the exam venue indicated that these pairs were seated in close proximity, and thus would have had the chance to share answers. The first pair of students in the table above had only a single response difference over all items (the **D** column), and gave the same incorrect answer on 25 items (**EEIC**). RSA's "**Sigma**" statistic found this to be a very unlikely result by chance alone -- the two students were interviewed, confessed to cheating, and were suspended for one semester.

To read up on RSA, see this paper, this Lelp topic, and the dataset used in this example.

2.3.6 Reliability and consistency

Coefficient alpha, eigenvalues, scree tests, omega reliability

The plot below displays the eigenvalues from a ten-item rating scale (eigenvalues are also known as "**latent roots**").



The value of the first eigenvalue was almost 3.83, after which there is a sharp drop to the "**scree**" -- the second eigenvalue has a magnitude just a bit over 1.00 and from there on the eigenvalues decline in a regular, nearly linear manner. (Some researchers would accept this as an indicator of a "single-factor" scale.

There is an interesting, useful relationship between the magnitude of a test's or scale's first eigenvalue and coefficient alpha. This is discussed further in a technical paper.

A practical example based on using the first eigenvalue and its related eigenvector in the development of a rating scale is presented in this topic. Turn to **Lelp** (Lertap help) for more comments on eigenvalues, **principal components**, and how to get Lertap 5 to produce them.

Lertap 5 will form a complete inter-item correlation matrix, putting **SMCs** (squared-multiple correlation values) on the diagonal if wanted. **Tetrachoric correlations** may also be output. Comments on these options may be reviewed at this topic, and assistance with getting **SPSS** to pick up Lertap's correlation matrix is covered here.

An empirical comparison of alpha and coefficient **omega**, with more about eigenvalues and test reliability, is to be found in this research report.

2.3.7 IRT, item response theory

Historically, Lertap 5's main focus has related to the application of "**CTT**", classical test theory. But it certainly has support for IRT users, to be sure.

Those interested in the **Rasch** model will find support in Lertap5's RaschAnalyis1 routine. Other IRT models are supported via the EIRTanalysis1 Excel add-in. (Note: RaschAnalysis1 will work with both Windows and Macintosh computers, but EIRTanalysis1 may only be used with Windows.)

An option in the System worksheet to turn on "**experimental features**" will get Lertap 5 to add three columns to one of its standard reports, "**Stats1b**". These are the "bis.", "b(i)", and "a(i)" columns seen in the screen snapshot below:

Lertap5 brief item stats for "Prueba de MATEMATICA", created: 15/08/2002.

Res =	A	B	C	D	other	diff.	disc.	?	bis.	b(i)	a(i)
Item 1	5%	8%	69%	7%	10%	0.69	0.32		0.42	- 1.17	0.47
Item 2	31%	13%	21%	12%	22%	0.31	0.38		0.50	0.98	0.58
Item 3	20%	17%	41%	9%	13%	0.41	0.13	A	0.16	1.45	0.17
Item 4	8%	22%	21%	20%	28%	0.22	0.24	D	0.33	2.30	0.35
Item 5	5%	53%	16%	15%	10%	0.53	0.26		0.33	- 0.26	0.35
Item 6	50%	7%	8%	14%	21%	0.50	0.38		0.47	- 0.01	0.54

Data / CCs / Freqs / Scores / Stats1f / **Stats1b**

a(i) and **b(i)** are estimates of the **2PL** IRT parameters, while **bis.** is the biserial correlation coefficient. The derivation of these statistics is the focus of this paper.

Lertap 5 has extensive, well-developed support for users of ASC's Xcalibre IRT program, and it will also output data ready for input to Bilog-MG. (The data file made for use with Bilog-MG has a format also useful to quite a number of other programs, such as ConQuest and RUMM.)

A webpage which demonstrates the strong relationship between CTT and IRT statistics is available at this link, and a technical paper, somewhat critical of the **Rasch** model, may be found by following this hyperlink. Remember to take in this 2015 document when time permits, a document describing other historical IRT developments in Lertap 5.

2.3.8 Supply items

Scoring open-ended and constructed-response items

Lertap 5 is particularly adept at processing multiple-choice quizzes and tests, and rating scales. However, it may also be used in situations where the number of responses to an item or question is not fixed in advance. We have two examples to recommend.

One example comes from a graduate student at the University of Minnesota who developed and administered a test having a mixture of multiple-choice and short-answer questions. Her work is described in this document.

Another example derives from a major project to develop a new high-school student aptitude inventory for use on a national scale. A mixture of multiple-choice and constructed-response items were used. An interesting aspect of this project had to do with the development of two equivalent forms of the instrument, making it possible to examine **parallel-forms reliability**. This project is discussed in another technical paper; the paper as well as authentic data from this project are available as one of our sample datasets.

A third example comes from "FIMS", the First International Mathematics Study -- a mixture of supply/constructed-response and multiple-choice items were used. (This example has links to actual items.)

2.4 Videos

The video resources referenced in this topic stem from three sources. The first are by an instructor not at all associated or affiliated or (thus far) even known by those of us here at Lertap Central in Western Australia. He has put together a series of **YouTube** videos having to do with the installation and application of Lertap 5. Quite a nice job he's done, too -- he speaks clearly and his examples are excellent. Here are links to three of his productions; once you're in one of them there may be links to more Lertap-related videos:

YouTube video TD1: Installing Lertap5

YouTube video TD2: Using Lertap 5

YouTube video TD3: Processing a survey

In 2019 **EC Software** released a new authoring utility program called "HelpXplain". The following topic page links to sample "Xplain videos".

The remaining links on this page, immediately below, lead to some videos made at **Lertap Central in 2010** (Coffs Harbour). They require the Adobe Flash Player. Your

browser may inform you that it won't play one of these dynamic, totally-captivating videos until you activate Adobe Flash, something that's pretty easy to do.

What was formerly Video (1) has now been removed as it was too long, excessively boring, and obviously indicated a lack of video-making experience. Fortunately, the other videos are shorter and better and note: some of the videos have corresponding "stories/themes" -- they're straightforward webpages, without "movies". Also note that the references provided for each video are more current than the videos themselves, but the videos are still quite good demonstrations of how Lertap works.

The videos do not start at Lertap 101. That was Video(1)'s job. Now that Video(1) is gone, what we hope will serve as a good-enough basic introduction to how Lertap works is this PowerPoint show. The little HelpXplain animated slide show might be useful for Lertap beginners too, and then, how about this webpage?

