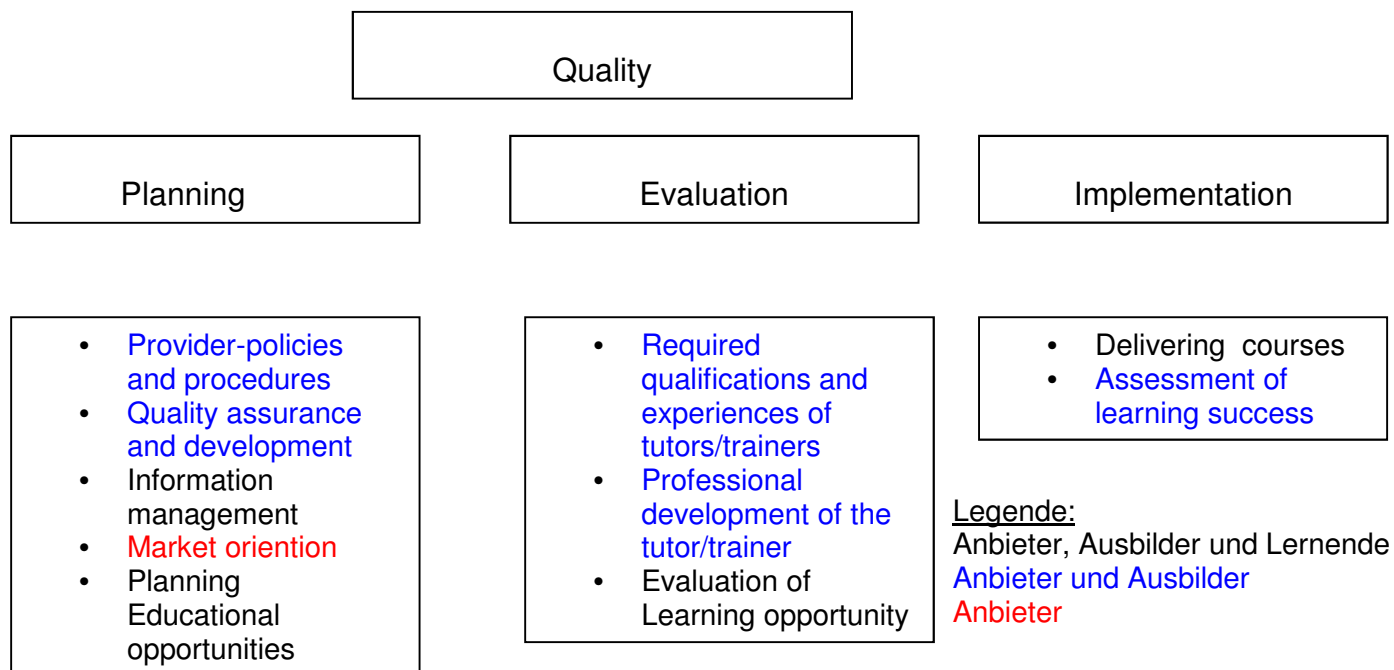


Das Online-Instrument

1. Der Aufbau des Fragebogens

Die Struktur des Fragebogens, seiner Skalen und der Items orientiert sich am Deming-Kreis, der unter anderem auch den Qualitätsrichtlinien der EU im Rahmen der Berufsbildung zu Grunde liegt (<http://www.eqavet.eu/index.html>). Eine erste Version des Instruments wurde auf dieser Grundlage und unter Bezugnahme auf bereits existierende Instrumente (SPEAK, Eduqua) erstellt und erprobt. Die Struktur des Instruments lässt sich der folgenden Übersicht entnehmen. Aus der farblichen Darstellung ist ersichtlich, welche Gruppen zu den jeweiligen Skalen Stellung nehmen konnten (s. die Legende innerhalb der Grafik).



