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 Consider a single factor model $X_j = \mu_j + \lambda_j F + \epsilon_j$. Which of the following is **not** a parameter?
Select one alternative:

- ϵ_j
- Ψ_j^2
- λ_j
- μ_j

Maximum marks: 2

2 The expected value of a random variable X is 5 and the expected value of a random variable Y is 3. What is the expected value of the random variable $Z = 4X + 5Y + 3$?

Select one alternative:

- 38
- 28
- 42
- 57

Maximum marks: 2

3 Which of the following statements is always true for a factor model?

Select one alternative:

- Residual variance is a reflection of systematic error.
- Factor loadings represent the sensitivity of items to measure the underlying construct.
- Factor scores are caused by item responses.
- Observed residual correlations are always zero.

Maximum marks: 2

4 For which of the following settings is equating necessary?

Select one alternative:

- When the same reading proficiency test is given at the same time to students with a learning disorder and students without the disorder, and the scores should be comparable between the groups.
- In a testing program designed to identify students at risk for dyslexia (defined as the 10% lowest performers in reading proficiency each year), where different tests are given each year.
- In an intervention study where repeated measurements of the same reading proficiency test is used to infer changes from pre-test to post-test.
- In an intervention study where changes in reading proficiency from pre-test to post-test are inferred from scores on different tests at each administration to avoid carry-over effects.

Maximum marks: 2

5 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:

- Content validity is always the first evidence category to consider in validation.
- All evidence categories are of equal importance, regardless of intended test score use.
- Intended use of test scores dictates what validity evidence is needed.
- Once a test has been validated, it may be used in any number of contexts.

Maximum marks: 2

6 A reading test was given to a large sample of students in grade 5 and grade 6. The classical true score model holds in each of the grades. The reliability coefficient for the sum score was estimated to be 0.9 in grade 5 and 0.8 in grade 6. What can you conclude from this?

Select one alternative:

- Either the error variance is higher for grade 6 students or the true score variance is lower for grade 6 students.
- The reading proficiency of grade 6 students is much higher than grade 5 students.
- The test has content better suited for grade 5 students.
- The test is biased against grade 6 students.

Maximum marks: 2

7 The following covariance matrix was observed for two variables X and Y:

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

What is the estimated correlation between X and Y?

Select one alternative:

- 0.05
- 0.25
- 0.75
- 0.5

Maximum marks: 2

8 Consider an item score for which a single factor model is appropriate for both grade 8 and grade 10 students. What would be an indicator of item bias?

Select one alternative:

- The variance of the factor score differs between grades.
- The variance of the item score is different between grades.
- The factor mean is the same in both grades, whereas the mean item score differs.
- The mean item score differs between grades.

Maximum marks: 2

9

A test measuring word reading (X) and a test measuring spelling (Y) were administered to a large random sample as part of a study of literacy in fifth graders. The estimated coefficient alpha was 0.75 for each of the tests. The observed correlation coefficient between the sum score of the two tests was 0.6 while the adjusted correlation obtained from $\frac{\hat{\rho}_{X,Y}}{\sqrt{\hat{\alpha}_X \hat{\alpha}_Y}}$ was 0.8.

Based on the above information, discuss the relationship between the constructs word reading and spelling in the population of interest. State the assumptions made in the interpretations of the relationship.

Fill in your answer here

Format

Words: 0/200

Maximum marks: 2

10 An algebra test had a standard error of measurement of 3. A respondent of the test received an observed score of 20. Estimate an approximate 95% confidence interval for the true score of this respondent. Interpret the obtained confidence interval and state the assumptions underlying the procedure used.

