

# Lertap Item-Criterion Correlation Calculations

Larry Nelson

Curtin University of Technology

Document date: 9 November 2004<sup>1</sup>

*A common question I receive from Lertap users goes this way: "Lertap's results don't agree with those produced by the program we used to use! What's up?"*

*This document is meant to provide an answer to this question – the usual reason for the discrepancy has to do with whether or not the "part-whole" correction has been applied to the calculations.*

This will be a very practical discussion, one which will require a small data set to get started. I'll use the "[ChemQuiz](#)" sample data set found on the Lertap website.

Table 1 displays the answers fifteen students gave to a ten-item multiple-choice class quiz on chemistry.

| Table 1: Student answers |          |    |    |    |    |    |    |    |    |    |     |
|--------------------------|----------|----|----|----|----|----|----|----|----|----|-----|
| No.                      | ID       | I1 | I2 | I3 | I4 | I5 | I6 | I7 | I8 | I9 | I10 |
| 1                        | Klien    | D  | B  | B  | C  | D  | D  | A  | C  | A  | B   |
| 2                        | Lampton  | B  | B  | B  | B  | D  | B  | C  | C  | B  | B   |
| 3                        | Mercurio | B  | B  | A  | B  | D  | B  | A  | C  | B  | D   |
| 4                        | Nelson   | B  | B  | C  | B  | D  | C  | B  | C  | C  | B   |
| 5                        | Oldfelt  | B  | B  | A  | B  | B  | C  | C  | C  | B  | B   |
| 6                        | Primo    | B  | A  | B  | B  | D  | C  | A  | C  | B  | A   |
| 7                        | Regalado | D  | D  | A  | B  | D  | C  | A  | C  | B  | A   |
| 8                        | Smith    | B  | B  | B  | D  | D  | C  | C  | D  | B  | B   |
| 9                        | Terace   | B  | B  | A  | D  | B  | C  | B  | C  | A  | A   |
| 10                       | Uptown   | B  | D  | B  | C  | A  | B  | C  | C  | B  | B   |
| 11                       | Virgo    | D  |    | B  | A  | D  | C  |    | C  | A  | A   |
| 12                       | Westphal | B  |    | A  | A  |    | C  |    | A  | B  | A   |
| 13                       | Xeno     | B  |    | B  | D  | A  | C  | C  | A  | B  | B   |
| 14                       | Yalso    | A  | C  | A  | D  | B  | C  |    | A  | D  | A   |
| 15                       | Zenu     | A  | D  | A  | A  | A  | C  | A  | C  | B  | A   |

The items in Table 1 have been labelled I1 to I10. Each item used {A,B,C,D} as options. Some students failed to answer some of the items – their "responses" are seen in Table 1 as empty cells; for example, Virgo did not answer I2; Yalso did not answer I7.

---

<sup>1</sup> Hyperlinks updated 14 December 2022

The answer key is shown in Table 2.

| Table 2: Correct answers |    |    |    |    |    |    |    |    |     |
|--------------------------|----|----|----|----|----|----|----|----|-----|
| I1                       | I2 | I3 | I4 | I5 | I6 | I7 | I8 | I9 | I10 |
| B                        | B  | B  | B  | D  | C  | A  | C  | B  | B   |

With the answer key in hand, I can score the students on each item. Klien incorrectly answered I1, equal to an item score of 0 (zero). Klien answered I2 and I3 correctly, getting one point for each.

Table 3 shows the item scores for all fifteen students.

| Table 3: Item scores |          |    |    |    |    |    |    |    |    |    |     |
|----------------------|----------|----|----|----|----|----|----|----|----|----|-----|
| No.                  | ID       | I1 | I2 | I3 | I4 | I5 | I6 | I7 | I8 | I9 | I10 |
| 1                    | Klien    | 0  | 1  | 1  | 0  | 1  | 0  | 1  | 1  | 0  | 1   |
| 2                    | Lampton  | 1  | 1  | 1  | 1  | 1  | 0  | 0  | 1  | 1  | 1   |
| 3                    | Mercurio | 1  | 1  | 0  | 1  | 1  | 0  | 1  | 1  | 1  | 0   |
| 4                    | Nelson   | 1  | 1  | 0  | 1  | 1  | 1  | 0  | 1  | 0  | 1   |
| 5                    | Oldfelt  | 1  | 1  | 0  | 1  | 0  | 1  | 0  | 1  | 1  | 1   |
| 6                    | Primo    | 1  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0   |
| 7                    | Regalado | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 0   |
| 8                    | Smith    | 1  | 1  | 1  | 0  | 1  | 1  | 0  | 0  | 1  | 1   |
| 9                    | Terace   | 1  | 1  | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0   |
| 10                   | Uptown   | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 1  | 1   |
| 11                   | Virgo    | 0  | 0  | 1  | 0  | 1  | 1  | 0  | 1  | 0  | 0   |
| 12                   | Westphal | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0   |
| 13                   | Xeno     | 1  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 1  | 1   |
| 14                   | Yalso    | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0   |
| 15                   | Zeno     | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 0   |

Note: This quiz's items have been scored on a simple right/wrong basis, with 1 (one) point for right answers, and 0 (zero) points for wrong answers; unanswered items have been scored as incorrect. This is the most common type of scoring for cognitive items, but, when you want it to, Lertap will score the items in other ways – for example, it's possible to have more than one right answer to an item, and it is possible to give more than one point for correct answers.

