ⁱ Exam information

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.

Good luck!

/ Björn

¹ RV1

Which of the following is a random variable?
Select one alternative:

○ The observed correlation between scores of the same test given to the same group of people at two time points, equal to 0.7.

○ The maximum likelihood estimator of coefficient omega.

○ $\omega = \frac{(\sum_{j=1}^{J} \lambda_j)^2}{\sigma_Y^2}$

○ The observed coefficient omega equal to 0.8, obtained from estimated factor model parameters.

Maximum marks: 2

2 RV2

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

○ 18

○ 71

○ 23

○ 66

Maximum marks: 2

3 Factor model1

Which of the following statements is always true for a single factor model?
Select one alternative:

○ Some, but not all, of the residual correlations between the observed variables are zero.

○ The variance of the sum score is equal to the sum of the variance of each observed variable.

○ The true score variance for each item must be larger than the error variance for that item.

○ The variance of the sum score can be decomposed into two parts: one construct-relevant part and one construct-irrelevant part.

Maximum marks: 2

4 Equating

For which of the following settings is equating necessary?
Select one alternative:

○ The same reading proficiency test is given at the same time to students in two different grades, and the scores should be comparable between the grades.

○ In a longitudinal study where repeated measurements of the same reading proficiency test is used to infer the trend of reading proficiency across multiple grades.

○ In a national study of trends in reading proficiency, conducted with a new sample every two years where reading proficiency is measured with different tests at each administration.

○ In a testing program designed to identify the top 5% students in reading proficiency each year, where different tests are given each year.

Maximum marks: 2

5 Validity1

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 evidence categories are of equal importance for any given test score use.

○ Validation comprises construct validity, content validity and predictive validity.

○ The validity evidence to be considered depends on the specific use of the test scores.

○ Construct validity is equivalent to evidence based on internal structure.

Maximum marks: 2

6 Factor model2

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

Select one alternative:

○ The test has content better suited for grade 6 students.

○ The reading proficiency of grade 6 students is higher.

○ The test is biased against grade 5 students.

○ Either the error variance is lower for grade 6 students or the true score variance is higher for grade 6 students.

Maximum marks: 2

7 Correlation

The following covariance matrix was observed for two variables X and Y
$$\begin{pmatrix} 9 & 3 \\ 3 & 16 \end{pmatrix}.$$
What is the estimated correlation between the variables?

Select one alternative:

○ 0.12

○ 0.25

○ Not enough information is available.

○ 0.5

Maximum marks: 2

8 Item bias

Consider an item score for which a single factor model is appropriate for both men and women. What would be an indicator of item bias?

Select one alternative:

○ The variance of the factor score is different for men and women.

○ The mean item score is different for men and women while the factor mean is the same for men and women.

○ The variance of the item score is different for men and women.

○ The mean item score is different for men and women.

Maximum marks: 2

9 Relation

A test measuring satisfaction with life and a test measuring depression severity were administered to a large random sample as part of a national survey. The estimated coefficient alpha was 0.5 for each of the tests. The observed correlation coefficient between the sum score of the two tests was -0.4 while the adjusted correlation obtained from $\frac{\text{Cor}(X,Y)}{\sqrt{\alpha_X \alpha_Y}}$ was -0.8.

Based on the above information, discuss the relationship between the constructs satisfaction with life and depression severity in the population of interest.

Fill in your answer here

Format ⌄	**B** *I* U x₂ x² I̶ₓ	⎙ ⎘	↩ ↪ ⟲	1≡ ≔	Ω ⊞	✎	Σ

Words: 0/200

Maximum marks: 2

10 SEM

A reading literacy test had a standard error of measurement of 2. A respondent of the test received an observed score of 10. 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 | B | I | U | x₂ | x² | Iₓ | ⬚ | ⬚ | ↩ | ↪ | ⟳ | ⬚ | ⬚ | Ω | ⊞ | ✎ |

