

## i Instructions

Dear student,

Some guidelines for the MAE4011 exam:

- The exam consists of a combination of multiple choice questions and open-response questions.
- It is totally fine to respond in bullet points, as long as no important information gets lost.
- For the computational tasks, it is sufficient to show how you *would* obtain the respective coefficients. You will find formulas in the provided external resource in the left corner.
- Read the questions carefully and write down if you do not understand something about a question. I will visit the exam site for during the first two hours of the exam to answer clarifying questions

Good luck!

Esther

## 1 Classical Test Theory I

Which of the following statements are always true for the classical true score model? (1 point)

**Select one alternative:**

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

---

Maximum marks: 1

## 2 Classical Test Theory II

In classical test theory, we denote with  $X$  the observed score of a test, with  $T$  its true score and with  $E$  its error score. (4 points)

1. Derive the covariance between  $X$  and  $T$ , drawing on the axioms of classical test theory and variance and covariance algebra.
2. Use these derivations to show that the squared correlation between true and observed scores  $\left(\frac{\text{cov}(T,X)}{\text{sd}(X)\text{sd}(T)}\right)^2$ , or, equivalently  $\left(\frac{\sigma_{T,X}}{\sigma_X\sigma_T}\right)^2$  is equal to the proportion of variance in observed scores going back to true score variation

Fill in your answer here

---

Maximum marks: 4

## 3 Operationalization

A rhetoric coach is interested in studying public speaking anxiety. For measuring public speaking anxiety, she is deciding between two measures: one self-report measure and one physiological measure. As a self-report measure, the coach considers using the Personal Report of Public Speaking Anxiety (PRPSA); a questionnaire where participants rate their agreement with statements such as "I feel anxious when I have to speak in front of a group" on a 5-point Likert scale. As a physiological measure, she considers recording the participant's hear rate while delivering a short (3-5 minute) speech in front of a small audience.

Name and briefly explain one advantage and one disadvantage of each public speaking anxiety measure. (4 points)

Fill in your answer here

---

Maximum marks: 4

#### 4 Variance-Covariance

We have the following variance-covariance matrix between items  $X_1$  and  $X_2$  forming the scale  $Y = X_1 + X_2$  and items  $X_2$  and  $X_3$  forming the scale  $Z = X_2 + X_3$ . (3 points)

1. Obtain the sample variance of the scale score  $Y$ .
2. Obtain the sample covariance between the scale score  $Y$  and the scale score  $Z$ .
3. What would be the sample covariance between the scale score  $Y$  and the scale score  $Z$  if we would upweigh the contribution of  $X_2$  to the scale score, i.e., define  $Y = X_1 + 2X_2$ ?

(It is sufficient to write down how you would the respective coefficients. You do not have to provide the exact numbers.) (3 points)

|       | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|-------|-------|-------|-------|-------|
| $X_1$ | 1.00  | 0.70  | 0.30  | 0.20  |
| $X_2$ | 0.70  | 1.00  | 0.40  | 0.10  |
| $X_3$ | 0.30  | 0.40  | 1.00  | 0.80  |
| $X_4$ | 0.20  | 0.10  | 0.80  | 1.00  |

Fill in your answer here

Maximum marks: 3

#### 5 Test Scores

An e-sports team is developing a new test to evaluate players' strategic skills in order to select the 10 best participants for their trainee program. (5 points in total)

1. The team is debating whether to use norm-based or reference-based scoring for interpreting players' test results. Provide an example interpretation of both norm-based and reference-based scores in this context. Which approach to test score interpretation would be more appropriate for this test in the given context, and why? (3 points)
2. To evaluate the utility of the test, the e-sports team investigates the correlation between the selected players' strategic-skills scores and the number of wins during their first season. The team notices that the variability in strategic-skills scores among selected participants ( $\text{var} = 1.5$ ) is much lower than among the applicants ( $\text{var} = 6$ ). It also notices that the correlation between the strategic skills scores and the number of wins of their trainees is rather low ( $r = 0.20$ ). State why the team might observe a low correlation in the selected sample even if the test is a strong predictor in the applicant population, and how the team could use statistical methods to re-evaluate the predictive utility of the strategic-skills test. (2 points)

Fill in your answer here

Maximum marks: 5

## 6 Item Statistics

The table below displays item difficulties  $p$ , item total correlations  $r$ , and correlations  $c$  with a relevant criterion for 10 items. Which items would you remove if you were to construct a shorter measure with 7 items only and why? (3 points)

| Item | $p$  | $r$  | $c$  |
|------|------|------|------|
| 1    | 0.97 | 0.39 | 0.52 |
| 2    | 0.72 | 0.49 | 0.39 |
| 3    | 0.59 | 0.68 | 0.61 |
| 4    | 0.49 | 0.53 | 0.42 |
| 5    | 0.52 | 0.66 | 0.49 |
| 6    | 0.40 | 0.51 | 0.38 |
| 7    | 0.29 | 0.20 | 0.55 |
| 8    | 0.31 | 0.51 | 0.34 |
| 9    | 0.30 | 0.52 | 0.16 |
| 10   | 0.23 | 0.38 | 0.47 |

