UiO : Universitetet i Oslo

CANDIDATE

184116

TEST

MAE4011 1 Principles of Measurement

Subject code	MAE4011
Evaluation type	Individuell skriftlig prøve
Test opening time	20.12.2022 09:00
End time	20.12.2022 13:00
Grade deadline	--
PDF created	26.12.2022 19:33
Created by	Tony Clifford Austin Tan

9 SR1H22

A scale to measure depression severity was developed and data were collected from a large group of students, along with the scores of an existing scale for satisfaction with life.

You observed the following covariance matrix for the scores of the two scales, where X denotes the depression severity scale scores and Y denotes the satisfaction with life scale scores:

$$\mathbf{\Sigma} = \begin{pmatrix} 10 & -7 \\ -7 & 10 \end{pmatrix}.$$

Based on these observations, how would you characterize the relationship between depression severity and satisfaction with life?

State the assumptions made in the interpretations of the relationship.

Fill in your answer here

Corr(X,Y)= COV(X,Y)/ sqrt of varianceX, sqrt of varianceY
= (-7)/sqrt10 x sqrt10 = -7/10 = -0.7

The correlation between X and Y is moderately strong. There is a negative, linear relationship between the two.
Assumption is large enough sample and single factor model holds.

Words: 44

Answered.

11 SR3H22

X and Y are two random variables where $\mathbf{Var}(X) = 2$, $\mathbf{Var}(Y) = 3$ and $\mathbf{Cov}(X, Y) = 1$.

1. Calculate $\mathbf{Var}(Z)$, where $Z = X - Y$. Show your work.
2. Calculate $\mathbf{Var}(U)$, where $U = X + 2Y$. Show your work.

Fill in your answer here

Var(Z) = Var(X-Y) = Var(X) + Var(Y) - 2COV(X,Y) = 2 + 3 - 2(1) = 3

Var(U) = Var(X + 2Y) = Var(X) + 2^2Var(Y) + 2COV(X,Y) = 2 + 4(3) + 2(1) = 16

Words: 36

Answered.

12 SR4H22

Let m be the number of items on a test. For a five-item test, the common factor loading λ was 1 and the variance of the sum score Y was 10. Compute coefficient alpha

$\alpha = m\frac{\lambda^2}{\mathrm{Var}(Y)}$

and interpret it. State the assumptions underlying the interpretation.

Fill in your answer here

alpha = 5 (1^2)/10 = 0.5

Given there is tau-equivalence with the common factor loadings, and if the single factor model holds, we can say that the coefficient alpha, which represents the lower bound reliability, is equivalent to the reliability coefficient of the true score. However, the value of 0.5 is below the threshold of 0.8 of acceptable.

Words: 58

Answered.

13 SR5H22

The *Standards for Educational and Psychological Testing* (2014) state that it is useful to consider ways in which the test scores can be influenced by either (1) too much or (2) too little.

A three-domain test is administered for the purpose of measuring Norwegian 15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges. The test is a low-stakes test for the respondents since individual assessment is not of interest.

Provide **one example** of a way in which the test-scores might be influenced by too much, and **one example** of how the test-scores might be influenced by too little.

Fill in your answer here

Too much : extraneous variables like anxiety, illness
Too little: inadequate representation of items to measure the construct

Words: 18

Answered.

14 SR6H22

For two tests of reading comprehension, X and Y, the linear equating function was estimated to be $\mathbf{eq}(Y) = 1.2X + 6$. The cut score for passing test Y was determined to be 30.
Give the cut score for pass in terms of the test X scores, based on the estimated equating function. Present and explain how the result was obtained.

Fill in your answer here

eq(Y) = 1.2X + 6
30 = 1.2X + 6
30 -6 = 1.2X
24/1.2 = 1.2X/1.2
20 = X

The cut score for test X = 20

Words: 28

Answered.

15 SR7H22

Item scores on a test of mathematics and a test of interest in mathematics were given to the same group of students. A two-factor model with correlated factors (one factor measured by the mathematics test items and the other by the interest in mathematics items) was estimated, yielding the model fit indices:

GFI	0.95
RMSEA	0.05
SRMR	0.06

The correlation between the sum scores of the respective tests was 0.2 while the estimated factor correlation was 0.5. Explain why there is a difference in the factor correlation and the sum score correlation in this context.

Fill in your answer here

The thresholds for good fit indices are :
GFI > 0.95
RMSEA < 0.05
SRMS < 0.08
Other includes: mean residual = 0

In the 2-factor model, it seems that the model has a good fit. The estimated factor correlation should be higher than the correlation between sum scores of each test, which is the case here. There is moderately strong relationship between the test of mathematics and test of interest in mathematics and it's positively linear. However, the correlation between the sum scores are low which indicates some measurement error. This could benefit from correction with attenuation.

