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 to answer clarifying questions.

Good luck!

/ Björn

1 Which of the following is an estimate?
Select one alternative:

- The sample covariance.
- Coefficient alpha.
- The MLE of the factor loading for an item score in a single factor model.
- The observed sample variance of the sum score, equal to 3.4.

Maximum marks: 2

2 The covariance for two random variables X and Y is $\mathbf{Cov}(X, Y) = 2$. What is $\mathbf{Cov}(2X, 0.5Y)$?
Select one alternative:

- 2
- $\sqrt{2}$
- 4
- 1

Maximum marks: 2

3 Which of the following statements is always true for an item score that conforms to the single factor model?

Select one alternative:

- Higher factor score variance implies higher reliability of the item score.
- The reliability is higher when the item score has medium difficulty.
- The factor score variance is higher than the error score variance.
- The factor score is uncorrelated with the item score.

Maximum marks: 2

4 A single factor model was appropriate for a number of item scores. Which of the following statements is a consequence of this?

Select one alternative:

- Coefficient alpha is equal to or higher than the reliability coefficient.
- The reliability of the sum score can be expressed as a function of the factor loadings and the sum score variance.
- The covariances between the item scores are all equal.
- The reliability of the sum score is at an acceptable level.

Maximum marks: 2

5 In which of the following settings is equating necessary?

Select one alternative:

- In a low-stakes reading exam, with purpose to assess the performance of different regions in a country.
- In a high-stakes graduate school entrance exam meant to identify the top 10% most suitable students graduating from college that year.
- When using a scale to evaluate the level of depression severity among the teenage population.
- When using two test forms of a cognitive scale different time points to measure cognitive decline over time in a group of old persons.

Maximum marks: 2

6

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:

- The type of evidence required for validation depends on whether it is predictive or construct validity that is to be established.
- When a test score is used in multiple ways, each of these uses must be validated.
- Criterion-related evidence is always the gold standard validity evidence.
- All evidence categories are required when conducting a validation study.

Maximum marks: 2

7 Assume that a single factor model is appropriate for each of two groups (Grade 5 students and Grade 6 students) for the analysis of several items. For a particular item, which of the following is evidence of item bias?

Select one alternative:

- The error variance of the item score is different between the groups.
- The variance of the item score is different between the groups.
- The mean of the item score is different between the groups.
- The difficulty parameter μ_j is different between the groups.

Maximum marks: 2

8 You want to shorten a scale from ten items to five while maintaining high reliability. A single factor model fits the data well. How should you select the five items?

Select one alternative:

- Select items with low item score variance.
- Select items with high information.
- Select items with medium difficulty.
- Select items with high item score variance.

Maximum marks: 2

9

A scale to measure interest in mathematics was developed and data were collected from a large group of students, along with the scores of existing tests of mathematics proficiency and reading comprehension.

You observed the following covariance matrix for the scores of the two scales, where X denotes the interest in mathematics scale scores, Y denotes the mathematics proficiency test scores, and Z denotes the reading comprehension test scores:

$$\mathbf{\Sigma} = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) & \mathrm{Cov}(X,Z) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) & \mathrm{Cov}(Y,Z) \\ \mathrm{Cov}(Z,X) & \mathrm{Cov}(Z,Y) & \mathrm{Var}(Z) \end{pmatrix} = \begin{pmatrix} 10 & 2 & 5 \\ 2 & 10 & 8 \\ 5 & 8 & 10 \end{pmatrix}.$$

Based on these observations, how would you characterize the relationship between interest in mathematics and mathematics proficiency, and the relationship between mathematics proficiency and reading comprehension?

State the assumptions made in the interpretations of the relationships.

Fill in your answer here

Format

Words: 0/200

Maximum marks: 3

10

Let m be the number of items on a test. For a ten-item test, the factor loading estimate was 2 and the observed sample variance of the sum score Y was 50. Compute the estimate of coefficient alpha

$$\hat{\alpha} = m\frac{\hat{\lambda}}{\hat{\sigma}_Y^2}$$

and interpret the quantity. State the assumptions underlying your interpretation.

Fill in your answer here

Format

Words: 0/150

Maximum marks: 2

11 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 numeracy test is administered for the purpose of measuring Norwegian grade 4 students ability to use numeric skills in everyday contexts. The test is a low-stakes test for the respondents since individual assessment is not of interest but the test scores are used by municipalities to evaluate their teaching practices.

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. *Note:* **Only give one example of each type.**

Fill in your answer here

Format

Words: 0/150

Maximum marks: 2

12 For two tests of English reading skills, X and Y, the linear equating function was estimated to be $\mathrm{eq}(Y) = 0.8X + 4$. The cut score for passing test Y was determined to be 20.

a) Present the cut score for pass in terms of the test X scores, based on the estimated equating function. Explain how the result was obtained.

b) A student obtained a score of 20 on test X. Does the student receive a pass or fail? Explain why or why not.

Fill in your answer here

Format B I U x₂ x² Tx Ω Σ

Words: 0/200

Maximum marks: 3

13 A six-item scale measuring social interaction and a five-item scale measuring neighbourhood satisfaction were given to a large random sample of people living in central Oslo.

