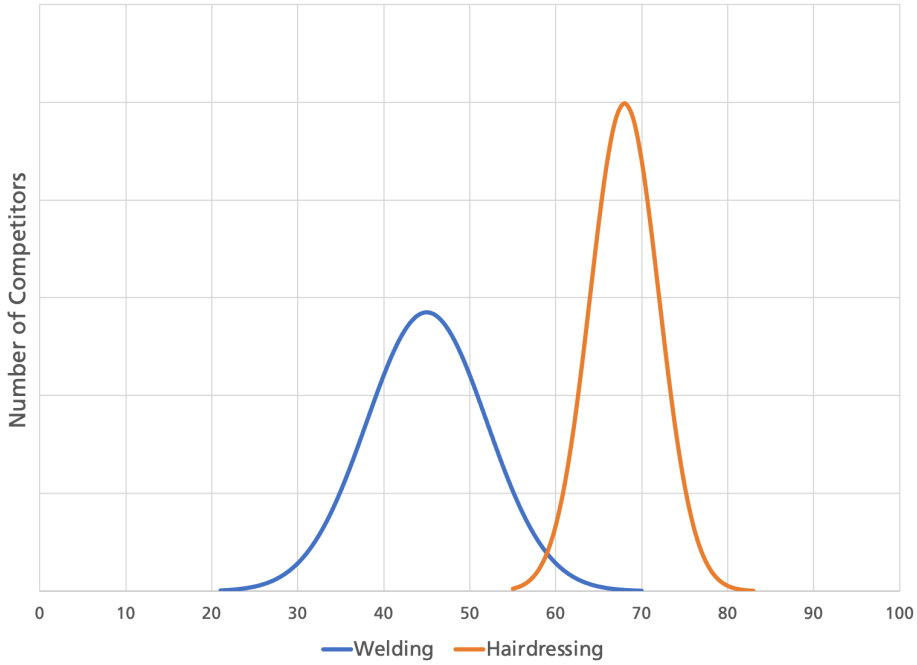
WorldSkills Scale

What is the WorldSkills Scale?

At a WorldSkills Competition, Competitors in each skill competition complete a Test Project and receive a mark out of 100 for their work. This determines the rank of each Competitor within their skill competition.



Mark distributions for Welding and Hairdressing

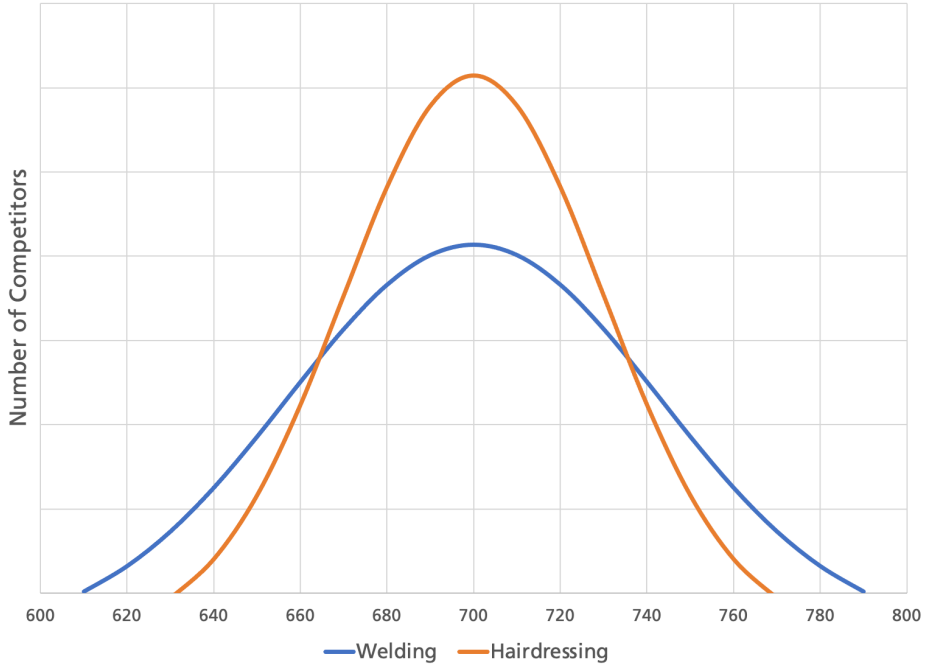
For example, in the chart above we can see the mark distributions for two skills. The top mark for Hairdressing is 83.55, while the top mark for Welding is 70.10. Each of these Competitors are the best in their skill and will receive a gold medal. But how do we determine if the hairdresser or the welder was the better overall Competitor? We cannot use their 100 scale marks because they completed different Test Projects with different Marking Schemes. One of the projects may have been more difficult than the other, or one set of markers may have marked more leniently than the other.