Σ | ⤢

Words: 0/200

Maximum marks: 2

11 Model-implied

Consider the factor model that is represented in the following graph.
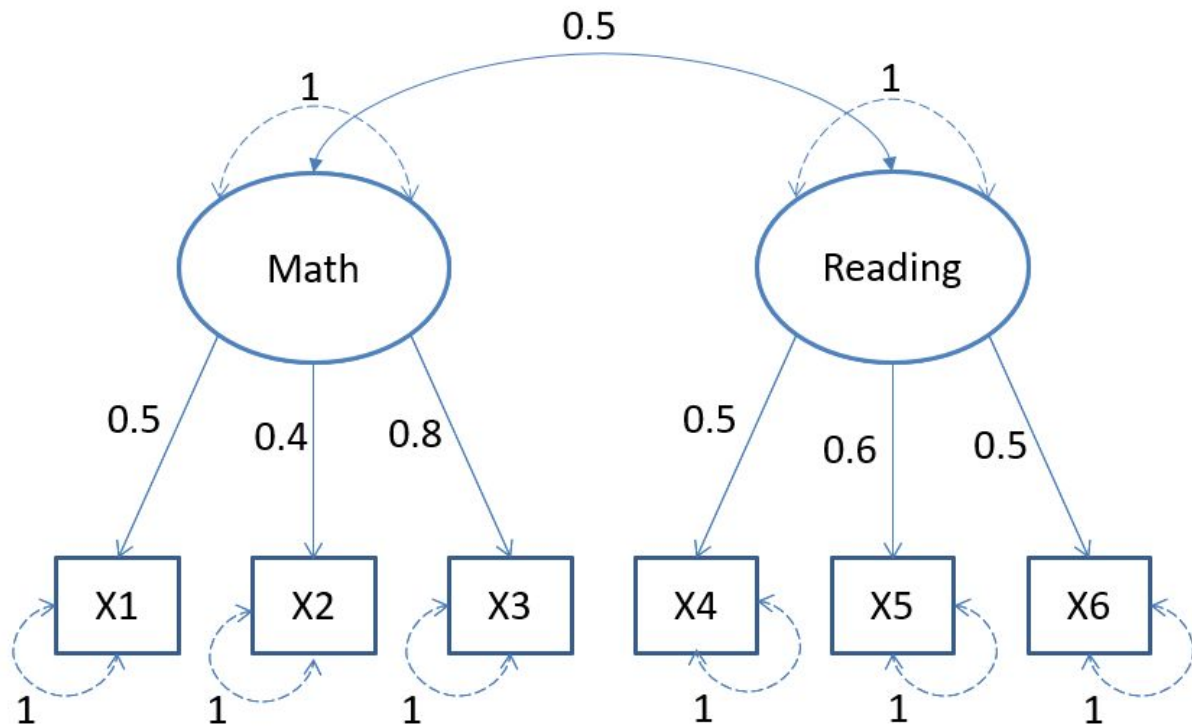
Compute the model-implied covariances for X_2, X_3 and X_4 (that is, the upper diagonal entries of the covariance matrix for X_2, X_3 and X_4 based on the model parameters).

Fill in your answer here

Format ▾ | B *I* U x₂ x² | Iₓ | ⎙ ⎙ | ↶ ↷ ⟳ | ≔ ≔ | Ω ⊞ | ✎ |

Σ | ⤧

Words: 0/200

12 Variance

For a random variable Y, $E(Y) = 2$ and $E(Y^2) = 6$. What is the variance of Y? *Hint:* The variance of a random variable X is defined as $E[(X - E(X))^2]$.

Fill in your answer here

Words: 0/200

13 Test-retest

The same environmental awareness test was given to a group of students at two time points. The reliability of the sum scores was estimated with the test-retest correlation coefficient. Explain under which conditions this an appropriate method for estimating the reliability of the sum scores.

Fill in your answer here

Words: 0/200

Maximum marks: 2

14 Alpha

Coefficient alpha is defined as $\frac{m}{m-1}\left(1 - \frac{\sum_{j=1}^{m}\sigma^2_{X_j}}{\sigma^2_Y}\right)$, where m is the number of items. A single factor model with a standardized factor was estimated for three item scores X_1, X_2 and X_3. The model fitted well and yielded factor loading estimates $\hat{\lambda}_1 = \hat{\lambda}_2 = \hat{\lambda}_3 = 1$ and error variance estimates $\hat{\Psi}^2_1 = 1, \hat{\Psi}^2_1 = 0.5$ and $\hat{\Psi}^2_1 = 1.5$. The variance of the sum score was 12. Present the estimated coefficient alpha and interpret it.

Fill in your answer here

| Format ⌄ | B | I | U | x₂ | x² | I_x | ⧉ | ⧉ | ⬅ | ➡ | ⟳ | ⦂≡ | ⦂≡ | Ω | ⊞ | ✎ |

Σ | ⤢

Words: 0/200

Maximum marks: 2

15 Equating

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

	Girls	Boys
$\hat{\alpha}_X$	0.80	0.85
$\hat{\alpha}_Y$	0.84	0.89
$\hat{\mu}_X$	24	20
$\hat{\mu}_Y$	28	24
$\hat{\sigma}_X$	8	10
$\hat{\sigma}_Y$	8	10

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

Fill in your answer here

Format ▾ | **B** *I* <u>U</u> x₂ x² | Iₓ | ▢ ▢ | ↩ ↪ ↺ | ≔ ≔ | Ω ▦ | ✎ |

Σ | ⤢

Words: 0/400

Maximum marks: 4

16 Evidence sources and data

A scale is being developed to measure depression severity with the intended purpose to use the scale in a longitudinal national survey to estimate the trend of depression severity in the country. The scale consists of Likert items with seven categories. According to the underlying theory of depression severity, it is a unidimensional attribute. The theory also states that depression severity is positively associated with self-harm and negatively associated with general quality of life.