If you try these videos, be patient when they start as it can sometimes take a few minutes for enough of the video to come through before it will start to play. And note, please: the URLs in these videos will often point to "lertap.curtin.edu.au" (a Curtin University internet domain). As of the year 2019, "lertap5.com" became the domain. A URL such as `lertap.curtin.edu.au/lelp.htm` would now be `lertap5.com/lelp.htm`.

---

(2) SMA Inggris item fixes (14 minutes)

The scene: this "story" involves adjusting the scoring of three multiple-choice items so that they have more than one correct answer. Multiple subtests are set up so that the resulting new test scores can be compared to the original ones. Also demonstrated in the story is the easiest way to exclude items from scoring.

The dataset used in the video is a carefully-developed 50-item English test from Indonesia.

The themes featured in the story may be seen here.

References: giving an item more than one correct answer involves the use of ***mws** control "cards". These are discussed in Chapter 3 of the manual, and also in Lelp (where you'd want to look at Example C10). Excluding an item may be accomplished in numerous ways; see Lelp. Scoring a test in multiple ways is accomplished by using multiple subtests; each subtest's ***col** control "card" will be the same.

The last part of the "B-Science" study, also referred to as the "Lertap tips and tricks demo", covers similar material, does not involve a video, and is more recent.

---

(3) Trigonometry exam demonstration (20 minutes)

This video involves a 40-item test with particularly good statistics (reliability of 0.92, and only two problematic items). Both types of quintile plots are extensively demonstrated and discussed. Well into the video, at the 14:28 mark, a discussion of the "**IStats**" report begins, with mention of such statistics as the "SMC", "tetrachoric correlations", "eigenvalues", and "principal components".

References: For more about IStats, see Chapter 10 of the manual, and, especially, Lelp. More about eigenvalues? Have got? Yes, try this read. Quintile plots are featured here.

---

(4) Affective scale demonstration (16 minutes)

The scene: this "story" is based on an affective instrument with 48 Likert items (*strongly disagree* to *strongly agree*). The instrument had very low reliability when it was originally scored by a hasty graduate student. Detecting how to correct the scoring involved the use of a Lertap principal-components analysis (had the student paid attention to the wording of the original items, the p-comps analysis would have been unnecessary). The reliability of the instrument jumped to 0.90 once the new scoring procedure was in place.

The themes featured in the story may be seen here.

References: reversing the scoring of Likert items involves the use of a ***pol** control "card", as discussed in Chapter 6 of the manual, and also in Lelp. Lertap's support for principal components is mentioned in Lelp (where reference is also made to factor analysis); a captivating, inspiring discussion of principal components, scree plots, and their relationship to reliability, as indexed by coefficient alpha, may be found in this paper.

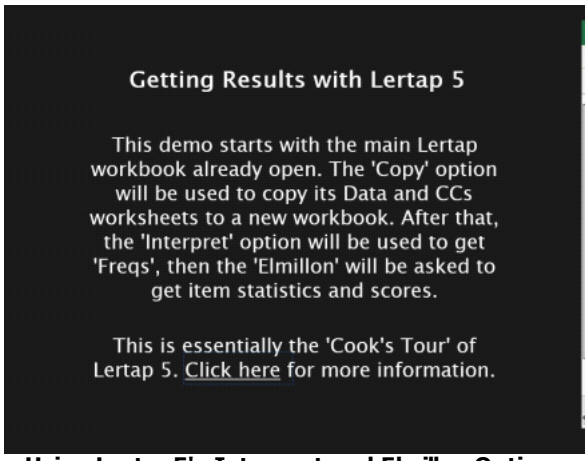
---

(5) Looking for group differences (29 minutes)

The scene: this group adventure tour uses a 50-item cognitive test from Indonesia, involving almost 17,000 high school students. The "Breakout score by groups" option is applied to see if there were differences in test performance by gender. Answering this question involves an analysis of variance, followed by a boxplot. This is followed, at the 11-minute mark, by use of "Ibreaks", the "Item responses by groups" option, used to see if there may have been response differences, by gender, on each of the 50 items. Response plots are created, more analysis of variance tables are drooled over, and then differential item functioning (DIF) methods are applied to index the extent of gender response differences, per item. (To see an example of how to create empirical DIF graphs, scroll in to just beyond the 24-minute mark.)

References: Lelp! Look to Lelp for the friendly web pages which underpin this tour. The "Breakout score by groups" option is here, and the material on "Ibreaks" is here. A truly top-flight reference, one of the hottest Lertap reference documents ever, can be had herewith.

2.4.1 HelpXplain



Using Lertap5's Interpret and Elmillon Options

2.5 For classes

Special notes for instructors and measurement classes

Students and instructors often find that the free "Mini" version of Lertap 5 will serve as an effective introduction to Lertap, giving students a chance to tackle the exercises suggested in the topics found below. There are also a number of other spots with possible exercises and pertinent discussions for classes; see, for example, the so-called "Erudite epistles" page, and the sample datasets website. The manual has a fairly extensive discussion of the various statistics used in classical test theory and how they're interpreted, particularly in Chapter 7.

Two "practical exhibits" showcase examples of Lertap5 in action, one for a cognitive test and another for an affective scale on depression. They are primarily introductory in nature but do delve a bit into more advanced topics. The cognitive test is called "Test13" and is available here. The depression inventory (or scale) is a very popular, true-to-life scale and may be accessed via this link.

A unique option, "To halve and hold", can be used to divide any dataset into two random halves, setting up the possibility of having classes look at parameter stability, or, perhaps, calibration and validation. A paper by Nelson (2008) reflects this idea in the context of **Rasch** scaling and the **Winsteps** program. Other related suggestions and challenges may be seen in this special exercises page.

The matter of linking Lertap 5 to other programs, such as **SPSS** and item-response theory (**IRT**) programs, is discussed in this topic. A theme popular in many classes and research endeavors is **factor analysis**. It might be suggested that students

select one of the sample datasets, such as "Mente 2010", and then apply the "To halve and hold" option mentioned in the previous paragraph to create two samples. In each sample an inter-item correlation matrix could be formed using Lertap 5's "IStats" option. The two matrices could then be picked up by SPSS's factor analysis routines and results compared. This type of exercise readily leads to topics such as the number of common factors and dimensionality; it can be particularly interesting when students compare their results. (Exercises of this sort can keep a class going for 60 to 90 minutes, giving the instructor a chance to slip out for a coffee break and a check of last night's lotto results.)