To compare one Competitor's performance to another Competitor's performance they both need to be measured on a common scale. This common scale allows Competitors to be compared across skill competitions, either in the same event or between any event that uses the same common scale. For example, a Competitor's performance at a WorldSkills International Competition could be compared to a Competitor's performance at a previous Competition, provided that all the conditions are similar. This allows Members to track elements of their progress over time in each skill competition.

To place a Competitor's result on this common scale, their performance is compared to the rest of the Competitors in their own skill competition. By this measure the Competitor who is furthest from the middle of the cohort in their skill competition can be considered the best overall performer. For this procedure to work optimally, the Test Projects and Marking Schemes must be able to adequately distinguish between the Competitors' capabilities and determine where they lie on the spectrum between competent and excellent.

The common scale used by WorldSkills is referred to as the WorldSkills Scale. The Competition Information System (CIS) applies a formula to the marks of each skill competition to produce a result for each Competitor on the WorldSkills Scale that can be used across all skill competitions. Results on the WorldSkills Scale will generally

fall between 600 and 800 points[1], although there may be outliers that fall outside this range. WorldSkills International also uses this common scale when awarding medals and awards such as the Albert Vidal Award for the Competitor with the highest number of points on the WorldSkills Scale.



WorldSkills Scale score distributions for Welding and Hairdressing

Returning to the earlier example, the chart above shows the marks for Hairdressing and Welding converted to WorldSkills Scale results. When comparing the top results in each skill it is evident that the Welder achieves a higher score and is therefore the better performing Competitor, even though their 100 scale mark was lower than the Competitor from Hairdressing.

Why have a WorldSkills Scale?

Put simply, as part of the WorldSkills Competition, WorldSkills and its Members want to be able to:

- compare the performance of Competitors at the same Competition
- compare certain elements of the performance of Competitors with previous Competitions
- celebrate and recognize the Competitor that achieves the overall best result out of all Competitors in all skill competitions.
- gather valuable management information to advance the WorldSkills Competition and model best practice in vocational assessment.

In order to do this, at a WorldSkills Competition, Competitor marks out of 100 are converted to a common scale used across all skills. This common scale is used to award medals, and is also used for comparing the performance of Competitors across different skill competitions and different events.

[1] For information on the difference between marks, scores and points please read the guidance note <WSI_SC_guidance_note_marks_scores_points> available at https://worldskills.org/internal/learning-resources/guidance-notes/

History of the WorldSkills Scale

Prior to the introduction of the WorldSkills Scale (introduced for WSC2017), results at a WorldSkills Competition were calculated using the "500 scale" calculation. The 500 scale relied on the use of the statistical mean and standard deviation to create a z-score – an indicator of how much better or worse a Competitor performed compared to the average Competitor in their skill competition. To produce a result that was more meaningful, these z-scores were multiplied by 30 and added to 500 to create a result that usually fell between 400 and 600 points with the average mark achieving a result of 500.

As the Membership of WorldSkills grew the competence level of Competitors attending the WorldSkills Competition became more widely varied. There were increasing numbers of poor performing Competitors and this exposed some weaknesses with the 500 scale calculation. The use of the mean and standard deviation is vulnerable to the influence of outliers (for example where one or more Competitors get a very low mark) and skewed data sets (those not conforming to the traditional "bell curve"). This can cause a distortion of the results for all Competitors in these skill competitions and prevent high performing Competitors from achieving a high number of points on the 500 scale.

These apparent weaknesses prompted WorldSkills International to commission a report from an independent statistician to review the 500 scale. The report contained a proposal for a new calculation to convert Competitor marks – the WorldSkills Scale. The report was thoroughly reviewed by the WorldSkills Secretariat and the calculation tested with the data from numerous competitions. The report was reviewed and approved by the Board and then presented to the Competitions Committee at GA2016, with its first implementation at WSC2017.

The WorldSkills Scale replaces the z-score calculation with a more robust one to create a "modified z-score". Rather than using the mean to measure the central tendency of marks, the WorldSkills Scale makes use of the median, which is not influenced by outliers and is largely immune to asymmetrical data. To measure the dispersion of marks from the median, the Median Absolute Deviation (MAD) is used. The MAD is then multiplied by a standard statistical "consistency constant" of 1.4826 to approximate the standard deviation. This data is used to create a "modified z-score", which gives an indicator of Competitor performance in a similar way to the conventional z-score in the 500 scale calculation.

