GUIDE TO ITEM ANALYSIS FOR FSW EXAMINERS

INDEX GUIDE

Note: This guide focuses on the analysis of test questions in Remindo only, to ensure a smooth transition to the new test platform Ans. The guide does not focus on test analysis, nor on other types of tests (written and oral examinations, and final papers).

A more comprehensive guide for examiners, including instruction relating to these test forms, can be expected in Q3-Q4 of 2024.

SHORT STEP-BY-STEP GUIDE TO ITEM ANALYSIS

Below is a brief roadmap, which will be explained in more detail in the manual.

1. ***Identify questions in the Excel/Remindo that may need revision,*** based on orange and red p'- and rir-values and response options chosen by students you can form an opinion about the quality of the test questions. Is the question (much) too easy -or rather too difficult-, is the answer model incorrect, is it a trick question, is the wording of the question and/or response options vague, etc.? Then action by the examiner is required, see step 2b-c.
2. ***For each question, choose whether you want to keep it with or without changes,*** *or whether you do not want the question to be transferred to Ans.*
   a. Retaining the question without adjustments: no action unless the question has a status other than 'approved'. In that case, approval of the question is required.

b. Do not transfer the question to Ans: indicate this in the Excel file according to the instructions in the mail. Evt. you can (using your table of specifications[1]) directly create a new question in Remindo. However, it is recommended to do this in Ans after migration.

c. Keep the question with adaptations: depending on your analysis in step 1, you will adapt the answer model and/or the test question and/or the response options in Remindo. You can use the guidelines providing in this document.

LITERATURE

The following sources were used in the preparation of this guide (all in Dutch):

- Checklist meerkeuzevragen. LLInC.
- Kwaliteitsborging toetsing – Een handreiking voor Examencommissies
- MC Exam Workshop Board of Examiners
- Overzicht p-rir-waarden (LLInC, Universiteit Leiden)
- Paragin informatie over toetsanalyse in Remindo: Toetsanalyse in RemindoToets - Paragin
- Tips bij toetsen (juli 2016). Geschreven door: Maarten Bergwerff, Marlous Dekker, Floris van Blankenstein, Daan Romein, Marie Klaren. *Faculteit der Geesteswetenschappen*, Universiteit Leiden.
- Toetsen in het hoger onderwijs (2017). Van Berkel, Bax en Joosten. Hoofdstuk 2.4 Psychometrische analyse: het berekenen van indicatoren.
- Toetsing en toetsanalyse (2008). Geschreven door: D.N.M. de Gruijter. *ICLON*, Universiteit Leiden.

More information on the practicalities of testing via Remindo at FSW can be found here: Timeline - Digital Assessment with Remindo (FSW) | Rise 360 (articulate.com)

PRACTICAL INFORMATION

Here are some practical pointers on (re)viewing question banks:

### HOW DO I REVIEW (THE PSYCHOMETRIC INFORMATION ABOUT) MY QUESTIONS?

The Excel file of questions contains the information needed to assess the quality of the questions; you can also indicate whether you do not want to transfer questions in this document.

---

[1] Table of specifications (for an exam) means the overview/table in which you connect learning objectives and learning outcomes, including the weighting of the learning objectives and the number of questions asked per learning objective. See for an example: voorbeeld-toetsmatrijs.pdf (ugent.be)

If you want to adjust one or more questions (based on the Excel file), go to the question bank in Remindo:

➢ Log in to Remindo's test management environment: https://universiteitleiden.remindotoetsmanager.nl/ .
➢ In the top-right of your screen select the tab 'Questions'.
➢ Then view psychometric information at the 'Indicators', shown as two squares (see Figure 1). If one or both squares are coloured red, follow step 2 of the roadmap in this guide.

FIGURE 1. INDICATORS IN THE QUESTION BANK
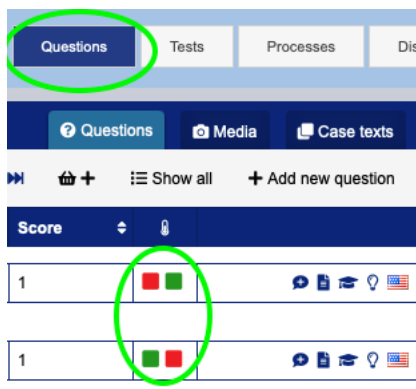
ITEM ANALYSIS: PRINCIPLES AND GUIDELINES

Below is a step-by-step description for how the test questions can be viewed by examiners during the switch from Remindo to Ans. The examiner looks at the individual test questions, then determines what action is needed regarding the individual test questions.

STEP 1. ASSESS THE QUALITY OF THE TEST QUESTIONS

First review the individual test questions based on the psychometric information in the Excel file or in Remindo. The questions with the worst values (negative rir or very low p') should be assessed first, keeping in mind whether revision is necessary. In doing so, the step-by-step plan below can be followed.

WHICH QUESTIONS MAY NEED REVISION?

RemindoToets highlights some questions that show anomalous values, this information was also added to the Excel file. When looking at the information, note *very high or low p'-values and/or negative rir-values*.