A two-factor independent-clusters model with correlated factors was estimated to the observed item scores, giving an RMSEA of 0.04 and a SRMSR of 0.05. The estimated correlation between the two factors was 0.47 while the estimated correlation between the sum scores was 0.30.

Structure your answer from a) to d) as follows:

a) Give an overall judgement of the fit of the estimated model based on the reported fit statistics.

b) Outline one further aspect regarding the model fit that you would like to assess.

c) Explain why the estimated sum score correlation is lower than the estimated factor correlation.

d) State which of the two correlations (sum score correlation and factor correlation) best represent the relationship between social interaction and neighbourhood satisfaction.

Fill in your answer here

Format

Words: 0/350

Maximum marks: 4

14 A scale is being developed to measure social participation with the intended purpose to use the scale in national survey to identify which factors are associated with social participation in the population. The scale consists of Likert items. According to the underlying theory of social participation, it is a unidimensional attribute. The theory also states that social participation is expected to have differences based on home ownership status but that no gender differences are expected.

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

Format

Words: 0/500

Maximum marks: 6

15 A bifactor model with one general factor and two subfactors (all factors independent) was estimated for six subtests of a Natural science test with two subdomains (Chemistry and Physics). The model fit was excellent and yielded the following factor loading estimates:

Subtest	General	Chemistry	Physics
1	4	2	0
2	2	1	0
3	3	0	1
4	2	0	1
5	2	0	2
6	1	0	1

In a previous study, the sum score of all items was used to assess Natural science competency of the respondents.

a) Based on the estimated factor loadings, would you recommend using the sum score of items? Justify your answer.

b) Describe an alternative method for scoring (not based on the sum score of all items) and provide **one benefit** of such a scoring method and **one downside** of such a scoring method.

Fill in your answer here

Format

Words: 0/350

Maximum marks: 4

16 You have been asked to assist the test development process to find an appropriate cut-score for the citizenship test in Norway. 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, you have access to detailed documentation regarding what is considered required knowledge to become a Norwegian citizen.

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

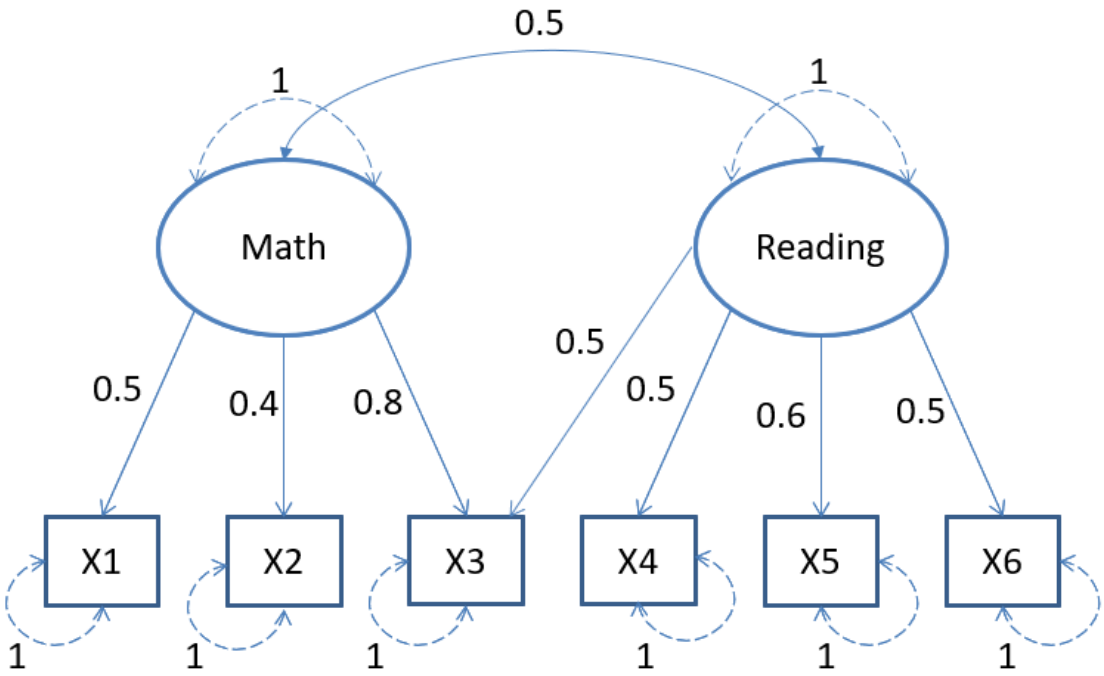
Fill in your answer here

Format

Words: 0/350

Maximum marks: 4

17 Consider the graph below and answer the following questions.



a) What is the equation which describes the model for the item score X6? Write down the equation with an explanation of the parameters and variables included. (2p)

Enter text here

b) What is the covariance between item scores X3 and X4 according to the model? (2p)

Enter text here

c) What is the reliability coefficient for item score X2? (2p)

Fill in your answer here

Maximum marks: 6