Auf der Grundlage dieser Erfahrungen, die in einer ersten Pilot-Phase gemacht wurden, der dabei erhobenen Kommentare der Probanden und einer ersten Item-Analyse wurde der Fragebogen auf dem Treffen in Durham im März 2012 eingehend diskutiert und revidiert. Die Skala 'Funding and quality of educational outcome' wurde aus dem Fragebogen entfernt und viele Items neu formuliert. Dabei wurden nicht nur sprachliche und begriffliche Probleme beseitigt, sondern es wurde vor allem auch auf eine verbesserte Korrespondenz der Items zwischen den verschiedenen Zielgruppen der Befragten geachtet. So wurden innerhalb der Skalen, die von mehreren Gruppen zu bearbeiten waren, die Fragen so formuliert, dass sie aus der jeweiligen Perspektive beantwortet werden konnten. Damit wurde nicht nur einem häufig genannten Kritikpunkt der Rückmeldungen Rechnung getragen, es wurde so auch explizit eine neue Auswertungsmöglichkeit geschaffen, die sich von der Sache her am sog. 360-Grad-Feedback orientiert (wovon hier dann lediglich 270 Grad realisiert wurden).

Das so überarbeitete Instrument wurde noch einmal getestet, um seine Eigenschaften zu dokumentieren. Dazu wurden insgesamt 120 Anbieter, 51 Ausbilder und 58 Lernende befragt. Die Eigenschaften des Instruments werden im Folgenden anhand zweier Kriterien beschrieben. Erstens wird anhand einer Itemanalyse die interne Konsistenz der Skalen ermittelt und mit Hilfe einer

Faktorenanalyse die unterstellte Struktur der Skalen (s. obige Grafik) geprüft. In einem zweiten Schritt werden die Kommentare der Befragten zusammengefasst.

2. Eigenschaften des Fragebogens

2.1 Itemanalyse

Die Itemanalyse wurde nach der sog. Klassischen Testtheorie durchgeführt, wie sie im Programm SPSS in der Prozedur 'Reliability' implementiert ist. Die folgende Tabelle gibt eine Übersicht über die Ergebnisse getrennt nach den drei befragten Gruppen. Angegeben ist jeweils der Reliabilitätskoeffizient (Chronbach's Alpha) und die Zahl der Fragebögen, die in dessen Berechnung eingingen (in Klammern).

Skala	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)

Auffällig ist zunächst die große Differenz zwischen den jeweiligen Teilnehmern an der Befragung (z.B. 120 Personen auf der Seite der Anbieter) und der Zahl der Fragebögen, die in der Auswertung Berücksichtigung fanden (zwischen 47 und 68). Das liegt zum einen daran, dass einige Teilnehmer die Fragebögen nur äußerst spärlich beantwortet haben und zum anderen daran, dass lediglich Fragebögen in der Auswertung berücksichtigt wurden, die in der betreffenden Skala lediglich einen fehlenden Wert aufwiesen. Dieser fehlende Wert wurde durch den Mittelwert der anderen Items einer Skala ersetzt. Per Konvention kann ein Chronbach-Alpha Wert von 0,6 als (gerade noch) zufriedenstellend angesehen werden, ein Wert über 0,8 als gut bis sehr gut. Die Übersicht zeigt, dass sie Skalen, wie sie von den Ausbildern und den Lernenden beantwortet worden waren, durchweg als gut bis sehr gut angesehen werden können. Auf der Seite der Anbieter ist die Skala 'Assessment of learning success' nicht zufriedenstellend. Hier zeigt sich, dass durch das Weglassen des Items 'Where learning/training is non-accredited procedures are in place to recognise and record progress and achievement' der Koeffizient auf 0,601 - also dem unteren Schwellenwert - verbessert werden könnte. Angesichts des Inhalts des Items erscheint der Befund plausibel. Bei der Skala 'Information Management' ergibt sich durch das Weglassen des Items 'Are partners engaged and consulted about the required educational provision and course content e.g Employers/Sector Skills Councils?' ein

Koeffizient von 0,678. Da die Skala 'Market orientation' lediglich aus zwei Items besteht, kann eine Verbesserung des Reliabilitätskoeffizienten durch Weglassen eines der beiden Items nicht erzielt werden. Inhaltlich erscheinen allerdings die beiden Items wichtig, so dass die Skala (gerade noch) im Fragebogen verbleiben kann.