Once we have a modified z-score, we can then create a more meaningful result for Competitors using the same technique as the 500 scale, by multiplying the result by 30 and adding 700. This will produce results that generally fall in the range of 600 to 800 points, although there will be a few that fall outside this range. This range of points was selected to create an obvious distinction to the 500 scale. This combination of the new calculation and central point of 700 is called the WorldSkills Scale. Results on the WorldSkills Scale cannot be compared to results on the 500 scale because they use different calculations. Having points in a different range makes this clear.

WorldSkills Scale Formula

$$WorldSkills\ Scale\ Score = \left(\left(\frac{100\ scale\ mark - median}{MAD\ \times\ 1.4826}\right) \times\ 30\right) + 700$$

Median – If all marks are arranged in order, the median is the middle mark. For example, if there are 23 Competitors the median mark belongs to the Competitor that came 12th.

MAD – The Median of Absolute Deviations. To determine the MAD value, the differences between each Competitor's mark and the median mark are collected and sorted from lowest to highest. The MAD is the median of that set of numbers.

1.4826 – This is a recognised adjustment factor used to approximate the standard deviation.

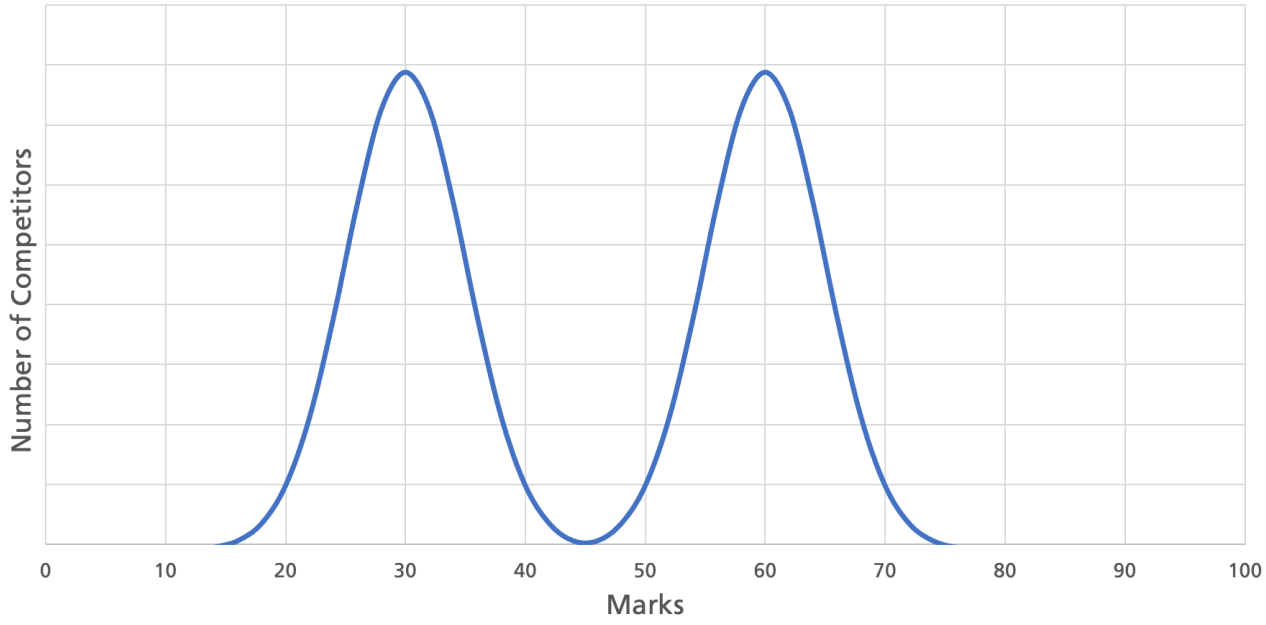
30 – Results are scaled by 30 to spread the results and reduce the chance of ties when rounding to a whole number.

700 – The median mark will receive a result of 700 points. All other results will be distributed around 700 points.

Weaknesses of the WorldSkills Scale

There are some unusual arrangements of marks that can prove problematic when being transformed to the WorldSkills Scale. These also caused problems for the 500 scale calculation that was used prior to the WorldSkills Scale, and indeed would be problematic for any calculation used.

Bi-modal mark distributions



Example of a bi-modal mark distribution

When there are two distinct clusters of marks as in the chart above, it is referred to as a bi-modal distribution. This can occur for a number of reasons, such as having a large number of Competitors who are not sufficiently competent in their skill, or having a Test Project that gives an advantage to a particular group of Competitors.

In these mark distributions the median mark usually falls in one of the two clusters, and there is a good chance that the MAD value will be quite small. This will significantly affect the results of the Competitors in the other cluster.