Instructors may request that the Mini version be "unlocked" to the full version for use by their students during a specified study period. Write to lertap5@gmail.com with your request.

The following resources have handouts and exercises for use with classes. While they're largely designed to be used with Lertap 5, a system such as SPSS can be applied to some of the exercises, as can Excel just by itself (without Lertap 5). Nelson's 2004 paper, "*Excel as an Aide in Teaching Measurement and Research Methods*" references some of these resources and exercises at greater length. Nate Thompson's little "CITAS" Excel item analysis spreadsheet may be another useful aid; other CTT and IRT systems are covered in a very useful Wikipedia webpage.

As an additional resource, this paper discusses SAS University and jMetrik as options to Lertap 5 -- these two programs will work on Macintosh computers as well as Windows and they're both free.

- The Chemistry Quiz & the "CEQ" -

We've got two little exercises made especially for use in classes. Each set comes with actual answer sheets which may be printed to simulate an authentic mini-research project -- students enter data in Excel and then get the results needed to answer the questions posed by the exercises.

The first example is from a chemistry quiz, the second involves the "CEQ", a computers-in-education survey given to a small sample of school principals. Each example can take up to 60 minutes for students to complete; having them work in pairs may make the data entry phase easier for them.

- The Lertap Quiz (November 2006 version) -

No doubt your grandparents have mentioned the world-renowned Lertap Quiz data set -- it's been with us since 1973. The original version of the dataset is described in the appendix to the manual. Here we give you the chance to download a modified version, one whose second column contains a school type code. This

modified version has been created in order to provide more challenges to beginning users; the presence of a school type code can lead to the use of, for example, the 'Breakout scores by groups' option on the Run menu (this dataset may be downloaded from this link: Excel xls file).

There is also a brief **codebook** which goes with the Excel workbook. It may downloaded by clicking here (Word document).

We've also created some tasks to help ward off brain addle. Try your hand, for example, at our Research Questions Set A.

Those tasks were too easy, you say? Fair enough, mate, avagowiththese: Research Questions Set B.

- The Ed 501 Data Set -

Here we've got a very practical exercise built on a real test from a university **educational psychology class**. Use this exercise to remind yourself (and students?) how a Lertap data set is formed when given just a worksheet of item responses, and a list of the correct answers to test items. (Includes a set of research questions to answer.)

- Cognitive Holding Power Questionnaire -

This sample stems from the work of **John Stevenson** of Griffith University (Australia). The CHPQ is an **affective instrument** with thirty items. Our sample includes an Excel data set. (Web page, will open in a new browser window.)

- Concurrent validity study -

Relates results from an authentic concurrent validity study undertaken by a masters student. Has numerous samples of Lertap output, and exemplifies how to apply some of Lertap's less-used options. Demonstrates use of an **external criterion** score. (Web page, will open in a new browser window.)

- Parallel-forms reliability study -

Discusses the development of an **academic aptitude test** for high school students, based on authentic data. Gets into some practical data processing problems; shows

how to use Excel's Data Filter, a powerful utility. Talks about speeded tests, and practice effects (based on the two equivalent forms developed in the study). Shows how short-answer **constructed-response items** may be used in Lertap 5. This is a typical example of classical test development methods, with many snapshots of Lertap results. (Web page, will open in a new browser window.)

- Importing data -

A fairly technical discussion of downloading sample datasets from the Internet in typical "text" format, and making them ready for Excel and Lertap. Shows use of "The Spreader" . Discusses general matters related to importing data, using samples from Iowa University. (Web page, will open in a new browser window.)"

2.6 Downloads

Get down and loaded with great stuff by using the links below.

- Software -

Note: Lertap 5 is an Excel "app". It requires a version of Microsoft Excel in order to run.

- Copies of Lertap 5 for use with **Windows** and **Macintosh** computers may be downloaded from this website.
- Copies of Lertap 5 for use by instructors with test and measurement classes may also be requested by sending a message to support@lertap.com.
- Related information about getting Lertap 5 is available at this topic.

- Documentation -

- The original manual, and individual chapters from it, may be downloaded from the links found at this topic.
- Resources for downloading in test and measurement classes are referenced here.

- Numerous special-interest technical topics with down-loadable datasets are covered in the "erudite epistles" topic.

- Samples -

- A set of sample Excel workbooks set up for use by Lertap 5 and other data analysis programs may be downloaded from this website.
- Practical examples with smaller sample sizes and companion work sheets are available here.

2.7 References

References cited in Lertap 5 documents

Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory.* Monterey, California: Brooks/Cole.

Angoff, W.H. (1993). Perspectives on differential item functioning methodology. In P.W. Holland and H. Wainer (Eds.), *Differential Item Functioning* (pp. 3-23). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Bandalos, D.L. (2018). *Measurement theory and applications.* New York, NY: The Guilford Press.

Berk, R. A. (1978). Empirical evaluation of formulae for correction of item-total point-biserial correlations. *Educational and Psychological Measurement, 38,* 647-652.

Berk, R.A. (1980). A consumer's guide to criterion-referenced test reliability. *Journal of Educational Measurement, 17,* 323-350.

Berk, R.A. (1984). Selecting the index of reliability. In R.A. Berk (Ed.) *A guide to criterion-referenced test construction.* Baltimore, Maryland: The Johns Hopkins Press.

Berk, R.A. (2000). *Ask Mister Assessment Person*. In Teachers: Supply and demand in an age of rising standards. Amherst, MA: National Evaluation Systems, Inc. (Note: in May,

2013, the papers in this series could be found as individual PDF files by searching the internet.)

Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch Model* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Bond, T.G. & Fox, C.M. (2015). *Applying the Rasch Model* (3rd ed.). New York, NY: Routledge.

Brennan, R.L. (1972). A generalized upper-lower discrimination index. *Educational and Psychological Measurement, 32*, 289-303.

Brennan, R.L. (1984). Estimating the dependability of the scores. In R.A. Berk (Ed.) *A guide to criterion-referenced test construction.* Baltimore, Maryland: The Johns Hopkins Press.

Brennan, R.L. & **Kane**, M.T. (1977). An index of dependability for mastery tests. *Journal of Educational Measurement,* 14, 277-289.

Brown, M.B. (1977). Algorithm AS 116: the tetrachoric correlation and its standard error. *Applied Statistics,* 26, 343-351.

Camilli, G. & **Shepard**, L.A. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage Publications.

Carr, N.T. (2011). *Designing and analyzing language tests.* Oxford: Oxford University Press.

