

# ConTest ReportBook user guide

---

Version: September 2018

You will receive three documents for your paper-based multiple choice test analysis:

- *ConTest ReportBook user guide*
- ReportBook.xls
- Raw data testdata.csv

This Guide consists of four parts:

1. Quick start guide: [answers to the most urgent question](#) (see below)
2. Test result analysis and how to use it
3. Explanation of the other tabs in ReportBook.xls
4. General information on ConTest

Ctrl + click to navigate quickly to the sections listed below.

## Contents

1. Quick start guide .....	2
<b>Where do I find the answers of the students?</b> .....	2
<b>Where do I find the points per student?</b> .....	2
<b>How does ConTest calculate the grade?</b> .....	3
<b>What is the quality of my test and of my questions?</b> .....	3
<b>How to use the test results to adjust your scoring and grading?</b> .....	4
2. Test result analysis data and how to use it .....	4
Meta data report .....	4
Test analysis report (entire test): Cronbach's alpha .....	4
Score frequency report (entire test score) .....	5
Item analysis report .....	5
Item advice report .....	7
3. Content of the other tabs in ReportBook.xls .....	9
Version mapping report .....	9
Participant score report .....	9
Grading report .....	9
Explanatory notes .....	9
4. General information on ConTest .....	10

What is ConTest?.....	10
Do I need a ConTest account?.....	10
How do I get a ConTest account?.....	10
Overview of the steps in the ConTest process.....	10
Questions and Contact.....	11

## 1. Quick start guide

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

You can find the unprocessed answers of the students in testdata.csv (see Table 1 for an example).

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Enabled	Participant Form Id	Batch	PgNo	Received	Version	1	2	3	4	5	6	7	8	9	10	
2	1	25282	82	1	29-8-2018	1 D	A	A	D	A	C	B	B	C	A		
3	1	25282	82	18	29-8-2018	1 D	B	B	B	D	C	B	B	A	A		
4	1	25282	82	24	29-8-2018	1 D	B	C	D	A	C	A	B	A	A		
5	1	25282	82	30	29-8-2018	1 A	C	A	D	A	C	C	A	C	A		
6	0	0	25282	82	0	29-8-2018	1 D	A	A	D	A	C	B	B	C	A	
7	1	25282	82	2	29-8-2018	1 D	B	C	D	A	C	B	D	A	A		
8	1	25282	82	3	29-8-2018	1 D	B	A	D	A	C	B	B	B	A		
9	1	25282	82	11	29-8-2018	1 D	B	C	D	D	C	A	B	D	A		
10	1	25282	82	12	29-8-2018	1 D	B	C	D	A	C	B	B	C	A		
11	1	25282	82	16	29-8-2018	1 D	B	A	A	A	C	B	B	C	A		
12	1	25282	82	19	29-8-2018	1 D	C	A	D	A	C	B	B	B	A		
13	1	25282	82	22	29-8-2018	1 D	B	A	D	A	C	B	B	B	A		
14	1	25282	82	28	29-8-2018	1 D	A	A	A	A	C	B	A	B	A		
15	1	25282	82	29	29-8-2018	1 A	B	C	D	A	C	C	B	B	A		
16	1	25282	82	31	29-8-2018	1 D	C	D	C	A	C	A	B	B	A		
17	1	25282	82	7	29-8-2018	1 D	B	A	D	A	C	C	B	B	A		
18	1	25282	82	8	29-8-2018	1 D	A	A	D	A	C	B	B	C	A		

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

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

#### 1. Participant Score Report tab

Table 2: Example of the Participant Score Report in ReportBook.xls

test006 (shuffle-test) - 03.09.2018							
Participant score reportTU Delft\01-Test-Programme/Opleiding\Course-1\test006 (shuffle-test) - 03.09.2018							
Entire Test							
Scores ordered by participant				Scores ordered by score			
Participant	Score	Open	Percentile	Participant	Score	Cum.Freq.	Cum.Perc.
1234567	4,00	0	25,00	1234570	1,00	1	25,00
1234568	6,00	0	50,00	1234567	4,00	2	50,00
1234569	6,00	0	50,00	1234568	6,00	4	100,00
1234570	1,00	1	0,00	1234569	6,00	4	100,00

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.

## 2. Grading Report tab

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

test006 (shuffle-test) - 03.09.2018									
Grading reportTU Delft\01-Test-Programme/Opleiding\Course-1\test006 (shuffle-test) - 03.09.2018									
Entire Test									
Participant	Name		Score			Grade			
			Open	Bonus	Tot.	Bonus	Tot.		
1234567	Student, Ana	✓	0	4,00	0,00	4,00	1,0	0,00	1,0
1234568	Student, Beatrix	✓	0	6,00	0,00	6,00	1,3	0,00	1,3
1234569	Student, Charly	✓	0	6,00	0,00	6,00	1,3	0,00	1,3
1234570	Student, Mo	✓	1	1,00	0,00	1,00	1,0	0,00	1,0

## 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.

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.

