****How to use Contest’s test result analysis?

This document is intended to help you to interpret Contest’s test result analysis and using it for adapting the grading at option, question and test level.

*Version: August 2021*

In order to do this, you will need the document Reportbook.xlsx.

* If you do not have a Contest account, this document is sent to you by email. This document is named **REPORT*\_<course\_code+exam\_name+date>.***
* If you do have a Contest account, you can download Reportbook.xlsx as follows:

**Reportbook.xls**

1. Go to the **Reports tab.**
2. In the dropdown menu **Report Modules**, make sure that the **Grading report** is selected.
3. Click the **Generate button.**
4. In the dropdown menu **Export to the selected format**, choose **Excel Worksheet.**
5. Click on the text **Export**. This will export the .xlsx-file.

## Quick links / Frequently asked questions

* 1. [Where do I find the answers of my students?](#_Where_do_I)
  2. [How do I know whether my students left questions unanswered?](#_How_do_I_2)
  3. [Where do I find the points and grade per student?](#_Where_do_I_1)
  4. [How does ConTest calculate the grade?](#_How_does_ConTest)
  5. [How do I get a first impression of the quality of my questions?](#_How_do_I_3)
  6. [How do I improve the scoring based on the item analysis?](#_How_do_I_4)
  7. [Where do I find more detailed information on the quality of my test and of my questions?](#_Where_do_I_2)
  8. [I’m confused by the terminology. Could you enlighten me?](#_I’m_confused_by)

# Content

[Quick links / Frequently asked questions 1](#_Toc535931360)

[1. Content 2](#_Toc535931361)

[2. Quick start guide 3](#_Toc535931362)

[2.1 Where do I find the answers of my students? 3](#_Toc535931363)

[2.2 How do I know if my students left questions unanswered? 3](#_Toc535931364)

[2.3 Where do I find the points and grade per student? 3](#_Toc535931365)

[2.4 How does ConTest calculate the grade? 4](#_Toc535931366)

[2.5 How do I get a first impression of the quality of my questions? 5](#_Toc535931367)

[2.6 How do I improve the scoring based on the item analysis? 6](#_Toc535931368)

[2.7 Where do I find more detailed information on the quality of my test and of my questions? 7](#_Toc535931369)

[2.8 I’m confused by the terminology. Could you enlighten me? 8](#_Toc535931370)

[3. Test result analysis data and how to use it 9](#_Toc535931371)

[3.1 How do I use the test results to adjust my scoring and grading? 9](#_Toc535931372)

[3.2 Detailed steps in using item analysis to improve scoring (see 2.7) 9](#_Toc535931373)

[4. Details per tab in ReportBook.xls 12](#_Toc535931374)

[4.1 Meta data report 12](#_Toc535931375)

[4.2 Test analysis report (entire test): Cronbach’s alpha 12](#_Toc535931376)

[4.3 Score frequency report (entire test score) 13](#_Toc535931377)

[4.4 Item analysis report 14](#_Toc535931378)

[4.5 Item advice report 15](#_Toc535931379)

[4.6 Version mapping report 17](#_Toc535931380)

[4.7 Participant score report 17](#_Toc535931381)

[4.8 Grading report 18](#_Toc535931382)

[4.9 Explanatory notes 18](#_Toc535931383)

[5. Adjusting scoring and grading in Contest 19](#_Toc535931384)

[5.1 How do I adjust the scoring per question in Contest? 19](#_Toc535931385)

[5.2 How do I give bonus points to individual students? 20](#_Toc535931386)

[5.3 How do I adjust the grade calculation in Contest? 20](#_Toc535931387)

# Quick start guide

## Where do I find the answers of my students?

You can find the unprocessed answers of the students in testdata.csv (see Table 1 for an example). If you do not have a ConTest account this file will be sent to you with the name **DATA\_*<course\_code+exam\_name+date>.***

**CONTEST ACCOUNT**

You can find the answers per student (testdata.csv-file) here:

* Go to the **Data tab**
* Click on the button **Export**. This will export a.csv-file (**testdata.csv**) with the answers of students.

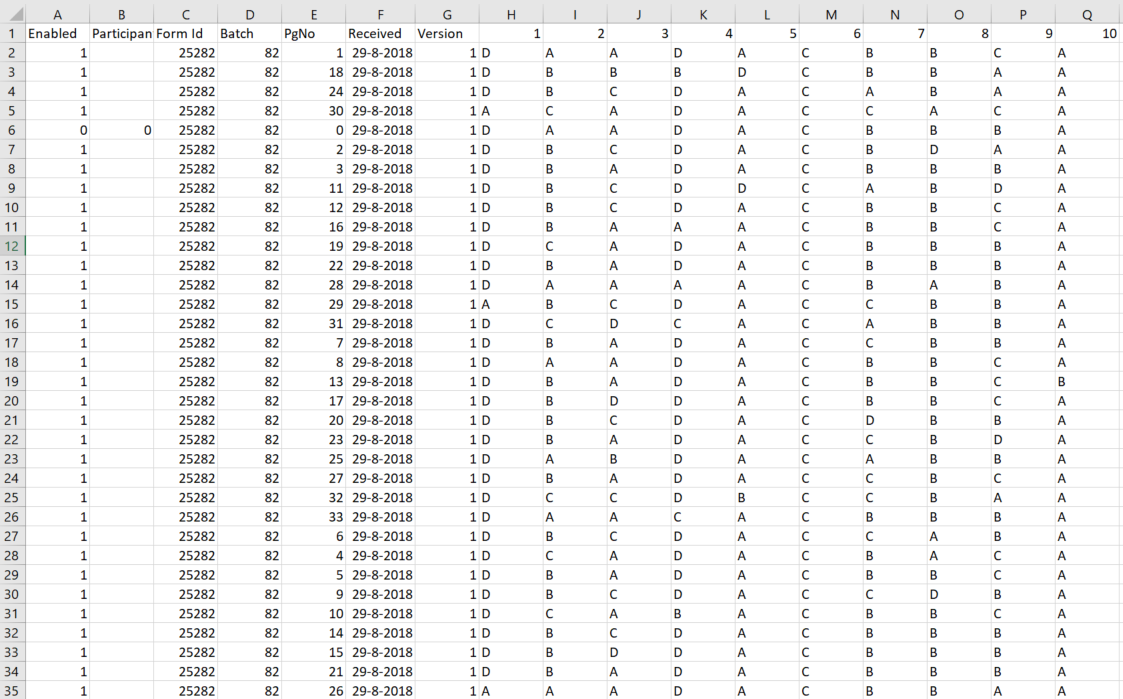


Table 1: Example of testdata.csv. The student numbers are left blank in this example, except for the key form 000000.

## How do I know if my students left questions unanswered?

In the tab **Participant Score Report** (see Table 2) and in the **Grading Report** (see Table 3) in the Reportbook.xlsx, the columns **Open** (see green squares) indicate the number of unanswered questions per student.