Case, S.M. & **Swanson**, D.B. (1998). *Constructing written test questions for the basic and clinical sciences.* Philadelphia: National Board of Medical Examiners.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research,* 1, 245-276.

Cizek, G.J. (2001). *An overview of issues concerning cheating on large-scale tests.* Paper presented at the annual meeting of NCME, the National Council on Measurement and Evaluation, April 2001, Seattle, Washington.

Clauser, B.E., & **Mazor**, K.M. (1998). An NCME instructional module on using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice,* 17(1), 31-44.

Cody, R. & **Smith,** J.K. (2014). *Test scoring and analysis using SAS.* Cary, NC: SAS Institute Inc.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20, 37-46.

Crocker, L.M. & **Algina**, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart, and Winston.

Dai, S (2021). Handling missing responses in psychometrics: methods and software. Psych 2021, 3, 673-693 (https://doi.org/10.3390/psych3040043).

Dawber, T. (2004). *Robustness of Lord's formulas for item difficulty and discrimination conversions between classical and item response theory models.* Edmonton, Alberta: unpublished doctoral dissertation, University of Alberta (also see following reference).

Dawber, T., **Rogers**, W.T., & **Carbonaro**, M. (2004). *Robustness of Lord's formulas for item difficulty and discrimination conversions between classical and item response theory models.* Paper presented at the annual meeting of AERA, the American Educational Research Association, April 12, 2004, San Diego, California.

de la Harpe, B.I. (1998). *Design, implementation, and evaluation of an in-context learning support program for first year education students and its impact on educational outcomes.* Perth, Western Australia: unpublished doctoral dissertation, Curtin University of Technology.

DeMars, Christine (2010). *Item response theory.* New York: Oxford University Press, Inc.

Dimitrov, D.M. (2003). Reliability and true-score measures of binary items as a function of their Rasch difficulty parameter. *Journal of Applied Measurement,* 4(3), 222-233.

Dorans, N.J. & **Kulick**, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential performance on the Scholastic Aptitude Test. *Journal of Educational Measurement,23,* 355-368.

Dorans, N.J. & **Holland**, P.W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland and H. Wainer (Eds.), *Differential Item Functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Dorans, N.J. & **Kulick**, E. (2006). Differential item functioning on the Mini-Mental State Examination: an application of the Mantel-Haenszel and standardization procedures. *Medical Care, 44(11),* S107-S114.

Dunn, T.J., **Baguley**, T., & **Brunsden**, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399-412. doi:10.1111/bjop.12046

Eason, S. (1991). Why generalizability theory yields better results than classical test theory: a primer with concrete examples. In B. Thompson (Ed.), *Advances in Educational*

Research: Substantive findings, methodological developments (Vol. 1, pp. 83-98). Greenwich, CT: JAI.

Ebel, R.L. & **Frisbie**, D.A. (1986). *Essentials of Educational Measurement* (4th ed.). Sydney: Prentice-Hall of Australia.

Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement,* 58 (3), 357-381.

Feldt, L.S. (1984). Some relationships between the binomial error model and classical test theory. *Educational and Psychological Measurement,* 44, 883-891.

Frederiksen, N., **Mislevy**, R.J., & **Bejar**, I.I. (Eds.) (1993). *Test theory for a new generation of tests.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Garrett, H.E. (1952). *Testing for teachers.* New York: American Book Company.

Geldhof, G.J., **Preacher**, K.J, & **Zyphur**, M.J. (2013). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods.* doi:10.1037/a0032138.

Gil Escudero, G., **Suárez Falcón**, J.C., y **Martinez Arias**, R. (1999): Aplicación de un procedimiento iterativo para la selección de modelos de la Teoria de la Respuesta al Item en una prueba de rendimiento lector. *Revista de Educación,* 319, 253-272.

Glass, G.V & **Stanley**, J.C. (1970). *Statistical methods in education and psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Glass, G.V & **Stanley**, J.C. (1974). *Metodos estadisticos aplicados a las ciencias sociales.* London: Prentice-Hall Internacional.

Green, J. (1999). *Excel 2000 VBA programmer's reference.* Birmingham, England: Wrox Press.

Gronlund, N.E. (1985). *Measurement and evaluation in teaching* (5th ed.). New York: Collier Macmillan Publishers.

Gulliksen, H. (1950). *Theory of mental test scores.* New York: John Wiley & Sons.

Haladyna, T.M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Haladyna, T.M. & **Rodriguez**, M.C.(2013). *Developing and validating test items.* New York: Routledge.

Hambleton, R.K. & **Jones**, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice,* 12(3), 38-47.

Hambleton, R.K. & **Swaminathan**, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer.

Hambleton, R.K., **Swaminathan**, H., & **Rogers**, H.J. (1991). *Fundamentals of item response theory.* Newbury Park, California: Sage Publications.

Harpp, D.N. & **Hogan**, J.J. (1993). Crime in the classroom- detection and prevention of cheating on multiple-choice exams. *Journal of Chemical Education,* 70(4), 306-311.

Harpp, D.N., **Hogan**, J.J., & **Jennings**, J.S. (1996). Crime in the classroom- Part II, an update. *Journal of Chemical Education,* 73(4), 349-351.

Hattie, J., **Jaeger**, R.M., & **Bond**, L. (1999). Persistent methodological questions in educational testing. *Review of Research in Education,* 24, 393-446.

Hays, W.L. (1973). *Statistics for the social sciences.* London: Holt, Rinehart and Winston.

Hills, J.R. (1976). *Measurement and evaluation in the classroom.* Columbus, Ohio: Charles E. Merrill.

Hopkins, K.D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Boston: Allyn & Bacon.

Hopkins, K.D. & **Glass**, G.V (1978). *Basic statistics for the behavioral sciences.* Englewood Cliffs, NJ: Prentice-Hall.

Hopkins, K.D., **Stanley**, J.C., & **Hopkins**, B.R. (1990). *Educational and psychological measurement and evaluation* (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hoyt, C.J. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6,* 153-160.

Kaplan, R.M. & **Sacuzzo**, D.P. (1993). *Psychological testing: principles, applications, and issues.* Pacific Grove, California: Brooks/Cole.

Kelly, T.L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology,* 30, 17-24.

Kerlinger, F.N. (1973). *Foundations of behavioral research* (2nd ed.). London: Holt, Rinehart, and Winston.

Kline, P. (2015). *A handbook of test construction*. New York: Routledge.

Kolen, M.J. & **Brennan**, R.L. (1995). *Test equating: methods and practices.* New York: Springer-Verlag.

