

## Appendix C. Why do we standardise grades?

### Standardised grades make comparisons easier

In Formas' calls, applications are assessed by several reviewers. All reviewers receive the same instructions on how to grade applications. Even so, reviewers tend to use the grading scale in different ways: they may spread their grades differently and use different parts of the scale.

To adjust for these differences, Formas uses standardised grades. In practice, standardisation captures whether a reviewer has graded an application higher or lower than the other applications they reviewed in the call. This means different reviewers' assessments can be compared in a more consistent way.

### Standardised grades describe grades in relative terms

Put simply, a standardised score shows how far above or below each reviewer's average score a particular score is. If the standardised score is positive, it means the reviewer has ranked the application as better than average (see Table A.2). The higher the value, the higher the reviewer has ranked the application. If the standardised score is 0, the application is of exactly average quality. If the reviewer instead judges the application as below average, the value becomes negative. The lower the value, the lower the reviewer has ranked the application. This also applies to the average standardised score, which combines the standardised scores from all reviewers into a single value (Table A.5).

### How we calculate standardised grades

#### Step 1: Standardise all grades from each reviewer

For each reviewer, we calculate the reviewer's average grade and the standard deviation of their grades, based on all applications the reviewer assessed in the call (Table A.1). Standard deviation is a statistical measure of how much values vary around an average. Technically, a value of +1 means a grade is one standard deviation above the reviewer's average. The dataset includes all grades a reviewer has given in a call. If a reviewer assesses 20 applications against four criteria, the standard deviation will be based on 80 values.

**Table A.1. Example of reviewers' average grades and standard deviations, based on all applications the reviewer assessed.**

| Reviewer | Reviewer's average grade | Reviewer's standard deviation |
|----------|--------------------------|-------------------------------|
| Anna A   | 5,45                     | 1,32                          |
| Benny B  | 4,71                     | 0,88                          |
| Conny C  | 5,90                     | 1,03                          |

When the reviewer's average score and standard deviation have been calculated, the next step is to produce a standardised value for all scores the reviewer has given. This is done by taking the original score for a criterion minus the reviewer's average score, and then dividing the result by the reviewer's standard deviation.

Summarised as a formula, it looks like this:

$$z = \frac{x - \mu}{\sigma}$$

- $z$  is the standardised grade
- $x$  is the original grade
- $\mu$  is the average of all grades the reviewer has given
- $\sigma$  is the standard deviation of all grades the reviewer has given

Table A.2 provides an example calculation for the reviewer Anna A and for an application where she gave the original grades ( $x$ ) of 5, 3, 5 and 4 for the criteria. Table A.1 shows that Anna A's average grade across all reviewed applications ( $\mu$ ) is 5.45 and her standard deviation ( $\sigma$ ) is 1.32.

For criterion 1, Anna's standardised grade ( $z$ ), is therefore  $= (5 - 5,45) / 1,32 = -0,34$ . In other words, Anna's grade for criterion 1 is slightly lower than her average grade for this criterion.

**Table A.2. Example of how the standardised grade for each criterion is calculated.**

|             | Original grade | Standardisation                 | Standardised grade |
|-------------|----------------|---------------------------------|--------------------|
| Criterion 1 | 5              | $z = (5 - 5,45) / 1,32 = -0,34$ | -0,34              |
| Criterion 2 | 3              | $z = (3 - 5,45) / 1,32 = -1,86$ | -1,86              |
| Criterion 3 | 5              | $z = (5 - 5,45) / 1,32 = -0,34$ | -0,34              |
| Criterion 4 | 4              | $z = (4 - 5,45) / 1,32 = -1,10$ | -1,10              |

We repeat this for all grades given by all reviewers until every grade has been converted into a standardised grade. Table A.3 shows an example of what the standardised grades can look like for an application assessed by four reviewers.

Note that the standardised grades are expressed on a different scale from the original grades. This means they cannot be combined with, or compared to, the original grades.

**Table A.3. Example of standardised grades for reviewers across criteria, produced by applying the formula.**

| Case no.   | Reviewer | Standardised grades for criteria 1–4 |       |       |       |
|------------|----------|--------------------------------------|-------|-------|-------|
|            |          | 1                                    | 2     | 3     | 4     |
| 2030–12345 | Anna A   | -0,34                                | -1,86 | -0,34 | -1,10 |
| 2030–12345 | Benny B  | 1,46                                 | -0,80 | 0,33  | 0,33  |
| 2030–12345 | Conny C  | -0,87                                | -0,87 | 0,10  | 1,07  |
| 2030–12345 | Donna D  | 1,37                                 | 1,37  | 0,68  | 1,37  |
| ...        | ...      |                                      |       |       |       |

## Step 2: Calculate an average standardised grade for each reviewer

In the next step, we calculate an average standardised grade for each reviewer, based on the standardised grades for each criterion (Table A.4).

**Table A.4. The average standardised grade from a reviewer is calculated by adding the standardised values for each criterion and then dividing the total by the number of criteria. The reviewer's average standardised grade is shown in the rightmost column.**

| Case no.   | Reviewer | Standardised grades for criteria 1–4 |       |       |       | Average standardised grade |
|------------|----------|--------------------------------------|-------|-------|-------|----------------------------|
|            |          | 1                                    | 2     | 3     | 4     |                            |
| 2030–12345 | Anna A   | -0,34                                | -1,86 | -0,34 | -1,10 | -0,91                      |
| 2030–12345 | Benny B  | 1,46                                 | -0,80 | 0,33  | 0,33  | 0,33                       |
| 2030–12345 | Conny C  | -0,87                                | -0,87 | 0,10  | 1,07  | -0,14                      |
| 2030–12345 | Donna D  | 1,37                                 | 1,37  | 0,68  | 1,37  | 1,20                       |
| ...        | ...      |                                      |       |       |       |                            |

## Step 3: Calculate an average standardised grade across all reviewers

Once the average standardised grades have been calculated for all reviewers for an application, we combine these into an average standardised grade across all reviewers (Table A.5).

**Table A.5.** The average standardised grade from each reviewer for an application is combined into the average standardised grade across all reviewers. The average standardised grade for an application is the mean of the reviewers' average standardised grades, in this case  $(-0.91 + 0.33 - 0.14 + 1.20) / 4 = 0.12$ .

| Case no.   | Reviewer | Average standardised grade from each reviewer | Average standardised grade across all reviewers |
|------------|----------|---|---|
| 2030–12345 | Anna A   | -0,91   | } 0,12  |
| 2030–12345 | Benny B  | 0,33  |   |
| 2030–12345 | Conny C  | -0,14   |   |
| 2030–12345 | Donna D  | 1,20  |   |

It is possible to convert standardised grades back to the usual 1–7 grading scale, but that would mean applications are spread across the full 1–7 range based on how good their standardised grades are. The application with the lowest standardised grade would then get a grade of 1—even if reviewers' original judgement was closer to a 3.

## Standard deviation of standardised grades

### The standard deviation of standardised grades shows disagreement

When you standardise scores, the standard deviation usually decreases. But it does not disappear completely. And in some cases, it can still be relatively high.

For the standard deviation of the original scores, it is impossible to know whether a high standard deviation reflects disagreement between reviewers – or whether it is the result of reviewers using the scoring scale differently. When we standardise grades, we remove the part of the uncertainty that comes from different use of the scoring scale. This means that a high standard deviation in the standardised scores indicates that reviewers do not agree about the application's qualities. In other words, the standard deviation of the standardised scores is an effective way of identifying disagreement between reviewers.