In case you **do** **not** have a Contest account and want to make sure they really did not fill in something, you will need to find their answer sheets in the paper pile of answer sheets, which you can collect from the ESA desk after you have received an email with the data and the report.

In case you **do**  have a Contest account, you can use the search option **in Contest**:

* Login to **Contest** and open the exam
* Go to the **Data** tab.
* In the field **Participant number**, fill in the student number and click **Search**.
* The data will show.
* Click on the document icon to open the student’s data, including the scan.
* You can **zoom** in by hovering the mouse over the scan, or **download** the scan as pdf.
* You can **edit** the recorded student answer by pressing the **Edit button.**
* Click **Save** to save.

See ‘[How do I create a test result analysis in ConTest](#_top)?’ on how to create a new test result report yourself.

## 

## Where do I find the points and grade per student?

In the Excel file **ReportBook.xls**, you can find this information in two separate tabs:

### Participant Score Report tab (without student names, without grades)

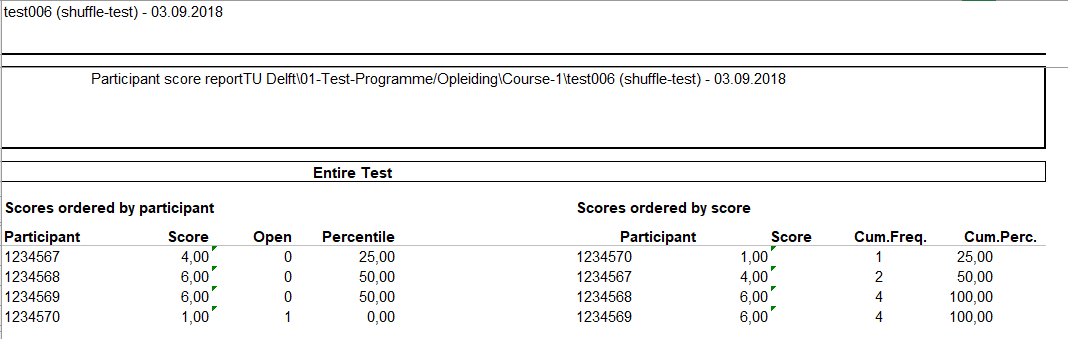


Table 2: Example of the Participant Score Report in ReportBook.xls. Red squares indicate the two scores ways scores are ordered; Green square: The open column indicates the unanswered questions per student.

Content:

* The total score per participant number.
* Student names are not included.

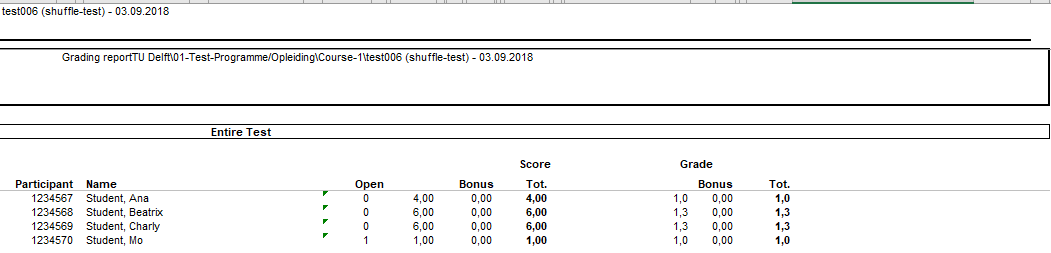
Feature:

* Ordered by participant in the left part of the table.
* Ordered by score on the right part of the table.

### Grading Report tab (including student names)

In the grading report tab (see **Table 3** for an example), the student numbers are listed together with the names of the students. It also contains the total score, bonus points awarded and the grades (refer to the next section to see how it is calculated). Open questions can only be added manually in ConTest.

Table 3: Example of the Grading Report in ReportBook.xls Left red square: score information. Right red square: grade information. Green square: information on the number of unanswered questions per student.



## How does ConTest calculate the grade?

ConTest draws a straight line between the average guessing score, corresponding to a grade of 1.0, up to the maximum score, which is awarded a grade of 10.0. Up to the average guessing score, the students are awarded with a 1.0.

You can find a plot of the score-grade transformation on the **Meta Data Report** tab. See Figure 1 for an example, of the score-grade transformation for 22 simple MCQs with 4 options. You could use the plot to communicate the calculation of the grade to your students.

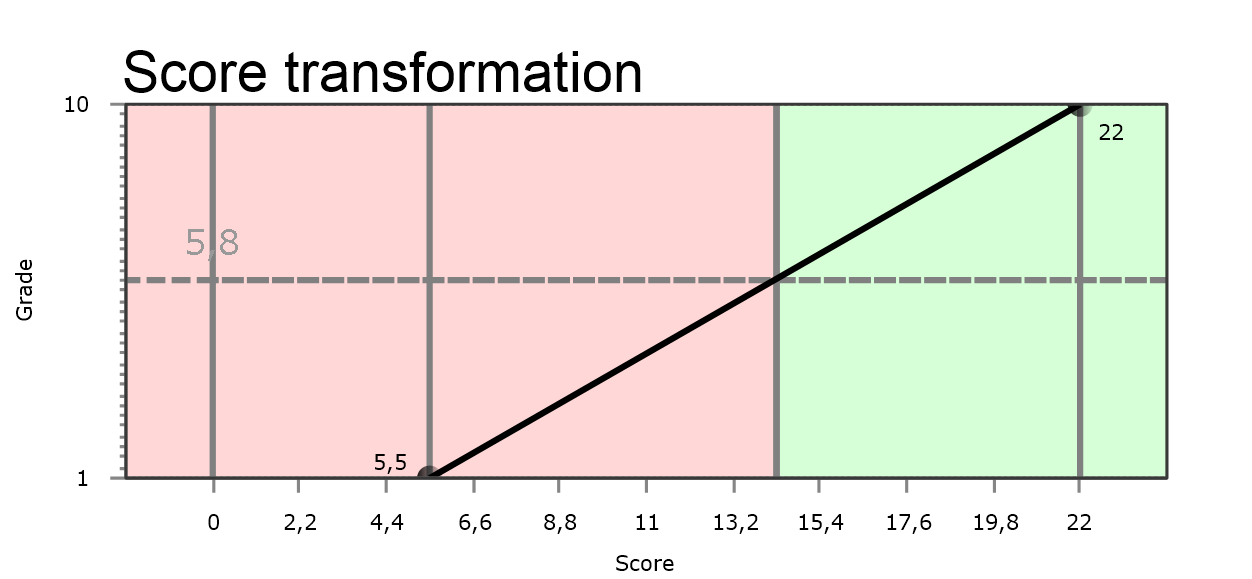


Figure 1: Score-grade transformation for an exam with 22 questions

## How do I get a first impression of the quality of my questions?

Open the **Item advice report** tab and scroll down for advice based on the difficulty of the question (p’) and on whether the score for this question correlates with the scores for all other questions (Rir). For an example, see below in Table 4. A detailed description of how p’ and Rir are calculated can be found in chapter 4.4.