Lawson, S. (1991). One parameter latent trait measurement: Do the results justify the effort? In B. Thompson (Ed.), *Advances in Educational Research: Substantive findings, methodological developments* (Vol. 1, pp. 159-168). Greenwich, CT: JAI.

Lindeman, R.H. & **Merenda**, P.F. (1979). *Educational measurement* (2nd ed.). London: Scott, Foresman and Company.

Linn, R.L. & **Gronlund**, N.E. (1995). *Measurement and assessment in teaching* (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Hillside, NJ: Lawrence Erlbaum Associates.

Lord, F.M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement,* 21(3), 239-243.

Lord, F.M. & **Novick**, M.R. (1968). *Statistical theories of mental test scores.* Reading, Massachusetts: Addison-Wesley.

MacDonald, P. & **Paunonen**, S.V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement,* 62(6), 921-943.

Magis, D., **Beland**, S., **Tuerlinckx**, F. and **De Boeck**, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42*, 847-862. doi: 10.3758/BRM.42.3.847

Magnusson, D. (1967). *Test theory.* London: Addison-Wesley.

McDonald, R.P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mehrens, W.A. & **Lehmann**, I.J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). London: Holt, Rinehart and Winston.

Meyer, J. Patrick (2010). *Reliability.* New York: Oxford University Press, Inc.

Meyer, J. Patrick (2014). *Applied Measurement with jMetrik.* New York: Routledge (Taylor & Francis Group).

Michaelides, M.P. (2008). An illustration of a Mantel-Haenszel procedure to flag misbehaving common items in test equating. *Practical Assessment, Research & Evaluation,* 13 (7). Available online: http://pareonline.net/getvn.asp?v=13&n=7

Nandakumar, R. (1994). Assessing dimensionality of a set of item responses—Comparison of different approaches. *Journal of Educational Measurement,* 31, 17-35.

Nelson, L.R. (1974). *Guide to LERTAP use and interpretation.* Dunedin, New Zealand: Department of Education, University of Otago.

Nelson, L.R. (1981). *PLATISLA, an introduction to applied social science statistical methods.* Dunedin, New Zealand: Department of Education, University of Otago. (Click here to link to sample data from Platisla.)

Nelson, L.R. (1984). Using microcomputers to assess achievement and instruction. *Educational Measurement: Issues and Practice,* 3(2), 22-26. (Click here for a PDF copy.)

Nelson, L.R. (2000). *Item analysis for tests and surveys using Lertap 5.* Perth, Western Australia: Curtin University of Technology (www.lertap5.com). (Click here to select chapters.)

Nelson, L.R. (2004). Excel as an aide in teaching measurement and research methods. *Thai Journal of Educational Research and Measurement (ISSN 1685-6740)*: 2(1), 43-55. (Click here to link to a copy of this paper; the final journal article is here.)

Nelson, L.R. (2005). Some observations on the scree test, and on coefficient alpha. *Thai Journal of Educational Research and Measurement (ISSN 1685-6740):* 3(1), 1-17. (Click here to link to a copy of this paper; the final journal article is here.)

Nelson, L.R. (2006). Using selected indices to monitor cheating on multiple-choice exams. *Thai Journal of Educational Research and Measurement (ISSN 1685-6740):* 4(1), 1-18. (Click here to link to a copy of this paper; the final journal article is here.)

Nelson, L.R. (2007). Some issues related to the use of cut scores. *Thai Journal of Educational Research and Measurement (ISSN 1685-6740):* 5(1), 1-16. (Click here to link to a copy of this paper; click here for the final journal article.)

Nelson, L.R. (2008). Rasching an achievement test. *Thai Journal of Educational Research and Measurement (ISSN 1685-6740):* 6(1), 1-22. (Click here to link to a copy of this paper; the final journal article is here.)

Nelson, L.R. (2016). Coefficients alpha and omega, an empirical comparison. DOI: 10.13140/RG.2.1.1957.5929 (Click here to link to a copy of this paper.)

Nelson, L.R. (2017). Item analysis software for classes. DOI: 10.13140/RG.2.2.32532.71049 (Click here to link to a copy of this working paper.)

Nelson, L.R. (2020). Assessing the invariance of cognitive item statistics. DOI: 10.13140/RG.2.2.36829.97769 (Click here to link to a copy of this paper.)

Nelson, L.R. (2021). Computing alpha and omega reliability estimates. DOI: 10.13140/RG.2.2.15727.97440 (Click here to link to a copy of the working paper.)

Online Press, Inc. (1997). *Quick course in Microsoft Excel 97.* Redmond, Washington: Microsoft Press.

Oosterhof, A.C. (1990). *Classroom applications of educational measurement.* Columbus, Ohio: Merrill.

Pedhazur, E.J. & **Schmelkin**, L.P. (1991). *Measurement, design, and analysis: an integrated approach.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Peng, C-Y.J. & **Subkoviac**, M.J. (1980). A note on Huynh's normal approximation procedure for estimating criterion-referenced reliability. *Journal of Educational Measurement, 17,* 359-368.

Pintrich, P.R., **Smith**, D.A.F., **Garcia**, T. & **McKeachie**, W.J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, Michigan: the University of Michigan.

Popham, W.J. (1978). *Criterion-referenced measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Qualls-Payne, A.L. (1992). A comparison of score level estimates of the standard error of measurement. *Journal of Educational Measurement,* 29(3), 213-225.

Revelle, W. (2016). Using R and the psych package to find ω. (February 2017: seen at http://personality-project.org/r/psych/HowTo/R_for_omega.pdf)

Revelle, W. & **Condon,** D.M. (2019). Reliability from Alpha to Omega: A tutorial. *Psychological Assessment,* 31(12):1395-1411.

Revelle, W. & **Zinbarg**, R.E. (2009). Coefficients alpha, beta, omega, and the glb: comments on Sijtsma. *Psychometrika,* 74(1):145-154. doi:10.1007/s11336-008-9102-z. (September 2020: seen at http://personality-project.org/revelle/publications/rz09.pdf)

Roussos, L.A., **Schnipke**, D.L., & **Pashley**, P.J. (1999). A generalized formula for the Mantel-Haenszel differential item functioning parameter. *Journal of Educational and Behaviorial Statistics,* 24(3), 293-322.

Sanders, D.H. (1981). *Computers in society.* New York: McGraw-Hill.