Fill in your answer here

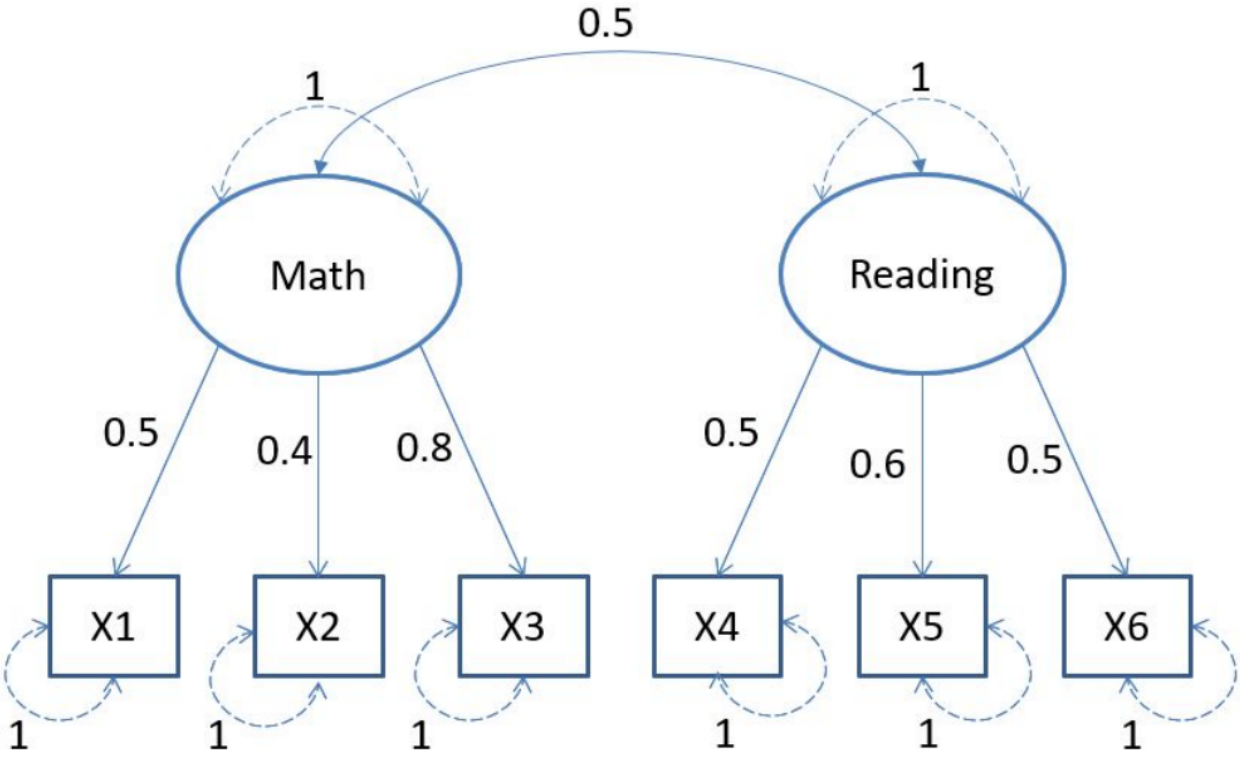
Format

Words: 0/200

Maximum marks: 2

11

Consider the factor model that is represented in the following graph.



The correlation between two item scores X_j and X_k is $\frac{\text{Cov}(X_j,X_k)}{\sqrt{\sigma^2_{X_j}\sigma^2_{X_k}}}$. Compute the correlation between X_1 and X_3 and the correlation between X_1 and X_4 with information from the graph. Show how you arrived at the answers.

Fill in your answer here

Words: 0/200

Maximum marks: 2

12 For a random variable Y, $E(Y) = 3$ and $E(Y^2) = 12$. What is the variance of Y? Show how you arrived at the answer. *Hint:* The variance of a random variable X is defined as $E[(X - E(X))^2]$.

Fill in your answer here

Format

Words: 0/200

Maximum marks: 2

13 As part of an intervention study, a vocabulary test was given to a group of students at two time points before the intervention and at two time points afterwards. Using test scores from the first two time points, the researchers estimated the test-retest correlation coefficient. Explain under which conditions this approach is appropriate for estimating the reliability of the sum scores and explain under what conditions it provides a valid reliability coefficient for the post-intervention time point.

Fill in your answer here

Format

Words: 0/200

Maximum marks: 2

14 You have been asked to assist a group of teachers of Norwegian as a foreign language to develop performance descriptors of scores on a test of Norwegian proficiency. The test consists of binary-scored items.

As part of the process, the test was piloted with a representative sample of the intended population and the results for each item score is available to you. In addition, an established framework describes the content of each of the items in the test.

Outline how a scale anchoring procedure can be used to find performance descriptors for low, medium, and high-proficiency in Norwegian.

Fill in your answer here

Format

Words: 0/400

Maximum marks: 4

15 The following summary statistics were obtained for two tests of mathematics proficiency (X and Y) in two groups (girls and boys).

	Girls	Boys
$\hat{\alpha}_X$	0.75	0.78
$\hat{\alpha}_Y$	0.72	0.74
$\hat{\mu}_X$	30	27
$\hat{\mu}_Y$	26	23
$\hat{\sigma}_X$	11	9
$\hat{\sigma}_Y$	11	9

The linear equating function is estimated by $eq(Y) = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y} X + \left(\hat{\mu}_X - \frac{\hat{\sigma}_X}{\hat{\sigma}_Y}\hat{\mu}_Y\right)$. Evaluate the usage of linear equating in terms of the equating criteria: equal reliability, equating symmetry and population invariance (consider Girls and Boys as different subpopulations).

