

i Instruction

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 Validity Evidence

Which of the following are **not** sources of validity evidence listed in the Standards for Educational and Psychological Testing? (select all that apply)

- ☐ Evidence based on response processes
- ☐ Evidence based on standardized test administration procedures
- ☐ Evidence based on test content
- ☐ Evidence based on internal consistency

Maks poeng: 2

2 Opportunity to Learn

According to the 2014 Standards for Educational and Psychological Testing, "opportunity to learn" refers to:

- ☐ The time and resources allocated to ensure that all students can master the content assessed by a test.
- ☐ The extent to which an examinee has been exposed to instruction, knowledge, or experiences assumed by the test developer and/or user.
- ☐ The extent to which a test assesses a broad range of abilities that students have been taught during their education.
- ☐ The alignment between the content of a test and the specific learning outcomes of a curriculum.

Maks poeng: 1

3 Cohen's kappa

A university admissions committee wants to evaluate their rating system for motivation letters for applicants to a Master's program. Two committee members independently classify 50 applicants as either "Promising" or "Less Promising" based on their motivation letters. **(4 points)**

1. Use the data in the table below to describe how you would compute Cohen's kappa for agreement between the two committee members. You do not need to provide the exact number.
2. The resulting Cohen's kappa is rather low. Which quality criterion according to the Standards for Educational and Psychological Testing is likely violated?

		Member 2		
		Promising	Less Promising	Total 1
Member 2	Promising	17	8	25
	Less Promising	6	19	25
Total 2		23	27	50

Maks poeng: 4

4 Violation of Quality Standards

In a study evaluating a standardized mathematics test used for placement into advanced secondary-school courses, researchers examine the relationship between test scores and students' end-of-year mathematics grades. They find that the test strongly predicts grades for students who attended schools with an academic track, but shows a substantially weaker relationship for students from vocational tracks. (2 points)

- What is this type of phenomenon referred to as?
- Which quality criterion for educational and psychological tests is being violated in this scenario?

Maks poeng: 2

5 MTMM

In the context of the Multitrait-Multimethod (MTMM) matrix, which of the following provides strong discriminant validity evidence?

- ☐ High correlations between the same trait measured using different methods.
- ☐ Low correlations between the same trait measured using different methods.
- ☐ High correlations between different traits measured using different methods.
- ☐ Low correlations between different traits measured using the same method.

Maks poeng: 1

6 Threats to Validity

A researcher wants to measure anxiety during a stressful task using skin conductance (sweaty palms) as the sole indicator. **(3 points)**

- Explain how this measure could lead to (a) construct underrepresentation and (b) construct irrelevant variability in the measurement of anxiety.
- Can both threats to validity occur at the same time and why?

Maks poeng: 3

7 Validity Evidence

A test battery is being developed to measure academic motivation in secondary school students, with the intended purpose of tracking motivation development throughout the first three years of secondary school. The battery includes three multi-item sub-tests designed to measure intrinsic motivation, extrinsic motivation, and self-efficacy for learning. According to theory, intrinsic motivation, extrinsic motivation, and self-efficacy are separate but related dimensions of academic motivation. Additionally, intrinsic motivation is theorized to be positively related to class participation, while self-efficacy is positively related to homework completion and academic performance. **(6 points)**

1. Describe at least three sources of evidence you would consider in order to evaluate the validity of the scale scores for their intended purpose.
2. Give a brief overview of how you would design a study to evaluate the validity of the scale score use. Your answer should include a description of the data you would like to collect and the analysis methods you would suggest.
3. For each source of validity evidence, describe the specific empirical results you would consider as evidence supporting the validity of using the scale scores for their intended purpose.

Maks poeng: 6

8 Parallel Items

Two parallel items have a correlation of 0.40. (3 points)

- What is the reliability of a single item?
- How many of those items do you need to construct a test with a reliability of 0.80?

Skriv svaret ditt her. Endringer blir lagret automatisk.

Maks poeng: 3

9 CTT Axioms

We denote with X the observed score of a test, with T its true score and with E its error score. Which of the following does **not** follow from the axioms of classical test theory?

