

i **General information MAE4011 HT19**

Some guidelines:

- The exam consists of a combination of multiple choice questions and free-response questions. Note that the multiple choice questions have only one category which is the correct one.
- When answering questions that ask you to interpret the results, you can obtain partial credit even if you did not calculate the quantities correctly. As long as your interpretation is consistent with your calculations you will obtain at least partial credit.
- Read the questions carefully and write down if you do not understand something about a question.
- The responsible teacher will visit the exam site for a short period during the first two hours of the exam.

Good luck!

/ Björn, Henrik and Haakon

1 **Parameters 1**

Which of the following is a parameter?

Select one alternative:

- The error score.
- The error variance.
- The observed sum score.
- The observed item score.

Maximum marks: 2

2 **Classical test theory 1**

Which of the following statements are always true for the classical true score model?

Select one alternative:

- The true score has a normal distribution.
- The error variances of different observed scores are the same.
- Higher true score variance means lower reliability.
- True scores and error scores are uncorrelated.

Maximum marks: 2

3 Coefficient alpha 1

When is coefficient alpha equal to the reliability coefficient?

Select one alternative:

- When the error variance is larger than the true score variance.
- When the item scores all have the same mean.
- When the items have equal factor loadings.
- When the error variances are all equal.

Maximum marks: 2

4 Single factor model 1

You have fitted a single factor model to a number of mathematics items. What would be evidence for using the sum score as an indicator for a single attribute?

Select one alternative:

- That the factor loadings are all equal.
- That the goodness-of-fit index is above 0.95.
- That the error variances are similar in magnitude.
- That the mean item scores are the same for all items.

Maximum marks: 2

5 Factor model 2

Consider the factor model

$$X_j = \mu_j + \lambda_j F + E_j,$$

where F has variance 1. The estimated parameters for an item were $\mu_j = 1$, $\lambda_j = 1$ and $\Psi_j^2 = 2$,

where Ψ_j^2 denotes the variance of E_j . What is the variance of X_j according to the estimated parameters?

Select one alternative:

- 3
- 4
- 5
- 6

Maximum marks: 2

6 Item bias 1

Assume that a single factor model is appropriate for the analysis of several items. Which of the following statements would be evidence of item bias?

Select one alternative:

- The mean observed item scores are different for men and women.
- The variance of the observed item scores are different for men and women.
- The reliability coefficient for the sum scores are different for men and women.
- None of the above.

Maximum marks: 2

7 Equating 1

For which of the following settings is equating necessary?

Select one alternative:

- When using an environmental awareness scale to identify 15-year old students who lack knowledge of the environment.
- When using a scale to evaluate the level of depression among teenagers.
- In a low-stakes national exam where the scores at the school level are used to infer differences in performance over time.
- In a high-stakes college entrance exam meant to identify the top 10% of students in a given final year class.

Maximum marks: 2

8 Validity 1

Which of the following statements represents the 2014 Standards for Educational and Psychological Testing's definition of "Construct"?

Select one alternative:

- An observable property of individuals which is the cause of test performances or responses.
- The concept or characteristic that a test is designed to measure.
- A useful theoretical construction used for categorizing test-taking behaviors.
- A label of convenience denoting the measurement operations performed by the test user.

Maximum marks: 2

9 Validity 3

Which of the following statements is most in line with the validity theory offered by the 2014 Standards for Educational and Psychological Testing?

Select one alternative:

- All sources of evidence should always be given equal attention when validating the interpretation of a test-score for a specific use.
- Factor analysis is required to establish construct validity and can be considered a method for investigating content validity.
- The type of evidence required for validation depends on whether it is content, criterion, or construct validity that is to be established.
- When a test-score is to be interpreted in several ways, each of those interpretations must be subjected to validation.

Maximum marks: 2

10 Scales_02_MC

Please select the relevant scale-types.

	Nominal	Ordinal	Interval	Ratio
The score on this exam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Celsius	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maximum marks: 2

11 Validity 4

The Standards state that validation practice can be guided by considering alternative hypotheses about interpretations of test-scores for particular uses. To aid generation of alternative hypotheses, the Standards state that it is useful to consider ways in which the test scores can be influenced by either (1) too much or (2) not enough.

An exam is administered for the purpose of grading students on their mastery of a given curriculum. 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

Format
-
B
I
U
 x_2
 x^2
 I_x

 Ω

 Σ
ABC

Words: 0/50

Maximum marks: 2

12 Item statistics 1

Consider the discrete random variable X which can take the values 0, 1, 2 and 3 with probabilities:

$$P(X=0) = 0.1$$

$$P(X=1) = 0.2$$

$$P(X=2) = 0.5$$

$$P(X=3) = 0.2$$

The mean of X is .