Table 4: Item advice report in words

## How do I improve the scoring based on the item analysis?

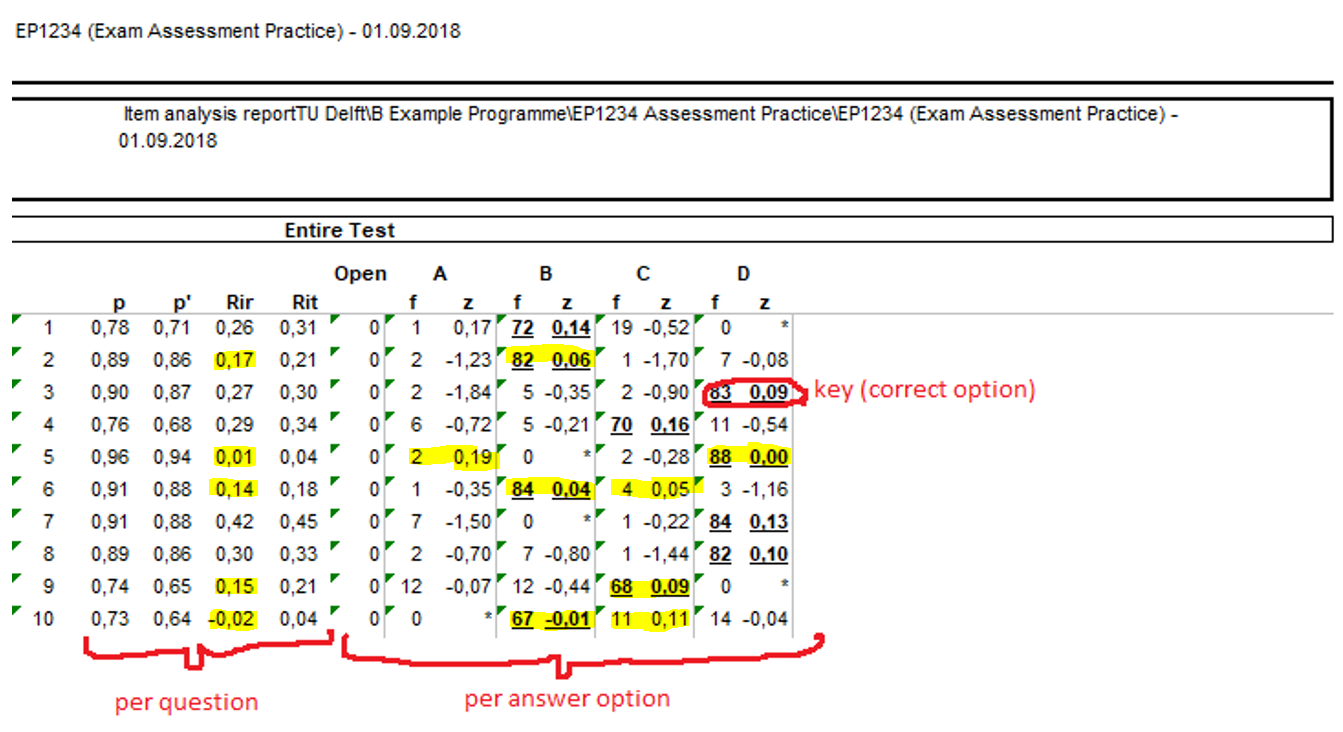


Figure 2: Step 2.1 mark all correct options (underlined) with a z-value below 0.2, and 2.2: marking all incorrect options with positive z-values. You can find the Item analysis report in the Reportbook.xlsx.

1. Print the **item analysis report** (see Figure 2)
2. Mark all Rir-values below 0.2  Check what is going on in the following steps, especially when the number of students affected (f) is relatively high.
   1. Mark all **correct options** (underlined data) with a **negative z-value or a z-value below 0.2**  check whether these options were incorrect. If so, [change the scoring](#_How_do_I).
   2. Mark all **incorrect options** that have a **positive z-value**  check whether these options were correct, or partially correct. If so, [change the scoring](#_How_do_I).
3. If some low/negative Rir-values were unexplainably low by the above steps[[1]](#footnote-1):
   1. In case of a high p: [Give all students full points](#_How_do_I).
   2. In case of a low p: [Remove the question from the grading](#_How_do_I), and [give bonus points](#_How_do_I_1) to the few (good performing) students who got the answer correctly.

More information on the steps can be found in [3.1 How do I use the test results to adjust my scoring and grading?](#_How_do_I_5)

## Where do I find more detailed information on the quality of my test and of my questions?

If you want more information check the following tabs:

* Test analysis report (entire test): Cronbach’s alpha: you can find Cronbach’s alpha here, which is a measure for the reliability of the test. Cronbach’s alpha should ideally be at least 0.8 when this tests is determining the final grade for your course alone. Otherwise, it should be at least 0.7.
* Score frequency report (entire test score): this is a histogram of the scores.
* Item analysis report: Here you can find the average score per question (p-value), the correlation of each option with the total score (z-value)

## I’m confused by the terminology. Could you enlighten me?

Here is an overview of confusing terms:

Correlation: Whether two things are related. Example: question 1 and question 2 correlate if students who correctly answer question 1, also correctly answer question 2. And if students who *incorrectly* answer question 1, also incorrectly answer question 2. Examples of correlations are Rir, Rit, and z.

Difficulty: How difficult a question was. This is measured by the p-value, which is the percentage of students who answered this question correctly, or the average score. However, if the p-value is high, the difficulty is low.

Distractor: Option that *is not* the correct answer.

Frequency: Number of times that something occurs. Example: the frequency of option C is the number of students that chose option C. The frequency of 6.15 points is the number of students that had a total score of 6.15 points.

Grade: The grade for the test that is communicated to the student. The grade is calculated from the total score by a formula, the *item-grade transformation*.

Item: Question. For open questions, it is a subquestion. An item is the smallest unit in a test.

Key: Option that *is* the correct answer.

Reliability: How reliable the grade is, based on an estimation of the measurement error. Cronbach’s alpha is measure for the reliability. The higher alpha, the higher the reliability. Reliability and reliability are two different things.

Score: The number point(s) that a student earns per question (*item score*) or for the whole test. *Total score* is the sum of all item scores. Score is expressed in points.

Validity: To what extent does the test represent the learning objectives that it is supposed to cover, at the right level? This is not covered in the test result analysis. You could make an assessment matrix to check whether your test is valid. *Representativity* (of the learning objectives) is a synonym.