> ➢ You look at the difficulty of the question through the *p'-values*: which questions are very well or poorly made? The p'-value indicates the proportion of correct answers for a question, this value ranges between 0 (all students had 0 points on the question) and 1 (all students had the maximum number of points on the question). In other words, the higher p', the easier the question. Note that characteristics of the student and/or the test also influence the p'-value: questions at the end of a test often yield a lower value, also tests with many resits can lead to low p'-values.
>
> ➢ You look at discriminating properties of the question using the *rir-value* (item residual correlation). This is a correlation between the scores on the question and the scores on all other questions within a test.
>> o A **positive correlation means good discriminatory power**: students who made the question well also score well on the other questions and students who made the question poorly also score poorly on the other questions. So, the question distinguishes well between good and bad students.
>> o However, if the correlation is negative, then students who did well on the question actually score poorly on the remaining questions and students who did poorly on the question actually score well on the remaining questions. **A negative correlation means low discriminating power** of the question - then there may be something wrong with the question.[2]

---

[2] Taken verbatim from 'Tips bij toetsen' (FGW)

- When looking at the rir-values, you can also use RemindoToets' interaction analysis to look directly at the a and rar values. This allows you to see which response options were chosen frequently. And you can find a possible cause of the anomalous values.
- Third, try to explain the scores, e.g. by asking yourself the questions below:
  - Does the question test general knowledge (as opposed to a specifically formulated learning objective)?
  - Is the question too easy?
  - Were students already familiar with the question?
  - Is the question/are the response options open to multiple interpretations?
  - Is it coincidence (for a small number of students)?
  - For an open-ended question: does the answer model prevent assessor effects (sufficient)?
  - Does the answer model contain an error? See instructions to correct it here: [Answer key and changing the response model - Digital Assessment with Remindo (FSW) | Rise 360 (articulate.com)](#)

See also Table 1 to interpret the p' and rir-values [3]. Note: this overview is based on multiple choice questions (with four response options), for open questions the lower limit is set at p'= 0,25.

TABEL 1. INTERPRETATION P'- AND RIR-VALUES, INFORMED BY BERKEL, BAX AND JOOSTEN (2017)

	Rir <0,15/negative	Rir between 0,15-0,25	Rir > 0,25
P' < 0,4 **(difficult question)**	There is a possibility that this is a bad question: Students who score well on the test actually score poorly on this question. *Inspect question and answer model. Consider adjusting answer-model or remove question if inspection suggests.*	Possibility that this is a bad question (e.g., mismatched learning objectives, detail question, ambiguous question, multiple answers correct) *Inspect question and answer model. Consider adjusting answer model or removing question if inspection suggests.*	This question probably mainly discriminates between the best-scoring students. *Question can be retained in principle. Note: if there are relatively many such questions in the test then this is potentially problematic.*
P' between 0,4 and 0,8 **(average question)**	Chances are that this is a bad question: students who score well on the test actually score poorly on this question.	The question seems to have limited discriminating properties: *Inspect question and answer model. Consider adapting*	This seems to be a good question. *No need for further inspection. The question can be upheld.*

---

[3] Overview p-rir-values (LLlnC) and 'Toetsen in het hoger onderwijs' (Van Berkel, Bax & Joosten, 2017)

	Inspect question and answer model. Consider adjusting answer-model or remove question if inspection suggests.	*answer model if inspection suggests it.*	
P' > 0,8 (easy question)	This is probably a bad question: students who score well on the overall test actually score poorly on this relatively easy question. *Inspect question. Remove the question only if there is a clear reason to do so upon inspection.*	Possibility that this is a bad question: the question is easy and does not discriminate (e.g., question tests general knowledge, the distractors are not plausible). *Inspect question. Remove the question only if there is clear reason to do so upon inspection.*	This question probably mainly discriminates between the worst scoring students. *Question can be retained in principle. Note: If there are relatively many such questions in the test then this is potentially problematic.*

THE QUALITY OF A QUESTION DEPENDS ON THE CONTEXT

Ultimately, the key issue is whether a question is of sufficient quality according to the examiner based on the pre-established table of specification, i.e. with the learning objectives in mind. Sometimes, from a didactic point of view, it can be useful to ask a relatively very easy, or, on the contrary, very difficult question - and still preferable to keep a question despite low discriminating properties.

➤ Perhaps the easy question tests a learning objective that has been addressed repeatedly, and which students should have achieved if they have studied independently. If the question still seems too easy, it is possible to test the learning objective at a higher level, or to give the question less weight.

➤ Possibly the difficult question tests an aspect of the material that has been added as a challenge. In doing so, consider whether the material has been covered sufficiently (accessibly) in teaching.