The median of X is .

Maximum marks: 2

13 **Statistics 4**

Let X be a random variable with finite expected value and variance. Define $Y = aX + b$, where a and b are constants. Show that $E(Y) = aE(X) + b$ and that $Var(Y) = a^2 Var(X)$.

Fill in your answer here

Maximum marks: 2

14 **Covariance 2**

The following covariance matrix was observed for two tests where one measured mathematics and one measured reading literacy:

$$\begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$$

Based on this covariance matrix and possible calculations from it, do you consider mathematics literacy and reading literacy to be highly related? Justify your answer and list the assumptions made. Note that the correlation between two random variables X and Y is $\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$.

Fill in your answer here

Maximum marks: 2

15 **Coefficient alpha 3**

Coefficient alpha is defined as:

$$\alpha = \frac{m}{m-1} \left[1 - \frac{\sum_{j=1}^m \sigma_{X_j}^2}{\sigma_Y^2} \right]$$

The following proportions correct were observed for three binary items: 0.5, 0.5, 0.5. The sum score variance was 1.5.

a) Show that the observed coefficient alpha is equal to 0.75.

Fill in your answer here

b) Interpret the observed coefficient alpha value.

Enter text here

Maximum marks: 2

16 **Scales 2**

Trends in International Mathematics and Science Study (TIMSS) is an international large-scale assessment administered every four years. Norway has participated in five cycles of TIMSS since 1995, and the mean science results for Norwegian fourth-graders from 1995 to 2015 is shown in the table below.

Year	Score
1995	504
2003	466
2007	477
2011	494
2015	493

The TIMSS score reporting scale in science achievement has been kept consistent since 1995, when the mean was defined as 500 and one standard deviation was defined as 100. The standard error of measurement is approximately 3 points.

Based on the above table, describe how the science proficiency of Norwegian fourth graders has developed between 1995 and 2015.

Maximum marks: 2













i Validity 5

Kobrin et al. (2008) provided a validation study where high-school grade-point average (HSGPA) and SAT scores were evaluated with respect to their use as selection-criteria for admission into higher education programmes. As a criterion of success, they used the first-year grade-point average (FYGPA). They found that both the HSGPA and the SAT predicted FYGPA, and that when taken together, the HSGPA and SAT provided more accurate prediction of FYGPA than either of them by themselves. The authors concluded that both the HSGPA and the SAT should be used for selection into higher education.

17 Validity 5a

With reference to the evidence-categories of the Standards, provide arguments for and against the use of FYGPA as a criterion.

Fill in your answer here

Format | **B** | *I* | U | x_2 | x^2 | I_x |  |  |  |  |  |  |  |  |  |  |  | 











Words: 0/100

Maximum marks: 2

18 Validity 5b

Imagine that the authors concluded that only the SAT should be used for selection into higher education. Drawing on concepts from the Standards, consider potential consequences of such a change on the system as a whole, both in terms of the acceptability of the test use and the validity of test score interpretations for the intended use.

Fill in your answer here

Format | **B** | *I* | U | x_2 | x^2 | I_x |  |  |  |  |  |  |  |  |  | Σ | ABC | 

Words: 0/250












Maximum marks: 4

19 Standard_Setting_01_Long

You have been asked to assist a group of driving instructors to find the appropriate cut-score for a theoretical driving test. 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, a regulatory body has provided a statement describing what is an acceptable level of theoretical driving knowledge.