# Test result analysis data and how to use it

## How do I use the test results to adjust my scoring and grading?

[In the next chapter](#_Details_per_tab), you will find an overview of the tabs in the **ScoreBook.xls** document that you can use to adjust your scoring (per item) and grade calculation for this test.

Figure 3: Test result analysis on three levels: 1. assessment level, 2. item level, and 3. option level

### 1. Analysis

You will analyse the test on three levels:

* **Assessment level**: what impression does the overall test give?
* **Item level**: are there questions (items) need your attention? You will analyse
* whether the question was reasonably difficult (p-value), and
* whether an individual question was able to discriminate between ‘good performing students’ (students with a high total score) and ‘bad performing students’ (students with a low total score).
* **Option level**: within a question (item), were there for example options that were chosen by relatively good performing students and that are (in hindsight) also correct or partially correct?

### 2. Adjusting scoring

Based on the analysis, you may choose to change the key (which answer is correct) and/or grant points to options that were correct too, or that were partially correct. If that is not enough, you might choose to give all students full points for an item, and/or give bonus points for students who gave the correct answer.

### 3. Adjusting grading

If the test was made really badly, and did not reflect the level of the students, you might want to increase the grade by adjusting the score-grade transformation.

## Detailed steps in using item analysis to improve scoring (see 2.7)

Here is a more detailed version of How do I use the test result analysis for changing the scoring in a straight-forward way?

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Analysis | Adapt scoring | Next year |
| 1. | Print the item analysis report. (see Figure 2) |  |  |
| 2. | For all items with Rir-values below 0.2, execute 2.1 and 2.2 (see Figure 2): |  |  |
| 2.1 | Search for **correct options** (underlined) with **z-values below 0.2**:  Especially if chosen by a substantial amount of students, check whether these options were incorrect (incorrect key). | If so, change the scoring. | Rephrase next year. |
| 2.2 | Search for **incorrect options** that have a **positive z-value:**   Especially if chosen by a substantial amount of students, check whether these options were correct, or partially correct. | If so, change the scoring. | Rephrase next year. |
| 3. | If some low/negative Rir-values were unaccounted for by 2.1 and 2.2[[2]](#footnote-2): |  |  |
| 3.1 |  Check whether p was high. | If so, give all students full points. | Rephrase next year. |
| 3.2 |  Check whether p was low/negative. | If so, remove the question from the grading, and give bonus points to the few (good performing) students who got the answer correctly. | Rephrase next year. |
| 4. | Mark all **implausible options** (all f-values below 10% of the ‘incorrect students’ and the ones that are indicated by2 (not chosen at all)). |  | Next year, either **rephrase** these implausible options, or **reduce the number of options**[[3]](#footnote-3). |

Table 5: How to use the item analysis for the adaption of grading.

You can use the step-by-step procedure in Table 5, to use the item analysis for the adaption of grading. In step 2, you first filter the worrisome items, based on whether or not their score correlates with the total score of the other items (Rir).

*The underlying idea is that all test questions try to answer the same question, namely to what extend students master the course’s learning objectives (LOs). Since learning objectives are the goals that students try to master during a course by doing exercises that lead to these LOs, this level of LO mastering depends on how actively students participated in the course. Assuming that students participated equally in all LO preparing activities, we expect the scores on a single item to correlate with the sum of the other items. This correlation is called the Rir (item-rest correlation).*

*If this correlation is low (close to 0, below 0.2), this indicates that the score on this item does not contain information on how they did on the rest of the test. If the Rir is (quite) negative, answering this question correctly actually indicates that you did not do very well on the rest of the test. This might indicate that the question was a trick question for the well performing students.*

In 2.1, you search for correct answers that were chosen by relatively bad-performing students, indicated by a negative z-value. Also, correct answers that do were chosen by both good and bad performing students are studied (z-value below 0.2). Negative or low z-values are indications that the key is incorrect (e.g. D was the correct answer, not C), or that multiple options might be (partially) correct, respectively.

At the same time, in 2.2, you search for incorrect answers (distractors) that were chosen by relatively good-performing students, indicated by a positive z-value. If these answers make sense, these could be the (partially) correct answers.

If 2.1 and 2.2 do not answer the question why there is little or negative correlation between the score for this question and the total score for all other questions, the action you could take depends on how many students gave a correct answer.

3.1: If many students answered the question correctly, the few students who did not answer the question correctly, were probably otherwise good scoring students. In other words, the question gives misinformation on the performance of the student. Therefore, you want to remove the question from the grade calculation. In order to do that without disadvantaging the students who answered the question correctly, you can give all students full points.

3.2: If only very few students answered the question correctly, these were probably not very good performing students. again, the question gives misinformation on the performance of the student and you want to remove the question from the grade calculation. In order to do that without disadvantaging the students who answered the question correctly, you could remove the question, and give the few students who gave a correct answer a bonus point, or give all students full points, depending on your preference.

4: For this year, you are ready. You can use the above information to improve the questions next year. However, there is one extra indicator that can help you to improve the formulation of distractors in next year’s exam:

If no or very few students choose a distractor, this distractor was not plausible. You can choose to either reduce the number of distractors (3 options is optimal, i.e. 2 distractors and a key), or to make the distractor more plausible. How many students are ‘very few students’? This depends on the number of students that chose incorrect options (including ‘no answer’) altogether. You could choose 10% of the not-correctly answering students as a minimum to call an option ‘plausible’.

# Details per tab in ReportBook.xls

## Meta data report

Here you will find an overview of the minimum and maximum score per question, and the question type:

* 0: No scoring
* 1: Multiple choice with single correct answer option
* 2: Essay question
* 3: Multiple choice with composite correct answer option
* 4: Multiple choice with composite correct answer option (revised)
* 5: Multiple choice with incremental score per answer option
* 6: Multiple choice with custom scoring

Furthermore, the pass rate is mentioned and the score-grade transformation is illustrated in a table and in a graph (see Figure 1 on page 5).

## Test analysis report (entire test): Cronbach’s alpha

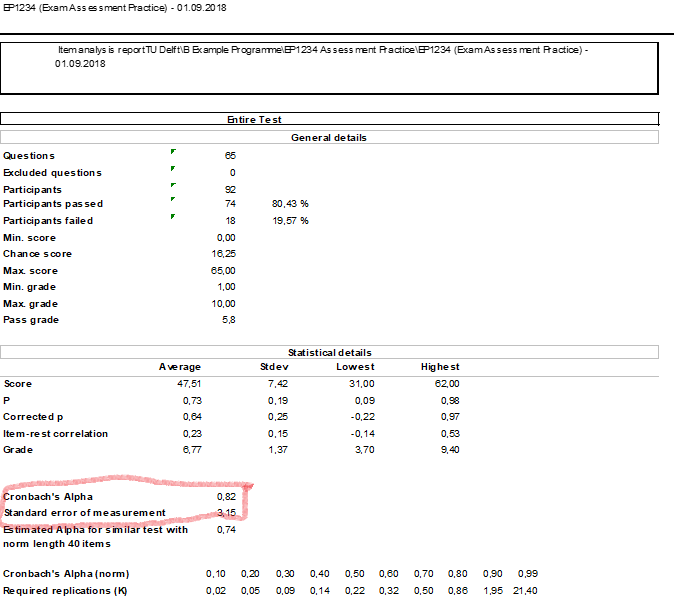


Table 6: Example Test analysis report

Here you will find the analysis of the test as a whole. Cronbach’s alpha indicates how reliable the grade is, and is used to calculate the standard error of measurement.