Die Faktorenanalysen wurden nicht mit über sämtliche Items des Fragebogens durchgeführt, sondern lediglich über die Skalen bzw. Items einer der vier Elemente des Deming-Kreises. Diese Vorgehensweise liegt vor allem in der für diese Analyse doch geringe Stichprobengröße begründet. Insgesamt konnten jeweils zwischen 60 und 80 Prozent der Itemvarianz durch Faktoren reproduziert werden - je nach Deming-Element und Zielgruppe, allerdings wurden die Items aufgrund der Analyse nicht immer denselben Faktoren zugeordnet, wie dies aus theoretischer Sicht plausibel erschien. Es gab dabei Unterschiede zwischen den Zielgruppen. So z.B. geht die Faktorenanalyse bezüglich des Elements 'Evaluation' bei der Analyse der Anbieter von vier Faktoren aus. Dabei werden die ersten beiden ('Required qualifications and experience of tutors/trainers' und 'Professional development of the tutor/trainer') reproduziert, der dritte ('Evaluation of learning opportunity') wird in zwei Faktoren aufgespalten, wovon der erste auf die Einhaltung von Qualitätsstandards bezieht, der zweite auf Feedback an Trainer. Die Auswertung derselben Fragen anhand der Antworten der Ausbilder ergibt zwei Faktoren, was allerdings vor allem der Tatsache geschuldet sein dürfte, dass die erste Skala hier lediglich aus zwei Items besteht.

Die Ergebnisse bezüglich des Deming-Elements 'Planung' sind etwas differenzierter zu betrachten, gehen aber ebenfalls in eine plausible Richtung. Aus vorwiegend pragmatischen Gründen möchten wir an der theoretisch getroffenen Einteilung festhalten, da sich die Reliabilität der Skalen spätestens durch das Weglassen eines Items auf einem zumindest zufriedenstellenden Niveau sicherstellen lässt. Dabei muss auch berücksichtigt werden, dass die Absicht des Instruments nicht darin besteht, Qualitätsmerkmale zu testen, sondern dass es in explorativer Absicht vor allem darum geht, Handlungsbedarfe für Qualitätssicherung zu ermitteln. Daher müssen die Qualitätskriterien an das Instrument nicht so streng gehandhabt werden wie bei Tests.

2.1 Bewertung

Die Kommentare der Teilnehmer bezüglich des Fragebogens sind für die Weiterentwicklung des Instruments sehr aufschlussreich. Zunächst eine Übersicht über die skalierten Rückmeldungen (zu berücksichtigen ist hierbei, dass in der Test-Version nicht bei allen Personen, die einen Fragebogen ausgefüllt haben, entsprechende Erhebungen gemacht wurden):

	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

Es zeigt sich, dass das Instrument im Wesentlichen gut bewertet wurde. Abstriche sind zu machen bezüglich der Klarheit der Fragen und der Relevanz und Anwendbarkeit für die Zielgruppen. Betrachtet man die Antworten näher - insbesondere bei denjenigen Fragebögen, die durchgängig schlechte Bewertungen abgaben, so wird deutlich, dass die Abwertungen vor allem mit der Sprache zusammenhängen; kritisiert wird namentlich ein jargon-artiger Sprachstil, der zu kompliziert sei und sich teilweise einer schlechten Ausdrucksweise bediene. Die Kritik bezieht sich vor allem auf die englische Version des Fragebogens und ist zunächst überraschend, da sämtliche Versionen des Fragebogens von Mitarbeitern aus dem jeweiligen Land kontrolliert und verabschiedet wurden und weil in den Arbeitstreffen gerade darauf geachtet wurde, eine verständliche und konkrete Sprache zu benutzen bzw. nach beobachtbaren Sachverhalten zu fragen. Eine nähere Betrachtung der Antwortbögen zeigt, dass die kritischen Stimmen vor allem in denjenigen Bögen vorkamen, die in frühen Testdurchgängen des Instruments erhoben wurden. In späteren Erhebungen dagegen gab es diese Kritik nicht mehr – im Gegenteil – in einigen Bögen wurde die Anwenderfreundlichkeit (easy to use) betont. Da diese Bögen aufgrund dessen, dass durch den Ausfall des irischen Partners dort diese Erhebungen erst sehr spät durch andere Personen durchgeführt werden konnten, liegt die Interpretation näher, dass die anfänglichen Übersetzungsprobleme ausgeräumt werden konnten. Allerdings kann aufgrund dieses Umstands keine länderspezifische Auswertung vorgenommen werden, da zeitbezogene Aspekte mit länderspezifischen konfundiert sind.

