



Analysis of reader reliability using LISREL

In an experiment (Votaw, 1948) to establish methods of obtaining reader reliability in essay scoring, 126 examinees were given a three-part English Composition examination. Each part required the examinee to write an essay, and for each examinee, scores were obtained on the following:

1. the original part 1 essay, represented by the variable 'Original part1'
2. a handwritten copy of the original part 1 essay ('Written copy')
3. a carbon copy of the handwritten copy in (2) ('Carbon copy'), and
4. the original part 2 essay, represented by the variable 'Original part2'.

Scores were assigned by a group of readers using procedures designed to counterbalance certain experimental conditions. The investigator would like to know whether, on the basis of this sample of size 126, the four scores can be used interchangeably or whether scores on the copies (2) and (3) are less reliable than the originals (1) and (4).

The covariance matrix of the four measurements is given in the command file below. The hypothesis to be tested in this example is that a one-factor congeneric measurement model describes these data well.

The LISREL command file for this analysis is (**EX31A_16.LIS**):

```
Analysis of Reader Reliability in Essay Scoring Votaw's Data
Congeneric model estimated by ML
DA NI=4 NO=126
LA
'ORIGINAL PART1' 'WRITTEN COPY' 'CARBON COPY' 'ORIGINAL PART2'
CM
25.0704
12.4363 28.2021
11.7257 9.2281 22.7390
20.7510 11.9732 12.0692 21.8707
MO NX=4 NK=1 LX=FR PH=ST
LK
```

```
'Essay ability'
PD
OU SE ND=2
```

The DA command specifies four observed variables and a sample size of 126; the MA default is assumed, so the covariance matrix will be analyzed. Labels for the input variables follow the LA command. The CM command indicates that a covariance matrix is to be input. Because an external file is not specified, the matrix follows in the command file. A format statement does not appear, so the input is in free format. The MO command specifies four x -variables and one latent variable; the elements of λ are all free (LX = FR), and the latent variable is standardized (PH = ST). A label for the latent variable follows the LK command. Note that all observed variable names are given in single quotes, as all names have a blank space between parts of the name. The same holds true for the latent variable 'Essay ability'.

In the results of this analysis, the goodness-of-fit statistic

| | |
|--|--------------------|
| Degrees of Freedom for (C1)-(C2) | 2 |
| Maximum Likelihood Ratio Chi-Square (C1) | 2.298 (P = 0.3169) |
| Browne's (1984) ADF Chi-Square (C2_NT) | 2.236 (P = 0.3270) |

indicates that the hypothesis is acceptable.

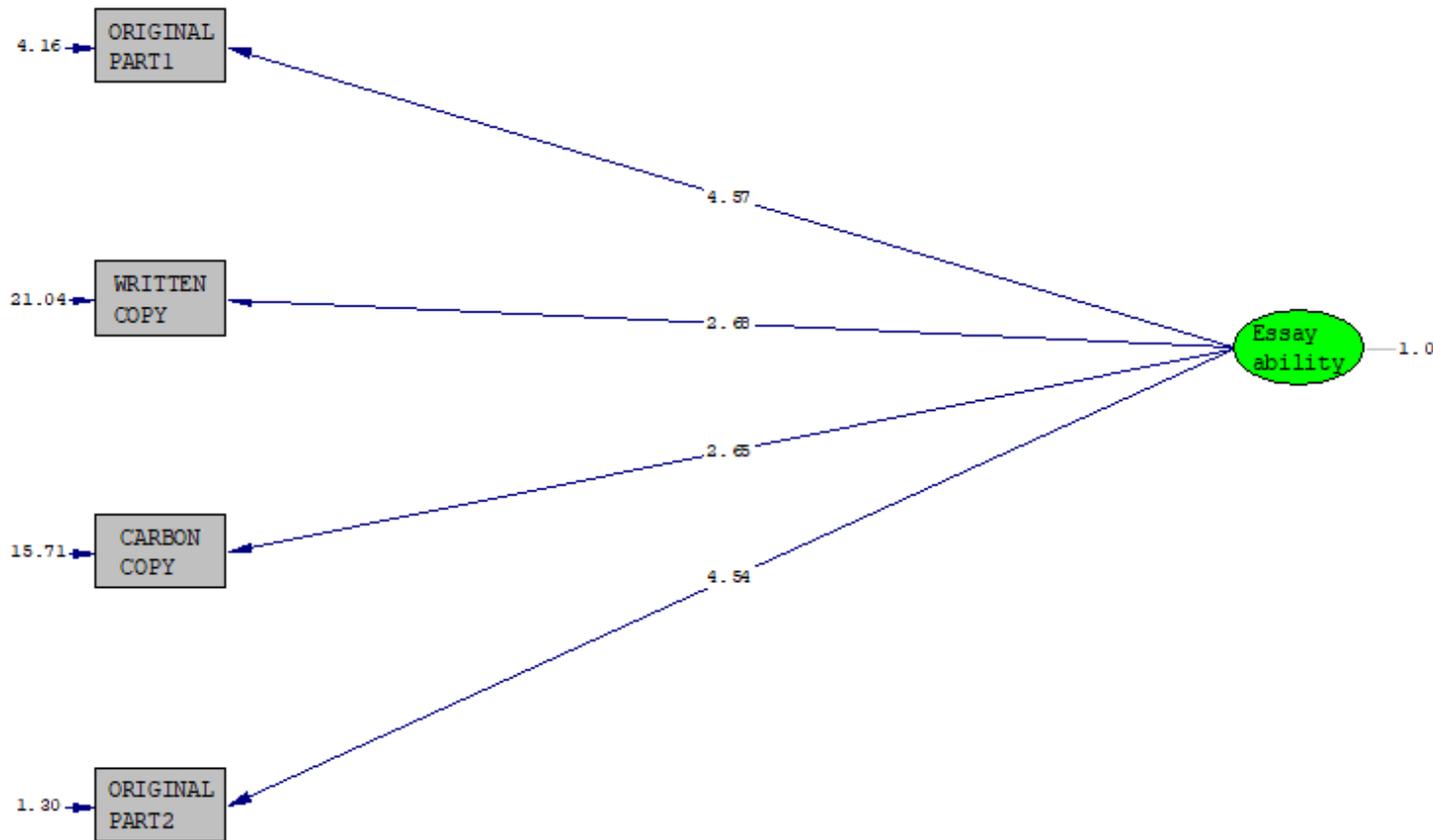
The results under the hypothesis are given in the table below. The three columns of this table can be read off directly from the output for the ML solution. The reliabilities in column 3 appear where the output says “squared multiple correlations for x -variables”.

Table: Essay scoring data: results for congeneric model

| I | $\hat{\lambda}_i$ | $se(\hat{\lambda}_i)$ | $\hat{\rho}_{ii}$ |
|-----|-------------------|-----------------------|-------------------|
| 1 | 4.57 | 0.36 | 0.83 |
| 2 | 2.68 | 0.45 | 0.25 |
| 3 | 2.65 | 0.40 | 0.31 |
| 4 | 4.54 | 0.33 | 0.94 |

Inspecting the different λ 's, it is evident that these are different even taking their respective standard errors of estimate into account. Comparing the reliabilities in the last column, one sees that they are high for scores (1) and (4) and low for scores (2) and (3). Thus, it appears that scores obtained from originals are more

reliable than scores based on copies. The path diagram obtained for this model is shown below.



Chi-Square=2.30, df=2, P-value=0.31690, RMSEA=0.034