STEP 2. DECIDE WHICH QUESTIONS SHOULD BE RETAINED, MODIFIED AND/OR NOT TRANSFERRED

After analysing the test questions, determine which of the following decisions applies:

a. **If you want to keep a test question (with status 'approved') without modification, you do not need to do anything further.** The question will be transferred to Ans by the DT team.
   o Note: questions with a different status (e.g. Need repair/concept) are not included in the Excel file. It may pay to look at these questions in Remindo's management

environment. If you want to transfer the question to Ans, it is necessary to adjust the question status to 'approved', see:[Managing your question bank - Digital Assessment with Remindo (FSW) | Rise 360 (articulate.com)](#)

b. **If you do not want one or more test question(s) transferred to Ans,** indicate this in the Excel file you receive from the DT team via e-mail. The DT-team ensures that these question(s) will not be transferred to your new question bank in Ans. In due course, this may mean that you will have to make (a) new examination question(s) using your table of specification.

c. **If you want to adapt one/more test question(s),** see the suggestions under 'Improving the quality of a multiple choice question' and/or 'Improving the quality of an open question'. Here you will find suggestions for adjusting the wording of the question and/or response options directly in Remindo.

   o **If you want to edit the answer model and/or make changes directly in Remindo,** see [Adding question(s), approving and editing - Digital Assessment with Remindo (FSW) | Rise 360 (articulate.com)](#).

IMPROVING THE QUALITY OF A MULTIPLE CHOICE QUESTION

You can find some guidelines for assessing the quality of a multiple choice question below. The 'Checklist for multiple choice questions' (LLInC), the MC training for Examination Boards (Political Science) and the Tips for toetsing (FGW) were used.

Writing the stem (=context and the question itself)
- Give clear instructions in the stem (why, when, what, how, etc.). The student should be able to answer the question even without the response options.
- Describe the context and central idea of the question.
- Keep the stem short and concise.
- Use positive wording, avoiding words such as NOT, NONE or PROHIBITED as much as possible. If it is necessary to use negative words, make sure they stand out by using capital letters or underlining them, for example.

Writing the alternatives (=the response options)
- Make sure one answer is truly correct - avoid overlapping alternatives. Also avoid the wording 'none (or all) of the preceding alternatives'.
- Keep response options independent; there should be no overlap between response options.
- Make sure response options are similar in content and grammar.
   o Avoid double negations.
   o Make alternatives grammatically consistent with the stem.
   o Use the same terminology consistently.
   o Avoid abbreviations.
- Ensure the length of the response options is similar.

- Be restrained in using wording like 'none of the above...' or 'all of the above...'
- Formulate response options positively; avoid negative wording such as NOT.
- Do not give clues towards the correct answer, such as:
  o Specific indications (always, never, totally, absolutely)
  o Associations with text in the stem; words identical to words in the stem or similar to words in the stem
  o Grammatical inconsistencies between response options
  o Multiple related response options
  o Obviously different response option
  o Absurd response option
- Ensure all distractors are essentially plausible.
- Make sure the question cannot be answered with general knowledge, but is really related to the material covered in the teaching. Where reference is made to an opinion, or where a quotation is used, include the reference to the source/article/book.
- Use typical thinking mistakes students make to write distractors (you can collect these during teaching...)
- Vary the location of the correct answer or always place answers in the same logical order (e.g. alphabetically).

IMPROVING THE QUALITY OF AN OPEN-ENDED QUESTION [4]

Some guidelines for assessing the quality of an open-ended question are described below.

Language
- Is the question grammatically correct?
- Does the question contain complicated sentence structures?
- Does the question contain a double negation?
- Does the question contain unnecessary insertions?
- Is the question unnecessarily negative?
- Can the wording of the question give rise to misunderstandings?
- Can shifting the emphasis give the question a different meaning?

Information
- Does the question contain sufficient information for answering?
- Does the question give enough information about the desired length and shape of the answer?
- Is it clear that an answer should be explained/illustrated?

---

[4] Taken verbatim from 'Tips bij toetsen' (FGW), using: CITO - Het construeren van open vragen
http://www.toetswijzer.nl/html/toetsenopschool/hoofdstuk7_herzien.pdf

- Are information and problem statement clearly separated?

Relevance

- Can the question be answered using knowledge other than that intended?
- Does the question suggest a problem that is not at issue?
- Does the question contain unintended hints to the answer?
- Is the difficulty level of the question respectively of the entire test acceptable?
- Is the difficulty of the question unnecessarily increased by irrelevant data?

Context

- Is the use of context (drawings, graphs, texts, images) functional?
- Is the context clear and correctly presented?
- Does the context contain irrelevant information?

Presentation

- Are the questions and question sections clearly distinguishable from each other?
- Is the numbering of the questions logical and clear?
- Are the conventions regarding spelling, symbol use, punctuation, etc. observed?
- Are the references in the question to texts, drawings, etc. correct?

SUPPORT AND CONTACT INFORMATION

For technical or substantive questions or comments on testing, please contact the Digital Testing team at digitaltesting@fsw.leidenuniv.nl.

Workshops to support examiners will also be organised by SOLO and LLInC:

- 4 April 11.00-12.00 How to improve your exam (questions) with psychometric data
- 4 April 12.00-13.00 Mastering the art of test question design
- 4 June 11.00-12.00 How to improve your exam (questions) with psychometric data
- 4 June 12.00-13.00 Mastering the art of test question design

You can sign up here: https://www.staff.universiteitleiden.nl/vr/social-and-behavioural-sciences/solo/information-for-lecturers/develop-your-teaching-skills/sign-up