Fill in your answer here

Maximum marks: 3

## 7 Quality Criteria

According to the 2014 Standards for Educational and Psychological Testing, “predictive bias” refers to (1 point)

Select one alternative:

- ☐ The degree to which a test produces inconsistent scores across subgroups with diverse cultural, linguistic, or educational backgrounds.
- ☐ The extent to which associations between test scores and other variables differs across groups.
- ☐ A situation in which two groups obtain different mean scores on a test because their true abilities differ.
- ☐ A procedure for statistically adjusting test scores to reduce subgroup differences and improve fairness.

Maximum marks: 1

## 8 Reliability I

You have a 40-item test with an estimated reliability of .30. If you expanded it to an 60-item test by adding parallel items, what reliability would you expect? It is sufficient to write down how you would compute reliability. You do not have to provide the exact number. (1 point)

Fill in your answer here

---

Maximum marks: 1

## 9 Reliability II

You have a 50-item test with an estimated reliability of .75. How many items would you need for the test to reach an expected reliability of .90? It is sufficient to write down how you would compute the required number of items. You do not have to provide the exact number. (1 point)

Fill in your answer here

---

Maximum marks: 1

## 10 Reliability III

Which of the following statements about Cronbach's alpha is *not* true?

Select one or more alternatives:

- ☐ Cronbach's alpha underestimates reliability when the assumption of tau-equivalence is violated.
- ☐ Cronbach's alpha is used to assess the internal consistency of a scale.
- ☐ A high value of Cronbach's alpha indicates that the scale is unidimensional.
- ☐ Cronbach's alpha always increases when more items are added to a scale.

---

Maximum marks: 4

## 11 Reliability IV

A low test–retest reliability coefficient may be due to (select all that apply)

**Select one or more alternatives:**

- ☐ The construct itself shows meaningful fluctuations over the retest interval.
- ☐ The mean level of the construct differs between the two measurement occasions, even though rank-ordering remains stable.
- ☐ The correlation between the two test scores is reduced because the sample has very high true-score variability.
- ☐ The measurement instrument contains substantial error score variance.

---

Maximum marks: 2

## 12 Validity I

A student uses to google's Gemini to support her learning process and prompts it for a definition of validity in psychological and educational measurement. Gemini replies that "*Validity is the extent to which a test measures what it is supposed to measure, evaluated based on empirical evidence and theoretical rationales.*" (3 points)

1. Identify one aspect of this definition that aligns with the current understanding of validity in educational and psychological measurement as described in the 2014 Standards for Educational and Psychological Testing.
2. Identify one aspect of this definition that does not align with the current understanding of validity as described in the 2014 Standards for Educational and Psychological Testing.

**Fill in your answer here**

---

Maximum marks: 3

### 13 MTMM

|               | Peer report |            | Parent report |            |
|---------------|-------------|------------|---------------|------------|
|               | Prosocial   | Aggression | Prosocial     | Aggression |
| Peer report   |             |            |               |            |
| Prosocial     | .85         |            |               |            |
| Aggression    | -.30        | .80        |               |            |
| Parent report |             |            |               |            |
| Prosocial     | .50         | -.15       | .75           |            |
| Aggression    | -.20        | .45        | -.10          | .70        |

The table represents a multitrait-multimethod (MTMM) matrix in which prosocial behavior and aggression are assessed via peer reports and parent reports. The diagonal displays the reliability (Cronbach's alpha) of each measure; the off-diagonal cells show the correlations among measures. (6 points)

1. Discuss the convergent and discriminant validity evidence for the parent report of prosocial behavior, referring explicitly to the values in the table.
2. Compute the disattenuated (true-score) correlation between the peer report and the parent report of prosocial behavior, accounting for measurement error in both measures.

**Write your answer in the box below. Changes are saved automatically.**

Maximum marks: 6

### 14 Confirmatory Factor Analysis

Which of the following assumptions and constraints are made in a congeneric confirmatory factor analysis model?

**Select one or more alternatives:**

- ☐ The model is assumed to be unidimensional
- ☐ All item loadings are constrained to be equal
- ☐ All item intercepts are constrained to be equal
- ☐ All item error variances are constrained to be equal

Maximum marks: 1