Stage, C. (1998). A comparison between item analysis based on item response theory and classical test theory: a study of the SweSAT Subtest READ. *Educational Measurement* No 30. Umeå, Sweden: University of Umeå, Department of Educational Measurement. (Possibly available at www.umu.se/edmeas/publikationer/index_eng.html.)

Stage, C. (2003). Classical test theory or item response theory: the Swedish experience. *Educational Measurement* No 42. Umeå, Sweden: University of Umeå, Department of Educational Measurement. (Possibly available at www.umu.se/edmeas/publikationer/index_eng.html; found at the following address January 2008: www.umu.se/edmeas/publikationer/pdf/em%20no%2042.pdf.)

Stevenson, J. (1998). Performance of the Cognitive Holding Power Questionnaire in schools. *Learning and Instruction,* 8(5), 393-410.

Stevenson, J.C. & **Evans**, G.T. (1994). Conceptualization and measurement of cognitive holding power. *Journal of Educational Measurement,* 31(2), 161-181.

Subkoviak, M.J. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement,* 13, 265-276.

Subkoviak, M.J. (1984). Estimating the reliability of mastery-nonmastery classifications. In R.A. Berk (Ed.) *A guide to criterion-referenced test construction.* Baltimore, Maryland: The Johns Hopkins Press.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications.* Washington, DC: The American Psychological Association.

Thompson, B. (2006). *Foundations of behavioral statistics: an insight-based approach.* New York: The Guilford Press.

Thorndike, R.L. (1982). Educational measurement: Theory and practice. In D. Spearitt (Ed.), *The improvement of measurement in education and psychology: Contributions of latent trait theory* (pp. 3-13). Princeton, NJ: ERIC Clearinghouse of Tests, Measurements, and Evaluations. (ERIC Document Reproduction Service No. ED 222 545.)

Tukey, J.W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Wainer, H. (1989). The future of item analysis. *Journal of Educational Measurement,* 26, 191-208.

Wainer, H. (2006). A psychometric cicada: Educational Measurement returns. *Educational Researcher,* 36(8), 485-486., DOI: 10.3102/0013189X07311288

Wesolowsky, G.O. (2000). Detecting excessive similarity in answers on multiple choice exams. *Journal of Applied Statistics,* 27(7), 909-921.

Wiersma, W. & **Jurs**, S.G. (1990). *Educational measurement and testing* (2nd ed.). Boston: Allyn & Bacon.

Wright, B.D. & Stone, M.H. (1979). *Best test design*. Chicago: Mesa Press.

Wu, M. et al. (2016). *Educational measurement for applied researchers*. Singapore: Springer Nature Singapore Ltd., DOI: 10.1007/978-981-10-3302-5_2

Wyse, A.E. & Babcock, B. (2016). Does maximizing information at the cut score always maximize classification accuracy and consistency? *Journal of Educational Measurement, 53(1),* 23-44.

Zieky, M. (2003). *A DIF Primer*. Princeton, NJ: Educational Testing Service.

Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: lagging rules, minimum sample size requirements, and criterion refinement. (Research Report. No. RR-12-08). Princeton, NJ: Educational Testing Service. Located here 8 January 2019.

3 Reviews and history

A few professionals reviewed Lertap 5 shortly after it first appeared, and, more recently, two leading textbooks have made reference to it.

It's possible to read most of what the reviewers have said to date by looking at a summary of their comments (links to a Microsoft Word document).

One of the reviewers was Professor **Nathan Carr**. In a fairly recent book (Carr, 2011, p.209), Carr referred to Lertap 5 as his "*... personal favorite for reliability and item analysis ...*".

The **Handbook on Measurement, Assessment, and Evaluation in Higher Education** (Secolsky & Denison, 2012, Table 20.3) referred to Lertap5 as "*... a sophisticated commercial item analysis program that works with Microsoft Excel .... is capable of calculating standard item analysis values as well as differential item functioning (DIF), distractor analysis, and response similarity analysis .... (it) is very flexible and robust and capable of handling up to 10,000 items and over one million students.*"

An excellent text and technical reference book by **Haladyna & Rodriguez** (2013, p. 387), stated that Lertap 5 is "*... a very versatile item analysis program with many options, including DIF ...* ".

Other references to Lertap may, of course, be found on the internet by using a search engine (such as Google).

3.1 History

The initial version of Lertap was developed for the **Venezuelan Ministry of Education** in the years 1971 through 1972. The Ministry was then embarking on a national assessment program, with emphasis on mathematics and language achievement. The *Kuhlmann-Anderson* aptitude, or "IQ", test was also used on a national scale by the Ministry, and a general-purpose item analysis program was required, one which could handle conventional achievement tests, and the Kuhlmann-Anderson.

At the time, the development of the Ministry's assessment centre was under the direction of *Rogelio Blanco*, with *Richard Wolfe*, of OISE (Ontario), overseeing the technical services part of the operation. Richard created a general front-end to set up data sets for subsequent analyses, using the PL/I programming language. I contributed what amounted to the first version of Lertap, programmed in FORTRAN II. It picked up data sets pre-processed by the PL/I program, and output classical item statistics. This initial Lertap, locally called "***DIEitem***", could not only handle the idiosyncrasies of the Kuhlman - Anderson test, but could also entertain multiple tests within the same data set. Thus one could submit a data set with results from the mathematics test, the Spanish-language test, and the Kulhmann-Anderson test, all mixed together.

Work on DIEitem was supported by the *Ford Foundation*, and by the *la Organización de los Estados Americanos, OEA*.

The first English-language version

In 1973 work on the second version began at the **University of Colorado**, home of **LER**, the Laboratory of Educational Research. The PL/I front end was replaced by another, written in FORTRAN, which featured the use of a set of free-form control cards to describe a job. These control cards included ones called *TST, *FMT, *SUB, and so on (I mention them as a reminder to those who spent much time with this version). Free-form control cards were not widely used in those days, and, in this regard, **Lertap 2** could be considered as being slightly ahead of its time.

Lertap 2 also introduced support for processing affective tests. The late *Bob Conry* of the University of British Columbia provided strong support for the "aff" subtest capability, while *Ken Hopkins* and *Gene Glass*, at LER in Boulder, did all possible to encourage the development of the overall package.

The work started at LER was continued when I assumed my first position at the **University of Otago**, in Dunedin, New Zealand, late in 1973. By the end of 1974 Lertap 2 was stable, and in use in a variety of centres in Canada and the United States. Lertap 2 was blessed, if I can use that term, with a complete user guide. I draw attention to this fact as the same could not be said of the IBM PC version which later emerged when the first microcomputers appeared.