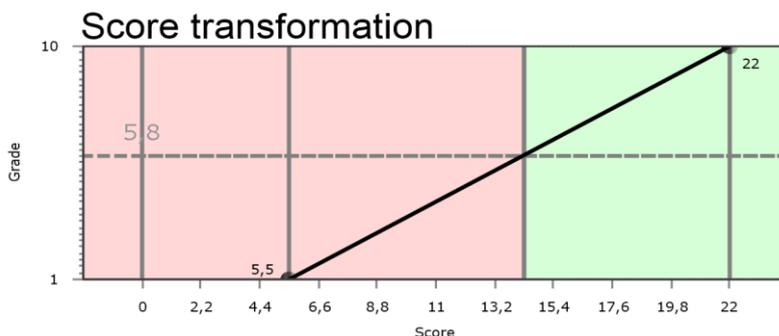


Figure 1: Score-grade transformation for an exam with 22 questions

## What is the quality of my test and of my questions?

Here, the following tabs are most important:

- **Test analysis report:** you can find Cronbach's alpha here, which is a measure for the reliability. 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:** 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)

You can find general information on the quality of the test in the **Test Analysis Report**.

## How to use the test results to adjust your scoring and grading?

In the next chapter, 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.

## 2. Test result analysis data and how to use it

### 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.

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

Here you will find the analysis of the test as a whole. Cronbach's alpha is of interest here. This 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 in case this is the only test in your course
- 0.7 if it is a partial test (for example, a midterm exam)

If Cronbach's alpha is low, check for items (subquestions) with a negative or relatively low  $R_{ir}$ -value, improve their answer model (for example by changing the key), or for example give all students full credits (i.e. basically not allowing the question to influence the final grade, without unfairly taking away points from students who answered the question correctly by deleting the question).

## Score frequency report (entire test score)

Here you will find the frequency per total score: for all existing total scores, the number of students who got exactly that score is plotted. Keep in mind that the maximum value at the horizontal axis is the maximal achieved score, and not the maximum possible score! The information in the graph is also plotted in a graph.

Use this distribution to decide whether or not you would want to correct the overall score, after improving the answer models for the individual questions. 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. 5.8) 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.

## Item analysis report

In this table, the most important indicators of the quality of your test items (questions) are displayed. Please note that below the table, you will find a graphical version, and that on the next tab (Item advice report), and in two pages, a version in written text that will indicate the main (possible) issues.

### P'-value

The p-value is the (normalized) average score for a question.

- For closed-ended questions with 0 or 1 as only possible scores, a p-value of 0.5 indicates that 50% of the students answered the question correctly.
- For open-ended questions, a p-value of 0.5 indicates that students on average got 50% of the score for this question. If the maximum score for this question was 8 points, this means that students on average scored 4 out of 8 points for this question.

p-values should not be too low (not below chance level), and too easy questions (i.e. too high p-values) may be rewritten next year.

### p-value for closed-ended questions

For closed-ended questions, students could get points when they are *only* guessing. For a multiple choice exam with 40 questions of 4 options each (i.e. ABCD), students that would do the exam without looking at the questions would on average score 10 points ( $40/4=10$  points). The p-value for a question where students would have guessed will be 0.25. When there would be 3 options, the p-value for guessing would be 0.33 ( $1/3$ ).

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  $\frac{3}{4}$  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).

For open-ended questions, the p'-value equals the p-value.

### 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 ((sub)questions). The Rir indicates whether students that perform well on this question are also performing well on this item, and whether the students that perform poor on this test, also perform poorly on this item.

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 (0.3 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. Questions with a negative RiR seem to be giving incorrect information about whether students master this course or not (since their score is contradicting the score of other questions). Therefore, you may consider not having them count towards the grade. This you can do by giving all students full points for the question, if reviewing the answer model did not help.
- In general, for open-ended questions, Rir-values tend to be much higher (around 0.6), since there is less noise than with closed-ended questions, where students who know the answer to a question partially, do not get partial credits, but either full or no points. This diminishes the correlation and therefore the Rir-value. In open-ended questions, the points can be awarded in a more precise way (partial credits), resulting in higher Rir-values. Therefore, in general, always look for the questions with the lowest Rir-values and check whether the answer model should be adapted.

### 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 for closed-ended questions

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

### z-value for closed-ended questions

The z-value is a measure of discrimination for all options. It 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.

