*The following table provides an overview of signal values and possible actions. The values are not indisputable – consider them indicative. The possible actions are always at the discretion of the examiner. Sometimes immediate action may be necessary, for example, when it turns out afterwards that the wrong key was used during marking (****a*** *was counted as correct but it should have been* ***c****) or, e.g., when a question proved impossible to answer due to ambiguity. In that case, a decision must be made whether the question will not count and the cutting score will be adjusted, or whether everyone will receive a point.
In general, one will strive for a well balanced exam, with some easy and difficult questions, but most questions in the middle range, with discriminating power.*

CELT-UT, T&A course material, June ‘25

| **P-value** | **Rit Value** | **Interpretation** | **Immediate action** | **Evaluation, for next time** |
| --- | --- | --- | --- | --- |
|  | Rit negative | Key-mistake? Unclear or confusing question?  | Needs review (What happened?). Possible adjustment of scoring.  | Adjustment, removal |
| Low < 0.4\* | High > 0.3 | (Very) Hard question but discriminates well | Consider: Too difficult? Especially if P-value below 0.2. Else (very) challenging item, but can be kept as difficult question  | Can be kept but good to keep an eye on next timeTopic might need more attention in class |
| Low < 0.4\* | Low < 0.3 | (Very) Hard question and doesn’t discriminate well, less reliable | Review (What happened? Guessing, confusing wording, etc.). If very low values, maybe adjustment of scoring. | Adjustment, removal |
| Mid-range0.4-0.9  | High> 0.3 | Ideal item (reasonable difficulty & good discrimination), contributes to reliability |   | KeepTopic was learned, understood well |
| Mid-range0.4-0.9 | Low< 0.3 | Moderate difficulty but not discriminating well, less informative |  | Review, may be poorly con-structed or irrelevant question. If high P-value – check distrac-tors, maybe easy to guess |
| High > 0.90 | High> 0.2 | Easy question, but still informative, contributing to reliability |  | Consider keeping as more easy question (e.g., for foundational concepts) |
| High> 0.90 | Low> 0.2 | Very easy question, doesn’t contribute much to reliability |  | Can be check for basic know-ledge, but not very informative. Might have been guessable and needs adjustment.  |
| **A-value** |
| Low < 0.05 |  | Distractor is not functional |  | Remove distractor or adjust for a next time |
| Higher than P-value |  | Distractor seemed more likely for many students | Needs review, especially if students doing well in total choose the distractor (What happened? E.g. two answers were correct; wrong key). Might need adjustment of scoring if severe problem, especially for high value. | Review, adjust question and/or distractors If question was constructed well: Consider teaching/learning process. Why do so many students make this mistake.  |

\* 0.4 is already quite low. If the value goes beyond 0.2 it really needs to look into, especially if the RIT is also very low, and maybe immediate action before grading.

Table is based on: [Testvision-Toets-en-itemanalyse-digitaal.pdf](https://husite.nl/toetsing-nieuw/wp-content/uploads/sites/299/2020/02/Testvision-Toets-en-itemanalyse-digitaal.pdf); ChatGPT 20-6-2025;prompt: *If you execute a psychometric analysis on an exam. How to interpret the P-value versus the Rit value in relation to each other to know whether and what measures should be taken?*