- ☐ $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$
- ☐ $\rho_{T,E} = 0$
- ☐ $\rho_{X,T} = 0$
- ☐ $\sigma_{XT} = \sigma_T^2$

Maks poeng: 1

10 CTT Reliability

We can use the observed correlation between scores on parallel tests $\frac{\sigma_{XX'}}{\sigma_X\sigma_{X'}} = \frac{\sigma_{XX'}}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_X^2}$ as an estimate of reliability. Why does the same not hold for tau-equivalent tests?

- ☐ Because $\sigma_{XX'} = \sigma_T^2$ does not hold when tests are tau-equivalent
- ☐ Because tau-equivalent tests do not assume equal total variance.
- ☐ Because error variances are assumed to be uncorrelated in parallel tests but not in tau-equivalent tests.
- ☐ Because tau-equivalent tests allow for different true score variances

Maks poeng: 1

11 Alpha and Omega

When do coefficient alpha and McDonald's omega yield the same conclusions regarding reliability?

- ☐ When the test is unidimensional
- ☐ When the items are parallel
- ☐ When items are essentially tau-equivalent
- ☐ When all items are equally correlated with one another

Maks poeng: 1

12 Test-Retest

A researcher is developing a new test to measure passive vocabulary knowledge of Norwegian as a foreign language. She administers the test twice to recent migrants to Norway, the first time two weeks after arrival and the second time six months later. Using the test scores from the two measurement occasions, the researcher estimates the test-retest correlation coefficient. Explain under which conditions this approach is appropriate for estimating the reliability of the sum scores and under which conditions another approach for estimating the reliability of the sum scores should be chosen. (2 points)

Maks poeng: 2

13 Model Comparison

A researcher wants to evaluate the properties of a newly developed three item measure. She compared two unidimensional confirmatory factor analysis models, one assuming a tau-equivalent model (fit.tau), and one assuming a parallel model (fit.par). The figure displays the results of a comparison between the two models. Which constraints are assumed to hold for the parallel model that do not need to hold for a tau equivalent model? Which model provides the better fit to the data? Briefly explain your reasoning. (2 points)

> anova(fit.GM.tau, fit.GM.par)

Chi-Squared Difference Test

```

              Df   AIC   BIC  Chisq Chisq diff  RMSEA Df diff Pr(>Chisq)
fit.GM.tau   4 24090 24119 201.29
fit.GM.par   6 24121 24138 236.33      35.048 0.08138      2 2.452e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Maks poeng: 2

14 Percentile Score

The table below displays transformations of the raw score Y of a standardized math test to percentile rank scores $Q(Y = y) = p(Y \leq y)$ (3 points).

1. How would you interpret a percentile rank score of 0.40 in this specific context?
2. Describe one advantage and one limitation of using percentile rank scores to compare individual performance in this specific context.

Y (Raw score)	Q(Y) (Percentile rank)
1	0.2
3	0.4
5	0.6
7	0.8
9	1.0

Maks poeng: 3

15 Variance-Covariance

We have the following variance-covariance matrix between items X_1 , X_2 , and X_3 forming the scale $Y = X_1 + X_2 + X_3$ and a criterion Z . (4 points)

- Obtain the sample variance of the scale score Y .
- Obtain the sample covariance between the scale score Y and the criterion Z .
- What would be the sample variance of the scale score Y if we would upweigh the contribution of X_2 and downweigh the contribution of X_3 to the scale score of, i.e., define $Y = X_1 + 2X_2 + 0.5X_3$?
- What would be the sample covariance between the scale score Y and criterion Z if we would upweigh the contribution of X_2 and downweigh the contribution of X_3 to the scale score of, i.e., define $Y = X_1 + 2X_2 + 0.5X_3$?

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

	X_1	X_2	X_3	Z
X_1	1.00	0.70	0.60	0.20
X_2	0.70	1.00	0.40	0.10
X_3	0.60	0.40	1.00	0.30
Z	0.20	0.10	0.30	1.00

Maks poeng: 4