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

Fill in your answer here

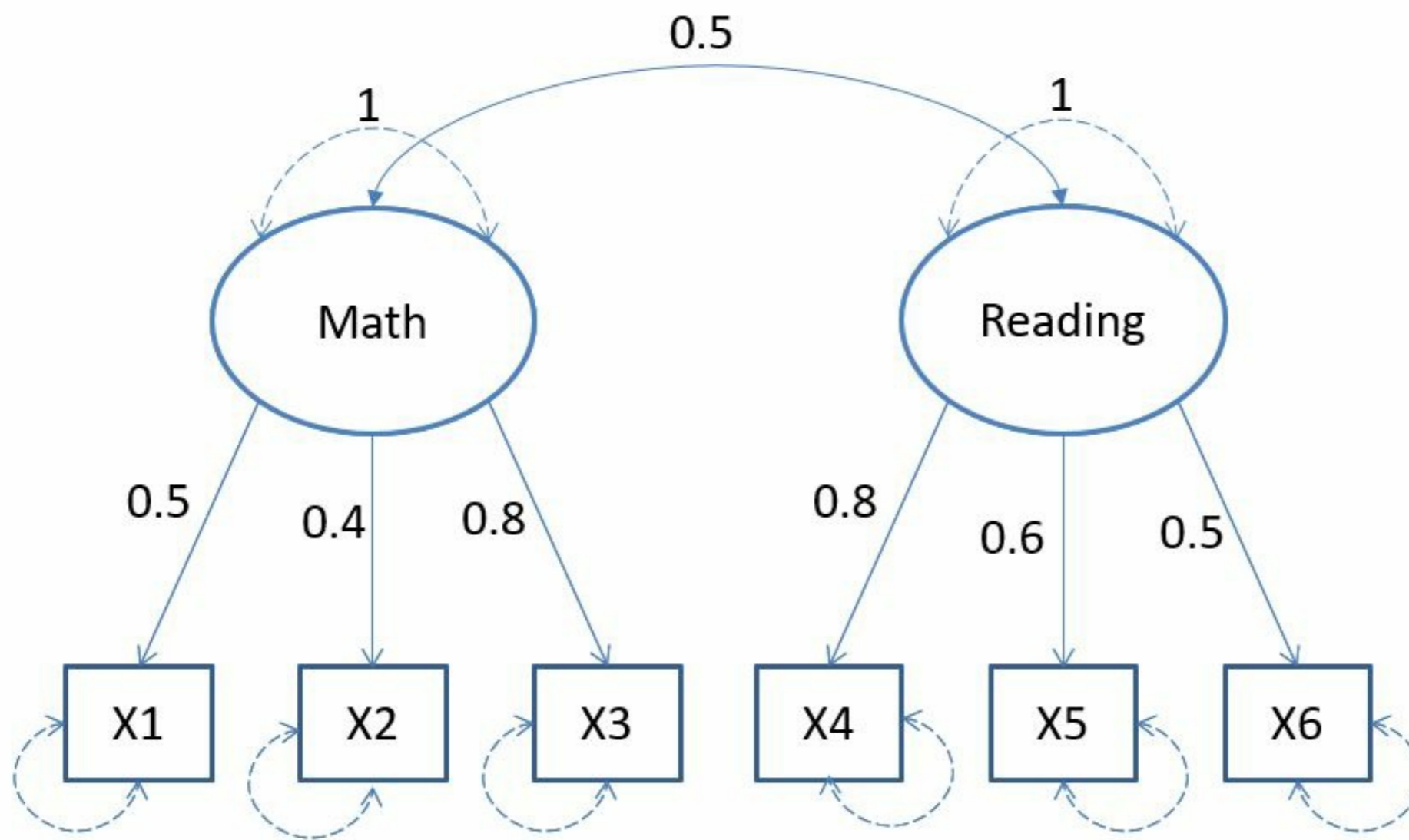
Format | **B** | *I* | U | x_2 | x^2 | I_x |  |  |  |  |  |  |  |  |  |  | ABC | 

Words: 0/200

Maximum marks: 4

20 **Factor model 4**

Consider the graph below and answer the following questions.



a) What are the factors (i.e. latent variables) in the model? (1p)

Fill in your answer here

b) Is the model identifiable from the data as illustrated in the graph? Motivate your answer. (1p)

Enter text here

c) What is the equation which describes the model for the item score X1? Write down the equation with an explanation of the parameters and variables included. (1p)

Enter text here

d) What is the covariance between item scores X1 and X4 according to the model? (1p)

Enter text here

Maximum marks: 4

21 **Factor model 6**

The following output was obtained from estimating a single factor model of five Likert-scale items from an attitude scale. The items consisted of a statement that the respondents were asked to indicate their agreement with. There were four response options: Strongly disagree, Disagree, Agree, Strongly agree.

	Factor loading (λ_j)	Error variance (Ψ_j^2)
Item 1	1.00	6.00
Item 2	2.00	3.00
Item 3	1.00	5.00
Item 4	1.00	10.00
Item 5	3.00	12.00

The model gave GFI = 0.95, RMSEA = 0.11 and a standard deviation of the residuals of 0.04.

a) Does the model fit the data well? Justify your answer. (1p)

Fill in your answer here

b) The coefficient omega is defined as:

$$\omega = \frac{\left(\sum_{j=1}^m \lambda_j\right)^2}{\left(\sum_{j=1}^m \lambda_j\right)^2 + \sum_{j=1}^m \Psi_j^2}$$

Calculate coefficient omega from the output provided. Comment on the reliability of the sum score in light of the result and provide a motivation for your answer. (1p)

Enter text here

c) Assume that a single factor model is appropriate. Which item is contributing the most to the reliability of the sum score and which item is contributing the least? Justify your answers. (1p)

Enter text here

d) From the description of the items above, give one criticism of applying the common factor model in this circumstance. (1p)

Enter text here

Maximum marks: 4