Now let me derive two scores for each student: S1 will be the score obtained from all ten items, while S2 will be the score on items I2 through I10. This will give me Table 4.

| Table 4: With test scores S1 & S2 |          |    |    |    |    |    |    |    |    |    |     |    |    |
|-----------------------------------|----------|----|----|----|----|----|----|----|----|----|-----|----|----|
| No.                               | ID       | I1 | I2 | I3 | I4 | I5 | I6 | I7 | I8 | I9 | I10 | S1 | S2 |
| 1                                 | Klien    | 0  | 1  | 1  | 0  | 1  | 0  | 1  | 1  | 0  | 1   | 6  | 6  |
| 2                                 | Lampton  | 1  | 1  | 1  | 1  | 1  | 0  | 0  | 1  | 1  | 1   | 8  | 7  |
| 3                                 | Mercurio | 1  | 1  | 0  | 1  | 1  | 0  | 1  | 1  | 1  | 0   | 7  | 6  |
| 4                                 | Nelson   | 1  | 1  | 0  | 1  | 1  | 1  | 0  | 1  | 0  | 1   | 7  | 6  |
| 5                                 | Oldfelt  | 1  | 1  | 0  | 1  | 0  | 1  | 0  | 1  | 1  | 1   | 7  | 6  |
| 6                                 | Primo    | 1  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0   | 8  | 7  |
| 7                                 | Regalado | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 0   | 6  | 6  |
| 8                                 | Smith    | 1  | 1  | 1  | 0  | 1  | 1  | 0  | 0  | 1  | 1   | 7  | 6  |
| 9                                 | Terace   | 1  | 1  | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0   | 4  | 3  |
| 10                                | Uptown   | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 1  | 1   | 5  | 4  |
| 11                                | Virgo    | 0  | 0  | 1  | 0  | 1  | 1  | 0  | 1  | 0  | 0   | 4  | 4  |
| 12                                | Westphal | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0   | 3  | 2  |
| 13                                | Xeno     | 1  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 1  | 1   | 5  | 4  |
| 14                                | Yalso    | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0   | 1  | 1  |
| 15                                | Zeno     | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 0   | 4  | 4  |

Ready? I next want to find the correlation between I1 and the total test score, S1.

There are numerous ways I could do this. I could get out my old statistics books and a calculator, and then simply follow the steps to compute the value of the Pearson product-moment correlation coefficient, using I1 scores as the "X" variable, and S1 scores as "Y".

Or, I could scratch my head, and recall something from Psych 210: there are computational shortcuts which will cover this case: I1 is a dichotomous variable, one which assumes just two values. I could use one of the formulas for the "point-biserial" correlation, such as that found in Glass and Stanley (1970), or that found in Crocker and Algina (1986) – these formulas will give the same result as that for the Pearson product-moment method, but they're easier to use.

Which of these approaches will I use? Neither. I'll put the data into Excel, and ask it to calculate the correlation.

Using Excel, I found the correlation between I1 and S1 to be 0.465. *This is the point-biserial correlation which some item analysis programs will display for I1.*

But not Lertap.

The value of 0.465 is inflated; correlating I1 with S1 will produce what Crocker and Algina (1986, p.317) refer to as "somewhat spurious" results. I1 is part of S1.

Lertap corrects for this part-whole inflation by correlating I1 with S2, with the modified test score formed by omitting I1 from S1.

Excel found the correlation between I1 and S2 to be 0.241. *This is the value of the point-biserial coefficient found in Lertap for I1.*

The correction for part-whole inflation makes quite a difference in this little example, doesn't it? But this won't always be the case – this example has only ten items; as the number of items increases, the inflation effect decreases. With 30 or more items the effect will usually be very small, and probably not worth much concern.

### **External Criterion Analysis**

The manual for Lertap 5 (Nelson, 2000, p.158) suggests that the inflation effect can be seen by running Lertap's external criterion analysis, an option available from Lertap's Run menu.

The manual has a few examples of external criterion runs (see pp.134-138 in the printed manual; the section title is "Using an external criterion"). To check on part-whole inflation effects, you'd follow the steps at the bottom of p.135, selecting the Scores column which corresponds to the subtest you want to check. For example, if I had just one subtest, called "ChemQuiz", I'd select the Scores column where ChemQuiz results appear, which would most likely be column 2. In the ECStats report which Lertap will then produce, the values of the uncorrected item correlations are found as the "**r/ec**" figures.