Desired values of Cronbach’s alpha are:

* 0.8 or higher in case this is the only test in your course
* 0.7 or higher if it is a partial test (for example, a midterm exam)
* 0.6 or higher for formative tests

A low Cronbach’s alpha implies that the grade is not precise enough for its goal. Adjusting the scoring might improve this. For information on how this is done in Contest, see Adjusting scoring and grading in Contest.

**Background information**: The standard error of measurement (just below Cronbach’s alpha, derived from Cronbach’s alpha) is the size of the 68% confidence interval (CI) around a grade (in the example of Table 6, the 68% CI of a student’s total score would be ±3.15 points, and the 95% CI of a student’s score would be ±6.30 points).

## Score frequency report (entire test score)

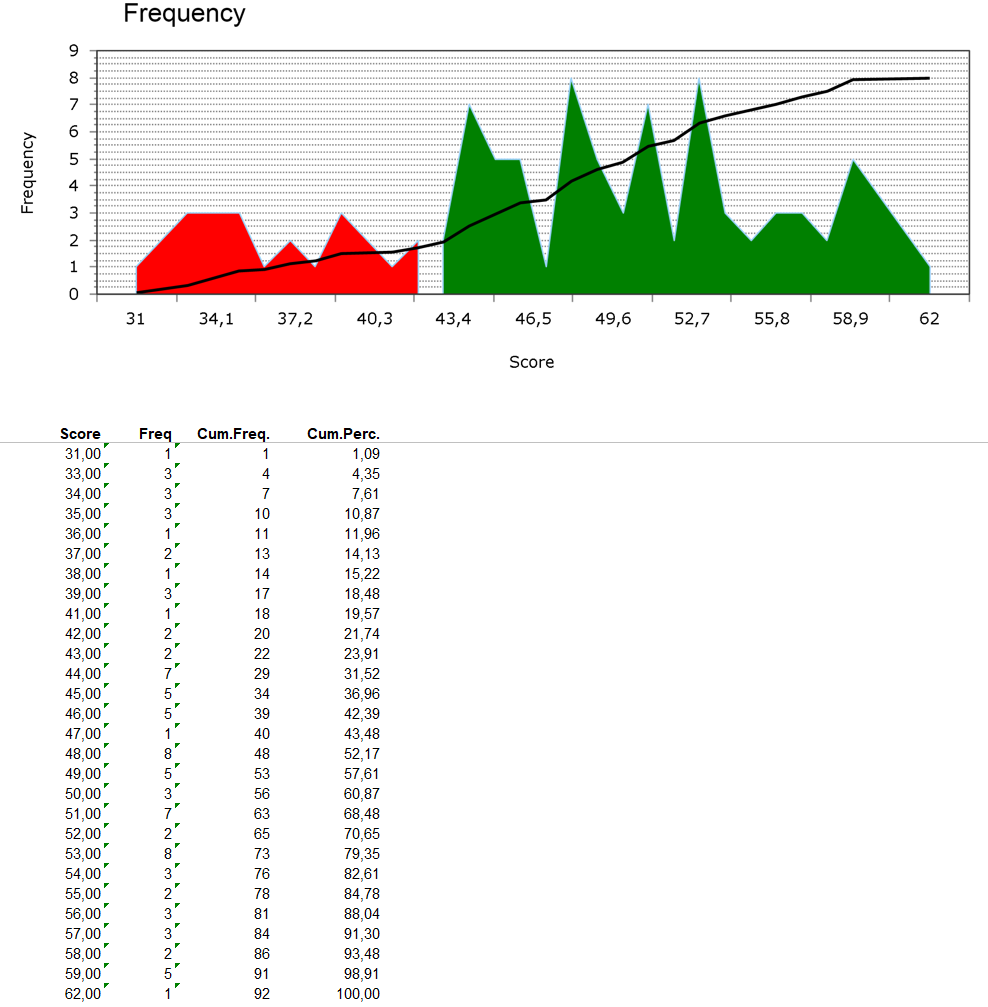


Figure 4: Example of a score histogram and table (in ‘Score frequency report’)

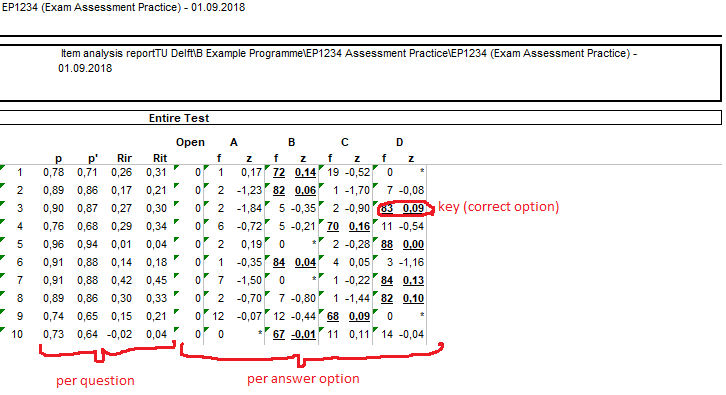
In ‘score frequency report’, you will find the frequency per total score: for all existing scores, the number of students who got that score is listed. You can find an example in Figure 4. The information in the table is also plotted in a histogram (the graph on top). Keep in mind that the maximum value at the horizontal axis is the maximal achieved score, and not the maximum possible score!

You can use this distribution to decide whether or not you need to increase the overall score, after improving the answer models for the individual questions. See 5. Adjusting scoring and grading in Contest.

## Item analysis report

In this table, the most important indicators of the quality of your test items (questions) are listed. Before we discuss this table, please note that on the next page, you will find a graphical representation of the item quality, and on the page after that, a summary of the main (possible) issues in plain text. We will first discuss the item analysis table.

Table 7: Example of (part of) Item analysis report Quality indicators are displayed per question (row) on question level (left) and per answer option. Indicators for keys (correct answers) are underlined.



### p-value

The p-value is the (normalized) average score for a question. For question 1 in Table 7, the p-value of .78 indicates that 78% of the students answered the question correctly.

For closed-ended questions, students can get points when they are *only* guessing. The p-value for a question where students would have guessed with 4 options will be 0.25. If there were 3 options, the p-value for guessing would be 0.33 (1/3).