I am grateful for the Lertap 2 support provided by several people at Otago, especially *Dan McKerracher*, Department of Education, and *Brian Cox*, Computing Centre. The user guide which emerged from Otago was a fun document to write, and seemed to be quite well received. Brian Cox saw that it had a fetching cover, featuring his favourite Burroughs computer, and some of his staff.

Use of Lertap grew steadily in the 70s. During these years I held posts at Boston University, and la Corporación Venezolana de Guayana.

An apple a day

> Late 1980 saw me back in Otago, experimenting with the new **Apple II** microcomputer. A year later I began the development of **Lertap 3** in earnest, using a CP/M card on an Apple, and then on an Osborne 1 system. I found the BASIC 80 language to be capable of speedy performance, and used it to produce a suite of interlinked modules which would load and unload themselves in just 56K of core memory.

In early 1983 a working version of Lertap 3 was ready, and it was accompanied by a new user guide every bit as complete as that created for Lertap 2. *Barbara Calvert* keenly supported the development of this version, putting some of the resources of the Department of Education behind the effort. A few hundred copies of the user guide were printed, and made ready for distribution.

Enter IBM's decision to produce a microcomputer of its own. Enter another LER alumnus, *Evelyn Brzezinski* of the **Northwest Regional Educational Laboratory**, Portland, and *Larry Erikson* of **National Computer Systems**, Minneapolis. By late 1983 Lertap 3 had been altered so as to operate within IBM micros, and National Computer Systems had purchased it. NCS repackaged the system as two stand-alone programs, MicroTest1, and MicroSurvey1.

I was free to continue the development of Lertap 3, and did so, at a reduced pace. By the late 80s this version was in use at many sites, with the NCS versions finding a home in many (many) more. It was unfortunate that a user guide for this version was never thoroughly developed. The guide printed at Otago covered the pre-IBM version, but the operation of this version differed much, and IBM users found it to be of limited use.

In 1987 I began to circulate a series of brief user help sheets from Curtin University. These were later assembled as a small book, Lertap 3 General Notes, printed by Curtin University's printing services.

Lertap 3 remained a potent data analysis system for many years. The scope of analyses it supported included those related to cognitive and affective tests, general surveys, and classroom gradebooks. Its data preparation facilities included a module

for complete date entry verification. And, it could handle results from the Kuhlmann-Anderson tests.

In 1992 *Piet Abik* translated Lertap 3 to the Indonesian language, and it was later purchased by Indonesia's Ministry of Education and Culture for country-wide use in secondary schools, with the support of *Bambang Irianto*.

Lertap 2, Lertap 3, and the NCS derivatives came to be used throughout the world. I am aware of some of the corners they've reached by references in publications, and by new would-be users emailing me to ask about the system's availability.

When Microsoft released the Windows 3 operating system, in 1992 (in Australia), it was soon clear that Lertap had to move to Windows. People started to write to ask if the Windows version would be ready soon. I worked at it, off and on, until 1997.

A dud, dudes

I refer to this work as **Lertap 4**. It was never finished, It came to have a nifty facility for processing survey results, but not much more. My error was in believing that I could, on my own, rebundle all of Lertap 3's power in a Windows package. In order to appreciate the scope of this desire it would be necessary to understand both the extent of Lertap 3's capabilities, and the nature of programming under Windows. Suffice it to say that the idea of building a stand-alone Windows version of Lertap 3 is one I had to give away.

Por fin, me pegó la luz: Excel

So, you might ask, what's this new version of Lertap, then? **Lertap 5**? Isn't it running within Windows? Yes, certainly. It will also run on a Macintosh system, or, for that matter, any platform which supports a recent version of Excel.

What is different about this new version is that it is built on Microsoft's spreadsheet flagship, Excel. Using Excel has freed me from having to develop user interfaces related to data entry and maintenance, and, to my pleasant surprise, writing output to Excel worksheets is entirely straightforward compared to writing output to generic Windows forms. There are numerous spots where data in a Lertap 5 worksheet are simply passed to Excel; this has freed me from having to re-structure Lertap 2 & 3 modules, and it has undoubtedly resulted in faster program execution.

I have, furthermore, gone back to Lertap 2's method of job definition. Instead of having users answer a multitude of dialog boxes, they define their jobs by using control "cards", ones which are nearly identical to those first seen in Lertap 2 almost 30 years ago. A retrograde step? I think not. The parsimony of using a control "card" job definition language is remarkable. I know that users can master this language; it's not that extensive, and it has a track record of success.

The result is a system which asks almost no questions of users. This is a vast change from both Lertap 3 and the design seen in Lertap 4. I did not intentionally set out to produce such a system; I didn't know it might be feasible. Yet here it is. Users enter their results in the Data worksheet, type up their control "cards" in a worksheet called "CCs" (for Control Cards), and click on the Run button. Not once are they asked a job definition question of any sort.

Lertap 5 credits

I gratefully acknowledge the support of **Curtin University** of Technology's Faculty of Education, and its Division of Humanities, which made it possible for me to set aside several months of development time in the year 2000 without having to be concerned with classes and committee work. A special thanks to *Graham Dellar*.

Nanta Politawanont of **Burapha University**, Thailand, and *Suchada Kornpetpanee*, also of Burapha, provided a home away from home in 2000, letting me use several of their Thai data sets to debug initial sections of Lertap 5 code.

Todd Rogers of the **University of Alberta**, along with his doctoral students, especially *Keith Boughton* and *Tess Dawber*, have provided invaluable guidance, putting Lertap 5 through the ringer on numerous occasions, and at times pointing the way to code modifications.

Nurhadi Amiyanto of the **Government of Central Java**, Indonesia, sponsored a Lertap workshop series in 2002 which saw the software tested on a few Jateng data sets; at the time, with over twenty thousand students, this was Lertap's most substantial data processing challenge.

Carlos Gonzalez, of la **Universidad Central de Venezuela**, has been behind Lertap 5 all the way, often sending sample data sets and testing new features.

After some initial reluctance, *David Weiss* of **Assessment Systems Corporation**, creator of Iteman, XCALIBRE and a vast quantity of other psychometric resources, saw the light, and gave Lertap 5 a prominent role in his shop 2001.

Professor Ken Hopkins

Finally, a dedication. It is an honour to dedicate this version to Professor **Kenneth D. Hopkins** of the University of Colorado. He has been teacher, mentor, and friend since 1971. His books and publications dealing with classical item analysis have provided the bases for Lertap's development, and his frequent feedback over the years has done much to bring the system to its present state. Many will be the number who join me in wishing him a long and pleasant retirement. Australians would say *"Onya, Ken, you done good"*. Indeed. LERTAP is, I hope, at least a small credit to his teaching and writing.

