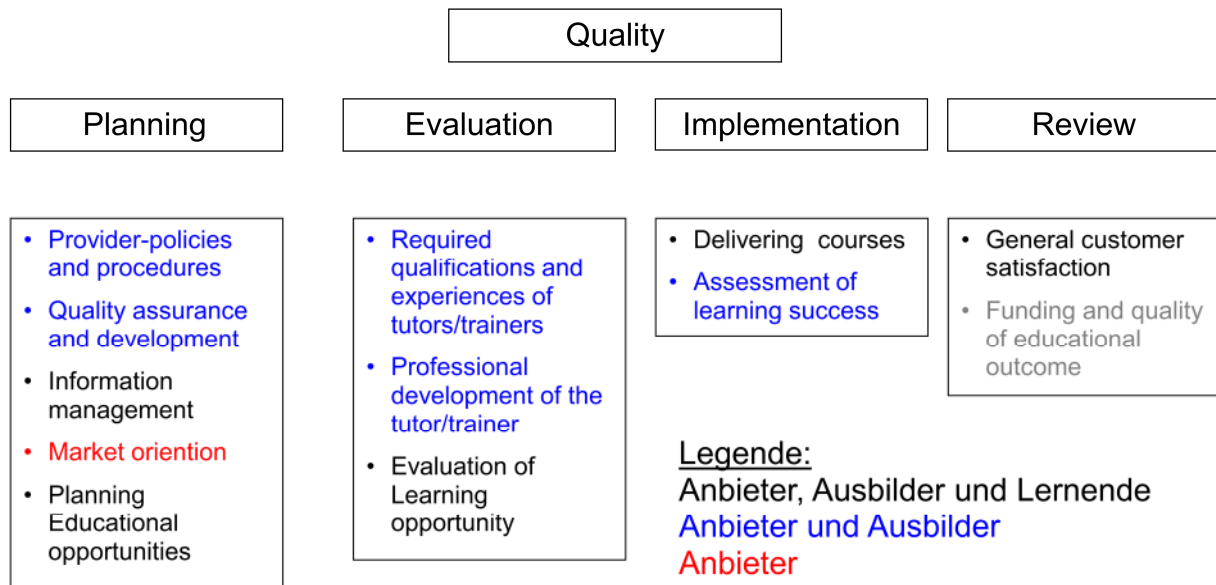


The Online-Instrument

1. The structure of the questionnaire

The structure of the questionnaire, its scales and its items is orientated at the Deming-wheel, which follows the guidelines of the EU concerning vocational education and training (<http://www.eqavet.eu/index.html>). The first version of the instrument was based on it also following existing instruments like SPEAK or Eduqua and was tested in a first step. You can get an overview of the structure following the next graphic. The color makes clear which target groups were asked to answer which questions (follow the legend within the graphic).



Basing on the experiences made in the first pilot phase and the feedback given there, a basic item-analysis was discussed and revised on our meeting in Durham in March 2012. The scale 'Funding and quality of educational outcome' was eliminated out of the questionnaire and many items were reformulated. In doing this not only the problems concerning the language and expressions were eliminated. Especially the correspondence of the items between the target groups was improved. Within the scales which were to be answered by different target groups the items were reformulated regarding their different perspectives. So one of the most serious point of critique was focused. Also a new possibility of analysis was made possible orientated at the 360 degree feedback method (of which just 270 degrees were realized).

The so revised instrument was tested once again to document its properties. For this 120 providers, 51 teachers and 58 learners participated. The properties of the instrument are described in the following alongside two criteria: The internal consistency of the scales is analyzed by an item-analysis. An additional factor analysis documents whether the empirical structure of the questionnaire meets its theoretical structure (see the graphic above). In the second step the feedback given by the test persons is summarized and analyzed.

2. Properties of the questionnaire

2.1 Item-Analysis

The item-analysis was done following the so called classical test-theory as it is implemented in the procedure 'Reliability' of SPSS. The following table gives an overview about the results for every target group. You can see the reliability-coefficient (Chronbach's Alpha) listed and the number of questionnaires (in brackets) which was the basis for the calculation.

Scale	Provider	Trainer	Student
Provider-policies and procedures	0,77 (68)	0,944 (35)	
Quality assurance and development	0,665 (59)	0,917 (31)	
Information management	0,573 (57)	0,874 (31)	0,796 (20)
Market orientation	0,601 (56)		
Planning Educational opportunities	0,766 (57)	0,828 (31)	0,725 (17)
Required qualifications and experience of tutors/trainers	0,868 (56)	0,913 (30)	
Professional development of the tutor/trainer	0,722 (56)	0,916 (31)	
Evaluation of learning opportunity	0,794 (56)	0,862 (28)	0,895 (17)
Delivering courses	0,791 (56)	0,872 (32)	0,874 (15)
Assessment of learning success	0,507 (47)	0,804 (32)	
General customer satisfaction	0,786 (56)	0,886 (32)	0,775 (18)

It is remarkable to find such a big difference in the number of people participating in the project (e.g. 120 providers) and the number of questionnaires being in the analysis (between 47 and 68). The reason is that some of the participants were answering very sparsely on the one hand and that only those questionnaires could be analyzed with not more than one missing answer per scale on the other. This missing answer was replaced by the mean of the other answers belonging to the same scale. Following common conventions a Chronbach's Alpha value of 0.6 can (just) be regarded as satisfactory, a value greater than 0.8 as good to very good. The overview shows that the scales as answered by trainers and learners throughout can be regarded as good or very good. On the side of the Providers the scale 'assessment of learning success' is not satisfactory. It turns out here that the coefficient is improving leaving the item 'Where learning/training is non-accredited procedures are in place to recognize and record progress and achievement' to 0.601 – that means up to the minimum of the threshold value. Looking at the text of the item this seems to be plausible. Dropping the item 'Are partners engaged and consulted about the required educational provision and course content e.g. Employers/Sector Skills Councils?' in the scale 'information management the coefficient improves to 0.678. Because the scale 'market orientation' only consists of two items the coefficient there cannot be improved by dropping an item. Looking at the text both questions seem to be important so that the scale can just be kept in the questionnaire.