Words: 98

Answered.

16 SR8H22

A bifactor model with one general factor and two subfactors (all factors independent) was estimated for an Norwegian test with two subdomains (reading and writing), yielding the following factor loading estimates:

Item	General	Reading	Writing
1	3	0.5	0
2	1	0.5	0
3	2	1	0
4	1	0	1
5	1	0	0.5
6	1	0	0.5

The model fit was judged to be satisfactory.

In a previous study, the sum score was used. Based on the estimated factor loadings, would you recommend doing this? Justify your answer.

Fill in your answer here

General factor loadings squared/n-1 = 9+1+4+1+1+1 = 17/5 = 3.4
The factor loadings of the subfactors ^2/n-1= 0.25+0.25+1+1+0.25+0.25/5=0.6

Given that the value of the one general factor is larger than that of the bifactor model, means that using the sum scores in this case was not justified because of multidimensionality.

Words: 50

Answered.

17 LR1H22

You have been asked to assist a group of teachers of Norwegian as a foreign language to find the appropriate cut-score for a test of Norwegian reading proficiency.

As part of the process, the test was piloted with a representative sample of the intended population and the results are available to you. In addition, an established framework describes the expected level of Norwegian reading proficiency.

Give a brief outline of how a standard-setting procedure could be used to find the cut-score for pass/fail on the Norwegian reading proficiency test.

Fill in your answer here

A method of standard-setting procedure is the Bookmark method where in it uses IRT to map items in proficiency distribution. A description of performance level is created, and a group of panelists are trained and go through a series of rounds. A pre-arranged Ordered Item Booklet is used wherein items are arranged based on difficulty. Each page contains 1 item. The panelist will then put a bookmark between two items such that a minimally performing student of the domain (Norwegian reading proficiency) is expected to have mastered the item below the bookmark and have not mastered the item above. An RV probability is used to determine the arrangement of items in the OIB.

Words: 113

Answered.

18 LR2H22

A scale is being developed to measure satisfaction with life with the intended purpose to use the scale in national survey to identify which factors are associated with high satisfaction of life in the population. The scale consists of Likert items. According to the underlying theory of satisfaction with life, it is a unidimensional attribute. The theory also states that satisfaction with life is expected to have differences based on gender.

With this information in mind, do the following:

- Describe what evidence sources you want to consider in order to evaluate the validity of the scale scores for their intended purpose
- Describe the data you would like to collect to conduct the validity study
- Describe the analyses you would do in the validity study
- Outline what results you would consider as evidence supporting the validity of using the scale scores in the national survey

Fill in your answer here

The evidence categories would be test content, internal structure and relation to external variables. The evidence for test content is of alignment, terms and wording should be operationalized. Expert judgement and opinion should be sought. The scale should contain items that represent the construct and the theories in the framework. The evidence for internal structure tests for intra-construct relationship. The correlation of the items with the construct has to be measured. Need to check factor loadings, error variances, if the item/s truly represent satisfaction with life and if a single factor model holds. The evidence in relation to external variables looks for inter-constuct relationship. Other scale that measure satisfaction with life can be compared with the current scale and look for convergent evidence, meaning the scale correlates highly with another scale said to be measuring the same construct as well as discriminatory evidence where it should not correlate with scales not measuring the construct. The test has to be administered to a sample representing the target population. Data on gender should be included. Tests for reliability and correlation are needed. A model can be fitted to check for indices and check for unidimensionality.

Words: 193

Answered.

19 LR3H22

The following output was obtained from estimating a single factor model to five 4-category Likert scale items from a scale measuring the environmental awareness of 15-year olds in Norway.

Item	Factor loading	Error variance
1	2.00	4.00
2	3.00	2.00
3	1.00	4.00
4	2.00	5.00
5	2.00	1.00

The residual correlation matrix was

$$\boldsymbol{\Sigma}_{\mathbf{res}} = \begin{pmatrix} \mathbf{0.000} & & & & \\ \mathbf{0.026} & \mathbf{0.000} & & & \\ \mathbf{0.017} & \mathbf{-0.035} & \mathbf{0.000} & & \\ \mathbf{-0.014} & \mathbf{0.072} & \mathbf{-0.019} & \mathbf{0.000} & \\ \mathbf{-0.025} & \mathbf{-0.039} & \mathbf{0.020} & \mathbf{0.009} & \mathbf{0.000} \end{pmatrix}.$$

Address the following in your response:

1. What validity evidence categories from the Standards for Educational and Psychological Testing are relevant in this analysis? (1p)
2. Based on your appraisal, does the single factor model fit well?
3. Assume that a single factor model is appropriate for the analysis of the five item scores. Which item contributes the most to the reliability of the sum score and which item contributes the least? Justify your answers. (1p)
4. From the description of the items above and the results of the estimated model, give **one reservation** against the use of the linear factor model in this case. (1p)

Fill in your answer here

The evidence category is internal structure.
Based on item information = λ^2/error variance:
Item 1 = $2^2/4 = 4/4 = 1$
Item 2 = $3^2/2 = 9/2 = 4.5$
Item 3 = $1^2/4 = 0.25$
Item 4 = $2^2/5 = 0.8$
Item 5 = $2^2/1 = 4$

The item that contributes the most to the reliability of ss is Item 2. While Item 3 contributes the least.

