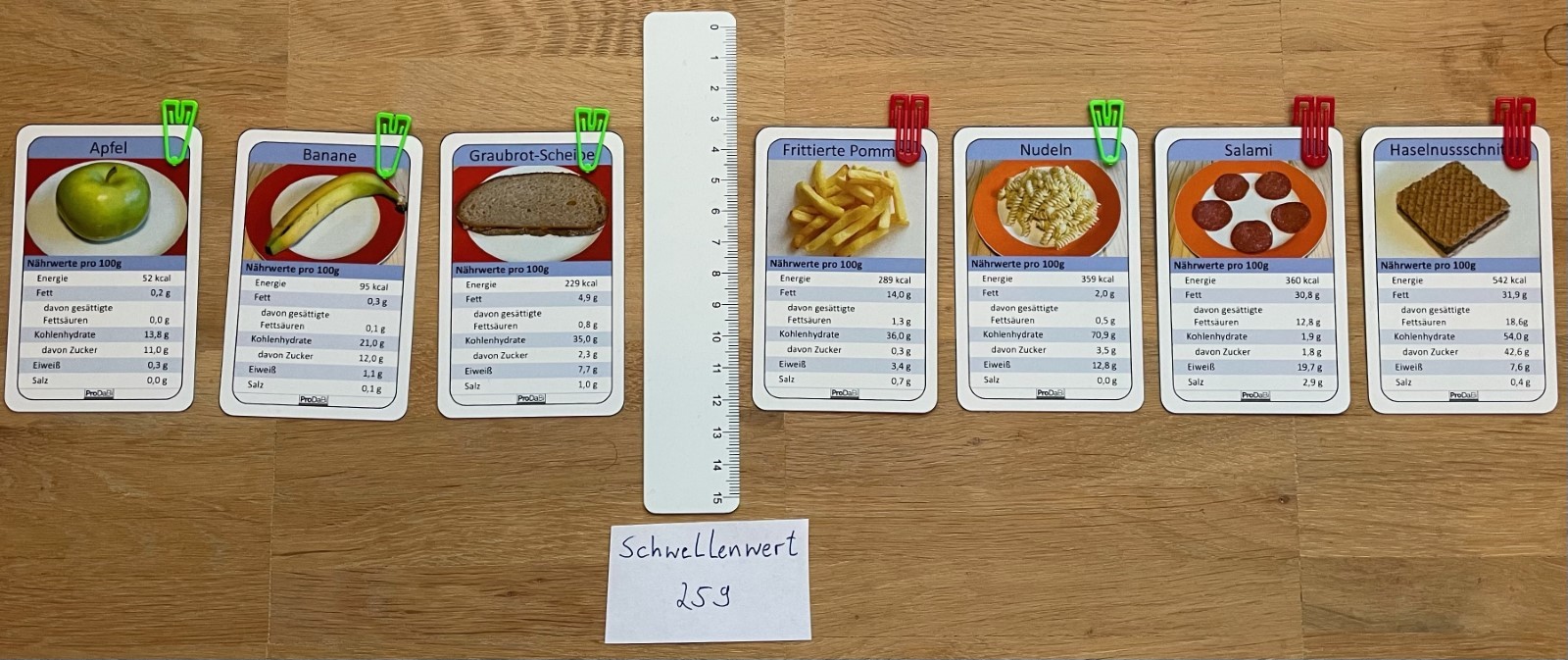
Further information on lesson 4

Notes on the ruler method

The ruler/pencil method is used to work systematically with data maps and the search for a suitable threshold value and to visually support the threshold value search. The procedure works as follows:

* The data cards are sorted in ascending order according to a characteristic (if there are too many cards, move 2 tables together if necessary).
* A ruler is placed in the distributions as a symbolic threshold value (between two cards) and then the number of correctly classified cards is counted. The majority value is formed on both sides. In the illustration, the majority value on the left is "rather recommendable" as the majority of the cards have green brackets. The majority value to the right of the ruler is "not recommended" because there are more red than green brackets. On both sides, the cards that deviate from the majority are classified as incorrect. There is only one misclassified card in the illustration: the noodles (green) on the right-hand side, so we have a misclassification count of 1.



Didactic note on the ruler method

The ruler method is used to determine the best threshold value from a set of considered threshold values. Not all possible threshold values need to be considered. Testing all threshold values (as a computer would do) is tedious with a large number of cards and is therefore not advisable for teaching practice. But with a sense of proportion and a little trial and error, you can also find a locally best threshold value, measured by the number of misclassifications.

Finding a threshold value by eye

To avoid having to try out all possible threshold values using the ruler method, it is advisable to work with a sense of proportion. A sense of proportion is built up through experience and can be supported by the following tips:

* You can select an initial threshold value by eye in order to compare various other threshold values based on this. The cards must always be sorted according to a characteristic.
* Various strategies can be used to select the first value:
  + The first threshold value selected does not necessarily have to be the best!
  + You can use the distribution of the colors (brackets) as a guide. There are usually many red cards further to the right and many green cards further to the left. This visual impression of the distribution can be used to set an initial threshold value so that the left partial data set is "as green as possible" and the right one "as red as possible". It is very important to calculate the misclassifications for each threshold value and not to rely on the visual impression. In this way you can
* Threshold values can then be tried out in the gaps next to the first threshold value.
* Studies have shown that students can use this strategy to make a preselection of favorable threshold values. The best threshold is then selected from the preselection using the misclassification criterion.

Set the value of a threshold

The ruler only indicates a gap between two cards and not yet a threshold value. In a further step, the threshold value must be selected from an interval. For example, the interval [229; 289] is shown at the top of the image. A value for the threshold value must now be selected from this. For the sake of simplicity, this can be any value of the half-open interval [229; 289], because the associated subgroups are formed from the set of data that is less than or equal to the threshold value and greater than the threshold value.

For an algorithm, for example, one of the following options is used to select the threshold value:

* The mean value (arithmetic mean) of the marginal values is selected.
* The smallest possible interval value is selected.