The factor analysis was not calculated using all the items of the questionnaire. Only those items were used belonging to the scales of an element out of Deming-wheel. The reason for that is that lies in

the small survey sample. Generally speaking between 60 and 70 percent of the item-variance could be reproduced by the factors – dependant on the Deming Element or the target group. But the items not always were allocated to the scale they should have been allocated to from the theoretical point of view. There were differences between the target groups. For example the factor analysis of the element ‘evaluation’ results in four factors of which the first two (‘Required qualifications and experience of tutors/trainers’ and ‘Professional development of the tutor/trainer’) are reproduced, but the third one (‘Evaluation of learning opportunity’) is divided into two factors. The first one is related to quality standards, the second one to feedback to the trainers. The analysis of the same questions given by the trainers results in two factors. But here there are less items.

The results concerning the Deming-element ‘Planning’ are to be regarded in a much more different way, but they also go in a plausible direction. For pragmatic reasons we want to keep the theoretical classification because the reliability of the scales would be too bad dropping one of the items. Besides this one should keep in mind that the intention of the tool is not to test quality features but to determine necessary actions in quality improvement. So the quality criteria of the tool do not have to be treated so strictly as in a test.

2.1 Assessment

The comments of the participants on the questionnaire are very informative for further developing the tool. First of all we present an overview of the scaled feedback (please keep in mind that the test-version not all of the persons filling out a questionnaire gave us a feedback).

	Excellent	Good	Adequate	Inadequate
Quality of questions	3	22	6	5
Clarity of questions	3	14	13	6
Simplicity of questions and lack of scientific jargon/ terminology	4	17	7	8
Comprehensibility of presented information	3	18	10	5
Relevance and applicability to the target groups	5	11	14	5
Time needed to complete it	8	16	7	2

It turns out that the instrument in general is rated rather good. Not so very good are the comments on the clarity of the questions and its applicability for the target groups. A closer look at the answers – especially at those coming from continuous bad judgements – shows that the devaluations resulted mainly out of a bad language. It was criticized that there was a jargon-like language style, which was too complicated and partly used bad expressions. The critique mainly focuses the English version of the questionnaire and is surprising at the first glimpse because all of the questionnaire-versions were controlled and adopted by mother-tongue collaborators and because we were very careful on our project meetings to use a language that is concrete and easy to understand. A closer look at the answers shows that the critical statements are coming from those participating in early rounds. Later such a critique could not be found any longer. Here the opposite – at least sometimes – was true: Some of the participants emphasized the user friendly construction of the instrument. Because our Irish partner got lost the Irish interviews were done by other persons at the end of the project. So it

is obvious that the language problems in the beginning were solved. Because of this incident a country specific analysis cannot be done because time-specific and country-specific aspects are confounded.

Besides the feedback to the questions the people were asked to judge the possible use of the instrument. Here the answers were very different. Some of the participants did not give any answer or just said yes or no. But some of them wrote a lot. For obvious reasons it is only possible here to analyze the given answers. But this was the majority. About half of the people asked (not answering!) felt encouraged to engage in actions of quality assurance more than they do now because they were motivated to think about quality, because they have learned to estimate the use of feedback or because they became aware of the possibility to compare several instruments to get to conclusions about quality which are more valid. Also the majority of the participants found the instrument helpful for their organization. The reasons for that were different: To participate the collaborates in reflection about quality, to stress models of good practice, to reflect changes or to collect important data. Most of all the accentuation of good practice was mentioned. The answers to the third question were merely disillusionating. Here the people were asked to recon how their collaborates think about benchmarking. Here the answers reached from very positive judgements (opportunity to take part in decision-making) over 'not everyone will be happy about that' to dismissive answers guessing that there would be an agreement on a verbal level but no enthusiasm on the action level. Corresponding to that the answers about a supposed support were different, too. They reached from a very positive agreement to a clear refusal.

Finally hints for further developing the instrument were given. The advice to improve the language was already mentioned. In addition there were a lot of compliments ('all aspects of evaluation are covered', 'it's easy to use'), critique (in handling the system; e.g. it was mentioned that the end of the questionnaire was not clearly indicated) and some proposals for new contents. E.g. it was proposed to implement a comparison between target and current status or the possibility to prioritize actions of quality assurance. Also it was criticized that there is no possibility to evaluate the date. This was implemented at the end of the project as it is mentioned in the next passage.

3. Possible evaluations

The possibilities in evaluating the data are orientated at the possible benchmarks and at the conception of the instrument. The following possibilities are implemented:

- 1 Analysis of change (time series analysis)
- 2 Analysis using a reference group
- 3 Analysis of correspondencies (=comparing answers generally) between providers, trainers and learners)
- 4 Analysis using a norm (benchmarking analysis)

All of the four possibilities are implemented in the web-site and there are documents explaining the evaluation procedures, how they work and how the results can be interpreted and documents explaining the way to get those results. Because those documents are available via our web-Site it is not necessary to explain the procedures and the way they can be used here.