Fill in your answer here

Format | B I U x_2 x^2 I_x | Σ

Words: 0/400

Maximum marks: 4

16 A test battery is being developed to measure morphological awareness, morphological analysis and morphological decoding with the intended purpose to use the battery to measure the development of morphological knowledge in Norwegian children through the primary school years. The battery consists of three sub-tests reflecting morphological awareness, morphological analysis and morphological decoding, each with 20 items scored 0-2.

According to theory, morphological awareness, analysis and decoding are separate, but related, dimensions of morphological knowledge. The theory also states that morphological analysis is positively related to general vocabulary, whereas morphological decoding is positively related to word reading.

Based on the above, structure your answer into the following three parts:

1. Describe what evidence sources you want to consider in order to evaluate the validity of the scale scores for their intended purpose.
2. Give an overview of how you would design a study to evaluate the validity of the scale score use. The answer should include a description of the data you would like to analyze and the analysis methods you would suggest using.
3. Outline what results you would consider as evidence supporting the validity of using the scale scores for their intended purpose.

Fill in your answer here

Format

Words: 0/600

Maximum marks: 6

17 Consider a test that serves as part of the certification for nurses. Obtaining at least 70% correct answers on the test is one requirement to obtain a nursing license. A study was done to evaluate the association between the scores of the test and future job performance as measured by a standardized assessment of job performance for nurses. The sample consisted of a large random sample of the nurses that graduated from Norwegian universities in one year. The participation rate was 76%.

A correlation analysis was done with job performance scores and exam test scores. The estimated Pearson correlation coefficient was 0.50.

What is your assessment of the validity of using the exam scores as part of the nurse licensing process in light of these results? *Hint:* Consider how the outcome used (job performance measure) is defined, how the study is designed and what level of evidence the study can be considered to provide for the intended use.

Fill in your answer here

Format

Words: 0/400

Maximum marks: 4

18 A scale consisting of six items meant to measure bullying in school was administered to a large sample of students at Oslo high schools. The scale consisted of questions concerning six statements with four categories to choose from for each statement: Never or almost never, A few times a year, A few times a month, and Once a week or more. The details of the scale are given below:

During the past 12 months, how often have you had the following experiences in school?

1. Other students left me out of things on purpose.
2. Other students made fun of me.
3. I was threatened by other students.
4. Other students took away or destroyed things that belonged to me.
5. I got hit or pushed around by other students.
6. Other students spread nasty rumours about me.

The proposed use of the scale was to monitor the level of bullying in Oslo schools to identify schools which were in need of support measures. However, a shortened scale of three items was desired in order to reduce the testing time.

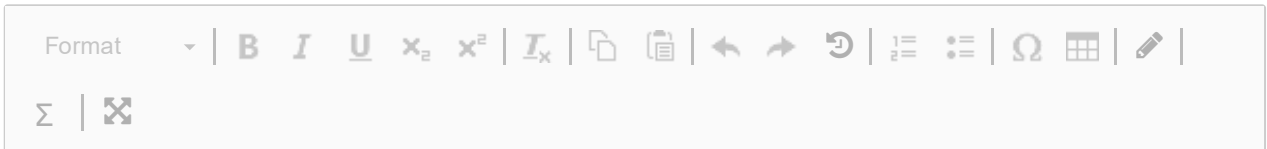
A single factor model was fitted (GFI = 0.96, RMSEA = 0.04 and SRMR = 0.05) and the estimated parameters were:

Item	Factor loading	Error variance
1	2.0	1.0
2	1.0	3.0
3	1.0	1.0
4	1.0	1.0
5	2.0	1.0
6	3.0	2.0

Structure your answer into three parts to address the following in your response:

1. Consider the respondents and the content-related validity of the scale (do not consider the estimated model). Give **one possible benefit** of shortening the scale to three items and give **one possible drawback** of shortening the scale to three items. (2p)
2. Construct a three-item version of the and describe your reasoning process for the construction. Estimate the reliability of the sum score of the three-item scale that you constructed. (2p)
3. Consider the intended usage of the scale scores and make an informed judgement of the implications of shortening the scale to the three items that you selected. (2p)

Fill in your answer here



Words: 0/600

Maximum marks: 6