The sum of the residuals = 0.026 + .017 - .014 -.025 -.035 + .072 - .039 - .019 + .020 + 0.009 = 0.012
The mean of the residuals = 0
Aside from mean residual = 0, we would need other indices to appraise the if the single factor model fit well.
omega = sum of factor loadings squared/sum of factor loadings square+sum of error variance
omega= 10^¨2/10^¨2+16 = 0.8
Based on the coefficient omega, the reliability is within the accepted threshold, assuming the single factor model holds.

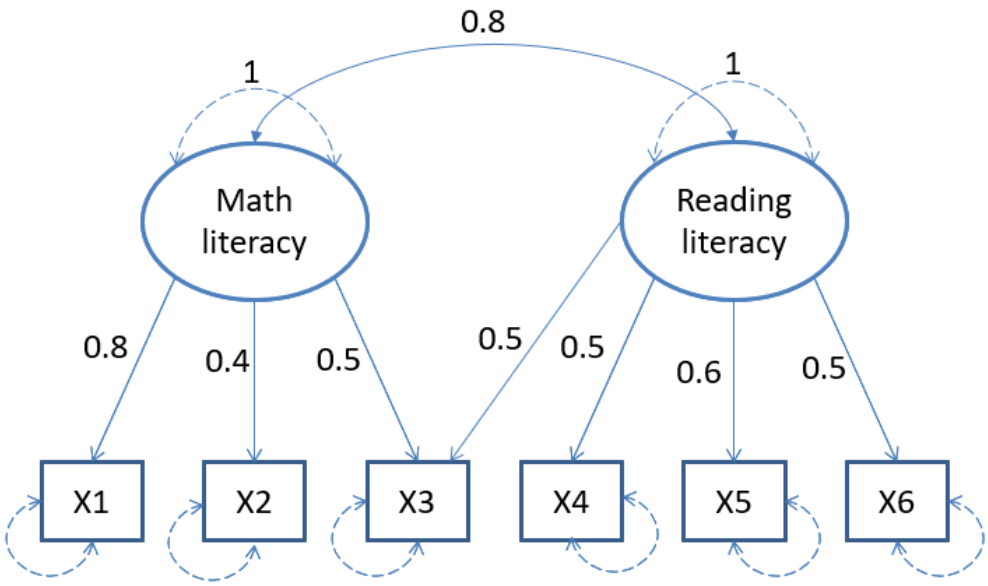
Given that it is a likert scale, thus ordinal, the use of linear factor model is not ideal.

Words: 175

Answered.

20 LR4H22

A multiple factor model is illustrated in the graph below. The latent variables and the observed variables are all standardized.



Answer the following questions based on the graph.

1. What is the equation which describes the model for the item score **X3**? Write down the equation with an explanation of the parameters and variables included. (2p)
2. What is the covariance between item scores **X3** and **X4** according to the model? Show your work and explain the steps taken. (2p)

Fill in your answer here

Item X3 = $\mu 3 + \lambda_{M3}M + \lambda_{R3}R + \epsilon_3$
Where $\mu 3$ is item difficulty, $\lambda M3$ and $\lambda R3$ are the factor loading on Math and Reading literacy respectively, M is the Math literacy factor and R is the reading literacy factor, $\epsilon 3$ is the error term of X3.

X4 = $\mu 4 + \lambda_{R4}R + \epsilon_4$
cov(X3, X4) = cov (mu3+lambdaM3M+lambdaR3R+e3, mu4+lambdaR4R+e4)
=cov(mu3,mu4)+cov(mu3,lambdaR4R)+cov(mu3,e4)+cov(lambdaM3M,lambdaR4R)+cov(lambdaM3M,mu4)+cov(lambdaM3M, e4)+cov(lambdaR3R,
mu4)+cov(lambdaR3R,lambdaR4R)+cov(lambdaR3R, e4)+
cov(e3, mu4)+cov(e3,lambdaR4R)+cov(e3,e4)
= 0+0+0+cov(lambdaM3M,lambdaR4R)+0+0+0+cov(lambdaR3R,lambdaR4R)+0+0+0+0
= lambdaM3lambdaR4 Cov(M,R)+lambdaR3lambdaR4varR
= (0.5)(0.5)(0.8)+(0.5)(0.5)(1)
= 0.2+0.25=0.45

Words: 82

Answered.