Under these conditions:

- Describe what evidence sources you want to consider in order to evaluate the validity of the scale scores for their intended purpose and describe the data you would like to analyze in doing so.
- 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

Words: 0/400

Maximum marks: 4

17 Validity of exam scores

Consider a test graded as pass/fail that serves as part of the certification for high school teachers. Obtaining a passing grade on the test is one requirement to obtain a teaching license. A study was done to evaluate the association between the scores of the certification exam and future job performance as measured by a standardized teacher quality assessment. A regression analysis was done with job performance scores as the dependent variable and exam test scores as the independent variable. The estimated standardized regression coefficient for the exam test scores was 0.30.

What is your assessment of the validity of using the exam scores as part of the teaching license process in light of these results? In your answer, you should consider how the outcome used (job performance measure) is defined and what level of evidence the study can be considered to provide for the intended use.

Fill in your answer here

Format ⌄	B	I	U	x₂	x²	Iₓ							Ω		

Words: 0/400

Maximum marks: 4

A factor model was estimated for four item scores X_1, X_2, X_3 and X_4. The equations for item scores X_1 and X_4 were:

$$X_1 = \mu_1 + \lambda_1 F_1 + \epsilon_1$$

and

$$X_4 = \mu_4 + \lambda_4 F_2 + \epsilon_4,$$

where $\mu_1 = 3$, $\mu_4 = 4$, $\lambda_1 = 2$, $\lambda_4 = 1$, $\mathrm{Var}(\epsilon_1) = 2$, $\mathrm{Var}(\epsilon_4) = 1.5$, $E(F_1) = E(F_2) = 0$, $\mathrm{Var}(F_1) = \mathrm{Var}(F_2) = 1$ and $\mathrm{Cov}(F_1, F_2) = 0.5$. The error terms ϵ_1 and ϵ_4 had means equal to zero and were independent of each other and of F_1 and F_2.

18 FM3a

What is $E(X_1 + X_4)$? Display and explain the steps used to compute the quantity.

Fill in your answer here

| Format ▾ | **B** *I* U̲ x₂ x² I͟ₓ ⌷ ⌷ ↰ ↱ ⟲ ⁞≡ ⁞≡ Ω ⊞ ✎ |
| Σ ⤢ |

Words: 0/200

Maximum marks: 1

19 FM3b

What is $\mathrm{Var}(X_1 + X_4)$? Display and explain the steps used to compute the quantity.

Fill in your answer here

Words: 0/200

Maximum marks: 1

20 FM3c

What is $\mathrm{Cov}(X_1, X_4)$? Display and explain the steps used to compute the quantity.

Fill in your answer here

| Format ▾ | B | I | U | x_2 | x^2 | $\mathcal{I}_x$ | ⎘ | 📋 | ← | → | ⟲ | 1≣ | ≔ | Ω | ⊞ | ✎ |
| Σ | ⤢ |

Words: 0/200

Maximum marks: 1

21 **FM3d**

Which item score, X_1 or X_4, has a higher precision of measurement? Also explain why this is. Display and explain the steps used.

Fill in your answer here

| Format ▾ | B | I | U | x_2 | x^2 | I_x | ⎙ | 📋 | ↰ | ↱ | ⟲ | ⅓≡ | ☰ | Ω | ⊞ | ✎ |

| Σ | ⤢ |

Words: 0/200

Maximum marks: 1

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	$\hat{\lambda}$	$\hat{\Psi}^2$
1	1.0	1.0
2	2.0	1.0
3	3.0	2.0
4	2.0	1.0
5	1.0	3.0
6	1.0	1.0

22 ValRela

Consider

- the administration setting for the scale
- the properties of the scale from the perspective of content

That is, do not consider the estimated model here.

Give **one possible positive aspect** of shortening the scale to three items and give **one possible negative aspect** of shortening the scale to three items.

Fill in your answer here

Words: 0/200

Maximum marks: 2

23 **ValRelb**

Construct a three-item version of the scale based on the estimated model parameters. Describe your reasoning process for the construction. Estimate the reliability of the sum score of the three-item scale that you constructed.

Fill in your answer here

| Format ▼ | B | I | U | X₂ | X² | Iₓ | ⧉ ▤ | ← → ⟲ | ≔ ⋮≡ | Ω ▦ | ✎ |
| Σ | ⛶ | | | | | | | | | | |

Words: 0/200

Maximum marks: 2

24 ValRelc

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.

Fill in your answer here

Format · | B *I* U x₂ x² | | | | | | | | | Ω | |

Σ |

Words: 0/200

Maximum marks: 2