## Item advice report

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

- **A: 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 revise the answer model accordingly.
- **B: Consider revising the answer model, and revising this question next year.**  
The question did not discriminate between good and poor performing students.
- **C: OK.**  
This easy question helped to discriminate between good performing and poor performing students.
- **D: Please check whether your answer model (key) was correct, and/or if there are multiple answers that may be correct.** This is especially relevant if more students chose one of the distractors, compared to the correct answer.  
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.
- **E: Consider revising the answer model now and revising this question next year.**  
The question did not discriminate between good and poor performing students.
- **F: OK.**  
This question helped to discriminate between good performing and poor performing students.
- **G: Please check whether your answer model (key) was correct, and/or if there are multiple answers that may be 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.
- **H: Consider revising the answer model or giving full points to all students. Revise the question next year.**

This difficult question did not discriminate between good and poor performing students and was made very poorly.

- **I: OK?**

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 report** on page 7).

### 3. Content of the other tabs in ReportBook.xls

#### 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 ConThest site](#).

#### Participant score report

See **Participant Score Report** tab on page 2.

#### Grading report

See **Grading Report** tab on page 3.

#### Explanatory notes

This used to contain the information in this document, and now refers to it.

## 4. General information on ConTest

### What is ConTest?

ConTest is a software based solution to facilitate the processing paper-based exams using answer sheets. This is a replacement for Sonate.

### Do I need a ConTest account?

In the following cases, you will need a ConTest account:

- If you want to use ConTest as a tool to grade and analyse your test results
- If you use multiple test versions
- If you want to change the grading, using ConTest
- If you want to access the scanned answer sheet online
- If you do not want to use a key form

### How do I get a ConTest account?

Mail to [fab-ConTest@tudelft.nl](mailto:fab-ConTest@tudelft.nl) to receive a ConTest account and add the following information:

- Programme / Opleiding
- Course code / Vakcode
- Title course / Titel vak
- Name instructor / Naam docent



**You will have to create your new tests in ConTest yourself.**

### Overview of the steps in the ConTest process

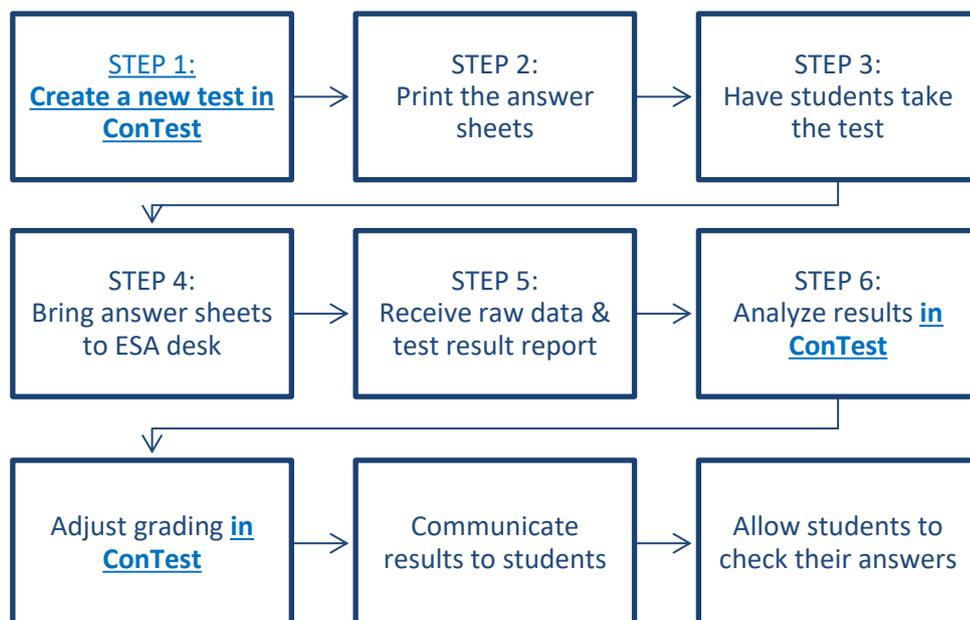


Figure 2. Overview of the process. Underlined parts are only applicable for ConTest account holders.

In Figure 2, you can find an overview of what steps you need to take when you work with paper-based multiple choice exams.

## Questions and Contact

Help with	Function	Contact	Email
General information on ConTest can be found here: <a href="http://www.icto.tudelft.nl/tools/ConTest/">http://www.icto.tudelft.nl/tools/ConTest/</a>			
If you have questions or need more support on educational aspects or interpretation of the test result analysis	Assessment advisor	1. Your faculty's ESA contact person 2. Assessment advisor at Teaching and Learning Services Lisette Harting: <a href="mailto:l.p.harting@tudelft.nl">l.p.harting@tudelft.nl</a>	
Questions about functionalities or bugs/errors in ConTest. Or if you need support on technical or account related issues	Functional application manager	Kasja de Jong	<a href="mailto:Fab-ConTest@tudelft.nl">Fab-ConTest@tudelft.nl</a> 015-2784590
Status of processing your tests after handing it over	ESA-desk	Leon van der Drift Wil van Leeuwen-Meyer	<a href="mailto:ConTest-os@tudelft.nl">ConTest-os@tudelft.nl</a>