# Draft of MAE4011 H22 Exam

## Part 1

### MC1

Which of the following is an estimate?

* The factor loading.
* The sample variance.
* The factor score.
* The observed sample mean.

### MC2

The variance of a random variable X is 2. What is the variance of the random variable Y = 2X + 1?

* 4
* 5
* 8
* 7

### MC3

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

* A lower difficulty implies higher reliability.
* A lower error score variance means lower reliability.
* The factor score variance is always larger than the error score variance.
* The error score is uncorrelated with the observed score.

### MC4

Assume that a single factor model is appropriate for a number of item scores. Which of the following statements is a consequence of this?

* The covariances between all item scores must be equal.
* Coefficient alpha is equal to the reliability of the sum scores.
* The reliability of the sum score can be decomposed into the function of the factor loadings and the error variances.
* The reliability of the sum score is above 0.8.

### MC5

Consider the factor model: , where . Consider an item with , and . What is the variance of ?

* 0
* 1/4
* 1/2
* 3/4

### MC6

Assume that a single factor model is appropriate for the analysis of several items. For a particular item, which of the following is evidence of item bias?

* The mean of the item score is higher for women than it is for men.
* The variance of the item score is higher for women than it is for men.
* The reliability of the item score is higher for women than it is for men.
* For men and women with the same factor score, the mean of the item score is higher for women than it is for men.

### MC7

For which of the following settings is equating necessary?

* When using a cognitive scale to screen for dementia in two groups defined by their gender identity.
* When using a scale to evaluate the level of anxiety among the teenage population.
* In a low-stakes national reading exam, with purpose to assess the performance at the regional level over time.
* In a high-stakes college entrance exam meant to identify the top 25% most suitable students graduating from high school that year.

### MC8

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

* The type of evidence required for validation depends on whether it is content, criterion, or construct validity that is to be established.
* When carrying out a validation study, one must consider all evidence categories.
* Evidence with regard to an external diagnostic criterion is always the gold standard validity evidence.
* When a test score is used in multiple ways, each of these uses must be validated.

## Part 2

### SR1

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 denotes the depression severity scale scores and denotes the satisfaction with life scale scores:

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.

### SR2

An item score that could take integer values betwee 0 and 2 had the following probability distribution in a population of students:

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 0.2 | 0.6 | 0.2 |

What is

a) the mean of ,

b) the mode of ,

c) the median of , and

d) the variance of ?

Only the answers are required.

### SR3

and are two random variables with , , and .

1. Calculate where . Show your work.
2. Calculate where . Show your work.

Credits are awarded to each step. Lone result with no working steps attracts no credit.

### SR4

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

and interpret it. State the assumptions underlying the interpretation.

### SR5

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) not enough.

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.

### SR6

For two tests of reading comprehension, and , the linear equating function was estimated to be . The cut score for passing test was determined to be 30.

Task: Based on the estimated equating function above, compute the cut score for passing test . Show your calculation steps.

### SR7

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, with model fit indices:

GFI = 0.95, RMSEA = 0.05 and 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.

### SR8

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 acceptable.

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