### p'-value

In order to make clear whether students scored above the p-value for guessing, p’ was introduced. p' is zero at the guessing score, 0.5 if halfway between guessing and full score, .75 at ¾ between guessing score and full score, etc. If p’ is negative (<0), this indicates that less students chose the correct option than the guessing score. This is a sign that something is wrong, especially in combination with a low RiR (see below).

### Discrimination index / Rir-value

As a teacher, you want each question to discriminate between good performing and poor performing students, in order to give them reliable grades. We use the Rir-value as a measure for discrimination of items (questions). The Rir indicates whether students that perform well on this question are performing well on the other questions, and whether the students that perform poorly on this question, also perform poorly on the other items.

The Rir value is the correlation (symbolised by R) between the score for this (sub)question (item = i) and the ‘rest-score’ (r). The rest-score is the total score for all questions, except this (sub)question.

The Rir can have a value between -1 and 1.

### Interpreting Rir-values:

* We want the Rir to be larger than 0.25 (0.2 in ConTest since we have to round off at the nearest tenth), the question can give us information about how well the student performed on the test.
* In case the Rir is negative (< 0), this is an indication that there might be a problem. Good students might have overthought the question and might have been tricked into choosing an incorrect option, while not so good performing students performed better. Or another option might in fact be the correct option (incorrect key), or might *also* be a (partially) correct option. If the other options were not correct, [consider giving full points to everyone.](#_How_do_I)

### Rit-value

The Rit-value is the correlation (R) between an item (i) and the total (t) score. It is less reliable than the Rir-value since it correlates with its own value, but since in early computer days it used to be much easier to compute the Rit-value than the Rir-value, it is still displayed.

### f-value

The f-value indicates how many students have chosen a certain option. The correct answer is indicated in **bold and is underlined**.

### z-value

The z-value is a measure of discrimination per option. It is the correlation between choosing this option, and the total score for the other questions. z should be positive and largest for the correct answer, and negative or close to zero for the distractors (incorrect options).

If the z-value is negative for the correct answer, and positive for a distractor, this is a very strong indicator that the answer key may be wrong, and that the distractor with the high z-value was actually the correct answer all along. Or that it is *also* (partially) correct. [Give (partial) points to (partially) correct options.](#_How_do_I)

## Item advice report

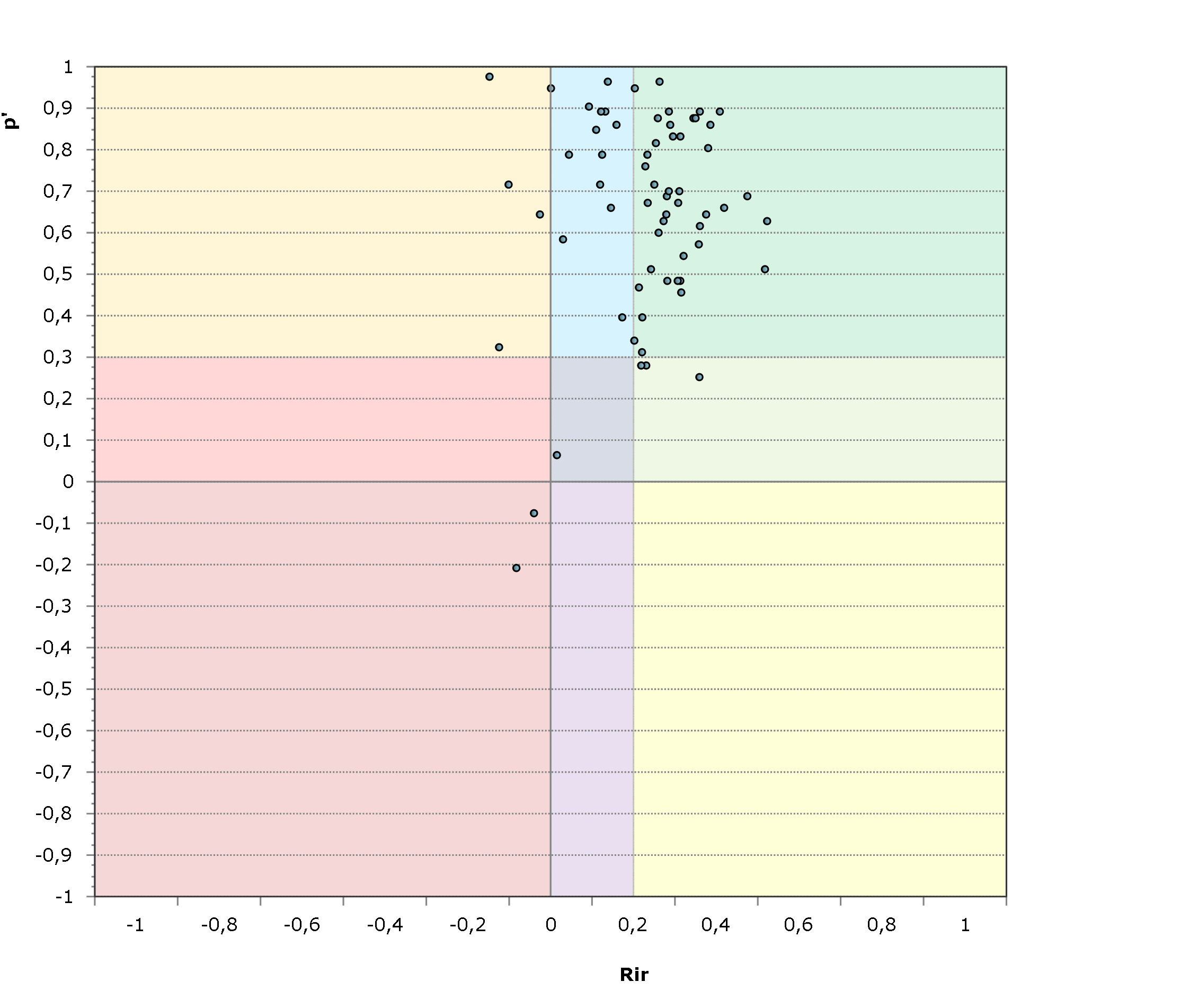
In the graph of the item advice report, the p’-value (see above) is plotted against its Rir-value. Below the graph, a legend is given, in which the letters refer to the more elaborate explanation below.

* **A: Trick question. Full points for all students or check whether options with a positive z-value were (also) correct.  
  Consider giving full points to everyone, and revising the question next year.**

The question did not discriminate between poor and well performing students. In other words: the question was answered incorrectly by the well performing students and/or correctly by the poor performing students. The question may have been a trick question that was overthought by well performing students, or there might be a hint that gave the correct answer away to poor performing students. Therefore, it does not seem to measure how well students are mastering the learning objectives and should preferably not have influence on the grade calculation.   
Check what happened and consider [giving full points to everybody](#_How_do_I). In this case, the people who did not get full points yet, are the the otherwise good performing students.