For example, if the median were to fall in the cluster with lower marks and the MAD value was low, it would exaggerate and push up the results of the Competitors in the higher cluster.

In general, there are no good statistical methods for handling bi-modal distributions, and in most cases the recommendation is to identify and separate out the different individual distributions present. Having two separate sets of results for a skill competition would not be a suitable solution.

The best way to guard against this is to have a good quality Test Project and Marking Scheme, and have a cohort of Competitors who are sufficiently competent.

Low number of Competitors

Skill competitions that have a low number of Competitors are particularly susceptible to unusual results on the WorldSkills Scale. This is because the median and the MAD value are determined by only a handful of data points. This increases the chances of a having both a small MAD value, and marks that are a significant distance from the median. The resulting effect is that there will be some extremely high or extremely low results on the WorldSkills Scale.

When trying to draw insights from any dataset, the more data points available the more reliable the results will be. When there are too few data points it is not possible to draw any meaningful conclusions from the data set. For a WorldSkills competition, having too few Competitors makes any comparative calculation (the WorldSkills Scale or any other calculation) unreliable.

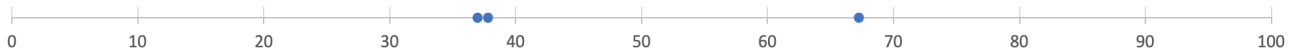
At a WorldSkills International competition, our Competition Rules specify a minimum of six Competitors[2] for a skill. So far this has proven to be an adequate number to achieve reliable results for a cohort of Competitors who are competent and where the Test Project is assessing excellence. For regional or national competitions where there is a wide variance in the skill level of Competitors it may be appropriate to have a higher minimum number of Competitors.

It is recommended that skill competitions with a low number of Competitors be excluded from any awards for the best performing Competitor of the Competition (similar to the Albert Vidal Award at the WorldSkills Competition).

EuroSkills Graz 2021

An example of the effect that low Competitor numbers can have on results took place at EuroSkills Graz 2021. The best in Europe (now called the Jos de Goey Award) went to the Gold medal winning team in the Mobile Robotics skill competition. They had a result of 1401 on the WorldSkills Scale. This disproportionately high number of points was the result of an unusual distribution of marks in a skill competition with a total of only three Competitors (or in this case three teams of Competitors).

The marks in the skill were 36.97, 37.82 and 67.25. This distribution results in a median (the middle mark) of 37.82. The median of absolute deviations (MAD) for this distribution is 0.85.



100 scale marks for Mobile Robotics at EuroSkills Graz

[2] Competition Rules v9.1 section 3.4

The WorldSkills Scale is a measure of how each Competitor performs in relation to the other Competitors in their skill competition. In this skill competition, it appears that one Competitor has performed significantly better than the rest of the cohort of Competitors. The top Competitor is 29.43 marks from the median mark, while the next closest to the median is 0.85 marks below it. The top Competitor is nearly 35 times further from the median than the next closest Competitor. If a Competitor outperforms their peers by this much they will obviously achieve a very high number of points on the WorldSkills Scale.

However, the fact is we do not have sufficient data to draw a conclusion about the performance of the top Competitor in this skill competition. As pointed out in the previous section, skill competitions with a low number of Competitors are susceptible to unusual mark distributions because of a lack of data. If there were more Competitors it would be highly likely that there would be a more even spread of marks and the MAD value would be significantly higher, giving us a better reflection of the Competitors' performance on the WorldSkills Scale.

These kinds of statistical anomalies are *possible* in any skill competition, but are much more likely to occur where there are low numbers of Competitors. If a Competitor achieved this level of separation from a larger cohort of say 20 Competitors, we would confidently say that they deserved their high result and were a worthy winner of the Jos de Goey Award.

As per the last paragraph of the preceding section, it is recommended that in future EuroSkills events, skill competitions with low numbers of Competitors be excluded from the Jos de Goey Award.

WorldSkills ASEAN 2023

A similar problem with a low number of Competitors was encountered at WorldSkills ASEAN 2023. In this case the distribution resulted in a Competitor receiving a very low result, due to a tight cluster of marks with a single outlier.

The marks for the skill were 99.25, 98.65, 97.55 and 45.75. This gives a median mark of 98.10 and coincidentally once again produces a MAD of 0.85. This puts the lowest Competitor more than 40 times the modified z-score lower than the median mark. In a normal distribution more than 99% of marks should fall within three z-scores either side of the middle of mark. The mark of the outlier Competitor should actually produce a negative result, but the CIS limited the result to zero.