These **r/ec** item correlation values can be compared with the item correlations found by your old item analysis program. I believe quite a number of present-day item analysis programs do not correct for part-whole inflation. The last time I looked in detail at *LXRTest*, for example, it did not correct its correlation coefficients, and Dave Weiss' veteran *Iteman* program also seems to ignore this problem. But, as mentioned above, the part-whole inflation effect is really only a potential problem when the number of test items is small (Crocker and Algina (1986) suggest that the effect might be noticed when there are less than 25 items).

What are the dangers of using uncorrected correlation figures? Well, the items are going to look better than they are. Look at the following Tables, 5 and 6:

**Table 5: Corrected r's****item discrimination bands**

**.00:** I3 I6 I7  
**.10:** I9  
**.20:** I1 I8 I10  
**.30:** I2  
**.40:** I5  
**.50:** I4  
**.60:**  
**.70:**  
**.80:**  
**.90:**

**Table 6: Uncorrected r's****r/ec correlation bands**

**.00:** I6  
**.10:**  
**.20:** I7  
**.30:** I3  
**.40:** I1 I8 I9 I10  
**.50:** I2  
**.60:** I5  
**.70:** I4  
**.80:**  
**.90:**

Table 5 displays corrected item correlations, while Table 6 features the same items, now wearing their uncorrected figures.

These are standard Lertap tables; Table 5 is always found close to the end of a Stats1f report, while Table 6 will appear towards the bottom of an ECStats1f report. The Stats1f report, or Stats2f, or Stats3f ... is always produced whenever the Elmillon option is taken from Lertap's Run menu. An ECStats1f, or ECStats2f, or ECStats3f report is produced when you use the Run menu to ask for an External criterion analysis.

You recall how to read these tables? Table 5 says that three items, I3, I6, and I7, had correlations less than **.10:**, while only three, I2, I5, and I4, had fairly strong correlations. But, when seen in Table 6, the items look quite a bit better, eh?

Table 7 provides more detail. The "corrected r" values come from Lertap's Stats1b report, while the uncorrected counterparts are from the ECStats1f report.

Of interest is the third statistic, "Difference", showing the effect of part-whole inflation for each item.

| <b>Table 7: Selected item statistics</b> |      |      |      |      |      |       |      |      |      |      |
|--|------|------|------|------|------|-------|------|------|------|------|
| - Statistic -                            | I1   | I2   | I3   | I4   | I5   | I6    | I7   | I8   | I9   | I10  |
| Corrected r                              | .241 | .312 | .073 | .557 | .448 | -.504 | .025 | .250 | .157 | .229 |
| Uncorrected r                            | .465 | .536 | .328 | .720 | .642 | -.323 | .269 | .459 | .391 | .467 |
| Difference                               | .224 | .224 | .255 | .163 | .194 | .181  | .244 | .209 | .234 | .238 |
| Item s.d.                                | .471 | .499 | .499 | .490 | .499 | .442  | .471 | .442 | .471 | .499 |
| Avg. r                                   | .092 | .119 | .025 | .210 | .168 | -.245 | .013 | .096 | .065 | .085 |
| SMC                                      | .676 | .731 | .773 | .780 | .780 | .415  | .639 | .558 | .606 | .677 |

Most items gained about .2 of a correlation, but you'll want to note that the gains were not exactly constant – some items gained a bit more than .2, some a bit less.

Why aren't the gains the same for each item?, you might ask.

I was hoping this question might occur to you. The size of the gain can be expected to relate to the item's variance, and to its correlation with the other items. In general, we'd expect items with greater standard deviation, and higher correlations with the other items, to gain the most from the inflation effect.

Did they? To find out, I used another option on Lertap's Run menu, "Output item scores matrix", to get the extra statistics shown at the bottom of Table 7.

Item s.d., standard deviation, is an indication of variance. "Avg. r" is the average item correlation with the other nine items, while SMC is the value of the squared multiple correlation between each item and the other nine.

I failed to detect any ready patterns in Table 7, so I asked Excel to correlate all the statistics. The greatest correlate with "Difference" was Item s.d., but it was only 0.25. The second-best was SMC, at a truly whopping 0.14. No wonder I didn't see a pattern in Table 7 – these figures are not large, but they do, at least, agree with theory. (Item s.d. and SMC were themselves highly correlated, at 0.89.)

## Summary

I have created this little paper to address a question often sent in by Lertap users: *"Why don't Lertap's results match those produced by the item analysis program I used to use?"*.

I hope the answer is now clear to you. Lertap corrects its correlations for the effects of part-whole inflation.

When a test has a small-ish number of items, say less than 30, the part-whole inflation effect will serve to noticeably boost an item's correlation with the criterion, an outcome which might quite possibly lead you to conclude that the items worked better than they actually did. In the example featured above, item correlations were found to gain approximately .2 from inflation effects.

## References

- Crocker, L.M. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.
- 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.
- Nelson, L.R. (2000). *Item analysis for tests and surveys using Lertap 5*. Perth, Western Australia: Curtin University of Technology.