Larry Nelson
School of Education
Curtin University
Created: May 2005

3.2 Changes in 2014

This topic is particularly relevant to experienced Lertap 5 users; it documents a variety of important changes implemented on the 1st of July, 2014.

As mentioned towards the end of the previous topic, Curtin University entered into a Lertap 5 marketing contract with Assessment Systems Corporation (ASC) in 2001. At that time ASC became the exclusive marketing agent for Lertap 5. It was the beginning of a partnership that lasted over a decade.

Up until 2008 or so, ASC had been what might be called a rather traditional software house, distributing their own systems, such as Iteman, Xcalibre, and FastTEST, and other systems too, such as Lertap 5 and software from SSI (including Bilog-MG). They also distributed technical books related to measurement and assessment, making them available for on-line purchase from their website.

During its tenure on the ASC website, Lertap 5 was the top income-generating product for non-ASC systems, and, until Iteman was upgraded to Iteman 4 in late 2009, ASC would recommend Lertap 5 over Iteman, regarding it as the most-capable classical test analysis system in its stable. (The new edition of Iteman incorporated a great variety of features previously found only in Lertap 5, bringing the two products, Iteman and Lertap, up to something of a level playing field. See this document for more comments.)

In the years 2010 and 2011 ASC began to increase its focus. New alliances and partnerships were made, and by late 2012 (or so) ASC was actively engaged in expanding into new markets, particularly those having to do with licensing and certification.

ASC got rid of their bookstore, and started to place restrictions on software sales. Any customer regarded as a potential competitor to ASC's commercial endeavors was to be prohibited from purchasing many of the systems on offer at the ASC website, including Lertap 5 and Xcalibre 4.

This effectively violated the original Curtin-ASC contract, and Curtin made a decision to terminate the contract at the end of June 2014. Curtin University subsequently agreed to assign its share of the intellectual property rights in Lertap 5 to me if I agreed to remove Lertap from Curtin's web servers. This I have done; I now own all of the rights to Lertap 5.

The internet domain www.lertap5.com was redirected to www.lertap5.com/lertap/ on the 1st of July 2014. (But www.lertap5.com will still work.)

I have created a new website, the one you're on now, to replace the old "Lertap resources" site at Curtin. Another new website, the "Lertap 5 store", has come on line too. Three of the other Lertap websites remain pretty much as they were as far as content goes. These are "Lelp (Lertap 5 help)", the very popular "sample datasets" website, and the "tips and tricks" website (also known as "B_Science").

But it's *not* back to business as usual. Users who have bookmarked favorite webpages in the past will find they don't work. A click on one of them, any of them, will bring users here, to this website. It will be necessary to get into the new websites, look for the content wanted, and bookmark them again. (*Sorry, but it had to be.*)

Users with versions earlier than 5.10.4 will find that the handy "little-h" online context-sensitive help links in Lertap do not work. All of them will route to the same place: this website. (An example of the little-h link is circled in red in the screen snapshot below.)

Lertap5 brief item stats for "Comfort with using LERTAP2", created: 9/07/2014.

Res =	1	2	3	4	5	other	pol.	mean	s.d.	cor.	h
Q26	13%	22%	25%	23%	17%		+	3.08	1.28	0.76	
Q27	5%	23%	37%	35%			-	2.98	0.88	0.55	
Q28	22%	45%	17%	13%		3%	-	3.75	0.94	- 0.14	
Q29	32%	35%	25%	5%		3%	-	3.93	0.89	0.44	
Q30	15%	33%	28%	13%	8%	2%	-	3.33	1.14	0.49	

We offer a $24 update option on the store for users who want to bring their Lertap up to the current version (this option will be useful *only* to users with Excel 2010 or Excel 2013). One of the main advantages to updating is that Lertap's many links to online help will lead to the correct pages.

For your information, a link to the "updates summary" page is here. And here's a link to the "24 smack-Os option" at the store.

Larry Nelson
Fremantle Village, Western Australia
July 2014

4 Help

Page ahead for assistance on some special topics, such as how to get Lertap, how to determine versions numbers, and how to find out about updates.

4.1 How to topics

Please use these links for how-to information of various sorts.

- **How to get Lertap**
- **How to install Lertap**
- **How to run Lertap**
- **How to upgrade Lertap**
- **How to determine version number**
- **How to find out about updates**

4.2 Contact us

The support team for Lertap5 is prone to wander around a bit. Our main base is in South Fremantle, Western Australia, just west of the city of Perth. Office hours are (generally) 9 a.m. to 4 p.m. Western Australia Standard Time (no daylight saving period for us).

Wherever we are, almost all of the time we're reachable via **email**. The best address is lertap5@gmail.com.

You may write to us in *any language* you'd like! However, if you want us to understand what you've written, it's best to write in Burmese, English, or Spanish.

Love to hear from you.

4.3 About the author

Larry Nelson is originally from the United States. He gained an BSc in Electrical Engineering from the University of Wisconsin in 1964, and an MSc in Educational Psychology from the same university, Wisconsin, in 1970. Awarded a National Science Foundation Fellowship in 1970, he then completed a doctorate in Educational Psychology / Psychometrics at the Laboratory of Educational Research, University of Colorado, in 1973. His major professors were Kenneth D Hopkins and Gene V Glass.

Lertap has been his largest and most enduring software development project. A history of Lertap may be seen via this link.

Larry has lived and worked in the United States, Venezuela, New Zealand, and Australia. Extensive teaching, consulting, and advising assignments have had him also spending years in Indonesia and Thailand.

Over years he has worked in both educational technology and educational measurement.

Some of his work, especially as related to educational measurement, is cited in the references.

At April 2021, besides heading up the Lertap.com business, he was holding part-time positions at two universities: Curtin University in Western Australia, where he was affiliated with the School of Education as a Research Fellow, and Burapha University in Thailand, where he held an adjunct professorship in the College of Research Methodology and Cognitive Science (RMCS).

Contact Larry at lertap5@gmail.com or l.nelson@curtin.edu.au

Index

- $ -

- * -

- A -

- B -

- C -

- D -

- E -

- F -

- G -

- H -

- I -

- J -

- K -

- L -

- M -

- N -

- O -

- P -

- Q -

- R -

- S -

- T -

- U -

- V -

- W -

- X -

- Y -