* **B: Consider adding (partially) correct options or give all students full points.**

The question did not discriminate between good and poor performing students.   
Students answered this question correctly, regardless of whether they studied for the course or not. Could students answer this question before taking the course?  
Consider [giving full points to everybody](#_How_do_I), and revising this question next year.



**B**

**H**

**I**

**G**

**D**

**F**

**E**

**C**

**A**



Figure 5: The p’ value and Rir value for all questions. Legend: Below

* **C: OK.**

This easy question helped to discriminate between good performing and poor performing students.

* **D: Key seems incorrect. Check whether options with a positive z-value were (also) correct.** This is especially relevant if more students chose one of the distractors, compared to the correct answer, and if its z-value was positive.   
  This difficult question did not discriminate between good and poor performing students. In other words: the question was answered incorrectly by the good performing students and/or correctly by the bad performing students. The question may have been a trick question that was overthought by good performing students, or there might be a hint that gave the correct answer away to poor performing students. Therefore, it does not seem to measure how well students are mastering the learning objectives and should preferably not have influence on the grade calculation.

Check what happened and [revise the answer model accordingly](#_How_do_I).

* **E: Students guessed. Consider adding (partially) correct options or give all students full points.**  
  The question did not discriminate between good and poor performing students. Whether students answered this question correctly or not does not depend on whether they studied for the course or not. If p’ is close to 0, they basically guessed. Was this part of the learning objectives? Did they practice with this type / topic of question?  
  Check whether some options did a positive z-value. Consider [adding more correct answers](#_How_do_I) (if applicable) or giving all students full points. Revise this question next year.
* **F: OK, though difficult.**

This question helped to discriminate between good performing and poor performing students.

* **G: Key seems incorrect. Check whether options with a positive z-value were (also) correct.** This is especially relevant if more students chose one of the distractors, compared to the correct answer. See page 4 and compare the z-values of the distractors.

This difficult question did not discriminate between good and poor performing students. In other words: the question was answered incorrectly by the good performing students and/or correctly by the bad performing students. The question may have been a trick question that was overthought by good performing students, or there might be a hint that gave the correct answer away to poor performing students. Therefore, it does not seem to measure how well students are mastering the learning objectives and should preferably not have influence on the grade calculation.

Check what happened and [revise the answer model accordingly](#_How_do_I).

* **H: Consider adding (partially) correct options or give all students full points.**

This difficult question did not discriminate between good and poor performing students and was made very poorly. Whether students answered this question correctly or not does not depend on whether they studied for the course or not. Students did worse than when they would have guessed. Was this part of the learning objectives? Did they practice with this type / topic of question? Is there a misconception that was not discussed in exercises or class?  
Check whether some options did a positive z-value. [Consider adding more correct answers (if applicable) or giving all students full points.](#_How_do_I) Revise the question next year.

* **I: OK? Check if the question was not too difficult.**  
  Check whether the question was not too difficult and whether at least some students gave the correct answer.

This difficult question helped to discriminate between good performing and poor performing students.

Below the graph, the graph is summarized in table form. You can use the letters to refer to the back information above Item advice reporton page 15).

## Version mapping report

If applicable, you will find the mapping from version 1 to the other versions. The use of different versions is only possible if you have a ConTest account. More information can be found in the ConTest manual [on the ConTest site.](http://www.icto.tudelft.nl/tools/contest/)

## Participant score report

See Participant Score Report tab on page 4.

## Grading report

See Grading Report tab on page 4.

# Adjusting scoring and grading in Contest

This chapter describes how to adjust the scoring per question, and the grade calculation based on the total score, in Contest. It also addresses how to give individual students bonus points, for example if you excluded questions from the grade calculation. Doing this yourself can only be done if you have an account in ConTest.

Figure 6: Improve the answer model 1st at the level of options, 2nd at the level of items, and 3rd at the level of the grades. Ideally, the adjustments are only at the option level.

In general, always adjust the scoring per question (steps 1 and 2 in Figure 6, before you decide on changing the grade calculation, or not. Ideally, the adjustments are only at option level.

## How do I adjust the scoring per question in Contest?

In case you decide to change how many points students will get per given answer, like giving full points to all students, you can do that in Contest. Each question with multiple (partially) correct answers, will (unfortunately) be removed from the item analysis by Contest[[4]](#footnote-4). Therefore, **save** a copy of the test result analysis, before adjusting the scoring per question.

* Open the **Scoring tab**.
* Click the **Edit button**.

For each question that you want to change the scoring for, you can either

1. Exclude the question from the grade calculation:
   * Tick the **Exclude tickbox**
   * Don’t forget to check whether the grade of individual students who answered this question correctly might decrease their grade. In that case, read the next section on how to **give students who answered the excluded questions correctly bonus points**.
2. Change the key (for example D was the correct answer, not C)
   * Tick the correct box (for example D)
   * Untick the incorrect box (for example C)
3. Add extra correct answers to the key (for example, D was also correct, not only C)
   * Tick the correct box (for example D)
   * The formula will now change (for example, from ‘C=1’ to ‘C=1;D=1’)
4. Add partially correct answers to the key (for example, D was also half correct, not only C)
   * In the **Type column**, select **6**.
   * Change the formula manually, by indicating the number of points per answer option, using fraction notation (e.g. 1/3 instead of 0.3333). You only need to indicate the options that will give students points. Separate the points by ;.
   * Examples:
   * C=1;D=1/2
   * C=1;D=1/3
5. Give all students full points, including the ones that did not answer the question
   * Tick all boxes (for example A, B, C, and D)
   * In the **column Empty**, fill in 1

 Do not forget to **press** the **Save button**!

## How do I give bonus points to individual students?

Use the following steps:

* Go to the **Data** tab. This is where the points per student can be found.
* Click on the **Participants** **button:** 
* Click on the **Edit button**
* Add **Bonus Scores** (which increases the score) or **Bonus Grades** (which directly increases the grade).
* Click **Save**.

## How do I adjust the grade calculation in Contest?

Reminder: For retake exams, do not increase the grades, since the group of retake students is not representative for the whole group of students. This group is expected to score lower.

### Cohen-Schotanus: give top 5% of your students a 10

You could for example give the top 5% of your students a 10, using the Cohen-Schotanus adjustment as follows:

* Go to the **Grading** tab.
* Click the **Edit** button.
* Scroll (down) to the **Score to Grade Transformation**.
* Set **Transformation template** to for example **Chance-score – Cohen-Schotanus – Percentile**.
* Click the **Update transformation** button.
* In the right, upper grey block (**Cohen-Schotanus**), click the **Edit** button.
* Set the percentile to **95**, indicating that the average score of the top 5% of the students will be set to the maximum grade (10).
* Set the factor (**c**) to 0,6, indicating that students will get a minimum pass grade (e.g. 6.0) when they have a score of 60% between chance and the top 5% score.
* Click **Save** in the grey box.
* Click the **Save** button at the top of the page.