Neben Kritik am Instrument wurde in den Feedbackbögen auch noch nach dem vermuteten Nutzen gefragt. Hier waren die Antworten sehr unterschiedlich. Zum Teil wurden keine Antworten gegeben oder nur mit ja oder nein geantwortet, zum Teil aber wurde auch sehr viel geschrieben. Letztlich kann hier nur über die ausführlichen Antworten berichtet werden, die allerdings die Mehrzahl ausmachte. Etwa die Hälfte der Befragten (nicht der Antwortenden!) fühlte sich durch den Fragebogen ermutigt, sich stärker an Maßnahmen zur Qualitätssicherung zu beteiligen, weil sie dadurch z.B. dazu angeregt wurden, über Qualität nachzudenken, weil sie den Wert von Feedback zu schätzen gelernt haben oder weil sie auf die Möglichkeit aufmerksam gemacht wurden, mehrere Instrumente einzusetzen, deren Ergebnisse zu vergleichen und so zu valideren Aussagen über Qualität zu gelangen. Ebenfalls die Mehrheit der Befragten sah das Instrument als hilfreich für ihre Organisation an. Dabei waren die Gründe vielfältig: Um die Mitarbeiter aktiv in die Reflektion um Qualität einzubeziehen, um Modelle guter Praxis hervorzuheben, Veränderungen zu reflektieren oder wichtige Daten zu sammeln. Am häufigsten wurde auf die Betonung vorbildlicher Praxisbeispiele abgehoben. Eher ernüchternd waren die Antworten zur dritten Frage, in der es darum ging einzuschätzen, wie die Mitarbeiter über Benchmarking denken. Hier reichten die Antworten von sehr positiven Einschätzungen (Gelegenheit zur Mitbestimmung) über ‚nicht jeder dürfte sich darüber freuen‘ bis zu eher ablehnend oder Antworten, die auf verbale Zustimmung, aber wenig Engagement tippten. Dem entsprechend unterschiedlich waren auch die Antworten zur vermuteten Unterstützung. Sie gingen von einer vermuteten hohen Zustimmung bis zu einer hohen Ablehnung.

Hinweise zur Weiterentwicklung des Instruments konnten abschließend gegeben werden. Über die bereits erwähnten Hinweise zur sprachlichen Überarbeitung hinaus gab es Lob („all aspects of evaluation are covered“, „it’s easy to use“), Kritik (an der Bedienung wie z.B. daran, dass nicht deutlich genug angezeigt wird, wann man fertig ist mit der Beantwortung der Fragen) und inhaltliche Hinweise. Gewünscht wurde hier z.B. die Implementierung eines Soll-Ist-Vergleichs oder die Möglichkeit der Priorisierung qualitätssichernder Maßnahmen. Auch wurde kritisch darauf hingewiesen, dass keine Auswertung der Daten möglich ist. Dies wurde erst zum Ende der Projektlaufzeit implementiert, wie im folgenden Abschnitt beschrieben.

3. Analysemöglichkeiten

Die Auswertungsmöglichkeiten orientieren sich an Vergleichsmöglichkeiten von Benchmarkings allgemein und der Anlage des Instruments speziell. Folgende Möglichkeiten sind vorgesehen:

- 1 Auswertung anhand von Veränderungen(sog. Zeitreihen-Analyse)
- 2 Auswertung anhand einer Referenzgruppe
- 3 Auswertung anhand von Korrespondenzen (=Vergleiche des generellen Antwortverhaltens)zwischen Anbietern, Trainern/Ausbildern bzw. Lernenden
- 4 Auswertung anhand einer Norm(sog. Benchmarking-Analyse)

Sämtliche vier Auswertungsmöglichkeiten wurden innerhalb der Web-Site implementiert und es existieren zwei Dokumente, die zum einen die Auswertungsprozeduren von ihrer Zielrichtung her erläutern und zum anderen beschreiben, wie man zu den jeweiligen Ergebnisübersichten kommen kann. Da diese Dokumente über die Web-Site erhältlich sind, muss hier auf die inhaltlichen bzw. formalen Auswertungsmöglichkeiten nicht weiter eingegangen werden.