### Adjust grade calculation by defining important points, like points needed for a 6.0 and points needed for a 10.

Another option would be to create your own linear transformation manually. You set a number of points through which Contest interpolates the score-grade transformation (bottom part of Figure 7).

In general, you are advised to define 3 points:

1. Score needed for grade 10.0
2. Score needed for grade 6.0 (this is called the *cut-off point*)
3. Maximum score that will still result in a 1.0 (the Chance score)

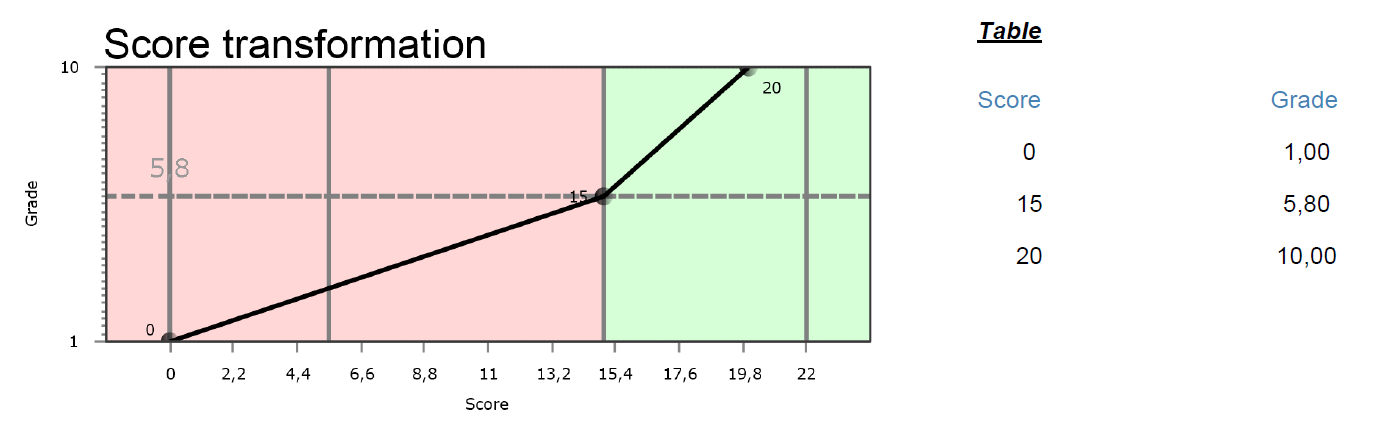


Figure 7: Score transformation (source: first page report). Visualizes which total score will lead to which grade.

This is how you can set the points (**Measure points**) in Contest:

* Go to the **Grading** tab.
* Click the **Edit** button.
* Scroll (down) to the **Score to Grade Transformation**.
* Set **Transformation method** to **Linear**.
* Set **Calculation method** to **Fixed scores (manual)**
* Set **Transformation template** to **None**.
* Set the number of **Measure points** for example to 3 (Contest will draw straight lines between these points to determine the relation between a student’s total score and their grade).
* Click the **Update transformation** button.
* Insert all the **Measure points** ((Score,Grade)-coordinates, 1 in Figure 8).
* Click **Save** in the grey box (2 in Figure 8).

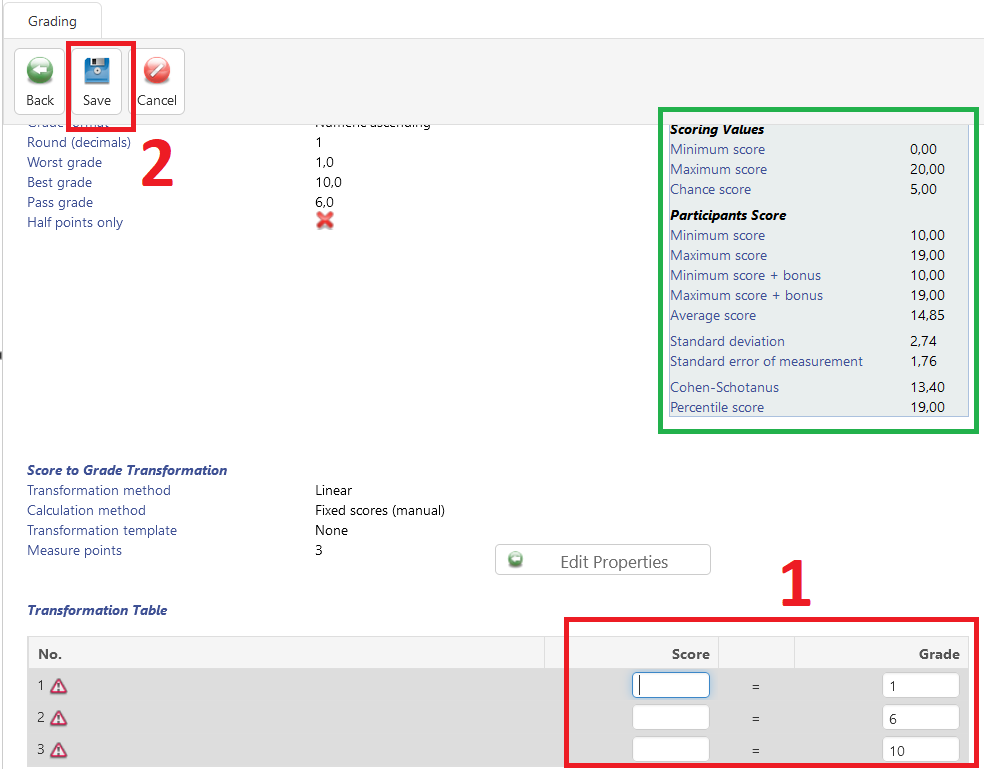


Figure 8: Creating linear transformation   
1: Fill in the Measure points here. 2: Save button. Green area: Information about the test, useful for deciding on the Measure points.

1. This question does not give information or gives misinformation about how well students master this course. Therefore, the question should not influence the grade. However, you might want to give students who got the answer right credits for their achievement. Consider the following in Contest or elsewhere: [↑](#footnote-ref-1)
2. This question does not give information or gives misinformation about how well students master this course. Therefore, the question should not influence the grade. However, you might want to give students who got the answer right credits for their achievement. Consider the following in Contest or elsewhere: [↑](#footnote-ref-2)
3. The latter would imply increasing the number of questions: As a rule of thumb, in case of decreasing the number of options, the total number of options should stay the same in order to keep the test reliability the same. In general, 3 options (including the correct option) are optimal, since it is a lot of work to come up with more than 2 plausible distractors. [↑](#footnote-ref-3)
4. Having more than one (partially) correct answer is only possible for question ‘type 6’. Contest does not calculate an item analysis for ‘type 6’ questions. [↑](#footnote-ref-4)